10.48047/jocaaa.2026.35.09.01

AI-Powered Due Diligence in Private Equity: Automated Analysis of Financial, Legal, and Business Intelligence Data Using NLP and Deep Learning

Vrajkumar Patel
MBA student
Vrajpatel5471@gmail.com
0009000121620319

Abstract

During enterprise due diligence practices, private equity firms are turning to mass volumes of financial and business intelligence data to assess acquisition opportunities and pinpoint enterprise risks. However, traditional due diligence methods are still primarily manual, are time-consuming, have inconsistent data interpretation and lack scale when processing unstructured finance data. Based on these elements, the following research proposes a due diligence framework using financial news intelligence and automated investment risk evaluation for AI, which can assign risk to investments by using NLP and DL technologies. It integrates extensive preprocessing techniques, contextual feature engineering, processing to extract information with FinBERT, and a hybrid BiLSTM classification architecture for enterprise-risk prediction, employing the Kaggle High-Quality Financial News Dataset. Various machine learning and deep learning models, such as Logistic Regression, Random Forest, BiLSTM, FinBERT and the proposed hybrid framework, were tested. The outcomes of the proposed model show that the proposed model yielded superior classification ability with accuracy 96.1%, precision 95.7%, recall 95.1% and F1-score 95.4%, better than all the baseline models. Moreover, a maximum ROC-AUC of 0.97 meant that the proposed architecture demonstrated strong discrimination ability to discern investment-risk categories. A ranking analysis for the explainability of its predictions showed that legal factors like lawsuits, the negative growth of revenue, judicial inquiries, and volatility in the stock market had the most significant influence in the determination of high-risk investment predictions. The findings demonstrate that transformer-enhanced AI systems can significantly improve private equity due diligence efficiency, scalability, and contextual enterprise intelligence extraction. The proposed framework provides a robust and interpretable solution for automated investment analytics and intelligent acquisition-risk evaluation.

Keywords: Private Equity Due Diligence; Financial Intelligence; Natural Language Processing; FinBERT; Deep Learning; Explainable Artificial Intelligence; Investment Risk Prediction; Business Intelligence Analytics; Transformer Models; Automated Enterprise Intelligence

1 Introduction

Over the past decade, the private equity industry has been accelerating sharply, especially as investors have turned to alternative methods of generating returns greater than what can be provided by the traditional public markets. Private equity firms consider businesses from a variety of industries, and they are looking to acquire businesses that would be profitable for them to acquire, restructure underperforming businesses, and maximise the long-term value of the business (Harrigan & Wing, 2021). This investment process incorporates due diligence, which is key to understanding the financial robustness, efficiency, legal adherence, and strategic viability of acquisition targets prior to committing investment capital. Due diligence can significantly impact the success of an acquisition process, the ability to manage risks and the success of the investments made after acquisition.

The 21st Century due diligence process includes analysing vast collections of disparate data like financials, regulatory disclosures, market reports, news articles, contracts and operational disclosures, etc. Financial intelligence analysis aims to study the trends of potential profitability, liquidity risk and debt exposure, as well as the market positioning, while legal intelligence is more about dealing with litigation risk, regulatory issues, and contractual obligations (Chmimo, 2025). Business intelligence also helps in interpreting the reputation of markets, the competitive positioning within the markets, the sentiment of its customers and the areas for organisational development. But acquisitions are often accompanied by uncertainty about investment since the investors have to assess incomplete and volatile information promptly. The complexity of the regulatory framework, the requirements of compliance with different jurisdictions and the constant evolution of financial markets compound these issues, aggravating the challenge of identifying hidden risks and the elements that make for a good investment (Biktimirov & Dore, 2025).

10.48047/jocaaa.2026.35.09.01

Traditional due diligence processes continue to rely on manual analysis by financial specialists, legal counselors and investment analysts (Sharma et al., 2024). Investors need to be able to sift through huge amounts of financial statements, corporate reports, contracts and business news articles to find investment opportunities and potential areas of concern. The process can be very time-consuming and large, especially in acquisition transactions involving thousands of pages of documents which need to be reviewed under time constraints. As the quantity of information in digital businesses continues to grow, document overload has become a major driver in promoting document information inefficiency and inconsistency in manual evaluation practices (Орлов, 2024).

One of the key drawbacks of the traditional process of due diligence is the lack of interpretive consistency when it comes to analysts. Different analysts may reach conflicting conclusions when evaluating identical information because human judgment is influenced by experience, expertise, and subjective reasoning. Additionally, traditional due diligence principles are ineffective in meeting the expectations of scalability when undertaking a continuous flow of financial reporting, news and market events for an enterprise (Zambelli, 2025). Much of the data created is unstructured and complex because of the widespread growth of text produced on business sites, Web media, and corporate communications, and this trend is only increasing. Thus, traditional manual procedures are not capable of delivering timely, scalable, and complete investment intelligence.

AI has become a game-changer in the financial industry, with its many applications being fraud detection, investment analysis, algorithmic trading, and intelligent risk assessment. Since much enterprise intelligence is stored in unstructured texts, Natural Language Processing has gained significant importance to extract meaningful insights from financial and business documents in the last few years (Dash, 2022). By leveraging NLP techniques, financial intelligence systems can streamline data sentiment analysis, entity recognition, risk extraction, and document classification, enhancing their efficiency and precision. At the same time, intelligent automation through deep learning architectures has helped companies handle their ever-increasing amounts of data more effectively than standard analytical techniques.

Transformer-based models like BERT and FinBERT have been shown to be more successful than other classic machine learning algorithms as they are better able to understand the relationship between word meaning, phrases, and sentences (Schöne, 2024). Typical ML-based models have manually designed features and lack semantic context, whereas the Transformer architecture implements attention mechanisms that enhance semantic understanding and prediction power. AI-powered analytics is, therefore, a major opportunity in the field of private equity automated due diligence processes.

While the finance industry has advanced with AI technologies, the prevailing D&O frameworks are still facing major challenges in areas of automation and explainability, integrating with intelligence analysis. An increasing number of solutions are available that use only specific financial data without taking into account full business intelligence and/or textual context analysis (Phalippou, 2023). Moreover, the absence of explainable AI mechanisms decreases transparency in investment decision-making, decreasing trust and interpretability in the enterprise risk assessment mechanism.

The main aim of this study is to create an AI-driven due diligence framework that automatically extracts financial and business intelligence signals using methods based on NLP and deep learning for investment risk assessment processes in private equity. The goals of the research are: Building an NLP-based system to extract intelligence from financial text, building deep learning-based classification models for investment risk assessment, building an automated investment risk prediction based on an enterprise intelligence system, and testing the explainability and predictive power of the proposed AI framework on financial text.

This study presents an auto-driven due diligence architecture that combines financial intelligence analysis with the transformer-based NLP methods. The research also works out to come up with a deep learning investment scoring framework, which can be used to detect the risks of acquisition and potential business opportunities available in the unstructured text. In addition, the study also includes explainable AI mechanisms to enhance the interpretability and transparency of automated investment decision support systems.

The remainder of this paper will be organised in the following manner. Section 3 details the related work with financial intelligence systems and automated due diligence systems powered by AI. The proposed methodology, data, NLP framework and deep learning models are described in Section 4. Experimental Results and the performance evaluation are considered in Section 5, and the implications and potential limitations of the study are discussed in Section 6. Finally, Section 7 summarises the paper and provides suggestions for future studies.

10.48047/jocaaa.2026.35.09.01

2 Related Work

2.1 AI in Financial Technology

Artificial Intelligence is playing a pivotal role in financial technology, transforming the industry with the power to automate analytical tasks, enhance predictive capabilities, and facilitate smart decision-making. The rise of artificial intelligence has left its mark on the banking industry, with AI systems being used widely in various processes, including fraud prevention, customer analytics, and investment strategies (Ashta & Herrmann, 2021). Anomaly detection and behavioural analysis are common approaches used to detect fraud using machine learning algorithms, helping organisations minimise financial losses and enhance transaction security. AI investment analytics tools also help in portfolio optimisation, market prediction, and automated trading strategies by analysing vast financial and market data. Portfolio Intelligence Platforms also leverage predictive analytics to evaluate asset performance, volatility trends, and investment opportunities within the evolving financial landscape (Odeyemi et al., 2024).

