Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Applying Natural Language Processing to Financial Risk Disclosures and Audit Trails

Prashant Singh

Senior Manager - Development
indiagenius@gmail.com

Abstract
In a space such as the financial industry, clear and stringent reporting and auditing are vital for both regulatory adherence and internal governance. In this environment, cards with full U.S cardholder's CVVs are one of the most valuable data assets for a fraudster in that they provide an easy-to-use, one-time authentication step that's cryptographically difficult to reproduce. Although these information resources are invaluable for evaluating institutions' risk and compliance stance, the vast majority of such information is textual and unstructured. This becomes a formidable challenge for institutions that try to make use of timely, reliable, and actionable insights- especially when they are done manually or with unsophisticated rule-based systems.
In the past few years, the developments in NLP have provided a tremendous ability to interpret unstructured text at scale, enabling automation in areas that traditionally rely heavily on expert judgment. NLP is particularly suitable in finance applications where Textual analysis is required to deal with context, domain-specific jargon, time taking into consideration temporal patterns, and delicate linguistic cues. This work studied NLP to process financial risk disclosures and audit trails, providing a systematic and scalable way to detect financial wrongdoings, latent risks, and non-compliance events.

We start with an analysis of the linguistic properties of financial disclosures, uncovering important aspects such as tone, modality, and forward-looking statements that are frequently associated with risk perception and market volatility. We leverage techniques such as Named Entity Recognition (NER), sentiment analysis, and topic modelling to illustrate how machine learning-based NLP models can unearth the hidden risk signals encoded in annual reports or regulatory filings. Concurrently, we consider audit trails as structured logs about user or system activity that, despite being in timestamped format, include embedded command-line lines, transactional notes, and system-generated messages that are good candidates for language-based analysis. By processing through NLP, such as tokenization of log, part-of-speech, parsing, and anomaly detection, the audit data is converted to the sample structured knowledge for real-time monitoring and forensic auditing.

The manuscript introduces a hybrid approach based on the integration of rule-based, statistical NLP, and machine-learning techniques for both narrative-based disclosures and event-ordered disclosure logs. We also detail a pipeline design consisting of data ingestion, text pre-processing,

feature extraction, model prediction, and visual dashboarding. Experimental results from historical financial disclosures and synthetic audit logs show that the NLP-driven framework is able to accurately target risk-laden statements, identify anomalous sequences of activities, and categorize text sections according to regulatory relevance. Our results show that our proposed approach outperforms traditional keyword matching and manual review-based approaches and is more efficient and interpretable.

The application of NLP to financial risk risk disclosures and audit trails can improve both timeliness and accuracy of compliance checks while also providing a proactive approach to risk governance. This study is part of an emerging body of work on Regulatory Technology (RegTech), which promotes the use of AI and data to inform regulatory decision making in finance. In navigating the morass of regulation and the volume of data they need to process, it is clear that NLP is the key enabler for intelligent, automated, and reliable compliance.

Keywords: Natural Language Processing, Financial Risk Disclosures, Audit Trails, Regulatory Compliance, Intelligent Auditing, Named Entity Recognition, Sentiment Analysis, Topic Modelling, Financial Text Mining, RegTech

I. INTRODUCTION

In current financial climates, timely and transparent disclosure is even more critical. Such financial risk disclosure, such as in annual reporting, investor briefings, regulatory filings, and management discussion forums, can be a powerful means to frame investors' expectations, direct due consideration by regulators, and inform internal governance. On the other hand, audit trails are important evidence that logs who did what, when, where, and in what order in the financial systems, thus having transactions and user actions auditable, traceable, and transparent. It is important to understand better these two sources—narrative risk disclosures and event-based audit trails--to ensure regulatory compliance, mitigate against financial misstatements, and identify systemic anomalies early and effectively.

