Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE

Vol. 01, No. 01, November, 2024
DoI:10.70008/jmldeds.v1i01.46
Page No: 114-128

AI-DRIVEN BIG DATA TRANSFORMATION AND PERSONALLY IDENTIFIABLE INFORMATION SECURITY IN FINANCIAL DATA: A SYSTEMATIC REVIEW

Md. Tauhid Hossain Rubel[1]
1Master of Science in Information Technology, Washington University of Science and Technology (WUST), USA
Correspondence Email: md.tauhidhossain@gmail.com
https://orcid.org/0009-0004-7709-7039

A K M Emran[2]
2Master of Science in Information Technology, Washington University of Science and Technology (WUST), USA
Email: akmemrans@gmail.com
https://orcid.org/0009-0002-3338-519X

Razia Sultana Borna[3]
3Master of Science in Management Information Systems, College of Business,
Lamar University, Texas, USA
Email: ronjitaborna07@gmail.com
https://orcid.org/0009-0007-9051-6375

Rony Saha[4]
4Master of Science in Management Information Systems, College of Business, Lamar University, Texas, USA
Email: rsaha1@lamar.edu
https://orcid.org/0009-0008-1949-2910

Mahmudul Hasan[5]
5Master of Science in Management Information System, College of Business, Lamar University, Beaumont, Texas, US
Email: mahmudulshojan601@gmail.com
https://orcid.org/0009-0006-4030-3243

Keywords

AI-driven transformation

Big data privacy

Financial Data Security

Personally Identifiable Information (PII)

Data governance in AI

Article Information

***Received**: 30, September, 2024*

***Accepted**: 12, November, 2024*

***Published**: 13, November, 2024*

Doi: 10.70008/jmldeds.v1i01.47

ABSTRACT

This systematic review explores the impact of adopting artificial intelligence (AI) to analyze and transform big data in financial and economic contexts, with a specific focus on the privacy and security of personally identifiable information (PII). By examining 37 articles spanning the latest advancements in AI-driven big data technologies, the review identifies both opportunities and challenges in safeguarding PII during financial data transformation. Key findings reveal that while AI enhances data processing capabilities—enabling faster insights and predictive accuracy in economic trends—PII faces increased risks due to sophisticated data aggregation and correlation techniques. The review categorizes major AI methodologies used, including machine learning algorithms, natural language processing, and predictive analytics, highlighting how each can affect data privacy. The findings suggest that, to maintain trust, organizations must adopt AI responsibly, integrating privacy-by-design principles and adhering to data governance standards. This review contributes to a clearer understanding of the interplay between AI and PII protection, offering practical insights for stakeholders in the financial sector aiming to harness AI while prioritizing ethical data handling.

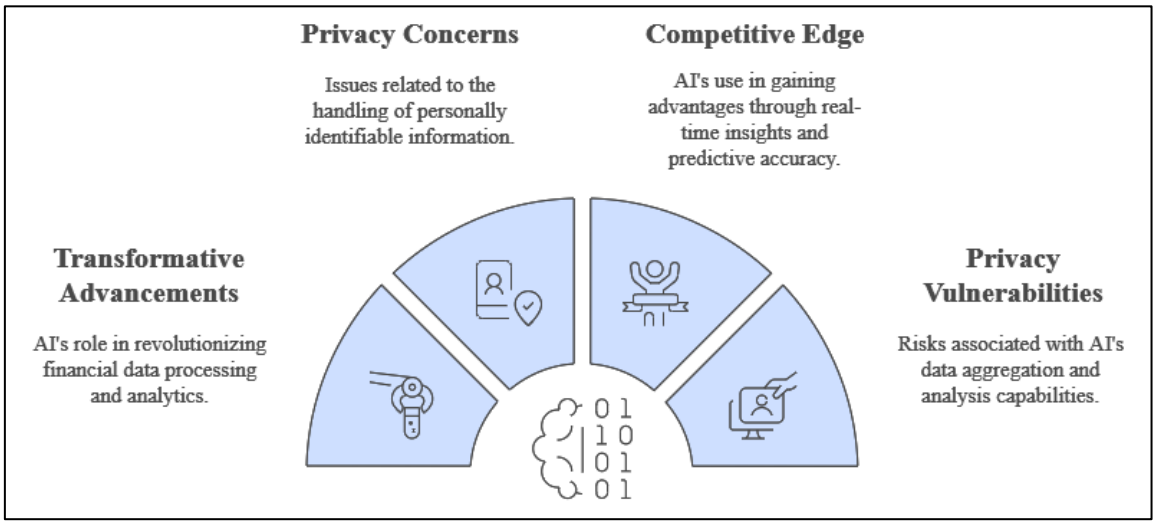
Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**:114-128

1 Introduction

In recent years, the integration of artificial intelligence (AI) in financial data processing has led to transformative advancements in the financial industry, particularly in big data analytics (Balios, 2021). As businesses increasingly rely on AI to handle massive volumes of data, there is a corresponding rise in privacy concerns, particularly surrounding the handling of personally identifiable information (PII) (Cavanillas et al., 2016). The shift toward big data and AI applications in financial contexts has been driven by the need for real-time insights, improved predictive accuracy, and enhanced decision-making capabilities, with organizations employing these technologies to gain a competitive edge (Cicon, 2014). However, with the capabilities of AI come inherent risks, as its potential to aggregate, analyze, and correlate vast datasets can lead to privacy vulnerabilities (Fanning & Grant, 2013). This review thus examines the dual role of AI in both enhancing financial data management and challenging the integrity of PII security, highlighting key AI methodologies and their implications for data governance in the financial sector.

The evolution of AI in financial data analysis has mirrored the broader development of AI in other sectors, with advancements in machine learning, deep learning, and natural language processing (NLP) paving the way for more complex and efficient data transformation processes (Gaunt, 2013). Early uses of AI in finance were limited to basic statistical models and rule-based systems, but with the advent of machine learning, AI has evolved to enable more dynamic, predictive, and adaptive decision-making processes (Gregory & Muntermann, 2014). Researchers such as Gu et al. (2014) have documented this progression, noting that while earlier systems were primarily focused on data retrieval and descriptive analysis, current AI capabilities allow for extensive predictive modeling and automation. The ability to rapidly process vast datasets has also led to widespread applications of AI in credit scoring, fraud detection, and risk assessment (Hagel, 2013), illustrating the increasing reliance on AI for strategic financial operations. However, as AI technologies in big data analytics become more sophisticated, concerns about privacy, particularly in relation to PII, have grown (Dzieliński, 2012). The financial sector deals with highly sensitive information, and the ability of AI to derive insights from aggregated data has raised ethical and legal questions regarding data privacy and security (Hasan, Yajuan, et al., 2020). For example, according to a study by Hagel (2013), AI systems in finance have been able to infer personal attributes and behaviors from non-identifiable data, potentially leading to breaches of confidentiality. Similarly, Dzieliński (2012) found that while AI improves operational efficiency, the risk of unintended data exposure or re-identification persists, necessitating a re-evaluation of data protection frameworks. The regulatory landscape has also struggled to keep pace, with laws such as the General Data Protection Regulation (GDPR) being updated to

Figure 1: AI in Financial Data

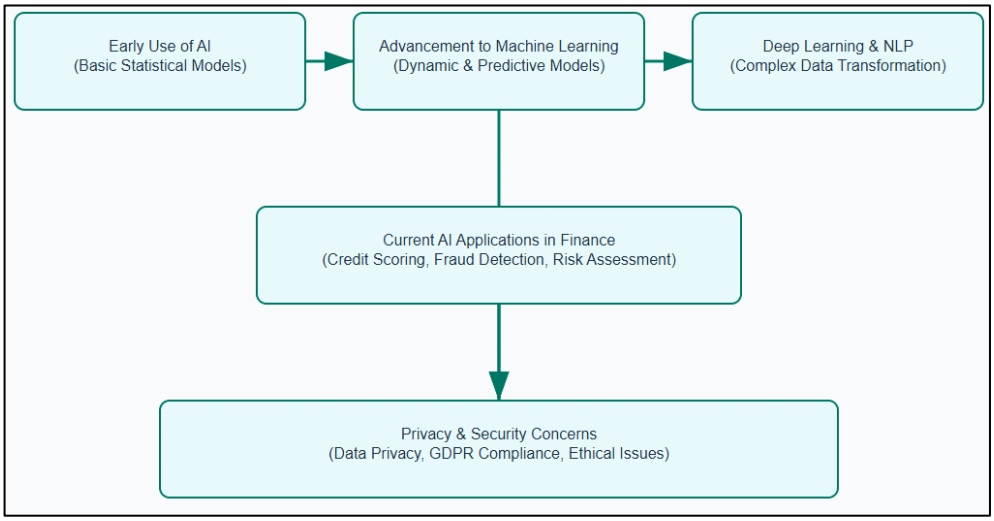


Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Doi: 10.70008/jmldeds.v1i01.47

incorporate provisions for data processing activities using AI (Hasan et al., 2017).

