FDLLaMa: Unlocking the Potential of Finance-Aware LLaMA in Combating Fraud

1st Qiufan Wu
School of Electronic and Computer Engineering
Peking University, Shenzhen, China
Shenzhen, China
qfwu@stu.pku.edu.cn

2nd Hui Li
School of Electronic and Computer Engineering
Peking University, Shenzhen, China
Shenzhen, China
https://orcid.org/0000-0001-5244-9473

3rd Aofan Liu
School of Electronic and Computer Engineering
Peking University, Shenzhen, China
Shenzhen, China
af.liu@stu.pku.edu.cn

***Abstract*—Financial statement fraud poses a significant threat to the stability of economic markets, causing substantial financial losses and eroding trust among stakeholders. Detecting such fraudulent activities remains a complex and resource-intensive task for regulators, auditors, and financial analysts. In this study, we investigate how combining pre-trained contextual language models with advanced domain-specific knowledge enhances the detection of financial statement fraud. Leveraging FinBERT, a model trained on financial text, we extract domain-specific embeddings from financial reports and integrate them with LLaMA, a state-of-the-art large language model. Through fine-tuning and knowledge distillation, we utilize embeddings from FinBERT as inputs alongside raw textual data, enabling LLaMA to inherit FinBERT's domain expertise. Our final model demonstrates superior performance, outperforming traditional machine learning models and standalone large language models in financial fraud detection. Specifically, our approach identifies a significantly higher proportion of fraudulent firm-year observations compared to baseline methods, optimizing the investigation process for regulators and stakeholders. These results highlight the potential of integrating domain-specific knowledge and large-scale language models to improve fraud detection, offering a practical and impactful solution for addressing financial irregularities.**

***Index Terms*—financial statement fraud detection, LLaMA, knowledge distillation**

I. Introduction

A. Background

Financial fraud refers to the use of deceitful, illegal, or deceptive practices to gain economic advantage. Fraudulent activities can occur across various sectors of finance, including banking, insurance, taxation, and corporations. Types of financial fraud, such as credit card fraud, tax evasion, financial statement fraud, money laundering, and other related offenses, have become increasingly significant challenges. Despite efforts to mitigate these issues, financial fraud continues to negatively impact businesses and society, resulting in billions of dollars in losses annually. These substantial financial losses significantly affect individuals, merchants, and financial institutions alike.

It is estimated that fraudulent activities have had a global financial impact of approximately \$5.127 trillion over the past two decades, with losses increasing by 56% in the last ten years [1]. These figures represent only the direct financial losses, excluding indirect costs such as reputational damage to stakeholders, including investors, creditors, and employees. In extreme cases, financial fraud can even result in organizational bankruptcy [2].

The scope of financial fraud is broad, encompassing corruption, asset misappropriation, and fraudulent financial reporting. This study focuses on financial statement fraud, which, according to Hajek and Henriques [3], involves the deliberate omission or misrepresentation of financial information in reports. Although asset misappropriation and corruption are more frequent, the consequences of financial statement fraud are far more severe, with a median loss of up to \$954,000. Such fraudulent activities often lead to plummeting stock prices, delisting from stock exchanges, and, in some instances, bankruptcy [4].

These observations underscore the extensive consequences of fraud and highlight the critical importance of minimizing its occurrence.

The Association of Certified Fraud Examiners (ACFE) reports that 10% of white-collar crime incidents involve falsified financial statements [5]. The ACFE categorizes occupational fraud into three types: asset misappropriation, corruption, and financial statement fraud. Among these, financial statement fraud results in the most significant financial losses. While asset misappropriation and corruption occur far more frequently, their financial impacts are considerably less severe.

A survey by EisnerAmper, a prominent U.S. accounting firm, further emphasizes the disproportionate damage caused by financial statement fraud [6]. Consequently, this study focuses specifically on financial statement fraud due to its devastating financial implications and widespread consequences.

Financial statements are documents that detail a company's business activities and financial performance. These records include information about revenue, expenses, profits, loans, potential future challenges, and management's commentary on business performance [7, 8]. All companies are required to publish financial statements quarterly and annually. These reports are essential for assessing a company's performance, financial health, and profitability potential.

