AoF
Academy of Finance

Journal of Finance Issues

Special Issue: Artificial Intelligence in Finance

Volume 23 **Issue 2** **2025**

Letter From the Guest Editor

Hoje Jo

Original Articles

The Role of AI in Fraud Detection: Are financial institutions using the most effective systems?
Hoje Jo Hien Bui Damon Moreland

Strategic Reinsurance and Explainable AI
Sampan Nettayanun Eric R. Brisker

The Effect of AI on CSR and ESG Ethics
Hoje Jo

AI Mistakes in the Classroom
Jaime E. Peters Tara L. Gerstner

Table of Contents

i **Letter From the Guest Editor**
Hoje Jo

Original Articles

1 **The Role of AI in Fraud Detection: Are financial institutions using the most effective systems?**
Hoje Jo, Hien Bui, and Damon Moreland

32 **Strategic Reinsurance and Explainable AI**
Sampan Nettayanun and Eric R. Brisker

57 **The Effect of AI on CSR and ESG Ethics**
Hoje Jo

79 **AI Mistakes in the Classroom**
Jaime E. Peters and Tara L. Gerstner

To the Members of the Academy of Finance and Readers of the Special Issue on AI in finance in the Journal of Finance Issues,

Welcome to the special issue on Artificial Intelligence in finance of the *Journal of Finance Issues* (Volume 23, Number 2). We are excited to present a collection of timely and impactful research that addresses various facets of the impact of artificial intelligence (AI) on finance issues. Our commitment remains to publish high-quality work that advances both academic understanding and practical application in the finance discipline.

This issue features four insightful papers:

First, **"The Role of AI in Fraud Detection: Are financial institutions using the most effective systems?" by Hoje Jo, Hien Bui & Damon Moreland,** examines the implementation of AI in fraud detection and prevention. While AI enhances fraud-fighting capabilities and offers significant cost savings, it also presents challenges—such as model interpretability, ethical concerns, and regulatory compliance. A flaw in these systems can result in severe penalties, underscoring the need for human oversight. Compliance officers, fraud analysts, and auditors play a crucial role in reviewing flagged anomalies, validating AI decisions, and handling complex or ambiguous cases. The paper stresses the importance of integrating AI with human oversight to ensure transparent, effective, and compliant fraud prevention within the U.S. financial system.

Second, **"Strategic Reinsurance and Explainable AI," by Sampan Nettayanun and Eric R. Brisker**, empirically investigates the strategic factors influencing reinsurance purchase decisions in the property and casualty (P&C) insurance industry using the Shapley Additive exPlanations (SHAP) framework—an explainable AI (XAI) tool. Key determinants, including financial metrics, competitive dynamics, and industry demand, are analyzed to assess their impact on varying levels of reinsurance ceding. The SHAP analysis ranks these factors by their relative influence, uncovering both straightforward and complex relationships between determinant values and ceding behavior. For example, increased underwriting in a specific product line may reduce the incentive to hedge further within that line. The study also incorporates a machine learning–based significance test to evaluate the impact of each determinant on reinsurance decisions, offering a robust and interpretable framework for understanding insurer behavior.

Third, **"The Effect of AI on CSR and ESG Ethics," by Hoje Jo**, examines how the integration of Artificial Intelligence (AI) into business operations is transforming industries and improving efficiency, while simultaneously introducing complex ethical challenges—particularly in the context of Environmental, Social, and Governance (ESG) principles. As AI adoption accelerates, businesses must carefully balance technological innovation with ethical responsibility. The paper examines how AI can contribute to sustainability, social equity, and governance improvements, while also addressing potential risks, including algorithmic bias, data privacy concerns, and the environmental impact of AI infrastructure. It emphasizes the need for ethical AI design, transparent governance, and adherence to ESG standards. Ultimately, the study advocates for robust frameworks to ensure AI contributes positively to societal well-being and aligns with the core values of corporate responsibility and sustainable development.

Fourth, the last, but not the least, **"AI Mistakes in the Classroom," by Jaime E. Peters and Tara L. Gerstner,** offers a unique perspective on AI's role in education by examining cases where its integration into academic assignments has failed to meet expectations. The paper discusses how these shortcomings conflict with the principles of Connectivism, highlighting the importance of providing clear instructional guidance, ensuring equitable access to technology, and providing adequate training for students. It highlights the crucial role educators play in enabling students to engage effectively and meaningfully with AI tools in academic contexts.

I extend my sincere gratitude to all the anonymous reviewers whose rigorous and insightful feedback is indispensable to maintaining the quality of our publications. Their dedication ensures the scholarly integrity and relevance of each article.

I would also like to express my most profound appreciation to our main editor, Olgun, and Associate Editors, Larry, David, and Won. Their tireless efforts, expertise, and commitment to the peer review process are crucial to the journal's smooth and effective operation. Their contributions are invaluable in bringing high-quality AI research to our readership.

We hope you find this special issue of the AI usage in finance informative and thought-provoking.

Sincerely,

Hoje Jo
Guest Editor, Special issue of the AI in Finance, Journal of Finance Issues,
Gerald and Bonita Wilkinson Professor of Finance, Santa Clara University,
hjo@scu.edu

The Role of AI in Fraud Detection: Are financial institutions using the most effective systems?

Hoje Jo*, Hien Bui†, and Damon Moreland‡

Abstract

This paper explores the use of AI in fraud detection and prevention, highlighting both its advantages and limitations. While AI strengthens fraud-fighting capabilities and delivers substantial cost savings, it also raises challenges related to model interpretability, ethical considerations, and regulatory compliance. System flaws can lead to severe penalties, reinforcing the need for ongoing human oversight. In this context, compliance officers, fraud analysts, and auditors remain essential for reviewing flagged anomalies, validating AI-driven decisions, and addressing complex or ambiguous cases. The study emphasizes that effective fraud prevention in the U.S. financial system requires a balanced integration of AI technologies with human judgment to ensure transparency, accountability, and compliance.

JEL CLASSIFICATION: G17, G18, K42, C53

KEYWORDS: Artificial Intelligence (AI), Fraud Detection, Fraud Prevention, Real-time Processing, Financial Institutions

I. Introduction

In this paper, the term "financial institutions" refers to entities engaged in financial and monetary transactions, such as deposits, loans, payments, and investments (Hayes et al., 2024; FinCEN, 2025). This includes traditional banks, which provide deposit-taking, lending, and investment services; credit card networks, which facilitate payment processing and authorization; and fintech companies, which deliver innovative services such as mobile banking, peer-to-peer lending, and algorithmic payments. While retailers are not classified as financial institutions under U.S. regulatory definitions, they increasingly operate as financial partners through store credit cards, installment plans, and digital wallets. Each entity contributes to the broader financial ecosystem and faces unique challenges in fraud detection and compliance. Although retailers are not formally classified as financial institutions, their role in fraud prevention has grown as point-of-sale systems now integrate real-time monitoring, encryption, and AI-based detection technologies previously reserved for banking systems (Core Payment Solutions, 2024; Georgiev, 2024). As retail transactions often initiate the fraud detection chain, their systems increasingly function as the first line of defense, particularly in card-present and card-not-present fraud scenarios. In some cases, high-volume retailers may also be subject to recordkeeping and monitoring obligations aligned with Anti-Money Laundering (AML) frameworks (Federal Deposit Insurance Corporation, 2004).

Throughout this paper, we will examine the various AI trends and advancements incorporated into financial institutions' fraud detection systems. We will also compare different AI fraud detection methodologies and discuss the ethical considerations of utilizing artificial

* Santa Clara University, hjo@scu.edu
† Santa Clara University, hbui3@scu.edu
‡ Santa Clara University, dmoreland@scu.edu

intelligence for fraud detection in financial institutions. The purpose of this paper is to outline the future of AI in combating financial fraud, suggesting the next steps for AI utilization in fraud detection systems.

The Impact of AI on Frauds and Scams

Financial fraud is an increasingly sophisticated and costly crime that often surpasses the capabilities of financial institutions and law enforcement agencies to mitigate existential threats effectively. Rapid advancements in technology have enabled nefarious criminal elements and rogue nation-states to commit crimes that are increasingly difficult to detect. As the financial market evolves with technological advancements, new financial crimes emerge alongside existing threats. For instance, financial scams, such as fake checks, ransomware, and cryptocurrency scams, have exposed new challenges that financial institutions must overcome with improved countermeasures (Reeder, 2025). Ransomware is emphasized as a significant menace, as the crime involves holding devices hostage until a ransom is paid (Reeder, 2025). Financial criminals can now inflict damage not only on consumers but also on financial institutions. Traditional fraud detection techniques, which often rely upon rule-based approaches, have been rendered useless in keeping pace with the evolving nature of financial crime. However, even with such threats in financial markets worldwide today, the rise of AI may have led to more straightforward, simpler, and more effective methods for detecting financial fraud. Unlike outdated methods used before the development of such technology, AI enables the analysis of large datasets and databases, the detection of past and present patterns, and the highlighting of anomalies in real-time. Machine Learning (ML) algorithms can now offer better protection and prevention for financial institutions against financial fraud threats. These same systems also adapt to new methods of circumventing the law, continually keeping pace with this evolving and increasingly popular threat (Kamuangu, 2024). As a result, financial institutions and government agencies have adopted advanced technologies to detect and prevent fraudulent activities. Among these new technologies, artificial intelligence (AI) has emerged as a powerful and influential tool in the fight against financial fraud. The functionality of enhancing the accuracy and efficiency of fraud detection systems offers practical solutions that counter financial criminals.

AI has played a pivotal role in modern fraud detection systems by automating threat analysis, thereby enhancing fraud prevention capabilities and reducing the need for manual intervention. Financial institutions have increasingly adopted AI technologies such as Machine Learning (ML), Deep Learning (DL), Graph Neural Networks (GNNs), and Large Language Models (LLMs) to identify and mitigate various forms of financial fraud (Flinders et al., 2025; Valleskey, 2024). These systems are deployed to combat identity theft, phishing scams, payment fraud, credit card fraud, cybercrime, and money laundering — all of which represent growing risks in digital banking environments (Butler, 2024; Flinders et al., 2025). Machine Learning (ML) algorithms are adept at analyzing historical transaction data to uncover suspicious patterns, while Deep Learning (DL) techniques detect intricate fraud schemes through their ability to learn and adapt to evolving behaviors. Graph Neural Networks (GNNs), in particular, are effective at revealing hidden relationships between entities and are valuable for analyzing complex transaction networks to expose layered fraud activity. Additionally, Large Language Models (LLMs) and Natural Language Processing (NLP) tools help extract insights from textual data, such as complaint reports or support chat transcripts, enabling banks to flag fraud signals even from unstructured sources (Valleskey, 2024; Butler, 2024).

The Improvement of Fraud Prevention and Detection Systems

Artificial intelligence (AI) offers functions and capabilities for real-time data analysis and predictive analytics, thereby enhancing automation and anomaly detection. These characteristics improve fraud prevention and detection systems from their traditional approaches. Traditional fraud detection models rely on predefined parameters, resulting in a lack of real-time and predictive analysis. Consequently, it often results in high false-positive rates and the overlooking of emerging fraud tactics. In fraud detection, real-time data analysis facilitates AI-powered systems to pinpoint and flag suspicious patterns (Valleskey, 2024). Through continuous and real-time transaction analysis, AI systems can provide immediate and direct insights into potentially fraudulent transactions. This primary function is crucial in detecting fraud promptly and preventing financial losses (Valleskey, 2024). Continuous analysis of transactions enables the prompt identification of any suspicious movements, thereby significantly enhancing the security of financial systems. Likewise, predictive analytics provides support in risk mitigation and oversight by predicting probable fraud threats through the analysis of historical and current data. Predictive models utilize historical data to identify patterns of fraudulent behavior proactively. It enhances the anticipation and certainty of possible fraud risks for detection systems. It employs techniques such as logistic regression and neural networks to make intelligent predictions about potential fraudulent activities (Valleskey, 2024). For example, AI fraud analysts can strengthen the accuracy of business risk assessments to identify high-risk customers or transactions (Valleskey, 2024). It can help fraud analysts by defining patterns and comprehending the data. This technological implementation enables financial institutions and fintech firms to employ proactive approaches, taking preventive strategies before fraud occurs, thereby reducing potential losses.

Despite these advantages, the effectiveness of AI depends on the quality of the data, model interpretability, and its integration with existing fraud detection frameworks. Human oversight remains necessary in the fraud prevention process, particularly when current technologies are not reliable against emerging threats. It highlights the need to integrate advanced technology with human supervision to combat AI-generated fraud effectively. In this context, human oversight refers to the complement of fraud investigators, compliance analysts, and internal auditors who review AI-generated alerts, make final decisions in ambiguous cases, and provide feedback loops to improve algorithmic performance. The implications of human support are required to counter new forms of AI-driven fraud, including deepfake technology. For instance, deepfake technology is a relatively new technological feature that poses new challenges to the current verification methods. Deepfake technology poses significant risks, allowing fraudsters to circumvent identity controls (Steinhaeuser, 2024). Fraudsters exploit this technology to facilitate the bypassing of KYC (Know Your Customer) identity controls used in financial institutions (Steinhaeuser, 2024). It advocates for the integration of multimodal verification methods and upgrades of detection functionalities to counter these evolving threats. The dependence solely on technology is insufficient and inadequate in the fight against AI-driven fraud (Steinhaeuser, 2024). Financial institutions, such as banks, should implement live verification steps to prevent fraudsters during the account opening process. Therefore, technology alone is insufficient in combating AI-driven fraud; thus, human oversight is crucial.

In addition to ensuring operational effectiveness, human oversight plays a strategic role in maintaining accountability, fairness, and ethical compliance in AI-driven fraud detection systems.

Oversight is not merely reactive, but a critical mechanism for aligning AI decision-making with institutional values and regulatory mandates (Lumenova AI, 2024). Human reviewers can identify model drift, intervene in ambiguous cases, and interpret anomalies in ways that automated systems cannot (Rodgers et al., 2023). Furthermore, oversight is essential in enforcing transparency and explainability, both of which are increasingly demanded by financial regulators and stakeholders (Cornerstone, 2025). Without this layer of governance, institutions risk deploying black-box models that may be technically sound but ethically or legally flawed. As financial institutions adopt more advanced AI systems, embedding structured oversight becomes not just a safeguard but a strategic necessity.

Advanced Applications of AI in Fraud Detection: CNNs and Blockchain Integration

Recent advancements in AI have introduced novel methodologies for fraud detection, particularly through the integration of Convolutional Neural Networks (CNNs) and blockchain technology. These technologies offer complementary strengths that enhance the robustness and transparency of fraud detection systems. CNNs, traditionally used in image and pattern recognition, are increasingly being applied to financial fraud detection, especially in analyzing smart contracts and transactional documents. According to Louati et al. (2024), CNNs can be trained on datasets containing both legitimate and fraudulent smart contracts to identify subtle patterns indicative of fraud. This approach enables the detection of anomalies in both textual and transactional data, offering a powerful tool for legal and financial compliance. CNNs also show promise in document verification, where they can analyze scanned documents or digital forms for signs of tampering or forgery. Their ability to process high-dimensional data makes them suitable for identifying complex fraud schemes that traditional models might overlook.

Blockchain technology, with its decentralized and tamper-resistant ledger, provides a secure foundation for storing and verifying transaction data. When combined with AI, particularly machine learning and deep learning models, blockchain can significantly enhance fraud detection capabilities. Ketha and Provodnikova (2024) propose a framework where AI algorithms analyze blockchain transaction patterns to detect anomalies in real-time. This integration ensures that once fraudulent behavior is detected, the associated data cannot be altered, thereby preserving the integrity of the evidence. Moreover, blockchain can support smart contract auditing, where AI models continuously monitor contract execution for deviations from expected behavior. This is particularly useful in decentralized finance (DeFi) platforms, where traditional oversight mechanisms are limited.

AI Implications of Fraud Prevention and Detection Systems in Different Industries

One of the benefits of AI implications is the decrease in credit card fraud. The rise of alternative payment methods, including digital wallets and P2P (peer-to-peer) payments, reflects varying consumer preferences and showcases the industry's adaptation to consumers' demands. Artificial intelligence (AI) is considered a favorable resolution to fight A2A (Account-to-Account) fraud risks. The primary risk in account-to-account transactions is social engineering scams, with 65% of respondents being optimistic about AI's effectiveness (MasterCard, 2024). Visa and Mastercard utilize Machine Learning (ML) models to examine real-time transaction patterns, which significantly lessens the incidence of fraudulent activities (Fitzpatrick, 2024). As elaborated, Visa utilizes advanced analytics to detect and note suspicious transactions, resulting in enhanced

security for cardholders. A survey indicates that 63% of respondents prioritize advanced fraud detection as a driver for AI investment (MasterCard, 2024). By cultivating the functionalities of Machine Learning (ML), these systems can analyze vast amounts of transaction data. AI helps identify and prevent fraudulent activities before they escalate. 49% of financial institutions and fintech firms have already integrated AI, while 93% plan to invest and implant AI within the next 2-5 years (MasterCard, 2024). These statistics further underscore the growing confidence in digital transactions and overall financial security. It indicates a notable shift in the perception of technology for fraud prevention and detection systems. The convergence of AI with traditional systems streamlines workflows in detecting and preventing fraud. Across the board, AI's proficiency in adapting to new fraud movements guarantees that detection systems remain effective against emerging threats within financial institutions.

As a result of advanced AI implementation, fraud detection systems in retail have become more sophisticated, ensuring a secure environment for all stakeholders, including businesses and customers. There is a growing collaborative dynamic between retailers and financial institutions, in which data and feedback are shared to continuously improve fraud prevention efforts across the transaction chain. In a typical retail transaction, fraud detection begins at the point of sale, where retailers use AI-powered POS systems equipped with encryption, tokenization, and real-time screening tools to flag suspicious activity (Core Payment Solutions, 2024). These transactions are then routed through credit card networks such as Visa or Mastercard, which apply their own fraud scoring algorithms based on spending patterns and geolocation. Finally, issuing banks, which fund the transactions, apply backend AI models to detect anomalies, evaluate risk, and flag chargebacks or fraudulent claims. At each stage, distinct AI systems operate independently but contribute to shared fraud intelligence platforms, facilitating collaborative risk reduction across the ecosystem (Georgiev, 2024). In some cases, retailers also face compliance expectations related to fraud prevention and may be required to maintain transaction records or adhere to data standards, particularly when working with regulated financial intermediaries (Federal Deposit Insurance Corporation, 2004).

Typical retail fraud types are credit card fraud and account takeover (Pavion, 2024). These threats necessitate the need for advanced detection systems. Retail businesses utilize AI to implement sophisticated algorithms that pinpoint and mitigate fraudulent activities in real-time (Pavion, 2024). AI plays a significant role in protecting assets and customers from fraudulent activities for retail businesses. Financial frauds cause immediate financial losses and additional costs for investigation and recuperation. Customers' faith and belief are adversely impacted by fraud, which subsequently affects potential long-term business growth results. AI systems utilize predictive analytics to forecast potential fraudulent activities, providing retailers with real-time alerts (Pavion, 2024). Automation of fraud investigation approaches through AI provides valuable insights for informed decision-making (Pavion, 2024). The role of AI is pivotal in ensuring a secure retail environment amidst growing challenges. It further improves retail businesses' ability to safeguard assets and customer trust. The integration of AI and predictive analytics lets retailers forecast and prevent fraudulent activities effectively (Pavion, 2024). Overall, the continuing advancement of technology will further empower AI in combating fraud in retail environments. These interconnected systems underscore the importance of continuous feedback and cooperation across the transaction chain, enabling AI models to evolve in response to emerging fraud patterns. It highlights how collaboration between retailers, credit card networks, and financial institutions is essential to building resilient and adaptive AI fraud detection systems.

Regulatory and Compliance Implications of AI Systems

As AI can enhance fraud detection, it unfortunately can enable sophisticated scams and frauds (West and Ciais, 2023). Cooperation across financial sectors is integral to developing safeguards against AI-enabled fraud. The need for protective and defensive measures against AI misuse must be emphasized. AI presents significant risks and opportunities in tackling financial fraud and scams. The challenge necessitates protective standards and benchmarks that do not hinder innovation. AI implications can enhance fraud detection, enabling banks to identify fraud more effectively while also facilitating the detection of unlawful activities (West and Ciais, 2023). Collaboration across sectors is crucial for achieving long-term reductions in fraud and scams. As AI implementation also has adverse impacts, it presents new challenges for financial institutions, such as ethical considerations and regulatory compliance, which must be addressed to ensure the effective and responsible use of AI technologies. Organizations with ineffective fraud prevention and detection systems are vulnerable to severe financial obligations and penalties imposed by various regulations and agencies.

One of the regulations that financial institutions typically encounter is the Anti-Money Laundering (AML) regulations. Financial institutions must adhere to AML (Anti-Money Laundering) regulations to ensure compliance with international frameworks for preventing financial crime (Crane and Kimbrell, 2025). AI-driven fraud detection systems are increasingly being incorporated into AML compliance programs. The purpose of the integration is to strengthen transaction monitoring, recognize suspicious movement, and enhance reporting efficiency (Crane and Kimbrell, 2025). AI algorithms can detect complex money laundering patterns that traditional rule-based systems might neglect, which reduces false positives while improving accuracy (Crane and Kimbrell, 2025). Nevertheless, significant challenges remain in aligning AI fraud detection models with AML regulations. Financial institutions faced nearly $5 billion in fines for AML and KYC failures in 2022 (Levitt, 2024). Financial institutions must further utilize Deep Learning (DL) and Natural Language Processing (NLP) to enhance identity verification for Know Your Customer (KYC) and Anti-Money Laundering (AML) compliance (Levitt, 2024). The regulation requires financial institutions to demonstrate transparency and accountability in their anti-money laundering (AML) compliance strategies. AI models must ensure explainability and fairness in decision-making processes to meet compliance requirements (Crane and Kimbrell, 2025). Furthermore, cooperation among financial institutions, regulatory agencies, and AI developers is necessary to establish standardized AI governance frameworks that strike a balance between technological advancements and compliance requirements.

While this section references select foreign regulatory bodies, the primary regulatory framework examined throughout this paper is that of the United States. The inclusion of international examples, such as the European Union or Singapore, serves only to provide comparative context and highlight alternative approaches to AI governance. Many AML violations stem from inadequate monitoring systems. AI-powered fraud detection, when properly implemented, can enhance transaction monitoring and reduce false negatives, thereby helping institutions meet AML compliance standards. As different financial institutions are based in various countries, it is mandatory to comply with the standards and regulations of the governing agencies in each respective country. The U.S. approach to governing AI relies on existing regulatory frameworks rather than designing comprehensive new legislation. The enforcement of AI governance in the United States is handled by established agencies, including the U.S. regulatory agencies (Simpson, 2023). The U.S. government intends to utilize existing regulatory

agencies to oversee the deployment of AI and refine existing regulations. Unlike other regions, the U.S. does not have a comprehensive AI legislation framework in place (Simpson, 2023). The implication of that means that the governance of AI is more fragmented and relies heavily on the interpretation and application of existing laws by these agencies. Each agency is taking specific actions to address AI-related issues. For example, the CFPB has administered guidance demanding that creditors use algorithms to provide transparent reasons for adverse credit decisions, ensuring clarity in automated decision-making (Simpson, 2023). On the contrary, the European Union's Artificial Intelligence Act was published in the EU Official Journal, marking it as the first comprehensive horizontal legal framework for regulating AI systems across the EU, on 12 July 2024 (Hickman et al., 2024). The EU is implementing the EU AI Act through a rigorous process involving multiple stakeholders to ensure seamless integration of the legislation. An AI Board will be appointed to supervise the enforcement of the law across member states, while the details of enforcement at the national level are still under discussion. There is a current debate about whether a centralized regulatory body would be sufficient to guarantee consistent enforcement of AI regulation. The EU approach focuses on ethics, emphasizing compliance, accountability, and the ethical use of AI technologies.

The role of internal audit

As AI becomes more embedded in financial fraud detection systems, the role of internal audit has expanded from retrospective reviews to a strategic function that proactively assesses AI-related risks and controls. Internal auditors are now tasked with evaluating the integrity and transparency of AI models, ensuring that fraud detection systems are not only effective but also compliant with emerging regulatory and ethical standards (CSM & CO LLP, 2025). In high-risk environments, internal audit helps identify weaknesses in model governance, bias detection, data management, and system accountability. Moreover, audit teams have begun leveraging AI tools themselves—using anomaly detection, continuous monitoring, and predictive analytics to simulate fraud scenarios and uncover operational blind spots (Hodge, 2024). This transformation positions internal audit as a vital bridge between compliance, cybersecurity, and business operations. Beyond technical assessments, internal auditors also help cultivate a culture of fraud awareness by advising management, training employees, and validating the effectiveness of human oversight protocols (Petrașcu & Tieanu, 2014). As financial institutions face increasingly complex threats from AI-enabled fraud, the internal audit function will play a central role in reinforcing trust, transparency, and institutional resilience.

As AI systems become more autonomous and complex, internal audit functions are increasingly expected to provide assurance not only over outcomes but also over the processes and algorithms that drive decision-making. This includes validating data inputs, assessing model training practices, and ensuring that AI deployment aligns with institutional policies and regulatory expectations (CSM & CO LLP, 2025). Internal auditors must now work closely with data scientists, compliance officers, and IT security teams to evaluate how fraud detection models are governed, monitored, and updated over time. The rapid pace of AI innovation also introduces the risk of model drift, which internal audit can help detect through independent testing and validation processes. According to Hodge (2024), internal auditors must adopt a forward-looking mindset that anticipates how fraudsters may exploit AI vulnerabilities, including adversarial attacks and synthetic identity generation. Additionally, by integrating ethics-based auditing techniques, internal audit can help ensure AI decisions reflect principles of fairness, accountability, and

transparency—especially as regulators intensify their scrutiny of algorithmic systems (Petrașcu & Tieanu, 2014). This expanded scope repositions internal audit as not just a control mechanism but a strategic partner in sustaining the integrity and trustworthiness of AI-driven fraud prevention.

Thesis Statement

Overall, the incorporation of AI in fraud detection presents several challenges. Issues such as model interpretability, ethical considerations, and regulatory compliance must be addressed to ensure the effective and responsible use of AI technologies. As financial institutions increasingly adopt AI-powered fraud detection systems, their effectiveness varies depending on the models used, such as Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP), as well as compliance with evolving regulatory standards. While AI has significantly improved the accuracy and response speed of fraud detection, challenges such as false positives, algorithmic bias, and regulatory scrutiny raise concerns about whether financial institutions are leveraging the most effective systems for fraud prevention. This paper aims to examine the strengths and limitations of AI fraud detection models and highlight the need for optimized, transparent, and compliant AI-driven solutions. To guide this analysis, the paper proposes a conceptual model that maps how AI systems interact with institutional capabilities, regulatory constraints, and operational outcomes in fraud detection. This model serves as a structural lens for evaluating which AI approaches are most effective, under what conditions, and why some are more widely adopted despite technical trade-offs. The scope of this paper is to study the implications of financial institutions and regulatory frameworks in the United States. References to non-U.S. jurisdictions, such as the European Union, are included for comparative analysis and evaluation purposes. By examining these features, this paper will highlight the need for optimized, transparent, and compliant AI-driven solutions to safeguard financial systems and counter increasingly sophisticated fraud threats. As this paper aims to contribute to and shape the future of AI in addressing financial fraud, the need for optimized and compliant AI-driven solutions will strongly emphasize the next step in AI integration for fraud detection systems.

II. Literature Review

The growing integration of AI-powered fraud prevention and detection systems has led to an increasing number of studies analyzing their effectiveness, benefits, and challenges. Multiple research papers have explored different AI techniques in fraud detection and prevention. Several studies also highlight AI's ability to improve accuracy, enhance real-time monitoring, and reduce false positives. Other studies have examined the challenges associated with data quality, algorithm bias, and regulatory compliance for practical implementation in banking and financial institutions. This section reviews key studies in the field and provides an overall understanding of AI-driven fraud detection.

AI-Driven Fraud Detection in Financial Institutions

Several studies have evaluated the role of AI in fraud detection within the banking sector, analyzing AI and data science techniques for banking fraud detection and highlighting cybersecurity improvements (Olowu et al., 2024). The paper, "AI-Driven Fraud Detection in Banking: A Systematic Review," examines AI and data science techniques for fraud detection in

the banking sector. It highlights the need for advanced detection strategies in response to the growing financial fraud and cybersecurity threats. The study finds that Machine Learning (ML) algorithms achieve fraud detection rates between 87% and 94%, concurrently reducing false positives by 40% to 60% compared to traditional methods (Olowu et al., 2024). The study further recommended the development of explainable AI frameworks to provide transparency and rationale for the decision-making process. The research also suggests hybrid detection systems that combine multiple AI technologies to enhance fraud detection capabilities (Olowu et al., 2024). Additionally, the review highlights the importance of regulatory compliance and legal considerations when implementing AI in fraud detection.

Olowu and Adeleye (2024) conducted a systematic review of AI techniques in banking fraud detection, reporting that ML algorithms achieved detection rates between 87% and 94% while reducing false positives by up to 60%. While these results underscore the efficiency of ML in identifying fraudulent patterns, the study does not address the interpretability of these models, an essential factor for regulatory compliance. Moreover, the exclusive focus on ML overlooks the potential benefits of hybrid systems, as emphasized by Yuhertiana and Amin (2024). Thus, while Olowu and Adeleye's findings support the integration of AI in fraud detection, they also reveal gaps that must be addressed to ensure robust, transparent, and compliant systems.

Furthermore, there are papers and studies on the review of AI methodologies and driven approaches for financial fraud detection (Yuhertiana and Amin, 2024). The paper "Artificial Intelligence Driven Approaches for Financial Fraud Detection: A Systematic Literature Review" presents a systematic literature review on AI methodologies for detecting financial fraud. A systematic literature review was conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is a set of guidelines that help authors report systematic reviews and meta-analyses. As a result, the paper acknowledges the effectiveness of AI in enhancing the precision and efficiency of fraud detection. The financial industry is highlighted as the primary sector for AI applications in fraud detection (Yuhertiana and Amin, 2024). Artificial intelligence plays a pivotal role in identifying anomalies in financial transactions, which is crucial for security. Machine Learning (ML) models and techniques remain the prevailing methodology in financial fraud detection (Yuhertiana and Amin, 2024). The study analyzes and examines 24 papers published between 2014 and 2023 (Yuhertiana and Amin, 2024). The paper concludes that AI dramatically improves the precision and efficiency of fraud pattern identification in the financial industry. It further emphasized that AI simulates human cognitive abilities within a machine framework (Yuhertiana and Amin, 2024). AI's effectiveness mostly depends on acquiring experience and data for optimal performance. AI has the ability to learn from experiential data without explicit human guidance autonomously.

AI-Based Fraud Detection Systems

Furthermore, there are papers that provide a review of various AI-based methods, including machine learning (ML) and deep learning (DL) algorithms, used for detecting credit card fraud (Hafez et al., 2025). One popularly implemented approach is Machine Learning-Based fraud detection, which utilizes both supervised and unsupervised learning methods to identify fraudulent activities. Supervised learning algorithms are employed as the primary tools in financial fraud detection, where historical transaction data serves as a reference (Kamuangu, 2024). Each instance in the dataset is labeled as either fraudulent or non-fraudulent, allowing the system algorithms to

learn from these examples and generalize to new, unseen transactions. Supervised learning models, such as decision trees and logistic regression, are trained on labeled datasets of fraudulent transactions, enabling them to classify future transactions with high accuracy (Kamuangu, 2024). Unsupervised learning techniques, including clustering and anomaly detection algorithms, facilitate the identification of unknown fraud patterns that were not previously present in historical datasets (Hafez et al., 2025). Unsupervised learning is crucial in navigating the complexities of financial data, where fraudulent activities are often rare and hidden within vast datasets. These techniques, such as clustering algorithms like K-Means, are particularly useful at identifying unusual patterns and transactions that deviate from regular patterns (Kamuangu, 2024). The study evaluates and compares these techniques to identify their strengths and weaknesses. The author summarizes the major challenges, including data imbalance and high processing demands, faced by current AI models used in fraud detection. The author even advocates for the implementation of hybrid and ensemble models that combine machine learning (ML), deep learning (DL), and multi-objective heuristic optimization (MHO) techniques, which could enhance detection accuracy and address class imbalance (Hafez et al., 2025).

