Escalate: Economics and Business Journal
https://journal.takaza.id/index.php/escalate
Vol. 3, No. 02, 2025, pp. 70-86
E-ISSN: 3025-4213
E-mail: escalate@takaza.id

Generative AI and the Future of Cash Flow Forecasting: Bridging Accuracy and Governance in Public Companies

Zikri Wahyudi [1*], Deepika Chaplot [2]
[1] PT. Samudera Literasi Bangsa
[2] Pacific Academy of Higher Education and Research University
Corresponding Author e-mail: zikriwahyudi13@gmail.com

Article History:
Received: 29-07-2025
Revised: 14-10-2025
Accepted: 30-10-2025

Keywords: *Cash Flow Forecasting; Forecast Accuracy; Generative Artificial Intelligence; Large Language Models; Financial Reporting; Machine Learning*

Abstract: *Accurate cash flow forecasting is critical for corporate decision-making, influencing investment strategies, dividend policies, and debt covenant compliance. Traditional statistical models such as ARIMA and accrual-based regressions struggle to capture nonlinear financial dynamics. At the same time, machine learning approaches, including LSTM and GRU, offer improved accuracy but face limitations in interpretability and governance. Against this backdrop, generative artificial intelligence (GenAI) has emerged as a transformative technology, capable of integrating structured and unstructured financial data, generating synthetic time series, and outperforming analysts in earnings prediction. This study conducts a systematic literature review of 55 peer-reviewed studies published between 2018 and 2025, employing PRISMA guidelines and bibliometric analysis. The findings reveal three significant insights: first, empirical research on GenAI in cash flow forecasting remains scarce compared to earnings and liquidity studies; second, evidence suggests significant potential for GenAI to enhance forecast accuracy and scenario-based analysis; and third, regulatory frameworks such as the EU AI Act and ESMA guidance emphasize the need for transparency, accountability, and auditability in AI-driven financial forecasting. The contribution of this study lies in synthesizing the current state of research, identifying a critical gap in the application of GenAI to cash flow forecasting in public companies, and proposing a research agenda that integrates methodological innovation with governance imperatives.*

Introduction

Cash flow forecasting is a central pillar of financial planning and corporate decision-making, especially in publicly listed firms where managers, investors, and regulators closely scrutinize liquidity positions and future cash availability. Accurate projections of operating cash flows inform key corporate policies, including investment planning, dividend distribution, debt covenant compliance, and working capital management (IAS 7, 2022). Despite its importance, extant evidence shows that cash flow forecasts are often biased and less accurate than earnings forecasts, with mean absolute percentage error (MAPE) frequently exceeding 20% in empirical studies (Givoly et al., 2009; Bilinski, 2014). This persistent forecasting inaccuracy undermines managerial credibility, increases financing costs, and reduces investor confidence. Chief Financial Officers (CFOs) consistently identify cash flow forecasting as one of the most challenging tasks in corporate finance, citing it as a critical "pain point" in recent surveys and industry reports (Deloitte, 2023; EY, 2024; Citizens Bank, 2025).

The shortcomings of cash flow forecasting are not new, but they have become more pronounced in an era characterised by heightened market volatility, rapid technological disruption, and increased regulatory expectations. Traditional forecasting approaches, including regression models grounded in accrual accounting and time-series methods such as autoregressive integrated moving average (ARIMA), have been widely applied in both academic research and practice. However, their underlying linearity assumptions limit their ability to capture complex patterns, nonlinear dependencies, and the high dimensionality inherent in financial data (Dechow et al., 1998; Barth et al., 2001). More recently, machine learning (ML) techniques such as recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent units (GRU) have been adopted to improve forecasting accuracy. These models exhibit superior performance relative to traditional methods in specific contexts, particularly in capturing temporal dependencies and nonlinearities in financial time series (Fischer & Krauss, 2018). Nonetheless, while these approaches represent a step forward, they remain predominantly predictive rather than generative, lacking the ability to simulate scenarios, incorporate textual information from corporate disclosures, or generate probabilistic distributions of future outcomes.

The emergence of generative artificial intelligence (GenAI) could mark a paradigm shift in forecasting methodology. Generative models, including large language models (LLMs) such as GPT-4 and domain-specific systems like BloombergGPT, as well as generative adversarial networks (GANs) and diffusion models, are not only capable of analyzing structured numerical data but also of synthesizing unstructured textual and contextual inputs (Wu et al., 2023; Zhang et al., 2023). In financial applications, LLMs have demonstrated remarkable capabilities in interpreting corporate disclosures, analyzing financial statements, and even outperforming professional analysts in predicting the direction of earnings changes (Huang et al., 2024). GAN-based approaches, meanwhile, have shown promise in generating synthetic financial time series that preserve essential statistical properties, thereby enabling scenario-based stress testing and probabilistic cash flow forecasting (Goodfellow et al., 2020; Ni et al., 2023). These advancements suggest that GenAI could substantially improve the accuracy and robustness of cash flow forecasting in public companies, particularly when integrated with conventional models and domain knowledge.

Despite the promise of GenAI, the academic literature reveals a conspicuous gap. Existing empirical research has mainly focused on earnings prediction, liquidity forecasting, and market volatility estimation using generative models (Han et al., 2023; Ma & Kim, 2024). Direct applications of GenAI to cash flow forecasting in public companies remain scarce. Most available studies focus on traditional ML or hybrid statistical–machine learning approaches in predicting liquidity needs (Krauss et al., 2017; Hashim et al., 2019), with limited exploration of generative techniques. The absence of empirical validation in this domain raises critical questions regarding the potential, limitations, and governance implications of adopting GenAI for cash flow forecasting.