Financial statement fraud involves falsifying financial reports to inflate a company's profitability, artificially raise stock prices, evade taxes, or secure bank loans.

The complexity and diversity of financial fraud methods—such as revenue falsification or understatement of liabilities—make fraud detection increasingly challenging [9].

Moreover, advancements in modern technologies, including the internet and mobile computing, have contributed to the rise of financial fraud in recent years [10]. Social factors, such as the increased issuance of credit cards, have not only led to higher spending but also to a surge in fraudulent activities [11]. Fraudsters continuously evolve their techniques, necessitating the development of detection methods that can adapt to these advancements [12].

B. Related Works

Traditional methods for detecting financial fraud often rely on external auditors, which have proven to be inefficient, time-consuming, and costly [13]. Additionally, these traditional approaches face ethical challenges, such as conflicts of interest when auditors have financial ties to the company being audited, and practical limitations in identifying complex fraudulent schemes [14]. Addressing financial fraud is urgent not only to recover economic losses but also to restore trust in the financial system and safeguard stakeholders. Developing advanced detection systems is critical for investors, audit firms, and regulatory bodies [15, 16]. Therefore, combating financial fraud is an immediate and pressing challenge that requires effective and timely solutions.

In financial statement fraud detection research, methodologies have evolved alongside technological advancements. Early studies primarily relied on traditional statistical learning techniques such as logistic regression. Logistic regression (LR) is a linear model-based method for statistical analysis of datasets. It operates by performing regression on a set of variables, making it a widely applicable approach for identifying patterns and elucidating relationships between one or more independent variables and a dependent variable [17].

Hajek and Henriques emphasized the use of Random Forest (RF) in detecting fraudulent transactions [3], highlighting its strength in managing imbalanced datasets through techniques like the Synthetic Minority Oversampling Technique (SMOTE).

Despite these successes, there is a growing need for a more holistic approach that integrates various financial data types alongside textual data, such as the Management Discussion and Analysis (MD&A) sections of reports.

Support Vector Machines (SVMs) have gained popularity due to their ability to handle both linear and non-linear classification problems. A study by Rizki et al. applied SVMs to detect financial fraud among Indonesian public companies [18]. The research revealed that feature selection significantly improved SVM accuracy, achieving a notable rate of 88.37%. However, this study relied solely on financial ratios, omitting the textual data from MD&A discussions.

To address this gap, Goel et al. investigated the use of SVMs for fraud detection within MD&A sections of annual reports [19].

While SVMs offer considerable flexibility in handling diverse classification problems, their reliance on feature selection and computational intensity can be limiting, especially when working with large datasets.

With the rapid expansion of computational power and data scale, Artificial Neural Networks (ANNs) and Deep Learning (DL) technologies have become increasingly mainstream.

The application of ANNs in financial fraud detection has shown varying levels of success across different datasets. In one U.S. study, a dataset comprising 208 fraud cases and 7,341 non-fraud cases, derived from corporate annual financial reports and the Compustat database, was analyzed [20]. The ANN model achieved an overall accuracy of 67.9% on the test dataset.

The growing recognition of unstructured data's value has led to widespread adoption of text mining and natural language processing (NLP) technologies. Text-based financial information, such as annual reports and disclosure documents, often contains crucial signals of fraud. NLP techniques enable the extraction of semantic and sentiment features from such text, while their integration with text mining tools provides robust support for identifying potential fraud signals in financial reports, corporate disclosures, and related documents.

Hierarchical Attention Networks (HAN) [20] have emerged as state-of-the-art models in financial fraud detection, particularly for analyzing text and financial data. HAN is designed to capture hierarchical language patterns by incorporating financial ratios alongside text. It stands out as one of the few deep learning models that utilize both textual and quantitative features for fraud detection. [20].

However, it is important to note that most comparisons of HAN's performance have been against traditional machine learning models or older deep learning architectures. While HAN has proven effective, its performance relative to newer specialized models like FinBERT remains largely unexplored.

FinBERT [21], a BERT-based model optimized for financial text, excels in tasks like sentiment analysis and text classification but is limited in fraud detection compared to larger models like LLaMA [22].

