International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Explainable AI for Financial Decision Systems: Improving Transparency and Trust in AI-Driven Finance

Krishna Prisad Bajgai[1](M.Phil. ICT-Scholar), Dr. Bhojraj Ghimire[2](PhD), Niraj Kumar Shah[3](MBA-Scholar), Netra Prasad Joshi[4](MA)
[1,2,3]Nepal Open University, Lalitpur, Nepal,
[4]Tribhuban University, Nepal.

Abstract- Artificial Intelligence (AI) and Machine Learning (ML) technologies are increasingly applied in financial institutions for credit scoring, fraud detection, algorithmic trading, and risk management. Although these techniques offer high predictive performance, many models operate as complex "black-box" systems whose decision-making processes are difficult to interpret. This lack of transparency creates challenges related to trust, fairness, and regulatory compliance. Explainable Artificial Intelligence (XAI) aims to provide transparency and interpretability to AI-based models by offering explanations for their predictions. This paper explores the role of explainable AI in financial decision systems, focusing on its applications in credit risk assessment, fraud detection, and financial forecasting. The study reviews existing explainability techniques such as SHAP, LIME, and interpretable models, and proposes a conceptual framework for integrating explainable AI into financial decision-making systems. The findings highlight that integrating explainability mechanisms improves trust, transparency, and regulatory compliance while maintaining model performance. The paper concludes with future research directions for developing trustworthy AI-driven financial systems.

Keywords: Explainable AI, Financial Decision Systems, Machine Learning, Credit Scoring, SHAP, LIME, Financial Risk Management.

I. INTRODUCTION

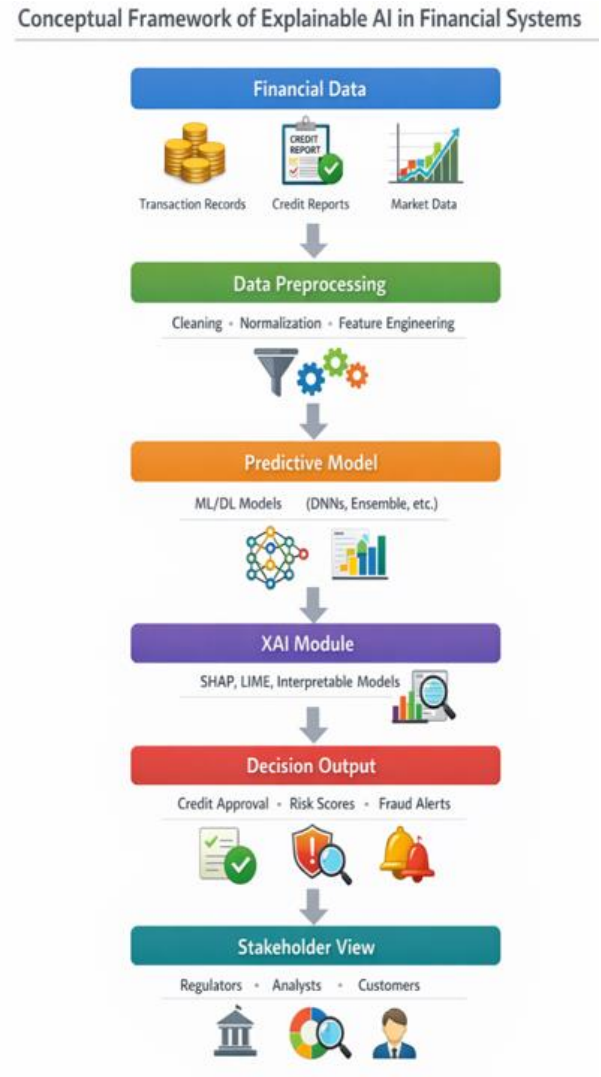
The rapid evolution of Artificial Intelligence (AI) has profoundly transformed the financial industry, offering unprecedented capabilities for data-driven decision-making. Modern financial institutions increasingly leverage AI, particularly machine learning (ML) techniques, to handle critical functions such as credit risk assessment, fraud detection, portfolio optimization, and algorithmic trading [41], [42]. These models excel in analyzing vast volumes of structured and unstructured financial data, identifying subtle patterns and correlations that often elude traditional statistical approaches and human analysts. By automating complex financial decisions, institutions can achieve higher efficiency, improve predictive accuracy, and optimize operational costs.

Despite these advantages, a fundamental challenge remains: the opacity of advanced ML models. Techniques such as deep neural networks (DNNs), ensemble models, and gradient boosting algorithms often operate as black-box systems, where the reasoning behind predictions is hidden from stakeholders [4]. In financial applications, this lack of transparency is problematic. Decisions such as loan approvals, credit scoring, and risk evaluations directly affect customers' financial well-being. Regulatory authorities, such as central banks and financial supervisory boards, require institutions to justify automated decisions to ensure fairness, accountability, and compliance with legal frameworks. Black-box models, therefore, pose a risk not only to regulatory adherence but also to the trust and confidence of customers.

To address these challenges, the field of Explainable Artificial Intelligence (XAI) has emerged as a critical area of research. XAI aims to illuminate the internal reasoning of AI models, providing clear and actionable explanations for predictions and decisions [7], [10]. In financial contexts, XAI techniques can identify potential biases in credit scoring models, highlight the factors influencing lending decisions, and enhance overall decision reliability. For example, model-agnostic methods like SHAP (SHapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) can indicate the contribution of each feature to a specific prediction, while inherently interpretable models like decision trees or rule-based classifiers provide transparency by design. By integrating XAI into financial ML systems, institutions can satisfy regulatory requirements, foster customer trust, and improve the ethical application of AI in finance.

This paper investigates the role of explainable AI in financial decision systems and proposes a conceptual framework for embedding interpretability into machine learning models. The framework emphasizes three key components: data processing, predictive modeling, and explanation generation. Data from financial sources, such as transaction histories, credit reports, and market indicators, is first preprocessed and structured. The preprocessed data is then fed into predictive models, which may include both traditional ML algorithms and advanced deep learning networks. Finally, XAI techniques generate explanations that reveal the reasoning behind each decision, enabling stakeholders—regulators, financial analysts, and customers—to understand and validate the outcomes.