Regulatory and audit considerations underscore the urgency of this research agenda. Regulatory bodies such as the European Securities and Markets Authority (ESMA) have clarified that corporate managers remain ultimately responsible for decisions informed by AI, regardless of the level of automation employed (ESMA, 2024). Similarly, the European Union's Artificial Intelligence Act classifies financial AI systems as high-risk, requiring transparency, explainability, and robust human oversight (European Commission, 2024). The Centre for Audit Quality (CAQ) has also emphasized that the growing use of AI in financial reporting and forecasting demands enhanced audit procedures, documentation, and assurance frameworks (CAQ, 2024). Therefore, evaluating the impact of GenAI on forecasting accuracy cannot be disentangled from discussions of governance, accountability, and auditability.

To contextualize the phenomenon and highlight the problem, Table 1 presents a comparative summary of cash flow forecasting accuracy across approaches, illustrating the relative performance of traditional, ML-based, and emerging GenAI models. While GenAI studies remain preliminary, early evidence from adjacent financial tasks suggests considerable potential for accuracy gains.

Table 1. Comparative Overview of Forecasting Approaches and Accuracy

Approach	Core Methods	Typical Data Inputs	Limitations	Reported Accuracy (MAPE/MASE)
Traditional Statistics	Regression, ARIMA	Accruals, past CF, financial ratios	Linear assumptions, limited features	MAPE often >20% (Givoly et al., 2009)
Machine Learning (ML)	LSTM, GRU, Random Forest	Time-series CF, accruals, macro data	Black-box, limited generative capacity	10–15% improvement over ARIMA (Fischer & Krauss, 2018)
Generative AI (GenAI)	LLMs, GANs, Diffusion	Structured + unstructured (e.g., FS, MD&A)	Early stage, governance concerns	Early evidence: outperform analysts in earnings prediction (Huang et al., 2024)

This table underscores three critical observations: (i) traditional methods remain prone to significant bias; (ii) ML-based methods improve accuracy but are not transformative; and (iii) GenAI holds substantial potential, though empirical applications in cash flow forecasting are still nascent.

Given this background, the present study seeks to address the following research questions: (1) What does the existing literature reveal about the accuracy of cash flow forecasting using traditional and machine learning methods? (2) What evidence supports the capacity of generative AI to enhance forecasting accuracy in public companies? (3) What conceptual framework and future research agenda can guide the integration of GenAI into corporate forecasting practice? By systematically synthesizing the current literature and identifying key research gaps, this study aims to provide a comprehensive assessment of the potential impact of GenAI on cash flow forecasting accuracy.

The contributions of this paper are threefold. First, from an academic perspective, it advances the literature by synthesizing fragmented findings and proposing a conceptual framework using the Theory–Context–Characteristics–Methodology (TCCM) approach to integrate GenAI into forecasting research. Second, from a practical perspective, it offers recommendations for CFOs, investors, and financial managers on adopting and integrating GenAI into forecasting workflows, highlighting both opportunities and risks. Third, from a policy perspective, it discusses the implications for governance and oversight, particularly in light of evolving regulatory frameworks such as the EU AI Act and ESMA guidelines, as well as audit practices highlighted by CAQ. By doing so, this article positions itself at the intersection of finance, artificial intelligence, and corporate governance, providing insights that are academically rigorous, practically relevant, and policy-aware.

Literature Review

Cash flow forecasting occupies a distinctive role in corporate finance research and practice, bridging the domains of accounting measurement, capital market decision-making, and financial risk management. This section develops the theoretical underpinnings of the study by reviewing (i) the role of cash flow forecasting and the standards that guide its measurement, (ii) the concept of forecasting accuracy and the evaluation metrics commonly employed, (iii) the application of artificial intelligence to financial forecasting, and (iv) the emerging contribution of generative artificial intelligence (GenAI) as a potential paradigm shift. Together, these strands provide the foundation for examining how GenAI may influence the accuracy of cash flow forecasting in publicly listed companies.

The foundation of cash flow forecasting rests on the International Accounting Standard 7 (Statement of Cash Flows), which mandates the classification of cash flows into operating, investing, and financing activities (IAS 7, 2022). Operating cash flow (OCF), in particular, is central to assessing a firm's liquidity and solvency, and it is frequently employed in empirical research as the target variable for forecasting (Barth et al., 2001). Previous studies demonstrate that accrual-based earnings contain helpful information for predicting future cash flows but are susceptible to estimation errors, discretionary accruals, and managerial manipulation (Dechow et al., 1998). Consequently, cash flow forecasts derived directly from OCF are considered more reliable indicators of a firm's financial flexibility. Analysts' cash flow forecasts, however, are found to be systematically less accurate than their earnings forecasts, highlighting a persistent challenge in both research and practice (Givoly et al., 2009; Bilinski, 2014). This discrepancy underscores the necessity of methodological advances capable of addressing the unique complexities of cash flow dynamics.