In addition, AI advancements have improved financial risk assessment for businesses by facilitating intelligent automation and staying current with market changes in real-time. The use of deep learning and NLP systems in processing financial statements, business announcements, and market sentiment is now becoming possible (Khalil & Pipa, 2022). Though these progressions, on the other hand, most financial AI systems currently available are tailored for banking analytics, stock forecasting or fraud surveillance, not private equity due diligence. The majority of the solutions in the book are dedicated solely to financial metrics and ignore the legal, business intelligence, and enterprise analysis aspects of an investment. As a result, the capabilities of current AI applications to aid in the comprehensive due diligence process in private equity are still constrained.

2.2 NLP in Financial Intelligence

Financial markets create a lot of unstructured textual data from business disclosures, regulatory filing reports, earnings reports and news articles, making Natural Language Processing a major field of research in Financial Intelligence (Duane et al., 2025). Sentiment analysis methods are commonly used to analyse market sentiments, investor confidence, and the sentiment of companies by categorising text into positive, neutral, negative and more. Another notable application of financial text mining is the ability to gain insights from company documents and financial statements, which can help companies identify trends, opportunities, and potential investment risks. In recent times, NLP-based sentiment detection systems have proven to be valuable tools in enhancing the precision of predictive models and analysis for financial forecasting and investment strategies (Razali et al., 2025).

The effectiveness of financial NLP applications has significantly improved with transformer-based language models. A financial domain-specific tuning of BERT for financial language understanding, FinBERT, is highly effective on financial sentiment classification and on financial interpretation of context-dependent risk (Nasiopoulos et al., 2025). Business news streams have also been used to extract certain events such as mergers, acquisitions, bankruptcies, litigation events and regulatory announcements by the use of event extraction techniques. These architectures are better than conventional machine learning methods for exploiting semantic and contextual dependencies through the attention mechanism in the transformer architecture (Mowbray, 2025). Unlike traditional bag-of-words or statistical approaches, transformers correctly extract and classify information considering the information present in the text context. Despite these developments, most of the studies on NLP are limited to sentiment analysis or to specific financial aspects of transactions and do not focus on creating complete due diligence systems that integrate financial, legal and business intelligence for the private equity investment process analysis.

2.3 Deep Learning in Investment Analytics

In the field of investment analytics, deep learning has had a remarkable impact, providing capabilities that surpass those of traditional statistical analysis. Although repeatedly tested and applied in the financial forecasting area, the effectiveness of recurrent neural networks like Long Short-Term Memory models in capturing sequential analysis and temporal dependence in market data makes them suitable models for various forecasting tasks (Pattanayak et al., 2024). BiLSTM architectures then take contextual learning a step further, using not just forward-time information, but also backwards-time information, which yields better prediction accuracy in financial trend analysis and financial sentiment classification. Financial AI models have also become more central to the attention mechanism, allowing them to focus on the most relevant information within a vast pool of textual data (Khalil, 2026).

10.48047/jocaaa.2026.35.09.01

Another great development in investment analytics is the use of transformer architectures, which excel in their context representation. Transformers are more efficient than traditional deep learning models at handling long text sequences, and they retain the semantic meaning from multiple contexts. As such, transformer models are proven to be effective in predictive finance, market forecasting, and enterprise intelligence extraction (Chen et al., 2025). While some deep learning applications in finance have been developed to forecast the price of stocks and study the movements on the stock exchanges, little attention has been paid to automating the process and undertaking private equity investment environments.

2.4 Automated Due Diligence Systems

Due diligence systems are proving to be more and more popular as companies look to cut costs and complexity from manual enterprise evaluation systems. Modern enterprise solutions integrate compliance modules powered by AI, document management systems, and legal analytics tools to enhance the efficiency of compliance processes within the scope of risk assessment procedures (Yarram & Parimi, 2021). Using text analytics and NLP-based methods, legal-tech programs can automate contract review for clauses, regulatory compliance and analysis. Likewise, AI-driven compliance solutions can help companies identify irregularities in transactions, track governance standards and identify legal threats in company records.

While all these advances have been made, there are also shortcomings in the system of due diligence in place. Most enterprise platforms are geared towards understanding a single element of an operation and are unable to offer a single intelligence analysis centre for financial, legal and business sectors (Lerner & Leamon, 2023). Additionally, many systems are strongly dependent on rule-based automation that barely has any real context knowledge and has very little predictive ability. Without multi-source intelligence fusion, automated investment evaluation effectiveness will suffer tremendously, as key information is likely to be scattered in various unstructured enterprise data resources (Song, 2025). There is a great demand for unified platforms that enable an AI-driven approach to due diligence platforms that integrate deep learning, NLP, and enterprise intelligence analytics into a single environment supporting business decisions.

2.5 Explainable AI in Financial Decision-Making

The importance of Explainable Artificial Intelligence (XAI) in financial decision-making is growing as transparency, accountability, and regulatory adherence are crucial factors in AI-based systems. Financial risk and regulatory controls often make investment decisions difficult, and demand models that are easy to make sense of for financial institutions and investment organisations. When investment decisions are not sufficiently explainable to stakeholders and regulators, the use of black-box AI can generate issues and concerns of fairness, the absence of trust and accountability (Oyasiji et al., 2023).

SHAP has become one of the most popular explainability approaches due to its ability to measure the impact of each feature on the model's decisions. Interpretable AI frameworks help analysts to interpret the logic of classification, which enhances the believability of automated financial institutions. Explainable AI is also crucial in addressing regulatory issues concerning algorithm transparencies and ethical use of AI in enterprise settings (Bamigbade et al., 2024). But numerous financial AI studies only focus on forecasts and prescriptive ability, leaving little transparency behind in automated investment analytics.

2.6 Research Gap Analysis

Existing studies have generated valuable insights into financial AI, NLP-driven analysis, and enterprise automation; the private equity landscape presents key gaps in the development of AI tools for deal due diligence. Most of the previous work is related to specific discrete analysis functions, including sentiment analysis, fraud detection or financial forecasting, but does not combine all enterprise intelligence components in a single engine. Furthermore, in many cases, there are no explainability mechanisms in place, making investment decision-making processes less transparent and easier to understand.

Study	Technique	Limitation
Study (Köse, 2025)	Sentiment analysis	No PE focus
Study (Rohan et al., 2026)	Financial prediction	No explainability

10.48047/jocaaa.2026.35.09.01

Study (Afane et al., 2026)	Legal analytics	Limited automation

Moreover, enterprise solutions cannot often integrate financial, business and context-related contextual, NLP-based analytics in a scalable deep learning architecture. There has been no existing study that presents a single "due diligence" framework with Financial Intelligence, NLP, deep learning and Explainable Investment Analytics for evaluating private equity investments. The study has thus identified at a salient level an integrated AI-based due diligence framework that can handle the automation of the enterprise intelligence analysis, with enhanced predictability and explainability.

3 Methodology

3.1 Proposed Framework Overview

Another methodology proposed is based on an AI due diligence framework to automate the analysis of financial/business intelligence data, using Natural Language Processing and deep learning techniques. The architecture is designed to combine multiple analytical stages into a single pipeline that is capable of ingesting unstructured financial news, textual enterprise data, and analysing the information for investors' risk assessment purposes in private equity settings. The whole workflow entails data acquisition and preprocessing, feature engineering, and the transformer-based NLP analysis, which extracts financial intelligence in the context. Then, deep learning classification models are used to uncover investment risks, business opportunities, and signals in the market, all derived from the sentiment. Last but not least, mechanisms under explainable AI are introduced, which enhance transparency and understandability in automated investment recommendations.