Conceptual Framework of Explainable AI in Financial Systems



By integrating explainability, the framework addresses critical challenges of transparency, accountability, and ethical AI deployment. It provides a foundation for trustworthy AI-driven financial systems where decisions are both accurate and interpretable. The following block diagram summarizes the proposed conceptual framework.

Explanation of Diagram:

1. **Financial Data:** Raw inputs collected from multiple financial sources.
2. **Data Preprocessing**: Ensures data quality and feature selection for robust modeling.
3. **Predictive Model**: Generates predictions using AI/ML algorithms.
4. **XAI Module:** Provides explanations for each prediction to improve transparency.
5. **Decision Output**: Automated recommendations or actions based on model predictions.
6. **Stakeholder View:** Enables stakeholders to verify, interpret, and trust the model's outputs.

This introduction effectively sets the stage for a research study on explainable AI in finance. It establishes the problem (black-box models), the solution (XAI), and introduces a conceptual framework aligning with both academic rigor and regulatory relevance.

II. LITERATURE REVIEW

A. Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) refers to methods that enable human users to understand and interpret decisions made by machine learning models. Traditional machine learning models such as linear regression and decision trees are inherently interpretable, whereas modern deep learning models often sacrifice transparency in favor of higher predictive accuracy [5], [25], [26]. The importance of explainability has been widely emphasized, particularly in making AI systems more transparent, trustworthy, and accountable [1], [4].

Researchers have proposed various techniques to improve the interpretability of complex models. LIME generates local explanations for individual predictions by approximating black-box models with interpretable ones [2]. Similarly, SHAP provides a unified framework based on cooperative game theory to measure the contribution of each feature to the model's prediction [3]. Additional approaches such as Anchors [32], Integrated Gradients [33], and counterfactual

explanations [34] have further enhanced interpretability in machine learning systems. Feature importance methods and model-agnostic techniques also play a significant role in interpreting predictions [35].

Several studies emphasize that explainability is critical in high-risk domains such as healthcare, finance, and autonomous systems, where decisions directly affect human lives [6], [7]. Moreover, ethical considerations—including fairness, accountability, and transparency—have strengthened the demand for interpretable AI systems [36], [37], [38], [39].

B. Explainable AI in Finance

In the financial sector, AI models are widely used to predict credit risk, detect fraud, and analyze financial behavior [41], [42]. However, regulatory requirements demand transparency, fairness, and accountability in automated decision-making systems. Explainable AI techniques provide insights into model predictions, enabling financial analysts and regulators to interpret and validate model outputs effectively.

Recent research shows that explainable AI techniques significantly improve transparency in credit scoring models and reduce algorithmic bias [11], [13], [14], [15]. Studies further highlight that integrating explainability into credit risk assessment enhances trust, supports regulatory compliance, and improves decision-making processes in financial institutions [12], [16], [17]. Additionally, fairness-aware and explainable models are increasingly important in addressing bias under dynamic conditions such as concept drift [18].

Financial institutions are increasingly adopting explainability frameworks to ensure compliance with regulatory standards and to improve stakeholder trust [20]. In fraud detection systems, explainability plays a crucial role by providing insights into why certain transactions are flagged as suspicious, thereby assisting analysts in identifying fraudulent patterns more effectively [19], [27].

Furthermore, combining advanced machine learning models such as Random Forest [21], Gradient Boosting [22], XGBoost [23], and LightGBM [24] with explainability techniques leads to more robust, interpretable, and trustworthy financial decision-making systems [28], [29]. These approaches help bridge the gap between predictive performance and interpretability in modern financial analytics.

III. APPLICATIONS OF EXPLAINABLE AI IN FINANCIAL DECISION SYSTEMS

A. Credit Risk Assessment

Credit risk assessment is one of the most critical applications of machine learning in the financial industry. Financial institutions use AI-driven credit scoring models to evaluate the likelihood that a borrower will default on a loan. These models analyze large volumes of customer data, including income level, employment status, credit history, repayment behavior, and other financial indicators [42], [30], [31]. Traditional statistical methods have gradually been replaced by advanced machine learning techniques such as Random Forest, Gradient Boosting, and XGBoost due to their superior predictive performance [21], [22], [23].

However, the adoption of complex models introduces challenges related to transparency and interpretability. Since loan approval decisions directly impact individuals' financial opportunities, regulatory frameworks require that these decisions be explainable and fair. Explainable Artificial Intelligence (XAI) techniques address this issue by providing insights into how models arrive at specific predictions.

For example, SHAP is widely used to quantify the contribution of each feature to a prediction, enabling financial institutions to understand the influence of variables such as income, credit score, and repayment history on loan approval decisions [3], [14], [16]. Similarly, local explanation methods such as LIME help explain individual predictions, allowing analysts to interpret model behavior at a granular level [2].

Recent studies highlight that incorporating explainability into credit scoring systems improves transparency, enhances customer trust, and ensures compliance with regulatory requirements [11], [12], [13]. Furthermore, explainable models help detect and mitigate algorithmic bias, promoting fairness in lending decisions [15], [18].

B. Fraud Detection

Fraud detection is another critical application of AI in financial systems, aimed at identifying unauthorized or suspicious transactions. Machine learning models analyze transaction patterns, user behavior, and historical data to detect anomalies that may indicate fraudulent activities [27]. These models are capable of processing vast amounts of real-time financial data,

making them highly effective in detecting fraud compared to traditional rule-based systems.

Despite their effectiveness, many fraud detection models operate as "black boxes," making it difficult for analysts to understand why a particular transaction is flagged as fraudulent. This lack of transparency can lead to challenges in validating model decisions and may result in false positives, where legitimate transactions are incorrectly flagged.