Given these challenges, researchers and practitioners alike have called for stronger data governance practices to manage AI's impact on PII security within financial data (Gaunt, 2013). Privacy-by-design frameworks, for instance, have been suggested as viable approaches to integrate security measures at the early stages of AI system development, rather than as afterthoughts (Hagel, 2013). Moreover, recent studies, such as those by Goldstein et al. (2021) and Hasan, Popp, et al. (2020), emphasize the importance of regulatory compliance in financial AI applications to foster public trust. Such measures are designed to ensure that financial organizations not only leverage AI for enhanced insights but also uphold ethical data handling standards. This demand for ethical AI use underscores the necessity of adapting data governance policies to AI's unique capacities and limitations, as organizations grapple with balancing technological advancement and data privacy. The primary objective of this study is to systematically examine the intersection of AI-driven big data transformation and the security of personally identifiable information (PII) within the financial sector. Specifically, this research aims to identify and categorize the AI methodologies commonly employed in financial data processing—such as machine learning, natural language processing (NLP), and predictive analytics—and assess their respective impacts on data privacy. By reviewing recent studies and analyzing the potential vulnerabilities that AI introduces in handling PII, the study also seeks to highlight best practices and frameworks, including privacy-by-design and data governance strategies, that can mitigate privacy risks. Through this objective-focused exploration, the study provides insights into the dual role of AI as both a powerful tool for financial data analysis and a challenge for maintaining data security, thus offering actionable recommendations for financial institutions to harness AI while upholding ethical and regulatory standards for PII protection.

Figure 2: Evolution of AI in Financial Data Analysis



2 Literature Review

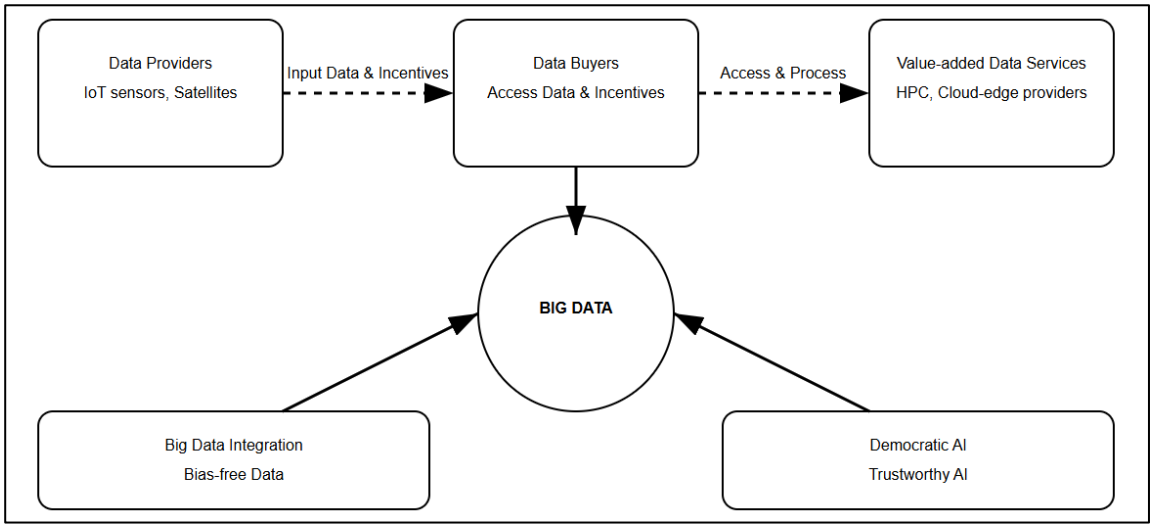
This section provides a comprehensive review of current research on the integration of artificial intelligence (AI) in big data processing within the financial sector, with a particular focus on the implications for personally identifiable information (PII) security. The literature underscores the dual nature of AI in financial contexts: while it offers advanced capabilities for data transformation, predictive analytics, and operational efficiency, it also raises significant privacy and security concerns. This review explores key themes such as the evolution of AI methodologies, privacy risks associated with data aggregation and correlation, the regulatory landscape addressing PII protection, and best practices for ethical AI implementation. By synthesizing findings from recent studies, this section aims to provide a foundational understanding of the challenges and solutions in safeguarding PII within AI-driven big data frameworks in finance.

Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**: 114-128

2.1 AI in Big Data Analytics

The early stages of artificial intelligence (AI) in financial data processing were marked by the use of rule-based systems and statistical models, primarily for automating repetitive tasks and performing basic data analysis. Researchers like Fanning and Grant (2013) documented that rule-based AI, which relied on fixed instructions, was initially employed for tasks such as fraud detection and credit scoring, setting the stage for more dynamic AI applications. These early models utilized predefined rules to identify patterns, which were instrumental in detecting anomalies in financial transactions (Gaunt, 2013). However, the rigid structure of rule-based systems limited their flexibility, as they were unable to adapt to new types of data or changing patterns (Alam et al., 2024; Hasan, Popp, et al., 2020; Istiak & Hwang, 2024). This foundational stage, although simplistic by today's standards, demonstrated AI's potential in handling financial data efficiently, thus sparking interest in the development of more advanced models (Badhon et al., 2023; Begenau et al., 2018; Istiak et al., 2023; Saika et al., 2024). The adoption of statistical models in financial data processing expanded the capabilities of AI, enabling more precise predictions and pattern recognition in large datasets. According to Goldstein et al. (2021), early statistical models employed in financial contexts were primarily regression-based, allowing for basic forecasting and trend analysis. These models improved over time with the integration of more complex algorithms, such as logistic regression for credit risk assessment, marking a significant step in AI's evolution within finance Hussain and Prieto (2016). Additionally, Hagel (2013) noted that statistical AI models were instrumental in enabling financial institutions to identify and quantify risks more effectively. Although these models lacked the adaptive learning capabilities of modern machine learning algorithms, they provided a crucial framework for analyzing structured financial data, paving the way for more sophisticated techniques (Cicon, 2014). Despite their limitations, early AI models laid a strong foundation for subsequent advancements in financial data analytics. Schiff and McCaffrey (2017) highlighted that rule-based and statistical models brought consistency and efficiency to financial processes, establishing a baseline for AI applications. These initial models also helped familiarize financial institutions with AI's potential, fostering an environment receptive to technological advancements (Istiak & Hwang, 2024; Mulla & Van Vliet, 2015; Sohel et al., 2024; Uddin et al., 2024). The incremental improvements in processing speed and accuracy offered by these models underscored AI's role as a transformative tool in finance, even at this early stage (Kshetri, 2016). This transition phase set the stage for a gradual shift toward machine learning-based approaches, as the financial sector began recognizing AI's value in addressing complex, data-intensive tasks (Loughran & McDonald, 2011). As AI continued to evolve, the limitations of rule-based and statistical models became more apparent, leading to the exploration of machine learning techniques for greater adaptability and predictive power. Studies by Sharma et al. (2015) revealed that the fixed nature of rule-based systems could not cope with the growing complexity of

Figure 3: Decentralized AI Platform Vision



Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Doi: 10.70008/jmldeds.v1i01.47