LLaMA [23], developed by Meta AI, combines scalability and efficiency, excelling in diverse NLP tasks with fewer resources.

This study integrates FinBERT's financial expertise with LLaMA's versatility using LoRA [24], fine-tuning LLaMA with domain-specific embeddings. The resulting model improves precision in financial fraud detection by blending domain knowledge with advanced language modeling.

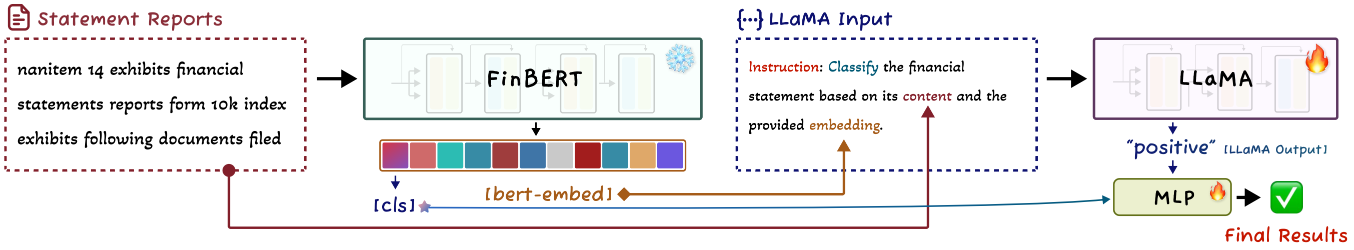
Statement Reports
nanitem 14 exhibits financial statements reports form 10k index exhibits following documents filed
FinBERT
[cls]
[bert-embed]
{···} LLaMA Input
Instruction: Classify the financial statement based on its content and the provided embedding.
LLaMA
"positive" [LLaMA Output]
MLP
Final Results

Fig. 1. The Pipeline of FDLLaMa

II. Method

In earlier studies on fraud detection, full financial texts were typically considered a less significant source of data. However, this research focuses on utilizing large language models (LLMs) to conduct an in-depth analysis of these texts, identifying fraudulent activities through a detailed examination of subtleties within the descriptions.

A. Problem Formulation

In conventional fraud detection methods, such as logistic regression, a textual context x is initially transformed into a vector V by techniques such as TF-IDF [25]. The regression model Φ then generates the prediction following this vectorized representation, formulated by:

$$y = \Phi\left(V\left(x\right)\right). \tag{1}$$

With the advancement of machine learning, traditional regression models have gradually been replaced by neural networks and adhere to the aforementioned paradigm Eq. 1.

As large language models have advanced, a variety of studies are now using their robust language understanding capabilities in financial analysis. Amit *et al.* recently employed FinBERT [26] for detecting fraud. In contrast to conventional techniques that convert words into vectors, these methodologies capitalize on the models' intrinsic embedding methods. FinBERT, a model tailored with comprehensive financial insights, is specifically crafted for financial analytic purposes. Nonetheless, owing to BERT's intrinsic sensitivity to data processing, FinBERT's classification performance is substantially inferior to that of conventional classifiers and neural network-based techniques [26].

This paper presents FDLLaMA, designed to address this limitation by employing a dual-residual architecture. This approach integrates FinBERT's comprehensive financial knowledge with LLaMA's effective data processing capabilities [27] for the purpose of fraud detection. We enhance the fundamental framework outlined in Eq. 1 by establishing FinBERT as a financial knowledge-augmented content encoder Ψ. Subsequently, the methodology of FDLLaMA can be expressed as:

$$y = \Phi\left(x, \Psi\left(x\right)\right). \tag{2}$$

Utilizing this approach, we detail the FDLLaMA pipeline, depicted in Fig. 1.

B. Statement-Embedding Combination

As detailed in Sec. II-A, LLaMA exhibits strong capabilities in text and data analysis. Nonetheless, an inadequate understanding of finance serves as a major limitation, potentially resulting in detection shortcomings. To overcome this, we implemented the *Statement-Embedding Combination* mechanism to integrate financial insights supplied by FinBERT.