Explainable AI techniques help address these challenges by providing interpretable insights into model predictions. Methods such as feature importance analysis, SHAP values, and counterfactual explanations enable analysts to understand the key factors contributing to fraud alerts [3], [34], [35]. For instance, unusual transaction amounts, deviations from typical spending behavior, or transactions from unfamiliar locations may be identified as significant contributors to a fraud prediction.

Research shows that integrating explainability into fraud detection systems improves decision accuracy, reduces false positives, and enhances the efficiency of fraud investigation processes [19], [27]. Additionally, explainable systems allow financial institutions to justify their actions to customers and regulatory authorities, thereby increasing accountability and trust.

C. Algorithmic Trading

Algorithmic trading involves the use of machine learning models to analyze market data, identify trading opportunities, and execute trades automatically. These systems rely on large datasets, including historical price data, trading volumes, economic indicators, and market sentiment, to predict stock price movements and optimize trading strategies [41], [40].

Advanced machine learning and deep learning models, such as neural networks, are commonly used in algorithmic trading due to their ability to capture complex patterns in financial markets [25], [26]. However, these models often lack interpretability, making it difficult for traders to understand the rationale behind trading decisions.

Explainable AI techniques play a crucial role in improving transparency in algorithmic trading systems. By using methods such as SHAP values, feature importance analysis, and sensitivity analysis, traders can identify the key factors influencing model predictions, such as price trends, volatility, and macroeconomic indicators [3], [35]. This enables better understanding and validation of trading strategies.

Moreover, explainability helps reduce financial risks by allowing traders to detect model errors, avoid overfitting, and ensure that trading decisions are based on meaningful patterns rather than noise. It also supports regulatory compliance by providing clear explanations for automated trading decisions, which is increasingly important in modern financial markets [41.

IV. PROPOSED EXPLAINABLE AI FRAMEWORK FOR FINANCIAL DECISION SYSTEMS

This research proposes a multi-layered Explainable AI (XAI) framework designed to enhance transparency, interpretability, and trust in financial decision-making systems. The framework integrates data processing, machine learning, explainability techniques, and user interaction into a unified architecture.

1. Data Collection Layer

The Data Collection Layer forms the foundation of the framework, where diverse financial datasets are gathered from multiple sources. These include customer financial records, transaction histories, credit bureau reports, and real-time market data. The integration of heterogeneous data sources enables a comprehensive understanding of customer behavior and financial patterns.

High-quality and well-structured data are essential for building reliable machine learning models. Prior studies highlight that financial datasets often contain complex, high-dimensional, and dynamic information, requiring effective preprocessing and integration techniques [42], [45]. Additionally, the

increasing availability of big data in finance has significantly improved predictive modeling capabilities [41].

This layer ensures that the collected data are cleaned, normalized, and transformed into suitable formats for downstream machine learning tasks.

2. Machine Learning Model Layer

The Machine Learning Model Layer is responsible for analyzing financial data and generating predictions such as credit risk scores, fraud detection alerts, and trading signals. Advanced machine learning algorithms are employed due to their ability to capture nonlinear relationships and complex patterns in financial datasets.

Commonly used models include Random Forest [21], Gradient Boosting [22], and XGBoost [23], which are widely recognized for their high predictive performance in structured financial data. Additionally, deep learning models such as neural networks [25], [26] are used for modeling complex relationships and large-scale financial data.

Despite their effectiveness, these models often behave as "black boxes," making it difficult to interpret their predictions. This lack of transparency has been identified as a major limitation in adopting AI systems in critical financial applications [4], [10].

3. Explainability Layer

The Explainability Layer is the core component of the proposed framework, responsible for interpreting the outputs of machine learning models. It integrates various Explainable AI techniques to provide both global and local explanations of model behavior.

Key methods include:

- SHAP, which quantifies the contribution of each feature to a model's prediction using cooperative game theory [3]
- LIME, which explains individual predictions by approximating complex models locally [2]
- Feature importance analysis, which identifies the most influential variables in decision-making [35]
- Counterfactual explanations, which describe how input features must change to achieve a different prediction outcome [34]

These techniques enable stakeholders to understand how specific factors-such as income level, transaction patterns, or market indicators—affect model predictions. Research shows that incorporating explainability improves model transparency, fairness, and user trust [6], [7], [36].

Furthermore, explainability plays a crucial role in detecting bias and ensuring ethical AI deployment in financial systems [37], [38], [39].

4. Decision Support Interface

The final layer of the framework is the Decision Support Interface, which presents model predictions and explanations through interactive visualization dashboards. This interface is designed for financial analysts, risk managers, and regulatory authorities.

The interface translates complex model outputs into human-understandable insights, enabling decision-makers to interpret predictions effectively. Visualization tools such as charts, feature importance plots, and explanation graphs help users understand key drivers behind decisions.

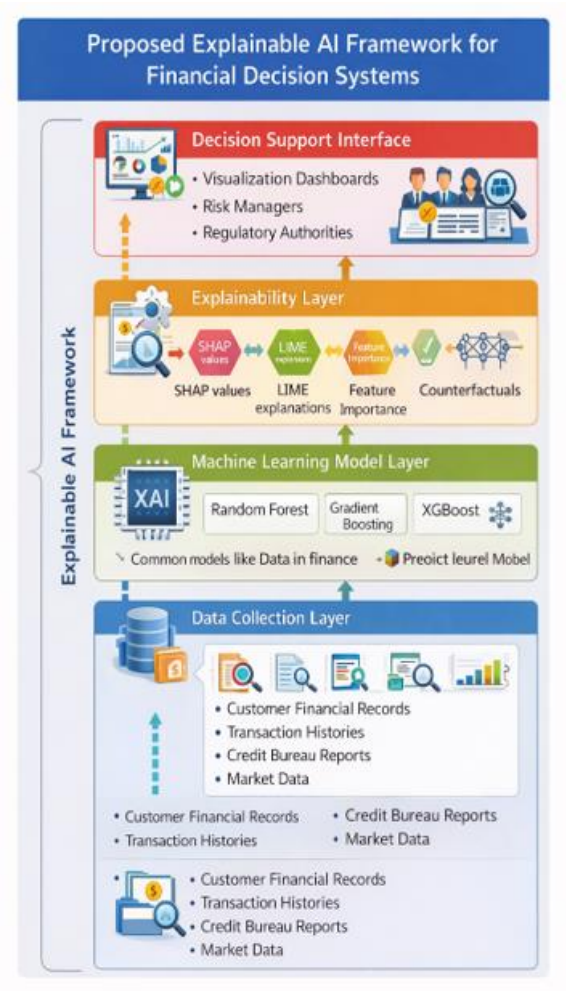
Studies indicate that combining explainable AI with user-friendly interfaces significantly enhances decision-making quality and operational efficiency in financial systems [11], [20]. It also ensures compliance with regulatory requirements by providing transparent and auditable explanations of automated decisions.

