Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Check for updates

A hierarchical review of multi-level sentiment analysis in financial texts: Advances in contextual embeddings and deep learning

Seun Ebiesuwa [1], Uzodinma Onwuchekwa C [1] and Funmilayo Sanusi A [2, *]

[1] *Department of Computer Science, Babcock University, Ilisan-Remo, Ogun State, Nigeria.*
[2] *Department of Software Engineering, Babcock University, Ilisan-Remo, Ogun State, Nigeria.*

Global Journal of Engineering and Technology Advances, 2026, 26(03), 114-127

Publication history: Received on 25 January 2026; revised on 08 March 2026; accepted on 11 March 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.26.3.0053

Abstract

People depend more than ever on unstructured financial texts, news stories, earnings calls, regulatory filings, and even social media to make decisions. That's pushed sentiment analysis into the spotlight. But old-school sentiment analysis just flattens everything out and misses the layers and nuance in financial language. Now, with advances in deep learning and contextual embeddings, researchers can dig into sentiment at different levels of text, capturing way more detail. This review pulls together what's new in hierarchical and multi-level sentiment analysis for financial texts, zeroing in on contextual embedding models and the latest deep learning setups. We systematically looked through the last five years of peer-reviewed studies, collecting papers from top databases like Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and SpringerLink. We only included studies that focused on hierarchical sentiment modeling, contextual embeddings, and deep learning within financial text analysis. Then we broke down the data, comparing models, embeddings, datasets, and how people measured results. What stands out? There's a big move away from lexicon-based and classic machine learning approaches transformer-based contextual embeddings now lead the way. Finance-focused models like FinBERT and Financial-RoBERTa are especially strong. Hierarchical and hybrid architectures almost always beat their flat counterparts, mainly because they keep track of sentiment connections across different levels and clear up context much better. Still, there are hurdles: adapting models to new domains, making them explainable, scaling them up, and the fact that there aren't standard benchmarks to compare results across different levels. This review points out why hierarchical modeling and contextual embeddings matter for financial sentiment analysis. It also calls out what's missing and where researchers should go next like building unified, explainable, and efficient sentiment systems. Bottom line: these insights give future researchers and anyone building AI-driven financial tools a solid place to start.

Keywords: Sentiment Analysis; Financial Texts; Deep Learning; Contextual Embedding; Artificial Intelligence

1. Introduction

1.1. Background and Motivation

Financial markets run on information, and a lot of that info comes wrapped in natural language—full of hints about what people expect, what worries them, and where the risks might be. That's why sentiment analysis has become such a valuable tool for financial decision-making. People use it for everything from event-driven trading and tracking portfolio risk to digging into corporate filings and keeping an eye on the market. Reviews in finance have made it clear: sentiment pulled from text adds something extra to the usual numbers, surfacing forward-looking beliefs and the way stories are told, stuff you just don't see in prices or balance sheets (Modak, 2023; Todd et al., 2024). There's meta-analytic evidence showing that text-based sentiment actually lines up, in a statistically significant way, with returns and

* Corresponding author: Sanusi Funmilayo A.

broader market moves. Of course, results change depending on where the data comes from, how long you look, and what models you use (Gric, 2021).

Lately, there's a ton of interest in building stronger sentiment models, and that's partly because there's just so much more text out there now. Beyond the old-school financial news, people are analyzing everything from earnings call transcripts and regulatory filings to analyst notes and even social media or investor forums. Each source has its own style and timing. For example, there are now big, open datasets of earnings call transcripts that researchers use to connect sentiment with real market outcomes (Roozen & Lelli, 2021). Surveys highlight how the explosion of text sources is pushing the field toward language models built specifically for finance, along with tougher benchmarks across mixed datasets (Du et al., 2024). More text means more opportunities to spot useful signals, but it also brings headaches like shifting language, noisy data, and inconsistencies from one source to the next. Even with all this progress, a lot of sentiment analysis in finance still feels pretty basic. Many systems stick with simple document-level labels, basic sentence polarity, or dictionary scores that just don't capture the layered, complex way financial language actually works. Real financial sentiment lives at several levels: word choices, sentence structure, which entities or aspects are mentioned ("revenue guidance," "liquidity"), and the broader story being told in a document. Research shows that in finance, sentiment depends a lot on domain-specific language and how words fit together. The usual lexicons and shallow models miss this (Xing et al., 2020). That's why the field is moving toward deeper methods: contextual embeddings and transformer models fine-tuned for finance, which can pick up on the subtle meanings in specialized texts (Liu et al., 2020). At the same time, aspect-aware and hierarchical models treat sentiment as something you infer at multiple levels, not just a single label. This is crucial, especially when decisions hang on exactly what the sentiment is about and where it pops up in the text (Lengkeek et al., 2023). All things considered, this review aims to pull together the latest advances and show how these multi-level, context-driven sentiment models can do what flat, old-school methods can't in the world of finance.

1.2. Hierarchical and Multi-Level Sentiment Analysis

Hierarchical sentiment modeling isn't just about slapping a single label on a piece of text. Instead, it digs deeper, treating sentiment as something layered, built from the ground up, word by word, phrase by phrase, all the way to the full document. This really matters in finance. Here, meaning doesn't show up all at once; it builds slowly, starting with a few loaded words or a carefully crafted phrase, then expanding to the way sentences are framed, and finally shaping the whole story that a document tells. Thanks to recent advances in deep learning, we can capture these different layers. These models pick up on tiny signals in the text, weaving them into bigger-picture sentiment judgments (Du et al., 2024; Lengkeek et al., 2023). Hierarchical systems usually stack different components or use attention mechanisms that mirror how text is organized. That way, what the model finds at the word or phrase level actually informs what it concludes at higher levels. Of course, the usefulness of this kind of multi-layered sentiment analysis depends on what you're after. At the word or phrase level, sentiment often comes down to financial jargon terms like "downgrade" or "liquidity squeeze" that carry specific meanings you won't find outside finance. When you zoom in on sentences, you start to see more localized opinions and stances, think of the back-and-forth in earnings calls or analyst reports. Aspect-level sentiment goes even further, zeroing in on how people feel about specific things like revenue growth, risk, or management's outlook. This kind of detail is gold for investment analysis (Lengkeek et al., 2023). Step back and look at the whole document, and you're dealing with its overall tone. Studies show that, when you get this right, you can actually link it to real-world outcomes like unusual stock returns or market volatility (Xing et al., 2020). Pull back even more, and you get market-level sentiment. Here, you're blending signals from all over different sources, different times to get a sense of the market's collective mood. Researchers in macro-finance and behavioral finance lean on this kind of analysis all the time (Goodell & Huynh, 2020).