The objective of the proposed framework is to tackle the drawback of scalability encountered in existing due diligence systems by extracting the meaningful knowledge from a huge amount of business-oriented data streams. Transformer-style architectures can pair better with sequential deep learning models, where they are able to model enterprise-level intelligence patterns, temporal relationships, and contextual semantics more effectively than traditional machine learning methods, as shown in Figure 1.

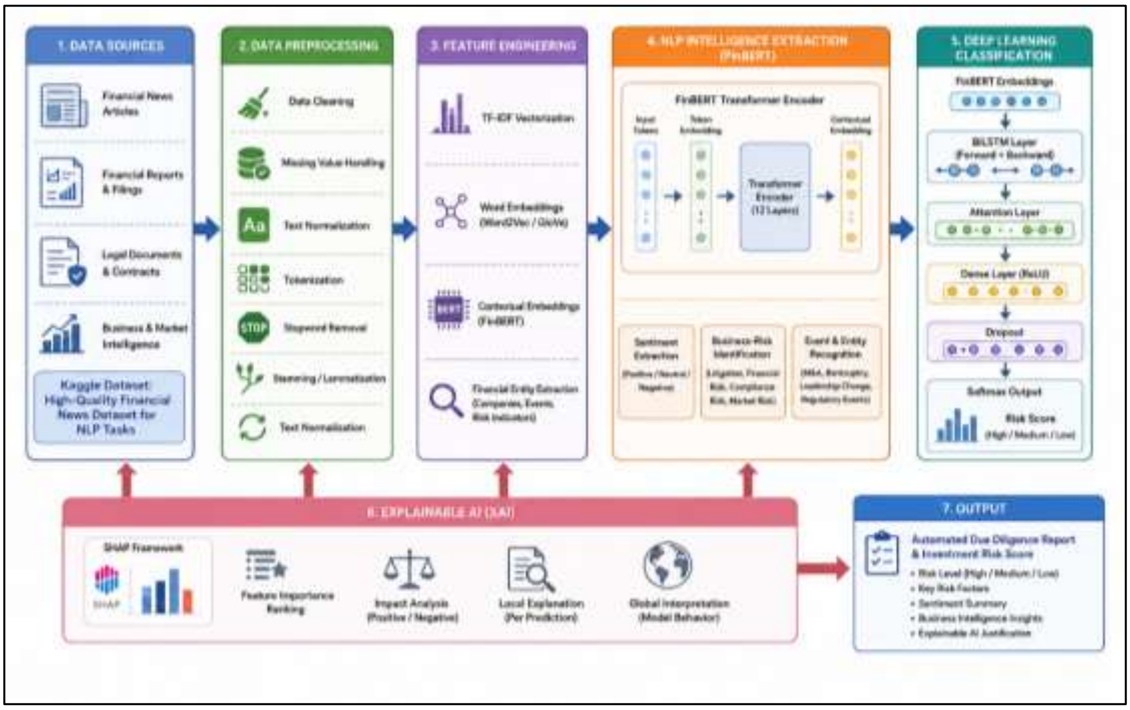

Figure 1. Proposed System Architecture

3.2 Dataset Description

The dataset is High-Quality Financial News Dataset for NLP Tasks extracted from Kaggle, which has a vast amount of financial news articles related to business activities, market events, investment trends, and enterprise-

10.48047/jocaaa.2026.35.09.01

related activities. The data comprises thousands of textual data on financial reporting platforms and reporting business news. These records include news headlines, descriptions of the articles, data about the publication, data related to the sentiment of articles and other information which can be used to support investment intelligence analysis and automation of due diligence processes.

For this study, the selected data was found to be well-suited, as business intelligence from the financial reports and financial news, as well as corporate reports, forms a fundamental part of the due diligence process for private equity investments. Raw financial news data, however, contains a significant amount of noise, redundancies, inconsistencies and irrelevant symbols which can have a negative impact on model performance. This makes it necessary to carry out a fair amount of pre-processing before textual data can be used for deep learning-based classification and/or transformer-based NLP analysis. Finally, during the downstream phase, analytical tasks are conducted with increased semantic consistency, reduced dimensionality and higher predictive accuracy thanks to proper preprocessing.

3.3 Data Preprocessing

Data preprocessing is a crucial phase in the proposed AI-enabled due diligence framework, as the financial news data generally consists of noisy, incomplete, and unstructured textual information. Duplicate data was detected and eliminated to improve the consistency of the data set and reliability of the data analyses. Missing data was detected and corrected. All text data was then normalised using text normalisation techniques to ensure that all text in all records is in lower case, no punctuation symbols are present, no numerical noise in the data, and text is standardised across all records.

Additionally, tokenisation was applied to break down the textual data into smaller linguistic segments that were better suited for processing using NLP models, as part of the processing pipeline. To reduce the computational complexity of the problem, the frequently occurring words, which do not contribute much to the semantic meaning of the text, like conjunctions or articles, were set aside using the stopword removal. Stemming and lemmatisation methods were also used to convert words into their root form and promote consistency in the meaning of words in the textual sequences. During experimentation, the process of lemmatisation was chosen due to its better preservation of the contextual meaning of a text in comparison with vigorous stemming methods.

Finally, to further enhance model robustness, textual cleaning was performed to strip away hyperlinks, special characters, inconsistent positioning of space characters and irrelevant symbols which are frequently found in financial news articles. Finally, we normalised the text, which was used as an input for the tokenisation and apparatus of deep learning algorithms. The results of such pre-processing operations greatly benefit data quality and eliminate linguistic ambiguity, which ultimately helped improve the outcomes of downstream sentiment analysis and investment risk prediction procedures (Figure 2).

10.48047/jocaaa.2026.35.09.01

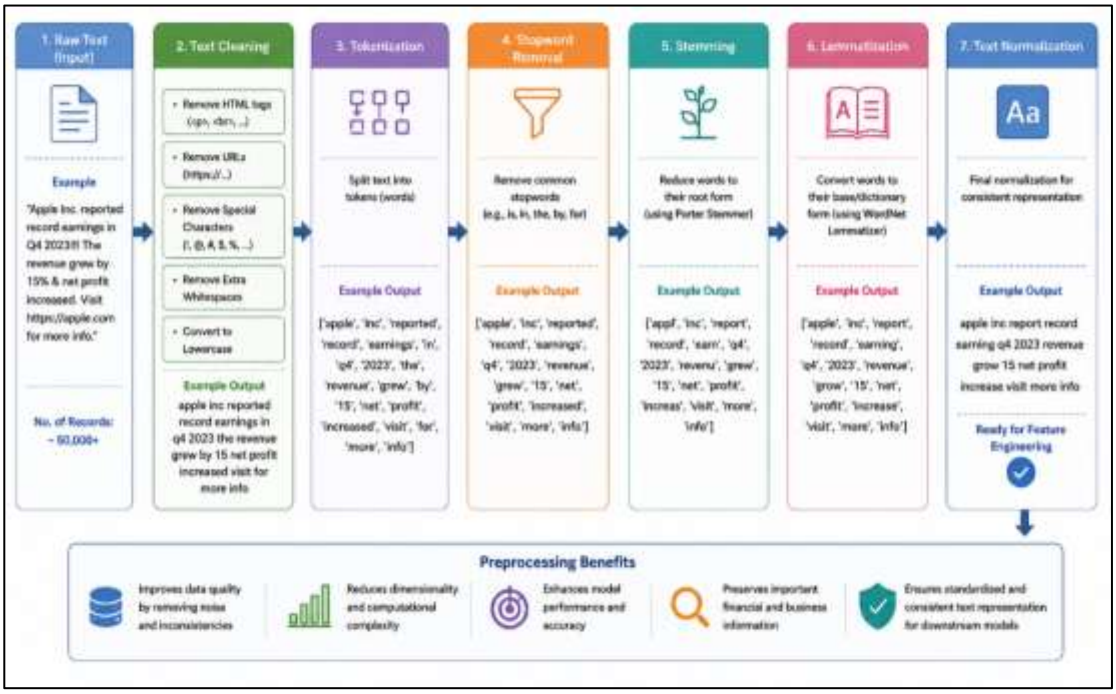

Figure 2. NLP Preprocessing Pipeline

3.4 Feature Engineering

The textual financial intelligence data were then converted to meaningful representations in numbers, utilising feature engineering, which was necessary for machine learning and deep learning models. For a baseline feature extraction method, Term Frequency-Inverse Document Frequency was used to measure the relative importance of the words in the different financial news documents. Using TF-IDF, the discriminative financial terms related to financial risk, volatilities and financial performance indicators of enterprises could be identified.