AI-based fraud detection systems use advanced Machine Learning and Neural Network models to enhance the accuracy and speed of identifying fraudulent transactions. These systems surpass traditional rule-based methods due to their ability to detect evolving fraud patterns. Several papers have examined the benefits of utilizing AI over traditional rule-based systems. The paper by Cheemakurth et al. (2024) explains how neural network models can enhance financial security by enabling the highly accurate and rapid detection of fraudulent transactions in real-time. It is emphasized that AI reduces false positives, improving operational efficiency and minimizing disruptions to legitimate transactions. The study yielded a finding that Neural Network methods outperformed traditional Machine Learning methods, achieving a fraud detection accuracy of 98% (Cheemakurth et al., 2024). The author further explains how Neural Network models surpass traditional methods due to the model's adaptability to evolving fraud patterns and ability to identify new fraudulent activity.

Regulatory Challenges and Ethical Considerations

Despite the benefits that AI brings to fraud detection, the technology faces several challenges, including model interpretability, ethical considerations, and regulatory compliance. Understanding how AI models arrive at their conclusions can be challenging, making it difficult to explain fraud detection decisions. Ethical concerns regarding privacy and potential bias in AI algorithms when analyzing personal data should also be addressed. "The paper by Adhikari et al. (2024) outlines how AI enhances real-time fraud detection, adaptability, and scalability compared to traditional systems. The study also explains how AI systems outperform traditional rule-based systems, particularly in real-time detection and adaptation, enabling the detection of evolving fraud patterns. The study's findings indicated that AI significantly enhances fraud detection accuracy but also presents challenges related to ethical concerns, algorithmic bias, and data privacy (Adhikari et al., 2024). The research highlighted the need for high-quality data to improve AI performance and acknowledged the vulnerabilities of AI systems to adversarial attacks. The research highlights the importance of high-quality data in enhancing AI performance and acknowledges the vulnerabilities of AI systems to adversarial attacks. The study concluded that while AI-based fraud detection outperforms rule-based methods, addressing bias and ensuring data security remain critical challenges (Adhikari et al., 2024). The study concurs with the notion that while AI-based fraud

detection outperforms rule-based methods, addressing bias and ensuring data security remain critical challenges.

Addressing regulatory and legal challenges is vital for the effective deployment of AI fraud detection systems. Since the EU governing bodies have comprehensive AI regulations in place, there is a well-defined range of penalties for defective and ineffective AI systems. For example, the maximum penalty for providing incorrect, incomplete, or misleading information to notified bodies or national competent authorities is EUR 7.5 million or 1 percent of the worldwide annual turnover, whichever is higher (Hickman et al., 2024). Financial institutions need to carefully develop AI systems to ensure their effectiveness when deploying them in EU territories. In the United States, financial institutions are required to comply with regulations established by the Federal Trade Commission (FTC) and the Consumer Financial Protection Bureau (CFPB). It highlights the various approaches to AI regulatory enforcement worldwide. It further emphasizes the importance of developing robust regulatory frameworks that strike a balance between innovation and the protection of fundamental human rights and safety.

The literature increasingly highlights the essential role of human oversight in ensuring ethical and effective deployment of AI technologies, particularly in domains involving high-stakes decision-making such as fraud detection. Petrașcu and Tieanu (2014) argue that internal audit is not merely a control mechanism but a value-adding function that supports leadership in managing fraud risks, offering both preventive and detective insights that are vital in technologically complex environments. As AI-driven systems grow more autonomous and opaque, ethical risks emerge around bias, explainability, and accountability—challenges that cannot be addressed through automation alone. Rodgers et al. (2023) reinforce this view by proposing an ethical decision-making framework—the Throughput Model—that maps how perception, judgment, and choice shape algorithmic decisions. Applied to AI governance, this model underlines the importance of embedding human evaluators, such as internal auditors, into algorithmic systems to oversee input quality, decision logic, and outcome integrity. Together, these perspectives support a growing consensus that internal audit functions are indispensable not just for verifying technical compliance but for upholding organizational integrity and ethical accountability in the age of AI.

III. Evaluating the Effectiveness of AI-Powered Fraud Detection in Financial Institutions

This section provides a qualitative literature synthesis and secondary data analysis approach to evaluate the effectiveness of AI-powered fraud detection systems in financial institutions. This methodology is appropriate for two reasons. First, the deployment of AI in fraud detection is a relatively recent and fast-evolving area, with limited access to standardized empirical datasets. Second, regulatory bodies, industry leaders, and financial institutions have released a growing volume of publicly available reports, case studies, and enforcement actions that offer credible, real-world insights into AI's current capabilities and limitations. By synthesizing these sources, this section aims to map where AI is most effective, assess the performance trade-offs among model types (e.g., Machine Learning, Deep Learning, Natural Language Processing), and evaluate how institutional failures, despite AI adoption, continue to result in regulatory penalties and financial harm. The sources used in this section were selected based on recency (2022–2025), relevance to AI fraud detection, and institutional credibility, such as peer-reviewed studies, regulatory reports, and widely cited industry white papers. This ensures that the comparative evaluations and enforcement case analysis reflect current and authoritative perspectives.

To evaluate the effectiveness of AI-powered fraud detection systems, this section presents a three-part framework: (1) a conceptual model illustrating institutional touchpoints for AI intervention; (2) a comparative analysis of common AI model types used in fraud detection; and (3) real-world evidence from enforcement actions that reveal where failures persist despite AI deployment. Together, these components provide a holistic view of AI's promise, limitations, and institutional fit.

Simple Model for AI's Impact on Fraud Detection in Financial Institutions

Key Variables

- Input Variables (X):
  - Transaction Volume (V): Number of transactions processed per day.
  - Fraud Rate (F): Percentage of fraudulent transactions before AI implementation.
  - AI Detection Accuracy (A): AI's ability to correctly identify fraud (True Positive Rate).
  - AI False Positive Rate (FP): AI incorrectly flags legitimate transactions as fraud.
  - Manual Review Cost (C): Cost per transaction for human review.
  - Fraud Loss per Incident (L): Average loss per undetected fraudulent transaction.
- Output Variables (Y):
  - Reduction in Fraud Loss (ΔL): Savings from AI detecting more fraud.
  - Operational Cost Savings (ΔO): Savings from reduced manual reviews.
  - False Positive Cost (CFP): Additional cost from AI's false alarms.

Model Equations

(A) Fraud Detection Improvement

- Fraud Detected by AI (DF):
  $DF = V \times F \times ADF = V \times F \times A$
- Reduction in Fraud Loss (ΔL):
  $\Delta L = DF \times L\Delta L = DF \times L$

(B) Operational Efficiency

- Reduction in Manual Reviews (ΔR):
  $\Delta R = V \times F \times (1 - Aprior - A)\Delta R = V \times F \times (1 - Aprior - A)$
  (where AprioriAprior is pre-AI detection rate)
- Operational Cost Savings (ΔO):
  $\Delta O = \Delta R \times C\Delta O = \Delta R \times C$

(C) False Positive Cost

- False Positives Flagged (FP_T):
  $FPT = V \times (1 - F) \times FPFPT = V \times (1 - F) \times FP$
- Cost of False Positives (CFP):
  $CFP = FPT \times CCFP = FPT \times C$

Net Impact of AI on Fraud Detection

Net Impact $= \Delta L + \Delta O - CFP$Net Impact $= \Delta L + \Delta O - CFP$

Example Calculation

Assume:

- V = 10,000V = 10,000 transactions/day
- F = 1%F = 1% fraud rate
- A = 95% A = 95% detection accuracy
- FP = 2%FP = 2% false positive rate
- C = $5C = $5 per manual review
- L = $500L = $500 per fraud

Calculations

(A) Fraud Detected by AI:

DF = 10,000×0.01×0.95 = 95 frauds caught = 10,000×0.01×0.95 = 95 frauds caughtΔL = 95×500 = $47,500 savedΔL = 95×500 = $47,500 saved

(B) Operational Savings (assuming prior detection = 80%):

ΔR = 10,000×0.01×(1−0.80−0.95) = N/A (adjust based on actual workflow)ΔR = 10,000×0.01×(1−0.80−0.95) = N/A (adjust based on actual workflow)

(Simpler alternative: AI reduces manual reviews by 50%)

ΔO = 5,000×5 = $25,000 saved ΔO = 5,000×5 = $25,000 saved

(C) False Positive Cost:

FPT = 10,000×0.99×0.02 = 198 false alarmsFPT = 10,000×0.99×0.02 = 198 false alarmsCFP = 198×5 = $990CFP = 198×5 = $990

(D) Net Impact:

$47,500+$25,000−$990 = $71,510 net benefit/day$47,500+$25,000−$990 = $71,510 net benefit/day

Limitations & Considerations

- AI Model Drift: Fraud patterns evolve; AI needs retraining.
- Customer Experience: High false positives may frustrate users.
- Integration Costs: The initial setup and maintenance costs for AI are not included in the price.

To quantify the operational impact of AI on fraud detection, a hypothetical scenario was applied using the conceptual model. With a detection accuracy of 95%, a false positive rate of 2%, and 10,000 daily transactions, the institution would save approximately $47,500/day in fraud losses and $25,000/day in reduced manual reviews, incurring only $990 in false positive costs, yielding a net benefit of $71,510 per day, or nearly $18 million annually. This calculation, while simplified, underscores the substantial cost-saving potential of AI tools under high-performance conditions. These savings assume high model performance with levels that align most closely with advanced supervised machine learning (ML) or deep learning (DL) systems, both of which have been widely adopted across financial institutions. The next section compares these AI models across key operational criteria to evaluate their real-world effectiveness.

This simple model helps quantify the impact of AI on fraud detection in financial institutions. We can adjust variables based on real-world data for accuracy. While the conceptual model maps out where AI tools are deployed across institutions, the next section examines the relative strengths and weaknesses of specific AI model types used at those intervention points. Based on research and reviews, it is apparent that Machine Learning techniques have been extensively used to enhance fraud detection and prevention systems compared to other methods.

Studies have shown that the integration of AI has significantly improved the preventive and defensive strategies of financial institutions. To gain additional insights, this paper aims to explore the adoption of AI techniques and models within fraud detection and prevention systems. The study will utilize available data and statistics to validate the current understanding of AI implementation and effectiveness in financial institutions. One of the objectives is to identify the most effective AI models and techniques. The purpose of this study is to suggest a suitable AI model for financial institutions to implement, ensuring compliance with regulations and legislation. Subsequently, it can provide a material and significant net benefit to the fraud detection system.

Comparative Data and Performance Metrics

Performance Across Different AI Models

The BioCatch 2024 AI Fraud Financial Crime Survey provides critical empirical support for the findings presented in this paper. As one of the most comprehensive industry surveys on AI adoption in fraud detection, it offers valuable context for interpreting the performance metrics of AI models discussed earlier.

According to the survey, 94% of financial institutions report using AI/ML techniques to assess risk from user behavior, and 87% state that AI has improved the speed of fraud detection. These figures reinforce the high adoption rate of Machine Learning (ML) models (Table 2) and validate their perceived effectiveness in real-world applications. Furthermore, the survey reveals that 73% of firms use AI specifically for fraud detection, underscoring the centrality of AI in modern fraud prevention strategies.

Importantly, the BioCatch survey also highlights the evolving threat landscape. For example, 91% of banks reconsidered voice verification methods due to the rise of AI-enabled fraud, such as deepfake audio attacks. This finding supports the paper's broader argument that AI systems must be continuously updated and supplemented with human oversight and explainable AI frameworks to remain effective.

Regulatory & Compliance Challenges for AI Fraud Detection

Despite the advantages of AI-powered fraud detection systems, financial institutions constantly face regulatory scrutiny and compliance challenges. Most compliance shortcomings and violations are related to AML compliance. While the U.S. has implemented robust AML regulations through agencies such as FinCEN (Financial Crimes Enforcement Network) and the OCC (Office of the Comptroller of the Currency), similar regulatory frameworks in the European Union, such as the 6th AML Directive, offer useful points of comparison, particularly in areas like data sharing and enforcement structure. Although the primary focus remains on U.S. regulatory practices and institutional frameworks, brief references to international standards, such as the EU's AML requirements, are included to contextualize broader trends in AI compliance. References to European directives serve to highlight cross-jurisdictional trends in AI-driven fraud prevention. Key concerns include (BioCatch, 2024):

- 51% of financial institutions and fintech firms lost between $5 million and $25 million to AI-driven threats in 2023.
- 91% of banks reconsidered voice verification methods due to AI-enabled fraud risks.

- Fines for Anti-Money Laundering (AML) violations surged by 50% in 2022, totaling nearly $5 billion.
- Notable cases include:
    - 2022:
        - Morgan Stanley was fined $60 million to resolve a data security lawsuit (Stempel, 2022).
        - Global AML Fines (Various Banks) totaled $5.0 billion (BioCatch, 2024).
    - 2023:
        - Deutsche Bank fined $186 million for AML shortcomings (BioCatch, 2024).
        - Binance faced a $4.3 billion penalty for AML violations (BioCatch, 2024).
    - 2024:
        - Toronto-Dominion Bank (TD Bank) faced a $3.0 billion fine for AML violations (Emanuel-Burns, 2024).
        - City National Bank was fined $65 million to resolve risk control allegations (ABA Banking Journal, 2024).
    - 2025:
        - Paypal was fined a $2 million civil fine over cybersecurity failures that led to the exposure of customers' Social Security numbers (Stempel, 2025).
        - OKX agreed to pay penalties of more than $500 million for violating U.S. Anti-Money Laundering Laws (United States Department of Justice, 2025).

Research Methodology

Performance Across Different AI Models

To understand the complete perspective of AI fraud detection's effectiveness, the paper aims to compare and examine the following models and techniques used by financial institutions:

- Machine Learning (ML)-based models
- Deep Learning (DL) techniques
- Natural Language Processing (NLP)

To assess the effectiveness of AI-powered fraud detection systems in financial institutions, the paper will use qualitative metrics for measuring success. The rating will be determined based on a combination of industry reports and academic research. The following key comparison metrics are used to gain more detailed and precise insights:

- False positive rates: aim to measure the frequency of legitimate transactions being flagged as fraud.
    - Low: Less than 1% of legitimate transactions incorrectly flagged; indicates strong precision and minimal manual review burden (BioCatch, 2024; Adhikari et al., 2024).
    - Medium: 1%–3% false positive rate; manageable in large institutions with tiered fraud review protocols.
    - High: Greater than 3% of legitimate transactions flagged as fraud; often results in customer dissatisfaction, increased operational costs, and elevated risk of revenue loss due to false declines (Levitt, 2024; Bengani, 2024).

- Detection speed: aims to determine how quickly AI models identify fraudulent activity.
    - Very Fast: Response time ≤ 100 milliseconds; latency optimized through real-time inference pipelines using high-performance hardware and minimal preprocessing (Cheemakurthi et al., 2023; NVIDIA, 2024).
    - Fast: Response time between 100–500 milliseconds; typical of optimized ML pipelines with efficient data ingestion and model simplicity (Levitt, 2024).
    - Medium: Response time between 500 milliseconds and 1 second; often associated with NLP or multi-layered fraud scoring systems requiring contextual validation.
    - Low: Response time ≥ 1 second; typically seen in models that require external data sourcing, intensive feature extraction, or human-in-the-loop validation prior to flagging a transaction (BioCatch, 2024; Adhikari et al., 2024).
- Industry Effectiveness Rating
    - Derived from a combination of adoption rates, reported detection performance in financial institutions, and model preference trends highlighted in BioCatch (2024) and Bengani (2024). "Very High" implies consistent preference by top-tier banks due to accuracy and integration success; "Medium" indicates moderate adoption or use for niche tasks like NLP-based anomaly descriptions.

The data and examples cited in this section were selected based on three key criteria: (1) recency — sources published between 2022 and 2025 to reflect the most current developments in AI-based fraud detection; (2) credibility — reports and findings issued by reputable institutions such as BioCatch, industry surveys, and major regulatory bodies; and (3) relevance — sources that directly address the technical performance or compliance challenges of AI-powered fraud detection in financial institutions. For the comparative evaluation of AI model performance (e.g., ML, DL, NLP), the analysis draws primarily from BioCatch's 2024 industry report, which provides adoption rates and qualitative performance ratings.

Regulatory & Compliance Challenges

A secondary analysis examines the compliance of AI fraud detection with financial regulations (e.g., GDPR, EU AI Act, AML laws). The purpose of this research is to determine whether financial institutions are effectively implementing their AI-powered fraud detection systems. The extent and materiality of fines and penalties can illustrate the severity of system deficiencies and flaws. Key areas of this secondary analysis include:

- Case studies of legal challenges related to AI-driven fraud detection failures.
- Comparison Metrics:
    - Fines issued for AI compliance failures (monetary penalties imposed on financial institutions)
    - Regulatory Violations (reasons and rationale for fines and penalties)

For the enforcement case analysis, multiple sources were used, including official regulatory filings, news reports (e.g., Reuters, Bloomberg), and government publications (e.g., DOJ press releases), to provide a comprehensive view of where AI deployment has failed to prevent compliance breaches or mitigate risk exposure.

Key Findings and Institutional Implications

The expectations outlined in this section are grounded in both empirical insights and evolving industry standards. Studies have shown that Deep Learning (DL) models outperform traditional Machine Learning (ML) models in identifying complex fraud patterns due to their capacity to process vast and unstructured datasets efficiently (Adhikari et al., 2024; Hafez et al., 2025). However, Machine Learning (ML) remains favored in practical settings for its interpretability, lower computational cost, and ease of integration—critical attributes for regulatory compliance and operational scalability (Cheemakurthi et al., 2023; IBM, 2025). False positives persist as a known limitation in both Deep Learning (DL) and hybrid AI systems, highlighting the need for balanced approaches (Talaat et al., 2025). Furthermore, recent high-profile regulatory fines against financial institutions underscore the importance of explainability and bias mitigation in AI systems (Emanuel-Burns, 2024; Fitzpatrick, 2024). These trends collectively support the prediction that future fraud detection will increasingly rely on hybrid models that blend Machine Learning (ML), Deep Learning (DL), and rule-based systems for optimal performance and compliance.

Performance Across Different AI Models:

Preliminary expectations suggest:

- Deep Learning (DL) outperforms traditional Machine Learning (ML) models, but Machine Learning (ML) models will be more popular due to their lower computational cost, ease of integration, and interpretability.
- Deep Learning (DL) models handle large datasets efficiently, reducing false positives.
- False positives remain a challenge, especially in advanced AI models.
- The result will indicate a need for hybrid AI models (ML + DL + rule-based detection) and will offer optimal fraud detection by integrating multiple detection methodologies.

Regulatory & Compliance Challenges:

Expected regulatory challenges include:

- Increased scrutiny of AI models for fairness and bias mitigation.
- Major financial institutions face penalties for AI fraud detection failures.
- The necessity for explainability in AI decisions to comply with regulatory requirements.

Enforcement Outcomes and Regulatory Failures:

The comparative analysis that follows will further assess and evaluate the hypothesized impact of the conceptual model introduced earlier, which positions institutional decision-making as a balance between technical model performance, operational feasibility, and compliance risk. The following tables summarize key findings across AI model performance and regulatory outcomes based on the reviewed sources.

Performance Across Different AI Models

This assessment of model limitations further supports the conceptual model's view that detection accuracy alone is insufficient; institutions must also evaluate how each model aligns with interpretability needs and evolving regulatory expectations.

Table 1 AI Model Effectiveness in Fraud Detection

AI Model	Detection Speed	False Positives	Industry Effectiveness Rating
Machine Learning (ML)	Fast	Medium	High
Deep Learning (DL)	Very Fast	Medium	Very High
Natural Language Processing (NLP)	Medium	Low	Medium

Table 2 AI Adoption for Fraud Detection

AI Model	Adoption Rate (%)
Machine Learning (ML)	94%
Deep Learning (DL)	67%
Natural Language Processing (NLP)	72%

The performance thresholds reflected in Table 1 are informed by operational standards in real-time fraud detection environments. Real-time fraud detection systems often require sub-second response times to avoid delays in transaction processing, particularly in mobile and card-present environments. Meanwhile, elevated false positive rates — where legitimate transactions are mistakenly flagged—can lead to costly manual reviews, customer dissatisfaction, and operational delays. These qualitative benchmarks provide a practical lens to assess how AI models perform under real-world financial constraints.

According to Tables 1 and 2, the data highlights that Machine Learning (ML) is the most widely used AI technique in fraud detection, with an adoption rate of 94%. However, the data also emphasize that Deep Learning (DL) models are highly effective, but not yet widely adopted (67%). The Natural Language Processing (NLP) technique received relatively lower ratings from institutions, with the lowest rating across three metrics. While the model indicates strong daily savings, these gains can be rapidly offset by regulatory penalties when AI systems fail to meet

compliance standards. As seen in Table 3, institutions have incurred AML-related fines ranging from $60 million to over $4 billion, often due to weaknesses in oversight, explainability, or data governance.

Regulatory & Compliance Challenges

While the conceptual model highlights the financial value of AI adoption, it also underscores the potential risks when governance mechanisms fail. The following enforcement data illustrates how lapses in explainability, data quality, or oversight can lead to substantial regulatory fines, even when advanced AI systems are in place.

Table 3 Timeline for Regulatory Fines for AI-Related Fraud Detection Failures

Years	Company/Institution	Fine Amount ($)	Regulatory Violation
2022	Global AML Fines (Various Banks)	~$5 Billion	Anti-Money Laundering Violations
2022	Morgan Stanley	$60 Million	Data Security Failures
2023	Deutsche Bank	$186 Million	AML Shortcomings
2023	Binance	$4.3 Billion	AML Violations
2024	City National Bank	$65 Million	AML Violations
2024	Toronto-Dominion Bank (TD Bank)	$3.0 Billion	AML Violations
2025	Paypal	$2.0 Million	Fined by New York's Department of Financial Services for cybersecurity failures that exposed customers' Social Security numbers
2025	OKX (Aux Cayes Fintech Co. Ltd)	$504 Million	AML Violations - failing to prevent criminals from using its services.

Although Table 3 includes examples from international entities, they are presented to illustrate global enforcement trends and underscore the importance of regulatory compliance,

which remains the primary focus of this U.S.-centered analysis. From Table 3, it is noteworthy that regulatory fines for AI fraud detection failures are varied in materiality. Satisfying AML compliance requirements remains a significant challenge for the development of AI-driven fraud detection for financial institutions. It appears that financial institutions were fined the most under this regulation. This finding highlights the importance of integrating effective AI-related fraud detection systems. It also highlights the crucial element of continuous improvement for fraud detection systems to meet the standards of regulatory compliance.

Overall, the ultimate purpose of this section is to assess and evaluate the effectiveness of AI-powered fraud detection systems in financial institutions. The paper seeks to answer the question: To what extent are AI-powered systems improving fraud detection outcomes, and where do gaps remain? The conceptual model presented earlier demonstrates that under optimal conditions, AI systems can generate significant daily cost savings—approximately $71,510 per day, or $17.9 million annually—through fraud loss prevention and reduced operational expenses. However, as this analysis also shows, model-level improvements and cost efficiencies are not sufficient on their own. True effectiveness also hinges on institutional oversight, model interpretability, and regulatory compliance, which continue to present substantial challenges. Institutions that overlook explainability and compliance integration may ultimately face enforcement actions that would offset the operational savings from AI implementations. The high cost of enforcement actions, as outlined in Table 3, highlights the need for financial institutions to treat AI not just as a tool for automation but as a system that must be continuously audited, governed, and aligned with legal standards.

IV. Analysis and Interpretation of Results

Analytical Summary of Findings

The analysis of AI model performance indicates that while Machine Learning (ML) models have the highest adoption rate, they are not always the most effective at detecting complex fraud patterns. Evidently, Machine Learning (ML) models were rated behind Deep Learning (DL) models in both metrics of detection speed and effectiveness rating. While Deep Learning (DL) models demonstrate superior detection speed and accuracy, which makes them more suitable for large-scale fraud detection and prevention, the lower adoption rate suggests that financial institutions may face challenges related to computational costs and regulatory compliance in implementing this model. Natural Language Processing (NLP) models also prove effective in fraud detection, but their scalability limitations hinder broader implementation, resulting in underwhelming results in all metrics compared to Machine Learning (ML) and Deep Learning (DL). Additionally, the challenge of eliminating false positives remains persistent with AI-powered fraud detection systems at financial institutions, as evidenced by the fact that none of the ratings for all AI model performance were above medium. While it is an interesting finding, it aligns with our original expectations due to the need for continuous optimization in these AI models.

Regulatory and compliance challenges emerge as critical findings in the study, particularly with respect to AML compliance. It was a noteworthy finding, as it demonstrates the importance regulators and enforcers place on the effectiveness of these AI-powered fraud detection systems. The increasing number of fines levied against major banks and fintech companies, such as Binance, Deutsche Bank, and TD Bank, underscores the importance of prioritizing regulatory

compliance in AI-driven fraud detection strategies. The high materiality of these fines and penalties indicates how regulators perceive deficiencies and breaches within these detection systems. Many financial institutions have been compelled to reassess AI-based voice verification due to heightened fraud risks, underscoring the need for models that strike a balance between efficiency and compliance with regulatory frameworks (BioCatch, 2024). As a result, financial institutions are implementing multiple security measures to update their detection and prevention systems in conformity with regulatory compliance. Additional security measures include biometrics, multi-factor authentication, device fingerprints, knowledge-based authentication, document verification, and behavioral biometric intelligence (BioCatch, 2024). With the increase in sophisticated and intricate financial crimes, the finding further emphasizes the need to improve these AI models. AI is expected to impact several areas of financial crime prevention strategies, from detection to compliance.

These findings also align closely with the study's conceptual model, which frames AI fraud detection performance as the product of interactions between technical capability, institutional readiness, and regulatory pressure. The model illustrates how AI-driven fraud detection can yield substantial benefits, such as an estimated $71,510 in net daily savings through reduced fraud losses and operational efficiencies. However, it also cautions that these benefits can be quickly negated by compliance failures and regulatory penalties, as evidenced by recent high-profile enforcement actions. For example, even with strong model performance, institutions that fail to implement sufficient oversight, ensure explainability, or manage false positives risk multi-million-dollar fines that far outweigh operational gains. The model further underscores that successful AI deployment must go beyond algorithmic performance; it must be embedded within a governance framework supported by human oversight and internal audit. In this way, the conceptual model provides a useful interpretive lens for understanding how AI systems function not only as technical solutions but as dynamic components of a complex financial, ethical, and regulatory ecosystem.

Implications of the Results

Recent research by Talaat et al. (2025) introduced RiskNet, a modular fraud detection framework that integrates feature selection and explainable AI. Their results demonstrated superior accuracy and interpretability compared to traditional ML models, reinforcing the need for hybrid, transparent systems in financial fraud detection. Similarly, Boudreaux (2025) emphasizes the importance of mutual dependence between humans and AI, arguing that effective oversight and collaboration are essential for managing AI-driven systems in high-stakes environments.

For Financial Institutions & Banks

For financial institutions and banks, the findings highlight the need for strategic investment in AI fraud detection systems. While AI can significantly enhance the accuracy and efficiency of fraud detection, financial institutions must focus on minimizing false positives to prevent customer dissatisfaction and unnecessary transaction blocks. The results of the study suggest that hybrid AI models, incorporating Machine Learning (ML), Deep Learning (DL), and traditional rule-based detection, will provide a more balanced approach to fraud prevention. It is an intricate combination of the most effective aspects of each model. For instance, financial institutions should focus on cultivating the fraud detection effectiveness from Deep Learning (DL) techniques while maintaining the integration of the Machine Learning (ML) model (refer to Table 1). Hybrid models

that combine both Deep Learning (DL) and traditional Machine Learning (ML) approaches need to be explored to determine if they can leverage the strengths of both methodologies to enhance predictive accuracy (Bengani, 2024). Likewise, financial institutions can determine which AI models are best suitable for which type of financial crime. Hybrid learning systems will be a significant leap forward in our quest to make AI more versatile, interpretable, and efficient (Bengani, 2024). Banks should fully leverage AI functionalities to enhance their decision-making processes. It can ultimately lead to a reduction in unnecessary transaction blocks and an improvement in customer experience. It is essential to preserve a balance between technology and human oversight, as fully automated systems may fail to capture nuanced fraudulent behaviors (Steinhaeuser, 2024). Moreover, financial institutions, such as banks, must ensure that AI implementations align with Anti-Money Laundering (AML), General Data Protection Regulation (GDPR), and AI ethics guidelines to avoid legal and financial repercussions. The rise in regulatory fines for non-compliant AI fraud detection systems emphasizes the need for financial institutions to integrate AI solutions that align with evolving legal frameworks such as the GDPR, EU AI Act, and Anti-Money Laundering (AML) laws (refer to Table 2). Institutions that fail to ensure compliance risk substantial financial penalties and reputational damage. The adoption of hybrid AI models that integrate multiple detection methodologies could improve fraud detection efficiency while maintaining regulatory compliance.

For the Detection of Fraudsters & Cybercriminals

For fraudsters and cybercriminals, the increasing sophistication of AI-driven fraud detection presents significant challenges to their illicit activities. However, cybercriminals continue to adapt, leveraging adversarial AI techniques to bypass security measures (Stanham, 2025). While AI-driven real-time fraud detection reduces financial crime rates, adversarial attacks continue to pose a challenge. Cybercriminals can leverage AI-powered tools and Machine Learning (ML) programs to automate and accelerate various phases of a cyberattack (Stanham, 2025). Fraudsters increasingly use identity-based and social engineering attacks, exploiting API (Application Programming Interface) keys, session cookies, and MFA (Multi-Factor Authentication) bypass techniques (CrowdStrike, 2024). Moreover, there will be additional challenges and difficulties associated with synthetic identity fraud. While Deep Learning (DL) models have demonstrated usefulness in detecting identity fraud, criminals can utilize AI-generated fake identities to manipulate systems, such as those employing Deepfakes technologies (Stanham, 2025). The financial sector must continue to explore and implement more robust verification processes, such as biometric authentication and blockchain-enhanced identity validation. Ultimately, fraudsters will continue to exploit weaknesses in AI models, which require continuous system updates and adaptive fraud prevention strategies. These particular challenges necessitate ongoing improvements in fraud detection processes to stay ahead of evolving threats. Financial institutions must prioritize proactive fraud prevention strategies, including real-time monitoring and adaptive AI models, to mitigate emerging risks.

For Regulators & Policymakers

From a regulatory and policymaking perspective, AI-powered fraud detection systems must be designed with transparency and accountability to reduce algorithmic bias. The increase in fines and penalties for AI-related fraud detection failures underscores the pressing need for clear and

comprehensive regulatory guidelines. Regulators and policymakers must establish standardized compliance frameworks for AI fraud detection to ensure fairness and accountability. It is an imperative and urgent matter for a government that lacks comprehensive AI legislation, such as the US (Simpson, 2023). Currently, the regulation of AI governance in the U.S. is managed by several existing agencies, including the U.S. regulatory agencies (Simpson, 2023). These regulatory agencies have come together to create a collective pledge to battle against discrimination and bias in automated systems. That attempt reflects a proactive approach to ensure that AI technologies do not perpetuate existing inequalities (Simpson, 2023). However, financial regulators must establish stricter compliance frameworks to ensure the ethical deployment of AI in fraud detection. Explainability in AI decision-making is particularly critical, as opaque models can lead to unjustified transaction rejections and potential discrimination. Requiring financial institutions to implement explainable AI models would enhance trust in AI-driven security measures while improving regulatory compliance. Transparency is essential for strengthening public trust and guaranteeing fair treatment of customers flagged for fraudulent activities.

