Received 10 November 2025, accepted 26 February 2026, date of publication 31 March 2026
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.25.1.4490

Financial Video Analysis Using Natural Language Processing: An Empirical Study of Stock Forecast

OLEKSII IVANOV, VLAD IATSIUTA, VITALIY KOBETS
Department of Computer Science and Software Engineering (CSSE), Kherson State University, Kherson, Ukraine.

Corresponding Author: Vitaliy Kobets (e-mail: vkobets@kse.org.ua)

ABSTRACT The proliferation of unstructured financial video content presents a significant challenge for traditional investment analysis. Natural Language Processing (NLP) offers a promising solution for extracting value from this data. This empirical study investigates whether NLP technologies can automatically extract, structure, and validate actionable investment insights from financial videos. We propose an automated pipeline using video transcription services and Large Language Models (LLMs). The methodology was tested on 22 YouTube financial analysis videos focusing on Amazon (AMZN) and Google (GOOG). The ChatGPT-4 model processed transcripts to extract stock tickers, risk levels, and price forecasts into a structured JSON format. The system achieved 100% accuracy in company recognition and filtering irrelevant content. Empirical validation against actual market data revealed an overall forecast accuracy of 85% (90% for AMZN, 80% for GOOG). This NLP approach also outperformed traditional ARIMA time-series models. The findings confirm that NLP can feasibly automate the analysis of financial video, transforming unstructured media into validated, structured data to support investment decisions.

KEYWORDS natural language processing, unstructured data, financial data analysis, ARIMA, LLM.

I. INTRODUCTION

In today's business landscape, data volumes are expanding at an unprecedented rate – 402.74 million terabytes in 2024 [1]. A large portion of this data is unstructured, including emails, documents, images, audio, and video files. This information holds both irrelevant content and valuable insights crucial for strategic decisions, process optimization, and business analysis [2]. However, traditional methods face major difficulties in handling such data. The core challenge lies in interpreting context, identifying hidden relationships, and transforming unstructured data into structured formats. Conventional tools such as Elasticsearch, Solr, IBM SPSS Text Analytics, and SAS Text Miner often struggle to handle the diversity and volume of modern information flows.

Natural Language Processing (NLP) enables automated text analysis, including keyword extraction, semantic interpretation, and document classification. Similarly, computer vision supports automated recognition and analysis of visual data such as images and videos, offering new opportunities for media content analysis. The integration of AI into these domains accelerates large-scale data processing, improves analytical accuracy, and uncovers patterns beyond human detection [3]. For instance, NLP can analyze customer feedback and social media sentiment, enabling companies to respond swiftly to shifts in demand or reputation. It also facilitates the analysis of press releases, public statements, and media coverage to detect events affecting a company's financial outcomes [4].

According to the Stanford AI Index 2025 report, as of 2024, artificial intelligence has surpassed human capabilities in most basic cognitive domains, including clinical diagnostics, competitive mathematics, and complex PhD-level scientific problems. However, human experts still retain an advantage in solving the most complex logical problems and executing long-term projects with extensive timeframes, where current AI systems cannot yet match human capabilities [5]. Integrating NLP into unstructured data processing has become both a competitive advantage and a necessity for businesses aiming to maintain industry leadership. As technologies advance and costs decline, implementation becomes more accessible, enabling innovation and growth. According to OpenAI data, ChatGPT reached 1 million users within a week of its late 2022 release and surpassed 100 million monthly users by early 2023. Global Market Insights projects that the wearable AI market will reach $180 billion next year. Increasingly, businesses recognize AI's strategic value: Forbes reports that 83% of companies view AI as a top priority, with chatbots and automated emails among the most common applications. Analytics Insight notes that nearly 80% of retail executives expect AI automation adoption by 2025.

Investors unfamiliar with a particular language or lacking the time for detailed research can leverage automated tools to analyze vast amounts of multimedia and textual data efficiently. Automatic translation and NLP tools remove language barriers, providing access to global insights. Processed information can then be queried interactively, supporting informed investment decisions. Entrepreneurs can monitor trends, customer sentiment, and competitor activities; analysts and researchers can sift through large volumes of academic literature; and financial professionals can gain fast access to market analytics, forecasts, and asset indicators [6]. In all cases, NLP and unstructured data processing enhance analytical productivity and democratize access to intelligence, enabling timely responses to global and financial developments.

The main **research question** of this paper is whether NLP technologies can automatically extract and validate actionable investment insights from unstructured financial video content with sufficient accuracy to support decision-making. The study aims to analyze sentiment dynamics at the individual stock level using a data mining architecture integrated with large language models for detailed sentiment analysis.

This paper examines NLP's role in processing unstructured data, its potential applications, and related challenges. It proposes an interaction framework connecting users, AI systems, and datasets, detailing the key services and interaction points. The practical section presents examples of video data processing for business analysis, illustrating how these results support problem-solving. Real-world use cases are discussed, along with prospects for this rapidly evolving field.

II. LITERATURE REVIEW

The application of natural language processing (NLP) methods for forecasting financial instrument prices has demonstrated significant effectiveness, particularly when utilizing unstructured data sources (such as regulatory reports, financial news, and social media) and for long-term predictions [7, 8].

Let us consider the effectiveness of NLP methods:

1. Forecasting stock market indices based on regulatory reports.

Much of the research focuses on using the text content of regulatory reports (ad hoc announcements) to forecast stock indices such as the German DAX, CDAX, and European STOXX Europe 600 [7].

Long-term effectiveness (24 months) [7]:

- NLP-based text models successfully reduce forecast errors (root mean squared error, RMSE) at a statistically significant level compared to baseline forecasts based solely on historical values.
- For the long-term horizon (24 months), a significant reduction in RMSE was achieved: DAX: 19.5% reduction in RMSE, CDAX: 19.4% reduction in RMSE, STOXX Europe 600: 35.6% reduction in RMSE.
- This confirms that the language embedded in regulatory reports has significant predictive power regarding future stock index levels in the long term.

Short-term effectiveness [7]:

- In short-term forecasts (e.g., 1 month), text models perform at a comparable level to baseline forecasts.
- The reduction in RMSE in the short term is less significant: 4.6% for the DAX, 4.4% for the CDAX, and 2.9% for the STOXX Europe 600.
- Short-term forecasts benefit significantly from machine learning, particularly with a restriction on large-cap companies, which helps to reduce the complexity of the input data.

2. Application of sentiment analysis and deep learning. NLP is a fundamental tool for sentiment analysis and extracting emotions from large data sets, which are then used as predictive features [8].

- Integration with quantitative data: Integrating sentiment analysis results with quantitative market data can significantly improve prediction accuracy [9].

For example, combining sentiment analysis (performed using the RoBERTa model) and quantitative data from the Tehran Stock Exchange (TSE) index in a Long Short-Term Memory (LSTM) model yielded the highest prediction accuracy compared to models using only quantitative data.