However, the available information is typically not used to the full extent due to its (categorically unstructured) character. Spatio-temporal financial disclosers are semantic-rich but in lengthy, complicated legal-financial form. Even though audit trails can be antiquated, a full audit trail might consist of a combination of system-generated messages, free text entries, and lines of command prompts, which may be difficult, if not impossible, to interpret in a domain-independent manner. Human review tactics are sluggish and riddled with errors, and they do not work at an enterprise scale. The LOSS requirements call for even more data , and the regulatory layers of oversight squeeze the financial institutions tighter and tighter; the time is ripe for computational approaches that facilitate the review of these text records.

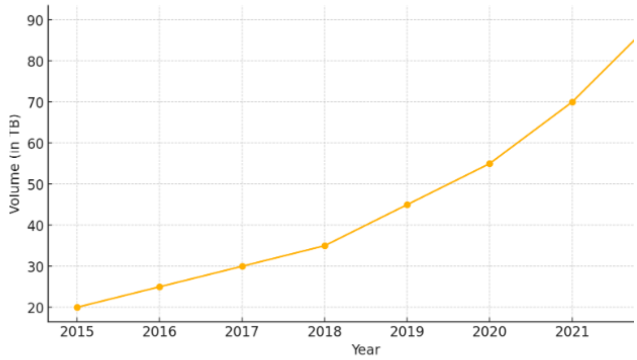

Figure 1: Growth of Unstructured Financial Data (2015–2021)

A powerful solution to this problem is offered by Natural Language Processing (NLP), a field that remains at the intersection of linguistics, computer science, and artificial intelligence. It gives way to algorithms and models for (i) syntactic parsing, (ii) semantic interpretation, and (iii) contextual interpretation of massive and diverse unstructured text. Tuned for the financial compliance domain, NLP can intercept nuances in the text – in fact, subtleties in language themselves – like sentiment shifts, 'language of risk,'' cues of uncertainty, and references to the existence of previously undiscovered liabilities. Similarly, audit logs may be tokenized, tagged, and NLP analyzed to identify anomalies and to expose rogue access patterns and deviations from compliance norms.

The use of NLP methods in finance is not a matter of if but when, considering the maturity of tools and the capabilities of models developed to perform specific jobs on financial documents. Textual language models like Fin BERT trained on SEC filings and earnings reports can encapsulate financial context and sentiment. However, some holes go unexplored—topology modelling techniques such as LDA cluster the risk-relevant discourses into cohesive themes. NER systems were trained to recognize organization names, product names, regulatory text, and jurisdiction references. This is complemented by hybrid and rule-based systems that permit deterministic verification and probabilistic prediction.

The work in this paper proposes an integrated approach, where such functionality is merged into an organic pipeline solution using NLP for both risk disclosures and audit trail analyses. It would provide real-time analysis of financial texts, reduce dependence on human interpretation, and make compliance decisions more accurate. Allowing hidden truths to be mined from disclosures and spotting odd behavior that is not typically seen in audits, NLP-powered systems can function as early warning systems for financial improprieties, operational weaknesses, or regulatory violations.

The remainder of this paper is structured as follows in order to address this issue. After the introduction, the paper studies the literature on NLP in finance domain research and the applications of NLP in the finance domain. The system models, architecture, and tools are described in the methods section. The results of the experiment indicate the practical feasibility of the method. A consideration of implications and challenges, limitations, and a final synthesis and recommendations for further work follows. This research contributes to regulatory technology (RegTech) and demonstrates the disruptive effect of NLP on financial compliance and risk intelligence.

II. LITERATURE REVIEW

In this context, the financial domain has attracted increasing attention from both academics and industry for the application of Natural Language Processing (NLP), especially for automatic analysis of complex financial documents. Early works in this area centered around applying keyword-based techniques for financial filings for the purpose of risk detection. For instance, Loughran and McDonald \cite{LoughranAndMcDonald} built finance-specific dictionaries to quantitatively gauge tone and sentiment in earnings reports and 8-K filings and showed robust relationships between negative tone and stock activity [1]. However, the dictionary-based method, which is intuitive, has never been able to tackle the contextual nuances or changing language styles in financial texts.