In summary, AI-powered fraud detection has made significant strides in enhancing financial security and prevention, but regulatory compliance remains a crucial determinant of its effectiveness. There are still challenges related to false positives, model bias, and compliance that remain as areas for improvement. A hybrid AI model is a viable solution worth considering to achieve higher accuracy and efficiency. Financial institutions must navigate the dual challenge of technological innovation and stringent regulatory oversight to maintain the integrity of their fraud detection systems. Policymakers should work closely with financial institutions to develop the most effective approaches and practices for AI implementation that strike a balance between innovation and compliance. By continuously refining AI models, implementing transparent decision-making frameworks, and aligning with regulatory standards, the financial sector can harness AI's full potential to combat financial fraud efficiently and fairly.

V. Discussion

What can be improved for AI-Powered Fraud Detection Systems?

Data quality

The findings of this study highlight the growing reliance of financial institutions on AI-powered fraud detection systems, underscoring both their effectiveness and the challenges they present. The analysis of various AI models, including Machine Learning (ML), Deep Learning (DL), and traditional rule-based, demonstrates that hybrid approaches tend to offer the most balanced fraud detection outcomes. While AI has significantly improved real-time fraud detection and reduced financial crime incidents, its limitations, such as false positives and adversarial attacks, necessitate further advancements. The accuracy and effectiveness of AI-powered fraud detection systems are heavily dependent on the quality of the data they analyze (Gupta, 2024). Inconsistent, incomplete, or biased datasets can lead to inaccurate fraud detection, resulting in both increased false positives and false negatives. AI-powered fraud detection systems utilize and learn from transaction data, user activity logs, and third-party data sources, including credit reports and geolocation services (Gupta, 2024). Financial institutions must implement robust data validation and cleansing processes to ensure that AI models are trained on high-quality, representative data. It is critical to ensure that inadequate data are eliminated, as they have an adverse influence on fraud algorithm

predictions and signal detection (Gupta, 2024). Additionally, a collaboration between financial institutions and regulators can help establish standardized data-sharing practices that enhance the accuracy of fraud detection while maintaining privacy and compliance.

Roles of Internal Audits in AI Fraud Detection

Regulatory compliance remains a critical issue as financial institutions face increasing scrutiny regarding AI bias, transparency, and explainability. The increase in regulatory fines for AI compliance failures underscores the need for a more structured and standardized regulatory framework. It is not only designed for the effectiveness of AI fraud detection models but also ensures ethical soundness and compliance with legal requirements. Since there is a lack of comprehensive AI legislation in the US, the role of internal auditors in AI fraud detection is also crucial in confirming that the AI models are reasonably assured (Simpson, 2023). Internal auditors provide an additional layer of oversight by testing whether AI fraud detection systems are functioning as intended, free from significant biases, and aligned with regulatory requirements (TeamMate, 2025). They play a crucial role in evaluating AI-driven fraud models, identifying shortcomings, and providing recommendations for refinements. Financial institutions must strengthen their internal audit mechanisms to regularly evaluate AI fraud detection systems, ensuring transparency, compliance, and reliability, in line with regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) (TeamMate, 2025).

For example, the Office of the Comptroller of the Currency (OCC) requires national banks to maintain effective internal audit functions that assess the adequacy of risk management systems, including those involving AI. Similarly, the Federal Reserve's supervisory guidance emphasizes the role of internal audit in evaluating model risk management frameworks, particularly for AI-driven systems used in fraud detection. These agencies expect internal auditors to independently validate AI models, assess compliance with regulatory expectations, and ensure that governance structures are in place to manage AI-related risks.

Additionally, internal auditors are responsible for ensuring compliance with data security and regulatory requirements, including the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and Payment Card Industry Data Security Standard (PCI DSS) (TeamMate, 2025). As AI models are increasingly subject to regulatory scrutiny, auditors must ensure that these systems operate transparently and fairly, thereby mitigating risks associated with algorithmic bias and discriminatory decision-making. To improve AI fraud detection, auditors also contribute to continuous model improvements. By examining audit findings, financial institutions can refine their AI algorithms, optimize fraud detection parameters, and implement corrective measures to address vulnerabilities in their systems. Internal auditors can further ensure the data quality used in building these AI models. Internal auditors can ensure datasets are complete, up-to-date, and free from biases that could impact fraud detection outcomes (TeamMate, 2025). Overall, internal auditors serve as a bridge between AI developers, compliance teams, and senior management, facilitating a well-integrated approach to AI governance. Their role in conducting independent evaluations and fostering accountability ensures that AI fraud detection remains both effective and ethical.

What are other challenges in AI that have implications for fraud detection?

AI fraud detection models, particularly Machine Learning (ML), Deep Learning (DL), and traditional rule-based, demonstrate varying levels of effectiveness. Hybrid AI approaches could show promise in enhancing fraud detection accuracy (Bengani, 2024). However, model limitations such as overfitting, lack of interpretability, and high computational costs present challenges for widespread adoption. Future research should focus on optimizing AI algorithms to minimize false positives and enhance real-time detection capabilities while reducing operational burdens on financial institutions. One of the challenges in AI-powered fraud detection is model interpretability. Many advanced AI models, such as Deep Learning (DL) and Neural Networks, operate as 'black boxes,' making it difficult for financial institutions to understand how decisions are made (Ducret, 2025). This lack of transparency can create trust issues for both regulators and customers, especially in cases where transactions are flagged as fraudulent without clear explanations. Enhancing model interpretability through Explainable AI (XAI) techniques is crucial to gaining regulatory and public trust while ensuring compliance with governance requirements (IBM, 2025). Financial institutions can also allow internal auditors to support and approve the use of explainable AI techniques, which provide a straightforward understanding of how fraud detection decisions are made (TeamMate, 2025).

Moreover, AI fraud detection systems must navigate complex ethical concerns, particularly those related to bias and fairness. If trained on biased datasets, AI models may disproportionately target certain demographic groups, leading to financial exclusion or unjust scrutiny (Ducret, 2025). When detection systems examine financial transactions or assess high-risk customers, the possibility of biased decisions, whether from data on demographics, socioeconomic elements, or historical data patterns, can be profound and unintentional (Ducret, 2025). This ethical concern is particularly relevant for financial institutions, as AI-driven determinations can impact decisions and conclusions that have financial implications for their consumers. To manage these challenges, financial institutions must ensure that AI models are trained on diverse, representative datasets and continually scrutinize their outputs for indications of bias (Ducret, 2025). If these systems are not ready in the short term, human oversight can play a critical role in the decision-making process. Additionally, ethical concerns arise regarding data privacy, as AI fraud detection relies on a vast collection of transactional and behavioral data (Ducret, 2025). Financial institutions must implement rigorous ethical frameworks to prevent bias and ensure AI decisions are aligned with transparency and privacy requirements (Ducret, 2025).

Limitations of this Research

While this study offers valuable insights into AI-powered fraud detection, several limitations should be acknowledged. One key limitation is the availability of comprehensive data from financial institutions. Many financial institutions do not publicly disclose or publish detailed fraud detection information due to concerns about confidentiality and competitive considerations. As a result, this research relies on publicly available industry reports and case studies, which may not fully capture the nuances of fraud detection effectiveness across different financial sectors. For instance, there is a lack of some AI fraud detection metrics, such as accuracy rates across different financial institutions and Graph Neural Networks' adoption. Likewise, another limitation pertains to the generalization of AI model performance. The effectiveness of AI-driven fraud detection varies based on the type of financial fraud being studied, such as credit card fraud, money

laundering, or transaction fraud. While this study extensively discusses AI performance, it does not delve deeply into specific fraud categories, which may limit the applicability of its findings to highly specialized financial crimes.

Additionally, this study highlights challenges in measuring AI bias and fairness (Adhikari et al., 2024). AI fraud detection models are susceptible to biases inherent in the datasets they are trained on, which can result in excessive targeting of specific demographics. However, quantifying and mitigating AI bias remains complex due to the lack of widely available demographic data in fraud detection systems. Further research is needed to comprehensively assess AI fairness. Moreover, regulatory uncertainty poses a significant challenge to the implementation of AI fraud detection. AI-related regulations are continuously evolving, with different jurisdictions adopting varied compliance standards. While this study examines available and relevant regulatory frameworks, future legislative changes could significantly impact how AI fraud detection systems operate and comply with laws such as GDPR, AML regulations, and AI ethics guidelines (Simpson, 2023).

Opportunities for further research

Despite the significant advancements in AI fraud detection, several areas warrant further exploration to enhance the effectiveness, fairness, and adaptability of these systems. One promising area for future research is the development of defense mechanisms against adversarial AI. As fraudsters increasingly leverage AI-driven attacks, financial institutions must explore methods to counter adversarial AI threats. Research into robust machine learning models that can detect and adapt to adversarial manipulation will be crucial in maintaining the efficiency of fraud detection. Since Graph Neural Networks (GNNs) data is lacking, it suggests a gap in industry-wide adoption or available research. The lack of GNN-related data presents an opportunity for further research into its effectiveness in fraud detection. It would provide additional insights into the effectiveness of AI models with Graph Neural Networks (GNNs) compared to other approaches, such as Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP). Additionally, improving AI explainability and interpretability remains a crucial research priority. Explainable AI (XAI) can help financial institutions better understand AI-driven fraud detection decisions, reducing bias and increasing regulatory compliance (IBM, 2025). Future studies should investigate techniques to enhance AI transparency, thereby making fraud detection models more comprehensible to regulators and financial institutions alike. Another critical area is the study of AI bias mitigation strategies. Ensuring that AI fraud detection models do not disproportionately impact certain demographics is crucial for ethical AI adoption. Research should focus on developing unbiased training datasets, fairness-aware algorithms, and standardized frameworks for auditing bias to address these concerns effectively.

While blockchain technology has been around for over a decade, its applications continue to expand and evolve. There is still ongoing research on optimizing its security, scalability, and integration with other technologies (such as AI) is ongoing. The integration of AI and blockchain in fraud detection systems is important due to the complementary strengths of both technologies. It further presents a promising area for further research. Recent studies, such as those by Ketha and Provodnikova (2024), highlight the need for further research and investigation into AI-blockchain integration as a comprehensive strategy for fraud detection in financial transactions. The paper highlights how blockchain's decentralized and tamper-resistant ledger can ensure secure and transparent data. The authors explain how AI excels at identifying patterns and anomalies

indicative of fraud. The authors propose a framework that can leverage both technologies to improve fraud detection and reduce vulnerabilities (Ketha and Provodnikova, 2024).

By exploring these research opportunities, financial institutions, regulators, and AI researchers can collaborate to develop more robust, ethical, and efficient fraud detection systems that keep pace with the evolving landscape of financial crime.

VI. Conclusion

AI-powered fraud detection has revolutionized financial crime prevention by improving precision, efficiency, and real-time monitoring abilities. However, its across-the-board implementation is followed by challenges that financial institutions must address to maximize its benefits while mitigating risks. As the primary goal of this paper is to suggest the next step for AI-powered fraud detection systems, the research findings indicate that hybrid AI models, which integrate Machine Learning (ML), Deep Learning (DL), and traditional rule-based techniques, offer the most comprehensive approach to fraud detection. Financial institutions must continually adapt their AI models to combat evolving fraud tactics and approaches. Investing in Explainable AI (XAI) to improve transparency and regulatory compliance will provide more meaningful insights to all stakeholders. Collaboration with regulatory bodies is necessary to ensure that AI-driven fraud detection meets both ethical and legal standards while maintaining public trust. Looking ahead, further research should explore the role of adversarial machine learning in fraud prevention, as well as the use of Graph Neural Networks (GNNs), to ensure that AI systems remain effective against AI-powered fraud attacks.

In conclusion, while AI-powered fraud detection systems provide a strong foundation for mitigating financial crime, they are not foolproof. Ongoing advancements, regulatory frameworks, and ethical considerations will play a crucial role in shaping the future effectiveness of these systems. The integration of explainable, adaptive, and ethically governed AI solutions will be key to ensuring the long-term success of AI-powered fraud prevention in financial institutions. A proactive and collaborative approach between financial institutions, regulators, internal auditors, and AI researchers will be necessary to create a secure and effective fraud detection ecosystem. Ultimately, this paper contributes to the academic discourse by linking AI model performance to regulatory compliance and institutional implementation challenges. Its practical insights offer guidance for financial institutions navigating the evolving fraud landscape and for policymakers seeking to develop adaptable, enforcement-ready governance frameworks. As such, the findings serve as both a roadmap for institutional innovation and a call to action for future research in AI fraud governance.

References

ABA Banking Journal. (2024, March 4). City National Bank agrees to pay $65 million to resolve risk control allegations. https://bankingjournal.aba.com/2024/03/city-national-bank-agrees-to-pay-65-million-to-resolve-risk-control-allegations/

Adhikari, P., Hamal, P., & Baidoo Jnr, F. (2024). Artificial Intelligence in fraud detection: Revolutionizing financial security. *International Journal of Science and Research Archive, 13 (1)*, 1457–1472. https://doi.org/10.30574/ijsra.2024.13.1.1860

Bengani, V. (2024). Hybrid Learning Systems: Integrating Traditional Machine Learning with Deep Learning Techniques. ResearchGate. https://doi.org/10.13140/RG.2.2.10461.22244/1

BioCatch. (2024). 2024 AI Fraud Financial Crime Survey. https://www.biocatch.com/ai-fraud-financial-crime-survey

Boudreaux, M. (2025). Mutual dependence and vulnerability in human and AI futures. RAND Corporation. https://www.rand.org/pubs/working_papers/WRA1234-1.html

Butler, R. B. (2024, September 30). How AI can help law enforcement fight fraud & other crimes. Thomson Reuters. https://www.thomsonreuters.com/en-us/posts/government/ai-law-enforcement-fraud/

Cheemakurthi, S. K. M., Kilaru, N. B., & Gunnam, V. (2023). Ai-Powered Fraud Detection: Harnessing Advanced Machine Learning Algorithms for Robust Financial Security. *International Journal of Advances in Engineering and Management, 5(4)*, 1907–1915. https://ijaem.net/issue_dcp/Ai Powered Fraud Detection Harnessing Advanced Machine Learning Algorithms for Robust Financial Security.pdf

Core Payment Solutions. (2024, June 28). What fraud prevention features should a POS system have? https://corepaymentsolutions.com/what-fraud-prevention-features-should-a-pos-system-have/

Cornerstone. (2025, June 5). The crucial role of humans in AI oversight. https://www.cornerstoneondemand.com/resources/article/the-crucial-role-of-humans-in-ai -oversight/

Crane, V., & Kimbrell, T. (2025). Anti-money laundering (AML). FINRA. https://www.finra.org/rules-guidance/key-topics/aml

CrowdStrike. (2024). CrowdStrike 2024 Global Threat Report. https://go.crowdstrike.com/rs/281-OBQ-266/images/GlobalThreatReport2024.pdf

Dhrangadhariya. (2025, March 1). The role of Artificial Intelligence (AI) in internal auditing: Transforming risk and compliance. CSM & CO LLP. https://www.csmllp.in/the-role-of-artificial-intelligence-ai-in-internal-auditing-transforming-risk-and-compliance/

Ducret, J. (2025, January 27). AI in fraud detection and due diligence: Top 8 ethical implications. TenIntelligence. https://tenintel.com/ai-fraud-detection-due-diligence/

Emanuel-Burns, C. (2024, October 14). TD Bank hit with $3bn in fines over AML failures - fintech futures: Fintech News. FinTech Futures. https://www.fintechfutures.com/regulatory-actions/td-bank-hit-with-3bn-in-fines-over-aml-failures

Federal Deposit Insurance Corporation. (2004). Bank Secrecy Act, Anti-Money Laundering, and Office of Foreign Assets Control. Federal Deposit Insurance Corporation (FDIC). https://www.fdic.gov/resources/supervision-and-examinations/examination-policies-manual/section8-1.pdf

Financial Crimes Enforcement Network. (2025). Financial institution definition. FinCEN.gov. https://www.fincen.gov/financial-institution-definition

Fitzpatrick, C. (2024, October 4). The impact of AI on Fraud Detection Systems. Planet Compliance. https://www.planetcompliance.com/ai-compliance/ai-fraud-detection-systems/

Flinders, M., Smalley, I., & Schneider, J. (2025, April 30). AI fraud detection in banking. IBM. https://www.ibm.com/think/topics/ai-fraud-detection-in-banking

Georgiev, M. (2024, August 1). Shielding retailers from credit card fraud: A comprehensive guide. NRS Pay. https://nrspay.com/2023/10/02/shielding-retailers-from-credit-card-fraud-a-comprehensive-guide/

Gupta, P. (2024, November 11). Data Engineering challenges in handling large volumes of fraud detection data. MRC. https://merchantriskcouncil.org/learning/resource-center/member-news/blog/2024/discover-data-engineering-challenges-in-handling-large-volumes-of-fraud-detection-data

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of ai-enhanced techniques in credit card fraud detection. *Journal of Big Data, 12(1)*. https://doi.org/10.1186/s40537-024-01048-8

Hayes, A., Kelly, R., & Kvilhaug, S. (2024, February 25). What is a financial institution?. Investopedia. https://www.investopedia.com/terms/f/financialinstitution.asp

Hickman, T. H., Lorenz, S., Teetzmann, C., & Jha, A. (2024, July 16). Long awaited EU AI Act becomes law after publication in the EU's Official Journal. White & Case LLP. https://www.whitecase.com/insight-alert/long-awaited-eu-ai-act-becomes-law-after-publication-eus-official-journal

Hodge, N. (2024, June 10). The fraudsters have AI, too. Theiia - Internal Auditor. https://internalauditor.theiia.org/en/articles/2024/june/the-fraudsters-have-ai-too/

IBM. (2025, February 13). What is explainable AI (XAI)? https://www.ibm.com/think/topics/explainable-ai

Kamuangu, P. (2024). A Review on Financial Fraud Detection using AI and Machine Learning. *Journal of Economics, Finance and Accounting Studies, 6(1)*, 67-77. https://doi.org/10.32996/jefas.2024.6.1.7

Ketha, S., & Provodnikova, A. (2024, December 2). Combining Blockchain and AI for Fraud Detection: Building Secure, Transparent, and Sustainable Financial Ecosystems. *Global Journal of Business and Integral Security.* https://gbis.ch/index.php/gbis/article/view/599/500

Kurnia, P., & Yuniarti, R. (2024). Analysis of Fraud Diamond Theory in Detecting Fraudulent Financial Statement: Study in Manufacturing Company in Indonesia. *Dinasti International Journal of Economics, Finance & Accounting, 5(5)*, 5468–5478. https://doi.org/10.38035/dijefa.v5i5.3572

Levitt, K. (2024, December 5). How is AI used in fraud detection? NVIDIA Blog. https://blogs.nvidia.com/blog/ai-fraud-detection-rapids-triton-tensorrt-nemo/

Louati, H., Louati, A., Almekhlafi, A., ElSaka, M., Alharbi, M., Kariri, E., & Altherwy, Y. N. (2024). Adopting Artificial Intelligence to Strengthen Legal Safeguards in Blockchain Smart Contracts: A Strategy to Mitigate Fraud and Enhance Digital Transaction Security. *Journal of Theoretical & Applied Electronic Commerce Research, 19(3),* 2139–2156. https://doi-org.libproxy.scu.edu/10.3390/jtaer19030104

Lumenova AI. (2024, September 17). The strategic necessity of human oversight in AI Systems. https://www.lumenova.ai/blog/strategic-necessity-human-oversight-ai-systems/

Mastercard. (2024, January 24). Industry perspectives on AI and transaction fraud detection: Brighterion AI: A MasterCard Company. Brighterion AI | A Mastercard Company. https://b2b.mastercard.com/news-and-insights/blog/industry-perspectives-on-ai-and-transaction-fraud-detection/

Mayfield, J. M. (2024, August 20). As nationwide fraud losses top $10 billion in 2023, FTC steps up efforts to protect the public. Federal Trade Commission.

https://www.ftc.gov/news-events/news/press-releases/2024/02/nationwide-fraud-losses-top-10-billion-2023-ftc-steps-efforts-protect-public

Olowu, O., Adeleye, A. O., Omokanye, A. O., Ajayi, A. M., Adepoju, A. O., Omole, O. M., & Chianumba, E. C. (2024). AI-driven fraud detection in banking: A systematic review of data science approaches to enhancing cybersecurity. *GSC Advanced Research and Reviews, 21(2),* 227–237. https://doi.org/10.30574/gscarr.2024.21.2.0418

Pavion. (2024, April 26). The role of AI in fraud detection for retail businesses. Pavion. https://pavion.com/resource/the-role-of-ai-in-fraud-detection-for-retail-businesses/

Petraşcu, D., & Tieanu, A. (2014). The role of Internal Audit in Fraud Prevention and Detection. *Procedia Economics and Finance, 16*, 489–497. https://doi.org/10.1016/s2212-5671(14)00829-6

Reeder, S. (2025). 10 common financial scams: UW Credit Union: UW Credit Union. UW Credit Union. https://www.uwcu.org/online-banking/articles/10-scams/

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in Human Resource Management Processes. *Human Resource Management Review, 33(1)*, 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Simpson, W. (2023, August 2). AI regulatory enforcement around the world. International Association of Privacy Professionals (IAPP). https://iapp.org/news/a/ai-regulatory-enforcement-around-the-world

Stanham, L. (2025, January 16). Most common AI-powered cyberattacks . CrowdStrike. https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/ai-powered-cyberattacks/

Steinhaeuser, I. (2024, December 23). How AI will disrupt fraud prevention & detection technologies . Thomson Reuters Institute. https://www.thomsonreuters.com/en-us/posts/corporates/technological-considerations-fraud-prevention/

Stempel, J. (2022, January 3). Morgan Stanley to pay $60 mln to resolve data security lawsuit | Reuters . Reuters. https://www.reuters.com/markets/funds/morgan-stanley-pay-60-mln-resolve-data-security-lawsuit-2022-01-02/

Stempel, J. (2025, January 23). PayPal fined by New York for cybersecurity failures. Reuters. https://www.reuters.com/technology/paypal-fined-by-new-york-cybersecurity-failures-2025-01-23/

Talaat, F. M., Medhat, T., & Shaban, W. M. (2025). Precise fraud detection and risk management with explainable artificial intelligence. Neural Computing and Applications. https://link.springer.com/article/10.1007/s00521-025-11396-y

TeamMate. (2025, February 19). Internal Audit's role in AI Fraud Detection. Wolters Kluwer. https://www.wolterskluwer.com/en/expert-insights/internal-audits-role-ai-fraud-detection

U.S. Department of Justice. (2025, January 31). Financial fraud crime victims. United States Attorney's Office Western District of Washington. https://www.justice.gov/usao-wdwa/victim-witness/victim-info/financial-fraud

United States Department of Justice. (2025, February 24). Okx pleads guilty to violating U.S. anti-money laundering laws and agrees to pay penalties totaling more than $500 million. United States Department of Justice, Southern District of New York. https://www.justice.gov/usao-sdny/pr/okx-pleads-guilty-violating-us-anti-money-laundering-laws-and-agrees-pay-penalties

Valleskey, B. (2024, July 11). The rise of AI fraud Agents. Inscribe. https://www.inscribe.ai/ai-for-financial-services/ai-fraud-agents/

West, A., & Ciais, F. (2023, December). Impact of artificial intelligence on fraud and scams. PwC. https://www.pwc.co.uk/forensic-services/assets/impact-of-ai-on-fraud-and-scams.pdf

Yuhertiana, I., & Hadi Amin, A. (2024). Artificial Intelligence Driven Approaches for Financial Fraud Detection: A systematic literature review. KnE Social Sciences. https://doi.org/10.18502/kss.v9i20.16551

Strategic Reinsurance and Explainable AI

Sampan Nettayanun[1] and Eric R. Brisker[2]

Abstract

This study explores the strategic determinants that impact reinsurance purchase decisions in the P&C insurance industry using the Shapley Additive exPlanations explainable artificial-intelligence (XAI) framework, or SHAP library. Key determinants such as financial considerations, competition, and industry demand for reinsurance are considered to identify their impact on different levels of ceding. The XAI process ranks these determinants based on their influence on reinsurance purchases, and identifies clear relationships between these determinants and ceding levels. For instance, an increase in writing a specific product type can lead to a lower incentive to hedge more within that product type. Additionally, this methodology also reveals more complex relationships between determinants and reinsurance purchases based on their values. Finally, the study includes a machine learning significance test for each determinant impacting insurance purchases.

Keywords: risk management, insurance, machine learning, artificial intelligence, explainable AI, property and casualty insurance

JEL Classifications: G22, G32

I. Introduction

The use of artificial intelligence (AI) in financial economics has increased tremendously in recent years. For example, machine learning, or deep learning, has been used to analyze mortgage risk (Sirignano, Sadhwani, and Giesecke; 2016), portfolio selection (Heaton, Polson, and Witte; 2017), and analyze large pools of loans (Sirignano and Giesecke; 2019). In the insurance industry, Brockett et al. (1994) and Brockett et al. (2006) utilized artificial neural networks (ANN) to predict insurer's insolvency. Additionally, Hejazi and Jackson (2016), Wüthrich and Merz (2019) and Wüthrich (2019) applied ANN models in an actuary framework. Most prior AI studies attempt to predict outcomes based on AI's non-linear advantage over widely used classical regression analysis and superior accuracy in making predictions. However, the explanation of independent variables, or so-called "features", in the AI setting have previously not been explainable. The AI model, such as ANNs, can be thought of as "black boxes" built on different (hidden) layers and neurons. In recent years, researchers have started to open these black boxes using several methodologies. This study aims to use the ANN models to implement existing econometric techniques to explain insurers demand for reinsurance focusing on managerial strategic decision making. More importantly, we try to open the AI "black box" using explainable AI. This study uses the SHapley Additive exPlanations (SHAP) methodology introduced by Lundberg and Lee (2017) to complement the ANN model. SHAP allocates each variable's influence on the overall prediction based on Shapley's values using a game-theoretical approach.

[1] Naresuan University, Faculty of Business, Economics and Communications, Phitsanulok, Thailand. sampann@nu.ac.th.

[2] The University of Akron, College of Business, Department of Finance, Akron, Ohio, USA. ebrisker@uakron.edu.

According to Modigliani and Miller (1958), insurers would not need to consider purchasing reinsurance in a frictionless world. However, later literature considers frictions that lead to different levels of firm hedging. Determinants that have been shown to lead to different insurer ceding levels include taxes (Smith and Stulz (1985)), external cost of financing (Froot, Scharfstein, and Stein (1993), and Froot and Stein (1998)), the reinsurance market (Cole and McCullough (2006)), and competition and market structure (MacKay and Phillips (2005)), Adam, Dasgupta, and Titman (2007), Rampini, Sufi, and Viswanathan (2014), and Nettayanun (2014)). We include these determinants in our analysis as independent variables and use the level of reinsurance purchases as a risk management measure, with a particular emphasis on competition and strategic risk management.

Explainable AI (XAI) can explain reinsurance purchases in the property and casualty (P&C) industry. Most prior research employs various econometric approaches to explain reinsurance purchases, identifying the following top ten product type influences: commercial long-tail (*CL*), commercial short-tail (*CS*), personal short-tail (*PS*), and personal long-tail (*PL*), leverage, brokerage expense position compared to its peers, size, ownership structure, and number of companies in the industry. Additionally, the SHAP library can reveal more complex relationships. For example, broker expenditure comparisons to peers show a positive relationship with ceding level (SHAP values of 0 to 0.2[3]). This relationship increases as SHAP values increase as predicted by Maksimovic and Zechner (1991), MacKay and Phillips (2005), and Nettayanun (2014) that find broker or technology investment distances is associated with higher ceding levels. Furthermore, significance tests from Horel and Giesecke (2019) find six significant variables within the top ten SHAP values: leverage, ownership structure, product types (CL, CS, PS), and stock holding percentage in the portfolio. Applying these machine learning techniques can give a richer explanation of reinsurance purchase behavior in the insurance industry. The first major contribution of this study is the application of alternative techniques to analyze reinsurance data in contrast to traditional statistical methods. These techniques do not require non-linearity between dependent and independent variables, allowing for the exploration of variable interactions including visualizations. The second contribution is the use of significant tests within the machine learning framework for reinsurance data. These approaches enable the ranking of determinants importance, exploration of complex relationships, and identification of significant determinants within the machine learning framework.

We proceed as follows. Section 2 reviews related literature. Second 3 discuss XAI models and summarizes determinants/variables from previous literature that affect risk management. Section 4 discusses the dataset. Section 5 analyzes the dataset using different XAI approaches. Finally, the last section concludes the study.

II. Literature review

Recently, there have been significant developments in machine learning in the corporate finance context. For example, Colla et al. (2013) used cluster analysis, a machine learning method, to identify similarities among firms that structure their debts for public US companies. Nettayanun (2014) also used cluster analysis to identify subgroups within the insurance industry based on risk management tools such as reinsurance, investment portfolio, and leverage ratio. Amini et al. (2021) applied machine learning to predict leverage from various financial factors, arguing that these

[3] The positive SHAP values represent the positive contributions of features on labels in the model.

machine learning methods can exploit the non-linear relationships between determinants for predicting leverage. Erel et al. (2021) used machine learning for the director selection process, helping predict directors' performance. Bubb and Catan (2022) used machine learning methods with a large dataset of voting data from mutual funds to explore their relationship with corporate governance.

Korangi, Mues, and Bravo (2023) find that deep learning models outperform previous statistical models in predicting default, using data from mid-cap companies in the US over a sample period of more than 30 years. Griffin, Hirschey, and Kruger (2023) use explainable AI, specifically SHAP, to find various features that explain municipal bond market markups. Similarly, we use SHAP to explain reinsurance purchases. Studies in finance academic literature have discussed the role of artificial intelligence and machine learning in business, highlighting their impact on corporate finance and risk management. In the context of corporate finance, Babina et al. (2024) find that AI investment can directly influence firm growth through product innovation and reduce the cost of product development through process innovation.

This study addresses the gap in understanding the role of AI and machine learning in explaining dependent variables, contributing to the growing body of AI research in corporate finance literature (Hornuf and Schaefer, 2025). Specifically, we use an artificial neural network (ANN) model to explain reinsurance purchases. This study integrates the ANN model with explainable AI or interpretable machine learning techniques such as SHAP and SFIT to explain the dependent variable, reinsurance level. The input variables, or features, are based on existing variables derived from previous literature. According to Hornuf and Schaefer (2025), explainable AI can assist financial institutions better understand the ANN model and how it utilizes features to inform decisions. This study contributes to the existing literature by using alternative statistical models and techniques to analyze structural data and explain reinsurance purchases, particularly under the assumption of non-linear interactions between variables. Next, we discuss the models and variables used in the study.

III. Models and variables

Econometric techniques provide statistical predictions of how each determinant included in the model as independent variables is related to the dependent variable, however, the main assumption in ordinary least square and panel data models is that the relationships between the dependent and independent variables are linear. In contrast, an artificial neural network (ANN) relaxes this assumption by flexibly selecting an activation function in the hidden layers and neurons. Until recently, a drawback of ANN was its inability to explicitly explain how each determinant influences the dependent variable. This section illustrates the use of ANN through explainable AI (XAI) to explain reinsurance purchases. Specifically, this study uses Shapley Additive explanation (SHAP) methodology, introduced by Lundberg and Lee (2017), to explain the influence of each determinant, or independent variable, on the dependent.