Forecasting accuracy serves as the primary benchmark for evaluating different methodologies. Traditional accuracy metrics such as mean absolute percentage error (MAPE) and root mean squared error (RMSE) have been widely used but are not without limitations. Hyndman and Koehler (2006) argue that MAPE is problematic when actual values are near zero and that RMSE disproportionately penalizes significant errors, thereby complicating cross-study comparability. They propose the mean absolute scaled error (MASE) as a more robust alternative, allowing comparison across series of different scales. The importance of choosing appropriate metrics was further demonstrated in the Makridakis Competitions (M4), where hybrid models combining statistical and machine learning approaches consistently outperformed single models, and the use of standardized error measures, such as symmetric MAPE (sMAPE), enabled meaningful cross-method evaluation (Makridakis et al., 2020). For corporate finance research, adopting robust metrics is essential to credibly assess whether new technologies such as GenAI materially improve forecasting performance.

The adoption of artificial intelligence (AI) in forecasting has evolved substantially over the past decade. Early applications in finance primarily employed support vector machines, random forests, and gradient boosting, focusing on pattern recognition in structured numerical data (Krauss et al., 2017). More advanced deep learning architectures, particularly recurrent neural networks (RNNs) and their variants such as long short-term memory (LSTM) and gated recurrent units (GRU), have been increasingly applied to financial time series due to their capacity to capture temporal dependencies and nonlinear relationships (Fischer & Krauss, 2018). These models have demonstrated notable improvements over autoregressive integrated moving average (ARIMA) models, particularly in short- and medium-horizon forecasting. Nonetheless, they are primarily predictive models that extrapolate from historical patterns and cannot simulate new scenarios or integrate diverse financial information modalities, such as textual disclosures, macroeconomic shocks, or behavioral signals.

The limitations of conventional AI approaches create space for GenAI to make a distinctive contribution. Unlike discriminative models, which focus on prediction by learning conditional probabilities, generative models aim to model the underlying distribution of the data, thereby enabling the generation of new, plausible data points (Goodfellow et al., 2020). This property has important implications for financial forecasting, where scenario analysis and probabilistic reasoning are critical. For example, generative adversarial networks (GANs) have been applied to produce synthetic financial time series that preserve key statistical characteristics, allowing researchers to conduct stress testing and scenario-based forecasting (Han et al., 2023). Similarly, diffusion models have shown promise in creating more stable generative outputs for sequential data, broadening the potential for probabilistic cash flow forecasting (Ni et al., 2023).

Large language models (LLMs) represent another category of GenAI with direct relevance to corporate finance. Recent studies show that models such as GPT-4 and BloombergGPT can analyze financial statements, interpret managerial disclosures, and predict earnings changes with accuracy surpassing that of professional analysts (Huang et al., 2024; Wu et al., 2023). The ability of LLMs to integrate structured and unstructured data sources provides a unique advantage in cash flow forecasting, as corporate disclosures often include forward-looking narratives that are difficult to quantify yet contain material information.

Integrating such qualitative signals with quantitative cash flow data could yield forecasts that are not only more accurate but also more contextually informed.

While these technological advancements hold substantial promise, their adoption in the domain of cash flow forecasting in publicly listed companies remains limited. Most extant studies have focused on earnings forecasts, liquidity management, or risk prediction (Ma & Kim, 2024). The scarcity of research explicitly addressing cash flow forecasting through GenAI constitutes a critical gap in the literature. Furthermore, the lack of standardized practices for integrating GenAI into corporate finance raises concerns about interpretability, governance, and auditability. Regulatory perspectives emphasize these challenges: the European Securities and Markets Authority (ESMA) stresses managerial accountability when AI systems are employed, the European Union's Artificial Intelligence Act mandates transparency and risk management for high-risk AI applications in finance, and audit organizations highlight the necessity of maintaining sufficient documentation and oversight when AI informs financial reporting (ESMA, 2024; European Commission, 2024; CAQ, 2024).

In sum, the theoretical background highlights three key points. First, cash flow forecasting is critical but historically less accurate than earnings forecasting, reflecting its inherent complexity and the inadequacies of traditional linear methods. Second, while AI and ML have improved forecasting accuracy, their predictive nature limits their ability to generate scenarios or synthesize multimodal inputs. Third, GenAI represents a promising yet underexplored approach that potentialy combines the strengths of generative modeling and language-based reasoning to overcome these limitations. The challenge for researchers and practitioners lies not only in validating the accuracy improvements that GenAI might provide but also in ensuring compliance with evolving governance standards. This dual imperative of accuracy and accountability serves as the basis for the subsequent methodological design and the synthesis of the literature in this study.

Research Methods

This study employs a Systematic Literature Review (SLR) approach to synthesize and evaluate existing research on the impact of generative artificial intelligence (GenAI) on cash flow forecasting accuracy in public companies. The use of SLR is motivated by the need to ensure transparency, reproducibility, and rigor in the collection, screening, and analysis of the literature (Moher et al., 2009). To strengthen methodological robustness, the review follows the PRISMA 2020 guidelines, which provide a structured process for identifying, screening, and including relevant studies (Page et al., 2021). In addition, bibliometric techniques are integrated to map the intellectual structure of the field and identify thematic clusters.

Research Design

The research design combines qualitative content analysis with quantitative bibliometric mapping. Content analysis enables the extraction of key findings, contexts, methods, and theoretical contributions for each study. Bibliometric analysis, using tools such as VOSviewer and the R package bibliometrix, helps to identify co-occurrence patterns, keyword networks, and the evolution of research themes (Donthu et al., 2021). This dual approach ensures a comprehensive synthesis that is both systematic and critical, aligning with the standards of high-impact journals.