- Use of deep learning: Deep learning combined with NLP demonstrates significant predictive power. In stock price forecasting (using stock prices as an example), the Convolutional Neural Network (CNN) model outperformed all other models (including ARIMA, SVR, MLP, and LSTM) across RMSE, MAE, and MAPE metrics [8].
- Contribution of the sentiment index: The sentiment index developed using text analysis (World Halal Tourism Composite Sentiment Index, WHTCSI) can be a significant predictive factor. The contribution of this index to forecast quality was found to be 35.55% in the CNN model for stock price forecasting [8].
- Impact of time lag: The study showed that lag sentiment has better properties for stock price forecasting than current-day sentiment [8].

3. Social media analysis and targeted aspect analysis. Social media, such as Twitter, are valuable sources of information used for market screening [10].

- High accuracy of emotion detection: The Targeted Aspect-Based Emotion Analysis system, based on NLP and machine learning, achieves over 90% precision in detecting the financial emotions of 'opportunity' and 'precaution' in Twitter messages about specific assets. This indicates a strong level of practical interest in supporting investment decision-making.
- Accuracy advantage: When analyzing financial decisions, high precision is the most relevant performance metric, as false recommendations are much more harmful than missed ones (low recall).

4. Methodological aspects of NLP and challenges. To achieve high efficiency, NLP is used in conjunction with machine learning, thereby addressing specific challenges of financial data:

- Processing 'broad' data: Text mining generates high-dimensional feature matrices, where the number of predictors (words) can significantly exceed the number of observations (so-called 'broad' data). This increases the risk of overfitting.
- Dimensionality reduction: To combat overfitting, strategies such as [7] are used:

Data-driven: Use of implicit feature selection and regularization methods (e.g., Lasso, Ridge regression, Elastic Net), as well as principal component analysis (PCA) and latent semantic analysis (LSA).

Knowledge-driven: Use of sentiment dictionaries, such as the finance-specific Loughran-McDonald dictionary, which replaces high-dimensional features with predefined domain knowledge.

- Dynamic Model Selection: Methods that use sequential modeling with NLP, such as the TimeSpeaks framework, treat the task of selecting the best model for forecasting as an NLP text completion task. These methods demonstrate better predictive performance and scalability than traditional dynamic ensemble and model selection methods, especially for long-term planning [11].

Natural Language Processing transforms business communication with AI-based solutions that can make more grounded decisions than using only structural datasets [12]. The advantage of integrating semantic analysis with financial indicators from business reports for predicting stock prices is demonstrated through the processing of 10-Q reports and time-series datasets [13, 14].

Thus, NLP is an effective tool for financial forecasting, particularly when combined with advanced machine learning and deep learning models, which can significantly reduce forecast errors, especially at long-term horizons.

Textual data represents a significant yet underutilized resource for enhancing comprehension of financial markets. Previous research has not examined sentiment dynamics at the individual stock level through data mining frameworks, a gap that this investigation addresses by employing large language models (LLMs) for granular sentiment extraction. The application of LLMs enables the extraction of nuanced insights from textual sources, thereby enhancing the precision of financial asset price prediction models.

Investor sentiment encompasses beliefs regarding prospective cash flows and investment uncertainty. Market fundamentals shape investor sentiment, and empirical evidence shows that sentiment affects market returns, asset valuations, and cash flow patterns. Analyzing the temporal patterns [15] and determinants of sentiment fluctuations in equity markets provides valuable insights into financial policy mechanisms, risk assessment frameworks, and portfolio construction strategies [16].

Measuring investor sentiment poses methodological challenges due to its multifaceted nature. Conventional sentiment indicators are generally categorized into three approaches: market-based metrics, survey-derived indices, and event-driven measures (such as pandemic-related disruptions). Market-based sentiment proxies often lack precision because they may conflate multiple underlying phenomena, complicating their interpretation.

Sentiment classification encompasses three categories [14]: positive sentiment reflects favorable information or upward trajectories that signal growth prospects, profit expansion, or business development opportunities; negative sentiment indicates adverse factors, including financial losses or elevated costs; neutral sentiment comprises descriptive content without direct price implications, typically providing contextual information. For sentiment generation, the LLM receives financial text with the classification prompt: ‘What is the sentiment of the following financial text? Is it positive, negative, or neutral?’

Quality assurance architectures in information systems vary according to the characteristics of the data source. Structured data sources feature organized information in predefined formats, where the primary challenge is mapping natural-language user queries to the formal query language of the data repository. Unstructured data sources lack systematic organization and often contain heterogeneous content types, necessitating the implementation of quality assurance systems that utilize natural language processing methodologies [17].

Given token limitations in LLM architectures when processing financial documents, researchers have developed approaches that first generate information summaries before conducting sentiment analysis [18]. This methodology enables more accurate stock return forecasting using 10-K filings. Trading strategies informed by LLaMA-2 sentiment analysis demonstrate substantially superior buy-and-hold returns compared to FinBERT and conventional models [19]. Research has shown that ChatGPT-managed S&P 500 portfolios outperform traditional investment strategies [20]. A comparative analysis of LLM architectures (OPT, BERT, and FinBERT) for sentiment analysis of U.S. financial news from 2010 to 2023 reveals that the OPT model, built on the GPT-3 architecture, substantially outperforms BERT and FinBERT, achieving 74.4% accuracy in predicting market returns [21].

Virtual Customer Assistant (VCA) quality control systems integrate techniques from natural language processing, artificial intelligence, information retrieval, and information extraction to support question-answering capabilities [22]. The generalized architecture of such systems is illustrated in Fig. 1.

VCA queries can be factual, requiring single-sentence responses that include specific entities (organizations, individuals, dates), which are computationally simpler and do not require advanced NLP techniques. Alternatively, queries may request entity lists, hypothetical scenario details, causal explanations, or binary confirmation responses.

Stock product recommendations pose particular complexity because they require forecasting future asset prices while accounting for investor risk preferences and return-optimization objectives [23].

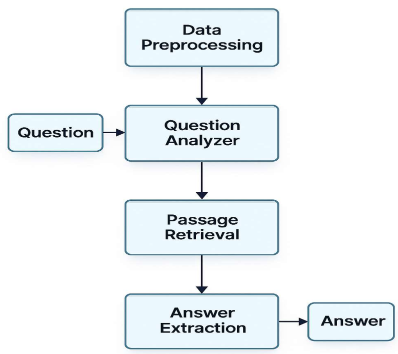

Figure 1. Architecture of the quality control system

Transformer models demonstrate superior prediction accuracy compared to the significantly slower traditional models, such as AutoARIMA, for enhancing retail demand forecasting [24]. The results show the superiority of the LSTM method over seasonal autoregressive integrated moving average (SARIMA) and triple exponential smoothing [25].