All in all, these hierarchical and multi-level approaches are a natural fit for finance. They let us match the way we read and make decisions layer by layer, from the tiniest detail to the biggest trend.

1.3. Rise of Contextual Embeddings and Deep Learning

Sentiment analysis in finance has changed a lot over the years. At first, people leaned on simple tools think hand-built word lists like Loughran McDonald or basic machine learning models using bag-of-words or TF-IDF. These early methods were easy to explain, but they just couldn't handle all the layers and shifting meanings packed into financial language. Words like "liability," "volatility," or "capital" don't always mean the same thing; their sentiment flips depending on the context, and those old systems usually missed that. Lately, though, things have taken a big leap forward. Contextual embedding models, especially those built on transformers have changed the game. Models like BERT, and finance-specific versions like FinBERT, don't just spit out one meaning for a word. They actually adjust depending on the surrounding text. That means they pick up on structure, nuance, and all the quirks of financial talk, making them far better at figuring out the real sentiment behind earnings calls, analyst notes, or news articles.

But the shift isn't just about scoring higher on benchmarks. Deep learning models, like transformers, hierarchical attention networks, and recurrent neural networks, can dig into text at multiple levels (Alsowaidi, 2024; Singh, Kumar & Gupta, 2024). They pick up on the flow and structure that's so typical in financial writing. You can also fine-tune them on industry-specific data, so they end up absorbing domain knowledge that used to require a ton of manual work. Studies back this up: these deep, contextual models consistently outshine the old classics, especially when dealing with long documents or the subtle, often implicit tone shifts you find in regulatory filings and forward-looking statements.

1.4. Objectives and Contributions of the Review

This review depicts the latest breakthroughs in hierarchical and multi-level sentiment analysis for financial texts, with a sharp eye on how contextual embeddings and deep learning are changing the game. Natural language processing keeps evolving fast, and financial decision-making leans more and more on unstructured data, so it's about time for a clear, critical look at what's out there. If you're a researcher or practitioner trying to get a grip on where financial sentiment analysis stands and where it's headed, this review aims to be your go-to guide. The paper also indicates how contextual embeddings and deep learning work when it comes to financial sentiment analysis. Instead of just listing which models perform best, this review gets into the nuts and bolts, looking at how choices like transformer-based embeddings, hierarchical attention networks, or hybrid setups shape things like meaning, scalability, and how easy it is to interpret results. There's a real focus on comparing general language models to ones trained specifically on finance, figuring out what each does well, where they fall short, and which situations they fit. This approach helps explain why some models pick up on subtle financial sentiment while others miss the mark.

The third contribution is about spotting what's missing in current research. By comparing recent studies, the review calls out big challenges like handling domain shifts, keeping up with changing concepts, making AI explainable in strict regulatory settings, and dealing with the lack of standard tests for multi-level sentiment. It also points to where the field should head next: things like unified frameworks, combining multiple types of data, and building explainable AI that works within financial regulations. All together, these insights make the review both a solid overview and a push forward, helping researchers get a better handle on how to tackle sentiment analysis in finance.

1.5. Review Methodology

This study takes a systematic scoping review approach to map out, evaluate, and pull together the latest research on hierarchical and multi-level sentiment analysis in financial texts. The focus is on how contextual embeddings and deep learning have shaped this field. By mixing systematic and scoping review methods, the study goes after two main goals: first, to find and assess peer-reviewed research in a clear, repeatable way; and second, to give an overview of new themes, popular methods, and gaps in the research as this area keeps changing fast. A plain systematic review usually sticks to narrow questions. Here though, adding the scoping part lets us cover a broad range of modeling styles, data types, and analytic levels that show up in today's financial sentiment analysis. The review only looks at studies from the past five years, since that's when transformer-based embeddings and advanced deep learning really took over in natural language processing.

The review includes research on sentiment analysis at different levels word, sentence, aspect, document, and even whole market level across financial settings like news, earnings calls, regulatory filings, analyst reports, and investor commentary. It brings in both methods papers and applied studies, as long as they deal directly with hierarchical or multi-level sentiment modeling.

Inclusion criteria for the review are as follows:

- Peer-reviewed journal articles and full conference papers published in English;
- Studies that focus on financial or economic text data;
- Research employing contextual embeddings, deep learning, or hierarchical modeling techniques for sentiment analysis; and
- Papers that report empirical evaluation, conceptual frameworks, or comparative analysis relevant to multi-level sentiment interpretation.

Exclusion criteria include:

- Studies published before the defined temporal window;
- Works focused solely on non-financial domains;

- Papers relying exclusively on lexicon-based or shallow machine learning methods without hierarchical or contextual modeling; and
- Non-scholarly sources such as editorials, short abstracts, preprints without peer review, or opinion pieces.

By clearly defining the review design and scope, this study ensures methodological transparency while enabling a comprehensive and structured synthesis of contemporary research in hierarchical financial sentiment analysis.

1.6. Literature Search Strategy

We took a careful, transparent approach to find studies on hierarchical and multi-level sentiment analysis in financial texts. Our main interest was in research using contextual embeddings and deep learning. To make sure we didn't miss anything important, we checked several top academic databases. We picked these because they cover a lot of ground journals and conference papers in computer science, AI, information systems, and computational finance.

We used Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and SpringerLink. Scopus and Web of Science helped us gather journal articles from different fields and keep track of citations. IEEE Xplore and ACM Digital Library were great for the latest conference papers and technical research in machine learning and NLP. SpringerLink gave us access to both journal articles and book chapters relevant to financial text mining and sentiment analysis. By pulling from all these sources, we built a solid, wide-ranging foundation for our review.