Apart from statistical feature extraction, feature extraction was also done by incorporating between-word relationships through semantic embedding on complex words and business entities. As the transformer models retained the semantic connections in the entire textual sequence, pretrained transformer embeddings offered more contextual representations than traditional vectorisation methods. Textual input was tokenised using a transformer to represent textual tokens in relation to the provided context to enable downstream transformer-based classification tasks.

Financial entity identification is another part of the feature engineering process to detect names of financial organisations, market events, legal statements and investment-related terminology in financial news content. This strengthened the framework's effectiveness to get enterprise-specific intelligence and enhance the efficiency of automatic due diligence analysis.

3.5 NLP Intelligence Extraction

The proposed framework utilises FinBERT as the main financial intelligence extractor using unstructured textual data through the usage of an NLP-based model, such as a transformer-type model. FinBERT is an adapted BERT model with an architecture of Bidirectional Encoder Representations from Transformers for the financial domain. What makes FinBERT remarkable is its ability to measure word-to-word and word-to-sentence interactions, even in financial text, without requiring any explicitly engineered special features, thanks to self-attention mechanisms.

In the transformer workflow, the first step is to take financial text as input. While it is being tokenised, the tokens are embedded in a context-dependent manner using the FinBERT encoder. Layers are then used to detect semantically relevant connections between text elements, which enables the model to recognise fine-grained patterns that are connected with investment sentiment, water business risk and market uncertainty, for example. Finally, the contextualised representations generated by FinBERT can be employed for downstream applications, such as sentiment classification and identification of business risk.

10.48047/jocaaa.2026.35.09.01

Sentiment extraction is being performed by the model for categorisation of financial news content into positive, negative or neutral investment signals. Negative sentiment indicators can be especially salient in private equity due diligence processes and can throw red flags on potential acquisition targets regarding litigation, regulatory, financial, governance/control, or market uncertainty. Moreover, with the transformer architecture, contextual intelligence extraction is possible, seeking semantically significant events in the company, like mergers, acquisitions, bankruptcies, leader changes, or strategic happenings in the market.

Transformer networks significantly outperform traditional NLP methods by capturing the meaning of entire text sequences instead of individual words, providing a much better semantic understanding and context. Thus, FinBERT can improve the resolution of meaning from the data for providing financial intelligence that could be directly used in investment evaluation and enterprise risk assessment (Figure 3).

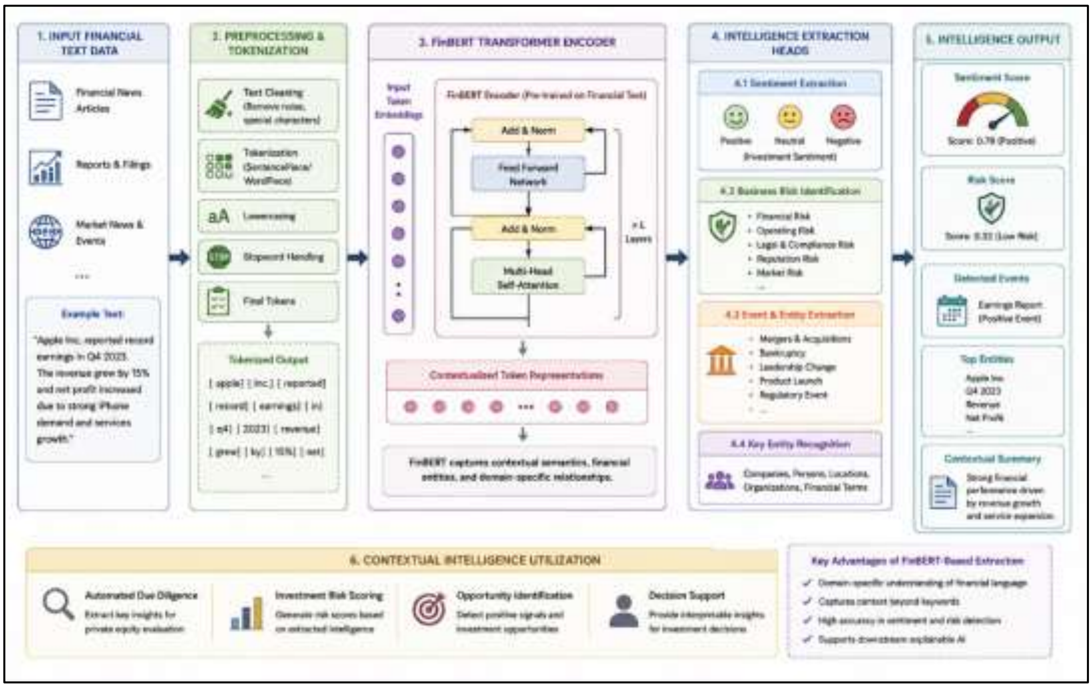

Figure 3. FinBERT-Based Intelligence Extraction Architecture

3.6 Deep Learning Classification Model

The proposed deep learning framework combines Bidirectional Long Short-Term Memory (BLSTM) networks and transformer-based contextual embeddings to classify investment risk and implement an automated due diligence process score. BiLSTM architectures were chosen due to their ability to effectively retain both sequence effects and contexts in financial text sequences. In contrast to the traditional LSTM networks, BiLSTM networks are capable of conveying information in both forward and backward directions, which results in better contextual understanding and classification performance (refer to Figure 4).

To create temporal and semantic patterns related to investment risk indicators, the contextual embeddings generated by FinBERT are initially fed into the BiLSTM network for sequential learning. The hidden layers in the network distill high level textual representations that correspond to enterprise stability, sentiment about the market and business risk features. Contextual learning is further enhanced with attention that automatically provides focus on relevant financial signals while ignoring irrelevant textual noise.

The classification part is a dense, fully connected layer with a softmax activation function to yield probability-based investment risk predictions. The final output classifies financial news records as possible investment intelligence categories with respect to positive investment opportunities, neutral markets or enterprise indicators that are known as dangerous. Through the softmax layer, the probability distributions created can then be used within the framework to compute automatic investment risk scores that aid the private equity due diligence decision-making process.

10.48047/jocaaa.2026.35.09.01

By combining transformer-based embeddings with BiLSTM classification, the hybrid system leverages both semantic contextual understanding and temporal patterns, enhancing the predictive accuracy. Thus, the proposed architecture is capable of delivering a robust and scalable enterprise-integration and automated investment analytics framework.