This approach involves two primary steps. Initially, the original statement reports x are processed by FinBERT, as illustrated in Fig. 1. FinBERT, being an encoder-centric model, is tailored for classification and outputs a class token y' along with embedding features r. Next, we formulate the LLaMA instruction input based on our established guideline depicted in the figure, incorporating both the statement reports x and the embedding features r obtained from FinBERT. Since LLaMA utilizes a decoder-oriented architecture, it produces a binary classification output y'', such as *positive* or *negative*.

C. Token-Output Combination

In contrast to earlier classification techniques, we present the *Token-Output Combination*, which empowers FDLLaMA to reach the final decision by combining the tendencies of both LLaMA and FinBERT, rather than relying solely on the classification outcomes of either model, enabling a more balanced and well-informed evaluation. Therefore, the ultimate determination by FDLLaMA can be expressed as:

$$y = \Omega\left(y', y''\right), \tag{3}$$

where Ω denotes a multi-layer perception (MLP) component.

D. Loss Function

Our FDLLaMA model possesses a substantial number of parameters. To enhance training efficiency effectively, we incorporate LoRA and employ block-wise training. In particular, we keep the parameters of FinBERT static to preserve its specialized financial knowledge, preventing it from being affected by binary fraud detection tasks. Concurrently, we also freeze LLaMA's parameters and utilize LoRA for fine-tuning. The loss function corresponding to this approach can be formulated as:

$$\mathcal{L}_{\texttt{LLaMA}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\log p(y_t^{(i)} \mid x^{(i)}, y_1^{(i)}, \ldots, y_{t-1}^{(i)}), \tag{4}$$

where N is the total sample count, T indicates the full length of the target sequence, x^i signifies the statement for the i-th sample, and y_t^i denotes the t-th token of the target.

For classification tasks using the MLP, the standard cross-entropy loss function is applicable:

$$\mathcal{L}_{\texttt{MLP}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right], \tag{5}$$

where N represents the total number of samples, y_i represents the actual label for the i^{th} sample, and p_i indicates the probability that the model classifies this sample as positive. Utilizing Eq. 4 in conjunction with Eq. 5, we construct an effective learning framework for our FDLLaMA.

III. Experiments

A. Dataset

Fraud detection is a challenging task due to the limited number of known fraudulent cases. The severe imbalance between positive (fraudulent) and negative (non-fraudulent) classes poses a significant obstacle for classification. For instance, in annual reports submitted to the SEC between 1999 and 2019, the ratio of fraudulent to non-fraudulent filings was approximately 1:250. In past research, the number of companies engaged in fraudulent activities within datasets has ranged from 12 to 788 [20].

This study utilizes the publicly available Financial Statement Fraud Data dataset from Kaggle [28]. This dataset comprises financial documents submitted by companies to the U.S. Securities and Exchange Commission (SEC). Each row in the dataset represents a unique filing, with columns including the document name, Central Index Key (CIK), filing year, sections containing textual information, company name, and a classification variable indicating fraudulent activity.

The dataset includes records from 85 companies involved in fraud cases and an equal number of companies with no fraudulent activity. Fraudulent companies were identified using the SEC Enforcement Division's FCPA (Foreign Corrupt Practices Act) list (www.sec.gov, 2BC), and CIK numbers were retrieved from SEC filings (Sec.gov, 2023). SEC data was chosen for its consistent format across company reports and its provision of legally mandated, reliable 10-K filings. These documents are particularly valuable for detecting fraudulent activities.

The final dataset comprises information from companies' MD&A (Management Discussion and Analysis) sections as well as historical financial statements declared on the SEC website. The MD&A section was specifically used to extract textual data, as investors commonly analyze this portion for informational advantages [29, 30, 31]. Holder-Webb [32] highlighted the MD&A section of 10-K reports as a formally recognized source of valuable information, signaling potential financial distress to investors.

Given the token limitations of large language models (LLMs), reducing the number of columns was necessary. The data was extracted using Python scripts (refer to the data extraction script) from Hugging Face's 10K-Fillings dataset library [35]. The dataset was structured to include specific columns (e.g., section_7 and section_8) that contained the aforementioned textual information. For use with machine learning, deep learning, and LLM models, the dataset was prepared with one column containing text data and another with labels indicating fraud or no fraud.