For the search itself, we put together targeted keyword strings, using Boolean logic to balance between finding enough studies and keeping them relevant. We grouped the keywords into three main themes: sentiment analysis, the financial domain, and modeling techniques. Some example searches we ran:

- ("sentiment analysis" OR "opinion mining") AND ("financial text" OR "financial news" OR "earnings call" OR "analyst report")
- ("hierarchical sentiment" OR "multi-level sentiment" OR "aspect-based sentiment") AND ("finance" OR "stock market")
- ("contextual embeddings" OR "transformer models" OR "BERT" OR "FinBERT") AND ("financial sentiment")
- ("deep learning" OR "neural networks") AND ("financial text mining")

AND and OR techniques were used to fine-tune results, and where possible, database filters were also used to limit papers to English, peer-reviewed publications from a specific time frame. We also went through the reference lists of selected articles by hand, just in case we missed any key studies with our initial search. This approach helped us gather a thorough, relevant set of literature for the review.

1.7. Study Selection Process

We picked studies for this review with a clear, step-by-step approach, aiming to include only the most relevant and high-quality work. After searching the selected databases, we dumped all the results into a reference management system. First thing duplicates had to go. Once we cleaned those out, we moved on to a multi-stage screening process to narrow things down even more. First, we checked the titles. If a study obviously had nothing to do with sentiment analysis, financial or economic text data, or computational and machine learning techniques, we dropped it right away. We didn't bother with anything outside those lines no unrelated topics, no non-textual financial analysis, and no general opinion mining unless it had a clear financial angle. Next, we looked at abstracts. Here, we focused on whether the work covered hierarchical or multi-level sentiment analysis, contextual embeddings, or deep learning methods in financial texts. If a study used only lexicon-based methods, basic machine learning without any hierarchical modeling, or focused on non-financial datasets, we left it out.

Finally, we read the full texts of the remaining papers. We stuck to our inclusion and exclusion criteria, paying special attention to how solid their methods were, how they handled hierarchical modeling, and how relevant they were to multi-level sentiment analysis. In the end, we kept only those studies that brought real empirical findings, solid conceptual frameworks, or meaningful comparisons that fit the review's focus.

To further ensure rigor, a quality assessment was applied to all included studies. Quality criteria included:

- Clarity of research objectives and problem formulation;
- Appropriateness and transparency of the methodology;
- Adequacy of dataset description and experimental setup;

- Validity of evaluation metrics and results; and
- Relevance and contribution to hierarchical or contextual sentiment analysis in finance.

Studies that did not meet minimum quality thresholds were excluded, resulting in a robust and credible evidence base for the review.

1.8. Data Extraction and Synthesis

Once studies that fit were selected, a clear, structured process was used to pull out the needed information. A standard template was created and filled it out for every study, making sure we grabbed all the important details about how they handled hierarchical and multi-level sentiment analysis in financial texts. This made it way easier to keep things organized and let us compare the studies and spot themes without getting lost in the details.

The extracted variables focused on core dimensions of sentiment modeling. These included:

- Model type, such as recurrent neural networks, hierarchical attention networks, transformer-based architectures, or ybrid and ensemble models;
- Embedding approach, distinguishing between static embeddings, general-domain contextual embeddings, and finance-specific contextual models;
- Level of analysis, identifying whether sentiment was modeled at the word, sentence, aspect/entity, document, or aggregated market level;
- Dataset characteristics, including data source (e.g., financial news, earnings calls, social media), size, language, and availability; and
- Performance metrics, such as accuracy, precision, recall, f1-score, regression-based sentiment scores, or market correlation measures. Capturing these variables enabled cross-study comparison and supported the identification of dominant methodological trends.

We brought together two main strategies to make sense of the research: thematic and comparative synthesis. With the thematic approach, we grouped the studies by patterns in their methods, design choices, and where they were applied. That's how we spotted big-picture trends like how transformer-based models keep popping up, or how more researchers are drilling down into aspect-level sentiment in finance. Then, with the comparative side, we lined up the studies and looked hard at their different performance, how well they scale, and where the methods run into trouble. Instead of just stacking models up by their reported numbers, we paid attention to the bigger picture things, like how tricky the datasets were or how finely the studies broke down sentiment.

Using both approaches together gave us a much deeper, more connected view of the field. It lets us move past just summarizing what's out there and start piecing together a more critical, thoughtful picture of what matters.

2. Hierarchical Modelling in Finance

Financial sentiment analysis is really important when it comes to understanding how people feel about money and markets. This is because the way people express their feelings about finance can be connected in ways. For example, the words and sentences used in a company's earnings report can give clues about how the managers feel, even if they do not directly say anything

The thing is, people often express their feelings about finance in ways, and you have to look at the whole picture to understand what they mean. This is where hierarchical models come in. These models are designed to look at how feelings and ideas connect across different levels, from individual words and sentences to the overall meaning of a document or even the whole market. Looking at finance, this way is helpful because financial language can be very complicated. Financial texts are often dense and technical, and the meaning of the words and phrases used can depend on the context. For instance, the same word can mean different things in different industries or at different times. Hierarchical models are good at dealing with these complexities because they can look at the context and the bigger picture at the same time.

This makes them very useful for analyzing documents like the reports companies have to file with the government and for looking at information from multiple sources. Overall, using models to analyze financial sentiment is a good way to understand how people really feel about money and markets, and it can help us make better decisions. In financial sentiment analysis, hierarchical modeling is key. It helps us capture the ways that people express their feelings about finance. Financial sentiment analysis is important, and hierarchical modeling is a part of that. By using models, we can

get a better understanding of financial sentiment and how it works. Financial sentiment analysis and hierarchical modeling go hand in hand.