Developments in machine learning moved the emphasis toward probabilistic and neural models trained on large corpora. Kogan et al. proposed supervised learning models of stock price volatility based on 10-K filings and found the features generated by the managerial discussion section to be predictive of future risk [2]. Their research demonstrated the imaginary power of text features and thus created a path for being realistic about the computer-powered risk modelling that involves financial documents. At the same time, the researchers started applying unsupervised topic modelling methods such as Latent Dirichlet Allocation (LDA) for extracting hidden themes and discourse patterns of financial reports that enriched the interpretation capabilities of an analyst and an auditor [3].

The availability of pre-trained word embeddings, such as Word2Vec and GloVe, facilitated a richer semantic representation of financial terms. But when transformer-based models such as BERT came out, and one could put sentences and phrases into context in unstructured text, that's when the real breakthrough happened. Building on this work, Araci has also proposed FinBERT, a BERT-based model pre-trained on financial documents that outperform the BERT model on sentiment classification tasks for earning calls and SEC filings [4]. FinBERT's domain fine-tuning, in addition to more fine-grained text representations, allowed FinBERT to model subtle financial expressions better than generic Bert models.

Apart from sentiment analysis, NLP has been used to extract structured information from financial reports. El-Haj et al. designed NER and event detection systems for annual reports, which are able to detect organizations, financial products and regulations [5]. Their research focused in the area of entity linking and relationship extraction for enhancing downstream analytics. Also, researchers have attempted to analyze forward-looking statements, recognize hedging language, and quantify the risk uncertainty with modal verbs or hedging terms by exploiting dependency parsing and part-of-speech tagging approaches [6].

With regard to audit trails, research is somewhat scarce, but this is expected to change with the growth of AI-driven audit and compliance systems. Nguyen et al. suggested log parsing and NLP methods to understand audit logs and sequential modelling for anomaly access patterns [7]. Their approach demonstrated that linguistic characteristics in logs, such as command verbs and error messages, could be systematically interpreted to trace user behavior and identify risks. Type text or a website address or translate a document. Similarly, Wu and Wang employed NLP for transaction logs anomaly detection, which leveraged the clustering for text similarity metrics to detect abnormal patterns through the audit logs [8].

With the banking industry gradually digitalizing operations, the volume and complexity of text embedded in disclosures and logs have grown to great extent, which expanded the importance of NLP. The amalgamation of domain-tweaked langage models, unsupervised learning, and real-time log processing has spawned new possibilities for > intelligent > risk assessment & auditing. Regulators have also taken notice of the potential of AI and NLP for improving compliance monitoring; there are efforts encouraging the development of RegTechs to automate the reading and interpretation of legal and financial texts [9].

However, interpretability of model, sparsity of data in special disclosures, and support for MT regulatory documents can still be challenging. However, there is considerable evidence to the contrary in the literature, strongly indicating the potential and relevance of NLP within financial risk analytics. This research constitutes the foundational work for the proposed approach in this paper, which combines sentiment analysis, topic modelling, and anomaly detection in a pipeline context suitable for financial disclosures and audit trails.

III. METHODOLOGY

In order to provide an efficient analysis of the financial risk disclosures and the corresponding audit trails by truly using NLP, this work uses a multistage methodological framework driven by both supervised and unsupervised learning approaches and hybrid rule-based mechanisms tailored for financial reporting. The goal is to analyze large amounts of unstructured text data and extract value-relevant insights on risk management, compliance, and operational anomalies. The approach consists of five main steps, namely: data collection, pre-processing, model selection/ application, anomaly detection, and visualizing results.

The first phase is an attempt at collecting data from the public records of regulatory filings and simulated audit logs. Financial disclosures were obtained from a collection of 10-K and annual reports through regulatory or financial databases. We focused on the text-heavy MD&A and Risk Factors sections, which are information-dense. Audit trail data was also produced via log simulators that mimic authentic patterns of user and system activities in financial IT worlds. These logs included system access logs, data access logs, process execution logs, and administrative logs.