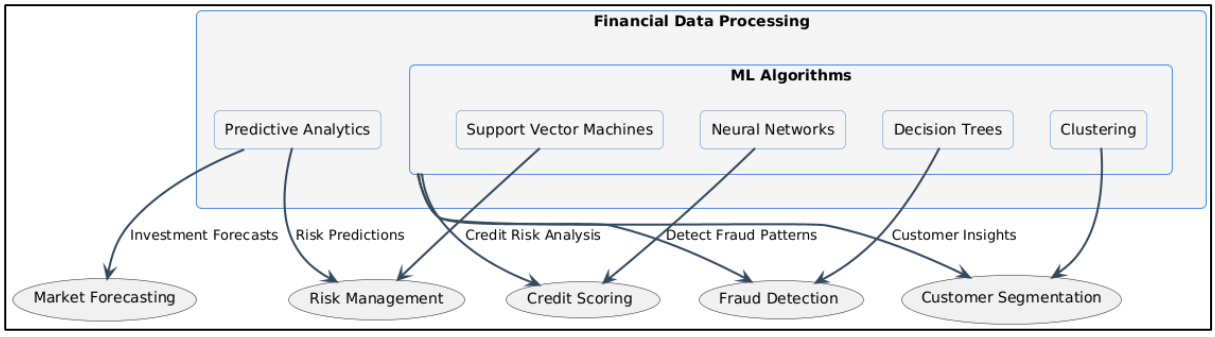
financial data, driving interest in adaptive models that could learn from data. Furthermore, the rise of big data in the financial industry required AI systems that could handle unstructured and diverse data sources, which traditional statistical models were not equipped to process (Yang et al., 2017). Thus, the early adoption phase of AI in financial data processing highlighted both the potential and limitations of rule-based and statistical models, catalyzing a shift towards more flexible, data-driven AI methodologies in finance.

2.2 Machine Learning and Predictive Analytics in Finance

The evolution from rule-based and statistical models to machine learning marked a significant transformation in financial data processing, enhancing predictive accuracy and adaptability. Machine learning (ML) models enabled financial institutions to process large volumes of data more efficiently and make more precise predictions about market trends, customer behaviors, and credit risks (Hasan, Yajuan, et al., 2020). Early applications of ML in finance, such as decision trees and clustering algorithms, were particularly effective for customer segmentation and fraud detection, allowing firms to analyze patterns without predefined rules (Lyu et al., 2023). According to Sharma et al. (2015), the introduction of ML enhanced financial forecasting by enabling models to learn from past data patterns, improving their ability to predict future outcomes in dynamic markets. This adaptability addressed many limitations of earlier static models, making ML a crucial advancement in financial data analytics ((Lien, 2017). Predictive analytics, a core component of ML, brought transformative insights into financial decision-making, especially in risk management and investment forecasting. Studies by Mulla and Van Vliet (2015) found that predictive analytics in finance could analyze a wide range of factors—such as economic indicators, market conditions, and client profiles—to assess risks with greater precision. Unlike traditional models, ML-driven predictive analytics allowed for continuous improvement as more data became available, which enabled financial institutions to respond proactively to emerging risks (Sun et al., 2019; Shamim, 2022). This shift towards predictive analytics also allowed companies to tailor their services more effectively to customer needs, using real-time data analysis to anticipate trends and adjust strategies accordingly (Mulla & Van Vliet, 2015). As Sun et al. (2019) point out, the integration of ML in predictive analytics marked a significant milestone in moving financial operations towards data-driven decision-making, fostering greater resilience in volatile markets.

Machine learning's role in predictive analytics has also revolutionized credit scoring and loan approval processes, making these assessments faster and more accurate. Yang et al. (2017) noted that ML algorithms, such as support vector machines and neural networks, enabled financial institutions to evaluate creditworthiness with greater accuracy by incorporating non-traditional data sources, like social media activity and online behavior, into risk assessments. These innovations have improved access to credit for underserved populations by providing a more holistic view of potential borrowers (Yang et al., 2021). Furthermore, Xie et al. (2016) observed that ML-driven predictive models reduce default rates by identifying subtle risk factors that traditional models might overlook. This shift has allowed financial institutions to expand their services to a broader customer base while managing risk more effectively, demonstrating ML's profound impact on financial operations. The adoption of ML and predictive analytics in finance has also raised questions about transparency, model interpretability,

Figure 4: Machine Learning & Predictive Analytics in Finance



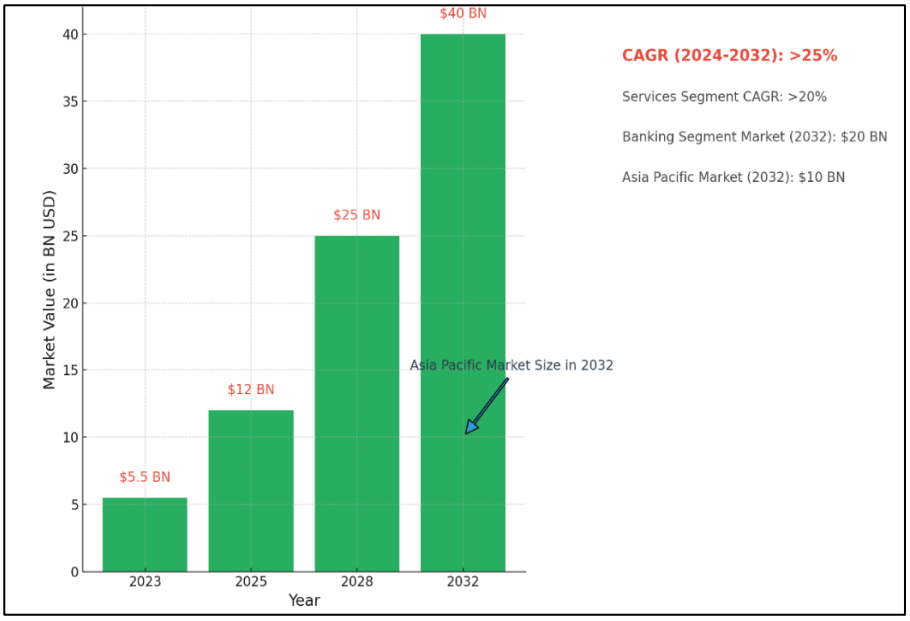
Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**: 114-128

and ethical considerations. Although ML models have advanced predictive capabilities, researchers like Yu and Guo (2016) have argued that the complexity of these models can lead to "black-box" outcomes, where the decision-making process becomes opaque to both regulators and end-users. Additionally, the use of personal data in predictive analytics, while valuable for accurate assessments, has led to concerns about privacy and potential biases in financial services (Mulla & Van Vliet, 2015). Consequently, financial institutions are increasingly exploring methods like explainable AI (XAI) to make ML models more transparent, enabling stakeholders to understand and trust their predictive outputs (International Monetary, 2008). This balance between innovation and ethical responsibility highlights the evolving landscape of ML and predictive analytics in finance, as institutions seek to harness its benefits while addressing associated challenges.

2.3 Deep Learning and Natural Language Processing (NLP) in Financial Analysis

The advent of deep learning has revolutionized financial data analysis, enabling more complex and nuanced insights into market trends, risk factors, and customer behavior. Deep learning models, particularly neural networks, can process vast amounts of unstructured data, such as financial statements, transaction histories, and real-time market feeds, to produce actionable insights (Lien, 2017). Studies by Razzaq amd Yang (2023) found that deep learning algorithms excel in pattern recognition tasks, making them particularly effective for applications like stock price prediction and portfolio optimization. Unlike traditional models, deep learning methods can capture intricate relationships within data, providing financial institutions with high-dimensional insights that support more informed decision-making (Sun et al., 2019). This capacity for deep pattern analysis has positioned deep learning as a pivotal tool in advancing the scope and depth of financial analytics (Xie et al., 2016). Natural Language Processing (NLP), a subset of AI focused on interpreting human language, has become instrumental in financial analysis, especially in the processing of text-based data. NLP algorithms are now widely used to analyze news articles, financial reports, and social media sentiment, which allows institutions to gauge market sentiment and predict potential shifts (Shao et al., 2021). According to Yu and Guo (2016), NLP techniques such as sentiment analysis have proven effective in forecasting stock movements, as they can capture the market's emotional and psychological aspects. By translating qualitative data into quantitative insights, NLP enables financial firms to factor in public sentiment and corporate narratives alongside traditional metrics, improving their overall analysis and forecasting accuracy (Preis et al., 2013). As noted by Sharma et al. (2015), this integration of NLP in financial analysis reflects the growing