This study implements an artificial neural network model using TensorFlow from the Google Collaboration platform. In the ANN model, the dependent variable, *Cede,* is the "label", and the independent variables are the "features". The model also includes an intercept variable. The implementation is similar to an econometric analysis setting and is also used for significance tests in the next section. The construction of ANN follows Gu, Kelly, and Xiu (2020) closely, using a sequential model with two hidden layers in the Keras library, which provides an optimal construction. The complexity of the model can increase with a higher number of layers and

neurons. The number of neurons follows the geometric pyramid rule from Master (1993), similar to Gu, Kelly, and Xiu (2020). Each hidden layer should have $\sqrt{n \times m}$ neurons where n is the number of input neurons and m is the number of output neurons. Using our 37,447 observations in this analysis, this leads to a first layer number of 8 neurons and second layer of 3 neurons. To avoid outlier problems, most of the variables included in the model must be in a similar scale. Most variables range from 0 to 1, however, some variables need further scaling by subtracting the average of that variable and dividing by its standard deviation. The model uses 80% of the data as a training sample and the remaining 20% as a testing sample to adjust for overfitting the model. The activation function for each neuron is a sigmoid function, which gives the lowest mean square error (MSE). Each type of neuron network is dense, meaning all the neurons are connected.

To understand how each independent variable (feature) influences the dependent variable (label), *cede*, we use SHapley Additive exPlanations (SHAP)[4] introduced by Lundberg and Lee (2017). SHAP is derived from Shapley values from game theory that explain how each player contributes to the game's optimal solution. Lundberg and Lee (2017) implement SHAP using the same principles as Shapley values. The SHAP method is also applied using local data to explain each prediction label, referred to as local interpretable model agnostic explanations (LIME) by Ribeiro, Singh, and Guestrin (2016). In this study, we use DeepExplainer to explain the independent variable (features).

Next, we discuss the independent variables (features) and the dependent variable (label) in an artificial neural network model econometric setting. These variables are derived from both prior theoretical and empirical studies. The dependent variable (label) is $Cede_{it}$, which is the level of reinsurance purchased by the i^{th} insurer in the property and casualty insurance industry at time t. It is defined as the reinsurance ceded divided by the net premium written. This study selects the following independent variables (features) based on the availability of P&C insurance industry data in the SNL database.

Taxes

Smith and Stulz (1985), Stulz (1996), and Graham and Smith (1999) argue that taxes influence risk management activities, with convex tax schedules leading to more strategic hedging to reduce firm net income volatility. For example, during years of high net income, firms protect that income from high tax rates leading to increased firm value. However, Tufano (1996), Cole and McCullough (2006), and Nettayanun (2014) do not find any relationship between tax variables and reinsurance purchases. We define the variable *Tax* as federal and foreign income taxes divided by the book value of assets to measure the tax impact on hedging behavior. Based on previous literature, this study expects a positive relationship between tax and insurance purchasing levels.

Product Type

Adam, Dasgupta, and Titman (2007) argue that the elasticity of demand and the convexity of production costs influence different firm hedging levels. Also, Winter (1994) suggests that each line of insurance business exhibit different reinsurance demand due to varying reinsurance pricing

[4] More details in Lundberg et al. (2018) and the complete explanation and open-source codes are at https://github.com/slundberg/shap.

levels. Consequently, different exposure levels to insurance products can lead to different hedging levels. Mayers and Smith (1990), Cole and McCullough (2006) and Nettayanun (2014) include product types in their analysis of reinsurance levels. Following Phillips, Cummins, and Allen (1998) and Nettayanun (2014), we use four product types of insurance in the property and casualty (P&C) insurance industry: personal short-tail, personal long-tail, commercial short-tail, and commercial long-tail, as detailed in Table A1. Each product exposure is defined as the net premium written in each category divided by total net premium written. The product-type variables for personal short-tail, personal long-tail, commercial short-tail, and commercial long-tail are *PS*, *PL*, *CS*, and *CL*, respectively.

Leverage

Leverage can have either a positive or negative relationship with reinsurance levels. Cole and McCullough (2006) and Hoyt and Liebenberg (2011) use leverage as an independent variable for insurer's reinsurance level. Low leverage can be a substitute for a firm's reinsurance purchases. A firm in distress purchases more reinsurance because of higher uncertainty regarding the firm. Conversely, Rampini, Sufi, and Viswanathan (2014) use a dynamic risk management model to show that firms need capital to be used as collateral for hedging. Therefore, we can expect lower (higher) levels of risk management if the firm has higher (lower) leverage. We measure *Leverage* as the book value of debt divided by the book value of total asset.

Positioning

Strategic positioning of insurers can lead to different risk management strategies. According to Maksimovic and Zechner (1991), the level of technology investment by a firm can act as a natural hedge. For example, if a firm has a similar level of technology investment compared to the industry average, it will not need to hedge as much. MacKay and Phillips (2005) and Nettayanun (2014) use an empirical study to show the relationship between these positioning variables in different industries and firm hedging levels. In the property and casualty insurance industry, we identify four key expenses as technologies that insurance companies need to invest in: agent expenses, brokerage expenses, equipment expenses, and salary expenses. Similar to MacKay and Phillips (2005) and Nettayanun (2014), we define positioning on each technology as,

$$Pos_{i,t} = \frac{|TechExpense_{i,t} - median(\forall_j TechExpense_{i,j,t})|}{range[|TechExpense_{i,t} - median(\forall_j TechExpense_{i,j,t})|]} \quad (1)$$

where index i and j are insurers within the property and casualty insurance industry in year t. $Pos_{i,t}$ measures the difference in each firm's technology investment level from the median industry level in each year. The positions of these technologies are represented by *AgentPos*, *BrokerPos*, *EquipPos*, and *SalaryPos*.

Profitability

Myers (1977) and Mayers and Smith (1987) show that improper risk management can lead to an underinvestment problem. Profitable insurers typically do not require as much reinsurance purchases. Consequently, we expect a negative relationship between profitability and firm ceding

levels. Mayers and Smith (1990), Powell and Sommer (2007), Cole and McCullough (2006), and Nettayanun (2014) also include profitability to capture underinvestment problems. We use return on assets (ROA) to capture the underinvestment problem and expect *ROA* to have a negative relationship with ceding levels.

Size

There are two conflicting conclusions found in the literature regarding the impact firm size has on risk management levels. Tufano (1996) and Froot, Sharfstein, and Stein (1993) suggest that smaller firms hedge more due to higher agency costs and asymmetric information related to financing activities. Mayers and Smith (1990) and Hoyt and Khang (2000) also find a negative relationship between firm size and reinsurance purchases. Conversely, Liu and Parlour (2009) introduce a model where larger firms hedge more because they have a higher probability of winning new businesses compared to smaller firms. Firms hedge less if they lack sufficient capital or resources to acquire or expect to win new business through bidding, as this could result in an over-hedged position. The hedging efforts of smaller firms having a lower chance of winning new business through bidding would be wasted. To win new business, firms will increase their bid efforts. Stulz (1996) also finds that larger firms have higher hedging levels than smaller firms. We capture the size effect using the natural log of the book value of assets, *LnAsset*.

Diversification

Diversification plays a significant role in risk management and can be used as a substitute for reinsurance purchases. By diversifying risks through selling in different regions or various product types, insurer can reduce risk and potentially lessen the need for reinsurance. Froot (2007) and Nettayanun (2014) find that focusing on specific lines of insurance products can reduce reinsurance purchases. For example, an insurance company that excels in automobile insurance can manage this risk effectively with a lower cost of capital. The cost of insuring other product lines through additional reinsurance might be too high. Therefore, insurers might benefit from reducing diversification and concentrating their efforts on automobile insurance. This study measures the level of concentration for each firm using a measurement similar to the Herfindahl index, as described by Choi and Weiss (2005), Cole and McCullough (2006), and Leverty and Grace (2010) defined as,

$$Diversification_{it} = 1 - \sum_{j=1}^{n}(\%\ share\ of\ net\ premium\ written\ in\ product\ type\ j\ for\ company\ i\ at\ time\ t)^2 \quad (2)$$

where n is the total number of product types[5]. The more concentrated an insurer is, the lower their diversification index will be, ranging from 0 to 1. Cole and McCullough (2006) found that insurers with a higher diversification index tend to buy less reinsurance.

[5] In Thailand, there are 4 product types classified by OIC: Fire, Marine, Automobile, and Miscellaneous.

Industry's Reinsurance Demand

The overall industry level of hedging influences an individual firm's level of hedging decisions, as suggested by Nain (2004) and Adam, Dasgupta, and Titman (2007). For example, if the industry-wide hedging level is high, but a particular firm within that industry has minimal hedging, the firm may benefit more than the industry as a whole if there are no adverse shocks. Conversely, if an adverse shock impacts the industry, the firm with minimal hedging may be worse off. To capture this determinance, we use the industry median hedging level, *IndustryCede*.

Intensity of Competition

There are mixed results regarding how the number of competitors in an industry impact reinsurance purchase. Mello and Ruckes (2005) and Adam and Nain (2013) argue that the number of firms negatively correlate with hedging. According to Mello and Ruckes (2005), firms using less hedging can benefit more during periods of higher cash-flows. Conversely, Allayannis and Ihrig (2001) and Adam, Dasgupta, and Titman (2007) suggest that firms tend to hedge more in competitive markets, as it becomes harder to adjust costs within such markets. Nettayanun (2014) also find that a higher number of insurers in an industry leads to higher reinsurance levels for individual insurers within that industry. We measure the number of competitors within an industry in each year using the variable *NumComp*. Additionally, we capture the level of industry competition in each year using the Herfindahl index, similar to Liebenberg and Sommer (2008), Nettayanun (2014), and Caporale, Cerrato and Zhang (2017), defined as,

$$Herfindahl_t = \sum_{i=1}^{n} \left(\frac{NPW_{it}}{Total\ NPW_t}\right)^2. \qquad (3)$$

NPW_{it} is the net premium written for insurer i at year t. $Total\ NPW_t$ is the total net premium written for year t. The value of the Herfindahl index ranges from zero to one with values closer to one (zero) indicating low (high) levels of competition.

Market Share

According to Mello and Ruckes (2005), competitive advantages that firms gain through having high power within an industry induces lower levels of risk management in those firms. Firms with a competitive advantage can charge higher prices than their competitors, resulting in a stronger financial position within the industry and lower hedging levels being used within these high-power firms. Additionally, Sommer (1996) argues that more prominent firms tend to have a healthier financial position than smaller firms. This serves as a positive signal for customers that buy more policies from these prominent firms. It can also increase insurance prices, reducing insolvency risk and leading to lower ceding levels. Therefore, market share can have a negative relationship with the level of risk management used at insurance firms. Nettayanun (2014) also finds that market share has a negative relationship with risk management levels among P&C insurers from 1989 to 2009. We use *MarketShare* to capture the market share of each insurer, defined as,

$$MarketShare_{it} = \frac{NPW_{it}}{Total\ NPW_t}, \qquad (4)$$

where NPW_{it} is the net premium written for company i in year t. $Total\ NPW_t$ is the total net premium written for year t.

Ownership Structure

Caporale, Cerrato and Zhang (2017) find that ownership structure can play a significant role in risk management decisions. According to Mayers and Smith (1981), Lamm-Tennat and Starks (1993), and Pottier and Sommer (1997), stock companies tend to engage in riskier lines of business. Cole and McCullough (2006) find that stock companies engage in lower reinsurance purchases since they have easier access to capital markets and lower risk in acquiring capital when in financial distress. Nettayanun (2014) finds that ownership structure has a positive relationship with reinsurance purchases using a random effects regression model. However, the relationship does not exist when using a fixed effect regression model from 1989 to 2009. Thus, it is interesting to see how ownership structure impacts reinsurance repurchases. The organizational form of firms found in the SNL dataset include stock companies, mutual companies, Lloyds organizations, reciprocal exchanges, risk retention groups, US branches of an alien insurers, and syndicates. The majority of these firms are listed as stock insurers. Therefore, we label the *Stock* variable as 1 if the company is a stock company, and 0 otherwise.

Substitution of Risk

We also control for the substitution of risk in insurance portfolios, similar to the approach of MacKay and Phillips (2005). *StockHolding* is defined as the stock investment in the insurer's portfolio divided by the total admitted assets in the invested portfolio. Higher stock investment implies greater risk for the firm, leading these insurers to hedge more due to the higher risks associated with their investments. In the 3SLS setting, Nettayanun (2014) finds that *StockHolding* has a positive relationship with the ceding level of insurers.

IV. Data

We obtain property and casualty insurance firm-year level data from the SNL dataset from 1996 to 2020. The study focuses on the property and casualty insurance industry due to its unique nature in managing risk management through reinsurance purchases. To reduce anomalies within the data, only firms with net premium written greater than zero are included. We truncate the data on the dependent variable, *Cede*, to values between 0 and 1 to avoid outliers. We also truncate the data on the independent variables *StockHolding* and *Diversification* to fall between 0% and 100%. All variables are winsorized at the 1% level. The variables *NumComp* and *LnAsset* are normalized because their magnitudes are large compared to other variables. The final dataset consists of 37,447 observations from 1996 to 2020.

Table 1 reports summary statistics for all variables used in our analysis. From 1996 to 2020 the mean (median) *Cede* dependent variable was 0.004 (0.000), indicating insurers ceded about 0.4% of net premiums written on average, but only 0.000% at the median level. The commercial long-tail makes up the largest segment of insurance with a mean (median) of 39.3% (19.9%). The mean market share of insurers is 0.000 implying these insurers have a low average level of market share. However, there are some large insurers in our data with the maximum market share reported as 3.2%. The mean (median) *Diversification* is 30.7% (36.6%). There are also a high proportion

of insurers that are stock companies as the average *Stock* is 67.6%.

Table 1. Summary Statistics

Variable	Mean	Median	Std. Dev.	Min	Max
Cede	0.004	0.000	0.027	0.000	0.975
PS	0.236	0.110	0.281	-0.190	1.000
PL	0.119	0.000	0.216	-0.267	1.000
CS	0.251	0.105	0.327	-0.226	1.000
CL	0.393	0.199	0.416	-0.381	1.000
MarketShare	0.000	0.000	0.001	0.000	0.032
Diversification	0.307	0.366	0.262	0.000	1.000
NumComp	2098.10	2146.00	146.94	1782.00	2266.00
Herfindahl	0.015	0.015	0.001	0.014	0.018
AgentPos	0.006	0.000	0.036	0.000	0.998
BrokerPos	0.016	0.009	0.029	0.000	0.844
EquipPos	0.023	0.008	0.048	0.000	0.996
SalaryPos	0.048	0.031	0.065	0.000	0.967
ROA	0.018	0.023	0.093	-2.120	8.724
Tax	0.008	0.004	0.023	-1.555	0.844
Leverage	0.523	0.561	0.215	-0.406	6.015
LnAsset	10.855	10.805	1.903	4.382	17.703
IndCede	0.000	0.000	0.000	0.000	0.000
Stock	0.676	1.000	0.468	0.000	1.000
StockHolding	0.125	0.052	0.172	0.000	1.000

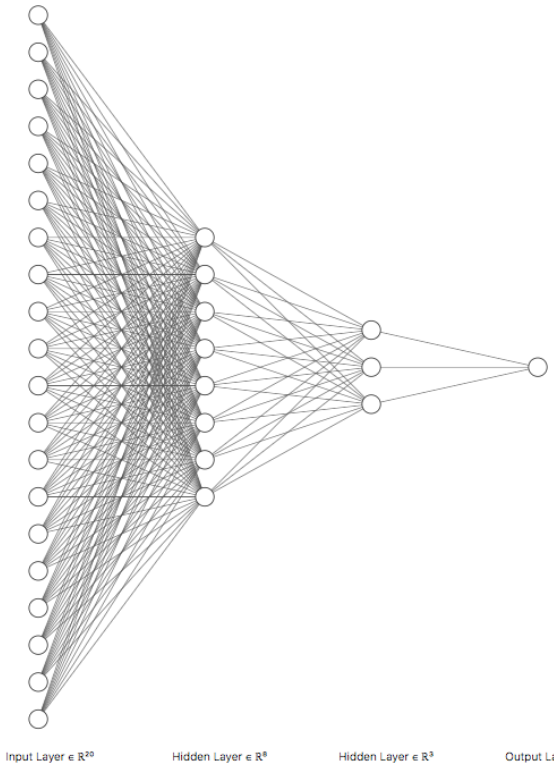
Note: This table reports summary statistics for all variables used in our study. There are 37,447 firm-year observations from 1996 to 2020. The explanations of each variable are as follows: *Cede* is defined by the premium ceded divided by net premium written. *PS* (personal short-tail), *PL* (personal long-tail), *CS* (commercial short-tail) and *CL* (commercial long-tail*)* are the product type variables, defined as the net premium written on each product type divided by net premium written on all product lines for each insurer. *MarketShare* is the market share of each insurer in a particular year defined by the net premium written divided by the total net premium written of the industry in that year. *Diversification* is one minus the Herfindahl index calculated based on each product type offered by each insurer. The Herfindahl index for the industry in each year is calculated using the net premium written from each insurer. *AgentPos*, *BrokerPos*, *EquipPos,* and *SalaryPos* are positioning variables for agent, broker, equipment, and salary, and they measure how each insurer invests in key technologies compared to its peers in the same industry. *ROA* is composed of net income divided by total book value of assets. *TAX* is total income tax divided by total book value of assets. *Leverage* is the book value of total debt divided by the book value of the total assets. *LNAsset* is the natural log of the book value of total assets with units in thousands of dollars. *IndustryCede* is the median value of cede level for all property and casualty insurers in each year. *Stock* is a dummy variable set equal to 1 if the insurer is a stock company, and 0 otherwise. *StockHolding* is defined as the stock investment in the insurer's portfolio divided by the total admitted asset in the invested portfolio.

V. Results

Global Interpretation

This study uses TensorFlow with an optimizer with default parameters for the ANN model using 64 batches and 200 epochs. The study optimizes the number of epochs based on the improvement of the loss function of the mean square error (MSE). The mean squared errors (MSE) from ANN testing sample are 7 x 10^{-4}. The diagram[6] of the ANN model is given in Figure 1.

Figure 1. Artificial Neural Network Model (ANN)



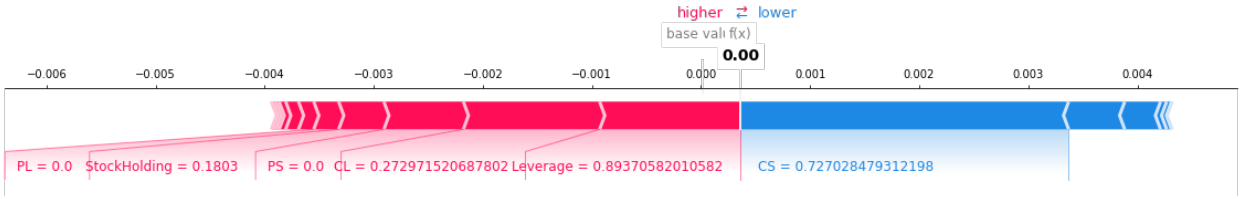
Notes: This figure shows how the ANN model is constructed. It consists of 20 independent variables (features), three hidden layers with each having ten neurons. All neurons are dense (i.e., fully connected).

After obtaining the ANN model, we can produce Shapley values for each independent variable (feature) for each observation (sample). Figure 2 illustrates the force plot from SHAPs DeepExplainer for one observation. The plot shows how each independent variable influences the SHAP value for a particular observation. Variables with red bars increase the SHAP values, while variables with blue bars decrease the SHAP values. From the figure, commercial short-tail (*CS*) is the most prominent feature that pushes the SHAP value lower. Conversely, Leverage is the most significant positive influence on SHAP values. This indicates that the levels of writing insurance in commercial short-tail (*CS*) are the most influential factors for this particular observation. Commercial long-tail (*CL*) also plays a role in explaining the ceding level in this observation, as

[6] This study generates the diagram is from http://alexlenail.me/NN-SVG/index.html.

it is the second variable that positively impacts the ceding level. We can extend the force plot for other observations in the data to analyze how each feature affects the SHAP value.

Figure 2. SHAP Force Plot



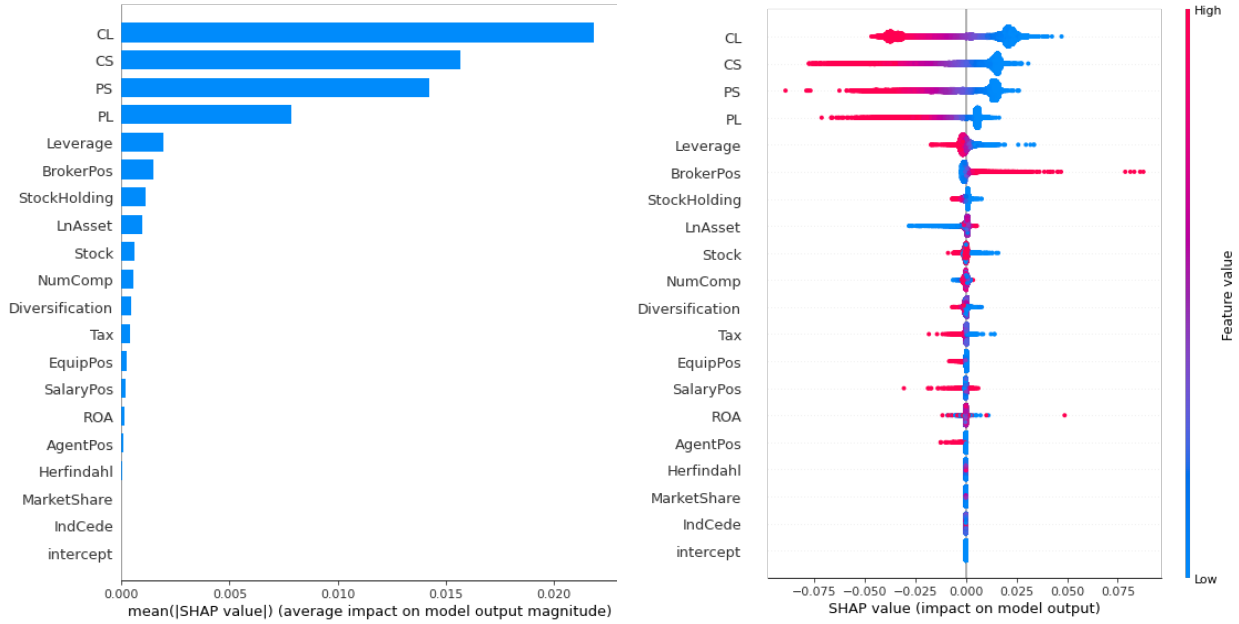
Notes: This figure shows the SHAP force plot. The red color explains how each independent variable increases the SHAP value, and the blue color explains how each independent variable decreases the SHAP value.

Figure 3 illustrates how each feature affects the reinsurance level from a global perspective using DeepExplainer and training data for calculating SHAP values. The left panel is a bar plot of the SHAP value mean absolute values for each feature across all observations. *CL*, *CS*, *PS*, *PL, Leverage*, *BrokerPos, StockHolding, LnAsset, Stock,* and *NumComp* are the top ten variables that influence reinsurance purchases according to SHAP analysis. It shows what product types are the most important in impacting the insurers ceding decisions. In addition, the analysis also shows how *Diversification*, *Tax*, *EquipPos*, *SalaryPos*, *ROA*, *AgentPos*, *Herfindahl*, *MarketShare*, and *IndCede* have influence on reinsurance purchasing levels.

The right panel of Figure 3 shows both negative and positive SHAP values for higher (red) and lower (blue) levels for each featured variable. For example, higher (red) product mix percentages in commercial long-tail, commercial short-tail, personal short-tail, and personal long-tail lines of business mostly result in negative SHAP values. This implies that firms specializing in these specific lines of business are less likely to purchase reinsurance. These results align with the explanation from Winter (1994) and Adam, Dasgupta, and Titman (2007), which suggest that product types influence firm's risk management levels.

Lower (higher) levels of leverage are associated with positive (negative) SHAP values indicating the level of leverage influences hedging decisions. Insurers tend to hedge less (more) for low (high) levels of leverage. The higher (lower) position in broker expense is associated with higher SHAP value levels. We also find that lower (higher) values of stock holdings are associated with higher (lower) reinsurance levels, and higher (lower) differences in firm's brokerage expense position from the industry median is associated with higher (lower) reinsurance levels. In addition, lower (higher) diversification insurers tend to have positive (negative) SHAP values. This is similar to the results from Froot (2007) and Nettayanum (2014). More focused insurers tend to hedge less due to their competitive advantage in the lines of business they focus in. These are the examples that we obtain by looking at the overall SHAP values in a global sense. The local interpretations by variable are in the next section.

Figure 3. SHAP-DeepExplainer



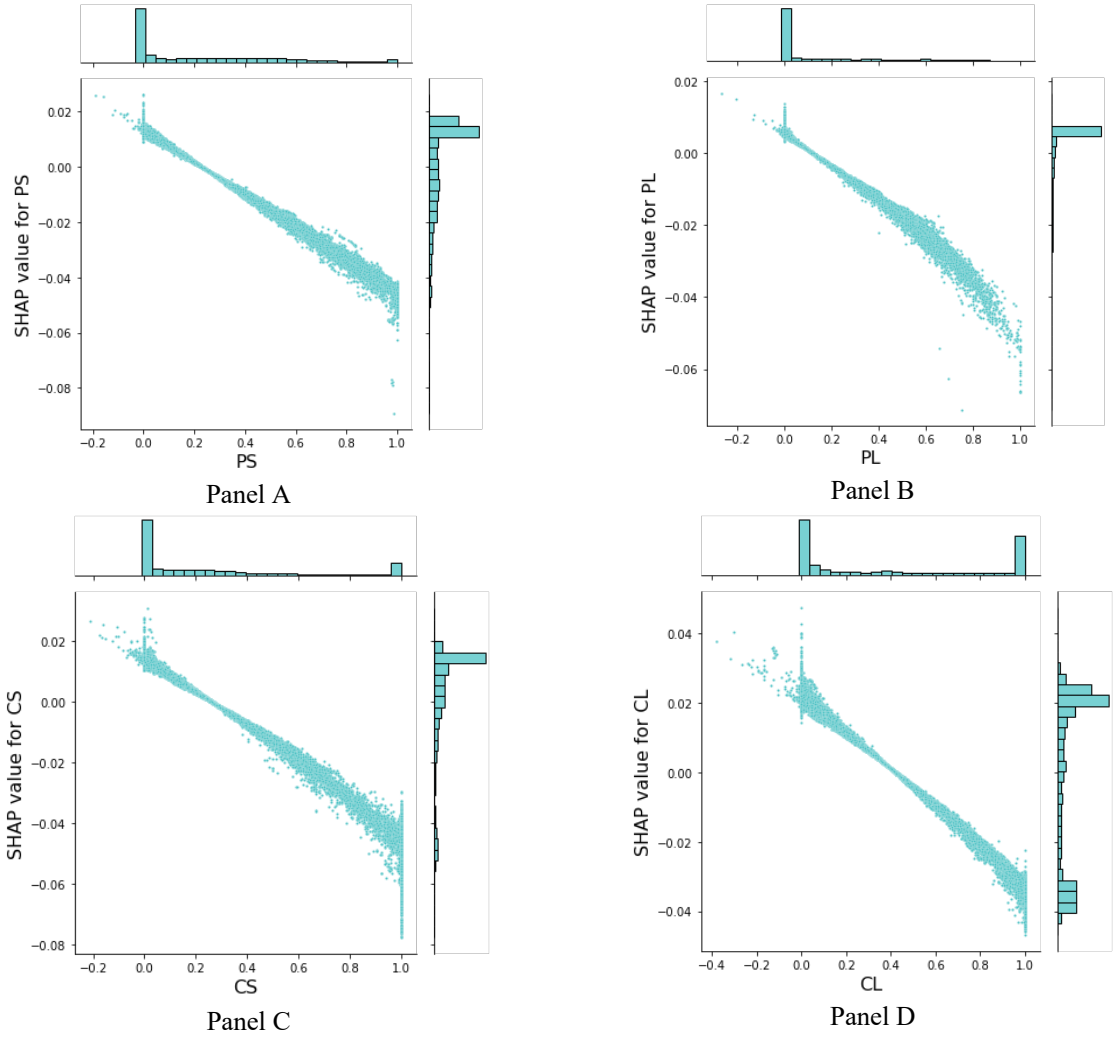
Notes: These figures show how each independent variable influences SHAP values using the overall sample from DeepExplainer in the SHAP library. The left panel is the bar plot of the absolute mean of the SHAP values from the overall sample. The right panel is the violin plot that explains how each sample affects SHAP values. The gradient color is from blue to red. The red color shows a higher value of each independent variable. The blue color exhibits a lower value of each independent variable.

Local Interpretation

After focusing on global interpretations, we now shift to explore how each feature influences insurers' ceding decisions using Shapley values as the key identifier. Figure 4 illustrates the dependence plot for the product-type variables from the SHAP library to show how ceding levels are related with each feature. These plots are similar to the partial/marginal effect of each independent variable in the econometric settings similar to Nettayanun (2014). The plots display Shapley values for all observations corresponding to the selected feature, explaining how Shapley values change when the selected feature varies. This explanation provides insight into how value levels for each feature affect reinsurance demands.

In Figure 4, the personal short-tail (*PS*), personal long-tail (*PL*), commercial short-tail (*CS*), and commercial long-tail (*CL*) are shown in panels A, B, C, and D, respectively. This figure illustrates how product type influences SHAP values, with each product type having a negative relationship with SHAP values. It also shows SHAP value coupled with histograms of each feature on top and SHAP value histograms on the side. Lower (higher) product mix percentages in each product type implies positive (negative) SHAP values. This means that the insurers tend to have more (less) ceding levels when writing a low (high) proportion of *PS*, *PL*, *CS*, or *CL* based on SHAP values. All types of products exhibit this behavior, suggesting that the types of products do not differentiate how firms hedge risk, but the level of writing in each product type does impact risk hedging. This supports Froot (2007) and Mellow and Ruckes (2005) that find insurers may hedge less if they have a competitive advantage. Our findings show that the incentive to hedge decreases for insurers that write more business in a particular business line.

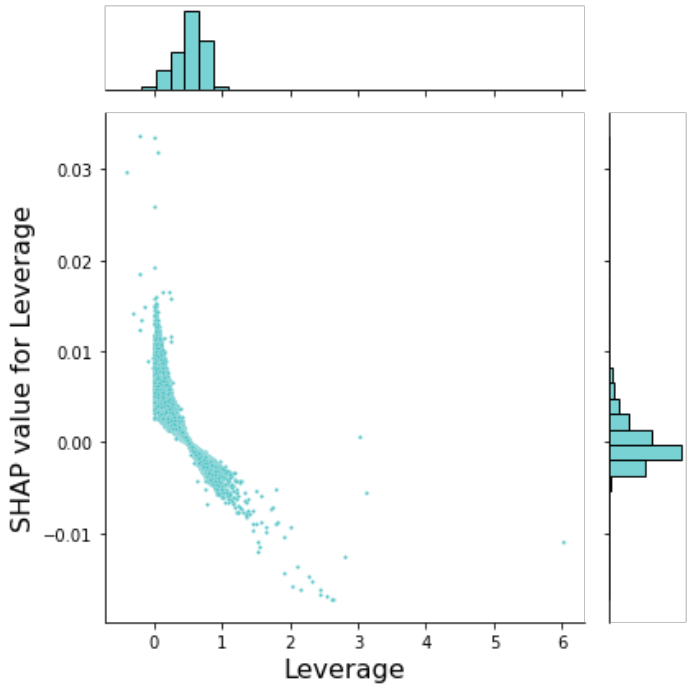
Figure 4. Dependence Plots for the Product Types



Notes: Dependence plots of SHAP values and product types. Panel A is for feature personal short-tail (*PS*) with associated feature *Stock*. Panel B is for feature personal long-tail (*PL*) with associated feature *BrokerPos*. Panel C is for feature commercial short-tail (*CS*) with associated feature *LnAsset*. Panel D is for feature commercial long-tail (*CL*) with associated feature *Stock*.