The proposed CNN-Correlation-based Attention Transformer model for forecasting financial time series (DJIA stock indices) demonstrated high accuracy (R-squared=0.9169 for DJIA), outperforming Deep-Transformer models [26]. Granger causality analysis revealed that sentiment indices have a statistically significant impact on changes in the stock market index (HSI) with a lag of 3 and 5 days [27]. The hybrid CNN-LSTM model with an attention mechanism achieved 98.7% accuracy in classifying sentiment from social networks (Twitter, Reddit) regarding cryptocurrencies. Emotional features extracted from reviews have a positive impact on predicting price movements of cryptocurrencies [28]. LLMs can serve as effective plug-and-play tools for subject-oriented classification tasks, even in the absence of large amounts of labelled data [29].

Language models are required to process text in multiple languages while considering context accurately for text analysis [30]. Knowledge-based AI systems can be categorized into three levels: Horizontal – recommendation systems that can be used today for decision-making; Partial information-based – systems where some information is available, and additional data is sought to aid decision-making for consumers or businesses; Fully autonomous systems – knowledge-based systems that currently operate primarily as recommendation systems but require further advancements to achieve complete decision-making control.

Success criteria for conversational AI agents, such as ChatGPT-like systems, constitute a major focus of research in this field, as they define the key dimensions for evaluating system effectiveness, user satisfaction, and real-world applicability.

- AI agents with domain-specific knowledge can help people from various industries and areas of expertise. People with other experiences and knowledge can satisfy their needs and requests using ChatGPT-like AI systems, as LLMs are trained on both domain-general and domain-specific knowledge.
- AI agents can prepare a quick response to the request because they can process large amounts of data better than humans. It can save time and improve performance metrics in personal and professional tasks. It increases human productivity. AI agents can perform the same tasks faster.
- Conversational AI agents can help automate an organization's business processes. It increases the organization's competitiveness in routine tasks and allows personnel to focus on tasks that are more creative.

Defining search strategies and knowledge sources is crucial for designing and developing effective AI systems. A key goal of AI systems has been to understand and work with NL, as humans widely use it to encode, express, and transfer knowledge. However, understanding NL is challenging because it requires extensive contextual knowledge. Most of this knowledge consists of common-sense reasoning, which we naturally acquire over time through interactions with the world. In recent years, new approaches to NL understanding that encode the structure of words in sentences using various quantitative methods have achieved great success. Today, many of these new approaches are integrated into large language models (LLMs). Modern AI systems increasingly leverage LLMs to enhance their capabilities, enabling powerful conversational experiences, such as those provided by ChatGPT and similar systems. These models demonstrate growing competence in understanding linguistic nuances, maintaining contextual awareness throughout interactions, and adapting their responses to user intent. Through continuous learning from vast datasets, LLMs improve their ability to interpret complex contexts and deliver more accurate, relevant, and human-like communication across diverse domains. (Fig. 2).

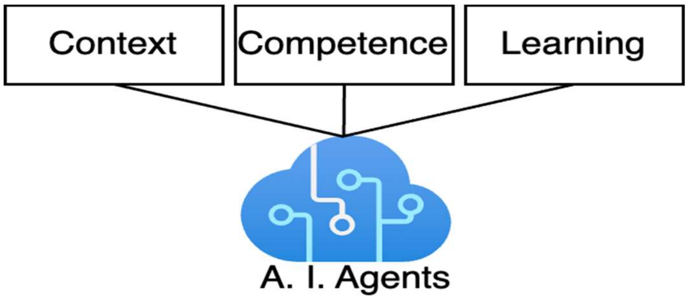

Figure 2. Development of AI Agents.

Current literature demonstrates that natural language processing technologies complement traditional quantitative approaches in supporting investment decisions. However, the application of NLP as a primary methodology for predicting financial instrument prices remains insufficiently explored. This research addresses this methodological gap in the existing literature.

III. SYSTEM ARCHITECTURE FOR UNSTRUCTURED DATA PROCESSING

The architecture discussion should begin with the data analysis module. Unstructured information consists of datasets lacking a defined format or organization and appears in multiple forms, each requiring a specific processing approach. Types of unstructured data include text (emails, articles, free-form documents, social media posts), video (recordings of conferences, presentations, seminars, webinars), audio (call recordings, podcasts, conference discussions), and images (scanned documents, graphs, photographs). The architecture for analyzing each format individually is illustrated in Fig. 3.

The displayed architecture involves a client submitting a query via an interface, which the orchestrator module divides into multiple sub-queries based on data type, then aggregates the responses before returning them to the user. However, this approach is inefficient, as it requires generating multiple queries to different AI modules (text, image, and speech recognition supported by NLP and LLM), each specialized in processing a particular data type. This leads to increased technical complexity, integration challenges, and a higher probability of errors.

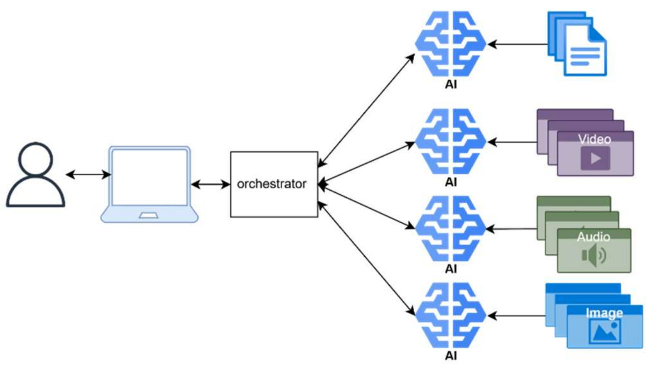

Figure 3. Architecture with subqueries based on data type.

Such an architecture is not optimal because, for each client request, the system must issue several separate queries to different AI modules, each focused on a specific data format. This causes significant drawbacks, including higher system complexity, difficulties in scaling and integration, and an elevated risk of errors arising from numerous module interaction points. Furthermore, once the responses are received, the orchestrator must accurately merge them, adding yet another layer of complexity to the process.

All data types should be converted into a unified format to optimize the analysis process. The most convenient and universal format for this purpose is text. The conversion for each data type can be approached as follows:

- Text data is already in the required format.
- Video content often includes embedded subtitles, and numerous AI-based services are available to generate subtitles if needed.
- Audio content can be transcribed into text using AI transcription services. Platforms like MS Teams and Zoom offer integrated AI transcription tools.
- Images are more complex. While scanned text documents can be processed relatively easily using Optical Character Recognition (OCR), analyzing non-text images, such as graphs, charts, or photographs, presents additional challenges. Although AI systems exist that generate descriptions for images, these non-textual sources are usually supplementary rather than primary and are often accompanied by corresponding text, audio, or video. Thus, in many cases, detailed image analysis provides limited practical value.

The updated architecture, in which all data types are converted into text before further processing, is illustrated in Fig. 4.