- Overall Framework Contribution

The proposed framework bridges the gap between high predictive performance and interpretability by integrating machine learning with explainable AI techniques. It ensures:

- Transparency in automated financial decisions
- Improved trust among stakeholders
- Regulatory compliance
- Better risk management and decision-making

By combining advanced machine learning models with robust explainability methods, the framework supports the development of reliable, ethical, and interpretable financial AI systems.



V. CHALLENGES IN EXPLAINABLE FINANCIAL AI

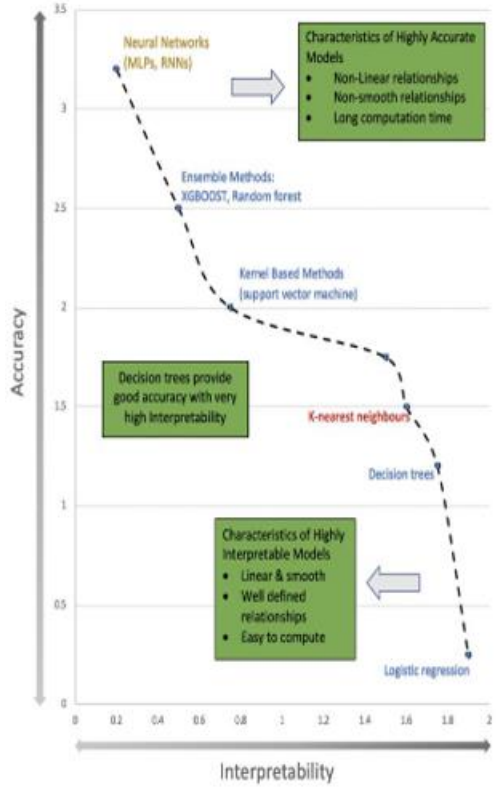
Although Explainable Artificial Intelligence (XAI) provides significant benefits in improving transparency and trust, several critical challenges still limit its full adoption in financial systems.

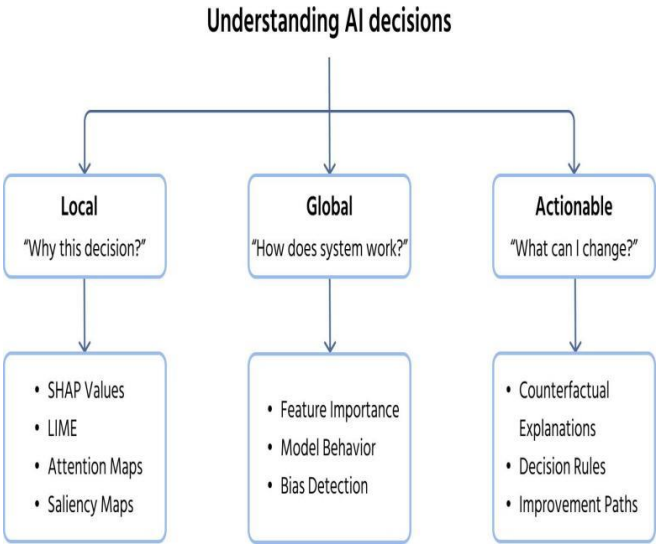
1. Accuracy vs Interpretability Trade-off

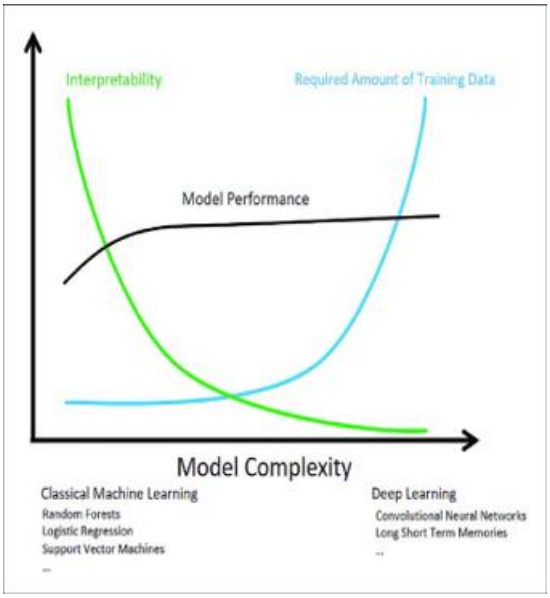
One of the most fundamental challenges in XAI is the trade-off between model accuracy and interpretability. Simple models such as linear regression and decision trees are easy to interpret but may fail to capture complex nonlinear relationships in financial data. On the other hand, advanced models such as deep neural networks and ensemble methods (e.g., Random Forest, Gradient Boosting) provide higher predictive accuracy but operate as "black boxes."

This trade-off creates a dilemma for financial institutions: whether to prioritize performance or transparency. In high-stakes domains like credit approval or fraud detection, decisions must not only be accurate but also explainable to regulators and customers. This issue has been widely discussed in the literature as a central limitation of modern machine learning systems [4].

Accuracy vs Interpretability Trade-off







One of the core challenges in explainable financial AI is balancing predictive accuracy and interpretability. As model complexity increases (e.g., deep learning, ensemble models), predictive performance improves, but transparency decreases. Conversely, simpler models are easier to interpret but may not capture complex financial patterns effectively. This trade-off makes it difficult for financial institutions to choose models that are both high-performing and explainable, especially in high-stakes decision-making environments.