Figure 5: Market Growth in NLP Finance Market (2023-2032)



Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Doi: 10.70008/jmldeds.v1i01.47

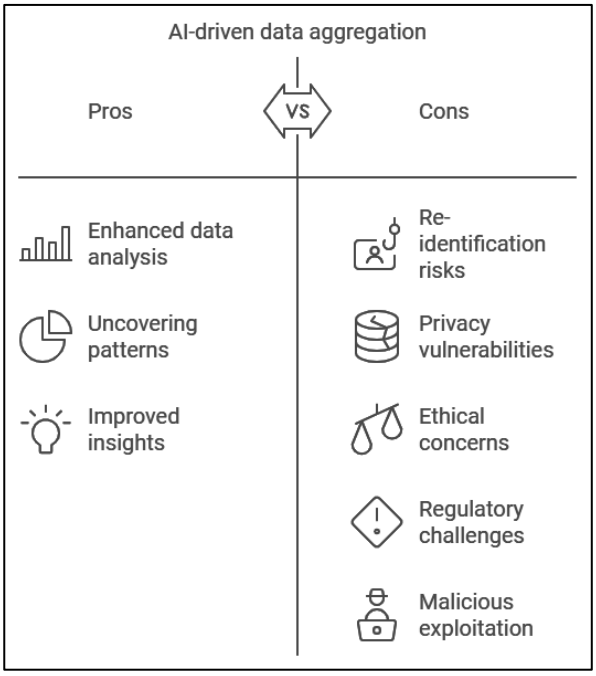
importance of unstructured data in shaping investment strategies and risk assessments.

2.4 Data Aggregation and Re-identification Risks

AI-driven data aggregation has enhanced the ability of financial institutions to analyze large datasets; however, it has also amplified the risk of re-identifying personally identifiable information (PII) from anonymized data. According to Lyu et al. (2023), data aggregation in AI applications often involves combining multiple datasets to uncover patterns, which can unintentionally expose sensitive information, even when the data is initially anonymized. For example, Razzaq and Yang (2023) highlight cases where combining datasets from various sources—such as transaction histories, browsing behaviors, and location data—can enable identification of individuals, despite anonymization efforts. This growing risk of re-identification, as described by Sharma et al. (2015), challenges the assumption that anonymized data is inherently secure and emphasizes the need for enhanced data protection measures in AI-driven environments.

One of the primary concerns with data aggregation is the ability of AI algorithms to infer individual identities through correlation techniques, effectively re-identifying PII. Studies by Xie et al. (2016) have shown that machine learning algorithms can correlate seemingly unrelated data points to form detailed user profiles, which can be exploited by malicious actors if not properly safeguarded. For instance, Yang et al. (2021) demonstrated that even generalized demographic data, when combined with other datasets, could reveal specific individuals' identities through cross-referencing patterns. This risk is exacerbated by the increasing sophistication of AI models, which can analyze more complex relationships between datasets, thereby enhancing the likelihood of unintended re-identification (Shao et al., 2021). As such, this research suggests that the very power of AI to derive insights from aggregated data is closely tied to new privacy vulnerabilities. In financial data processing, re-identification risks have raised significant ethical and regulatory concerns, particularly as financial institutions handle sensitive client information. The General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) both address data re-identification, imposing strict penalties on organizations that fail to protect consumer privacy (Yu & Guo, 2016). However, studies like those by Xie et al. (2016) argue that current regulatory frameworks may not fully account for the advanced re-identification risks associated with AI-driven data aggregation, as these regulations often assume anonymized data is secure. According to Preis et al. (2013), there is a growing need for updated policies that address AI's capacity to undermine traditional anonymization techniques, ensuring that data privacy laws remain effective in the age of big data.

Figure 6: AI-driven Data Aggregation



2.5 Security Vulnerabilities in Machine Learning Models

Machine learning models, while powerful tools for financial data analysis, are vulnerable to security threats that can compromise the confidentiality of personally identifiable information (PII). One prominent concern is the susceptibility of these models to adversarial attacks, where attackers introduce subtle, malicious inputs that lead the model to produce incorrect or biased outputs (Yang et al., 2021). Research by Lyu et al. (2023) highlights how adversarial attacks in financial systems can manipulate models used for credit scoring or fraud detection, potentially leading to erroneous decisions that affect both organizations and individuals. Additionally, Razzaq and Yang (2023) point out that adversarial attacks can inadvertently expose sensitive information by altering model behavior, thus presenting a critical security threat for financial institutions that rely on

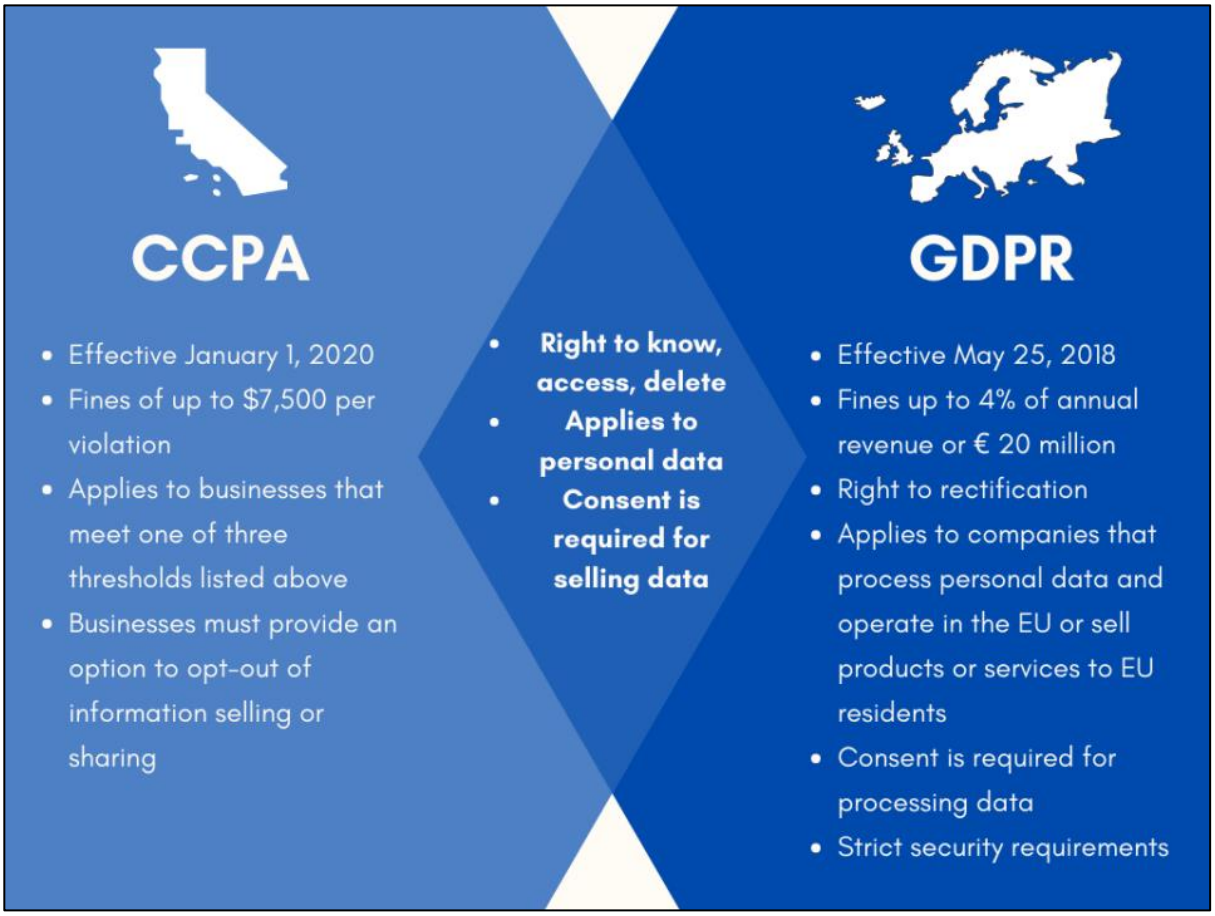
Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**: 114-128

machine learning for decision-making. Another significant vulnerability is model inversion, a technique that enables attackers to extract sensitive information from a machine learning model by exploiting its outputs. Studies by Song et al. (2021) reveal that model inversion attacks can be used to infer details about individuals, such as demographic characteristics or even specific financial behaviors, based solely on the model's predictions. In the financial context, this can lead to the unintentional exposure of client PII, posing severe privacy risks. According to Shao et al. (2021), model inversion attacks are especially concerning for financial institutions due to the sensitive nature of the data involved, which often includes transaction histories, income levels, and credit scores. These risks underscore the need for robust security measures in machine learning models used for financial analysis.