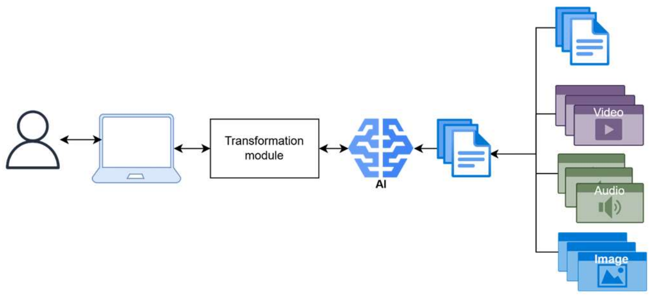

Figure 4. Architecture with subqueries according to the data type.

In this approach, all files are converted into text format and processed by AI with NLP support. The user submits a request through the interface to the processing module, which forwards it to the AI system. After receiving the response, the system returns it to the user. This method reduces potential costs and resource consumption while simplifying the overall architecture, thereby decreasing the likelihood of bugs and implementation errors. By optimizing efficiency and minimizing system complexity, it ensures more reliable performance, easier maintenance, and faster development cycles.

The need for precise transcoding: converting data into text requires high-quality technologies such as speech recognition, OCR, or image description systems to avoid content distortions. This approach is particularly critical when an audio recording contains significant noise, is spoken by a person with speech impairments, or is recorded by someone whose native language is not the language of the recording. Additionally, it is essential to remember that many words sound similar but have entirely different meanings:

- Principle: A fundamental truth or law (e.g., "The principle of compound interest drives growth.").
- Principal: The initial money invested or loaned (e.g., "You earn interest on the principal amount.").
- Capital: Financial resources or assets (e.g., "We need more capital for this investment.").
- Capitol: A building where legislative bodies meet (e.g., "The policy was debated at the state capitol.").
- Loan: Borrowed money to be repaid (e.g., "He took out a loan to start his business.").
- Lown: A less familiar term referring to calmness or quietness (rarely used in finance but phonetically similar).
- Tax: A mandatory financial charge (e.g., "The company must pay corporate tax.").
- Tacks: a strategic approach (e.g., "We need to change tacks in our investment strategy.").

Another important question arises regarding the political bias of responses. Can models like ChatGPT or Gemini provide objective and impartial information, especially for American companies? Could they hide specific facts or present them in a more favorable light? Similarly, it is worth considering how unbiased DeepSeek's responses would be when asked about Chinese companies. Given that, such language models are trained on specific datasets and may be subject to regulatory constraints, their answers could potentially contain hidden or explicit bias.

In the next section, we conduct an experimental study to validate the proposed approach by applying NLP tools to real-world financial video data.

IV. EXPERIMENTAL RESEARCH ON FINANCIAL VIDEO PROCESSING USING NLP

A. THE OBJECTIVE OF THE EXPERIMENT

The primary objective of this experimental study is to evaluate the ability of NLP technologies to automate the analysis of financial video content. The experiment aims to transform unstructured video data into structured information supporting investment decision-making. The focus is on extracting key elements from financial analyses, such as stock price forecasts, risk assessments, and influencing factors. The study also assesses the feasibility of integrating NLP-based automation into financial advisory systems and investment reporting processes.

B. HYPOTHESIS

It is hypothesized that modern NLP models can automatically extract key financial information from video content, transforming unstructured data into a structured format suitable for investment analysis. This approach is expected to significantly reduce video processing time, improve access to financial insights, and enhance the accuracy of stock market forecasts by leveraging aggregated sentiment and risk factors identified in the analyzed content.

C. METHODOLOGY

The experimental methodology was designed to evaluate the ability of NLP technologies to process a newly collected dataset of financial video content. The process consisted of several stages to ensure the quality and reliability of extracted insights.

1) Video Selection Criteria

YouTube was selected as the source of financial video content due to its widespread use among retail and institutional investors. Videos were chosen according to the following criteria:

- Relevance: Videos included content related to the financial analysis of Amazon (AMZN) or Google (GOOG).
- Topic Diversity: While most videos were recent (2024–2025), several older videos (e.g., from 2022) were also included to assess model robustness across different publication periods.
- Search Keywords: Queries such as “Amazon stock analysis,” “Google stock analysis,” and “investment outlook” were used.
- Duration: Videos ranged from 3 to 20 minutes to ensure sufficient informational content.

As a result, 22 videos were selected: 10 focused on Amazon, 10 on Google, 1 Google-related instructional video unrelated to stock analysis, and 1 completely unrelated video.

2) Obtaining Text Transcripts

The Restream "Transcribe Video to Text" service was selected due to its accessibility, support for automatic transcription of publicly available video links, and compatibility with the English-language financial content used in this study. Its ability to handle videos of varying lengths without requiring file downloads or manual segmentation made it suitable for streamlined data preparation within the research workflow.

The videos were transcribed into text using the Restream "Transcribe Video to Text" service [31]. The process included the following steps: Uploading video links to the service; Generating automatic text transcripts; Saving the transcripts in .txt format without additional editing; Verification of Transcription Accuracy.

Manual verification of the transcripts confirmed sufficient accuracy, ensuring the text reliably reflected the video content for subsequent NLP analysis.

3) Text Data Processing Using NLP

The transcripts were processed using the ChatGPT-4 model. A standardized prompt was applied to extract structured investment information. The target outputs included company recognition, investment risk level, stock price forecast, and explanatory notes.

The standardized JSON output format was as follows:

```
{
 "stock": "<stock_ticker>",
 "risk": "high/medium/low",
 "risk_notes": "",
 "price_forecast": "up/same/down",
 "forecast_notes": ""
}
```

4) LLM Prompting Strategy

A standardized prompt was applied to each transcript using the ChatGPT-4 model. The prompt was designed to guide the model in extracting structured investment signals without imposing subjective bias. The exact prompt used was as follows:

"You're a stock advisor video analyzer. I will be giving you a video transcription, and you should extract from this video stock about what is being discussed and provide an analysis of what they are talking about, for example, if in the video they are saying that they forecast pricing up and so on. You should provide it in strict JSON format: {"stock": "<stock_ticker>", "risk": "high/medium/low", "risk_notes": "", "price_forecast": "up/same/down", "forecast_notes": ""}.

The prompt intentionally avoided using sentiment terminology to reduce anchoring bias, instead relying on the model's contextual interpretation of the transcript.

Sentiment classification was derived implicitly from the price_forecast field: an "up" prediction corresponds to positive sentiment, reflecting expectations of growth or favorable market conditions; a "down" prediction corresponds to negative sentiment, indicating anticipated decline or adverse factors; and a "same" prediction corresponds to neutral sentiment, representing an absence of strong directional signals. This mapping aligns with the standard three-class sentiment categorization used in financial NLP research [14], while avoiding the need for a separate sentiment-extraction prompt, which could introduce additional model variance.

D. ANALYSIS OF RESULTS

The experiment results were analyzed against several key criteria, including the accuracy of company and ticker recognition, the accuracy of stock price forecasts, and the assessment of investment risks. Based on the evaluated videos, an aggregated analysis was also conducted to summarize the dominant market sentiment.