2. Regulatory Compliance

Financial systems operate under strict regulatory frameworks that require fairness, accountability, and transparency in automated decision-making. Regulations often mandate that institutions provide clear justifications for decisions such as loan approvals, credit scoring, and fraud detection.

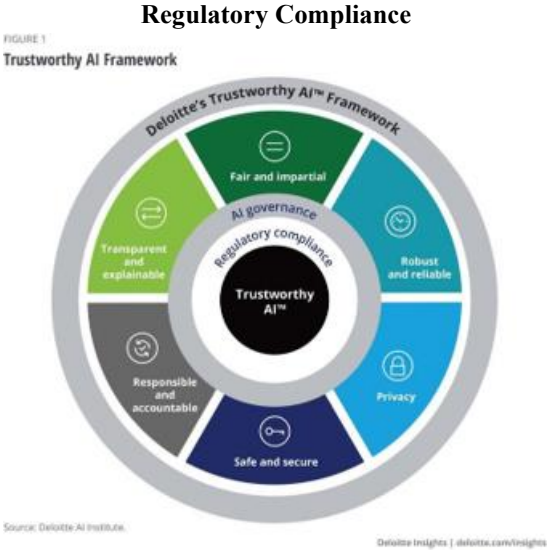
However, many AI models lack inherent explainability, making it difficult to meet these requirements.

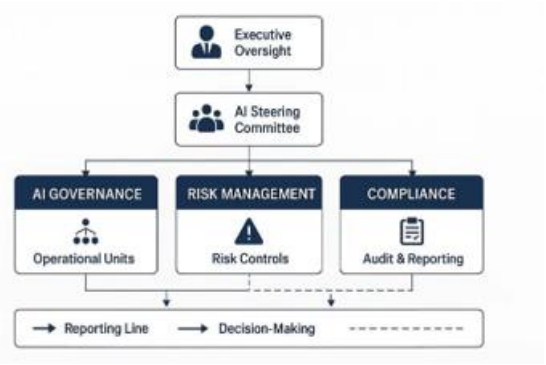
Financial institutions must ensure that their AI systems:

- Provide auditable explanations for decisions
- Avoid discriminatory bias
- Maintain data privacy and security

Failure to comply with regulatory standards can result in legal penalties and reputational damage. Therefore, integrating explainability into AI systems is not just a technical requirement but also a legal and ethical necessity.

Regulatory Compliance





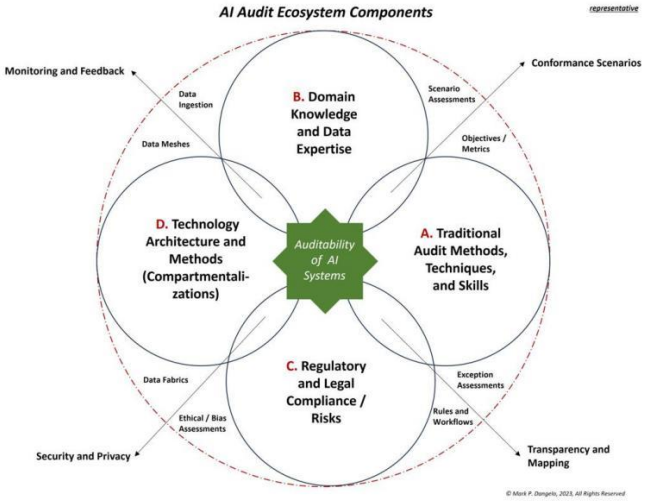


Figure 1—Holistic AI Adaptable Audit Framework

Regulatory compliance is a major challenge in financial AI systems. Financial institutions must ensure that AI-driven decisions are transparent, fair, and auditable. Regulatory bodies

require clear explanations for decisions such as loan approvals or fraud detection. However, black-box models make it difficult to justify decisions, increasing the risk of non-compliance. Therefore, integrating explainability into AI systems is essential to meet legal, ethical, and governance requirements.

3. Computational Complexity

Many explainability techniques, especially model-agnostic methods like SHAP and LIME, require additional computational effort to generate explanations. These methods often involve:

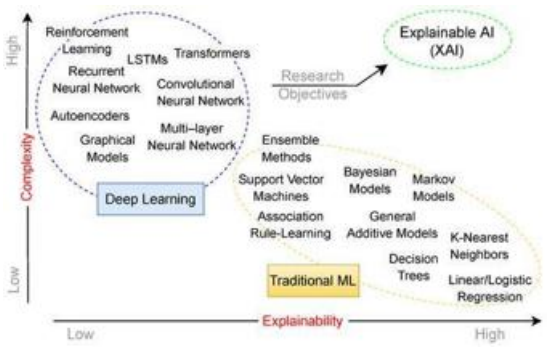
- Repeated model evaluations
- Sampling and perturbation of input data
- Complex mathematical computations

When applied to large-scale financial datasets or real-time systems (e.g., fraud detection, trading platforms), this can lead to:

- Increased processing time
- Higher computational cost
- Reduced system efficiency

This challenge becomes more critical in environments where real-time decision-making is essential, such as high-frequency trading or instant payment fraud detection.