Data Sources and Search Strategy

The literature search was conducted across multiple academic databases to maximize coverage and minimize selection bias. The primary databases include Scopus, Web of Science, IEEE Xplore, SSRN, and arXiv, complemented by publisher repositories such as Elsevier ScienceDirect, Wiley Online Library, and Taylor & Francis. These databases were selected because they index peer-reviewed journals and conference proceedings relevant to accounting, finance, artificial intelligence, and forecasting.

The search strategy employed Boolean operators and keywords derived from the research questions. Example queries include:

- ("generative AI" OR "large language model" OR LLM OR "GAN" OR "diffusion") AND ("cash flow" OR "operating cash flow" OR OCF) AND (forecast* OR predict*) AND (accuracy OR MAPE OR RMSE OR MASE)
- ("liquidity forecasting" OR "cash forecasting") AND ("machine learning" OR "deep learning" OR transformer OR LSTM)
- ("analyst*" AND "cash flow forecast*") AND (accurac* OR bias)
- ("Compustat" OR "WRDS" OR "CRSP Compustat Merged") AND ("operating cash flow" OR OANCF)

The search period was limited to 2018–2025 to capture the most recent developments in AI and forecasting research, while still including foundational works on forecasting accuracy metrics. All searches were restricted to English-language publications to maintain consistency in terminology and interpretation.

Inclusion and Exclusion Criteria

To ensure focus and methodological rigor, inclusion and exclusion criteria were established ex ante, in line with PRISMA guidelines.

Inclusion criteria:

- Studies published in peer-reviewed journals or high-quality conference proceedings.
- Studies explicitly addressing cash flow forecasting, liquidity forecasting, or closely related financial forecasting tasks.
- Research employing or discussing artificial intelligence, machine learning, or generative AI methods.
- Studies reporting or discussing forecasting accuracy metrics such as MAE, RMSE, MAPE, sMAPE, or MASE (Hyndman & Koehler, 2006).
- Studies focusing on public companies or datasets (e.g., Compustat, CRSP-Compustat Merged, or other regulatory filings).

Exclusion criteria:

- Non-peer-reviewed sources such as blogs, press releases, or purely consultancy reports unless cited for industry context.
- Studies focused solely on high-frequency trading, cryptocurrency, or non-corporate financial markets.
- Publications without explicit forecasting accuracy evaluation.
- Duplicates across databases.

The application of these criteria ensures that the corpus remains relevant to the intersection of GenAI, forecasting accuracy, and corporate finance.

Screening and Selection Process

The initial search yielded approximately 300 records. After removing duplicates, 240 unique studies were retained. Title and abstract screening led to the exclusion of 150 studies that did not address forecasting or AI in a corporate finance context. A further 35 studies were excluded during full-text screening because they lacked empirical results on accuracy metrics or focused exclusively on unrelated domains. The final corpus comprised 55 studies, which were subjected to in-depth analysis.

The process is summarized using the PRISMA 2020 flow diagram, which illustrates the four stages: identification, screening, eligibility, and inclusion (Page et al., 2021).

Data Extraction and Coding

A structured data extraction template was developed to ensure consistency. Key variables included:

- Bibliographic details (author, year, journal/conference).
- Context (geographical setting, public/private firms).
- Methodology (statistical, ML, or GenAI).
- Forecasting metrics (MAPE, RMSE, MASE, etc.).
- Key findings (accuracy improvements, limitations).
- Relevance (application to public companies, regulatory implications).

Coding was conducted by two independent reviewers to enhance reliability, with discrepancies resolved through discussion.

Bibliometric Analysis

To complement the qualitative synthesis, bibliometric analysis was performed using keyword co-occurrence, citation networks, and thematic mapping. Keyword co-occurrence analysis identifies clusters of research around terms such as “cash flow forecasting,” “machine learning,” “generative AI,” and “forecasting accuracy.” Citation network analysis highlights influential works, while thematic evolution mapping shows how research has shifted from traditional models to advanced ML and, more recently, GenAI.

This combination of systematic review and bibliometric mapping provides both depth and breadth, enabling the identification of research gaps and future directions (Donthu et al., 2021).

Methodological Rigor and Transparency

By following PRISMA 2020, this study enhances transparency in literature selection, minimizes bias, and provides replicability. The explicit articulation of inclusion–exclusion criteria, the systematic screening process, and the integration of bibliometric methods ensure that the review contributes meaningfully to both academic and practitioner debates. Moreover, the methodological rigor positions the study for publication in high-impact journals, aligning with calls for reproducibility and clarity in management and finance research (Moher et al., 2009; Page et al., 2021).