It is important to distinguish between two separate dimensions of accuracy evaluated in this study. The first is the NLP extraction accuracy - the system's ability to correctly identify companies, tickers, risk levels, and directional forecasts from raw transcript text. The second is the forecast validation accuracy - the degree to which the extracted forecasts aligned with actual subsequent market movements. The former reflects the quality of the NLP pipeline itself. At the same time, the latter serves as an indirect indicator of the system's practical utility: a pipeline that reliably extracts forecasts from credible financial analysts produces outputs that carry real investment signal value.

1) Company and Ticker Recognition Accuracy

For all 20 financial videos directly related to Amazon (AMZN) or Google (GOOG), the model correctly identified the companies and their corresponding stock tickers. In the two non-relevant cases (one Google-related instructional video and one unrelated video), the system correctly detected the absence of stock-related content. For such cases, the output returned “N/A” for risk and price_forecast, with notes stating: “Transcript does not contain relevant stock-related information.” This shows the system’s robustness in filtering irrelevant inputs.

2) Forecast Accuracy for Stock Prices (Up/Same/Down)

- Amazon (10 videos): Out of 10 forecasts, 9 correctly predicted stock price growth. Only 1 forecast predicted a decline, which proved inaccurate because the actual market trend was upward. This produced an overall accuracy of 90% for Amazon-related videos.
- Google (10 videos): Out of 10 forecasts, 8 correctly predicted stock price movements when the actual outcomes matched, including 7 correct growth forecasts and 1 correct recognition of decline.

The inclusion of older videos (e.g., from 2022) added variability, particularly for Google, where changing macroeconomic conditions and sector dynamics affected validation results.

3) Risk Assessment Results

The model consistently categorized risk levels [32] as high, medium, or low based on the transcript content:

- For Amazon, risks were typically associated with competition in cloud services (AWS vs. Microsoft Azure), employee stock-based compensation, and overall market valuation.
- For Google, risks often reflected concerns about advertising market volatility, regulatory scrutiny, and the impact of competition in artificial intelligence and cloud services.

4) Aggregated Market Sentiment Analysis

In this study, sentiment was not extracted through a dedicated classification prompt. Instead, it was derived from the price_forecast field in the structured JSON output: "up" forecasts were interpreted as positive sentiment, "down" as negative sentiment, and "same" as neutral sentiment. This approach ensures consistency between the extracted forecast signal and the sentiment label, eliminating potential discrepancies that could arise from running two separate prompts on the same transcript.

The aggregated analysis of the dataset revealed the following:

- Amazon: The dominant sentiment was strongly optimistic, supported by high accuracy in forecast validation (90%).
- Google: Sentiment was also predominantly positive, but the forecast validation revealed a 20% error rate, underscoring the challenges of extracting consistent predictive signals across different companies and timeframes.
- Irrelevant videos: The system correctly identified that no financial content was present, returning neutral outputs with N/A in risk and forecast fields.

The summarized results of the NLP-based analysis for the selected videos are presented in Table 1.

Table 1. Results of NLP-based Analysis of Financial Videos

Date	Video #	Ticker	Risk	Price Forecast	Actual change
02/02/2024	Video 1	AMZN	Medium	Up	up
02/02/2024	Video 2	AMZN	Medium	Up	up
26/02/2024	Video 3	AMZN	Medium	Up	up
26/02/2024	Video 4	AMZN	Medium	Up	up
07/04/2024	Video 5	AMZN	High	Down	up
09/04/2024	Video 6	AMZN	Medium	Up	up
23/06/2024	Video 7	AMZN	Low	Up	up
04/07/2024	Video 8	AMZN	Medium	Up	up
05/07/2024	Video 9	AMZN	Medium	Up	up
15/08/2024	Video 10	AMZN	Low	Up	up
12/05/2024	Video 11	GOOG	Medium	Up	up
24/08/2024	Video 12	GOOG	Medium	Same	Same
18/03/2024	Video 13	GOOG	Medium	Up	up
21/02/2025	Video 14	GOOG	Medium	Same	up
27/11/2024	Video 15	GOOG	Medium	Up	up
2/12/2024	Video 16	GOOG	Medium	Up	up
25/11/2022	Video 17	GOOG	Medium	Up	up
13/05/2023	Video 18	GOOG	Medium	Up	up
04/02/2025	Video 19	GOOG	Medium	Same	down
11/10/2024	Video 20	GOOG	Medium	Up	up

Future research will consider the application of multimodal sentiment analysis models that simultaneously process text, audio, and visual data to improve accuracy and capture nonverbal signals [33].

E. VALIDATION OF FORECASTS

The forecasts extracted from the financial video content were validated by comparing them with the actual stock price movements over the year following each video's publication.

The validation process included the following steps:

1) Collection of Actual Stock Price Data

Historical stock price data was collected from Yahoo Finance [34] to establish the ground truth for validation.

2) Classification of Real Outcomes

Each forecast was classified based on the actual price movement observed between the video publication date and 3 and 6 months later.

The classification criteria were Up – if the daily growth rate remained positive; Down – if the daily growth rate was negative.

3) Evaluation Metrics

The quality of the forecasts was evaluated using standard classification metrics:

- Accuracy – the proportion of correct forecasts relative to the total number of forecasts.
- Precision – the ratio of true positive predictions to the total predicted positives.
- Recall – the ratio of true positive predictions to all actual positive outcomes.
- F1-Score – the harmonic mean of precision and recall.

The results showed that out of 20 forecasts, 16 “up” predictions and 1 “same” prediction matched the real stock movements. In total, 17 out of 20 correct forecasts resulted in an overall accuracy of 85%.

4) Confusion Matrix

Based on the comparison between the forecasted and actual stock price movements, the confusion matrix was constructed and is presented in Table 2.

Table 2. Confusion Matrix for Forecast Validation

	Actual Up	Actual Down	Actual Same
Predicted Up	16	0	0
Predicted Down	1	0	0
Predicted Same	1	1	1

The matrix shows that in 16 cases, the forecasts correctly predicted an increase in the stock price; in 1 case, they correctly predicted no movement; in 2 cases, they predicted no movement, but the stock went up; and in 1 case, they predicted a downward movement, but the stock went up.

TP_total = 16+0+1=17
FP_total = 0+1+2=3
FN_total = 3+0+0=3
Precision = TP / (TP + FP)
Precision = 17/ (17+3) = 0.85
Recall = TP / (TP + FN)
Recall = 17 / (17+3) = 0.85
F1 = 2 * (Precision * Recall) / (Precision + Recall)
F1 = 2 * (0.85 * 0.85) / (0.85 + 0.85) = 0.85

F. METHOD LIMITATIONS

Although the experimental results demonstrated the feasibility and efficiency of using NLP technologies for financial video analysis, several limitations of the proposed approach were identified:

First, transforming multimedia content into plain text results in the loss of crucial non-verbal information, such as emotional intonations, sarcasm, and visual context, which can significantly affect the interpretation of financial messages. Second, the accuracy of automatic transcription remains a critical factor. External conditions such as background noise, speaker accents, and specialized financial terminology can introduce distortions in the transcribed text.