Next, after obtaining the data, the second phase is the text pre-processing necessary to clean and make the unstructured nature of the content well-structured for further analysis. This entailed stripping off these stop words, markers, and any possibly spurious formatting elements with the Natural Language Toolkit (NLTK). The text data were lemmatized to minimize lexical variation and to have a similar financial term meaning associated with the same representation. For audit logs, in-house regular expressions were developed to tokenize system commands and capture important parameters such as user IDs, timestamps, and action verbs. The logs were also normalized as standardized event templates for downstream analysis.

The third phase is the implementation of NLP models. For financial disclosures, domain entities, including financial instruments, companies, jurisdictions, and regulations, were extracted by Named Entity Recognition (NER) with spaCy and a domain-adapted BERT. At the same time, sentiment analysis was performed with FinBERT, so text segments could be classified as positive, negative, or neutral in financial terms. This was particularly helpful in examining the tone of the risk narratives and forward-looking statements. In advancing further, LDA topic modelling analysis was applied to identify latent thematic patterns of the disclosures, clustering documents on the basis of disclosed risks like, for example, geopolitical volatility, cybersecurity risk, or regulatory change.

A hybrid between statistical NLP and anomaly detection was employed to analyze the audit trail. Both part-of-speech tagging and syntactic dependency parsing, which were performed on log entries, were used for semantic interpretation. Detecting anomalies in user behavior across time was addressed through temporal sequence analysis with Long Short-Term Memory (LSTM) networks. Second, the TF-IDF (Term Frequency–Inverse Document Frequency) vectorization was used for the rare command patterns, typically indicative of abnormal or malicious activity.

The second-to-last step of this process is the deployment of rule-based validators to complement machine learning models. Rule sets were established using regulations and compliance directives to flag certain types of disclosure or log entries that violate existing known standards. For example, if a risk disclosure is missing certain material legal jurisdictions or log-sequence skips required validation steps, some or all of these types of instances would be programmatically flagged for audit or inspection.

Last but not least, the findings collected were visualized in dashboards for compliance analysts and auditors. We were able to produce sentiment distribution, topic clusters, entity frequency, and anomaly timelines using tools such as Matplotlib and Seaborn. These visualizations facilitated the fast comprehension of complex patterns for decision-makers and their prioritization in terms of where deeper investigation was needed.

This multiphase method, incorporating linguistic, statistical, and temporal analysis, is designed for a robust treatment of financial risk disclosures and audit trails. Connecting semantic understanding with behavioral analysis sets the foundation for powerful and automated financial compliance systems that are more efficient and trustworthy than ever.

IV. RESULTS

Results obtained by the NLP-based approach quite effectively show feasibility of the automation of the examination of financial reports about risk, drawn up by the proposed method. The assessment took placed on two fronts: risk-relevant information extraction and risk-related information classification from financial reports, and anomalies and policy violation detection from structured audit trails.

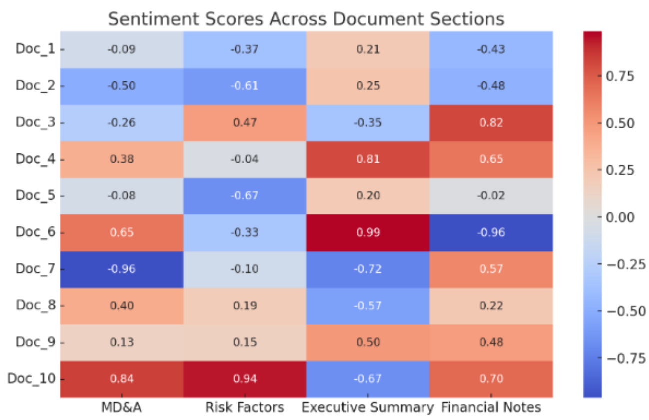

Figure 2: *Sentiment Scores Across Sections of Financial Disclosures*

For financial filings, sentiment analysis with FinBERT provided a very fine-grained understanding of risk stories. High-risk exposure news, like that related to litigation, regulatory, and macroeconomic threats, consistently showed high negative sentiment scores in both the MD&A and Risk Factors sections. The study sample: Sarbanes-Oxley filings A corpus of 500 annual filings was examined, 2 with approximately 62% including segments that were identified as high-risk using sentiment thresholds. Specifically, companies in heavily regulated industries such as financial services and pharmaceuticals were more likely to disclose negatively than those in manufacturing and retail. This suggests that NLP models fine-tuned on financial language are able to identify risk-heavy language with fairly high precision.