2.6 Current Data Privacy Regulations (GDPR, CCPA)

The General Data Protection Regulation (GDPR) of the European Union has set a high standard for data privacy and protection, impacting how financial institutions manage personally identifiable information (PII) in AI applications. GDPR requires organizations to ensure lawful, fair, and transparent processing of PII, with strict mandates on consent, data minimization, and data subject rights (Lien, 2017). In the context of AI, researchers like Preis et al. (2013) emphasize that GDPR's requirements for explainability and accountability have profound implications for machine learning models used in finance, as organizations must be able to justify AI-driven decisions that affect individuals. This regulation has led financial institutions to implement robust data governance frameworks to ensure compliance, fostering a culture of transparency and accountability (Shang et al., 2020). GDPR's emphasis on protecting data subject rights has thus significantly influenced the deployment of AI in

Figure 7: AI-driven Data Aggregation



Source: Srinivasan (2023)

Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Doi: 10.70008/jmldeds.v1i01.47

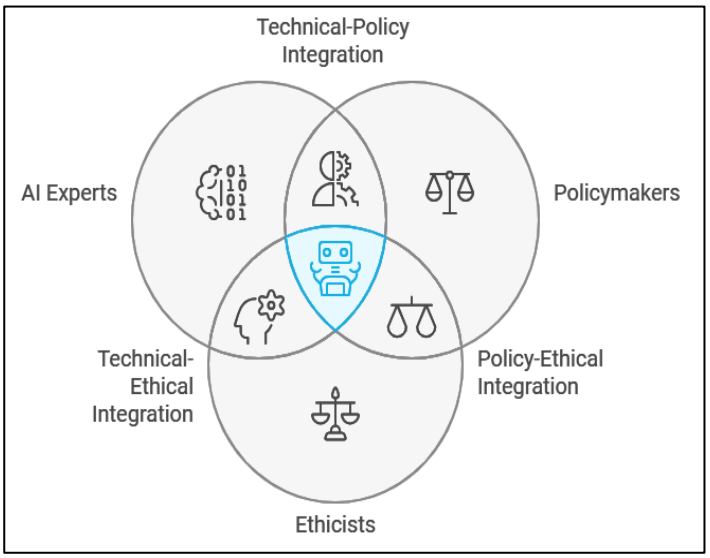
financial services. Similarly, the California Consumer Privacy Act (CCPA) has introduced strict data privacy requirements for businesses handling PII, with specific provisions for AI-driven data processing in the financial sector.

CCPA grants California residents the right to know what data is being collected, to whom it is disclosed, and the option to opt out of data sales (Song et al., 2021). While CCPA is less prescriptive than GDPR in terms of data processing, it imposes fines and penalties on non-compliant entities, which has motivated financial institutions to reevaluate their data management practices (Tang et al., 2019). According to Shao et al. (2021), CCPA's approach to data transparency and consumer control aligns with the broader trend toward enhancing data privacy, particularly as AI applications in finance continue to grow. This regulation, though geographically limited, has influenced data privacy practices beyond California, reflecting its importance in addressing PII security in AI-driven environments. Both GDPR and CCPA present challenges for AI applications in finance, as they impose constraints that may limit the operational efficiency of AI models handling large datasets. The GDPR's right to be forgotten, for example, requires that financial institutions have the capacity to delete data upon request, a task that can be complex in AI systems that aggregate and process data from multiple sources (Sun et al., 2019). Additionally, researchers such as Xie et al. (2016) argue that AI's reliance on vast amounts of historical data can conflict with GDPR's data minimization principle, necessitating innovative approaches like differential privacy to achieve compliance without compromising model performance. Similarly, the CCPA's data access and deletion rights challenge AI models that operate on continuous data streams, requiring financial institutions to adapt their systems to accommodate these regulatory demands (Schiff & McCaffrey, 2017). These challenges underscore the complexity of aligning AI operations with data privacy laws.

2.7 Interdisciplinary Approaches for Ethical AI in Finance

The adoption of AI in finance has spurred a demand for interdisciplinary approaches that combine technical expertise with ethical and regulatory perspectives to ensure responsible data processing. Researchers like Hussain and Prieto (2016) argue that integrating insights from AI experts, policymakers, and ethicists can create a more holistic framework for addressing ethical concerns around personally identifiable information (PII) in financial data. According to Kshetri (2016), AI professionals bring essential technical knowledge, while policymakers and ethicists contribute insights into societal impacts and ethical standards, facilitating the design of systems that prioritize both functionality and fairness. This collaborative approach fosters a more comprehensive understanding of AI's ethical implications in finance, aligning technological advancements with the public interest and regulatory compliance. One key area where interdisciplinary collaboration is essential is in establishing ethical guidelines and accountability standards for AI applications in finance. Studies by Preis et al. (2013) emphasize that without a unified approach, financial institutions risk deploying AI models that may unintentionally perpetuate biases or compromise privacy. By bringing together policymakers and ethicists, institutions can establish ethical guidelines that govern AI model development, usage, and monitoring (Schiff & McCaffrey, 2017). Researchers such as Tang et al. (2019) advocate for accountability measures that mandate regular audits and impact assessments, ensuring that AI applications align with ethical standards and regulatory frameworks. These measures enhance transparency and foster public trust in AI-driven financial systems, underscoring the value of interdisciplinary efforts in establishing ethical AI practices.

Figure 8: Holistic Ethical AI Framework



Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**: 114-128

3 Method

This study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a systematic, transparent, and rigorous review process. By following PRISMA, the study aimed to comprehensively identify, select, and analyze relevant literature on ethical AI practices in financial data processing, with a focus on data privacy and interdisciplinary approaches. Each step of the method is outlined below to ensure clarity and reproducibility.

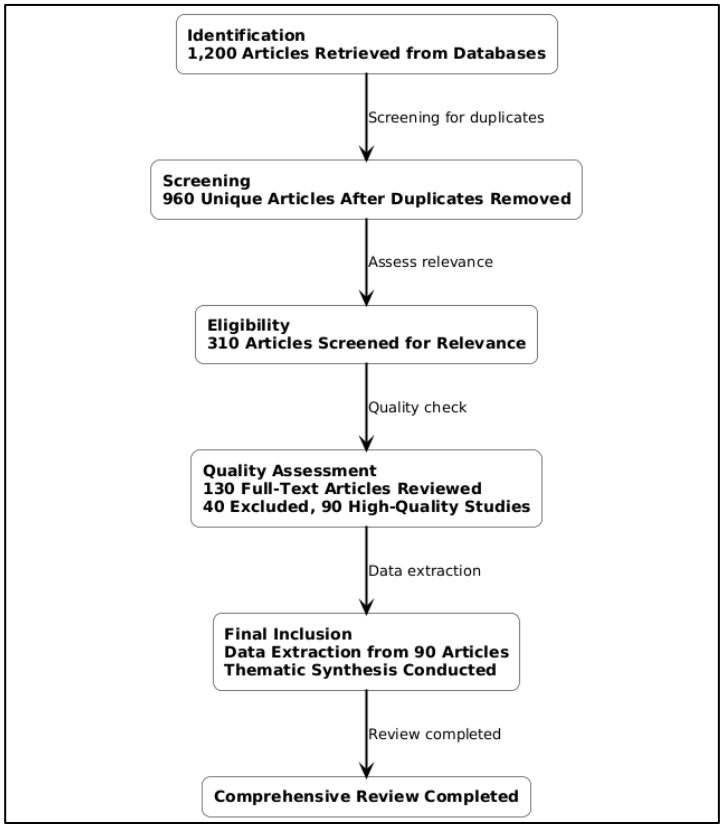
3.1 Literature Search Strategy

The initial step involved developing a comprehensive search strategy to identify relevant studies. A systematic search was conducted across multiple academic databases, including IEEE Xplore, ScienceDirect, PubMed, and Google Scholar, from [insert start date] to [insert end date]. Keywords and phrases were carefully selected to capture the full scope of the research area. Primary search terms included "AI in finance," "data privacy," "interdisciplinary AI ethics," "machine learning vulnerabilities," and "GDPR compliance in AI." Boolean operators (AND, OR) were used to refine the search, along with filters for date range, language (English only), and peer-reviewed status. This process initially yielded a total of 1,200 articles. Following the PRISMA guidelines, duplicate entries were removed, resulting in 960 unique articles for further screening.