3. Contextual Embeddings for Financial Sentiment Analysis

3.1. Limitations of Traditional Word Embeddings

The way we show text, like Bag-of-Words and Term Frequency-Inverse Document Frequency, was used a lot in the beginning for looking at how people feel about things. This was because it was simple and did not need a lot of computer power. These ways of doing things just look at a bunch of words without caring about the order they are in or how they work together. So, they do not do a job of understanding what the words really mean together, which is very important for texts about money where feelings are often hinted at and not directly said. Also, these ways of showing text are not very good for complicated texts about money. Then some new ways of showing words, like Word2Vec, GloVe and FastText came along. Were a big improvement. They made it so that words could be shown as vectors that capture how similar they are in meaning based on how they are used. Even with these improvements each word still only gets one fixed meaning no matter how it is used. This is a problem for texts about money because many words can have meanings depending on what is around them. For example, words like capital, risk or charge can mean things and show different feelings based on the text around them. FastText tries to deal with this a little by looking at parts of words. It still does not really understand the context.

Moreover, the old ways of showing words have trouble working with texts about money because they were trained on texts and may not work well without a lot of extra training. They also do not do a job of looking at how feelings are shown at different levels of text. These problems made people want to use ways of showing words that can change based on the context and work better with how texts about money are structured. The new ways of showing words like embedding models are better at understanding the context and the hierarchy of financial language. Contextual embedding models, such as these ways of showing words are very important for financial texts. Financial texts, like these need embedding models to really understand the feelings and meanings, behind the words.

3.2. Contextual Embedding Models

The way we understand language has changed a lot because of embedding models. They help us see how words are related to the words around them. This is a deal because it helps us figure out what words mean in different situations. At the heart of these models is something called the Transformer architecture. It was created to fix some problems with models that were not good at understanding how words far apart are related. One of these models is called BERT. It looks at how all the words in a sentence're related to each other which helps it understand what words mean in different contexts. This is really important for understanding how people feel about something because it depends a lot on the order of the words and the situation. Many studies have shown that BERT is better at understanding how people feel than models especially when the text is complicated or specific to a certain area.

Other researchers have taken this idea. Made models that are specific to finance and economics. One of these models is called FinBERT. It was trained on a lot of texts like reports and news articles. Because it knows a lot about language it is better at understanding how people feel about financial things than general models. There are models like this, such as Financial-RoBERTa which is also good at understanding financial texts. Then there is EconBERT, which is focused on economics and policy. These models are good at understanding language and can help us figure out how people feel about things in a more detailed way. These new models are the basis for systems that try to understand how people feel about things in a detailed way. They help us understand what words mean in situations and can give us a better sense of how people feel about financial things. This is a step forward because it helps us make more sense of what people are saying especially when it comes to finance.

- These models are good at understanding language in context
- They can help us figure out what words mean in situations
- They are especially good at understanding language
- They can help us understand how people feel about things in a detailed way
- They are the basis for systems that try to understand sentiment in a detailed way

Contextual embedding models, especially the finance-specific ones are really important for understanding language and sentiment. They help us make sense of what people're saying, especially when it comes to finance. By using these models we can get a sense of what people mean and how they feel about things. This is a deal because it can help us make better

decisions and understand the world around us. Contextual embedding models like BERT and FinBERT are leading the way, in this area.

3.3. Comparative Performance in Financial Domains

How well sentiment analysis models work in finance really depends on whether you use general-domain or finance-specific contextual embeddings. Take BERT or RoBERTa, for example—they're trained on huge, mixed datasets like Wikipedia or BooksCorpus, so they pick up on general language patterns pretty well. They're solid as a starting point. But when it comes to financial sentiment, they tend to fall short. They just don't handle the industry jargon and the unique way people talk about finance. That's not just theory—several studies back this up, showing these general models often struggle with the specialized language you see in finance (Xing et al., 2020).

Table 1 Comparative Performance of Sentiment Embedding Models in Financial Domains

Embedding Type	Training Domain	Relative Performance	Strengths	Limitations
Static embeddings (Word2Vec, GloVe)	General	Low–Moderate	Simple, low computational cost	No contextual awareness; poor disambiguation
General-domain contextual embeddings (BERT, RoBERTa)	General	High	Strong linguistic representation	Domain mismatch in finance
Finance-specific contextual embeddings (FinBERT)	Financial	Very High	Accurate financial terminology interpretation	Requires large domain data
Finance-specific robust transformers (Financial-RoBERTa)	Financial	Very High	Improved generalization and robustness	Higher computational overhead
Economic-domain embeddings (EconBERT)	Economic/Policy	High	Captures macroeconomic semantics	Limited micro-finance coverage

Finance-specific embeddings like FinBERT and Financial-RoBERTa get their edge from being trained on huge piles of financial documents—think earnings reports, analyst notes, and news stories straight from the markets. Because these models learn the language of finance, they pick up on all the little quirks and shades of meaning you'd miss with a general-purpose model. When you stack them up against standard embeddings, the finance-focused ones come out ahead almost every time, especially on tasks like figuring out sentiment in financial texts, breaking down sentiment by aspect, or predicting how the market will react (Liu et al., 2020; Du et al., 2024).

What really sets these models apart is how well they deal with tricky financial terms. Words like "risk," "exposure," "capital," or "charge" can mean a bunch of different things, and their tone shifts depending on where you find them. Models trained on financial data get this they learn to read these words in context, so they're less likely to mess up and misclassify something just because of a literal reading. This skill becomes even more important when you're running complex analyses, like figuring out sentiment at different levels or across specific aspects. Nail the details, and the bigger picture falls into place. In the end, finance-specific embeddings just give you a sturdier base for modeling sentiment in the messy, nuanced world of financial text.

3.4. Challenges with Contextual Embeddings

Contextual embedding models work well, but they hit some big roadblocks in financial sentiment analysis mostly around domain shift, data scarcity, and the heavy lift of fine-tuning. Domain shift stands out as a huge issue. Most of these models start out trained on broad, everyday language. But throw them into the world of finance, and things get messy fast. Financial language is packed with jargon, strict regulations, and subtle meanings that change with context. Even if you tweak a model for finance, it can still stumble when you move from, say, banking to crypto or insurance. Then there's the problem of data. Good labeled datasets in finance are hard to come by. Companies lock down their data

for privacy and competitive reasons. Getting experts to label sentiment is expensive and slow. So, people end up working with small, narrow datasets, and models start overfitting or just stop working well when faced with anything new.