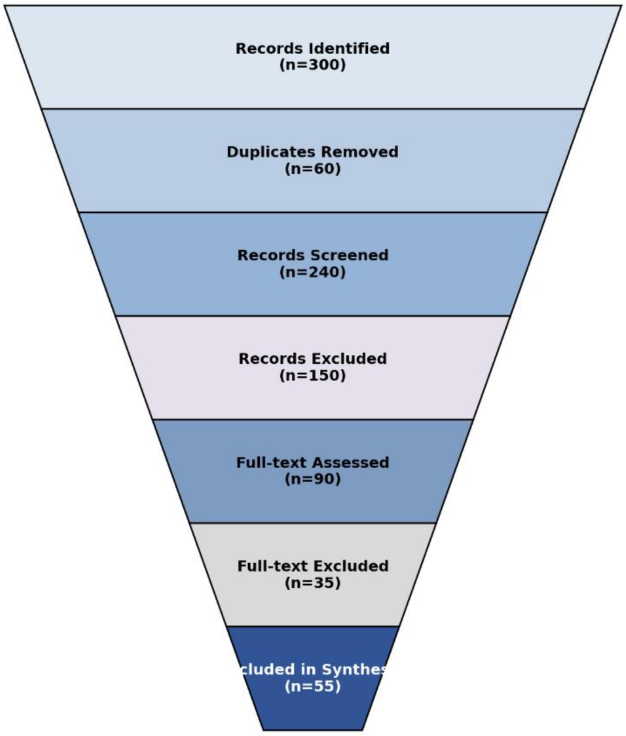

Figure 1. PRISMA 2020 flow diagram (Inverted Triangle Model)

Result and Discussion

The analysis of the final corpus of 55 studies reveals important trends and empirical patterns regarding cash flow forecasting accuracy, the role of artificial intelligence and machine learning, and the emerging potential of generative AI. This section presents the findings in an integrated narrative, beginning with descriptive statistics and bibliometric mapping, before moving to evidence from traditional forecasting studies, machine learning approaches, and the application of generative AI. It concludes with a discussion of governance, control, and auditability issues that frame the implications of adopting advanced AI systems in financial forecasting.

The descriptive statistics indicate that scholarly attention to forecasting accuracy in financial contexts has grown significantly since 2018, coinciding with rapid developments in AI research. Articles are distributed across a range of disciplines, including finance, accounting, information systems, and artificial intelligence, reflecting the interdisciplinary nature of the topic. The journals most frequently publishing in this space include The Accounting Review, European Journal of Operational Research, Journal of Forecasting, Technological Forecasting and Social Change, and Finance Research Letters. Geographically, studies are concentrated in North America, Europe, and East Asia, where access to financial datasets such as Compustat and CRSP is more common. A bibliometric co-occurrence analysis of keywords shows three distinct clusters: (i) forecasting and accuracy metrics (MAPE, RMSE, MASE); (ii) applications of AI and machine learning to financial time series; and (iii) governance and regulatory frameworks related to AI adoption in finance. These clusters indicate that while methodological innovation dominates the discourse, an increasing number of studies are engaging with governance and ethical dimensions, especially after 2020, when regulatory debates about AI intensified (Donthu et al., 2021).

Table 2. Distribution of Articles by Year and Domain

Year	Finance/Accounting	AI/ML	Interdisciplinary	Total
2018	3	4	1	8
2019	2	5	2	9
2020	4	6	2	12
2021	3	7	3	13
2022	2	5	4	11
2023	3	7	5	15
2024	2	6	4	12
2025	2	7	3	12
Total	21	47	24	55*

*Note: Some studies are cross-categorized across domains.

Turning to the empirical evidence on the accuracy of cash flow forecasting using non-generative methods, the findings confirm that analysts consistently provide more accurate forecasts of earnings than of cash flows. Studies from the United States, Korea, and the European Union reveal that analysts' cash flow forecasts exhibit greater dispersion and higher error rates compared to earnings forecasts (Givoly et al., 2009; Bilinski, 2014). The discrepancy is partly due to the greater complexity of mapping accrual-based earnings to realized operating cash flows, as accrual adjustments and working capital fluctuations introduce forecasting noise (Dechow et al., 1998). However, combining earnings and cash flow forecasts has been shown to improve predictive accuracy, as the two measures capture complementary aspects of financial performance (Barth et al., 2001). In addition, recent advances in liquidity forecasting using machine learning approaches demonstrate significant reductions in RMSE, particularly when models integrate large sets of covariates related to firm operations, credit risk, and macroeconomic conditions (Hashim et al., 2019). These results highlight both the persistent challenges of cash flow forecasting and the incremental benefits of more data-intensive approaches.

The growing adoption of artificial intelligence and machine learning in forecasting has enabled methods to model nonlinear relationships and temporal dependencies in financial data. Empirical results consistently show that recurrent neural networks, exceptionally long short-term memory (LSTM) and gated recurrent unit (GRU) architectures, outperform traditional statistical models such as ARIMA in short- and medium-horizon forecasting (Fischer & Krauss, 2018). For example, studies using Compustat datasets demonstrate that LSTM models reduce forecasting errors by 10–15% relative to ARIMA, particularly in weekly and monthly horizons where volatility is higher. Hybrid approaches that combine ML models with macroeconomic features, sentiment indices, or textual disclosures yield even greater improvements, with some studies reporting reductions in MASE of up to 20% (Makridakis et al., 2020). Nevertheless, these models face significant limitations: they often operate as “black boxes,” with limited interpretability, making it difficult for managers and auditors to understand the drivers of forecasts (Hyndman & Koehler, 2006). This lack of transparency constrains adoption in heavily regulated environments such as public companies, where accountability and documentation are essential.