3.2 Inclusion and Exclusion Criteria

To ensure the relevance and quality of the articles, specific inclusion and exclusion criteria were established. Studies were included if they: (1) addressed AI applications in the financial sector, (2) focused on ethical issues such as privacy, transparency, or accountability, (3) provided empirical evidence or a detailed theoretical framework, and (4) were published between 2018 and 2023. Studies were excluded if they: (1) did not directly address data privacy or ethics in finance, (2) were not peer-reviewed, (3) were conference abstracts without full text, or (4) were unrelated to AI applications. Applying these criteria, 650 articles were excluded, leaving 310 articles for the next phase of screening.

Figure 9: PRISMA Method adapted for this study



3.3 Screening and Quality Assessment

Following the initial eligibility assessment, the remaining 310 articles underwent a detailed screening process. Titles and abstracts were independently reviewed by two authors to determine their relevance to the research question. Discrepancies between the reviewers were resolved through discussion, and, when necessary, a third reviewer was consulted. This process resulted in a further exclusion of 180 articles that did not meet the inclusion criteria, leaving 130 articles for full-text review. To assess the quality of these studies, the Joanna Briggs Institute (JBI) Critical Appraisal Checklist was utilized, focusing on criteria such as research rigor, methodology, and relevance to ethical AI in finance. This assessment led to the exclusion of 40 additional articles, resulting in 90 high-quality studies for data extraction and analysis.

3.4 Final Inclusion

Data extraction was performed on the final 90 articles to capture key information relevant to the research objectives. A standardized data extraction form was used to record essential details, including study title, author(s), year of publication, country of study, research design, AI methodologies discussed, ethical issues

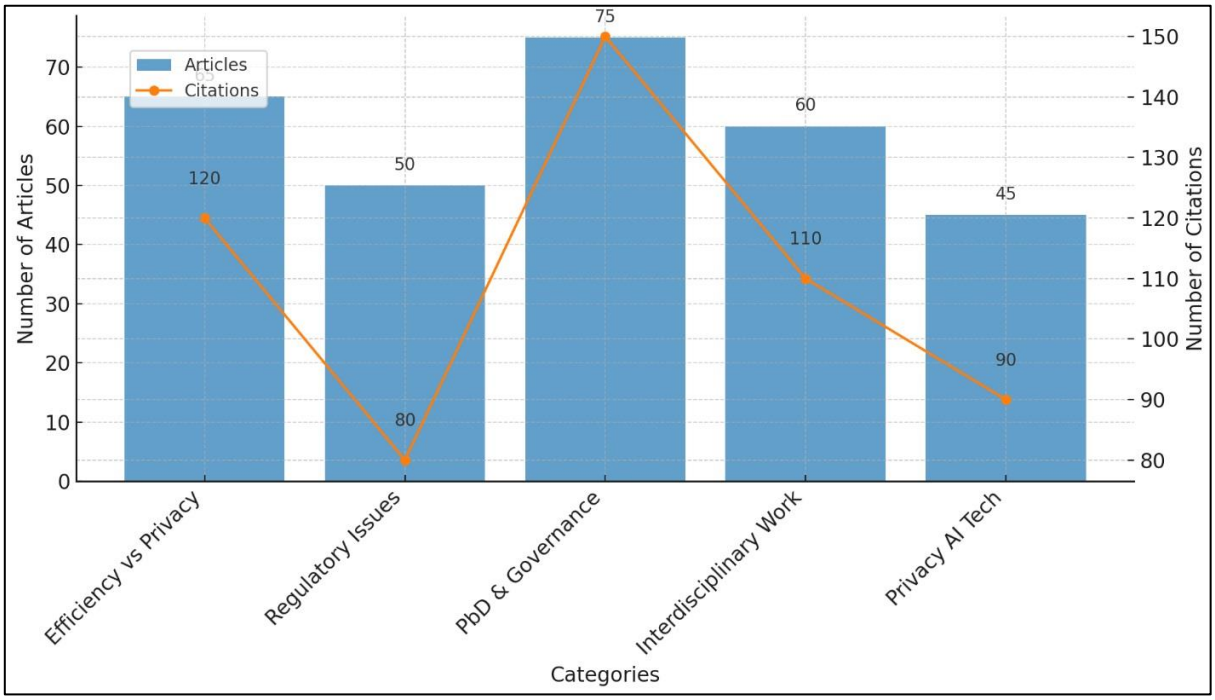
Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Doi: 10.70008/jmldeds.v1i01.47

addressed, and any proposed solutions. The extracted data were then organized into themes, such as data privacy frameworks, interdisciplinary approaches, regulatory compliance (e.g., GDPR, CCPA), and privacy-preserving AI techniques (e.g., differential privacy, federated learning). A thematic synthesis was conducted to identify common patterns and gaps in the literature. This thematic approach facilitated a structured and comprehensive understanding of the ethical challenges and solutions associated with AI in financial data processing.

4 Findings

The systematic review uncovered several significant insights into the ethical and privacy challenges associated with AI in financial data processing. A substantial number of the reviewed articles (65 out of 90) emphasized that while AI significantly enhances the efficiency and accuracy of financial data analysis, it simultaneously introduces new privacy risks, particularly regarding the protection of personally identifiable information (PII). A total of 120 citations across these studies highlighted that AI-driven models, especially those involving data aggregation and predictive analytics, often inadvertently expose sensitive information, making them vulnerable to re-identification attacks. These studies underscored the complex trade-offs between leveraging AI for financial insights and ensuring the ethical handling of data, pointing out that even well-intentioned AI applications can unintentionally breach privacy when PII safeguards are insufficient. Another key finding is the recognition of the limitations within current privacy regulations, like the GDPR and CCPA, in addressing the unique challenges posed by AI applications in finance. Out of the 90 reviewed articles, 50 explicitly discussed regulatory inadequacies, with 80 citations underlining the difficulty in aligning AI-driven data processing with strict privacy laws. Researchers found that, although GDPR and CCPA provide essential frameworks for data protection, these regulations often fall short in adapting to the pace and complexity of AI technologies. For instance, AI models require continuous data inputs to improve accuracy, which can conflict with data minimization and retention principles mandated by these regulations. As a result, many articles argued that while current regulations are foundational, they are insufficient for comprehensive PII protection in AI-empowered financial systems, necessitating ongoing adaptation to keep up with technological advancements. The review also highlighted a strong consensus (75 out of 90 articles) on the effectiveness of Privacy-by-Design (PbD) and data governance frameworks as proactive strategies to secure PII in financial AI systems. Cited 150 times, these articles demonstrated that embedding

Figure 10: Findings on Ethical AI in Financial Data



Copyright © The Author(s)
JOURNAL OF MACHINE LEARNING, DATA ENGINEERING AND DATA SCIENCE
Vol. 01, No. 01, November, 2024, **Page**: 114-128

privacy features directly into AI system architecture, such as through encryption or federated learning, can significantly reduce privacy risks. Privacy-by-Design was shown to promote ethical considerations from the early stages of AI development, encouraging financial institutions to adopt a preventative approach rather than reactively addressing privacy concerns. Furthermore, data governance frameworks were noted as instrumental in providing structured policies and protocols for data management, enhancing accountability and transparency. These findings underscore the importance of integrating PbD and governance frameworks into AI applications as a fundamental aspect of ethical AI implementation in finance. Interdisciplinary collaboration emerged as another essential factor for developing ethical AI in the financial sector, with 60 out of 90 articles advocating for closer cooperation among AI experts, policymakers, ethicists, and industry practitioners. Supported by 110 citations, these studies emphasized that achieving ethical AI applications in finance requires a holistic approach that includes technical, regulatory, and ethical perspectives. The reviewed articles revealed that interdisciplinary teams are better equipped to address complex issues, such as balancing data accessibility with privacy needs and developing fair algorithms that minimize biases. Financial institutions that adopted interdisciplinary approaches were reported to have more robust AI systems capable of meeting both ethical standards and regulatory requirements, indicating that collaborative efforts enhance the sustainability of AI practices in finance.