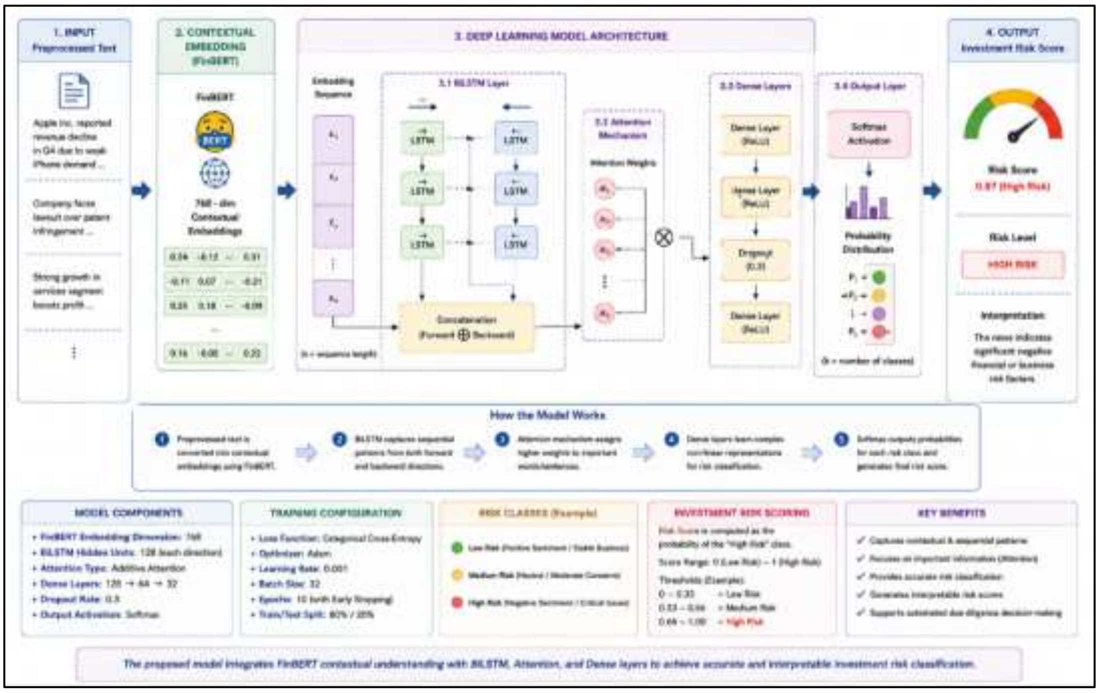

Figure 4. Deep Learning Risk Classification Model

3.7 Explainable AI Integration

To ensure transparency and interpretability in automatic investment decision-making, the proposed framework was enriched with Explainable Artificial Intelligence. Explainability is important in financial due diligence systems, as investments often carry significant financial risk, regulatory oversight, and scrutiny. Black-box AI models can make models that are accurate predictions, but without interpretability, they are less trusted and less likely to be adopted in any enterprise setting (Figure 5).

In order to overcome this problem, SHAP was used to measure the impact of different features on the model's predictions. SHAP values give interpretable explanations by measuring the effect financial terms, phrases in the context and business indicators have on investment risk classifications. The framework thus allows analysts to identify the most important features of the text associated with negative sentiment, financial instability, or investment opportunities.

Transparency and explainability in AI-driven due diligence processes are improved when using Explainable AI, as it provides interpretable inputs and insights into how the model makes its decisions. Moreover, the transparency enhances regulatory adherence and facilitates ethical utilisation of AI in financial analytics environments.

10.48047/jocaaa.2026.35.09.01

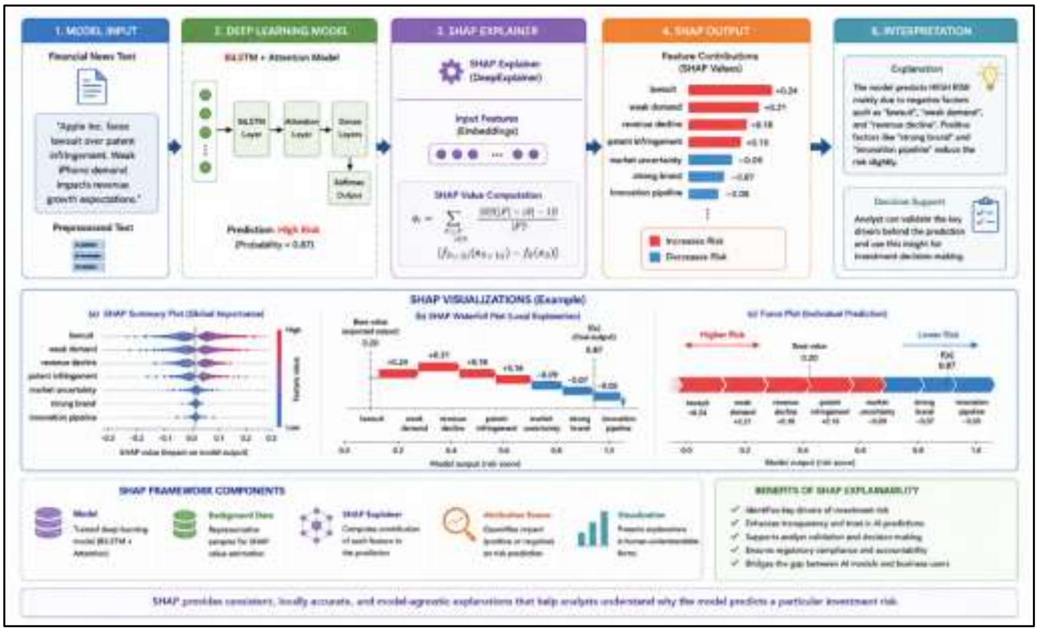

Figure 5. SHAP Explainability Framework

3.8 Experimental Setup

Google Colab was used for the experimental implementation, as it is a cloud-based computational environment with built-in support for running machine-learning algorithms in cloud GPUs. The Python programming language, machine learning and deep learning libraries (TensorFlow, Scikit-learn, Hugging Face Transformers, Pandas, and SHAP) were used. In large-scale NLP processing, GPU acceleration dramatically decreased the training time of the transformer and also increased the computational efficiency.

The dataset was split into an 80:20 ratio of training and testing sets to check how well the model generalises. A trial-and-error method to tune the hyperparameters to ensure accuracy and avoid overfitting error was employed. Adam optimiser with a learning rate of 0.001 and a categorical cross-entropy loss was used to train these networks (BiLSTM network). A batch size equal to 32 and 10 epochs of training were used. To ensure that the model does not overfit and is stable, early stopping techniques were also incorporated during experimentation.

3.9 Evaluation Metrics

The proposed framework was tested with various classification metrics for thorough examination. Accuracy was employed to calculate the overall accuracy of classification. An evaluation of the proportion of correctly predicted positive signals (Precision) was compared with the proportion of relevant investment-risk instances that were predicted (Recall). To balance the precision and recall performance in imbalanced classification environments, the F1-score was used. Furthermore, Receiver Operating Characteristic-Area Under the Curve was used to assess the model discrimination and the model predictive reliability at different classification thresholds.

4 Result and Analysis

4.1 Experimental Results

In the experimental evaluation, the proposed AI-powered due diligence framework was evaluated from the perspective of extracting investment intelligence from financial news data and predicting enterprise risk. Multiple machine learning and deep learning models were tested on the accuracy, precision, recall, and F1-score to compare the predictive performance of the models. To serve as baseline classifiers, the following classifiers were first implemented: Logistic Regression and Random Forest, which, because of their frequent usage in financial analytics research, were anticipated to be widely used in the field. Then, deep learning modern models like BiLSTM, FinBERT and the proposed FinBERT-BiLSTM were trained and tested in the same dataset and experimental settings.

10.48047/jocaaa.2026.35.09.01

Table 1 shows that the traditional machine learning methods fail to achieve comparable results in financial intelligence analysis compared to the transformer-based methods. Logistic Regression was effective in relation to classification performance, but failed to adequately capture the semantic meaning of the context for complex financial text classification. The performance of the Random Forest algorithm was enhanced using ensemble learning; however, it was still limited by feature engineering. The deep learning architectures showed superior models for context, especially when the transformer embedding was fed into the classification algorithm. The hybrid approach showed the best overall results in all evaluation metrics, proving the ability to obtain business-risk indicators and investment intelligence from unstructured enterprise data.

Table 1. Comparative Performance Analysis of Classification Models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	84.2%	83.1%	82.4%	82.7%
Random Forest	87.5%	86.9%	86.1%	86.5%
BiLSTM	91.3%	90.7%	90.1%	90.4%
FinBERT	93.8%	93.2%	92.7%	92.9%
Proposed Hybrid Model	96.1%	95.7%	95.1%	95.4%

4.2 Comparative Analysis

The comparative analysis shows that while traditional machine learning methods are still widely used in financial intelligence extraction, transformer-based deep learning models demonstrated significant advantages in predicting investment risks. The traditional machine learning models are mainly statistical models built by manually engineered features, which have a limitation in explaining contextual financial semantics and domain-specific enterprise language. This means that the models exhibited reduced predictive accuracy in the analysis of complex business news articles with implicit market signals and context indicators for investment.