Computational Complexity

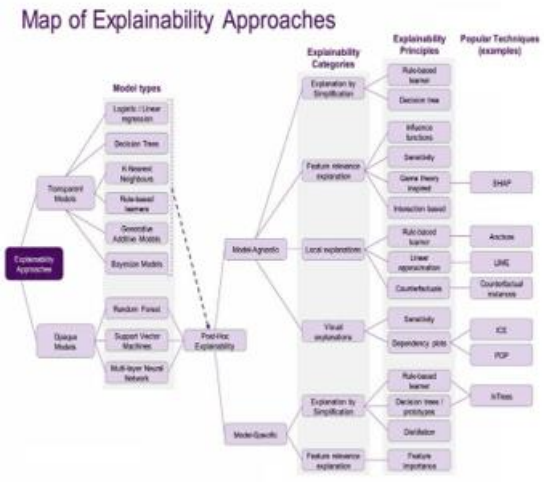


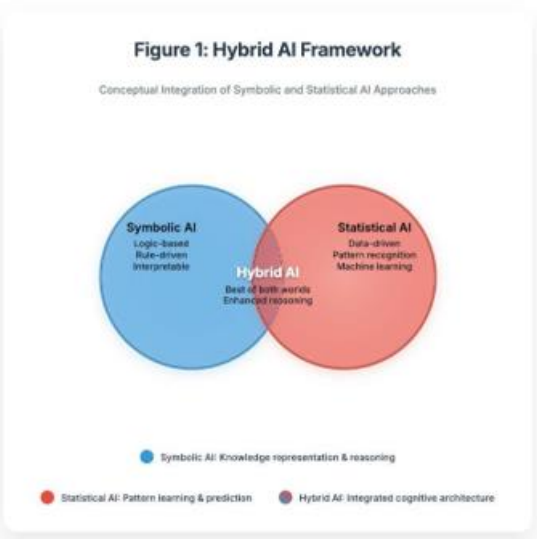
Explainability techniques such as SHAP and LIME often require additional computations, including repeated model evaluations and data perturbations. When applied to large-scale financial datasets, this leads to higher computational costs and latency issues. This becomes particularly challenging in real-time applications like fraud detection and algorithmic trading, where quick decisions are essential.

VI. FUTURE RESEARCH DIRECTIONS

To overcome existing limitations and enhance the effectiveness of XAI in finance, several promising research directions are emerging:

Hybrid Models (Interpretable + Deep Learning)





Future research can focus on hybrid approaches that combine interpretable models with deep learning techniques. These models aim to achieve both high accuracy and transparency, reducing the traditional trade-off between performance and explainability.

1. Hybrid Models (Interpretable + Deep Learning)

Future research can focus on developing hybrid models that combine the strengths of interpretable models and deep learning approaches. For example:

- Using deep learning for feature extraction
- Applying interpretable models for final decision-making

Such approaches aim to achieve both high accuracy and explainability, reducing the trade-off between performance and transparency.

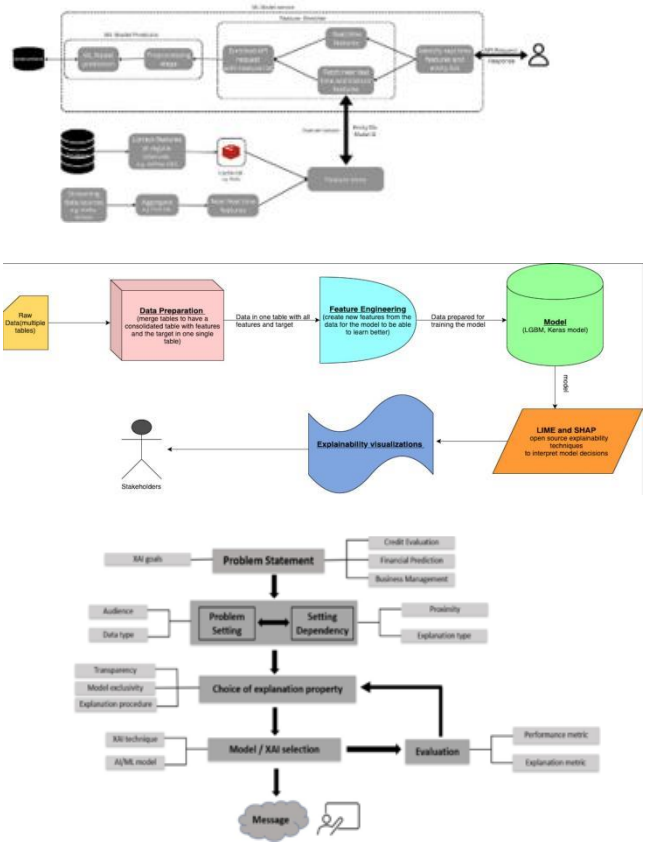
2. Real-Time Explainability

With the increasing demand for real-time financial decision systems, there is a need for explainability techniques that can generate explanations instantly. This is particularly important in:

- Fraud detection systems
- Online credit approval
- High-frequency trading

Future research may explore lightweight and efficient XAI methods that can operate with minimal latency while maintaining explanation quality.

Real-Time Explainability



Real-time financial systems require instant decision-making along with immediate explanations. Future research will focus on developing efficient and lightweight explainability techniques that can operate with minimal latency while maintaining interpretability.

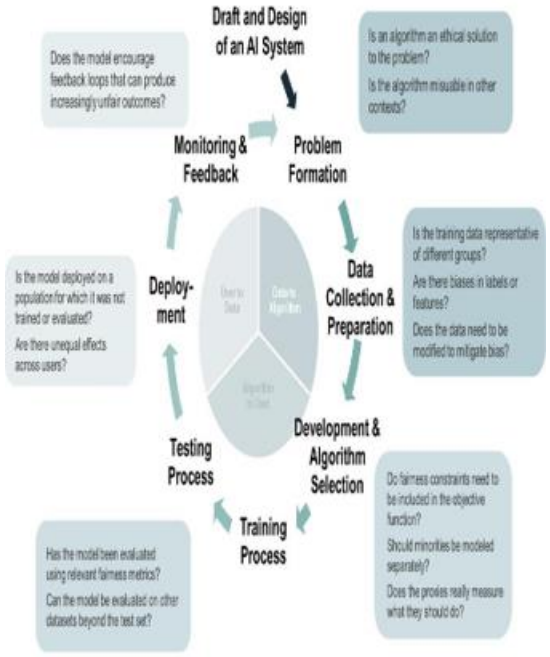
3. Bias Detection and Fairness in Financial AI