B. Evaluation Metrics

In the context of financial statement fraud detection, model performance is evaluated within a binary classification framework, to quantify the predictive power of the model, a combination of evaluation metrics is employed. This study assesses model performance using Accuracy, Precision, Recall and F1-Score.

In the domain of financial statement fraud detection, the cost of failing to detect fraud often exceeds the cost of misclassifying non-fraudulent cases.

Consequently, precision become critical determinants of a model's effectiveness.

C. Benchmark Testing

This study focuses on assessing the performance of large language models (LLMs) in detecting financial fraud from textual data. Its primary objective is to evaluate and compare the effectiveness of traditional machine learning models and language models in this domain.

The dataset used in this study was manually collected and labeled by the researchers [28], rather than sourced from providers like Compustat or aggregated from various other sources as mentioned in literature reviews.

The models trained on the same dataset include Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Hierarchical Attention Networks (HAN). These models were selected based on their prior success in financial fraud detection.

In addition to traditional models and neural networks, this study incorporates language models like FinBERT [21] and GPT-2 [33]. FinBERT is a domain-specific, pre-trained language model optimized for financial data, making it highly relevant to this research. GPT-2, on the other hand, excels in general text processing. Including these language models allows for insightful comparisons to determine whether pre-training on financial data significantly enhances fraud detection, particularly within the MD&A and financial sections of annual reports.

In financial statement fraud detection, minimizing false negatives is critical, as misclassifying a fraudulent company as non-fraudulent can result in billions of dollars in losses. Consequently, precision and recall are emphasized in evaluating model performance.

As shown in Table 1, the Random Forest model achieved the highest accuracy at 0.94, while the HAN model achieved the highest recall at 0.98. However, the proposed FDLLaMA model outperformed others in precision (0.92) and F1-score (0.93), demonstrating its superior ability to balance precision and recall in detecting financial fraud.

TABLE I
QUANTITATIVE COMPARISON OF CLASSIFICATION METRICS ACROSS 7 METHODS.
THE BEST SCORES FOR EACH METRIC ARE HIGHLIGHTED IN **BOLD**, WHILE THE SECOND-BEST SCORES ARE UNDERLINED.

Classifier		**Classification Metrics** [@.5]			
Type	Name	Accuracy ↑	Presicion ↑	Recall ↑	F1-Score ↑
Traditional	Logistic Regression	0.79	0.77	0.81	0.78
	Random Forest	**0.94**	0.85	0.88	0.86
	Support Vector Machine [SVM]	0.84	0.75	0.97	0.85
Neural Networks	Artificial Neural Network [ANN]	0.92	0.81	0.96	0.88
	Hierachical Attention Network [HAN]	0.47	0.47	**0.98**	0.64
Large Language Models	GPT-2 [33]	0.44	0.71	0.44	0.40
	FinBERT [34]	0.79	0.78	0.83	0.81
	FDLLaMA [Ours]	0.81	**0.92**	0.94	**0.93**

TABLE II
ABLATION STUDY OF KEY COMPONENTS IN OUR FDLLAMA MODEL.
✓ INDICATES THE INCLUSION OF THE CORRESPONDING COMPONENT.

Components		**Classification Metrics** [@.5]			
FinBERT	Residual	Accuracy	Presicion	Recall	F1-Score
	✓	0.64	0.68	0.82	0.74
✓		0.76	0.84	0.81	0.83
✓	✓	**0.81**	**0.92**	**0.94**	**0.93**

D. Ablation Experiment

As shown in Table 2, we conducted ablation experiments on two key modules in the FDLLaMA model:

1) Ablation of FinBERT Knowledge Distillation : To validate the benefits of using FinBERT for knowledge distillation, we designed an experiment where FinBERT was excluded from embedding tokens for textual data. Instead, LLaMA directly classified the text data. The experimental results, as shown in the table, revealed an accuracy of 0.64, precision of 0.68, recall of 0.82, and an F1-score of 0.74. All four evaluation metrics were lower than those of the proposed FDLLaMA model, demonstrating the advantages of incorporating FinBERT for knowledge distillation.