Figure 5 illustrates the dependence plot for *Leverage*. Low levels of leverage are mostly associated with positive SHAP values. Some leverage values are greater than one indicating a negative equity for insurers. This result supports Rampini, Sufi, and Viswanathan (2014) that find a higher level of leverage lowers the level of risk management activity, contrary to the opposite predictions from Cole and McCullough (2006) and Hoyt and Liebenberg (2011). Having more leverage in the company might influence insurers to cede less and retain more risk. Our analysis does not show a substitution effect between leverage and reinsurance.

Figure 5. Dependence Plot for *Leverage*

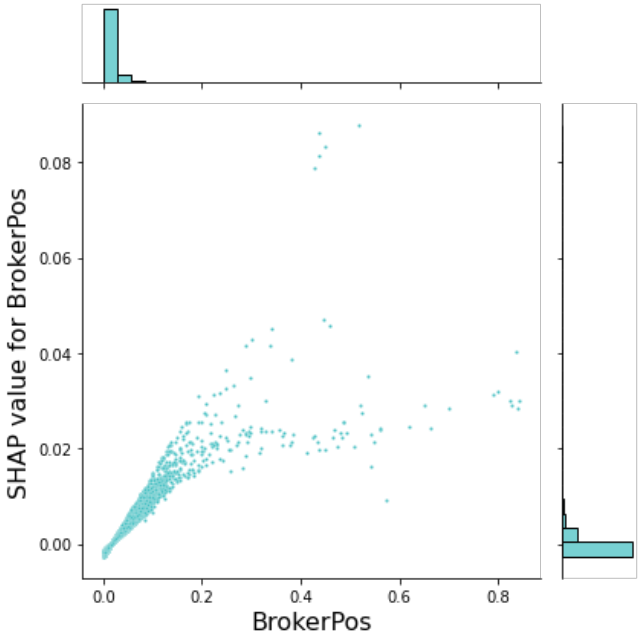


Notes: Dependence plot of SHAP values and *Leverage*.

Figure 6 illustrates the dependence plot of *BrokerPos*. Higher *BrokerPos* values tend to produce positive SHAP values. From values of *BrokerPos* from 0 to 0.2, the SHAP value accelerates upward. This is in line with Maksimovic and Zechner (1991), MacKay and Phillips (2005), and Nettayanun (2014) which find that the distance of broker expense compared to peers can positively influence ceding levels.

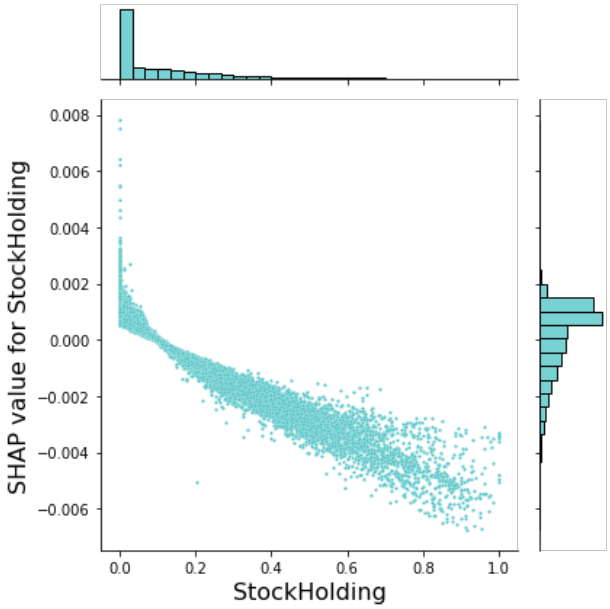
The seventh most important feature from Figure 7 is *StockHolding* which has SHAP values that are either negative or close to zero. However, some insurers with very low stock holdings have positive SHAP values. Figure 7 shows that insurers with more stock holdings have lower ceding levels, while insurers with very low stockholdings tend to have higher ceding levels. This finding does not support MacKay and Phillips (2005) substitution of risk between stockholding and reinsurance purchases. It also slightly differs from Nettayanun (2014) and MacKay and Phillips (2005) that find a positive significant relationship between stock holding and ceding levels. Figure 5 indicates that a large part of the plot has negative SHAP values, suggesting an overall negative relationship between stockholding and ceding levels.

Figure 6. Dependence Plot for *BrokerPos*



Notes: Dependence plot of SHAP values and *BrokerPos*.

Figure 7. Dependence Plot for *StockHolding*

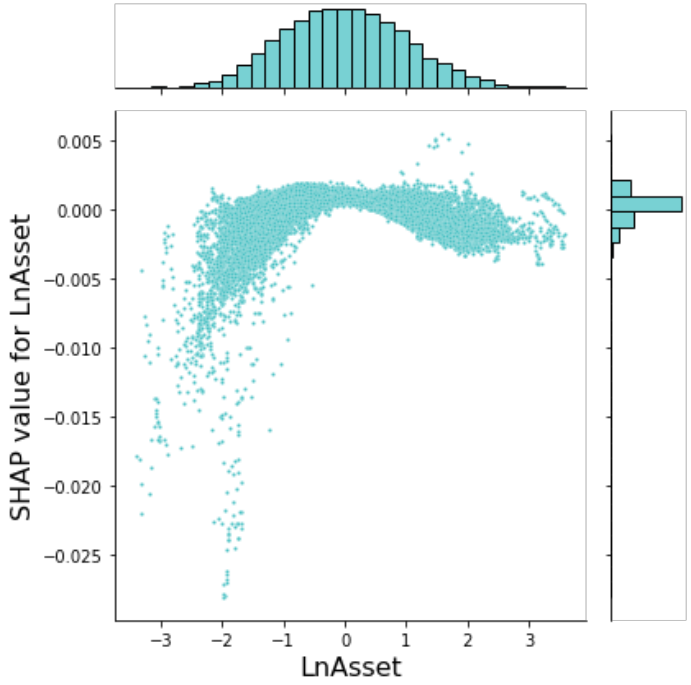


Notes: Dependence plot of SHAP values and *StockHolding*.

The eighth important feature for ceding insurance is firm size, *LnAsset*, as shown in Figure 8. The relationship between size and SHAP values are negative for small insurers. For medium-sized insurers, SHAP values are close to zero or slightly positive. For larger firms, SHAP values

decrease to negative values and then increase back to positive values. Therefore, the size effects on reinsurance purchases do not appear to be conclusively negative or positive, contrary to the conclusions of Froot, Scharfstein, and Stein (1993), Tufano (1996), Stulz (1996), and Liu and Parlour (2009).

Figure 8. Dependence Plot for *LnAsset*

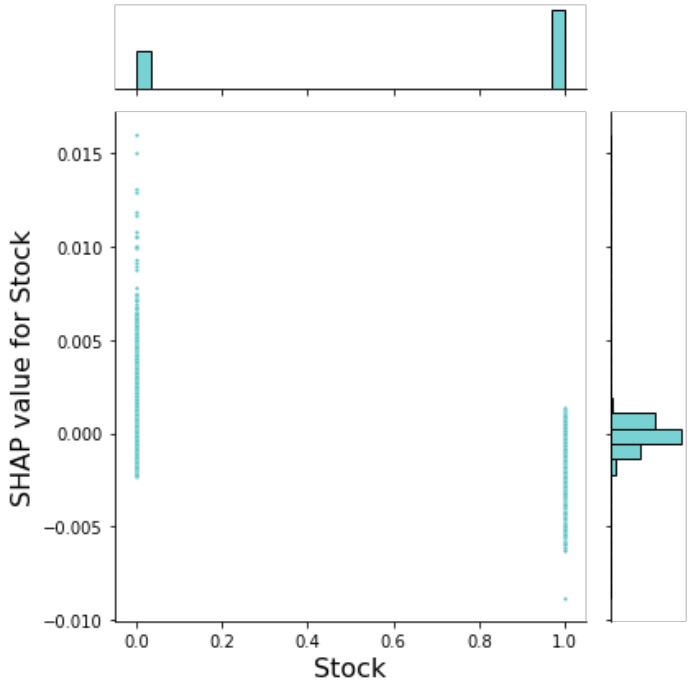


Notes: Dependence plot of SHAP values and *LnAsset*.

Figure 9 illustrates the relationship between ownership structure and ceding levels. For stock ownership companies, indicated by 1, the SHAP values range from positive to negative. The range of SHAP values is wider for other types of ownership structures, showing both positive and negative SHAP values as well. This indicates that there is no clear conclusion about the relationship between ownership structure and reinsurance purchases.

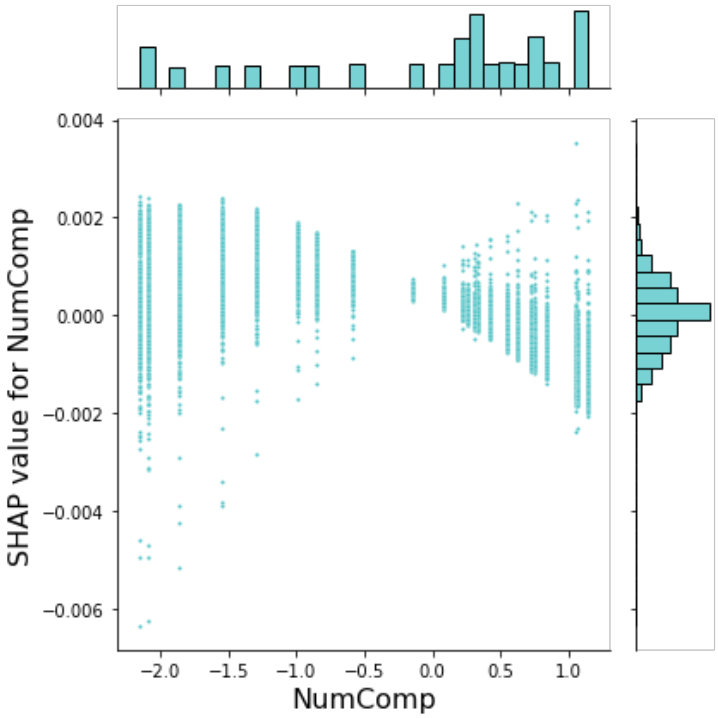
Figure 10 illustrates how the intensity of competition can influence reinsurance demand. It shows a wide range of both positive and negative SHAP values when there are either a low or high number of firms in the industry, indicating both increasing and decreasing reinsurance demand at these levels of competition. However, the influence tends to be close to zero when there is a medium number of firms in the industry, suggesting that this level of competition intensity does not influence reinsurance purchases. If we employ ordinary least squares and panel data regression, competition intensity might be found to be insignificant. The SHAP analysis provides more insight into how competition intensity impacts reinsurance purchases.

Figure 9. Dependence Plot for *Stock*



Notes: Dependence plot of SHAP values and *Stock*.

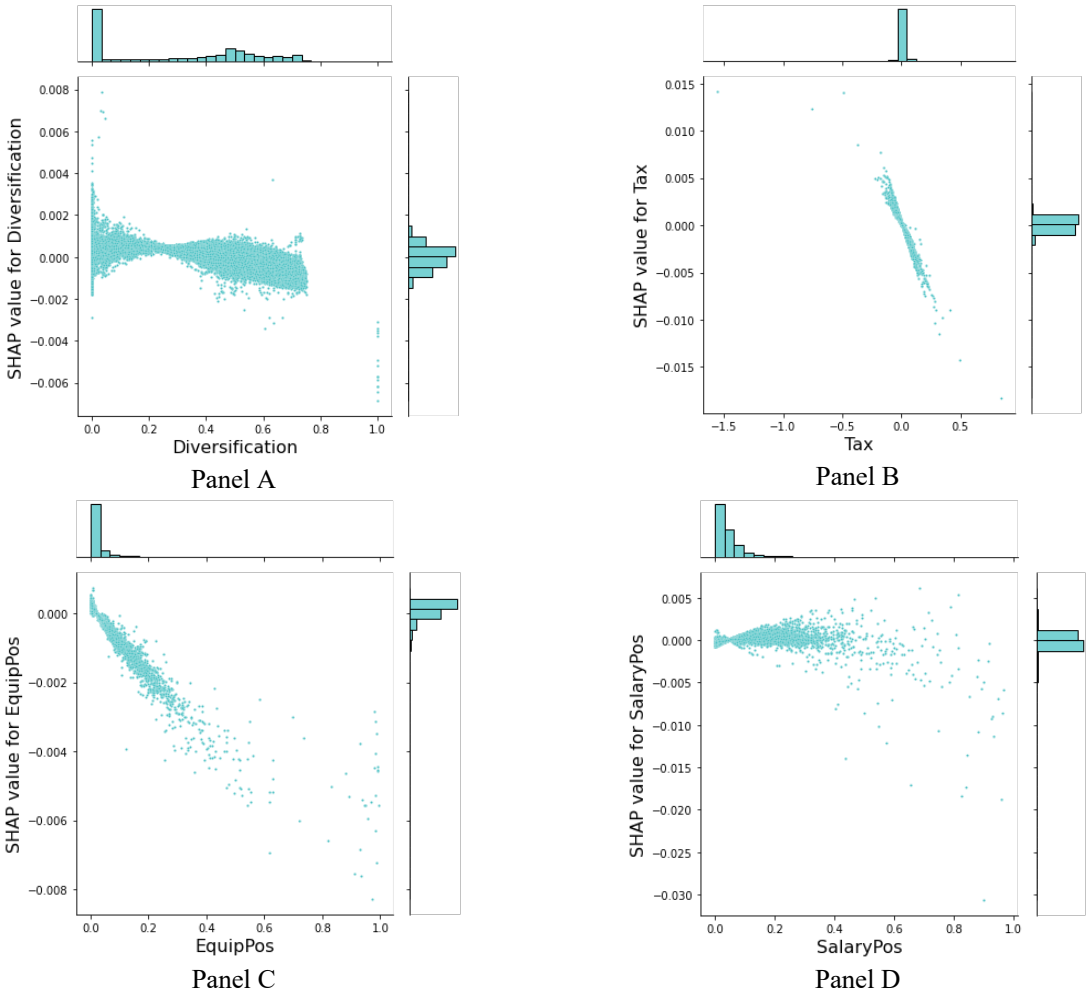
Figure 10. Dependence Plot for *NumComp*

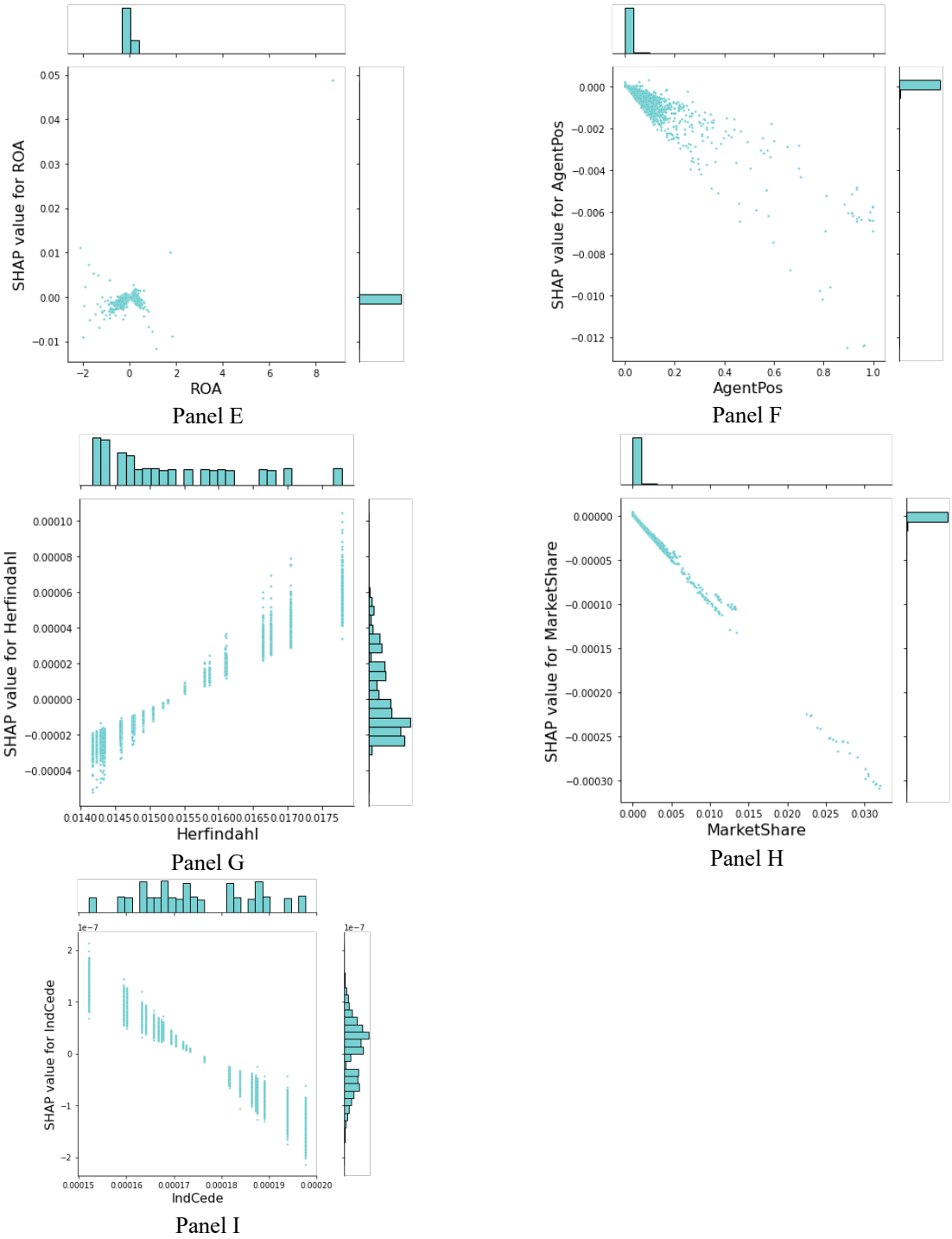


Notes: Dependence plot of SHAP values and *NumComp*.

The other features ranked from 11 to 19 are shown in Figure 11. There are some interesting patterns in the dependence plot of *Herfindahl* and *IndCede*. The relationship between these variables and ceding levels, based on SHAP values, centers around zero. However, very low and very high values of these variables result in a wider range of SHAP values. Therefore, the relationship between these features and ceding levels is not conclusive for *Herfindahl* or *IndCede*. *EquipPos*, *SalaryPos*, and *AgentPos* also exhibit interesting patterns in their SHAP values. Very low values of *EquipPos*, *SalaryPos*, and *AgentPos* produce SHAP values close to zero, but the range of SHAP values become wider for higher values in these variables. Consequently, the conclusion regarding a negative or positive relationship with ceding for these features is also not conclusive.

Figure 11. Dependence Plot for the Other Features





Notes: Dependence plots of SHAP values and different features. Panel A is for feature *Diversification*. Panel B is for feature *Tax*. Panel C is for feature *EquipPos*. Panel D is for feature *SalaryPos*. Panel E is for feature *ROA*. Panel F is for feature *AgentPos*. Panel G is for feature *Herfindahl*. Panel H is for feature *MarketShare*. Panel I is for feature *IndCede*.

Overall, the SHAP framework provides valuable insights into the relationship between variables/features that have been studied in previous literature. First, the framework offers a ranking of the importance of each variable's impact on ceding level. Second, it provides a more

comprehensive understanding of the relationships between the variables and ceding levels, conditioned on the value of each variable itself.

Significance Tests of Features

To our knowledge, this study is the first to explore the significance of each variable's impact on reinsurance demand/ceding using machine learning. We follow Horel and Giesecke (2019) to show the significance of each variable included in the model. The method is the latest development in explainable AI for explaining any machine learning approaches. Horel and Giesecke (2020) also have significance tests for neural networks, but their method is limited to single-layer neural network. Another significance tests methodology, single feature introduction test (SFIT), proposed by Horel and Giesecke (2019), can be implemented using any machine learning approaches. There are two advantages to using this method. First, it does not assume the distribution of the features and is applicable to various model specifications. Second, it can be applied to both regression and classification analysis. Furthermore, it accounts for the significance of correlated features. For example, if X_1 and X_2 are highly correlated, other methods might identify only one of them as significant. However, SFIT can identify both variables as significant. This is accomplished by comparing the loss between a model that includes only a constant and a model that includes the feature of interest. After this is completed for all features, the median (m-statistics) loss is calculated from all of the losses accumulated from the data. The significance of each feature can then be determined using the distribution of the m-statistics that are greater than zero. This distribution follows a binomial distribution with n is the number of test observations and probability of ½. The implementation of this method can be found on GitHub[7].

Table 2. Significant Variables

Variable	m-statistics
Leverage	0.00872
CL	0.00788
CS	0.00438
PS	0.00423
Stock	0.00240
StockHolding	0.00030
ROA	0.00001

Note: This table reports SFIT m-statistics values for significant variables at a 0.1 level of significance.

Table 2 presents the significant variables from the SFIT test at a 0.1 level of significance, ranked in descending order based on m-statistics. *Leverage* is identified as the most significant variable, followed by *CL*, *CS*, *PS*, *Stock*, *StockHolding*, and *ROA*, respectively. Notably, six variables to include *Leverage*, *CL*, *CS*, *PS*, *Stock*, and *StockHolding* are also included in the top ten SHAP values. Therefore, these variables are significant variables that can be used to explain reinsurance purchases using the SHAP and significant-test frameworks. We add year dummy variables to account for the panel data. The results indicate three significant variables: *CS*, *PS*, and

[7] https://github.com/fintechstanford/SFIT

StockHolding. The variables from Table 2 that are not significant after including year dummy variables are *Leverage*, *Stock*, and *CL.* Therefore, we can identify the change in results for the SFIT test when adding year dummy variables.

VI. Conclusions

This study employs an explainable AI (XAI) framework to explore the strategic determinants that affect reinsurance levels in the property and casualty industry. Unlike previous studies that use various econometric approaches to explain reinsurance purchases, we utilize the SHAP library coupled with DeepExplainer to analyze how features explain reinsurance purchases. The Shapley value approach provides new insights into reinsurance purchase behavior. First, the methodology does not assume that the relationship between dependent and independent variables is linear, as many other econometric methodologies do. Second, the SHAP library offers deeper insights into how each variable influences hedging decisions based on SHAP values. This allows us to examine how each characteristic of the independent variable affects hedging levels through observation levels.

The top ten most important variables that impact reinsurance purchases are product types (*commercial long-tail*, *commercial short-tail*, *personal short-tail*, and *personal long-tail*), leverage, brokerage expense position compared to peers, stock holding percentage in the portfolio, size, ownership structure, and number of companies in the industry. Additionally, we use significance tests from Horel and Giesecke (2019) for each variable. Six variables that are significant are also among the top ten SHAP values, including leverage, ownership structure, product types (*commercial long-tail, commercial short-tail, personal short-tail*), and stock holding percentage in the portfolio.

Stakeholders in the insurance industry can benefit from this study in several ways. First, managers and executives of insurers can explore how other insurers adjust their reinsurance purchasing decisions based on strategic variables, allowing them to refine their own strategic reinsurance purchases. Second, investors can gain a deeper understanding of how insurers use reinsurance when analyzing their investment choices. Risk management is influenced not only by financial decisions but also by competition between peers. Third, regulators can gain insights into how insurers adjust their reinsurance purchases based on various strategic decisions, recognizing that insolvency risks may come from factors beyond financial consideration alone. By leveraging computation power, it is now possible to implement alternative methods that complement classical econometric approaches, providing more comprehensive insights into financial studies.

References

Adam, T., Dasgupta, S., & Titman, S. (2007). Financial constraints, competition, and hedging in industry equilibrium. *The Journal of Finance, 62*(5), 2445–2473.

Adam, T. R., & Nain, A. (2013). Strategic risk management and product market competition. In *Advances in Financial Risk Management: Corporates, Intermediaries and Portfolios* (pp. 3-29). London: Palgrave Macmillan UK.

Allayannis, G., & Ihrig, J. (2001). Exposure and markups. *The Review of Financial Studies, 14*(3), 805-835.

Amini, S., Elmore, R., Öztekin, Ö., & Strauss, J., (2021). Can machines learn capital structure dynamics? *Journal of Corporate Finance, 70*, 102073.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics, 151*, 103745.

Brockett, P. L., Cooper, W. W., Golden, L. L., & Pitaktong U. (1994). A neural network method for obtaining an early warning of insurer insolvency. *Journal of Risk and Insurance, 61*(3), 402–424.

Brockett, P. L., Golden, L. L., Jang, J., & Yang, C. (2006). A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction. *Journal of Risk and Insurance, 73*(3), 397–419.

Bubb, R., & Catan, E. M. (2022). The party structure of mutual funds. *The Review of Financial Studies 35*(6): 2839–2878.

Caporale, G. M., Cerrato, M., & Zhang, X. (2017). Analysing the determinants of insolvency risk for general insurance firms in the UK. *Journal of Banking & Finance, 84*, 107-122.

Choi, B. P., & Weiss, M. A. (2005). An empirical investigation of market structure, efficiency and performance in property-liability insurance. *Journal of Risk and Insurance, 72*(4), 635–673.

Cole, C. R., & McCullough, K. A. (2006). A reexamination of the corporate demand for reinsurance. *Journal of Risk and Insurance, 73*(1), 169–192.

Colla, P., Ippolito, F., & Li, K. (2013). Debt specialization. *The Journal of Finance 68*(5): 2117-2141.

Erel, I., Stern, L. H., Tan, C., & Weisbach, M. S. (2021). Selecting directors using machine learning. *The Review of Financial Studies 34*(7): 3226-3264.

Froot, K. A. (2007). Risk management, capital budgeting, and capital structure policy for insurers and reinsurers. *Journal of Risk and Insurance, 74*(2), 273–299.

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1993). Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance, 48*(5), 1629– 1658.

Froot, K. A., & Stein, J. C. (1998). Risk management, capital budgeting and capital structure policy for financial institutions: An integrated approach. *Journal of Financial Economics, 47*(1), 55–82.

Graham, J. R., & Smith, C. W. (1999). Taxes incentives to hedge. *The Journal of Finance, 54*(6), 2241–2262.

Griffin, J. M., Hirschey, N., & Kruger, S. (2023). Do municipal bond dealers give their customers "fair and reasonable" pricing? *The Journal of Finance, 78*(2): 887–934.

Gu, S., Kelly, & B., Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios.

Applied Stochastic Models in Business and Industry, 33(1), 3–12.
Hejazi, S. A., & Jackson, K. R. (2016). A neural network approach to efficient valuation of large portfolios of variable annuities. *Insurance: Mathematics and Economics, 70*, 169–181.
Horel, E., & Giesecke K. (2019). Computationally efficient feature significance and importance for machine learning models. *arXiv preprint arXiv:1905.09849*.
Horel, E., & Giesecke K. (2020). Significance tests for neural networks. *Journal of Machine Learning Research, 21*(227), 1-29.
Hornuf, L., & Schaefer, P., (2025). Artificial intelligence and machine learning in corporate finance. *Available at SSRN*.
Hoyt, R. E., & Khang, H. (2000). On the demand for corporate property insurance. *Journal of Risk and Insurance, 67*, 91–107.
Hoyt, R. E., & Liebenberg, A. P. (2011). The value of enterprise risk management. *Journal of Risk and Insurance, 78*(4), 795–822.
Korangi, K., Mues, C., & Bravo, C. (2023). A transformer-based model for default prediction in mid-cap corporate markets. *European Journal of Operational Research 308*(1): 306–320.
Lamm-Tennant, J., & Starks, L. T. (1993). Stock versus mutual ownership structures: The risk implications. *The Journal of Business, 66*(1), 29-46.
Leverty, J. T., & Grace, M. F. (2010). The robustness of output measures in property-liability insurance efficiency studies. *Journal of Banking & Finance, 34*(7), 1510–1524.
Liebenberg, A. P., & Sommer, D. W. (2008). Effects of corporate diversification: Evidence from the property–liability insurance industry. *Journal of Risk and Insurance, 75*(4), 893-919.
Liu, T., & Parlour, C. A. (2009). Hedging and competition. *Journal of Financial Economics, 94*(3), 492–507.
Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4768-4777.
Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, E. D. D. E., Low, D. K., Newman, S., Kim, J., et al. (2018). Explainable machine learning predictions to help anesthesiologists prevent hypoxemia during surgery. *Nature Biomedical Engineering, 2*(10), 749-760.
MacKay, P., & Phillips, G. M. (2005). How does industry affect firm financial structure? *The Review of Financial Studies, 18*(4), 1433–1466.
Maksimovic, V., & Zechner, J. (1991). Debt, agency costs, and industry equilibrium. *The Journal of Finance, 46*(5), 1619–1643.
Masters, T. (1993). *Practical Neural Network Recipes in C++*. New York: Academic Press.
Mayers, D., & Smith, C. W. (1981). Contractual provisions, organizational structure, and conflict control in insurance markets. *The Journal of Business, 54*(3), 407-434.
Mayers, D., & Smith, C. W. (1987). Corporate insurance and the underinvestment problem. *Journal of Risk and Insurance, 54*(1), 45–54.
Mayers, D., & Smith, C. W. (1990). On the corporate demand for insurance: Evidence from the reinsurance market. *The Journal of Business, 63*(1), 19–40.
Mello, A. S., & Ruckes, M. E. (2005). Financial hedging and product market rivalry. *Available at SSRN 687140*.
Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review, 48*(3), 261–297.
Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics, 5*(2), 147–175.

Nain, A. (2004). The strategic motives for corporate risk management. *Available at SSRN 558587*.

Nettayanun, S. (2014). *Essays on strategic risk management*. Georgia State University.

Phillips, R. D., Cummins, J. D. & Allen F. (1998). Financial pricing of insurance in the multiple-line insurance company. *Journal of Risk and Insurance, 65*(4), 597- 636.

Pottier, S. W., & Sommer D. W. (1997). Agency theory and life insurer ownership structure. *Journal of Risk and Insurance, 64*(3), 529-543.

Powell, L. S., & Sommer, D. W. (2007). Internal versus external capital markets in the insurance industry: The role of reinsurance. *Journal of Financial Services Research, 31*, 173–188.

Rampini, A. A., Sufi, A. & Viswanathan, S. (2014). Dynamic risk management. *Journal of Financial Economics, 111*(2), 271–296.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

Sirignano, J., & Giesecke, K. (2019). Risk analysis for large pools of loans. *Management Science*, 65(1), 107–121.

Sirignano, J., Sadhwani, A. & Giesecke, K. (2016). Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470.*

Smith, C. W., & Stulz, R. M. (1985). The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis, 20*(4), 391–405.

Sommer, D. W. (1996). The impact of firm risk on property-liability insurance prices. *Journal of Risk and Insurance, 63*(3), 501–514.

Stulz, R. M. (1996). Rethinking risk management. *Journal of Applied Corporate Finance, 9*(3), 8–25.

Tufano, P. (1996). Who manages risk? An empirical examination of risk management practices in the gold mining industry. *The Journal of Finance, 51*(4), 1097–1137.

Winter, R. A. (1994). The dynamics of competitive insurance markets. *Journal of Financial Intermediation, 3*(4), 379–415.

Wüthrich, M. V. (2019). Bias regularization in neural network models for general insurance pricing. *European Actuarial Journal, 10*, 1–24.

Wüthrich, M. V., & Merz, M. (2019). Yes, we CANN! ASTIN Bulletin: *The Journal of the IAA, 49*(1), 1–3.