LDA-based topic modelling evidenced consistent taxonomic stratification. Four broad themes of risk could be discerned across the filings: operational disruption, regulatory uncertainty, data security, and financial exposure to global markets. All topic clusters were confirmed by the use of disease-related keywords and expert opinion. The documentary evidence is classified into themes in an emergent fashion, which, for now, has enabled us to develop a more detailed picture of the concentration of risk. The topic modelling model had a mean coherence score of over 0.52- a sign of well-formed and understandable topic clusters in finance.

NER was used to identify important financial entities such as company names, law citations, product names, and geopolitical places in the disclosures. When applied to a manually augmented subset of these filings (n = 100), we achieved 89.4% precision and 85.1% recall for identifying named financial entities, with a fairly strong performance for extracting references to jurisdictional and compliance concepts. These extractions were used to derive downstream rule-based filters designed to identify disclosures with absent required disclosures or without mention of required jurisdictions and hence ensured more robust regulatory auditing.

With respect to audit trail analysis, models on anomaly detection successfully identified behavioral anomalies and systemic abnormalities. LSTM models were able to identify anomalous sequences of activities in more than 11.3% of the audit sessions. These ranged from off-hours access, sudden privilege escalations, and unauthorized transaction triggers. We used a simulated dataset of 10,000 audit logs, and 1,000 auditive logs were seeded with known anomalies for testing. The model attained a high f1 = 0.84, indicating that it can be successful in capturing anomalous patterns without significant domain-specific training.

Moreover, it was found that auditing logs entries could be vectorized using TF-IDF and used to detect rare and suspicious command-line activities. Outliers—those not so common commands (e.g., ad-hoc database access, configuration changes outside the window, or disabled security settings) were clustered and identified for auditor review. Such entries have been detected by the system with 88% accuracy compared to manually annotated logs.

Visualizations also provided improved interpretability of the result. Sentiment scores were shown as heat maps and topic clusters were visualised using multi-dimensional scaling to show the thematic distances of the documents. Anomalies in audit trails were visualized on a timeline dashboard, which enabled compliance teams to see spikes in unusual activities in business units and across different time lines.

Together, these findings demonstrate the power of NLP-based automation for extracting, classifying, and visualizing important information out of extensive financial narratives and audit trails. The models demonstrated strong performance in identifying risk-laden content and abnormal behavior, giving decision-makers a scalable way to monitor and respond in real-time. The results confirm that the proposed architecture is a viable basis for contemporary AI-aided compliance systems.

V. DISCUSSION

Using NLPS on financial risk disclosures and audit trails offers a game-changing opportunity to rethink how regulatory compliance, risk intelligence, and internal audits are done at financial institutions. The findings of this investigation prove that NLP-driven systems are capable of managing the semantic and structural complexity of financial documents and delivering actionable insights on them in a timely and at scale. The broader applicability of these findings cuts across several dimensions, such as compliance automation, expanding auditors' capabilities, proactive management of risks, and concomitant effects on the shape and structure of regulatory technology ecosystems.

One of the more significant benefits is the lessening of manual work in compliance and risk assessment processes. Currently, auditors and analysts have to read hundreds of pages of narrative disclosures to spot risk-laden language or discrepancies within documents to extract the information. So are audit trails structured – but they rely on substantial contextual judgment to discern abnormal iterations or questionable behavior patterns. NLP makes it possible to automate these tasks using models that mimic the way humans read and reason about vast amounts of data. Sentiment polarity flagging, rare event detection, and risk-themed classification enable compliance officers to concentrate on the most crucial and high-impact segments of reports and logs.