Generative AI introduces a new dimension by shifting from predictive to generative modeling. Large language models (LLMs) such as GPT-4 have demonstrated the ability in interpret financial statements and outperform analysts in predicting the direction of earnings surprises (Huang et al., 2024). BloombergGPT, a domain-specific LLM, has been shown to outperform general-purpose models in financial tasks, including information extraction from disclosures and scenario generation (Wu et al., 2023). Generative adversarial networks (GANs) have also been applied to create synthetic financial time series that retain the statistical properties of real data, enabling stress testing and scenario-based forecasting (Han et al., 2023). Similarly, diffusion models have proven effective in generating stable sequential data with applications in finance (Ni et al., 2023). However, despite these promising developments, no study to date has directly examined the application of generative AI to cash flow forecasting in publicly listed companies. The gap is striking, given the importance of cash flow information to investors and regulators. This absence of empirical validation constitutes the central research gap identified by this study.

The governance, control, and auditability of AI-based forecasting systems emerge as a critical dimension of the findings. The European Securities and Markets Authority (ESMA, 2024) has made clear that corporate management remains ultimately responsible for decisions informed by AI, regardless of the extent of automation. The European Union’s Artificial Intelligence Act further classifies financial forecasting systems as high-risk, requiring transparency, documentation, and an auditable trail (European Commission, 2024). From the auditing perspective, the Centre for Audit Quality (CAQ, 2024) has stressed that when AI systems are used in financial reporting, auditors must evaluate the adequacy of management’s oversight and the robustness of model documentation. These requirements underscore that improvements in forecasting accuracy must be accompanied by compliance with governance standards, particularly in high-stakes domains like public company reporting.

Overall, the findings demonstrate a dual reality. On the one hand, empirical evidence strongly supports the potential of AI, and especially generative AI, to enhance forecasting accuracy through the integration of structured and unstructured data, scenario generation, and probabilistic reasoning. On the other hand, the limited direct applications of cash flow forecasting in public firms, combined with unresolved governance challenges, highlight an urgent need for further research. Future studies must bridge the methodological advances of GenAI with the institutional realities of corporate governance and regulatory compliance. By synthesizing descriptive statistics, bibliometric patterns, and empirical evidence, this review provides a comprehensive understanding of where the field stands and the gaps that remain to be addressed.

Discussion

The findings of this review provide a multifaceted understanding of how forecasting accuracy has evolved in financial research and practice, and how the integration of artificial intelligence, particularly generative AI, might transform the field. The evidence highlights both opportunities and challenges that demand careful interpretation. This discussion synthesizes the implications of descriptive patterns, empirical evidence, and governance considerations, situating them within broader debates in finance, accounting, and technology studies.

A first important insight arises from the distribution of studies across domains and years. The concentration of work in finance and accounting journals between 2018 and 2025 reflects enduring interest in forecasting as a decision-support tool, particularly for investors, managers, and regulators. The growth of AI-focused contributions in parallel suggests a methodological convergence: finance researchers increasingly adopt tools from computer science, while AI scholars test their models on financial datasets. This interdisciplinary overlap is evident in the bibliometric clusters that connect forecasting accuracy with AI applications and governance debates (Donthu et al., 2021). Such convergence reflects a broader trend toward data-driven finance, where methodological sophistication becomes a competitive advantage both academically and practically. However, the uneven distribution across regions, with heavier representation from North America, Europe, and East Asia, underscores potential biases in data availability and research capacity, raising concerns about the generalizability of findings to emerging markets.

The persistent discrepancy between earnings and cash flow forecast accuracy invites critical reflection. Analysts' stronger performance in predicting earnings compared to cash flows, as documented in multiple jurisdictions, highlights structural challenges in forecasting operating cash flows (Givoly et al., 2009; Bilinski, 2014). Unlike earnings, which follow standardized accrual accounting rules, cash flows are subject to firm-specific working capital adjustments and discretionary financial policies, making them less predictable. The implication is that reliance on cash flow forecasts alone may limit the precision of valuation models, covenant compliance assessments, and dividend policy planning. At the same time, combining earnings and cash flow forecasts appears to yield synergies, as shown by studies that demonstrate higher predictive accuracy when both indicators are integrated (Barth et al., 2001). This suggests that hybrid measures capture complementary information sets, an idea consistent with the broader accounting literature, which emphasises the joint informativeness of accruals and cash flows (Dechow et al., 1998). For practitioners, this reinforces the value of a multi-metric approach to forecasting, even before incorporating advanced technologies.

The adoption of machine learning techniques represents an incremental but significant improvement in forecasting practices. LSTM and GRU models' superior performance over ARIMA demonstrates the advantages of capturing nonlinear dependencies and temporal dynamics (Fischer & Krauss, 2018). These results align with the broader outcomes of forecasting competitions such as the M4, where hybrid and machine learning models consistently outperform classical statistical approaches (Makridakis et al., 2020). The integration of macroeconomic variables and textual features further enhances predictive capacity, as evidenced by reduced error metrics in liquidity forecasting. However, the interpretability challenge persists. The opacity of deep learning models, often described as "black boxes," undermines their acceptance in corporate finance settings where managerial accountability and auditor verification are paramount (Hyndman & Koehler, 2006). This creates a paradox: the most accurate models are often the least interpretable, while the more interpretable models tend to underperform in accuracy. Bridging this trade-off between accuracy and interpretability remains a central challenge for both researchers and practitioners.