Third, current NLP models, including LLMs like ChatGPT4o, demonstrate a limited capacity for deep contextual understanding. Complex or ambiguous financial discussions may be incorrectly interpreted if key contextual cues are missing.

Fourth, the risk of bias in AI models must be considered. LLMs may inherit biases from their training datasets, which can subtly affect the neutrality and accuracy of financial risk assessments and forecasts [30].

Finally, the system currently does not incorporate visual data analysis. Information in graphs, charts, and on-screen financial indicators remains unprocessed, leading to incomplete conclusions.

Additionally, a notable limitation of this study is the relatively small size of the experimental dataset, comprising only 22 videos covering two companies (AMZN and GOOG). While sufficient to demonstrate proof of concept, this scale limits the generalizability of the findings. Future research should expand the dataset to include a broader range of stocks across different sectors and market capitalizations, extend the time horizon beyond 2022–2025 to capture diverse market cycles, and incorporate multilingual video content to assess the pipeline's robustness across non-English financial media.

Future enhancements should focus on overcoming these limitations by developing multimodal models that can jointly process text, audio, and visual content to improve the quality and reliability of automated financial analysis [33].

G. AUTHORS' CONTRIBUTIONS

In contrast to prior studies that primarily focus on textual sources, such as earnings reports or financial news articles, our research investigates a less explored data type: financial video content. While earlier work has demonstrated the potential of NLP for sentiment analysis of tweets and structured corporate disclosures, few studies have examined the feasibility of extracting structured investment insights directly from multimedia sources.

In this experiment, we proposed and validated a fully automated pipeline for processing 22 YouTube videos, including financial analyses of Amazon (AMZN) and Google (GOOG), as well as two non-relevant cases. The system integrated transcription, NLP-based processing, and structured output generation, without manual intervention.

Unlike most theoretical approaches, the extracted forecasts were empirically validated against actual market performance. The results confirmed high reliability of Amazon forecasts (90% accuracy) and moderate accuracy of Google forecasts (80%), highlighting the model's strengths and limitations when applied across different companies and timeframes. The system also correctly handled irrelevant content by returning N/A outputs, demonstrating robust filtering of non-financial material.

Our contributions go beyond simple sentiment classification or keyword extraction. The system simultaneously identified multiple financial signals:

- company and ticker recognition (with 100% accuracy)
- investment risk categorization (high, medium, low)
- directional stock price forecasts,
- explanatory notes for each output,
- correct classification of non-relevant transcripts.

This multidimensional analysis provides investors with a more comprehensive view than traditional one-dimensional sentiment scores. Furthermore, by including both recent (2024–2025) and older (e.g., 2022) videos, the experiment also tested the robustness of NLP-based forecasting across different periods, making the study more realistic.

The architecture was designed for scalability, supporting multilingual content and batch processing of larger datasets. This offers a viable alternative to labor-intensive manual analysis, paving the way for integrating NLP-driven video analytics into financial advisory systems and decision-support platforms.

Collectively, these contributions distinguish our study by showing how state-of-the-art NLP technologies can transform diverse, partially irrelevant multimedia content into structured, actionable intelligence – while also revealing the limitations that must be addressed in future research.

To verify the changes identified using NLP, we will compare the results obtained with classic ARIMA models used for forecasting time series, and select the highest-quality forecast model using the Akaike criterion. Based on the video forecasts' dates, we will select six months to build the forecast (up to the forecast date) and a six-month period to verify its quality (after the forecast date). Six months were chosen to avoid traders' short-term speculative motives. The results obtained are presented in Table 3.

Table 3. ARIMA for Forecast Validation of AMZN

Previous Period	Prediction Period	Model	Price Forecast	Actual change	Z-test*
06/08/2023 02/02/2024	02/02/2024 31/07/2024	ARIMA(0,1,0) AIC=646.67	Same (video 1,2)	up	z=34.43*
30/08/2023 26/02/2024	26/02/2024 24/08/2024	ARIMA(0,1,0) AIC=587.25	Same (video 3,4)	up	z=0.973
10/10/2023 07/04/2024	07/04/2024 04/10/2024	ARIMA(0,1,2) with drift AIC=573.73	Up (video 5)	up	z=19.81*
12/10/2023 09/04/2024	09/04/2024 06/10/2024	ARIMA(0,1,2) with drift AIC=570.18	Up (video 6)	up	z=19.39*
26/12/2023 23/06/2024	23/06/2024 20/12/2024	ARIMA(0,1,2) with drift AIC=592.53	Up (video 7)	up	**z= -2,837***
06/01/2024 04/07/2024	04/07/2024 31/12/2024	ARIMA(2,1,0) with drift AIC=579.34	Up (video 8)	up	z=10.51*
07/01/2024 05/07/2024	05/07/2024 01/01/2025	ARIMA(2,1,0) with drift AIC=579.34	Up (video 9)	up	z=10.61*
17/02/2024 15/08/2024	15/08/2024 11/02/2025	ARIMA(2,1,2) AIC=611.27	Same (video 10)	up	z=13.27*

*The z-test confirms a statistically significant change in the actual share price (α=0.05 in a one-tailed test). The result of the ARIMA forecast check, which was not confirmed and turned out to be the opposite, is highlighted in bold.

The results show that the main limitation of ARIMA models is their inability to forecast an unchanged exchange rate when NLP predicts an increase in the share price, yet the exchange rate actually rises. At the same time, during one period, the ARIMA forecast proved more accurate than the NLP forecast (Fig. 5). After April 7, 2024, the AMZN share price increased, whereas NLP made the opposite prediction.

Table 4. ARIMA for Forecast Validation of GOOG

Previous Period	Prediction Period	Model	Price Forecast	Actual change	Z-test*
14/11/2023 12/05/2024	12/05/2024 08/11/2024	ARIMA(0,1,0) AIC=673.87	Same (video 11)	up	**z= -4.271***
26/02/2024 24/08/2024	24/08/2024 20/02/2025	ARIMA(0,1,0) AIC=619.83	Same (video 12)	up	z=7.46*
20/09/2023 18/03/2024	18/03/2024 14/09/2024	ARIMA(0,1,0) AIC=572.56	Same (video 13)	up	z=30.91*
25.08.2024 21.02.2025	21.02.2025 20.08.2025	ARIMA(0,1,0) AIC=636.28	Same (video 14)	up	**z= -3.055***
31.05.2024 27.11.2024	27.11.2024 26.05.2025	ARIMA(0,1,2) AIC=596.76	Same (video 15)	up	z=3.78*
05.06.2024 02.12.2024	02.12.2024 31.05.2025	ARIMA(0,1,2) AIC=592.7	Same (video 16)	up	z=3.861*
29.05.2022 25.11.2022	25.11.2022 24.05.2023	ARIMA(0,1,0) AIC=597.17	Same (video 17)	up	**z= -7.684***
14.11.2022 13.05.2023	13.05.2023 09.11.2023	ARIMA(0,1,0) AIC=547.75	Same (video 18)	up	z=43.03*
08.08.2024 04.02.2025	04.02.2025 03.08.2025	ARIMA(0,1,0) AIC=612.91	Same (video 19)	up	**z= -2.367***
14.04.2024 11.10.2024	11.10.2024 09.04.2025	ARIMA(0,1,0) AIC=613.77	Same (video 20)	up	z=6.91*

*The z-test confirms a statistically significant change in the actual share price (α=0.05 in a one-tailed test). The result of the ARIMA forecast check, which was not confirmed and turned out to be the opposite, is highlighted in bold.