Fine-tuning adds another layer of hassle. Transformer models are massive think hundreds of millions of parameters. Training and running them takes serious hardware, usually GPUs or TPUs, and not everyone has access to that kind of power. Plus, fine-tuning is tricky. If you don't get the settings and training approach just right, you can actually make the model worse instead of better. All this makes it tough to scale these models up or use them in situations where you need results fast, like real-time trading. So, if you want to use contextual embeddings for financial sentiment, you have to wrestle with making them more efficient, robust, and less hungry for labeled data. That's really where the research needs to focus.

4. Deep Learning Architectures for Hierarchical Sentiment Analysis

4.1. Recurrent Neural Networks and Variants

Recurrent Neural Networks (RNNs) have been a building block in deep learning for sentiment analysis, mostly because they're good at understanding sequences in text. But standard RNNs run into some trouble—the vanishing and exploding gradient problems make them struggle with longer financial documents. That's where Long Short-Term Memory (LSTM) networks, and their bidirectional cousins (BiLSTM), come in. LSTM and BiLSTM models handle long-range dependencies much better since their gating mechanisms help them keep the important stuff and forget the rest. This makes them a good fit for finding sentiment in financial texts, like earnings calls or news articles.

To dig deeper into the structure of sentiment, Hierarchical Attention Networks (HANs) build on RNNs by adding attention at different levels. HANs usually use word-level LSTMs to encode sentences and sentence-level LSTMs to represent whole documents. Attention layers step in to highlight which words or sentences matter most for sentiment. This setup just makes sense for financial text, which often has layers of meaning. By letting the model zero in on the most telling words and sentences, HANs pull together a clearer picture of sentiment. Thanks to this combination of RNNs and hierarchical attention, these models are easier to interpret and tend to perform better on financial sentiment analysis.

4.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have shown real promise in sentiment analysis because they can pick out local features like key phrases or modifiers no matter where they appear in the text. Financial sentiment analysis often uses CNNs to grab important patterns at both the sentence and document level, with convolutional filters picking up n-grams that hint at positive or negative tone. Since CNNs are easy to run in parallel, they can chew through massive financial datasets quickly, which makes them appealing for jobs where speed and scale matter.

But here's the catch: CNNs are great at spotting local features, yet they don't really capture long-range connections across a document. To fill that gap, researchers have started blending CNNs with RNNs in hybrid models. In these setups, the CNN layers first extract the strongest local signals, and then RNN or LSTM layers step in to connect everything together, making sense of the sequence and broader context. These hybrids strike a good balance they're strong at picking out local details but also keep track of the big picture, which really helps when you're dealing with complex financial reports or regulatory filings.

4.3. Transformer-Based Architectures

Transformer-based models have taken over modern sentiment analysis, and for good reason. Their self-attention mechanisms let them look at every word in a sequence at once, figuring out which parts matter most for understanding sentiment. In financial sentiment analysis, this means transformers can weigh the importance of different words, sentences, or even whole sections, picking up on subtle cues that signal positive or negative tone. Unlike RNNs, transformers don't read text one step at a time they process everything in parallel, which makes them especially good at handling long, information-packed documents like annual reports or analyst notes.

One of the standout features of transformers is multi-head attention. This lets the model look at the text from several angles at once, with each attention head picking up on different linguistic or semantic details maybe tone, maybe context, maybe specific entities. Because of this, transformers can capture sentiment at lots of different levels, from individual words to whole documents, often without being explicitly told to do so. All in all, transformer-based models are flexible, powerful, and well-suited for digging into the layers of sentiment in complex financial writing.

4.4. Hybrid and Ensemble Approaches

Hybrid and ensemble methods are changing the game for sentiment analysis in finance. By mixing things like embeddings, attention mechanisms, and layered structures, these models tap into the best of each technique. Take hybrid models, for example. They might blend rich contextual embeddings with hierarchical attention networks or even mix CNNs with Transformers. The result? You get deeper semantic understanding, can keep track of sentiment shifts at different levels, and focus on the parts of the text that actually matter for sentiment.

Ensembles push things even further. They combine predictions from several models, which cuts down on variance and helps avoid the usual biases you get from relying on just one approach. But there's a catch. All these benefits come with heavier computational demands, more memory usage, and trickier training setups. So, rolling out these systems in real-time or in places where resources are tight isn't always easy. In the end, it's all about finding the right balance accuracy, interpretability, and efficiency all matter if you want to get the most out of these advanced sentiment analysis tools.

5. Evaluation Metrics and Benchmark Datasets

5.1. Commonly Used Financial Sentiment Datasets

Benchmark datasets have really pushed financial sentiment analysis forward. They let researchers actually compare models and see what works. Take the Financial PhraseBank, for example. It's a go-to dataset full of sentences from real financial news, each one tagged by experts as positive, negative, or neutral. That makes it perfect for testing how well models handle sentence-level sentiment in finance, especially when you're working with language models built for this domain.

Then there's the FiQA dataset. It doesn't just stick to news articles. You get news headlines, social media posts, and all sorts of user-generated content, all labeled for sentiment and relevance. FiQA covers both classification and regression tasks, so you can test out all kinds of models, especially those that rely on context. And let's not forget data from places like StockTwits or finance threads on Reddit. Sure, these posts are messier and less polished, but they come fast and reflect what real investors are thinking right as it happens. They're great for tracking market moods in real time or making short-term predictions. Put together, these datasets give researchers a lot of options. They encourage experiments with both simple and sophisticated models, and they keep the field moving forward.

5.2. Evaluation Metrics

When you want to judge how good a financial sentiment analysis model is, you can't just look at one thing. You need to check how well it classifies things, how it handles scores that fall on a spectrum, and how useful it actually is in the market. For classification jobs, people usually look at accuracy, precision, recall, and F1-score. Accuracy tells you how often the model gets it right overall, but precision and recall dig into the details like, does the model confuse positives and negatives? That matters a lot in finance since a single mistake can cost you more in one direction than the other. The F1-score is handy when your data is skewed, since it balances out precision and recall. If you're working with sentiment that's more of a sliding scale than just positive or negative, you'll want to use regression metrics. Here, mean squared error or measuring how closely the predictions line up with human ratings do the trick. This helps when you care about how strong the sentiment is, not just which side it falls on.