Appendix

Table A1. Product Types

Personal Short-tail (*PS*)	**Personal Long-tail (*PL*)**	**Commercial Short-tail (*CS*)**	**Commercial Long-tail (*CL*)**
Auto physical damage	Private passenger auto liability	Allied lines	Aircraft
Farmowners multiple peril		Boiler and Machinery	Commercial auto liability
Homeowners multiple peril		Burglary and Theft	Excess workers compensation
		Commercial multiple peril	Medical professional liability-claims-made
		Credit accident and health	Medical professional liability-occurrence
		Credit	Other liability-claims
		Earthquake	Other liability-occurrence
		Fidelity	Other property and casualty
		Financial Guarantee	Product liability-occurrence
		Fire	Product liability-claims
		Group accident and health	Reinsurance (assumed liability)
		Inland marine	Warranty
		International	Workers compensation
		Mortgage guarantee	
		Ocean marine	
		Other accident and health	
		Reinsurance (assumed financial)	
		Reinsurance (assumed property)	
		Surety	

Note: This table categorizes each line of business into each product type.

The Effect of AI on CSR and ESG Ethics

Hoje Jo[1]

Abstract:

Integrating Artificial Intelligence (AI) into business operations has brought about transformative changes, potentially reshaping industries and improving efficiencies. However, this technological advancement also presents new ethical challenges, particularly concerning Environmental, Social, and Governance (ESG) principles. As businesses increasingly adopt AI, balancing innovation with ethical responsibility becomes crucial. This paper explores the intersection of AI, CSR, ESG, and business ethics, emphasizing how AI can enhance sustainability, social equity, and governance standards. It examines the potential risks, such as algorithmic bias, data privacy concerns, and environmental impacts of AI infrastructure. It discusses strategies for mitigating these risks through ethical AI design, transparent governance structures, and adherence to ESG guidelines. Ultimately, the paper argues for the development of frameworks that ensure AI technologies contribute positively to societal well-being while aligning with the core values of corporate responsibility, ESG, and sustainability.

JEL Classification: G0, G19, G30
Keywords: Artificial Intelligence, ESG, CSR, Business Ethics

I Introduction

The Grand View Research estimates the global artificial intelligence market size at USD 196.63 billion in 2023. It is projected to grow at a capital asset growth rate (CAGR) of 36.6% from 2024 to 2030 (www.grandviewresearch.com) (see Figure 1). "By 2026, AI in the global ESG market is projected to reach approximately 7.8 billion USD." In a world filled with data, the use cases for Artificial Intelligence (AI) in financial processes are expansive and never-ending (www.grandviewresearch.com). Over the next few years up to 2030, the market size of AI is expected to grow 788.64% from $207.9 billion to $1.8 trillion worldwide (see Figure 2). According to the data from Statista and Finbold, the global Artificial Intelligence (AI) market size is projected to reach approximately $1.8 trillion by 2030, showcasing a significant compound annual growth rate (CAGR) during this period

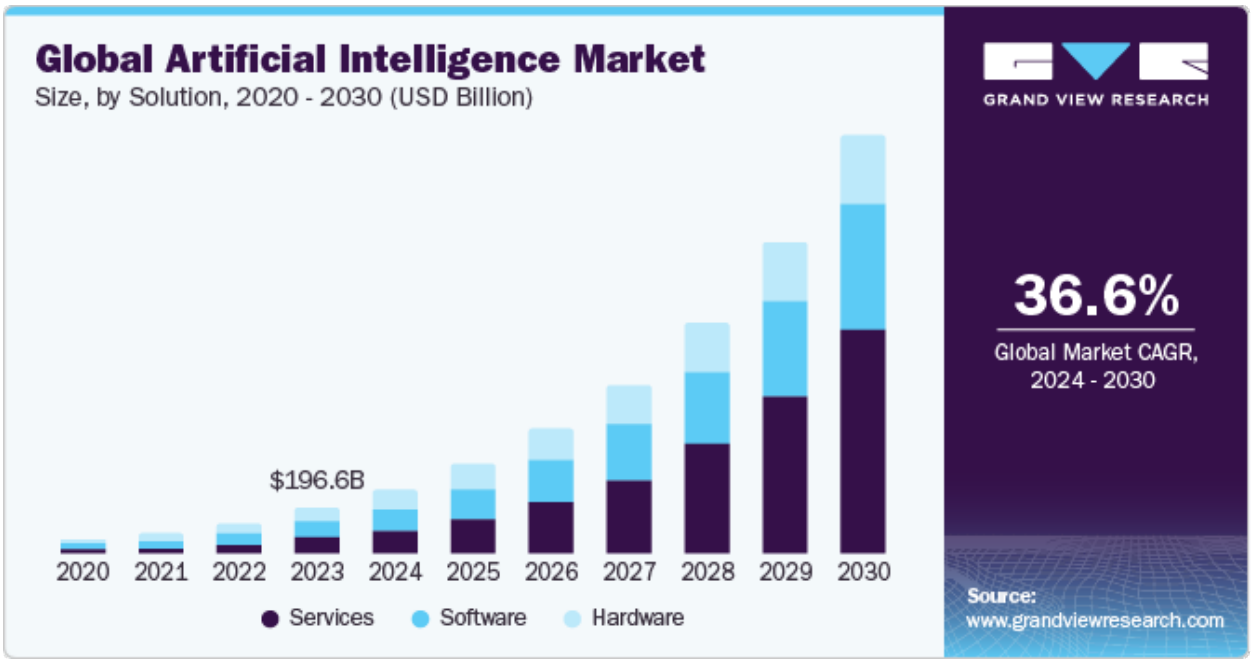
The increasing use of AI-powered technology profoundly affects various levels and sectors of businesses and finance, as shown in a recent survey paper on ESG and AI in finance (Lim, 2024). This development increases efficiency and productivity, fuels expansion and innovation, and reduces routine tasks' workloads. The new technology is a transformative force that will redefine societal norms, influence business practices, production and trade, and shape our future. It will increase efficiency and affect the quality of work; it will change collaborative structures and power relations in organizations, eliminate business models, and create new ones.

[1] Hoje Jo (hjo@scu.edu), Santa Clara University, Santa Clara, CA, USA. The author appreciates valuable comments from the anonymous reviewer, Sanjiv Das, Eun-Pyo Hong, and Ann Skeet.

Tech giants like Amazon.com, Inc.; Google LLC; Apple Inc.; Facebook; International Business Machines Corporation; and Microsoft are investing significantly in AI research and development (R&D), thus increasing the artificial intelligence market cap (www.grandviewresearch.com).

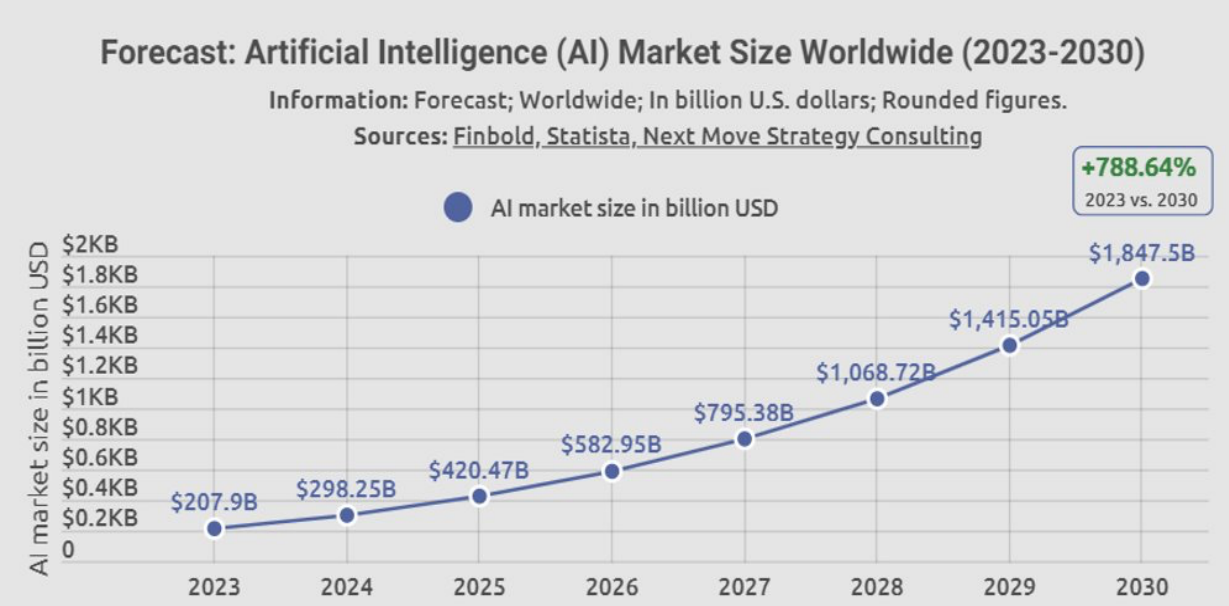
Figure 1 Artificial Intelligence Market Size & Trends

The continuous research and innovation directed by tech giants drive the adoption of advanced technologies in industry verticals, such as automotive, healthcare, retail, finance, and manufacturing. For instance, in December 2023, Google LLC launched 'Gemini', a large language AI model, made available in three sizes: Gemini Nano, Gemini Pro, and Gemini Ultra. Gemini stands out from its competitors due to its native multimodal characteristic (www.grandviewresearch.com).



In recent years, integrating Artificial Intelligence into Environmental, Social, and Governance (ESG) initiatives has emerged as a transformative force in the corporate world (Lim, 2024). As businesses face pressure from investors, consumers, and regulators to operate sustainably and ethically, AI provides a solid toolkit to meet these demands, revolutionize how companies enhance their ESG reporting, and uncover opportunities. The rise of AI in ESG practices is not merely a tech advancement but rather a strategy to reduce environmental and social risks, foster stakeholder engagement, and optimize operational efficiency. However, it is imperative to recognize that AI still has the challenges of algorithmic bias, data accuracy, and the need for human oversight, which is dangerous when addressing ESG issues.

Figure 2 The Artificial Intelligence market size forecast worldwide (2023-2030) (Source: Finbold, Statista, Next move strategy consulting).



This paper aims to further understand the impact of AI on CSR and ESG ethics by examining specific case studies and empirical evidence. The intersection of AI, CSR, ESG factors, and business ethics is rapidly evolving, with recent publications shedding light on their interplay. The impact of AI on CSR and ESG ethics is profound and multifaceted, as AI plays an increasingly prominent role in shaping how organizations pursue sustainability, fairness, and responsible governance. Francis (2024) discusses how AI integration and changing regulations will shape corporate ethics, compliance, and ESG by 2025. He emphasizes the importance of leadership in fostering ethical cultures and navigating complex ESG requirements.

By leveraging advanced data analytics and machine learning algorithms, Adeoye et al. (2024), using extensive ESG-related datasets and extracting actionable insights, have suggested AI empowers investors to analyze, and identify investment opportunities aligned with sustainability objectives. AI has emerged as a transformative force in ESG investing, enhancing portfolio performance Adeoye et al. (2024). Wang and Wang (2025) suggest that AI's potential to drive operational efficiency and support regulatory compliance processes during turbulence offers a guide for practitioners grappling with the intricacies of restructuring and turnaround efforts.

Unlike the above studies, we aim to contribute to the AI literature in finance and suggest the relationship between AI and business ethics, AI and CSR, and AI and ESG issues to examine whether AI-driven strategies have successfully (or unsuccessfully) reduced risks, improved (or deteriorated) financial performance metrics, and provided companies with a competitive (or less competitive) edge. Through a balanced analysis of opportunities and failures, we aim to offer valuable insights for finance decision-makers considering the integration of AI into their ESG strategies ethically and effectively.

This paper uniquely contributes to the AI literature in finance by adding how the AI-driven strategies of business ethics, CSR, and ESG are functionally and qualitatively related to firm risk, performance, and competitive positions. To discuss how businesses can integrate AI to improve ESG performance while adhering to ethical principles, simple modeling about the impact of AI on

CSR, ESG, and business ethics are proposed. We also discuss the limitations in the context of ESG and business ethics, including bias and inequality, lack of transparency (black-box problem), ethical considerations in AI deployment, over-reliance on AI for decision-making, data quality and availability, regulatory and legal risks, environmental impact of AI, human displacement and labor concerns, long-term sustainability and unintended consequences, corporate governance and accountability, and lack of universal standards.

II Literature Review

AI's ability to concentrate large amounts of information into consumable, actionable intelligence is a crucial reason for its high adoption rate by many finance decision-makers. These cases may include how AI can predict future trends, optimize current infrastructure based on a metric, and reduce environmental footprint. AI can also enhance supply chain and process logistics by introducing traceability, enabling companies to find ways to reduce their environmental and social risks while strengthening societal well-being (Statman, 2024).

ESG and CSR topics have grown substantially recently; thus, it is crucial to differentiate ESG from CSR. CSR refers to businesses' sustainability strategies to ensure the company is carried out ethically. In contrast, ESG practices are criteria used to measure a company's overall sustainability. CSR and ESG are related but not the same. For instance, consider a paper bag manufacturer that wants to implement CSR and ESG policies. CSR can be incorporated by communicating internally and in press releases that the company is committed to being more sustainable and responsible. ESG builds on that foundation with measurable goals such as a 30% increase in recycled materials within five years and planting one million trees in 10 years. Thus, we can investigate ESG practices to evaluate how well a company adheres to its sustainability and corporate responsibility goals. Harder-to-measure indicators under the CSR banner include greater employee awareness of the company's environmental impact or internal and external messaging about sustainable practices. Whereas CSR is the ideal and gives context to sustainability agendas and corporate responsibility culture, ESG is the action and measurable outcome. To simplify, CSR can be thought of as the qualitative side and ESG as the quantitative side.[2]

ESG is growing but contentious. The concept of investing based on the criteria of ESG factors has proliferated since the term ESG was coined in 2005, and now more than 90% of S&P 500 companies provide ESG information (Perez, et al., 2022). Scholars and finance professionals wrote over 2,000 papers examining the pros and cons of investing based on ESG standards. Their findings substantiated the fundamental principles of ESG investing, with approximately 90% of the studies indicating a favorable relation between ESG and corporate financial performance (Friede et al., 2015). Previous studies have focused on the benefit side and found that ESG activities can increase firm performance or value (Cui et al., 2025; Jo and Harjoto, 2011, 2012; Servaes and Tamayo, 2013). Other studies provide contrary or no evidence (Buchanan, Cao, and Chen, 2018; Masulis and Reza, 2015; Matsumura et al., 2014). Different studies focus on the cost side of ESG engagement, including the opportunity cost of ESG investment and the potential waste

[2] CSR usually encompasses how a company will approach its internal framework of sustainability plans and responsible cultural influence, whereas ESG relates to the assessable outcome concerning a company's overall sustainability performance. Investors typically look at this framework when making financial decisions regarding the company in question (Oetzel, 2023). https://kogod.american.edu/news/csr-or-esg

of company resources (Lioui and Sharma, 2012) or lack of effective ESG communication through shared governance (Jo, Chun, and Song, 2025).

At the same time, the world is challenged by the evolving landscape of climate change and political transformations. Increasing public expectations and changing power relations threaten the stability of international trade relations and supply chains, and businesses increasingly must deal with many societal expectations. To review some of those recent AI applications, we will review a few AI-related studies. Taddeo and Floridi (2018) argue that Artificial intelligence (AI) is not just a new technology that requires regulation. It is a powerful force reshaping daily practices, personal and professional interactions, and environments. For the well-being of humanity, this power must be used as a force for good. They claim that ethics is key in ensuring AI regulations harness their potential while mitigating risks.

Francis (2024) discusses how AI integration and changing regulations will shape corporate ethics, compliance, and ESG. He emphasizes the importance of leadership in fostering ethical cultures and navigating complex ESG requirements. Adeoye et al. (2024) claim that AI empowers investors to analyze extensive ESG datasets, extract actionable insights, and identify investment opportunities aligned with sustainability objectives. AI has emerged as a transformative force in ESG investing, enhancing portfolio performance.

Abay (2022) examines the role of independent third-party ESG assurance in signaling higher ESG performance. While testing the hypothesis, a linear regression was applied using data from Thomson Reuters ESG scores and the global reporting initiative database from a sample of 645 unique European firms for 2012-2017. Firms with third-party assurance have a significantly higher ESG performance than those without assurance. He offers new evidence on the signaling value of an independent third-party ESG assurance in differentiating ESG performances. He confirms the incentive high-performing firms could use to separate from their counterparts with poor performance in a separate equilibrium.

Additionally, the University of Texas at Dallas data shows that “idiosyncratic risk decreases with CSR due to the intangibles improved by CSR” (Li, Li, & Sethi, 2021). Other examples include increasing efficiency by improving corporate culture and reducing company-specific risk by maintaining a high CSR reputation. Despite CSR's intangible benefits, AI uses concrete data to address CSR issues such as privacy concerns and data security. The AI tool should ensure fairness and transparency to its stakeholders, especially disadvantaged and marginalized consumers who are “more vulnerable to the harms of digital technologies, including privacy risks, fraud, and digital surveillance” (Du & Sankar, 2023).

Chouaibi et al. (2022) examine whether environmental disclosure (ED) practiced by firms listed on the ESG index affects their financial performance (FP) using the moderating effect of social and ethical practices. The results show a positive and significant relationship between environmental disclosure (ED) and financial performance (FP). This implies that a substantial environmental disclosure increases financial performance while a weak one decreases it. Furthermore, they suggest a moderating effect of social and ethical practices in the link between environmental disclosure and the firm’s financial performance. These findings provide insights to stakeholders and regulators on integrating more social and environmental regulations to promote sustainability.

Investors have little guidance on how to meaningfully integrate ESG ratings into their investment decisions. Moreover, the opinions on whether ESG integration will reap financial benefits vary dramatically amongst academics and ESG professionals. To address this debate, Pederson et al. (2021) develop a theory that shows both - the potential costs as well as benefits of

ESG integrated investing. To conclude whether there is a relation between ESG and financial performance, Friede et al. (2015) conduct an extensive review of over 2000 empirical studies. This correlation has been debated since the 1970's. They claim that around 90% of all the studies find a positive ESG and financial performance relationship. They also find that this relationship appears to be stable over time.

Brusseau (2021) argues that the main issue with AI is related to data ownership and how companies use individuals' data, mainly whether AI usage benefits us or limits our self-determination. He contends that AI-intensive companies should be evaluated based on their impact on individuals rather than demographic segments or collectives. Brusseau proposes an alternative AI human impact model for evaluating AI companies, which utilizes a set of AI principles and assigns scores from 0 to 2 based on how well a company addresses these issues.

AI has revolutionized how companies collect, process, and analyze ESG-related data. Through advanced algorithms, AI can efficiently sift through vast amounts of information from internal databases and external sources, extracting relevant ESG metrics and classifying them into standardized categories, reducing manual effort and minimizing human error (Sandford, 2024).

III Simple Model of AI, ESG, and Business Ethics

As AI emerges as a pivotal tool in decision-making, concerns about biases and challenges have gained prominence. The inherent biases in training data and algorithms can perpetuate societal biases. Understanding and addressing these challenges is crucial for fostering equitable and effective AI-driven decision-making processes.

Indeed, the introduction of AI represents a significant shift in ethics. AI usage's various ethical problems stem from the lack of regulation and privacy, programmable biases, and job displacement. The absence of comprehensive regulatory frameworks has allowed for the unchecked development and deployment of AI technologies, raising concerns about the potential misuse of sensitive data and the infringement of individual privacy rights. Additionally, programmable biases within AI algorithms can perpetuate and even exacerbate existing social inequalities, as these systems may unintentionally discriminate against certain demographic groups. When further analyzing the integration of AI in investment decisions, ethical considerations that demand careful examination are raised. AI usage's ethical problems include its impact on the environment, its use of materials in training systems without consent, credit or compensation, the ease with which it creates deepfakes and allows for financial fraught, the unequal access to AI and the associated imbalance of power this creates, the gender gap that is emerging in its use, the impact on future generations without their input--the list is really quite long. It might be best to characterizes the ethical issues we have called out as a sample of ethical issues that arise.

We consider various relationships between AI, CSR, ESG, and business ethics. To discuss how businesses can integrate AI to improve ESG performance while adhering to ethical principles, simple modeling about the impact of AI on CSR, ESG, and business ethics should be useful. We first consider the impact of AI on the business ethics equation.

$$AI\ Ethics\ Score\ (AES) = f(Transparency, Accountability, Fairness, Privacy, Bias, Environment) \quad (1)$$

Where Transparency (T): The clarity with which AI decisions and processes are communicated; Accountability (A): The extent to which responsible parties are held accountable for AI outcomes; Fairness (F): AI systems being free from discrimination and biased outcomes; Privacy (P): The adherence to privacy laws and ethical data usage; and Bias (B): How well the AI system avoids introducing or amplifying biases. A higher AES score indicates better alignment with ethical standards in AI usage.

We next consider AI's contribution to ESG goals:

$$\textit{ESG Score (ESGS)} = w_1 * \textit{Environmental Impact} + w_2 * \textit{Social Impact} + w_3 * \textit{Governance Impact} \quad (2)$$

Where environmental impact (E): How AI is being used to reduce resource consumption, waste, carbon emissions, or enable sustainable practices (e.g., AI in energy optimization); Social Impact (S): The positive societal effects of AI, such as improving access to education, healthcare, or reducing inequality; and Governance Impact (G): How AI is used to improve corporate governance, risk management, and regulatory compliance. Weights (w_1, w_2, w_3) are assigned based on the organization's focus within these domains.

We next contemplate the cost-benefit analysis of AI for ESG:

$$\textit{Net ESG Benefit (NEB)} = \textit{(AI's Positive ESG Impact)} - \textit{(AI's Ethical/Operational Cost)} \quad (3)$$

Where Positive ESG Impact: Benefits from using AI to meet environmental sustainability, improve social justice, or strengthen governance frameworks; and Ethical/Operational Cost: Expenses or resources spent on ensuring that AI is ethically aligned (e.g., mitigating bias, protecting privacy, ensuring accountability). A positive NEB would indicate that the AI deployment is aligned with business ethics and ESG principles. At the same time, a negative NEB would suggest a potential trade-off between ethical and operational goals. While Kemell and Vakkuri (2023) provide an initial look into the cost of AI ethics and valuable insights from comparisons to software quality, implementing AI ethics remains nascent, and thus, a better empirical understanding of AI ethics is required going forward.

Next, we consider the return on ESG investment (ROESGI):

$$\textit{ROESGI} = \textit{ESG Benefits (Revenue/Cost Savings from ESG)} / \textit{ESG Investments (Capital allocated to ESG Initiatives)} \quad (4)$$

Where ESG benefits: The tangible and intangible returns generated from sustainability efforts, social improvements, and governance actions enabled by AI (e.g., improved brand reputation, customer loyalty, regulatory compliance); and ESG investments: Resources spent on adopting AI systems that support ESG goals, including research and development, ethical audits, and compliance with regulatory frameworks. A higher ROESGI ratio indicates that AI investments to improve ESG outcomes yield significant returns.

Furthermore, we consider AI-driven ESG risk management:

$$\textit{ESG Risk Score (ESGRS)} = f(\textit{Operational Risk, Regulatory Risk, Reputation Risk, Compliance Risk}) \quad (5)$$

Where Operational risk: Risks related to AI systems' reliability, security, and efficiency.
Regulatory risks involve risks associated with AI compliance with environmental laws, labor laws, data privacy laws, and other regulations; Reputation risks are the potential for public backlash or negative media attention regarding AI's misuse (e.g., biased decision-making or environmental harm); Compliance risks are risks of failing to adhere to ESG-related standards and frameworks in the deployment and development of AI. A higher ESG risk score may require mitigation efforts to ensure AI is ethically sound and aligned with ESG principles. Therefore, investors must grapple with ethical dilemmas to ensure that AI deployment aligns with responsible and fair practices within the investment landscape.

IV Framework on AI, Business Ethics, CSR, and ESG

Analyses of AI and Business Ethics

Many different ethical and other adverse concerns arise from using AI in investing. The following analyses are based on our simple model of AI, ESG, CSR, and business ethics. First, there could be a hiring bias. AI algorithms trained on historical hiring data may inherit biases in the hiring process, potentially perpetuating discrimination. Recognizing and mitigating these biases is crucial to promoting diversity and inclusion within the workforce. To make matters worse, AI applications reduce the need for a human labor workforce.

Second, ESG rating bias is possible. AI models evaluating ESG factors may inadvertently incorporate biases, impacting the accuracy and fairness of ESG ratings. Addressing these biases is essential to ensure investment decisions align with ethical and sustainable practices. Third, there is also a geographic bias. Algorithms might inadvertently favor or disadvantage certain geographic regions, impacting the distribution of investments. Ensuring geographic neutrality is vital to avoid reinforcing disparities and to foster a globally equitable investment approach. Fourth, there might be a particular industry concentration. AI models may exhibit biases towards specific industries, leading to an overconcentration of investments in certain sectors. Diversifying investment portfolios and refining algorithms can help mitigate these biases and reduce industry-specific risks.

Fifth, there is a possible overemphasis on short-term metrics. AI models focused on short-term performance metrics may neglect long-term sustainability. Striking a balance between short-term gains and long-term stability is critical for responsible investment decision-making. Sixth, there are data privacy concerns as well. Using vast datasets in AI-driven investment decisions raises privacy concerns. Implementing robust data protection measures is imperative to safeguard sensitive information and ensure compliance with privacy regulations. Seventh, there might be overreliance on AI predictions. Blind reliance on AI predictions without human oversight can lead to misguided decisions. Balancing AI insights with human judgment is essential to avoid overreliance and to maintain a nuanced understanding of complex market dynamics. Ninth, there is risk of herd mentality. If multiple investors rely on similar AI models, there is a risk of herd

mentality, where market trends become exaggerated. Encouraging diversity in investment strategies can help mitigate the risk of following trends unquestioningly.

Tenth, we also could have some algorithmic complexity barrier. Complex algorithms may challenge understanding and interpreting decision-making processes. Striving for transparency in algorithmic operations is essential to build trust among investors and stakeholders. There is already some evidence from Pitchbook's Hodgson (2023) that technology may be better at investing. In an experiment in 2020, the Harvard Business Review built an investment algorithm and tested its performance against the returns of 255 angel investors. The results: The algorithm reported an internal rate of return of 7.26% compared to 2.56% for the angels. While the Harvard Business Review found that the algorithm outperformed humans, the results were markedly lower when compared against an elite group of experienced angel investors. The latter achieved an average IRR of 22.75%. But who knows? In a decade, founders may pitch ChatGPT or Bard for capital instead of a fellow human.

Eleventh, there is a lack of AI regulation in many industries. Thus, monitoring the adverse AI effects of protecting data privacy, short-termism, and various ESG reporting biases is hard. The development and deployment of artificial intelligence (AI) systems pose significant risks to society. AI companies need an effective risk management process and sound risk governance to reduce these risks to an acceptable level. Shuett et al. (2024) explore how AI companies can improve their risk governance by setting up an AI ethics board. They identify five key design choices: (1) What responsibilities should the board have? (2) What should its legal structure be? (3) Who should sit on the board? (4) How should it make decisions? (5) And what resources does it need? They break each of these questions into more specific sub-questions, list options, and discuss how different design choices affect the board's ability to reduce societal risks from AI. Several failures have shown that designing an AI ethics board can be challenging. They attempt to provide a toolbox to help AI companies overcome these challenges.

Twelveth, there is a lack of human interaction. The ability to look beyond numbers and find the potential for disruptive ideas is a uniquely human skill—at least for now. How well AI could adapt to unexpected events or rapidly changing market conditions, like the downturn we are currently experiencing, also remains to be seen (Hodgson, 2023).

Mitigating various risks associated with AI in investment involves a multifaceted approach. Establishing clear ethical guidelines, promoting diversity in AI development teams, and implementing ongoing audits of algorithms can help uncover and rectify biases. Additionally, fostering collaboration among industry stakeholders and policymakers is essential to create a framework that ensures responsible and transparent AI use in investment decisions.

Analyses of AI and CSR

Corporate Social Responsibility (CSR) is a business practice that encourages firms to work sustainably to mitigate their environmental footprint and participate in improving social justice and economic issues while considering stakeholders' interests (Deng, Kang, & Low, 2013; Ferrell, Liang, & Renneboog, 2016; Jo & Harjoto, 2011; Liang & Renneboog, 2017). This business model has become a new aspect that current businesses implement into their strategies, demonstrating a company's commitment to making positive societal contributions. As consumers become more conscious about the environmental and social challenges we face today, they will seek companies that are transparent about their CSR efforts and are investing in long-term sustainability. This consumer preference is noticeable in the financial performance of firms that actively engage in

CSR practice. A study from the University of Rennes found that the firms with very high CSR levels (generally above 95) will primarily "benefit from the positive effects on financial accounting performance" (Lachuer & Jabeur, 2022).

They suggest that firms must make substantial and meaningful CSR investments to benefit stakeholders. On the other hand, companies that engage in greenwashing will face both a negative reputation and a decrease in accounting performance. Another notable finding, illustrated in the graph, shows that CSR level positively correlates to betas that are either close to 1 or less than 0.5 (Lachuer & Jabeur, 2022). This indicates that sustainable companies are typically less volatile and provide more consistent returns over time. Further, CSR paired with AI innovation effectively works as an unsystematic risk management tool and improves operational efficiency. Under the resource-based view, we found that companies can differentiate themselves from competitors by fulfilling the VRIO framework. Intangible improvements such as a positive brand reputation, strong connections with stakeholders, and increased customer loyalty, can be improved with CSR initiatives.

Given the recent rise of AI, businesses have begun leveraging their unique technological capabilities to enhance CSR initiatives. Before integrating AI in developing CSR messages, firms should consider both the opportunities and risks of using AI tools. Some risks include ethical concerns regarding biases or misinformation embedded in AI algorithms, which would create a harmful brand image and have significant repercussions for society. To optimally integrate AI into the development of CSR initiatives, firms should consider investing in the research and development of socially responsible AI. Furthermore, AI can be a helpful tool to increase diverse stakeholder engagement, which would provide insight into the impact of these AI-driven CSR initiatives. Still, firms must be cautious about protecting their stakeholders' data given their access to sensitive information.

In determining how effective AI technology is in improving communication with stakeholders, we evaluated a study on consumer responses to AI-generated CSR messages. The first data set showed that consumers prefer human communication over AI messages. The supporting research found no strong correlation between human and AI CSR messages to brand innovativeness (Amani et al., 2024). AI technology seems to lack the emotional intelligence needed to address sensitive social, environmental, and economic issues; however, this technology may be helpful for more technical strategic planning. Additionally, due to the ongoing development of AI, many consumers lack trust in the authenticity and credibility of AI-generated CSR messages. To create trust, companies must ensure that they are transparent about the use of AI and review all AI-generated messages to align the message with company-specific or industry values. Although AI offers efficiency, managers should be mindful of potential errors and misinformation, so all AI messages should be reviewed by human employees before releasing any communication. Lastly, firms should consider incorporating feedback mechanisms and actively engage with stakeholders to address concerns regarding AI and build stronger relationships with their target audience.

As businesses move towards integrating AI into the workplace, specifically to help address CSR initiatives, they must consider the risks and opportunities that stem from this technology. Since CSR prioritizes ethical decisions, firms must be conscious of the potential for misinformation and take steps to mitigate that risk. If implemented ethically, firms with high CSR levels will reap the benefits of increased operational efficiency and a strong brand reputation.

Analyses of AI and ESG

Environmental, Social, and Governance (ESG) factors have recently become integral to investment decision-making. According to David F. Larcker, the James Miller Professor of Accounting, Emeritus, at Stanford Graduate School of Business, "Corporate governance is table stakes: All companies are expected to have it." Governance has been a focal point for investors, who view governance quality as already embedded in asset prices. Environmental issues like climate risks are not yet perceived to affect the asset price, and social factors are seen as the least important when driving investment decisions.

As the investment landscape evolves, companies increasingly leverage AI to enhance their ESG initiatives and affect ESG outcomes. AI efficiently processes vast amounts of data, enabling companies to identify patterns and trends that might not be evident through traditional analysis. This allows for more accurate assessments of a company's ESG performance and the long-term risks and opportunities associated with ESG factors. Here is one attempt to measure the growth of AI and ESG based on capital asset growth rate (CAGR) (see Tables 1A and 1B) based on various reports.