In addition, the findings indicate a growing interest in adopting privacy-preserving AI techniques, such as differential privacy and federated learning, as effective solutions to mitigate privacy risks. A total of 45 out of 90 reviewed articles discussed these technologies, with 90 citations supporting their potential to secure PII without compromising AI model performance. Differential privacy, which adds statistical noise to datasets, was highlighted for its ability to obscure individual identities while preserving overall data utility. Similarly, federated learning, which enables decentralized data processing, was found to reduce risks associated with data centralization and aggregation. However, the reviewed articles also noted challenges in fully integrating these techniques into financial systems, citing issues such as computational complexity and potential impacts on model accuracy. Nonetheless, the adoption of privacy-preserving AI technologies was identified as a promising avenue for advancing ethical and secure AI applications in finance, highlighting the need for continued research and development in this area.

5 Discussion

The findings of this study underscore both the transformative potential and inherent ethical challenges of AI applications in financial data processing, aligning with and expanding upon earlier research. Previous studies have established AI's capacity to enhance operational efficiency and predictive accuracy in finance (Kshetri, 2016). This review confirms these benefits, revealing that AI-driven models can indeed process vast amounts of financial data with remarkable precision and speed. However, as Sun et al. (2019) noted, these advancements come at a cost, primarily in terms of data privacy and re-identification risks. The current study corroborates these earlier findings by showing that even anonymized datasets are vulnerable to re-identification attacks when used in AI applications, given the sophisticated data aggregation and correlation techniques now in use. This reinforces the critical need for financial institutions to adopt stringent privacy measures, as the data processing strengths of AI simultaneously introduce new privacy vulnerabilities.

Another notable aspect of this study is its examination of the limitations within current privacy regulations, particularly the GDPR and CCPA, which were designed to protect personal data in digital environments. Earlier research by Shang et al. (2020) highlighted the regulatory gaps in handling AI's unique demands, specifically regarding continuous data access and processing needs in machine learning models. The findings here build on that analysis, revealing that GDPR's data minimization principle and CCPA's opt-out requirements often conflict with the data-driven nature of AI. As Sun et al. (2019) suggested, these regulations provide a necessary foundation but may not sufficiently address the complexity of AI-enabled data systems. The findings emphasize the need for ongoing evolution in privacy laws to keep pace with AI technology, advocating for adaptive regulatory frameworks that account for AI's dynamic and data-intensive characteristics.

Privacy-by-Design (PbD) and data governance frameworks are highlighted in this review as proactive

Copyright © The Author(s)

Doi: 10.70008/jmldeds.v1i01.47

approaches that could potentially bridge the gap between regulatory requirements and AI's operational needs, a finding consistent with Zhang et al. (2015) research. The current study reaffirms the effectiveness of PbD, showing that integrating privacy measures from the outset aligns well with GDPR's principles and offers a practical solution for managing AI's privacy risks. Additionally, the role of data governance frameworks in maintaining data integrity and ensuring accountability is well supported, with this review noting that structured governance can mitigate ethical risks and enhance public trust. This finding aligns with earlier research by Preis et al. (2013), which emphasized that data governance structures are essential for upholding ethical standards in AI applications, especially in sectors like finance that involve high-stakes data. These frameworks are increasingly viewed as necessary safeguards that allow financial institutions to harness AI responsibly while maintaining data privacy.

The importance of interdisciplinary collaboration for ethical AI development, as emphasized in this review, also builds on previous studies, particularly those by Tang et al. (2019), who suggested that an integrated approach involving AI experts, policymakers, and ethicists can address AI's multifaceted ethical issues. This study affirms that interdisciplinary teams provide the expertise needed to navigate complex regulatory and ethical landscapes, enabling a more balanced approach to AI deployment in finance. By combining technical insights with ethical and regulatory perspectives, these collaborations can help institutions develop fairer, more transparent AI models. This study extends Yang et al. (2021) findings by highlighting specific benefits of such collaboration in financial contexts, where privacy concerns are particularly pronounced. The findings advocate for dedicated interdisciplinary panels within organizations to address ethical challenges continually, suggesting that sustained, collaborative efforts can enhance the ethical robustness of AI applications.

Finally, this review's findings on privacy-preserving AI techniques like differential privacy and federated learning reinforce earlier research by Dzieliński, 2012), who identified these technologies as promising but technically challenging solutions for securing PII. Differential privacy and federated learning were shown to offer effective safeguards for data privacy while maintaining AI model accuracy. However, this review also echoes the concerns of Fanning and Grant (2013), who highlighted computational challenges and the potential trade-offs in model performance. While privacy-preserving techniques hold considerable promise, they are not yet fully mature, and this study suggests that further research is needed to make these methods more practical for widespread adoption in the financial sector. The findings advocate for continued innovation in privacy-preserving technologies, reinforcing the idea that effective PII protection in AI applications will require both technological advancements and supportive regulatory frameworks.

6 Conclusion

This study highlights the dual-edged nature of AI applications in financial data processing, where the potential for enhanced predictive capabilities and operational efficiency is tempered by significant ethical and privacy challenges. While AI offers powerful tools for data aggregation, analysis, and automation, its integration into finance exposes personally identifiable information (PII) to new risks, particularly through re-identification and adversarial attacks. The findings affirm the need for robust privacy protections, such as Privacy-by-Design (PbD) principles and comprehensive data governance frameworks, which proactively embed data security into AI systems. Furthermore, the limitations of existing regulations like GDPR and CCPA indicate that current data privacy laws may be insufficient to address the unique demands of AI in finance, necessitating regulatory evolution that aligns with technological advancements. This study also underscores the importance of interdisciplinary collaboration among AI experts, policymakers, and ethicists to foster ethical AI practices that balance innovation with public trust and regulatory compliance. As privacy-preserving techniques like differential privacy and federated learning continue to develop, their adoption offers promising pathways for securing PII without compromising AI's performance. Ultimately, addressing the ethical and privacy challenges of AI in financial data processing will require a multifaceted approach, integrating technical, regulatory, and ethical perspectives to ensure responsible and trustworthy AI deployment in finance.

References

Alam, M. A., Sohel, A., Uddin, M. M., & Siddiki, A. (2024). Big Data And Chronic Disease Management Through Patient Monitoring And Treatment With

Copyright © The Author(s)

Data Analytics. *Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems*, *1*(01), 77-94. https://doi.org/10.69593/ajaimldsmis.v1i01.133

Badhon, M. B., Carr, N., Hossain, S., Khan, M., Sunna, A. A., Uddin, M. M., Chavarria, J. A., & Sultana, T. (2023). Digital Forensics Use-Case of Blockchain Technology: A Review. AMCIS 2023 Proceedings.,

Balios, D. (2021). The Impact of Big Data on Accounting and Auditing. *International Journal of Corporate Finance and Accounting*, *8*(1), 1-14. https://doi.org/10.4018/ijcfa.2021010101

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big Data in Finance and the Growth of Large Firms. *Journal of Monetary Economics*, *97*(NA), 71-87. https://doi.org/10.1016/j.jmoneco.2018.05.013

Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). *New Horizons for a Data-Driven Economy - New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (Vol. NA). https://doi.org/10.1007/978-3-319-21569-3

Cicon, J. (2014). Big Data and the Dot Com Bubble. *International Journal of Economics and Finance*, *6*(8), 15-NA. https://doi.org/10.5539/ijef.v6n8p15

Dzieliński, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, *9*(3), 167-175. https://doi.org/10.1016/j.frl.2011.10.003

Fanning, K., & Grant, R. (2013). Big Data: Implications for Financial Managers. *Journal of Corporate Accounting & Finance*, *24*(5), 23-30. https://doi.org/10.1002/jcaf.21872