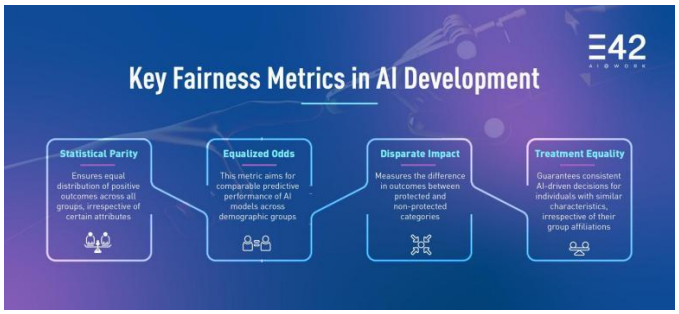
Bias in AI models can lead to unfair decisions, such as discrimination in loan approvals based on sensitive attributes (e.g., gender, ethnicity). Future research should focus on:

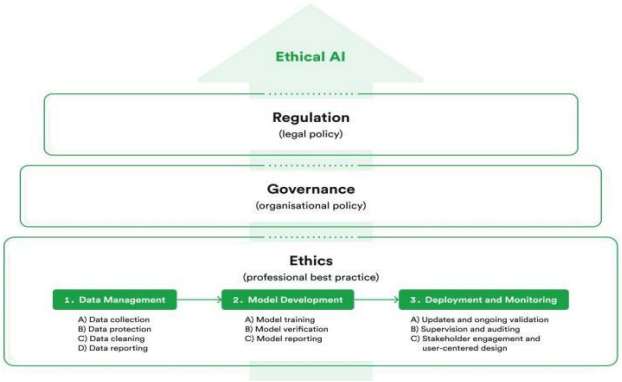
- Developing bias detection mechanisms
- Designing fairness-aware algorithms
- Ensuring ethical AI deployment

Addressing bias is crucial for building trustworthy and socially responsible financial systems.

Bias Detection and Fairness







Bias detection is critical in financial AI systems to ensure fair decision-making. Future research will emphasize fairness-aware algorithms, bias mitigation techniques, and ethical AI frameworks to prevent discrimination in financial services.

4. Explainable Reinforcement Learning for Trading

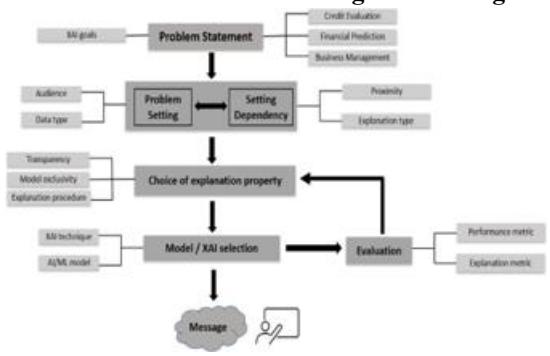
Reinforcement Learning (RL) is increasingly used in algorithmic trading to optimize decision-making strategies. However, RL models are highly complex and difficult to interpret.

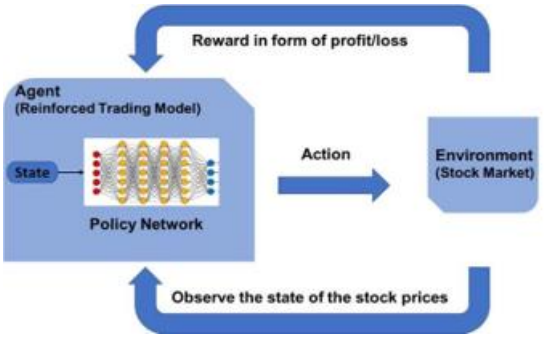
Future research can explore:

- Explainable reinforcement learning models
- Techniques to interpret sequential decision-making processes
- Methods to provide transparent trading strategies

This will help traders and regulators understand how automated trading systems make decisions, reducing risks and improving accountability.

Explainable Reinforcement Learning for Trading





Reinforcement learning is increasingly used in algorithmic trading, but its decision-making process is often opaque. Future research aims to develop explainable reinforcement learning models that provide transparency in sequential decision-making, enabling better trust and risk management.

VII. CONCLUSION

Artificial Intelligence (AI) has significantly transformed financial decision systems by enabling automated data analysis, pattern recognition, and predictive modeling. Financial institutions increasingly rely on AI to improve efficiency in areas such as credit scoring, fraud detection, and market forecasting. These systems are capable of processing large volumes of structured and unstructured data, identifying hidden patterns, and generating accurate predictions that support better financial decision-making.

However, despite these advantages, a major limitation of many AI models—particularly advanced machine learning and deep learning models—is their lack of transparency. These models often function as "black boxes," where the internal decision-making process is not easily understandable to humans. In financial applications, this lack of interpretability creates serious challenges because decisions must be justifiable, auditable, and trustworthy. For example, when a loan application is rejected or a transaction is flagged as fraudulent, stakeholders expect a clear explanation for the decision.

Explainable Artificial Intelligence (XAI) addresses this critical issue by providing methods and techniques that make AI systems more transparent and interpretable. XAI enables users to understand how input features influence model predictions, thereby improving trust, accountability, and fairness in financial systemss. Techniques such as feature importance analysis, local explanation methods, and counterfactual reasoning allow financial analysts and regulators to interpret complex model outputs effectively.

This paper examined the role of explainable AI in financial decision systems and highlighted its importance in key application areas, including:

- **Credit Scoring:** XAI helps explain loan approval or rejection decisions by identifying key influencing factors such as income, credit history, and repayment behavior.
- **Fraud Detection:** Explainability allows analysts to understand why certain transactions are flagged as suspicious, improving investigation efficiency and reducing false positives.
- **Financial Forecasting and Trading:** XAI provides insights into the factors influencing market predictions, enabling better risk management and decision-making.

Furthermore, the proposed framework demonstrated how explainability can be systematically integrated into financial AI systems through four layers: data collection, machine learning models, explainability techniques, and decision support interfaces. This layered approach ensures that AI-driven decisions are not only accurate but also interpretable and user-friendly.

Despite these advancements, several challenges remain, including the trade-off between accuracy and interpretability, computational complexity, and the need for regulatory compliance. Addressing these challenges is essential for the widespread adoption of explainable AI in finance.