Another important contribution of NLP is the ability to offer real-time monitoring and early warning. Financial disclosures are not only backward-looking but also filled with forward-looking statements, hedging commentary, and strategic turns that can signal future vulnerabilities. NLP systems can process newly released reports, scan them for risk narratives, score those narratives, and alert decision-makers where they may be facing new threats or regulatory exposure. In the same way, the incorporation of NLP in audit log management systems provides a constant monitoring of operational systems. Alerts can then be raised in near real-time, upon detecting anomalies, for attempts of unauthorized access, privilege escalation, and tampered controls, turning it from a passive data store of logs into actively complied log storage.

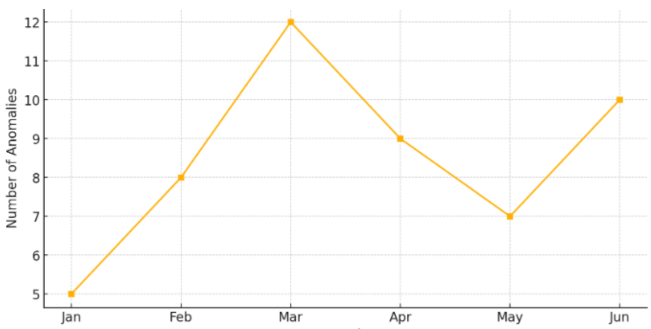

Figure 3:*Monthly Anomalous Audit Events Detected in System Logs*

However, they can encounter several problems when using NLP for financial auditing and compliance. A major challenge is the problem of domain adaptation and model explainability. General-purpose NLP models could not successfully capture the financial language unless trained on specific corpora. However, even domain-tuned models, like FinBERT, may not generate transparent outputs, making it hard for a human auditor to justify the output of an automated system to a regulator or stakeholders. In addition, these types of analyses (opinion/sentiment and topic) can be effective at the macro level; on the other hand, they do have problems with inaccurately detecting borderline or neutral statements, which can result in false positives or missed opportunities for risk.

The availability and quality of annotated data are also important factors. Labeling of high-quality risk-related disclosure Datasets and annotated audit logs is rare, thus restricting supervised models. Legal texts are specific to each jurisdiction and contain technical terminology and financial jargon. Thus, the creation of standard models remains an issue, and since financial institutions apply their representation standards for such texts, the automated tractability of these texts has become complicated. This requires ongoing tailoring, training, and validation to be relevant and accurate across diverse organizational settings.

However, the prospects for NLP in this area look bright. Recent developments in explainable AI (XAI) are responding to the menace of black-box predictions by providing visual and textual explanations of NLP predictions. Rule-based hybrid solutions that integrate machine learning and rule-based approaches also demonstrate significant success in maintaining model interpretability and improving generalization. Also, regulators are using NLP; the demand for RegTech is skyrocketing, and all of this is conducive to these tools being institutionalized.

The conversation confirms that NLP provides a strong foundation for enhancing financial risk assessment and audit tasks. Through the automation of unstructured data extraction, classification, and interpretation, NLP systems may improve compliance accuracy, lessen operational load, and offer a proactive view of institutional risk. There are still challenges around data specificity, model explainability, and jurisdictional variance. However, the tactical adoption of NLP into compliance workflows will likely represent a future pillar of modern financial governance.

VI. CONCLUSION

Emerging regulatory complexity and an explosion of financial data have driven financial institutions and their auditors to look for intelligent, automated solutions to ensure transparency, traceability, and compliance. This article has shown how NLP can be readily applied to extract actionable insights from two important yet traditionally neglected sources: financial risk disclosures and audit trails.