The superior performance of FinBERT stems from its specialized capability in the financial domain, which provides LLaMA with highly refined domain-specific features. This significantly reduces LLaMA's learning burden and enhances its classification performance. Conversely, the direct use of LLaMA for classification yielded subpar results, primarily due to the model's lack of deep understanding of financial contexts, which hindered its feature extraction and classification capabilities. These findings highlight the potential of combining domain-specific models with general-purpose large models to achieve better performance in specialized tasks such as fraud detection.

2) Ablation of the Text Residual Module: In the proposed approach, both the tokens derived from FinBERT knowledge distillation and the original textual data (inputs) were fed into LLaMA. To verify the advantage of using text residual connections, we conducted an experiment where only the tokens output by FinBERT were used as LLaMA's input. The results, as shown in the table, indicated an accuracy of 0.76, precision of 0.84, recall of 0.81, and an F1-score of 0.83. Again, all metrics were lower than those of the FDLLaMA model, confirming the necessity of incorporating text residual connections.

While FinBERT's embedding tokens extract high-dimensional financial semantic features, they may lose some fine-grained information present in the original text. Text residuals (inputs) retain this original textual information, compensating for the deficiencies in FinBERT embeddings. This enhances the diversity and completeness of feature representations while leveraging LLaMA's ability to effectively fuse multiple inputs.

IV. CONCLUSION

Detecting financial statement fraud remains a challenging task due to the constantly evolving strategies of fraudulent entities. This study explores the integration of domain-specific textual knowledge with advanced deep learning models to enhance fraud detection. By leveraging FinBERT for domain-specific knowledge extraction and distillation, and fine-tuning LLaMA, the proposed FDLLaMA framework significantly outperforms traditional machine learning models, neural networks, and standalone language models across key performance metrics.

As shown in Table 1, FDLLaMA achieves notable improvements in precision and F1-score compared to classifiers like Random Forest and SVM, as well as neural network-based methods. The results underscore the value of combining domain-specific embeddings with large language models to capture the nuanced linguistic patterns and subtle cues in financial reports. Unlike traditional approaches that often miss such complexities, FDLLaMA effectively integrates textual and residual information, delivering a more robust fraud detection solution.

References

[1] J. Gee, M. Button, and G. Brooks, “The financial cost of fraud,” *MacIntyre Hudson, Milton Keynes*, 2009.

[2] Z. Rezaee, “Causes, consequences, and deterence of financial statement fraud,” *Critical perspectives on Accounting*, vol. 16, no. 3, pp. 277–298, 2005.

[3] P. Hajek and R. Henriques, “Mining corporate annual reports for intelligent detection of financial statement fraud–a comparative study of machine learning methods,” *Knowledge-Based Systems*, vol. 128, pp. 139–152, 2017.

[4] J. Macey, M. O’Hara, and D. Pompilio, “Down and out in the stock market: The law and economics of the delisting process,” *The Journal of Law and Economics*, vol. 51, no. 4, pp. 683–713, 2008.

[5] T. W. Singleton and A. J. Singleton, *Fraud auditing and forensic accounting*. John Wiley & Sons, 2010, vol. 11.

[6] S. K. Dutta, *Statistical techniques for forensic accounting: Understanding the theory and application of data analysis*. FT Press, 2013.

[7] F. H. Glancy and S. B. Yadav, “A computational model for financial reporting fraud detection,” *Decision support systems*, vol. 50, no. 3, pp. 595–601, 2011.

[8] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, “Detection of financial statement fraud and feature selection using data mining techniques,” *Decision support systems*, vol. 50, no. 2, pp. 491–500, 2011.

[9] G. L. Gray and R. S. Debreceny, “A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits,” *International Journal of Accounting Information Systems*, vol. 15, no. 4, pp. 357–380, 2014.

[10] I.-C. Yeh and C.-h. Lien, “The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients,” *Expert systems with applications*, vol. 36, no. 2, pp. 2473–2480, 2009.

[11] D. Sánchez, M. Vila, L. Cerda, and J.-M. Serrano, “Association rules applied to credit card fraud detection,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3630–3640, 2009.

[12] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.

[13] A. Dyck, A. Morse, and L. Zingales, “Who blows the whistle on corporate fraud?” *The Journal of Finance*, vol. 65, no. 6, pp. 2213–2253, 2010.