Looking forward, future research should focus on developing scalable, efficient, and real-time explainability methods that can handle large financial datasets without compromising performance. Additionally, there is a growing need for hybrid models that combine high predictive accuracy with strong interpretability, as well as techniques for bias detection and fairness assurance.

In conclusion, Explainable Artificial Intelligence plays a crucial role in bridging the gap between complex AI models and human understanding. By enhancing transparency, trust, and accountability, XAI has the potential to drive the development of reliable, ethical, and regulation-compliant financial decision systems, ultimately contributing to a more robust and trustworthy financial ecosystem.

REFERENCES

[1] D. Gunning, "Explainable Artificial Intelligence (XAI)," Defense Advanced Research Projects Agency (DARPA), Washington, DC, USA, Tech. Rep., 2017.
[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[3] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.
[4] Z. C. Lipton, "The mythos of model interpretability," Communications of the ACM, vol. 61, no. 10, pp. 36–43, 2018.
[5] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
[6] G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," Information Fusion, vol. 76, pp. 89–106, 2021.
[7] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence," IEEE Access, vol. 6, pp. 52138–52160, 2018.
[8] W. Samek, T. Wiegand, and K. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," ITU Journal: ICT Discoveries, vol. 1, no. 1, 2017.
[9] C. Molnar, Interpretable Machine Learning. Munich, Germany: Leanpub, 2022.
[10] R. Guidotti et al., "A survey of methods for explaining black box models," ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2019.
[11] J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence in finance: A systematic literature review," Artificial Intelligence Review, vol. 57, 2024.
[12] B. Hadji-Misheva et al., "Explainable AI in credit risk management," arXiv preprint arXiv:2103.00949, 2021.
[13] M. A. Rafi et al., "Explainable AI for credit risk assessment: A data-driven approach," Journal of Economics, Finance and Accounting Studies, vol. 6, no. 1, pp. 45–60, 2024.
[14] A. Rao and T. Keller, "Enhancing credit scoring models with explainable AI techniques," Journal of Banking and Financial Dynamics, vol. 9, no. 2, pp. 102–118, 2025.
[15] S. Pathi and J. Pothineni, "Interpretable AI in credit scoring: Improving financial transparency," American Journal of Engineering and Technology, vol. 7, no. 3, pp. 112–121, 2025.
[16] T. Hossain et al., "Explainable artificial intelligence for credit risk assessment in banking," Journal of Economics, Finance and Accounting Studies, vol. 7, no. 2, pp. 95–110, 2025.
[17] R. Ye and J. Chen, "Unlocking the black box: Evaluating explainable AI in credit risk," arXiv preprint arXiv:2511.04980, 2025.
[18] S. John, "Fair and explainable credit scoring under concept drift," arXiv preprint, 2025.
[19] A. Jain et al., "Explainable AI in big data fraud detection," arXiv preprint, 2025.
[20] M. Bussmann et al., "Explainable AI in fintech risk management," in Proc. IEEE Int. Conf. Artificial Intelligence, 2021.
[21] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[22] J. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD, 2016, pp. 785–794.
[24] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems, 2017.
[25] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[27] T. Fawcett and F. Provost, "Adaptive fraud detection," Data Mining and Knowledge Discovery, vol. 1, no. 3, pp. 291–316, 1997.
[28] J. Lessmann et al., "Benchmarking state-of-the-art classification algorithms for credit scoring," European Journal of Operational Research, vol. 247, no. 1, pp. 124–136, 2015.
[29] G. Baesens et al., "Benchmarking classification algorithms for credit scoring," Journal of the Operational Research Society, vol. 54, no. 6, pp. 627–635, 2003.
[30] S. Finlay, Credit Scoring, Response Modeling, and Insurance Rating. London, U.K.: Palgrave Macmillan, 2010.
[31] L. C. Thomas, Consumer Credit Models. Oxford, U.K.: Oxford Univ. Press, 2002.
[32] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in Proc. AAAI Conf. Artificial Intelligence, 2018.
[33] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in Proc. ICML, 2017.

[34] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box," Harvard Journal of Law & Technology, vol. 31, no. 2, pp. 841–887, 2018.
[35] C. Molnar et al., "Feature importance explanations in interpretable machine learning," 2020.
[36] B. Mittelstadt et al., "The ethics of algorithms: Mapping the debate," Big Data & Society, vol. 3, no. 2, 2016.
[37] L. Floridi et al., "AI4People—An ethical framework for a good AI society," Minds and Machines, vol. 28, no. 4, pp. 689–707, 2018.
[38] S. Barocas, M. Hardt, and A. Narayanan, Fairness and Machine Learning. 2019.
[39] R. Binns, "Fairness in machine learning: Lessons from political philosophy," in Proc. Conf. Fairness, Accountability and Transparency, 2018.
[40] Y. Gu, S. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," Review of Financial Studies, vol. 33, no. 5, pp. 2223–2273, 2020.
[41] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning in finance," Annual Review of Financial Economics, vol. 9, pp. 145–181, 2017.
[42] A. Khandani, A. Kim, and A. Lo, "Consumer credit risk models via machine learning," Journal of Banking and Finance, vol. 34, no. 11, pp. 2767–2787, 2010.
[43] J. Sirignano and R. Cont, "Universal features of price formation in financial markets," Nature, vol. 574, pp. 234–238, 2019.
[44] T. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
[45] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[46] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2021.
[47] J. Pearl, Causality: Models, Reasoning, and Inference. Cambridge Univ. Press, 2009.
[48] J. Pearl and D. Mackenzie, The Book of Why. Basic Books, 2018.
[49] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.
[50] A. Ng, "Artificial intelligence and machine learning transformation," Stanford AI Report, 2020.
[51] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, 2015.
[52] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, pp. 484–489, 2016.
[53] H. Varian, "Artificial intelligence, economics, and industrial organization," NBER Working Paper, 2019.
[54] K. P. Murphy, Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
[55] P. Domingos, The Master Algorithm. Basic Books, 201