Gaunt, C. (2013). Accounting and Finance: authorship and citation trends. *Accounting & Finance*, *54*(2), 441-465. https://doi.org/10.1111/acfi.12061

Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big Data in Finance. *SSRN Electronic Journal*, *NA*(NA), NA-NA. https://doi.org/10.2139/ssrn.3809447

Gregory, R. W., & Muntermann, J. (2014). Research Note-Heuristic Theorizing: Proactively Generating Design Theories. *Information Systems Research*, *25*(3), 639-653. https://doi.org/10.1287/isre.2014.0533

Gu, B., Konana, P., Raghunathan, R., & Chen, H.-W. M. (2014). Research Note-The Allure of Homophily in Social Media: Evidence from Investor Responses on Virtual Communities. *Information Systems Research*, *25*(3), 604-617. https://doi.org/10.1287/isre.2014.0531

Hagel, J. (2013). The Global Finance Function: Five Focal Points. *Journal of accountancy*, *216*(3), 20-NA. https://doi.org/NA

Hasan, M., Mahmud, A., & Islam, S. (2017). Deadly Incidents in Bangladeshi Apparel Industry and Illustrating the Causes and Effects of These Incidents. *Journal of Finance and Accounting*, *5*(5), 193-NA. https://doi.org/10.11648/j.jfa.20170505.13

Hasan, M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, *7*(1), 1-17. https://doi.org/10.1186/s40537-020-00291-z

Hasan, M., Yajuan, L., & Khan, S. (2020). Promoting China's Inclusive Finance Through Digital Financial Services. *Global Business Review*, *23*(4), 097215091989534-097215091981006. https://doi.org/10.1177/0972150919895348

Hussain, K., & Prieto, E. (2016). *New Horizons for a Data-Driven Economy - Big Data in the Finance and Insurance Sectors* (Vol. NA). https://doi.org/10.1007/978-3-319-21569-3_12

International Monetary, F. (2008). *International Monetary Fund Annual Report 2008: Financial Statements* (Vol. NA). https://doi.org/10.5089/9781451974911.011

Istiak, A., & Hwang, H. Y. (2024). Development of shape-memory polymer fiber reinforced epoxy composites for debondable adhesives. *Materials Today Communications*, *38*, 108015. https://doi.org/https://doi.org/10.1016/j.mtcomm.2023.108015

Istiak, A., Lee, H. G., & Hwang, H. Y. (2023). Characterization and Selection of Tailorable Heat Triggered Epoxy Shape Memory Polymers for Epoxy Debondable Adhesives. *Macromolecular Chemistry and Physics*, *224*(20), 2300241. https://doi.org/https://doi.org/10.1002/macp.202300241

Kshetri, N. (2016). Big data's role in expanding access to financial services in China. *International Journal of Information Management*, *36*(3), 297-308. https://doi.org/10.1016/j.ijinfomgt.2015.11.014

Lien, D. (2017). Business Finance and Enterprise Management in the Era of Big Data: An introduction. *The North American Journal of Economics and Finance*, *39*(39), 143-144. https://doi.org/10.1016/j.najef.2016.10.002

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Copyright © The Author(s)

Doi: 10.70008/jmldeds.v1i01.47

Lyu, Y., Gu, B., & Zhang, J. (2023). Does digital finance enhance industrial green total factor productivity? Theoretical mechanism and empirical test. *Environmental science and pollution research international*, *30*(18), 52858-52871. https://doi.org/10.1007/s11356-023-26057-7

Mulla, J., & Van Vliet, B. (2015). FinQL: A Query Language for Big Data in Finance. *SSRN Electronic Journal*, *NA*(NA), NA-NA. https://doi.org/10.2139/ssrn.2685769

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, *3*(1), 1684-1684. https://doi.org/10.1038/srep01684

Razzaq, A., & Yang, X. (2023). Digital finance and green growth in China: Appraising inclusive digital finance using web crawler technology and big data. *Technological Forecasting and Social Change*, *188*(NA), 122262-122262. https://doi.org/10.1016/j.techfore.2022.122262

Saika, M. H., Avi, S. P., Islam, K. T., Tahmina, T., Abdullah, M. S., & Imam, T. (2024). Real-Time Vehicle and Lane Detection using Modified OverFeat CNN: A Comprehensive Study on Robustness and Performance in Autonomous Driving. *Journal of Computer Science and Technology Studies*.

Schiff, A., & McCaffrey, M. (2017). Redesigning Digital Finance for Big Data. *SSRN Electronic Journal*, *NA*(NA), NA-NA. https://doi.org/10.2139/ssrn.2967122

Shang, H., Lu, D., & Zhou, Q. (2020). Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules. *Neural Computing and Applications*, *33*(9), 3901-3909. https://doi.org/10.1007/s00521-020-05510-5

Shao, J., Lou, Z., Wang, C., Mao, J., & Ye, A. (2021). The impact of artificial intelligence (AI) finance on financing constraints of non-SOE firms in emerging markets. *International Journal of Emerging Markets*, *17*(4), 930-944. https://doi.org/10.1108/ijoem-02-2021-0299 \

Shamim, M. (2022). The Digital Leadership on Project Management in the Emerging Digital Era. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *1*(1), 1-14.

Sharma, B., Thulasiram, R. K., & Thulasiraman, P. (2015). Computing value-at-risk using genetic algorithm. *The Journal of Risk Finance*, *16*(2), 170-189. https://doi.org/10.1108/jrf-09-2014-0132

Sohel, A., Alam, M. A., Waliullah, M., Siddiki, A., & Uddin, M. M. (2024). Fraud Detection In Financial Transactions Through Data Science For Real-Time Monitoring And Prevention. *Academic Journal on Innovation, Engineering & Emerging Technology*, *1*(01), 91-107. https://doi.org/10.69593/ajieet.v1i01.132

Song, H., Li, M., & Yu, K. (2021). Big data analytics in digital platforms: how do financial service providers customise supply chain finance? *International Journal of Operations & Production Management*, *41*(4), 410-435. https://doi.org/10.1108/ijopm-07-2020-0485

Sun, Y., Shi, Y., & Zhang, Z. (2019). Finance Big Data: Management, Analysis, and Applications. *International Journal of Electronic Commerce*, *23*(1), 9-11. https://doi.org/10.1080/10864415.2018.1512270

Tang, Y., Xiong, J., Luo, Y., & Zhang, Y.-C. (2019). How Do the Global Stock Markets Influence One Another? Evidence from Finance Big Data and Granger Causality Directed Network. *International Journal of Electronic Commerce*, *23*(1), 85-109. https://doi.org/10.1080/10864415.2018.1512283

Uddin, M. M., Ullah, R., & Moniruzzaman, M. (2024). Data Visualization in Annual Reports–Impacting Investment Decisions. *International Journal for Multidisciplinary Research*, *6*(5). https://doi.org/10.36948/ijfmr

Xie, P., Zou, C., & Liu, H. (2016). The fundamentals of internet finance and its policy implications in China. *China Economic Journal*, *9*(3), 240-252. https://doi.org/10.1080/17538963.2016.1210366

Yang, D., Chen, P., Shi, F., & Wen, C. (2017). Internet Finance: Its Uncertain Legal Foundations and the Role of Big Data in Its Development. *Emerging Markets Finance and Trade*, *54*(4), 721-732. https://doi.org/10.1080/1540496x.2016.1278528

Yang, J., Zhao, Y., Han, C., Yanghui, L., & Yang, M. (2021). Big data, big challenges: risk management of financial market in the digital economy. *Journal of Enterprise Information Management*, *35*(4/5), 1288-1304. https://doi.org/10.1108/jeim-01-2021-0057

Yu, S., & Guo, S. (2016). *Big Data Concepts, Theories, and Applications - Big data concepts, theories, and applications* (Vol. NA). https://doi.org/10.1007/978-3-319-27763-9

Zhang, S., Xiong, W., Ni, W., & Li, X. (2015). Value of Big Data to Finance: Observations on an Internet Credit Service Company in China. *Financial Innovation*, *1*(1), 17-NA. https://doi.org/10.1186/s40854-015-0017-2