But let's be real just being good with words isn't enough. You also need to see if these sentiment signals actually matter in the market. So, people track how well sentiment lines up with stock returns, volatility, or trading volume. If those numbers move together, you've got something valuable proof that your model isn't just smart, it's useful in the real world.

5.3. Cross-Level and Cross-Dataset Evaluation Issues

One big problem with financial sentiment analysis is that everyone seems to use their own rules. There's no single benchmark for sentiment levels or data sources, just a patchwork of different datasets, each with its own way of labeling, level of detail, and focus. That makes it tough to compare results between studies. Take models trained to spot sentiment in financial news at the sentence level. They often stumble when you throw them into earnings reports or try to use them for market-wide sentiment. With so much fragmentation, it's hard to know what "good" performance even means in this field. Generalization and reproducibility are headaches, too. A lot of research is built on proprietary data or really narrow collections, so other people can't easily check the results or test them in new settings. Plus, everyone has their own process for cleaning data, fine-tuning models, and evaluating results, which makes it even trickier for anyone to reproduce the findings. So, reported improvements sometimes fall apart when you try them on different datasets or in

other corners of finance. What's really needed here is a set of open benchmarks, clear evaluation guidelines, and better transparency. Without that, progress in financial sentiment analysis will stay pretty scattered.

6. Explainability and Interpretability in Hierarchical Financial Sentiment Models

6.1. Importance of Explainability in Financial Applications

Explainability matters a lot in financial sentiment analysis, and not just because it sounds good on paper. Banks and other financial players have to follow strict rules, so their models need to show how and why they make decisions. When you're dealing with money trading, giving out loans, managing risk—you can't just trust a black-box model and hope for the best. If a model spits out a decision and nobody can say how it got there, regulators start asking tough questions. That's where explainable models come in. They let everyone from investors to analysts to regulators see how a model turns text into sentiment signals. This isn't just about checking boxes for compliance, either. People need to know these models are fair, reliable, and make sense, or they just won't trust them. Tools like attention maps and feature attribution actually show which words or phrases matter most to the model's prediction. This kind of transparency goes a long way. It builds trust, helps keep AI ethical, and makes sure nobody gets blindsided by weird or mysterious decisions in high-stakes financial settings.

6.2. Post-hoc Explainability Methods

Post-hoc explainability methods let you interpret sentiment analysis models without changing how the models are built. Some of the most popular tools here are SHAP, LIME, and Integrated Gradients. SHAP looks at each feature or token and shows how much it contributes to a prediction, all based on ideas from game theory. LIME works differently it builds a simple model right around the prediction you're interested in, so you can see why the original model made that call. Integrated Gradients, on the other hand, figure out which input features drive the prediction by looking at how the output changes as you move from a baseline to the actual input.

In financial sentiment analysis, these tools help you see which words or phrases sway the model's decisions. Still, their explanations usually focus on single examples and can miss how different parts of the text combine or interact at a bigger-picture level.

6.3. Limitations of Current Explainability Techniques

Explainability tools have their uses, but they run into real problems when it comes to financial sentiment analysis. Most of them only offer local explanations, missing the bigger picture and the way sentiment shifts across different parts of the text. Attention weights feel straightforward, but they don't always point to what actually matters, which can throw people off. And those post-hoc tools like SHAP and LIME? They eat up a lot of computing power and can react unpredictably if you tweak the settings. On top of that, the explanations themselves often turn out too technical, leaving non-experts confused. This makes it tough for these methods to really work in strict, regulated financial settings.

7. Open Challenges and Research Gaps

7.1. Domain Adaptation and Concept Drift

Financial sentiment models have a lot of trouble with something called domain adaptation and concept drift. The way people talk about money is different in areas like stocks or cryptocurrency and it also changes when the market is doing well or badly. So a model that is trained on news about stocks might not work well when you try to use it on news, about cryptocurrency or what the government is doing with the economy. The way people think about money also changes over time because the market and the rules are always changing and people are always changing how they invest their money. This means that models that do not change and learn can become useless unless they are always being updated. This shows that we need to find ways for these models to learn and adapt so they can keep working in different situations. Financial sentiment models really need to be able to adapt to these changes.

7.2. Sarcasm, Ambiguity, and Implicit Sentiment

Sarcasm, ambiguity, and subtle sentiment always get in the way of analyzing financial text. People in finance don't always say exactly what they mean, especially in things like regulatory filings and earnings calls, where the language can get cautious or even intentionally vague. On social media and investor forums, it gets even trickier. Sarcasm and irony are everywhere, so what someone says might be the exact opposite of what they actually mean. Even with all the

advances in contextual language models, picking up on these subtleties is tough. Models keep missing the mark, which leads to mistakes and makes their decisions harder to trust.

7.3. Multilingual and Low-Resource Financial Texts

Most research on financial sentiment sticks to English, which leaves out a lot of the world. In many emerging markets, financial disclosures and commentary come in local languages, but there just aren't enough labeled examples to train good models. This means we end up with tools that don't work well outside the usual English-speaking markets—and that introduces bias, both geographic and economic. To fix this, we need better multilingual embeddings, smarter transfer learning, and more annotated datasets for these overlooked languages.

7.4. Real-Time and Streaming Financial Data

These days, everyone wants up-to-the-second insights from real-time financial data. Social media, breaking news, live earnings calls, there's a flood of text coming in every second, and people expect instant analysis. The problem? Deep learning models, especially the big ones, need a ton of computing power and can be slow to update. That's just not going to cut it for real-time needs. We need faster, more efficient models and ways to keep them updated on the fly, or else they'll never keep up.