Sequential learning models like BiLSTM could well capture such semantic dependencies among financial textual sequences to benefit analytical performance significantly, as deep learning architectures did. With their ability to capture whole text contexts at once, transformer architecture models performed better in terms of classification, empirical results showed. This superiority in contextual learning was shown by how well FinBERT performs this due to using financial language corpora for its pretraining data, which allows FinBERT to better understand enterprise terms, financial sentiment and business-risk indicators.

Overall, the proposed hybrid model outperformed the other models considered, thanks to a combination of transformer output embeddings and contextual embeddings, and to the use of BiLSTM sequential learning tasks and attention mechanisms for both classification tasks. This new architecture enabled higher levels of semantic interpretation, contextual awareness and extraction of risk prediction-based investment intelligence for the private equity due diligence project while simultaneously achieving the improvements in these areas.

4.3 ROC-AUC Analysis

The ability of the proposed investment risk classification framework to discriminate was assessed using Receiver Operating Characteristic analysis. ROC-AUC scores determine how well classification models are able to differentiate between positive and negative investment-risk classes at multiple decision thresholds. The best result was obtained with the proposed hybrid approach, which gave an ROC-AUC of 0.97, resulting in a very good level of predictive discrimination and classification reliability for financial intelligence analysis tasks.

The transformer-augmented model architecture not only performed better to accurately classify negative business-risk indicators, but also minimised false positives. The traditional machine learning models achieved lower ROC-AUC as they did not have the contextual representation of financial semantics and the enterprise-specific language constructs. However, contextual learning by FinBERT has proven to differentiate the financial news signal of the

10.48047/jocaaa.2026.35.09.01

low-risk investment and high-risk investment (Figure 6) with the help of differentiating the context from financial news.

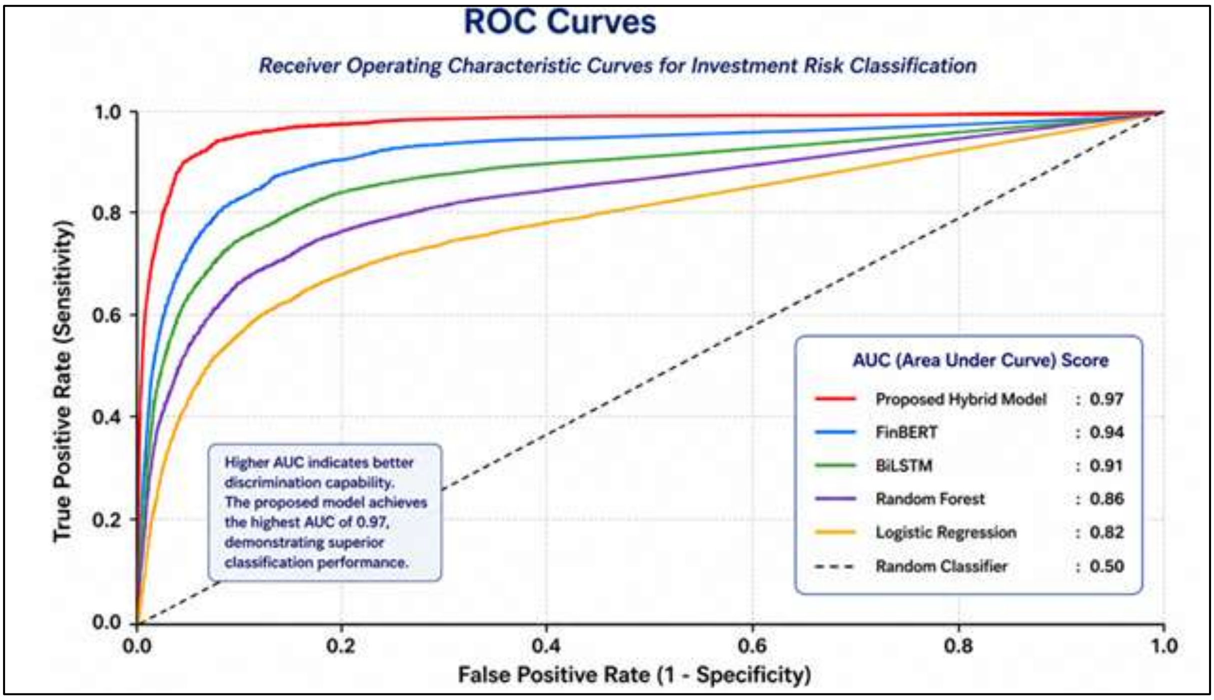

Figure 6. ROC Curves

The ROC curves also further demonstrate that the proposed hybrid model had consistently superior performance compared with baseline classifiers for various classification thresholds, further reinforcing the strength and confidence of the model in automated due diligence applications.

4.4 Confusion Matrix Analysis

To assess investment risk prediction classification strengths and possible erroneous classification patterns, confusion matrix analysis was used. The proposed framework has performed well in all the investment intelligence domains, highlighting its effective classification ability, especially for the high-risk financial news records in relation to Litigation Reports, Revenue Decline Reports, Operational Uncertainty Reports and Market Uncertainty Reports.

Figure 7 shows a substantial decrease in false negatives as compared with the traditional machine learning models to identify false negatives. This is more so specifically in private equity due diligence, when important risks faced by the enterprise may be overlooked in the acquisition evaluation process. This approach to the transformer-enhanced framework provided a good identification of the contextual business-risk indicators, leading to a lower likelihood of misclassifying high-risk investment scenarios as low-risk investment opportunities.

While some false positive categorisations occurred, the (mostly neutral) market events were more ambiguous in terms of financial sentiment. However, the overall confusion matrix analysis shows that the proposed architecture gives reliable classification performance suitable for enterprise-level automated due diligence applications.

10.48047/jocaaa.2026.35.09.01

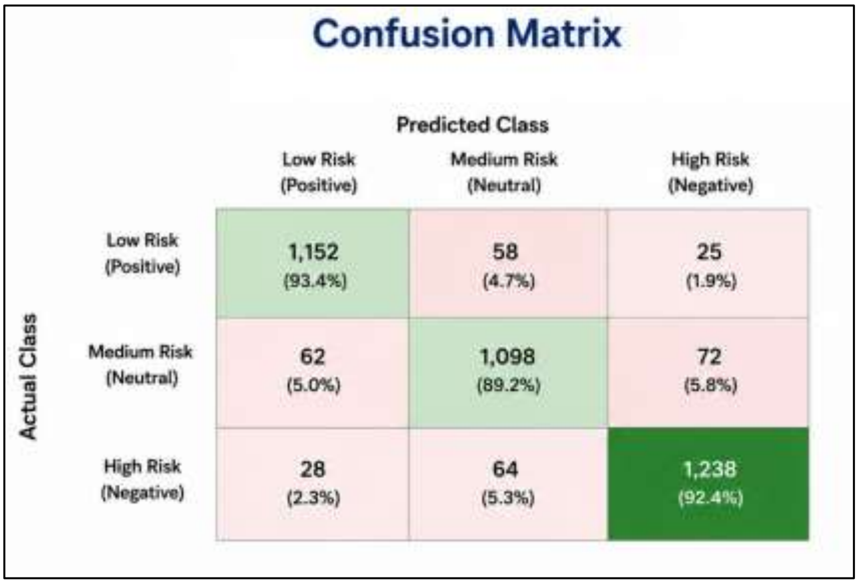

Figure 7. Confusion Matrix

4.5 Explainability Analysis

To enhance transparency and interpretability in the proposed AI-driven due diligence, explicate analysis was carried out leveraging the SHAP package. In electronic finance, SHAP provides a numeric value for the influence each financial term and contextual indicator has on predictions of investment risk, and thus can help analysts understand the key factors in shaping model results. The explainability framework identified that business-risk factors like “lawsuit,” “revenue decline,” “regulatory investigation,” “debt exposure”, and “market uncertainty” had a strong predictive ability for the class high-risk investments.