[14] R. Simnett, A. Vanstraelen, and W. F. Chua, “Assurance on sustainability reports: An international comparison,” *The accounting review*, vol. 84, no. 3, pp. 937–967, 2009.

[15] A. Abbasi, C. Albrecht, A. Vance, and J. Hansen, “Metafraud: a meta-learning framework for detecting financial fraud,” *Mis Quarterly*, pp. 1293–1327, 2012.

[16] W. S. Albrecht, C. Albrecht, and C. C. Albrecht, “Current trends in fraud and its detection,” *Information Security Journal: a global perspective*, vol. 17, no. 1, pp. 2–12, 2008.

[17] J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” *Computer and Security*, vol. 57, pp. 47–66, 2016.

[18] A. A. Rizki, I. Surjandari, and R. A. Wayasti, “Data mining application to detect financial fraud in indonesia’s public companies,” in *2017 3rd International Conference on Science in Information Technology (ICSITech)*. IEEE, 2017, pp. 206–211.

[19] S. Goel, J. Gangolly, S. R. Faerman, and O. Uzuner, “Can linguistic predictors detect fraudulent financial filings?” *Journal of Emerging Technologies in Accounting*, vol. 7, no. 1, pp. 25–46, 2010.

[20] P. Craja, A. Kim, and S. Lessmann, “Deep learning for detecting financial statement fraud,” *Decision Support Systems*, vol. 139, p. 113421, 2020.

[21] Y. Yang, M. C. S. Uy, and A. Huang, “Finbert: A pretrained language model for financial communications,” *arXiv preprint arXiv:2006.08097*, 2020.

[22] Y. Nie, Y. Kong, X. Dong, J. M. Mulvey, H. V. Poor, Q. Wen, and S. Zohren, “A survey of large language models for financial applications: Progress, prospects and challenges,” *arXiv preprint arXiv:2406.11903*, 2024.

[23] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, “Llama: Open and efficient foundation language models,” *arXiv preprint arXiv:2302.13971*, 2023.

[24] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” *arXiv preprint arXiv:2106.09685*, 2021.

[25] P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, “Fraud detection with natural language processing,” *Machine Learning*, vol. 113, no. 8, pp. 5087–5108, 2024.

[26] A. Kedia, “Financial fraud detection using llms,” https://github.com/amitkedia007/Financial-Fraud-Detection-Using-LLMs, 2025, accessed: 2025-01-08.

[27] J. Zhao, Z. Zhang, L. Gao, Q. Zhang, T. Gui, and X. Huang, “Llama beyond english: An empirical study on language capability transfer,” *arXiv preprint arXiv:2401.01055*, 2024.

[28] A. Kedia, “Financial statement fraud data,” https://www.kaggle.com/datasets/amitkedia/financial-statement-fraud-data, 2023, accessed: 2025-01-08.

[29] A. Durnev and C. Mangen, “The spillover effects of md&a disclosures for real investment: The role of industry competition,” *Journal of Accounting and Economics*, vol. 70, no. 1, p. 101299, 2020.

[30] T. Loughran and B. McDonald, “Textual analysis in accounting and finance: A survey,” *Journal of Accounting Research*, vol. 54, no. 4, pp. 1187–1230, 2016.

[31] V. Muslu, S. Radhakrishnan, K. Subramanyam, and D. Lim, “Forward-looking md&a disclosures and the information environment,” *Management Science*, vol. 61, no. 5, pp. 931–948, 2015.

[32] L. Holder-Webb and J. R. Cohen, “The association between disclosure, distress, and failure,” *Journal of Business Ethics*, vol. 75, pp. 301–314, 2007.

[33] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, “Language models are unsupervised multitask learners,” *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[34] D. Araci, “Finbert: Financial sentiment analysis with pre-trained language models,” *arXiv preprint arXiv:1908.10063*, 2019.

[35] L. Loukas, M. Fergadiotis, I. Androutsopoulos, and P. Malakasiotis, “Edgar-corpus: Billions of tokens make the world go round,” *arXiv preprint arXiv:2109.14394*, 2021.