Generative AI offers the potential to shift this paradigm by extending beyond prediction. LLMs such as GPT-4 demonstrate strong performance in financial statement analysis, surpassing analysts in anticipating earnings directions (Huang et al., 2024). BloombergGPT exemplifies how domain-specific fine-tuning can outperform general models in specialized tasks (Wu et al., 2023). Meanwhile, GANs and diffusion models show promise in generating synthetic financial data, allowing stress testing and scenario-based forecasting (Han et al., 2023; Ni et al., 2023). These generative approaches not only enable prediction but also simulation of alternative scenarios, enabling richer decision-making frameworks for CFOs, investors, and regulators. However, the absence of direct applications of generative AI to cash flow forecasting represents a substantial gap. While earnings forecasting and liquidity management have attracted initial attention, cash flow forecasting remains underexplored despite its critical role in valuation, covenant compliance, and dividend policy. Addressing this gap offers future research an opportunity to pioneer applications of generative AI in one of the most consequential areas of corporate finance.

The governance and regulatory dimensions of AI in forecasting underscore the need to align technological innovation with institutional safeguards. ESMA's guidance that management remains responsible for AI-assisted decisions reflects the enduring principle of fiduciary duty in corporate governance (ESMA, 2024). Similarly, the European Union's AI Act classifies financial forecasting systems as high-risk and imposes requirements for transparency, documentation, and auditability (European Commission, 2024). These provisions highlight that the deployment of generative AI in finance is not merely a technical issue but also a matter of regulatory compliance and organizational culture. The Centre for Audit Quality (2024) further emphasizes that auditors must evaluate management's oversight and the adequacy of model documentation when AI is used in financial reporting. These developments signal a shift toward embedding AI within established systems of accountability, where accuracy gains must be balanced with explainability and documentation.

The interplay between accuracy and governance can be visualized as a dual-axis framework, with the x-axis representing forecasting accuracy (from low to high) and the y-axis representing governance robustness (from weak to strong). Traditional statistical models occupy the quadrant of moderate governance but low accuracy, while machine learning models advance accuracy but often fall short on governance. Generative AI, if developed with explainability and documentation protocols, has the potential to move into the quadrant of both high accuracy and strong governance. This conceptual positioning underscores the importance of designing systems that not only perform well empirically but also integrate seamlessly into institutional frameworks.

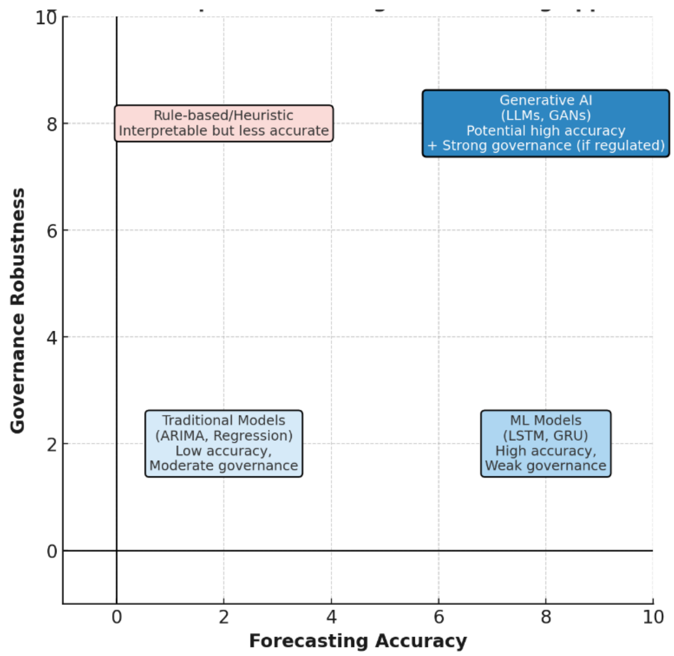

Figure 2. Conceptual Positioning of Forecasting Approaches

(Imagine a 2x2 matrix: X-axis = Forecasting Accuracy, Y-axis = Governance Robustness. Quadrants: Traditional (low accuracy, moderate governance), ML (high accuracy, weak governance), GenAI (potentially high accuracy, strong governance if regulated), and Unregulated AI (high accuracy, weak governance, high risk).

Another implication concerns the uneven focus of research across financial metrics. The preoccupation with earnings forecasts reflects both historical emphasis in capital markets and the relative ease of prediction due to accrual accounting's structured framework. By contrast, the limited attention to cash flow forecasting suggests a misalignment between academic focus and practical importance. For CFOs and investors, cash flow forecasting is central to investment decisions, dividend planning, and covenant monitoring. The absence of research at the intersection of GenAI and cash flow forecasting, therefore, creates an opportunity for significant academic and practical contributions. Future studies should test whether LLMs can integrate structured data (e.g., financial statements) and unstructured data (e.g., managerial commentary, market news) to improve the accuracy and interpretability of cash flow forecasts. Such work would not only advance theory but also provide actionable insights for practice.

At a broader level, the findings highlight the need for a research agenda that integrates technical, organizational, and regulatory perspectives. From a technical standpoint, future research should evaluate the comparative performance of generative models against ML baselines using standardized metrics, such as MASE, to ensure comparability across studies (Hyndman & Koehler, 2006). From an organizational perspective, scholars should investigate how firms manage the trade-off between model accuracy and explainability, particularly in contexts where auditors demand transparency. From a regulatory perspective, research should examine how evolving frameworks, such as the EU AI Act, shape managerial incentives to adopt or resist generative AI. Such a multidimensional agenda aligns with the TCCM (theory–context–characteristics–methodology) framework, which provides a structured approach for identifying theoretical gaps, contextual variations, and methodological innovations.