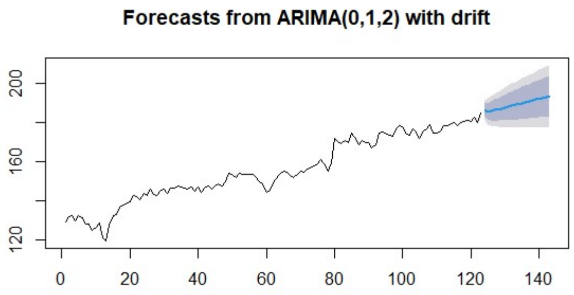

Figure 5. Forecast from ARIMA(0,1,2), period 10.10.2023-07.04.2024.

For NLP and ARIMA models, short-term forecasts are generally more accurate, but in the medium term, they may become less accurate. At the same time, the number of false positives in NLP is lower than in traditional time-series forecasting methods, such as ARIMA models, demonstrating the potential of combining structured and unstructured data processing methods in financial forecasting.

V. CONCLUSION

This paper presents an experimental study of the application of NLP technologies to the analysis of financial video content. The proposed approach transformed unstructured multimedia data into a unified text format, enabling automated extraction of key investment insights such as stock price forecasts, risk assessments, and company recognition.

The experiment was based on 22 YouTube videos published between 2022 and 2025, including 10 on Amazon (AMZN), 10 on Google (GOOG), and 2 non-relevant cases. The results demonstrated: 100% accuracy in company and ticker recognition, confirming the reliability of entity extraction; high forecast accuracy for Amazon videos (90%), where 9 of 10 predictions were correct; moderate accuracy for Google videos (80%), with 8 correct predictions and higher variability caused by changing market conditions and the inclusion of older content; correct handling of irrelevant videos, where the system returned N/A values and explanatory notes, showing robustness in filtering non-financial material.

When comparing forecasts against actual stock price movements, the system achieved 85% overall accuracy (17 correct out of 20), confirming the feasibility of using NLP-driven automation for financial video analysis. These findings highlight the potential of such systems to support investment decision-making while also revealing limitations: difficulty handling long-term market volatility, imperfect interpretation of neutral predictions, and a lack of multimodal processing of visual cues, such as charts and non-verbal signals. The NLP approach outperforms ARIMA-based time-series price predictions for the considered financial instruments.

The novelty of this research lies in demonstrating that NLP can extract structured, validated financial forecasts not only from recent multimedia sources but also across different timeframes, including older market analyses. The inclusion of irrelevant content further confirmed the robustness of the methodology.

Future research should focus on developing multimodal models that combine text, audio, and visual signals, and on expanding validation to larger, more diverse datasets. Such improvements are expected to enhance contextual accuracy, reduce bias, and increase the reliability of automated financial forecasting based on unstructured media sources.

DECLARATION ON GENERATIVE AI

During the preparation of this article, the authors used Grammarly for grammar and spelling checks, Paraphrase and rewording, and ChatGPT for sentiment analysis of selected financial videos. After using these tools/services, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] F. Duarte, "Amount of Data Created Daily (2024)," *Exploding Topics*, Feb. 23, 2025. https://explodingtopics.com/blog/data-generated-per-day.

[2] T. Tian, R. Cooper, A. Vasilakos, J. Deng, Q. Zhang, "From Data to Strategy: A Public Market Framework for Competitive intelligence," *Expert Systems With Applications*, vol. 296, 129061, 2025. https://doi.org/10.1016/j.eswa.2025.129061.

[3] J. Chou, K. Lin, T. Pham, "AI-fused construction portfolio investment system with risk hedging using machine learning and long-short strategies," *Applied Soft Computing*, vol. 183, 113555, 2025. https://doi.org/10.1016/j.asoc.2025.113555.

[4] K. Lipianina-Honcharenko, C. Wolff, A. Sachenko, O. Desyatnyuk, S. Sachenko, I. Kit. "Intelligent information system for product promotion in internet market," *Applied Sciences*, vol. 13, issue 17, 9585, 2023. https://doi.org/10.3390/app13179585.

[5] N. Maslej, L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J.C. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlasci, J.B. Lotufo, A. Rome, A. Shi, S. Oak, "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, *Stanford University*, Stanford, CA, April 2025. https://doi.org/10.48550/arXiv.2504.07139.

[6] Y. Zhang, Y. Li, W. Wang, "Comparative analysis of investor sentiment and attention in the U.S. and Chinese stock markets: Implications for financial stability," *International Review of Financial Analysis*, vol. 109, 104751, 2025. https://doi.org/10.1016/j.irfa.2025.104751

[7] S. Feuerriegel, J. Gordon, "Long-term stock index forecasting based on text mining of regulatory disclosures," *Decision Support Systems*, vol. 112, pp. 88–97, 2018. https://doi.org/10.1016/j.dss.2018.06.008.

[8] M. Abdullah, Z. Sulong, M.A.F. Chowdhury, "Explainable deep learning model for stock price forecasting using textual analysis," *Expert Systems With Applications*, vol. 249, 123740, 2024. https://doi.org/10.1016/j.eswa.2024.123740.

[9] S. S. Mehrkian, H. Davari-Ardakani, "An integrated model of sentiment analysis and quantitative index data for predicting stock market trends: A case study of the Tehran stock exchange," *Expert Systems with Applications*, vol. 269, 126298, 2024. https://doi.org/10.1016/j.eswa.2024.126298.

[10] S. García-Méndez, F. De Arriba-Pérez, A. Barros-Vila, F. J. González-Castaño, "Targeted aspect-based emotion analysis to detect opportunities and precaution in financial Twitter messages," *Expert Systems with Applications*, vol. 218, 119611, 2023. https://doi.org/10.1016/j.eswa.2023.119611.

[11] J. Jiang, K. K. Bandeli, K. Srinivasan, "Dynamic model selection in enterprise forecasting systems using sequence modeling," *Decision Support Systems*, vol. 193, 114439, 2025. https://doi.org/10.1016/j.dss.2025.114439.

[12] V. Iatsiuta, V. Kobets, "Transforming business communication with solutions based on artificial intelligence technologies with support for natural language processing," *Science and Innovation*, vol. 21, issue 5, pp. 126–143, 2025. https://doi.org/10.15407/scine21.05.126.