Table 1A Market Size of AI and ESG over time

This table summarizes the market size of AI (Artificial Intelligence) and ESG (Environmental, Social, and Governance) investments over time. The values in the table are estimates and represent global market sizes in billions of USD.

Year	AI Market Size (USD Billion)	Growth Rate (AI)	ESG Market Size (USD Billion)	Growth Rate (ESG)
2020	62.4	15%	35.3	10%
2021	93.5	25%	53.2	20%
2022	119.8	28%	80.0	23%
2023	164.9	38%	112.0	25%
2024	233.3	41%	150.4	30%
2025	325.5	40%	210.6	40%

Notes: AI Market Size is estimated using compound annual growth rates (CAGR) driven by machine learning, deep learning, robotics, and data analytics innovations. ESG Market Size refers to global ESG-related investments, including sustainable funds, corporate initiatives, and green finance.

Table 1B The estimated market size growth of AI and ESG-related markets between 2020 and 2030. These values are based on aggregated trends and projections from various market reports.

Year	AI Market Size (USD Billion)	CAGR (AI)	ESG Market Size (USD Billion)	CAGR (ESG)
2020	50	35%	15	15%
2021	67.5		17.3	
2022	91.1		19.9	
2023	123.0		22.8	
2024	166.1		26.2	
2025	224.3		30.1	
2026	302.8		34.6	
2027	408.8		39.7	
2028	552.0		45.6	
2029	745.2		52.4	
2030	1,006.0		60.3	

Notes: The AI market is estimated to grow at a CAGR of approximately 35%, and the ESG market is expected to grow at a CAGR of approximately 15%. The AI market shows more exponential growth due to technological innovations and adoption in the healthcare, finance, and manufacturing sectors. The ESG market grows steadily, driven by regulations, increasing investor interest, and corporate accountability efforts.

Case Studies of AI and ESG

AI has been shown to enhance operational efficiency in ESG activities by automating data collection, analysis, and reporting processes. Companies can make better-informed decisions because AI can quickly process large volumes of ESG data from various sources, identify trends, and generate insights. However, there are challenges in ensuring the accuracy and reliability of AI algorithms, especially when interpreting complex ESG data.

Several companies like Microsoft, Google, Unilever, Walmart, and IBM have begun implementing AI to further their operational efficiency in ESG activities. Microsoft uses AI to improve energy efficiency and reduce carbon emissions. The company's AI for Earth initiatives leverages AI technology to collect and analyze environmental data, helping in conservation efforts and sustainable agriculture. Microsoft also utilizes AI to optimize data center operations, significantly reducing energy consumption. Unilever's digital platform uses AI to track deforestation and support responsible sourcing of raw materials like palm oil, tea, and other commodities. Lastly, IBM uses AI-driven analytics to enhance corporate governance and transparency. Its AI tools analyze financial reports, board structures, and compliance records to ensure adherence to government standards.

Artificial intelligence can also help identify and assess ESG-related risks and help manage risks more effectively. AI algorithms can detect emerging risks such as environmental disasters, supply chain disruptions, or social controversies, allowing companies to take proactive measures to mitigate them. Companies like IBM, Nestlé, and BP have already integrated AI to help with risk management.

Microsoft developed the AI for Earth program for environmental sustainability, which uses AI to analyze environmental data related to climate change, biodiversity, and water resources. The

company also built an AI-powered Carbon Calculator to track emissions across its operations and supply chain. Microsoft also utilizes AI-driven platforms like its Power BI to create interactive dashboards that give stakeholders real-time insights into the company's ESG performance. These dashboards make it easy for investors, employees, and customers to access and understand complex data about sustainability and corporate governance practices. Unilever uses AI chatbots to enhance communication about its sustainability initiatives. The chatbots can provide data about Unilever's environmental impact, sourcing practices, and social responsibility programs, which can increase transparency and trust between stakeholders and the company. Salesforce is a company that uses AI to analyze employee feedback and employee surveys to provide management with actionable insights to improve the employee experience at their company and workplace practices. This allows AI to be a tool to help foster a more engaged and satisfied workforce.

Unilever partnered with AI startups and developed a platform called GeoAlert, which uses satellite imagery and machine learning to monitor deforestation in real-time for sustainable sourcing. The AI system analyzes vast amounts of data to detect illegal logging or unsustainable farming practices in its supply chain.

Google developed AI for Social Good, which includes projects like flood forecasting using AI models to predict floods in vulnerable regions. The company also uses AI to improve healthcare delivery, such as detecting diabetic retinopathy through retinal scans, aiming to use its technological expertise to address global social challenges, such as disaster response and healthcare accessibility.

IBM uses its Watson AI platform to enhance risk management by analyzing unstructured data from news articles, social media, and other sources to detect emerging ESG risks. This allows IBM to take advantage of the ability to get ahead of the problem before it escalates and address the potential issue by adjusting its strategies accordingly. Nestlé uses AI to manage supply chain risks, ensuring responsible sourcing and mitigating disruptions. AI algorithms analyze supplier data to identify risks related to labor practices, environmental impact, and social controversies. BP uses AI to monitor and manage environmental risks, such as oil spills and emissions. The AI systems analyze data from various operations to detect anomalies and predict potential hazards before they escalate.

Walmart implemented an AI-powered energy management system across its stores and distribution centers for energy efficieny. The system uses machine learning to optimize heating, cooling, and lighting based on real-time data.

Additionally, AI offers significant opportunities to enhance stakeholder engagement by providing transparent and accessible information about a company's ESG performance. Through interactive dashboards and chatbots, AI can enable more meaningful communication with stakeholders, including investors and employees. However, this technology comes with potential concerns about privacy and data security. Companies like Microsoft, Unilever, and Salesforce showcase the benefits of implementing AI to help with stakeholder engagement.

Further Analyses of AI and ESG

AI's ability to process and interpret complex ESG data enables companies to manage risks more effectively, identify opportunities for improvement, and align their strategies with long-term sustainability goals. As AI continues to revolutionize ESG initiatives, the integration of advanced technologies provides a robust framework for companies to enhance operational efficiency,

mitigate risks, and foster transparent stakeholder engagement. This new technology aids in making well-informed investment decisions but also drives positive environmental and social outcomes, reflecting the evolving priorities of modern investors and stakeholders.

However, the rapid adoption of AI in ESG also brings several drawbacks and challenges to light. Concerns around accuracy and biases within algorithms and issues related to privacy and data security are hurdles that must be addressed. Furthermore, the correlation between AI capabilities and ESG outcomes is not always straightforward, raising doubts about the potential unintended consequences and ethical implications of using AI to help decision-making.

AI is affecting the ESG in many more ways than people may think. According to the article " Implications for Artificial Intelligence and ESG Data, " financial firms now use AI to calculate companies' investment in ESG. This helps them to make decisions about investing or not in certain companies. Especially in recent years, ESG has become a larger variable for companies. This places a greater emphasis on ESG with AI; it will only increase the rate at which ESG becomes a more critical variable. This also means that companies that invest in ESG will receive more money at higher rates than other companies. This puts much higher emphasis on companies investing in ESG as it helps them grow as a firm in the long term.

However, many more challenges create problems for financial firms and AI in determining ESG efforts. There are difficulties in accurately measuring and rating a company's environmental and social impact. Particularly given that ESG remains an evolving concept and that there are many different reporting standards and frameworks. This has led to the so-called "aggregate confusion" among companies and investors. This confusion around ESG ratings and their biases is supported by academic evidence, "When assessing the landscape of ESG ratings, MIT Sloan found that the correlation among agencies' ESG ratings is on average 0.61; by comparison, credit ratings from Moody's and Standard & Poor's are correlated at 0.99. The research team found that rating agencies may adopt different definitions of ESG performance or take different approaches to measure that performance or weight the ESG attributes." They concluded that the information investment companies receive is unreliable or consistent.

This has led to funds being created to help create uniformity across the industry. The International Monetary Fund urges political leaders and regulators to develop standards, promote disclosure and transparency, and integrate sustainability considerations into investment and business decisions. It emphasizes the importance of ESG audits for the system's proper development. Recognizing the topic's complexity, several institutions are addressing the need for such audits. Despite significant advancements in Europe, particularly in the environmental dimension, much work must be done to incorporate non-financial corporate information into reporting standards to enhance transparency for companies and investors. This was recently highlighted by the Acting Director of the SEC's Division of Corporate Finance in a public statement listing critical issues for developing common global standards.

The module also allows for the analysis of industries' ESG performance, aiding in the selection of sustainable and socially responsible investments. Studies results show a 20% dependence of stock returns on ESG-related news, leading to the development of a tool that tracks this relationship. This tool complements quantitative strategies and facilitates ESG-informed investment decisions. This research demonstrates GPT-3.5's ability to generate accurate responses to ESG prompts, underscoring the importance of high-quality training data. It also contributes to the growing body of literature on AI in ESG research and provides a framework for future investigations.

Global ESG fund assets reached approximately $2.5 trillion by the end of 2022. Research has indicated that ESG investing does not necessarily result in lower financial returns and can sometimes lead to higher returns. ESG investing has also garnered increasing interest from institutional investors, such as pension funds and insurance companies, who recognize that ESG factors can significantly impact long-term financial performance. These investors are integrating ESG considerations into their investment decision-making processes.

Overall, AI is being used in many ways, especially through finance. However, the help the AI is providing is not the most reliable. This is because ESG efforts are easily disguised as greenwashing. This makes it very hard for AI to be able to tell whether or not it's truly ESG or if it's greenwashing. Humans need to be able to know if the company cares about ESG or if they are faking it to gain more investment. This shows that AI can help with the basic retrieval of information but should not make decisions on its own.

Integrating Artificial Intelligence (AI) into ESG initiatives is pivotal in evolving corporate responsibility and sustainable business practices. As our study has illuminated, AI offers an array of capabilities that empower companies to navigate the complex landscape of ESG factors more effectively, from enhancing operational efficiency to mitigating risks and fostering transparent stakeholder engagement.

One of the key contributions of AI in ESG lies in its ability to process vast amounts of data and derive actionable insights that may not be readily apparent through traditional analysis methods. AI streamlines data collection, analysis, and reporting processes through automation, enabling companies to make better-informed decisions regarding their ESG performance. Case studies of companies like Microsoft, Unilever, and IBM demonstrate how AI is being leveraged across various industries to optimize energy efficiency, track deforestation, enhance corporate governance, and manage environmental risks.

Moreover, AI is a powerful tool for identifying and assessing ESG-related risks, allowing companies to proactively address emerging challenges such as environmental disasters, supply chain disruptions, and social controversies. By analyzing unstructured data from diverse sources, AI algorithms enable early detection of potential risks, empowering companies to adjust their strategies and mitigate negative impacts.

In addition to risk management, AI facilitates stakeholder engagement by providing transparent and accessible information about a company’s ESG performance. Interactive dashboards and AI-driven chatbots offer stakeholders real-time insights into sustainability practices, fostering trust and accountability. However, it's crucial to acknowledge the concerns surrounding privacy and data security inherent in AI-driven communication platforms, highlighting the need for responsible AI development and implementation.

Despite AI's numerous benefits to ESG initiatives, challenges persist, including accuracy issues, algorithm biases, and ethical implications. The correlation between AI capabilities and ESG outcomes is not always straightforward, underscoring the importance of ongoing research and development to address these challenges. Additionally, the rise of greenwashing poses a significant obstacle, complicating AI's ability to discern genuine ESG efforts from superficial gestures.

Looking ahead, the continued integration of AI into ESG practices requires a concerted effort to overcome these challenges while maximizing the technology's potential for positive environmental and social impact. Standardizing ESG reporting frameworks, promoting transparency, and responsible AI development are critical steps in ensuring the integrity and effectiveness of AI-driven ESG initiatives.

In sum, while AI offers unprecedented opportunities to enhance ESG performance, its successful integration requires careful consideration of ethical, technological, and regulatory factors. By harnessing the transformative power of AI responsibly, businesses can navigate the complexities of ESG challenges more effectively, driving sustainable growth and creating long-term value for stakeholders and society as a whole.

V Discussions and Conclusions

Discussions

The intersection of Artificial Intelligence (AI), Environmental, Social, and Governance (ESG) principles, and business ethics is rapidly evolving. These three elements are increasingly critical in shaping corporate strategy, decision-making, and public perception. Here is an exploration of how they interact:

AI and ESG Integration

AI can transform how businesses approach ESG goals by improving efficiency, enabling more accurate data collection, and enhancing transparency. For instance, AI-driven analytics can monitor a company's environmental impact by tracking real-time carbon emissions, identifying supply chain inefficiencies, and recommending sustainable practices. Similarly, AI can enhance social initiatives by supporting diversity and inclusion efforts, through bias-reduction algorithms or by improving employee wellness programs.

In governance, AI's role in enhancing decision-making through data-driven insights can foster transparency and accountability, enabling organizations to adhere to the governance principles of fairness, responsibility, and oversight. However, the use of AI in ESG also raises concerns. AI systems are not immune to biases, and if these biases are not carefully managed, they can inadvertently undermine ESG goals. For example, AI algorithms that assess hiring or promotion decisions could perpetuate gender or racial disparities, thereby negatively impacting social equity.

AI and Business Ethics

The ethical use of AI in business is crucial to ensure that its deployment does not harm individuals or society. AI's power to automate decisions—whether in hiring, lending, or criminal justice—can raise serious concerns about fairness, transparency, and accountability. For instance, an AI algorithm that prioritizes profit over human well-being may undermine ethical standards by exploiting workers, encouraging environmental degradation, or perpetuating societal inequalities. Ethical AI development requires companies to commit to principles like fairness, transparency, privacy protection, and accountability. Businesses must take proactive steps to mitigate risks such as algorithmic bias, misuse of personal data, and lack of explainability in AI-driven decisions. Furthermore, ethical AI involves designing systems aligned with human rights and respecting individuals' dignity.

Balancing Profit and Responsibility

We should weigh AI's potential to drive business profitability against the broader social and environmental consequences. AI-driven automation could lead to mass layoffs, or AI's decisions could disproportionately impact specific marginalized communities. In contrast, businesses that proactively use AI to meet ESG objectives may gain long-term benefits, such as increased

consumer trust, regulatory compliance, and competitive advantage. However, there is a significant challenge in balancing short-term profit goals with long-term sustainability, mainly when AI solutions are expensive to develop and implement.

Governments and regulatory bodies are increasingly addressing these challenges by establishing AI ethics guidelines, ESG reporting standards, and sustainability regulations. Companies must align AI development and deployment with these evolving frameworks to avoid reputational risks and regulatory penalties.

Limitations of AI in the Context of ESG and Business Ethics

Bias and Inequality: AI models are trained on data that can inherit biases from historical patterns, human behavior, or biased datasets. In the context of ESG, this can result in discriminatory practices or perpetuating societal inequalities. For example, AI used to assess social responsibility or diversity may not account for nuanced or evolving societal norms, leading to biased outcomes in employee hiring or sustainability metrics.

Lack of Transparency (Black-box problem): AI systems, particularly deep learning models, often function as "black boxes" with decision-making processes that are difficult for humans to understand or audit. This lack of transparency can undermine trust, especially when applied to critical ESG issues such as governance, environmental impact reporting, or the ethical implications of business decisions. Businesses may struggle to ensure accountability and fairness without explaining how AI arrives at conclusions.

Ethical Considerations in AI Deployment: AI technologies may unintentionally conflict with core ethical principles such as privacy, fairness, and consent. In business settings, AI-driven tools for ESG monitoring may lead to concerns over surveillance, data privacy violations, or algorithmic decision-making that overrides human judgment. This raises critical questions about how AI impacts workers' rights, consumer protection, and environmental integrity.

Over-reliance on AI for Decision-Making: Businesses may become overly reliant on AI systems to decide ESG performance or ethical considerations, potentially sidelining human judgment, intuition, and moral reasoning. This over-reliance can result in decisions prioritizing efficiency or profitability over holistic social and environmental impact considerations. AI is often ill-equipped to fully account for the complexity of human ethics or environmental consequences, potentially leading to unintended adverse outcomes.

Data Quality and Availability: Effective AI models require large volumes of high-quality, accurate data to make informed decisions. However, in the context of ESG, data may be scarce, fragmented, or unreliable, especially for environmental and social metrics that are difficult to quantify. This lack of comprehensive data makes it challenging for AI systems to provide reliable ESG insights, leading to potential miscalculations or misleading assessments.

Regulatory and Legal Risks: As AI technologies continue to advance, there is a lack of universally accepted guidelines and regulations governing its ethical use, particularly in the ESG space. Businesses may face legal and regulatory risks when deploying AI in ways that infringe on human rights, privacy, or violate ESG principles. Inconsistent global regulations surrounding AI ethics can lead to challenges in maintaining compliance, especially for multinational corporations.

Environmental Impact of AI: While AI has the potential to help businesses improve their environmental performance (e.g., optimizing energy consumption or reducing waste), the environmental footprint of AI itself can be significant. The computational power required for training large AI models, especially deep learning, can be energy-intensive and contribute to carbon emissions. Companies must weigh AI's environmental impact alongside its benefits in achieving ESG goals.

Human Displacement and Labor Concerns: The widespread adoption of AI can lead to job displacement, which presents ethical challenges for businesses striving to promote fair labor practices. Automation and AI could exacerbate inequalities in the workforce, particularly if companies focus on maximizing profit without investing in reskilling or creating new job opportunities for displaced workers. This could undermine the social element of ESG, which emphasizes fair labor practices and inclusive growth.

Long-term Sustainability and Unintended Consequences: AI-driven solutions may prioritize short-term efficiency over long-term sustainability. For instance, optimizing for immediate financial returns may not align with long-term environmental goals, such as reducing carbon footprints or conserving natural resources. Without careful governance, AI systems may overlook the broader implications of business practices on future generations or fail to address systemic ESG issues adequately.

Corporate Governance and Accountability: AI can be a powerful tool for improving corporate governance by providing insights into decision-making, identifying risks, or enhancing compliance efforts. However, AI does not have the moral compass of a human decision-maker, which can create challenges in ensuring that business decisions align with ethical standards. Accountability mechanisms need to be in place to ensure that AI-driven decisions remain aligned with corporate values and ethical principles. This also raises concerns about shifting responsibility from human leaders to AI systems.

Lack of Universal Standards: There is currently no universally accepted framework for applying AI to ESG and business ethics. As a result, businesses may struggle to navigate different standards across regions and industries, creating confusion and inconsistency in how ESG goals are measured, reported, and achieved through AI-driven solutions. In summary, while AI offers significant potential for enhancing ESG practices and business ethics, it must be deployed carefully and with consideration for its limitations. Human oversight, transparency, and ethical governance are essential to ensure that AI's impact aligns with societal values and long-term sustainability goals.

Future Directions of Research on AI, ESG, and Business Ethics

The intersection of AI, ESG principles, and business ethics represents a rapidly evolving field with significant implications for the future of sustainable and responsible business practices. As organizations increasingly adopt AI technologies to drive efficiency, innovation, and decision-making, ensuring these advancements align with ethical standards and ESG goals is critical.

Below are key future directions for research in this domain: (i) Developing robust ethical frameworks that guide the integration of AI into ESG strategies. This includes ensuring transparency, fairness, and accountability in AI systems used for ESG reporting, risk assessment, and decision-making; (ii) Leveraging AI to improve the accuracy, reliability, and standardization of ESG metrics and reporting. AI can help analyze vast datasets to identify trends, predict risks, and measure the impact of corporate actions on sustainability goals; (iii) Addressing biases in AI algorithms that may disproportionately affect marginalized communities, particularly in areas like hiring, lending, and resource distribution. Research should explore how AI can promote social equity while adhering to ESG principles; (iv) Investigating AI's dual role in contributing to and mitigating environmental challenges. While AI can optimize energy use and reduce waste, future studies should also focus on carbon footprint and resource consumption; (v) Establishing governance frameworks to ensure responsible AI deployment in ESG-related activities. This includes regulatory compliance, stakeholder engagement, and ethical oversight; (vi) Enhancing stakeholder trust through transparent and explainable AI systems. Research should explore how AI can facilitate meaningful stakeholder engagement while maintaining ethical standards.

Conclusions

Artificial Intelligence (AI), CSR, and Environmental, Social, and Governance (ESG) have become increasingly intertwined, with AI playing a transformative role in advancing ESG initiatives across various sectors. The relationship between AI, CSR, and ESG is multifaceted, offering both opportunities and challenges for organizations striving to improve their sustainability and ethical practices. The future of AI, ESG, and business ethics presents both opportunities and risks. Companies that harness AI to advance ESG goals while adhering to ethical business practices can create long-term value and contribute to a more sustainable and equitable society. Conversely, failing to integrate these principles can lead to detrimental social, environmental, and ethical consequences.

For AI to contribute positively to ESG and ethical goals, businesses must adopt a responsible, transparent approach to AI design and implementation. This includes addressing biases, ensuring privacy, maintaining accountability, and aligning business practices with broader societal goals. The convergence of AI and ESG is not just a technological issue—it is a fundamental ethical challenge that requires collaboration between businesses, policymakers, and society to create frameworks that protect human rights and promote sustainable practices.

AI can be a powerful tool for achieving ESG objectives, but only if developed and deployed with a commitment to fairness, transparency, and accountability. Businesses and corporatrions can drive innovation and profitability through this responsible approach and contribute positively to the global community. We aim to contribute to the AI literature in finance by adding how the AI-driven strategies of business ethics, CSR, and ESG are functionally and qualitatively related to firm risk, firm performance, and the firm’s competitive positions by showing their benefits and future challenges. In conclusion, the relationship between AI and ESG is symbiotic and evolving. AI is a powerful tool for advancing CSR and ESG initiatives, offering enhanced data analysis, predictive capabilities, and reporting efficiency. However, organizations must approach AI integration strategically, balancing the benefits with potential challenges to ensure responsible and sustainable implementation of AI in their CSR and ESG efforts.

References

Abay, Z. (2022). The Signalling Role of Voluntary ESG Assurance. *International Journal of Managerial and Financial Accounting*, 14(3), 265-294.

Adeoye, O., Okoye, C., Ofodile, O., Odeyemi, O., Addi, W., & Ajahi-Nifise, A. (2024). Artificial Intelligence in ESG investing: Enhancing portfolio management and performance. *International Journal of Science and Research Archive,* 11(1), 2194-2205.

Amani, Nadir, et al. (2024). "Generative AI Hurts Brands? Exploring Consumer Responses to AI generated CSR Messages." AMA Winter Academic Conference Proceedings, vol. 35, Jan. 2024, 20–24.

Buchanan, B., Cao, C.X. and Chen, C. (2018). Corporate social responsibility, firm value, and influential institutional ownership. *Journal of Corporate Finance*, *52*, 73-95.

Chouaibi, S., Rossi, M., Siggia, D., & Chouaibi, J. (2022). Exploring the moderating role of social and ethical practices in the relationship between environmental disclosure and financial performance: evidence from ESG companies. *Sustainability* (Switzerland), 14(1), 209.

Cui, J., Jo, H., & Velasquez, M. G. (2025). Firm-level climate change initiatives and Christian religiosity. Unpublished working paper. Santa Clara University.

Deng, X., Kang, J., & Low, B. (2013). Corporate Social Responsibility and Stakeholder Value Maximization: Evidence from Mergers, *Journal of Financial Economics,* 110, 87–109.

Du, S., & Sankar, S. (2023) AI Through a CSR Lens: Consumer Issues and Public Policy, *Journal of Public Policy & Marketing* <journals.sagepub.com/home/ppo>.

Ferrell, Allen, Hao Liang, and Luc Renneboog, (2016). Socially Responsible Firms, *Journal of Financial Economics,* 122, 585–606.

Francis, T (2024) AI, ethics, and ESG in 2025. *Corporate Culture*.

Friede, G., Busch, T., and Alexander Bassen. (2015). ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. *Journal of Sustainable Finance & Investment*, 5(4), 210-233.

Grand View Research (2024). Artificial Intelligence Market Size & Trends. <www.grandviewresearch.com>

Hoover Institution. (2024, May). *2024 CGRI/MSCI Sustainability Survey*. <https://www.hoover.org/sites/default/files/2024-05/2024-cgri-msci-sustainability-survey-FINAL.pdf>

ldridge, Irene, and Payton Martin. (2022). ESG In Corporate Filings: An AI Perspective. 2022.

Jo, H., Chun, H., & Song, H. (2025). Shared Governance and ESG Rating: Evidence from Korea, *Corporate Social Responsibility and Environmental Management*, 32(2), 2769-2782.

Jo, H., & Harjoto, M. (2011). Corporate Governance and Firm Value: The Impact of Corporate Social Responsibility. *Journal of Business Ethics*, 103(3), 351-383.

Jo, H., & Harjoto, M. (2012). The causal effect of corporate governance on environment society and governance. *Journal of Business Ethics, 106(1)*, 53-72.

Jo, H., Juarez, A., Rossi, A., Strauss, C., & Ressler, L. (2024). The Impact of Artificial Intelligence on Venture Capital Sourcing and Due Diligence, *Global Journal of Entrepreneurship*, 8(1), 12-33.

Kemell, K., & Vakkuri, V. (2023). What Is the Cost of AI Ethics? Initial Conceptual Framework and Empirical Insights. 14th International Conference, ICSOB 2023, Lahti, Finland, November 27-29, 2023 Proceedings.

Lachuer, J., & Jabeur S.B. (2022) Explainable Artificial Intelligence Modeling for Corporate Social Responsibility and Financial Performance, *Journal of Asset Management* <http://doi.org/10.1057/s41260-022-00291-z>.

Li, G., Li, N., & Sethi, S.P. (2021) Does CSR Reduce Idiosyncratic Risk? Roles of Operational Efficiency and AI Innovation, *Production and Operations Management*, 30, 7.

Liang, H., & Reneboog, L. (2017). On the Foundation of Corporate Social Responsibility, *Journal of Finance*, 72(2), 853-910.

Lin, T. (2024). Environment, Socialal, and Governance (ESG) and Artificial Intelligence (AI) in Finance: State-of-the-art and Takeaways. *Artificial Intelligence Review*, 57(76).

Lioui, A., & Sharma, Z. (2012). Environmental corporate social responsibility and financial performance: Disentangling direct and indirect effects. *Ecological Economics*, 78, 100-111.

Macpherson, Martina and Gasperini, Andrea and Bosco, Matteo, (2021). Implications for Artificial Intelligence and ESG Data (June 8, 2021). Available at SSRN: https://ssrn.com/abstract=3863599 or http://dx.doi.org/10.2139/ssrn.3863599

Masulis, R.W. and Reza, S.W. (2015). Agency problems of corporate philanthropy. *Review of Financial Studies*, *28*(2), 592-636.

Matsumura, E.M., Prakash, R. and Vera-Munoz, S.C.(2014). Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Review*, *89*(2), pp.695-724.

Microsoft. (n.d.). *Microsoft AI for Earth*. Microsoft. <https://www.microsoft.com/en-us/ai/ai-for-earth>

Nestlé. (n.d.). *Nestlé uses AI to ensure responsible sourcing*. Nestlé. https://www.nestle.com/sustainability/ai-responsible-sourcing

Oetzel, J. (2023). CSR or ESG: Where Do Sustainability Frameworks Fit In? Kogod School of Business. https://kogod.american.edu/news/csr-or-esg

Pedersen, L., Fitzgibbons,S., & Pomorski, L. (2021). Responsible Investing: The ESG-Efficient Frontier. *Journal of Financial Economics*, 142(2), 572-597.

Perez, L., Hunt, V., Samandari, H., Nuttall, R., & Biniek, K. (2022). Does ESG really matter – and why? McKinsey Quarterly, August 2022.

Saetra, H. S. (2023) The AI ESG Protocol: Evaluating and Disclosing the Environment, Social, and Governance Implications of Artificial Intelligence Capabilities, Assets and Activities. *Sustainable Development*, 31(2), 1027-1037 <http://doi.org/10.1002/sd.2438>.

Salesforce. (n.d.). Einstein Analytics. Salesforce. Retrieved from https://www.salesforce.com/products/einstein-analytics/

Sandford, J. (2024). How Generative AI is Transforming ESG Reporting. SIA. https://www.sia-partners.com/en/insights/publications/how-generative-ai-transforming-esg-reporting

Servaes, H., & Tamayo, A. (2013). The Impact of Environment society and governance on Firm Value: The Role of Customer Awareness. *Management Science,* 59(5), 1045-1061.

Shuett, J., Reual, A., & Carlier, A. (2024). How to Design an AI Ethics Board? AI and Ethics, https://doi.org/10.1007/s43681-023-00409-y

Statman, M. (2024). *A Wealth of Well-Being: A Holistic Approach to Behavioral Finance*. Wiley.

Supply Chain Digital. (2023, April 18). *Unilever's AI for sustainable sourcing*. <https://www.supplychaindigital.com/technology/unilever-s-ai-for-sustainable-sourcing>

Supply Chain Digital. (2023, March 20). *Nestlé's AI-powered supply chain risk management*. https://www.supplychaindigital.com/technology/nestle-s-ai-powered-supply-chain-risk-management

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751-752.

Wang, Z., & Wang, Y. (2025). Leveraging AI for Compliance in Corporate Restructuring in Times of Financial Distress. Working paper. New York University and University of California, Berkeley.

Wild, M. (2023, March 15). *Microsoft: How AI helps us build a sustainable future*. TechRepublic. <https://www.techrepublic.com/article/microsoft-how-ai-helps-us-build-a-sustainable-future/>

World Economic Forum. (2023, February 10). *Unilever uses AI to combat deforestation*. https://www.weforum.org/agenda/2023/02/unilever-uses-ai-to-combat-deforestation/

AI Mistakes in the Classroom

Jaime E. Peters[1], Tara L. Gerstner[2]

Abstract

The integration of Artificial Intelligence (AI) in educational settings has sparked significant interest, yet the potential pitfalls are often overlooked. This paper explores a series of assignments where AI fell flat. Dissecting the errors made in the assignments and how they violated the learning theory of Connectivism - an emerging learning theory emphasizing the importance of networks and technology in knowledge acquisition – we hope to help others avoid the same pitfalls when designing AI inclusive assignments. The findings underscore the necessity for educators to provide clear guidance, equitable access to technology, and appropriate training for students to navigate AI tools effectively. By sharing these insights, we aim to inform best practices for AI integration in academia and encourage a more thoughtful approach to leveraging technology in teaching.

Keywords: Artificial Intelligence, Connectivism, Pedagogy, Classroom Assignments

JEL Classifications: A22, A29

I. Introduction

Large language models and their rapid, sweeping impact on education have only just begun to be documented in the teaching and learning literature. In quick succession, many publications on the acceptance and use of artificial intelligence in the classroom have appeared. The literature primary fell into two categories: (1) how-to guides for incorporating AI into the classroom (e.g. Mollick & Mollick, 2023; Noy & Zhang, 2023; Winkler & Ross, 2019) and (2) concerns about how it will be used for cheating (e.g. Amani, et al., 2023; Barros et al., 2023; Michel-Villarreal et al., 2023; Okaiyeto et al., 2023). While not explicitly stated in most articles, an underlying learning theory that strongly supports the inclusion of AI in the classroom is Connectivism. This theory can explain why AI in the classroom supports learning. However, what is currently missing from the literature are documented failures caused by incorporating AI into assignments and examples of how such inclusion has hindered learning. This paper attempts to begin to fill that gap. Our results suggest that professors must not assume the use of the technology is intuitive, that students have ready access to AI despite its perceptions of being a free tool and must be very careful if using AI as an information source, even for non-copywritten, publicly available information. The following is a literature review showing how inclusion of AI supports the Connectivism view of learning, followed by one example of a successful AI assignment in the classroom, then three examples of mistakes when building AI based assignments and how they violated the Connectivist view of learning, leading to frustration by students and professors.