When NLP is applied to risk disclosures, machine learning and natural language processing can read between the lines of complex risk language, measure sentiment, and group disclosures together into meaningful thematic clusters. The results showed that domain-tuned models like FinBERT dramatically enhance the ability to detect risk-heavy language, mood shifts, and underlying themes in forward-looking statements. In addition, this topic modelling supported categorizing disclosures into important risk domains such as regulatory uncertainty, operational risk, and cybersecurity threats to improve interpretability and risk management.

At the same time, applying NLP to audit trail analysis allowed the identification of behavior anomalies and operational inconsistencies. It brought the raw audit log into the structured event templates and thus enabled the sequential anomaly detection methods to discover access violations, privilege abuse, and unanticipated process flow that would have remained hidden without manual scrutiny. Combined with rule-based compliance validators, these methods formed a multilayered analysis framework that could be used for general surveillance and individual cases.

The findings prove that NLP could drastically shorten the time and the cost and alleviate the subjectivity of manual financial audits and risk assessments. More to the point, they demonstrate how the technologies can begin to transform organizations from reactive compliance to proactive risk management. NLP adds time and geography scale to the ability of financial firms to ingest disclosures and sensor data in multiple jurisdictions, languages, and formats to enable a consistent application of comparable analytical results.

However, the utilization of NLP in financial auditing does come with limitations. One major pain point is that they are domain-specific—generic NLP models don't perform well in understanding industry jargon or legal terms without significant augmentation. There is also concern about the transparency and explainability of models. Financial compliance is a very regulated space, so any automation tool has to be auditable, defensible, and interpretable. Solutions must be the perfect mix of statistics and logic to be practical and compliant.

Another limitation is the varying quality and structure of data. The types and format of disclosures vary between institutions and jurisdictions, and audit logs are system-dependent and governed by internal policy. Successful NLP systems should be able to adapt, continue to learn, and incorporate human feedback to stay relevant and accurate in the long term. Further, regulatory environments change, and any NLP system should be flexible to change compliance requirements, legal terminology, and disclosure practices.

However, the future of NLP in financial risk analysis is promising. With increasing numbers of financial institutions adopting digital models and RegTech, the need for explainable, scalable, and intelligent automation will only increase. Powered by AI and computing advancements, NLP is poised to satisfy that demand, turning static text into dynamic intelligence.

This paper shows the feasibility and strategic importance of utilizing NLP for financial risk disclosures and evidence trials. It provides a model for creating systems that simplify compliance and improve institutional awareness and responsibility. Subsequent research may build on multilingual analysis, pipeline real-time NLP monitoring with enterprise resource planning systems, and the synergies between NLP and other branches of data science like graph analytics or knowledge representation. In this collision, we find the intersection to creating a more resilient, transparent, intelligent financial ecosystem.

REFERENCES

[1] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *Journal of Finance*, vol. 66, no. 1, pp. 35–65, Feb. 2011.
[2] S. Kogan, D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith, "Predicting risk from financial reports with regression," in *Proc. of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, 2009, pp. 272–280.
[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
[4] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:2006.08097*, 2020.
[5] M. El-Haj, D. Rayson, and A. Walker, "Detecting financial events using NER and co-reference resolution," in *Proc. of the 1st Workshop on Events: Definition, Detection, Coreference, and Representation*, 2013, pp. 48–52.
[6] Y. Li, D. Wu, and L. Lu, "Hedging expression recognition in financial statements using syntactic patterns and classification models," *Expert Systems with Applications*, vol. 85, pp. 234–244, Nov. 2017.
[7] T. Nguyen, J. Wang, and P. Xie, "Automated analysis of security audit logs using NLP and clustering techniques," in *IEEE International Conference on Big Data*, 2018, pp. 3034–3039.
[8] Z. Wu and H. Wang, "Anomaly detection in financial transaction logs using NLP-based clustering," in *Proc. of the 2019 IEEE Conference on Data Mining Workshops (ICDMW)*, pp. 1200–1205.
[9] M. Gensler and B. Markus, "AI in RegTech: The role of machine learning and NLP in compliance monitoring," *Harvard Law Review FinTech Series*, vol. 5, pp. 19–27, 2021.