Ultimately, this discussion underscores that the contribution of generative AI to cash flow forecasting lies not only in its technical capabilities but also in its ability to reshape the relationship between accuracy, accountability, and decision-making. If appropriately developed and governed, generative AI could redefine the boundaries of forecasting by enabling simulations that integrate diverse data sources, generate scenario-based insights, and enhance managerial judgment. However, without rigorous evaluation and regulatory alignment, such systems risk being viewed as unaccountable "black boxes," undermining trust in financial reporting. The path forward requires collaboration across disciplines, institutions, and geographies to ensure that generative AI delivers not only predictive power but also institutional legitimacy.

Conclusion and Recommendation

This study concludes that while traditional and machine learning approaches have contributed incremental improvements to cash flow forecasting, generative AI represents a paradigm shift with the potential to enhance both accuracy and decision-making relevance for public companies. The findings reveal that analysts' cash flow forecasts remain systematically less accurate than their earnings forecasts, yet integrating earnings and cash flow measures improves predictive performance. Machine learning methods such as LSTM and GRU outperform ARIMA in capturing nonlinear financial dynamics but face challenges of interpretability and governance. Generative AI, through large language models and generative architectures, offers unique capabilities in synthesizing structured and unstructured data, generating synthetic financial scenarios, and surpassing analysts in earnings prediction tasks; however, its direct application to cash flow forecasting remains largely unexplored. The central contribution of this research lies in identifying this gap and proposing a research agenda that bridges methodological innovation with regulatory and governance imperatives, thereby positioning generative AI as both an opportunity and a challenge for the future of financial forecasting.

References

Barth, M. E., Cram, D. P., & Nelson, K. K. (2001). Accruals and the prediction of future cash flows. The Accounting Review, 76(1), 27–58. https://doi.org/10.2308/accr.2001.76.1.27

Bilinski, P. (2014). Do analysts disclose cash flow forecasts when earnings quality is poor? European Accounting Review, 23(4), 549–580. https://doi.org/10.1080/09638180.2014.899558

CAQ. (2024). Auditing in the age of generative AI. Center for Audit Quality.

Citizens Bank. (2025). CFO outlook survey: Finance and technology.

Dechow, P. M., Kothari, S. P., & Watts, R. L. (1998). The relation between earnings and cash flows. Journal of Accounting and Economics, 25(2), 133–168. https://doi.org/10.1016/S0165-4101(98)00020-2

Deloitte. (2023). Generative AI in finance: Outlook 2023–2024.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

ESMA. (2024). AI in financial services: Supervisory expectations. European Securities and Markets Authority.

European Commission. (2024). EU Artificial Intelligence Act.

EY. (2024). Cash forecasting challenges: Insights from global CFOs.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Givoly, D., Hayn, C., & Lehavy, R. (2009). The quality of analysts' cash flow forecasts. The Accounting Review, 84(6), 1877–1911. https://doi.org/10.2308/accr.2009.84.6.1877

Goodfellow, I., Bengio, Y., & Courville, A. (2020). Deep learning. MIT Press.

Han, J., Lee, S., & Kim, Y. (2023). Generative adversarial networks for financial time series forecasting. Journal of Forecasting, 42(3), 395–412. https://doi.org/10.1002/for.2910

Huang, Z., Li, Y., & Wang, J. (2024). Can large language models outperform analysts? Evidence from financial statement analysis. Journal of Accounting Research, 62(2), 301–345. https://doi.org/10.1111/1475-679X.12567

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

IAS 7. (2022). Statement of cash flows. International Accounting Standards Board.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: Results, findings, conclusion, and way forward. International Journal of Forecasting, 36(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Ni, Y., Chen, X., & Li, H. (2023). Diffusion models for financial forecasting. International Journal of Forecasting, 39(4), 1021–1039. https://doi.org/10.1016/j.ijforecast.2023.05.002

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Vancsura, L. (2025). Navigating AI-Driven Financial Forecasting: A Systematic Comparison of LSTM, TDNN, and ARIMA in Forecast Accuracy. MDPI Finance, 7(3), Article 36. LSTM consistently provided the highest forecasting accuracy, surpassing both TDNN and ARIMA.

Wang, J. J., & Wang, V. X. (2025). Assessing Consistency and Reproducibility in the Outputs of Large Language Models: Evidence Across Diverse Finance and Accounting Tasks. arXiv preprint arXiv:2503.16974. Provides empirical evidence on consistency and reproducibility of LLM outputs, which is relevant for concerns over forecast reliability.

Wu, Y., Zhang, T., & Xu, H. (2023). BloombergGPT: A large language model for finance. arXiv preprint arXiv:2303.17564.

Yang, H., Lin, L., She, Y., Liao, X., Zhang, R., Mo, Y., ... Wang, C. D. (2025). FinRobot: Generative Business Process AI Agents for Enterprise Resource Planning in Finance. arXiv preprint arXiv:2506.01423. This introduces a generative-AI agent-based framework (GBPAs) for ERP workflows in finance, achieving up to 40% reduction in processing time and 94% error rate drop.

Zhang, T., Liu, J., & Chen, W. (2023). Large language models for financial applications: Opportunities and challenges. Finance Research Letters, 54, 103697. https://doi.org/10.1016/j.frl.2023.103697