7.5. Ethical Issues and Bias in Financial Sentiment Models

Bias and ethics aren't just buzzwords; they're real problems in financial sentiment analysis. If you train a model on old data, it can easily pick up biases based on company size, industry, or location, which ends up hurting certain players in the market. On top of that, if the model's decisions aren't transparent, it's hard to hold anyone accountable, especially where regulations matter. Tackling these issues means we have to find and fix bias, be open about how models work, and make sure they're in line with both the rules and what society expects. All together, these challenges make it clear: financial sentiment analysis needs to be robust, flexible, and above all, ethical.

8. Future Research Directions

8.1. Unified Hierarchical Modeling Frameworks

It's time to focus on building unified hierarchical modeling frameworks that bring all levels of sentiment analysis together from single words and aspects, all the way up to full documents and even market-wide trends. Right now, most methods split these levels up and work on them separately, which leads to gaps and lost information. A unified approach keeps sentiment consistent as it moves through different layers of analysis, helps standardize how we evaluate models, and makes it easier to share knowledge between financial topics and data sources.

8.2. Multi-Modal Financial Sentiment Analysis

Financial sentiment isn't just about text anymore. People use their voice and visuals like tone in earnings calls or images in financial news to express market feelings. When we blend text, audio, and visuals, we get a much deeper and more accurate read on what's really happening. There's still a lot to figure out when it comes to combining all these data types, but the payoff for prediction and analysis could be huge.

8.3. Integration with Market Prediction Models

Another important direction is to tightly connect sentiment analysis with market prediction models. Instead of tacking on sentiment as just another feature, future systems should weave sentiment insights especially those rich, hierarchical ones right into the heart of forecasting, risk, and trading tools. Doing this not only makes economic sense, but it also gives sentiment-driven insights more practical value.

8.4. Lightweight and Efficient Deep Learning Models

Big, heavy transformer models eat up too many resources for real-time finance work. So, there's a real need for lighter, faster deep learning models think model compression, distillation, and smarter fine-tuning. The goal? Hit the sweet spot between accuracy, speed, and scalability, so these models actually work in real-world financial settings.

8.5. Explainable and Regulation-Aware Sentiment Systems

Finally, future sentiment systems can't just be high-performing black boxes. They need to be explainable and built with regulatory rules in mind, so people can trust them and banks can actually use them. Making explainability and compliance part of the model from the start is the only way financial institutions will adopt these tools responsibly.

9. Conclusion

This study has taken a look at the recent advances in hierarchical and multi-level sentiment analysis in financial texts. It pays attention to how contextual embeddings and deep learning architectures work. The review shows that financial sentiment is made up of different levels from words and sentences to aspects and documents and the market as a whole. Contextual embedding models those made for finance like transformers have done a great job of understanding what words mean and figuring out sentiment. They are better at this than methods that use lexicons and traditional machine learning. Hierarchical and hybrid architectures also do a job by keeping track of how different levels are connected and letting sentiment be combined in a structured way.

This review of financial sentiment analysis has made some contributions. It brings together all the ways to model hierarchical sentiment compares embedding models that are general and those that are specific to finance and looks at the results from many different datasets and architectures. The review also points out some challenges that're still present, like adapting to new domains being able to explain results, being able to handle large amounts of data and being consistent in evaluations. The review has implications for people who do research and for people who work in finance. For researchers, the study shows that we need frameworks that can handle level benchmarks that are standard and investigations that can work with many languages and not much data. For people who work in finance, the review emphasizes the importance of models that are adapted to the finance domain, can be explained, and do not use much computer power. This is important because these models will be used in places where finance is regulated.

Financial sentiment analysis that uses levels is a very important area of research. As models get better at adapting to data being interpretable and working with systems that predict the market, they will play a bigger role in making financial decisions based on data and in doing research. The idea of sentiment analysis will be very important in the future of financial text analytics. Hierarchical sentiment analysis will keep being a direction for advancing financial text analytics and financial sentiment analysis.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alsowaidi, S. M. (2024). Sentiment analysis for airline Twitter based on machine learning and deep learning. Global Journal of Engineering and Technology Advances, 20(02), 125–134.

[2] Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. ArXiv.org. https://arxiv.org/abs/1908.10063

[3] Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The Long-Document Transformer. ArXiv.org. https://arxiv.org/abs/2004.05150

[4] Dao, T., Fu, D. Y., Ermon, S., Rudra, A., & Ré, C. (2022). FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. ArXiv.org. https://arxiv.org/abs/2205.14135

[5] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 4171–4186.

[6] Du, K., Zhang, S., Liu, Y., & He, Q. (2024). Financial sentiment analysis: Techniques and applications. ACM Computing Surveys, 56(4), 1–38.

[7] Fedus, W., Zoph, B., & Shazeer, N. (2021). Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. ArXiv.org. https://arxiv.org/abs/2101.03961

[8] Goodell, J. W., & Huynh, T. L. D. (2020). Did Congress trade ahead? Considering the reaction of US industries to COVID-19. Finance Research Letters, 36, 101578.

[9] Gric, Z. (2021). Does sentiment affect stock returns? A meta-analysis of the literature (CNB Working Paper No. 10/2021). Czech National Bank.

[10] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. ArXiv.org. https://arxiv.org/abs/2106.09685

[11] Koh, P. W., Sagawa, S., Henrik Marklund, Xie, S. M., Zhang, M., Akshay Balsubramani, Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B., Haque, I., Beery, S. M., Leskovec, J., Anshul Kundaje, & Pierson, E. (2021). WILDS: A Benchmark of in-the-Wild Distribution Shifts. PMLR, 5637–5664. https://proceedings.mlr.press/v139/koh21a.html

[12] Lengkeek, M., van den Bos, J., & Frasincar, F. (2023). Leveraging hierarchical language models for aspect-based sentiment analysis in finance. Information Processing & Management, 60(2), 103181.

[13] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. ArXiv.org. https://arxiv.org/abs/2005.11401

[14] Li, X. L., & Liang, P. (2021). Prefix-Tuning: Optimizing Continuous Prompts for Generation. ArXiv.org. https://arxiv.org/abs/2101.00190

[15] Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A., Newman, B., Yuan, B., Yan, B., Zhang, C., Cosgrove, C., Manning, C. D., Ré, C., Acosta-Navas, D., Hudson, D. A., & Zelikman, E. (2022). Holistic Evaluation of Language Models. ArXiv.org. https://arxiv.org/abs/2211.09110

[16] Liu, Z., Shin, B., Xu, Y., & Hu, J. (2020). FinBERT: A pre-trained financial language representation model for financial text mining. Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 4513–4519.