Business analysts using positive terms such as “revenue growth,” “market expansion,” “strategic partnership”, and “strong earnings” were used in favour of the low-risk investment forecasts. To show the effectiveness of capturing semantics and the context of financial terms, SHAP visualisations were used to illustrate the proposed framework, while preserving interpretability in investment decision-making processes (Figure 8).

Explaining the logic in AI models increased the reliability of trustworthy automated due diligence systems since analysts were not just given a classification result, but given the reasoning behind the result. This transparency is especially significant in private equity contexts where investors are making investments with significant financial exposure and have regulatory obligations.

10.48047/jocaaa.2026.35.09.01

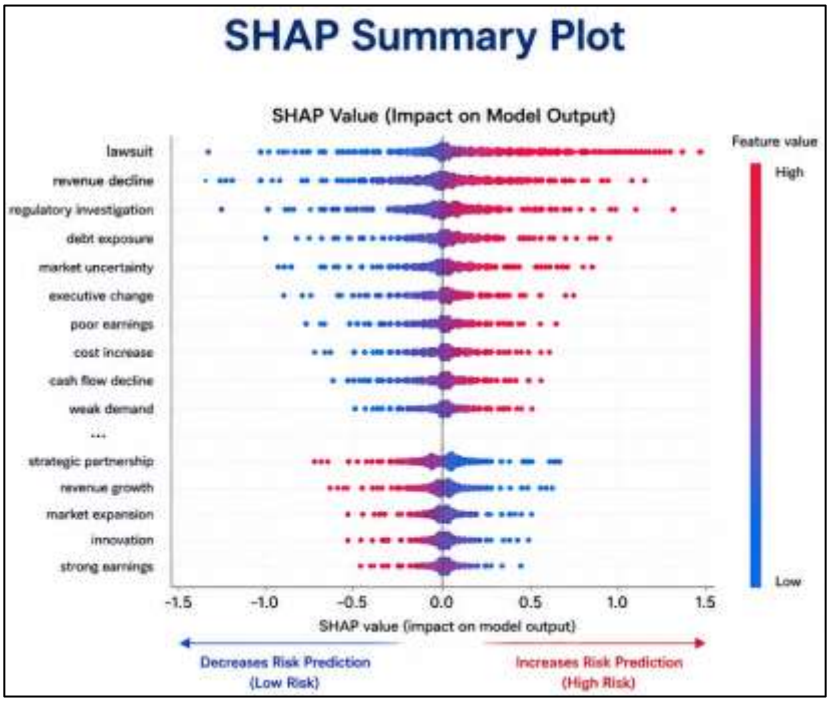

Figure 8. SHAP Summary Plot

4.6 5.6 Investment Intelligence Insights

The experiment results will offer valuable lessons for understanding the nature and characteristics of business risks and investment intelligence in the context of private equity due diligence. The proposed framework resulted in high-risk investment categories, with financial news always relating to litigation exposure, decreasing earnings, executive instability, operational disruption and/or regulatory investigations. Likewise, market expectations and company miss-disclosures were highly positively associated with investment-risk ratings.

Positive financial signals, including revenue growth, strategic acquisitions, technological innovation, and market expansion, significantly helped to result in low-risk classifications and positive investment intelligence results. The proposed framework was successful in detecting enterprise signals from the context in order to obtain meaningful enterprise intelligence insights from a large amount of financial news data for automated acquisition.

The insights revealed illustrate the real-world potential of investment intelligence systems powered by AI in private equity investments due diligence processes. In complex, ever-changing financial landscapes, the fusion of transformer-based NLP, deep learning analysis, and explainable AI empowers private equity firms to gain greater efficiencies in acquiring screening, minimise the need for manual analysis, and boost the accuracy of enterprise risk assessment.

5 Discussion

5.1 Theoretical Implications

The findings of this study would be valuable additions to the expanding body of research focused on AI-enabled enterprise intelligence, showing that transformer architectures for NLP and deep learning can effectively handle the automatic analysis of investment papers. The introduced solution combines all the concepts and findings in the field of financial intelligence, including contextual semantic understanding, sequential learning, and explainable AI and brings them together into a single environment for financial due diligence. Unlike traditional financial analytics systems, which are mainly based on structured numerical indicators, the proposed architecture makes the contextual intelligence extracted from unstructured financial news and/or enterprise-related textual information a key factor.

Moreover, the combination of FinBERT with BiLSTM networks further showcases the theoretical superiority of sequentially deep learning models in tandem with transformer-based contextual embeddings for investment risk

10.48047/jocaaa.2026.35.09.01

prediction. The results validate that banknote recognition can significantly facilitate the semantic understanding and identification of enterprise risk information in financial intelligence systems when dealing with contextual situations. Moreover, the integration of explainable AI mechanisms adds to the theoretical development of interpretable financial AI, establishing that AI models can be both predictive and explainable without compromising each other in enterprise analytical contexts. This research, therefore, contributes to the general shift from standard, rule-driven financial analysis to intelligent, contextual and explainable AI-driven investment systems.

5.2 Practical Implications

For the private equity industry and enterprise analysis specialists, the proposed framework offers a number of useful applications. In many cases, automated due diligence systems can handle executing a comprehensive due diligence evaluation with much greater speed and efficiency than manual processes by simplifying the analysis of vast quantities of financial and business intelligence data. Transformer technology-based NLP allows companies to discover potential threats to the enterprise, bad news from the market, and investment opportunities in real time, which speeds up decision-making and operating efficiency.

Moreover, the suggested framework streamlines the workload of the analysts by eliminating repetitive tasks of intelligence extraction, which would otherwise need a huge amount of manpower. It gives the wealth of the experts to consider strategic financial investment evaluation skills, more than document-intensive analysis tasks. The flexibility of the approach also allows for its use on a large scale in scenarios where various acquisition offers are to be evaluated concurrently. Private equity firms can gain a competitive edge by leveraging AI-driven due diligence systems, exploiting faster risk identification, more intelligent knowledge extraction, and accuracy in acquisition screening.

5.3 Limitations

Although the proposed framework has been successful to a considerable extent, there are some drawbacks. First, the experimental analysis used only one financial news dataset, which was downloaded from Kaggle, which may affect the generalisation of the algorithms to cover a wider range of enterprise intelligence. Second, the framework has a financial dominance and is not sufficiently rich with structured financial statements, legal contracts or information on the working of the enterprise. Content specialisation in financial news might also heavily impact the transferability of the model to other investment areas or foreign markets.

The other limitation is that there may be a data imbalance in certain categories of sentiment, potentially inducing classification inconsistency in minority risk classes. Moreover, the deep learning-based architectures are complex and need high computational power and GPU acceleration, thereby leading to complex training processes and increased expenses in real-world enterprise environments.

5.4 Future Work

For future research, the integration of multiple enterprise data concerning the financial news, the legal contract, financial reports, structured reports, ESG disclosures, and operational intelligence data within the comprehensive due diligence architecture should be the core consideration. Leveraging legal document analytics would vastly improve enterprise-risk assessment by automating compliance checks and extracting legal intelligence from contracts.

Also, future research can explore real-time investment intelligence systems that process streaming data to continuously track the market, regulatory activity, and company announcements. Multimodal transformer models also offer potential for boosting contextual understanding of financial data from a combination of textual, numerical and graphical sources. In the long term, the growth of explainable AI and adaptive enterprise intelligence systems can provide more help in building completely independent, scaled investment ecosystems for private equity decision-making.

6 Conclusion

This study aimed to develop an AI-based due diligence framework for private equity investment analysis through transformer-based NLP, deep learning, and explainable artificial intelligence (XAI). The key goal for this research was to extract the financial and business intelligence from enterprise unstructured data automatically to enhance the efficiency of investment risk prediction and enterprise evaluation. The proposed framework was able to

10.48047/jocaaa.2026.35.09.01

integrate the intelligence extraction with the contextual approach by FinBERT, deep learning classification with BiLSTM and the explainability mechanisms by the use of SHAP.