[13] O. Ivanov, V. Kobets, "Integrating semantic analysis and financial indicators of business reports for predicting stock prices," In *Communications in Computer and Information Science*, 2025, pp. 297–308, 2025. https://doi.org/10.1007/978-3-031-81372-6_22.

[14] O. Ivanov, V. Kobets, "Future financial impact analysis from sentiment and indicators analysis," *Computational Economics*, vol. 66, pp. 4959–4985, 2025. https://doi.org/10.1007/s10614-025-10891-7.

[15] P. K. Naik, J. Soni, "Asymmetric and temporal effects of investor sentiment and institutional trading behavior on returns and volatility: evidence from the Indian stock market," *Review of Behavioral Finance*, vol. 18, no. 2, pp. 186-214, 2026. https://doi.org/10.1108/rbf-09-2025-0387.

[16] T. Boufateh, Z. Saadaoui, Z. Jiao, "On the time-varying responses of Fintech stock returns to geopolitical, financial and market sentiment shocks," The *Quarterly Review of Economics and Finance*, vol. 101, 101951, 2025. https://doi.org/10.1016/j.qref.2024.101951

[17] V. Iatsiuta, V. Kobets, "Generalized AI-based solutions with NLP support for processing business requests," In *Communications in computer and information science* pp. 233–245, 2025. https://doi.org/10.1007/978-3-031-81372-6_17.

[18] I-C. Chiu, M.-W. Hung, "Finance-specific large language models: Advancing sentiment analysis and return prediction with LLaMA 2," *Pacific-Basin Finance Journal*, vol. 90, 102632, 2025. https://doi.org/10.1016/j.pacfin.2024.102632.

[19] A. H. Huang, H. Wang, and Y. Yang, "FinBERT: A large language model for extracting information from financial text," *Contemporary Accounting Research*, vol. 40, no. 2, pp. 806-841, 2022. https://doi.org/10.1111/1911-3846.12832.

[20] O. Romanko, A. Narayan, and R. H. Kwon, "ChatGPT-based investment portfolio selection," *Operations Research Forum*, vol. 4, no. 4, 2023. https://doi.org/10.1007/s43069-023-00277-6.

[21] K. Kirtac and G. Germano, "Sentiment trading with large language models," *Finance Research Letters*, vol. 62, pp. 105227–105227, 2024. https://doi.org/10.1016/j.frl.2024.105227.

[22] F. Sayyed, R. Argiddi and S. Apte, "Generating recommendations for stock market using collaborative filtering," *Int. J. Comput. Eng. Sci.*, vol. 3, pp. 46-49, 2013.

[23] Q. Liu, X. Wang, and Y. Du, "The weekly cycle of investor sentiment and the holiday effect – An empirical study of Chinese stock market based on natural language processing," *Heliyon*, vol. 8, no. 12, e12646, Dec. 2022. https://doi.org/10.1016/j.heliyon.2022.e12646.

[24] J.M. Oliveira, P. Ramos, J.M. Oliveira, P. Ramos, "Evaluating the effectiveness of time series transformers for demand forecasting in retail," *Mathematics*, vol. 12, 2728, 2024. https://doi.org/10.3390/MATH12172728.

[25] Y. Ensafi, S.H. Amin, G. Zhang, B. Shah, "Time-series forecasting of seasonal items sales using machine learning – a comparative analysis," *International Journal of Informational Management Data Insights*, vol. 2, issue 1, 100058, 2022. https://doi.org/10.1016/j.jjimei.2022.100058.

[26] L. Feng and A. Sinchai, "Deep context-attentive transformer transfer learning for financial forecasting," *Peer Journal of Computer Science*, vol. 11, e2983, 2025. https://doi.org/10.7717/peerj-cs.2983.

[27] S. W. K. Chan and M. W. C. Chong, "Sentiment analysis in financial texts," *Decision Support Systems*, vol. 94, pp. 53-64 2017. https://doi.org/10.1016/j.dss.2016.10.006.

[28] D. Tiwari *et al.*, "Attention-augmented hybrid CNN-LSTM model for social media sentiment analysis in cryptocurrency investment decision-making," *Springer Nature*, vol. 15, 33201, 2025. https://doi.org/10.1038/s41598-025-18245-x.

[29] A. Erfani and H. Khanjar, "Large language models for construction risk classification: A comparative study," *Buildings*, vol. 15, no. 18, p. 3379, 2025. https://doi.org/10.3390/buildings15183379.

[30] D. Noor Mohammadzadeh Maleki, S. Abbaskhani, A. Farhadi, A. Zamanifar, "Artificial intelligence methods for financial market prediction: A systematic review," *Computers & Electrical Engineering*, 134, 111110, 2026. https://doi.org/10.1016/j.compeleceng.2026.111110

[31] "Create and Multistream Live Video," *Restream*. [Online]. Available at: https://restream.io.

[32] O. I. Beley, S. I. Sachenko, "The information system of control risks," In *Proceedings of the International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*. IDAACS'2001 (No.01EX510), pp. 270-274, 2001. https://10.1109/IDAACS.2001.942029.

[33] P. Q. Dao, M. Roantree, T. B. Nguyen-Tat, and V. M. Ngo, "Exploring multimodal sentiment analysis models: A comprehensive survey," Aug. 2024, https://doi.org/10.20944/preprints202408.0127.v1.

[34] "Amazon.com, Inc. (AMZN) Stock Historical Prices & Data - Yahoo Finance," [Online]. Available at: https://finance.yahoo.com/quote/AMZN/history.

***VLAD IATSIUTA** received his Master's degree in Informatics from Kherson State University in 2010 and his Master's degree with honors in Information Systems and Technologies in 2023. From 2008 to 2010, he worked as a developer at the Laboratory of Integration of Pedagogical Software. Since 2010, he has been working at DataArt, where he has advanced to the positions of Senior Software Developer, Team Lead, and Architect. He has more than 15 years of professional experience in software development across various business domains, including travel, finance, and healthcare.*

***OLEKSII IVANOV** received his Master's degree in Software Engineering from Kherson State University in 2018 and his Master's degree in Information Systems and Technologies in 2023. From 2015 to 2016, he worked as a developer at Kherson State University. Since 2018, he has worked at various IT companies in the position of Senior Software Developer. He has over 8 years of professional experience in software development across different business domains.*

***VITALIY KOBETS** has been working at Kherson State University and has risen from Assistant Professor to Professor. Since 2019, he has been a Professor of Computer Science and Software Engineering Department at Kherson State University.*

He has published over 80 peer-reviewed publications in international journals and conferences, indexed by Scopus and WoS.

Research interests include evolutionary microeconomics, robo-advisors, business analytics, machine learning, and financial technology.

He is a member of the International Association for Technological Development and Innovations IATDI, a permanent specialized academic council member (Kyiv), a member of 3 editorial boards, a program committee member of International Conferences, and a reviewer for International Journals.