[17] Morris, J., Lifland, E., Yoo, J. Y., Grigsby, J., Jin, D., & Qi, Y. (2020). TextAttack: A Framework for Adversarial Attacks, Data Augmentation, and Adversarial Training in NLP. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 119–126. https://doi.org/10.18653/v1/2020.emnlp-demos.16

[18] Ribeiro, M. T., Wu, T., Guestrin, C., & Singh, S. (2020). Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4902–4912. https://doi.org/10.18653/v1/2020.acl-main.442

[19] Roozen, D., & Lelli, F. (2021). Stock values and earnings call transcripts: A dataset suitable for sentiment analysis.

[20] Todd, A., Goodell, J. W., & others. (2024). Text-based sentiment analysis in finance.

[21] Web, C. (2024). Loughran-McDonald Master Dictionary w/ Sentiment Word Lists. Software Repository for Accounting and Finance. https://sraf.nd.edu/loughranmcdonald-master-dictionary/

[22] Xing, F., Cambria, E., & Welsch, R. (2020). Financial sentiment analysis: An investigation into common mistakes and remedy. Proceedings of the 28th International Conference on Computational Linguistics (COLING), 336–347.

[23] Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., & Liu, T.-Y. (2021). Do Transformers Really Perform Bad for Graph Representation? ArXiv.org. https://arxiv.org/abs/2106.05234

[24] Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontanon, S., Pham, P., Ravula, A., Wang, Q., Yang, L., & Ahmed, A. (2020). Big Bird: Transformers for Longer Sequences.

[25] Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document transformer. arXiv (2004.05150).

[26] Carlini, N., Tramer, F., Wallace, E., et al. (2021). Extracting training data from large language models. Proceedings of the 30th USENIX Security Symposium, 2633–2650.

[27] Dao, T., Fu, D. Y., Ermon, S., Rudra, A., & Ré, C. (2022). FlashAttention: Fast and memory-efficient exact attention with IO-awareness. arXiv (2205.14135).

[28] Fedus, W., Zoph, B., & Shazeer, N. (2021). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. arXiv (2101.03961).

[29] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. In Proceedings of ACL 2020 (pp. 8342–8360).

[30] Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M.-W. (2020). REALM: Retrieval-augmented language model pre-training. arXiv (2002.08909).

[31] Hu, E. J., Shen, Y., Wallis, P., et al. (2021). LoRA: Low-rank adaptation of large language models. arXiv (2106.09685).

[32] Karpukhin, V., Oguz, B., Min, S., et al. (2020). Dense passage retrieval for open-domain question answering. In Proceedings of EMNLP 2020 (pp. 6769–6781).

[33] Koh, P. W., Sagawa, S., Marklund, H., et al. (2021). WILDS: A benchmark of in-the-wild distribution shifts. In Proceedings of ICML 2021.

[34] Lepikhin, D., Lee, H., Xu, Y., et al. (2020). GShard: Scaling giant models with conditional computation and automatic sharding. arXiv (2006.16668).

[35] Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. In Proceedings of EMNLP 2021 (pp. 3045–3059).

[36] Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv (2005.11401).

[37] Li, X. L., & Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation. In Proceedings of ACL 2021 (pp. 4582–4597).

[38] Liang, P., Bommasani, R., Lee, T., et al. (2022). Holistic evaluation of language models. arXiv (2211.09110).

[39] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65.

[40] Modak, R. (2023). Market sentiment analysis using multimodal transformers: Integrating earnings calls, social media, and technical indicators. Global Journal of Engineering and Technology Advances, 15(3), 228–237.

[41] Morris, J. X., Lifland, E., Yoo, J. Y., Grigsby, J., Jin, D., & Qi, Y. (2020). TextAttack: A framework for adversarial attacks, data augmentation, and adversarial training in NLP. In Proceedings of EMNLP 2020: System Demonstrations (pp. 119–126).

[42] Pfeiffer, J., Rücklé, A., Poth, C., et al. (2021). AdapterFusion: Non-destructive task composition for transfer learning. In Proceedings of EACL 2021 (pp. 487–503).

[43] Pineau, J., Vincent-Lamarre, P., Sinha, K., et al. (2021). Improving reproducibility in machine learning research: A report from the NeurIPS 2019 reproducibility program. Journal of Machine Learning Research, 22(164), 1–20.

[44] Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and transparent dataset documentation for responsible AI. In Proceedings of FAccT 2022 (pp. 1285–1298).

[45] Rampášek, L., Galkin, M., Dwivedi, V. P., et al. (2022). Recipe for a general, powerful, scalable graph transformer. arXiv (2205.12454).

[46] Ribeiro, M. T., Wu, T., Guestrin, C., & Singh, S. (2020). Beyond accuracy: Behavioral testing of NLP models with CheckList. In Proceedings of ACL 2020 (pp. 4902–4912).

[47] Singh, P., Kumar, A., & Gupta, R. (2024). Comprehensive analysis of natural language processing techniques including sentiment classification. Global Journal of Engineering and Technology Advances, 19(01), 083–090

[48] Wallace, E., Feng, S., Kandpal, N., Gardner, M., & Singh, S. (2021). Concealed data poisoning attacks on NLP models. In Proceedings of NAACL 2021 (pp. 139–150).

[49] Yang, Y., Downey, D., & Tsioutsiouliklis, K. (2020). FinBERT: A pretrained language model for financial communications. arXiv (2006.08097).

[50] Ying, C., Cai, T., Luo, S., et al. (2021). Do transformers really perform badly for graph representation? Advances in Neural Information Processing Systems, 34.

[51] Zaheer, M., Guruganesh, G., Dubey, A., et al. (2020). Big Bird: Transformers for longer sequences. Advances in Neural Information Processing Systems, 33.