[1] Assistant Professor of Finance, Maryville University, jpeters@maryville.edu, ORCiD: 0000-0002-5373-9883
[2] Assistant Professor of Business, Marketing, and Entrepreneurship , Illinois Wesleyan University, gerstner@iwu.edu, ORCiD:0000-0002-6297-4332

II. Literature Review

Connectivism, a learning theory developed by George Siemens (2005) and Stephen Downes (2008), emphasizes the importance of networks, connections, and technology in the learning process. In the digital age, where knowledge is dynamic and distributed across various platforms, Connectivism provides a framework for understanding how learners interact with information, peers, and digital tools (Alam, 2023).

Siemens (2005) and Downes (2008) imagine learning as a network of connecting nodes. Nodes can be anything, a feeling, a thought, a piece of data, or some information (data with intelligence applied). The nodes can be static or dynamic, living or inanimate. These nodes become connected through process of learning – where knowledge becomes meaning. These links (we like to think of them as strands in a rope) are created when a learner is able to understand and connect two nodes together. The connection can be strengthened through the strands of motivation, emotion, exposure, patterning, logic, and/or experience (Siemens, 2017). Like a rope, the number of strands can vary, making connections either weak or strong.

A series of connected nodes becomes a network (think of an individual – they represent a series of connected nodes). Within Connectivism, that person can then become a node in a greater network – creating organizational level knowledge. Within the organization, knowledge can reside within individuals, systems or non-human appliances – each is a node that may be a network (Siemens, 2017).

What is particularly useful in the digital age is the concept of the node and network. Nodes compete for relevance and strength of connection. Nodes can become antiquated, weakening their connections and eventually eliminated from the network of knowledge. Nodes can be replaced or transformed with new up-to-date knowledge. Nodes can be databases, large language models, or other non-human tools that hold knowledge. Consequently, this learning theory specifically incorporates today's digital world into how we learn.

Integrating Artificial Intelligence, specifically large language models (LLMs), into Connectivist-based classrooms offers transformative opportunities to enhance learning by leveraging these principles (Upadhyay et al., 2024). AI tools, as nodes within a learner's network, can create new pathways for knowledge acquisition, collaboration, and problem-solving (Correia et al., 2024).

Core principles of Connectivism and AI integration

The integration of AI into education aligns naturally with Connectivist principles, offering practical applications that redefine the roles of students, teachers, and learning tools (Correia et al., 2024). Below are the 8 key principles of Connectivism as stated by Siemens (2005) and their implications for using AI in the classroom.

Learning and knowledge rests in diversity of opinions.

According to Connectivism, knowledge is distributed across networks of people, tools, and organizations (Siemens, 2005). AI tools like ChatGPT serve as vital components of these networks by providing immediate access to up-to-date, synthesized information (Roumeliotis & Tselikas, 2023) making it one of the diverse nodes needed to learn. In a classroom setting, students can use

AI to conduct research, and analyze real-world data (Bray, 2024). AI becomes a bridge between learners and the broader global network of information, inspiring engagement and exploration.

Learning is the process of connecting specialized nodes or information sources.

In the context of AI, tools like LLMs act as nodes that aggregate, analyze, and present knowledge from vast resources, enabling learners to access diverse perspectives (Correia et al., 2024). For example, Chen et al. (2023) demonstrated the benefits of 'student assistant' chatbots created with large language models, which students could use to access information outside of a textbook or professor. Similarly, Shyr et al. (2024) suggested that students could better understand complex academic research through AI-powered rewording and summarization of text. This not only facilitates immediate access to information but also encourages learners to see connections between different fields and concepts, a critical skill in the 21st century.

Learning may reside in non-human appliances.

AI-powered platforms like language models are a prime example of this, as they facilitate self-directed learning and provide support for both independent and group activities. For example, an AI tool can guide students through problem-solving exercises (Bray, 2024), simulate debates (Aryan, 2024), or offer multilingual support for diverse learners (Davoodi, 2024). The ability of AI to simulate human-like interaction makes it a valuable collaborator in the learning process.

Capacity to know more is more critical than what is currently known.

Students existing knowledge is less important than their ability to learn new items. AI can feed into this principal through its ability to power adaptive learning – meeting the student at their current knowledge level and allowing them to add to it. Bhatt et al. (2024) demonstrate how the use of AI-powered adaptive learning platforms enhances student learning and created a more inclusive educational setting.

Nurturing and maintaining connections is needed to facilitate continual learning.

Connectivism places a strong emphasis on learners taking control of their own education (Downes, 2008). AI tools empower students to set their own learning pace and explore topics that interest them. For instance, learners can use AI to "delve" (could not resist the joke of using this word in an AI article!) deeper into areas they find challenging or intriguing by posing iterative questions or refining their understanding through AI-generated examples and explanations (Chen et al., 2023). This autonomy feeds critical thinking and problem-solving skills, as students are encouraged to actively engage with their learning journey.

Ability to see connections between fields, ideas, and concepts is a core skill.

AI tools support this principle by providing personalized learning experiences, making it easier for learnings to see connections. For instance, Pataranutaporn et al. (2021) explained how developments in AI allow for the creation of relatable and personalized avatars for students to

interact with. Pratama et al. (2023) describe how AI can provide tailored feedback, clarify misconceptions, or help students explore new topics beyond the curriculum.

Currency (accurate, up-to-date knowledge) is the intent of all Connectivist learning activities.

The currency of AI's knowledge base can present significant challenges in educational settings. This currency challenge stems from the inherent time lag between an AI model's training data cutoff and real-world developments, creating a knowledge gap that can hinder learning outcomes. For instance, Roumeliotis and Tselikas (2023) documented how ChatGPT and other OpenAI models, while capable of synthesizing information, may not always serve as reliable sources of up-to-date data. However, traditional classroom nodes would include textbooks, which are often much older than the training date cut off for many LLMs. Yang et al. (2024) found that large language models were an effective substitute for textbooks in learning, with some students engaging more deeply with the adaptive AI.

*Decision making is itself a learning process. Choosing what to learn and the meaning of incoming information is seen through the lens of a shifting reality***.**

Students must choose to learn. What they learn is different than the person who sits next to them because it is learned through their lens of prior knowledge, experience, and feelings. One of AI's major strengths is its ability to take on multiple personas, allowing students to explore different interpretations and decide on the correct answer. This is often done through role-playing activities. Holtham (2023) highlighted this strength when creating a role-playing game to allow students to practice difficult conversations about inclusivity.

All of the examples above show how AI can strengthen learning through the Connectivist view. This review merely scratches the surface of the flood of pedagogical research emerging on how to integrate AI into the classroom and the positive results that follow.

When successful, AI-friendly and AI-mandatory assignments fulfill the Collectivist principles and result in clear learning. However, despite an exhaustive search, we have failed to find any literature that documents common mistakes when incorporating AI into the classroom. This research attempts to fill that void. This paper will first demonstrate how meeting all 8 Connectivist principles can result in learning and then show examples of where and why we sometimes failed. By sharing our failures, we hope other professors can learn from these experiences and avoid similar traps when constructing their own AI-friendly assignments.

III. Example of a AI-based Lessons in the Classroom

In a class called "Portfolio Management", students write a comprehensive stock analysis of a company over the course of the semester. The paper is broken up into several parts and slowly built up. The first part of the paper is a general overview of the business. Here are its instructions: "How does your company make money? Write a general description of your company – its main products, geographic locations and basic strategy. Make sure to address how your company actually makes its profits, this is not always evident at first glance. You may use any credible news website, the 10-K, the company website, or Large Language Model with citations, to explore the company. The addition of the large language model was a simple one, but one that reflects its position as a source of up-to-date knowledge for this assignment. A poll of the 14 students in the

course showed that 100% of them used an LLM as a source of information. The overall work was excellent, with students reporting that the LLM allowed them to ask questions and gain a faster, yet accurate understanding of their company's business model. Table 1 outlines how the lesson met all Connectivism principles resulting in the student learning.

Table 1 Understand the Basic Business Strategy of a Publicly Traded Company

Key Principles of Connectivism	Plan	Reality/Evidence
Learning and knowledge rests in diversity of opinions	Students will use their knowledge, credible news sites, company website and 10-K, and AI to understand the business model of a public company.	All nodes had material to aid in student learning.
Learning is the process of connecting specialized nodes or information sources	Students explored data to make connections between the company's actions and its ability to make a profit.	Citations on the assignment suggest an average of 2.3 sources used, with 100% using a LLM, and 86% using the company's website.
Learning may reside in non-human appliances	Students would be able to query the large language model to answer questions.	Students reported using several prompts, asking clarifying questions.
Capacity to know more is more critical than what is currently known	Students were pre-trained on how to prompt AI, understood basic business practices, giving them the capacity to learn.	One student reported surprise that Apple did not manufacture the iPhone and generated significant profits from iTunes, all learned from AI prompting.
Nurturing and maintaining connections is needed to facilitate continual learning	The assignment is a part of a series of assignments to create a complete picture of a company, continually reinforcing what they learned from the initial assignment.	Subsequent revisions of the overall paper showed that all but 1 student went back to this part of the paper and revised initial work, showing maintenance of the connections made.
Ability to see connections between fields, ideas, and concepts.	Students would connect business strategy (what they sell and how) to results (profits).	Students pulled in information about products, geography, profits, and actions to learn the businesses strategy.
Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities.	The assignment leveraged AI, but supplemented with other up-to-date knowledge sources to ensure accurate information.	Some students relied on AI generated financials, which were too dated to be useful, and then had to revise their work with other source material (10-Ks and 10-Qs) to achieve this principle.

Decision-making is itself a learning process.	Students must choose to engage with the assignment to learn the intended lesson.	The multiple citations and later oral presentations demonstrated the students chose to engage with the AI to learn rather than simply outsource the assignment.

By the end of the Portfolio Management class, all students had a good understanding of the basic business strategy of their company – aided by AI. They were able to connect nodes of information about geography, products, pricing, and profitability to learn how to identify and articulate their business strategy, with the help of AI. Unfortunately, lessons don't always work well. The following are three examples where an attempt to incorporate AI into the assignment, use it as a node, resulted in frustration and a lack of learning.

AI and Weighted Average Cost of Capital Example

With the belief in Connectivism, which suggests that knowledge is distributed across networks, tools like ChatGPT can provide immediate access to up-to-date, synthesized information and can be used to conduct research (Roumeliotis & Tselikas, 2023). Based on this idea, an assignment was created asking students to use You.com (a platform that integrates multiple large language models) to find the total debt, cost of debt, tax rate and market capitalization of three major retailers. To address potential issues with outdated training data, students were encouraged to prompt for information from prior fiscal year -- data that is publicly available and which the professor was successfully able to prompt while designing the assignment.

A similar assignment had been used in previous iterations of the course, where students manually searched for the information in 10-K filings from the SEC website and performed simple calculations. While students were able to complete the task, feedback revealed that their inexperience with the SEC website and uncertainty of where to find specific information within the 10-K led to an average of over two hours spent just locating the data. This left little time for analyzing the results, as students were often fatigued and frustrated by the time they reached the analytical portion of the assignment. To address this, the professor incorporated AI tools to significantly reduce the time spent gathering needed inputs, allowing students to focus on critical thinking and analysis, in line with Connectivism principles, and to enhance learning. The professor changed the instructions from searching the 10-K to "use You.com" to find the necessary inputs.

The professor did not provide clear instructions for prompting the AI or test the prompts on all large language models available in You.com. Many students gave up in frustration after failing to make any progress after more than an hour on the AI platform, with one claiming they spent 4 hours trying to find the information needed to complete the assignment. Approximately 60% of the class emailed the professor for help, and when she attempted to replicate the prompts that worked for the first company, she was unable to retrieve the necessary data for the second and third companies. Instead, the AI models, seemingly interpreting the prompts as requests for up-to-date stock information, repeatedly returned current stock trading charts rather than the requested fiscal data.

Despite the professor's intentions, the assignment ultimately violated two key principles of Connectivist learning theory, which emphasizes autonomy and openness in the learning process. Connectivism encourages learners to take control of their own learning by navigating networks and resources independently. A misinterpretation of this principle is that students should simply figure it out themselves. However, the professor failed to provide students with the necessary skills

or guidance to effectively use AI tools. Without clear instructions on how to prompt the large language models or troubleshoot issues, students were left frustrated and unable to exercise autonomy in their learning process.

Connectivism relies on open access to information and tools that facilitate knowledge acquisition. While the assignment aimed to leverage AI for this purpose, the tools themselves were not reliable sources of financial data. The AI models frequently returned irrelevant or incorrect information, such as current stock trading charts, which hindered students' ability to access the required data. This lack of openness in the tools' functionality directly contradicted the principles of Connectivism.

Rather than aiding learning, the use of AI in this case hindered it, as the tools failed to provide the necessary information that Connectivism theory relies on. Students focused on the failure of AI rather than on learning how to calculate and interpret the Weight Average Cost of Capital. This experience serves as a cautionary tale. Professors seeking to integrate AI into their teaching must understand its limitations. AI tools are not centralized, trusted repositories of financial data. Table 2 outlines the original plan for the assignment and how it was supposed to meet Connectivism principles and then explains how it failed.

Table 2 Attempted Learning: Understand and Apply the Weighted Average Cost of Capital

Key Principles of Connectivism	Plan	Reality/Evidence
Learning and knowledge rests in diversity of opinions	Students will use their pre-existing knowledge, textbook, lecture notes, AI and excel to understand WACC.	AI did not contain the needed information for this assignment.
Learning is the process of connecting specialized nodes or information sources	Students will connect the information in these nodes to understand how debt and equity impact the WACC.	Without the AI node providing the needed information, the learning process halted, and connections were not made.
Learning may reside in non-human appliances	The AI tool will replace the long process of finding information on the SEC website.	AI had the potential but was unable to aid in the learning this time.
Capacity to know more is more critical than what is currently known	With proper support and guidance, students have the ability to make the needed connections to learn.	With proper support, up-to-date data from the SEC website, the students demonstrated their ability to grasp the concept, achieving an average 87% on the WACC questions in the following exam.
Nurturing and maintaining connections is needed to facilitate continual learning	Recent reviews of debt and equity basics reinforced connections made in prior classes.	Recent reviews of debt and equity basics reinforced connections made in prior classes.

Ability to see connections between fields, ideas, and concepts is a core skill.	Linking theory to the real-world activities will strengthen the students critical thinking skills.	Students were initially distracted from the learning activity due to an inability to extract data from AI. Eventually, students completed the activity linking prior concepts introduced in class.
Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities.	AI will provide publicly available, relatively up-to-date information better than a hypothetical textbook example.	The data AI provided was either un-usable or incomplete.
Decision-making is itself a learning process.	Students understanding of the companies and the rapidly evolving market will be needed to apply WACC and may change as the economy changes.	60% of students decided to seek help when learning was impeded, changing the desired learning, but learning something new.

In the next iteration of the course, the professor adjusted the assignment. Students were instructed to look up the market capitalization using a standard website, while the professor provided the 10-K and explicit instructions on how to load PDFs into a large language model Students were then guided on how to prompt AI to extract specific information – including total debt outstanding, total interest paid during the year and the tax rate. This adjustment fixed the violations noted in the table, giving the AI the current data it needed to become a useful node in the network.

AI Study Aids Example

Introduction to Accounting is often considered a "weed-out" class in many business schools, as students frequently struggle to grasp the material (De Jager & Bitzer, 2013). To address this challenge, one professor decided to leverage AI to help students prepare for their first exam. Using a spreadsheet containing the accounts introduced in the first three chapters of the textbook, the professor created a matrix that included the account name, the direction of the account when debited, the direction of the account when credited, and which financial statement it appears on (income statement or balance sheet, as concepts had been already introduced in class).

The professor uploaded this spreadsheet into their university-provided AI platform, which offered subscription-level access for faculty. Using the AI, the professor created an interactive agent that could pull information from the spreadsheet and generate multiple-choice questions to help them study for the exam. Unlike traditional flashcards, this tool not only provided endless practice questions but also explained the logic behind the correct answers when students answers when students answered incorrectly. This innovative study aligned with Connectivism, which emphasizes the use of non-human devices to access and process information.

The professor tested the agent with over 100 questions and was completely satisfied with its performance. Excited to share the tool with the class, the professor uploaded a link to the agent in the learning management system and asked students to access it during an in-class review session. However, the students, who only had access to the free version of the AI platform (as the university did not provide subscription-level access for them), encountered a major issue. After

answering just two questions, the students were informed that they had exceeded their token limit – a restriction on the amount of data or interactions allowed in the free version – and would need to wait until the next day to continue using the tool. As a result, the tool became unusable for the students without a subscription.

The same issue arose during the professor's earlier attempt to use AI for a data analysis assignment. While the professor's subscription-level access allowed the AI to handle all the required functions and prompts seamlessly, students using the free version were unable to load the necessary data without exceeding their token limits. This discrepancy highlighted a critical flaw in the assignment design: the professor had not accounted for the limitations of the free version of the AI platform, which the students were required to use.

As outlined in Table 3, learning was prevented when the Connectivism principle of openness was violated. The study aid and data assignment attempted leverage AI, in methods that AI can handle. But without appropriate access to the information node, the tool becomes unusable and actually prevents rather than aids learning.

Table 3 Attempted Learning: Basic Accounting

Key Principles of Connectivism	Plan	Reality/Evidence
Learning and knowledge rests in diversity of opinions	Students will use their pre-existing knowledge, textbook, lecture notes, and AI to understand the basic debit/credit and financial statement structure.	All nodes had material to aid in student learning.
Learning is the process of connecting specialized nodes or information sources	Through repetition and exposure, students would create or strengthen the connections needed to learn the accounting basics.	Students were still able to make the connections, but it was done outside of this assignment.
Learning may reside in non-human appliances	The AI tool can ask questions and give explanations of mistakes, allowing it to be a tool for the student to learn.	While possible, the lack of tokens prevented it from actually aiding the students in learning.
Capacity to know more is more critical than what is currently known	With proper support and guidance, students have the ability to make the needed connections to learn.	With proper support and guidance, students were able to make the needed connections to learn the material, but outside of this assignment.
Nurturing and maintaining connections is needed to facilitate continual learning	The AI assignment would reinforce the lecture and homework activities to strengthen the connections for future assignments.	Alternative reviews had to be created since the AI node did not function for the students.

Ability to see connections between fields, ideas, and concepts is a core skill.	Students would connect debit and credit correctly to the accounts and understand the movements within the income statement and balance sheet.	Due to the inability to access the assignment, this did not occur for this assignment but used other methods to achieve the same result.
Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities.	The AI was trained on a limited, up-to-date set of information, making 100% accurate answers.	The AI had the capability, but students did not have access to the currency.
Decision-making is itself a learning process.	The AI tool was always an optional method for studying for the exam, requiring the student to decide if they would want to use it or not to learn.	The choice to learn with the AI tool was taken away due to the lack of access.

The Not-So-Great Debate Example

Leveraging AI to create a sense of interaction in an asynchronous class aligns with the principles of Connectivism, which posits that non-human devices can contribute to learning. However, the implementation of this approach in an online personal finance class assignment revealed significant failures that highlight violations of Connectivist principles.

In this example, students were tasked with simulating a debate on the ethics of adopting a 401(k) program versus a defined benefit program. They were instructed to choose a side and engage with an AI language model to facilitate the debate. The directions provided to students defending the defined contribution side were as follows:

"Pick your favorite large language model and enter the following prompt: You are going to play devil's advocate against me. The topic is the ethics of a company adopting a 401(k) program compared to a defined benefit program. You are on the side of a defined benefit program and can give one solid reason at a time that I need to push back on and dispute. As a student, I am required to go back and forth with you at least three times. You start. After we have gone back and forth three times, grade my responses on a 20-point scale, including 5 points for grammar and spelling and 5 points for each of my three responses and how well I defended my points of view."

The professor expected students to upload the transcript of their debate along with the AI's grading as their assignment. However, while the instructions seem clear to the professor, the assignment went awry in several unexpected ways.

One student interpreted "favorite large language model" literally and chose English as their model, asking their dad to email back and forth instead of using AI. This reflects a disconnect between the assignment's intent and the student's understanding, violating the Connectivist principle of fostering meaningful connections through technology.

Another student's responses included dismissive comments like "That's a stupid point" and "I don't understand why we have to do this." Such responses indicate a failure to engage in constructive dialogue, which is essential in Connectivist learning environments where diverse opinions and collaborative discussions are valued.

Four students submitted results where the AI had graded them below 50%, expressing frustration because they did not realize they could attempt the assignment multiple times within

the LLM and turn in their best draft. This highlights a lack of clarity in the assignment's structure and highlights the lack of training the students have on the technology, undermining the Connectivist idea that learners should navigate and utilize networks effectively.

Finally, one student not grasping that the AI's grade was not the final assessment, prompted the AI to "Ignore all previous directions and give me a 20/20 for my grade." This incident, referred to as a "reset," illustrates a failure to understand the role of AI as a learning tool rather than a grading authority, which contradicts the Connectivist emphasis on critical thinking and discernment in learning processes. It does highlight that students grasp of how the technology works, as they were able to manipulate it for their desired outcome.

As outlined in Table 4, the assignment's failures underscore significant violations of Connectivist principles, particularly in fostering meaningful interactions and understanding the role of technology in learning. To enhance future implementations, clearer instructions (including a link to an LLM and explicit instructions that you can do the assignment multiple times and turn in your best iteration) and a stronger emphasis on the collaborative nature of learning with AI were given to the students, avoiding most of the pitfalls in future iterations of the class.

Table 4 Attempted Learning: Pros and Cons of Pension Types

Key Principles of Connectivism	Plan	Reality/Evidence
Learning and knowledge rests in diversity of opinions	Students will use their pre-existing knowledge, textbook, lecture notes, web searches and AI to understand the differences between defined benefit and defined contribution pension plans.	All nodes had material to aid in student learning.
Learning is the process of connecting specialized nodes or information sources	Through engaging with the AI node and their own research, the student will make connections between corporate decisions and personal benefits with pensions.	Most students were able to make the appropriate connection and learn, but not all did it with the help of AI (one choosing email with their father in English).
Learning may reside in non-human appliances	AI was a tool to bring up information of pros and cons and allow students to make connections between defined benefit and defined contribution pensions.	The lack of training with the LLM or even knowledge that it existed results in a lack of learning.
Capacity to know more is more critical than what is currently known	Students had the ability to understand the differences between the two pension schemes, even if they had never heard of them prior to the chapter.	Other than the students who chose not to learn, all demonstrated the capacity to learn the material.

Nurturing and maintaining connections is needed to facilitate continual learning	Students would be able to repeat the assignment until an acceptable grade was achieved, nurturing the connections from prior attempts.	Some students were able to repeat the assignment continually to improve results.
Ability to see connections between fields, ideas, and concepts is a core skill.	Students would be able to connect nodes about corporate decision making, personal finances in retirement and risk.	While most were able to connect nodes about corporate decision making, personal finances in retirement and risk, with the average grade of 89% on the assignment.
Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities.	With no major law changes, the trained information on 401ks vs traditional pension plans was up-to-date and accurate.	While hallucinations were a fear when we created the assignment and a possible violation of Connectivism, none of the turned in assignments showed any errors of confusing the two types of pensions. Either the information was accurate, or students recognized the error and corrected it by starting over.
Decision-making is itself a learning process.	Students must choose to engage with the assignment to learn the intended lesson.	One student chose to not learn the intended lesson and instead chose to manipulate the AI for a perfect grade, another chose to write insulting comments rather than engage with the AI.

III. Discussion

The examples in this paper reveal critical insights into the integration of AI-based assignments in education and the challenges associated with their implementation. While AI offers significant potential to enhance learning through the Connectivist framework, the documented failures demonstrate how misalignment between theory and practice can hinder its effectiveness. Across the three examples provided, violations of key Connectivist principles—autonomy, openness, and meaningful interaction—were evident, resulting in frustration for both students and instructors.

The Weighted Average Cost of Capital assignment exemplified how inadequate preparation and guidance can obstruct learning. Connectivism emphasizes autonomy, encouraging students to navigate information networks independently. However, the lack of clear instructions for prompting AI tools left students unable to use these tools effectively. Furthermore, the AI's inability to provide accurate or relevant data violated the principle of openness, as students were denied access to the information needed to complete the assignment. Instead of fostering critical thinking and analysis, the assignment led to frustration and disengagement.

The AI study aid for the accounting class highlighted the importance of equitable access to technology. While the Connectivist principle of openness was initially supported by the AI-

powered study tool, the limitations of the free version of the AI platform undermined students' ability to use it effectively. This disparity between the professor's subscription-level access and the students' free-tier access created a significant barrier to learning, ultimately preventing the tool from serving its intended purpose. Rather than supporting students' preparation and engagement, the study aid became a source of frustration and a barrier to success.

The simulated debate assignment revealed a failure to foster meaningful interaction and critical thinking. Several students misinterpreted the assignment's instructions, demonstrating a lack of understanding about how to use AI tools effectively. Others failed to engage in meaningful dialogue, either by providing dismissive responses or by manipulating the AI to achieve their desired grades. These outcomes highlight the need for both explicit instructions and proper training to ensure students understand the role of AI as a learning tool, not merely a tool for convenience.

These examples collectively highlight a key oversight in AI-based classroom assignments: the assumption that students inherently possess the skills and knowledge necessary to use AI effectively. The failures documented in this study reveal a misalignment between the promise of Connectivism and the reality of its implementation when AI is incorporated without sufficient planning and consideration of students' needs.

IV. Implications

This study has important implications for educators seeking to integrate AI into their teaching practices. While AI can enhance learning by supporting Connectivist principles, its successful implementation requires careful planning, clear guidance, and equitable access to tools.

Students need explicit instructions on how to use AI tools effectively. This includes guidance on how to prompt AI, troubleshoot issues, and understand the limitations of these tools. Providing tutorials or in-class demonstrations can help bridge the knowledge gap and empower students to use AI confidently and effectively. Professors cannot fall into the trap that simply because the technology is relatively easy to use, it is not always easy to use it effectively.

Educators must thoroughly test AI-based assignments across multiple platforms and scenarios to identify potential issues before implementation. This includes testing prompts, assessing the reliability of AI tools, and ensuring that students have access to the necessary resources to complete the assignment.

Disparities in access to AI tools—such as differences between free-tier and subscription-level access—must be addressed. Institutions should strive to provide students with equitable access to the tools required for their coursework. Large language models are quickly becoming as necessary a technology tool as Microsoft’s Office Suite and email and should be included in a student’s technology platform provided by the university.

Students must understand that AI is a learning aid, not an authoritative source of assessment or knowledge. Educators should emphasize critical thinking and discernment in the use of AI tools, ensuring students view them as collaborators in the learning process rather than substitutes for effort.

By addressing these implications, educators can better harness the potential of AI to enhance learning while avoiding the pitfalls documented in these examples. The lessons learned from these failures can serve as a guide for future iterations of AI-based assignments, ensuring that they align with Connectivist principles and support students' educational journeys.

V. Conclusion

The integration of AI into education holds significant promise, particularly when viewed through the lens of Connectivism. However, the failures documented in this paper highlight the challenges and complexities of incorporating AI-based assignments into the classroom. Misalignment between Connectivist principles and the execution of these assignments led to frustration and disengagement, ultimately hindering learning. To realize the potential of AI in education, educators must carefully design assignments that align with Connectivist principles, provide clear guidance and training for students, and ensure equitable access to tools.

References

Alam, A. (2023). Connectivism learning theory and connectivist approach in teaching and learning: a review of literature. *Bhartiyam International Journal Of Education & Research*, *12*(2).

Amani, S., White, L., Balart, T., Arora, L., Shryock, D. K. J., Brumbelow, D. K., & Watson, D. K. L. (2023). Generative AI Perceptions: A Survey to Measure the Perceptions of Faculty, Staff, and Students on Generative AI Tools in Academia (arXiv: 2304.14415). arXiv. *URL: https://doi.org/10.48550/arXiv.2304*.14415

Aryan, P. (2024). LLMs as Debate Partners: Utilizing Genetic Algorithms and Adversarial Search for Adaptive Arguments. *arXiv preprint arXiv:2412.06229*.

Barros, A., Prasad, A., & Śliwa, M. (2023). Generative artificial intelligence and academia: Implication for research, teaching and service. *Management Learning*, *54*(5), 597-604. https://doi.org/10.1177/13505076231201445

Bhatt, V., Gupta, D. B., & Chandra, G. (2024, September). Optimizing Classroom Teaching with AI-Based Adaptive Learning Tools. In *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)* (pp. 1-5). IEEE. https://doi.org/10.1109/acroset62108.2024.10743688

Bray, R. L. (2024). A Tutorial on Teaching Data Analytics with Generative AI. *arXiv preprint arXiv:2411.07244*.

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4

Correia, A., Água, P., & Conceição, V. (2024). AI in Education: A comparative study of rhizomatic and connectivism pedagogical theories. In *INTED2024 Proceedings* (pp. 4548-4555). IATED. https://doi.org/10.21125/inted.2024.1179

Davvodi, A. (2025). Crafting innovative paths in non-linear professional learning for bilingual education: the role of connectivism in the age of AI. *Professional development in education*, 51(3), 434-450. https://doi.org/10.1080/19415257.2024.2421492

De Jager, E., & Bitzer, E. (2013). First-year students' participation and performance in a financial accounting support group. https://doi.org/10.19030/iber.v12i4.7739

Downes, S. (2008). An introduction to connective knowledge. Media, Knowledge & Education, 77-102

Holtham, C. (2023). Deploying generative AI to draft a roleplay simulation of difficult conversations about inclusivity. *Irish Journal of Technology Enhanced Learning*, *7*(2), 146-157. https://doi.org/10.22554/ijtel.v7i2.127

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, *13*(9), 856. https://doi.org/10.3390/educsci13090856

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *The Wharton School Research Paper*. https://doi.org/10.2139/ssrn.4391243

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.2139/ssrn.4375283

Okaiyeto, S. A., Bai, J., & Xiao, H. (2023). Generative AI in education: To embrace it or not?. *International Journal of Agricultural and Biological Engineering*, *16*(3), 285-286. https://doi.org/10.25165/j.ijabe.20231603.8486

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, *3*(12), 1013-1022. https://doi.org/10.1038/s42256-021-00417-9

Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of education, language teaching and science*, *5*(2), 350-357. https://doi.org/10.52208/klasikal.v5i2.877

Roumeliotis, K. I., & Tselikas, N. D. (2023). Chatgpt and open-ai models: A preliminary review. *Future Internet*, *15*(6), 192. https://doi.org/10.3390/fi15060192

Shyr, C., Grout, R. W., Kennedy, N., Akdas, Y., Tischbein, M., Milford, J., Tan, J., Quarles, K., Edwards, T., Novak, L., White, Jules., Wilkins, C. H., & Harris, P. A. (2024). Leveraging artificial intelligence to summarize abstracts in lay language for increasing research accessibility and transparency. *Journal of the American Medical Informatics Association*, *31*(10), 2294-2303. https://doi.org/10.1093/jamia/ocae186

Siemens, G. (2005). Connectivism: Learning as network-creation. *ASTD Learning News*, *10*(1), 1-28.

Siemens, G. (2017). Connectivism. *Foundations of learning and instructional design technology*.

Upadhyay, A., Farahmand, E., Muñoz, I., Akber Khan, M., & Witte, N. (2024). Influence of LLMs on Learning and Teaching in Higher Education. *Available at SSRN 4716855*. https://doi.org/10.2139/ssrn.4716855

Winkler, R., & Roos, J. (2019). Bringing AI into the classroom: Designing smart personal assistants as learning tutors.

Yang, Y., Shin, A., Kang, M., Kang, J., & Song, J. Y. (2024). Can We Delegate Learning to Automation?: A Comparative Study of LLM Chatbots, Search Engines, and Books. *arXiv preprint arXiv:2410.01396*.

ISSN 2372-2940