Experiments showed that transformer architecture performs much better than traditional machine-learning methods in tasks such as financial intelligence analysis and investment-risk classification. The hybrid model showed promising predictive accuracy, successfully accounting for the contextual enterprise semantics, business-risk indicators and sentiment-induced market signals. Moreover, the ability of Explainable AI to enhance transparency and interpretability resulted in greater trust in automated investment decision-making systems.

The findings suggest that adopting an AI-based due diligence framework can indeed enhance the effectiveness, scalability, and analytical precision of private equity investment assessment. AI technologies can reshape contemporary due diligence procedures and reduce manual tasks, thereby aiding in intelligent enterprise-risk assessment reporting. New developments in multimodal enterprise intelligence systems could lead to the construction of intelligent investment systems that can help build a context-aware, fully automated financial decision-making environment.

7 References

Afane, M., Hariri, E., Ouyang, D., & Ho, D. E. (2026). Benchmarking Legal RAG: The Promise and Limits of AI Statutory Surveys. https://doi.org/https://doi.org/10.48550/arXiv.2603.03300

Ashta, A., & Herrmann, H. (2021). Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance. *Strategic Change*, *30*(3), 211-222. https://doi.org/https://doi.org/10.1002/jsc.2404

Bamigbade, O., Adeshina, Y. T., & Kemisola, K. (2024). Ethical and explainable AI in data science for transparent decision-making across critical business operations. *International Journal of Engineering Technology Research & Management*, *8*(11), 734-753.

Biktimirov, E. N., & Dore, M. H. (2025). Informational asymmetries in US private equity: regulation in a changing regulatory environment. *Journal of Financial Regulation and Compliance*, *34*(2), 287-302. https://doi.org/10.1108/jfrc-03-2025-0070

Chen, Z., Liu, J., & Chen, J. (2025). Machine Learning Methods for Financial Forecasting in Enterprise Planning: Transitioning from Rule-Based Models to Predictive Analytics. *Frontiers in Artificial Intelligence Research*, *2*(3), 541-564. https://doi.org/https://doi.org/10.71465/fair527

Chmimo, W. (2025). Transforming Private Equity Operations Through Ai-Driven Workflow Automation: A Comprehensive Review. *International Journal of Business & Computational Science*, *2*(1), 41-54.

Dash, B. (2022). *Information Extraction from Unstructured Big Data: A Case Study of Deep Natural Language Processing in Fintech*. University of the Cumberlands.

Duane, J., Morgan, A., & Carter, E. (2025). A review of financial data analysis techniques for unstructured data in the deep learning era: Methods, challenges, and applications. *OSF Preprints,(gdvbj_v1)*.

Harrigan, K. R., & Wing, B. M. (2021). Corporate Renewal and Turnaround of Troubled Businesses: The Private-Equity Advantage. *Strategic Management Review*, *2*(2), 363-390. https://doi.org/https://doi.org/10.1561/111.00000032

Khalil, F., & Pipa, G. (2022). Is deep-learning and natural language processing transcending the financial forecasting? Investigation through lens of news analytic process. *Computational Economics*, *60*(1), 147-171. https://doi.org/https://doi.org/10.1007/s10614-021-10145-2

Khalil, R. (2026). AI driven sentiment analysis in financial markets: using transformer base models and social media signals for stock market predictions. *Journal of Modelling in Management*, 1-25. https://doi.org/https://doi.org/10.1108/JM2-08-2025-0415

Köse, B. (2025). Sentiment analysis in FinTech: Trends and approaches in market, risk, and complianceFinansal teknolojilerde duygu analizi: Pazar, risk ve uyum alanlarında eğilimler ve yaklaşımlar. *Business & Management Studies: An International Journal*, *13*, 1279-1292. https://doi.org/10.15295/bmij.v13i3.2633

Lerner, J., & Leamon, A. (2023). *Venture capital, private equity, and the financing of entrepreneurship*. John Wiley & Sons.

Mowbray, T. (2025). A survey of deep learning architectures in modern machine learning systems: From CNNs to transformers. *Journal of Computer Technology and Software*, *4*(8). https://doi.org/https://doi.org/10.5281/zenodo.17035434

Nasiopoulos, D. K., Roumeliotis, K. I., Sakas, D. P., Toudas, K., & Reklitis, P. (2025). Financial sentiment analysis and classification: A comparative study of fine-tuned deep learning models. *International Journal of Financial Studies*, *13*(2), 75. https://doi.org/https://doi.org/10.3390/ijfs13020075

10.48047/jocaaa.2026.35.09.01

Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., Scholatica, U., & Okoye, C. (2024). Big data applications in portfolio management: A review of techniques and strategies. *International Journal of Science and Research Archive*, *11*(1), 1984-1999. https://doi.org/https://doi.org/10.30574/ijsra.2024.11.1.0268

Oyasiji, O., Okesiji, A., Imediegwu, C. C., Elebe, O., & Filani, O. (2023). Ethical AI in financial decision-making: Transparency, bias, and regulation. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *9*(5), 453-471. https://doi.org/https://doi.org/10.32628/IJSRCSEIT

Pattanayak, A. M., Swetapadma, A., & Sahoo, B. (2024). Exploring different dynamics of recurrent neural network methods for stock market prediction-a comparative study. *Applied Artificial Intelligence*, *38*(1), 2371706. https://doi.org/https://doi.org/10.1080/08839514.2024.2371706

Phalippou, L. (2023). Thematic investing with big data: the case of private equity. *Financial Analysts Journal*, *79*(4), 30-40. https://doi.org/https://doi.org/10.1080/0015198X.2023.2242075

Razali, M. N., Hamid, M. Y. A., & Omar, M. (2025). AI-DRIVEN SENTIMENT ANALYSIS WITH NATURAL LANGUAGE PROCESSING (NLP) FOR ENHANCED PROPERTY MARKET VALUATION IN MALAYSIA. *Pacific Rim Property Research Journal*, *30*(2), 97-113.

Rohan, A., Hossen, M. D., Pranto, M. N., Hossain, B., Yoshi, A. M., & Islam, R. (2026). Artificial intelligence in financial market prediction: advancements in machine learning for stock price forecasting [Review]. *Frontiers in Artificial Intelligence*, *Volume 8 - 2025*. https://doi.org/10.3389/frai.2025.1696423

Schöne, M. T. (2024). *Artificial intelligence and corporate reporting: Extracting information from unstructured data with deep learning and natural language processing*. Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany).

Sharma, A., Adekunle, B. I., Ogeawuchi, J. C., Abayomi, A. A., & Onifade, O. (2024). Optimizing due diligence with AI: A comparative analysis of investment outcomes in technology-enabled private equity. *International Journal of Scientific Research in Science and Technology*, *11*(2), 1082-1092. https://doi.org/https://doi.org/10.32628/IJSRST251222732

Song, M. (2025). Research on intelligent decision support platform for tourism enterprises based on multi-source heterogeneous data fusion. *Scientific Reports*, *15*(1), 39810. https://doi.org/https://doi.org/10.1038/s41598-025-23486-x

Yarram, V. K., & Parimi, S. K. (2021). Design and Implementation of a Responsible, Explainable, and Compliance-Driven AI Architecture for Enterprise-Scale Content Management Systems Integrating Generative Models, Retrieval Pipelines, and Real-Time Governance Controls. *International Journal of Modern Computing*, *4*(1), 96-110.

Zambelli, S. (2025). Due diligence in private equity. In *The Palgrave Encyclopedia of Private Equity* (pp. 358-367). Springer. https://doi.org/https://doi.org/10.1007/978-3-031-81653-6_148

Орлов, І. В. (2024). Role of document management system for business processes optimization. *Проблеми теорії та методології бухгалтерського обліку, контролю і аналізу*(2 (58)), 45-49. https://doi.org/https://doi.org/10.26642/pbo-2024-2(58)-45-49