Abstract

This paper presents a systematic literature review on financial narratives and the methods used to model them. To support this review, we develop an algorithmic framework that improves the efficiency, reproducibility, and consistency of study selection. The framework uses Natural Language Processing (NLP) techniques, clustering algorithms, and interpretability tools to automate key steps of the screening and analysis process. We apply this approach to the study of financial narratives, a growing field in financial economics concerned with how structured interpretations of economic events shape market behaviour and asset prices. Drawing from the Scopus[1] database of peer-reviewed literature, the review identifies research practices used to model financial narratives with a range of NLP methods. Results show that although technical progress has been significant, the conceptualization of financial narratives remains fragmented and is often simplified to sentiment-based measures. A large share of studies relies on media sources, which is understandable given their availability, but the short length of these texts introduces noise and reduces predictive reliability. Using longer and more diverse textual materials would help to address this limitation and provide a broader view of narratives circulating among economic agents. The findings also highlight the importance of more comprehensive forms of narrative modeling that go beyond sentiment alone, and they illustrate the practical usefulness of the proposed algorithmic SLR methodology.

Keywords— Systematic Literature Review, Text Processing, NLP, Financial Narrative, Financial Market Dynamics

1   Introduction

The influence of narratives on financial markets has recently become a prominent area of study in both economics and finance. The analysis of narratives includes understanding how stories evolve, spread, and impact financial markets over time. By examining the mechanisms through which they form and propagate, researchers aim to uncover their role in shaping expectations, driving investor behavior, and consequently impacting market cycles. In social science, narratives are structured accounts or interpretive frameworks through which people understand and act in the social world. The work of Somers (1994) advances that they are not simply representations but ontological structures: social life itself is storied. Narratives constitute social identities, guide actions, and embed individuals in relational settings over time and space. These narratives are constituted through emplotment (causal linking of events), selective appropriation, and their temporal and spatial connectivity.

In economics, narratives have gained traction as a bridge between behavioral and informational theories of markets. Grossman (1980) shows that when information is costly, prices cannot be perfectly informationally efficient, as perfectly revealing markets would remove incentives to acquire information. This insight departs from the view of fully efficient markets and acknowledges the role of delayed or selective information diffusion. Thus, if information diffusion is neither instantaneous nor uniform, then the stories through which it is interpreted play an important role in how beliefs form and prices adjust. The work of Shiller (2017) highlights this point by demonstrating that contagious economic stories can drive fluctuations in aggregate activity. Additionally, Shiller (2019) goes further and argues that economic fluctuations cannot be fully understood through quantitative models alone, as narratives play a critical role in shaping market movements. Similarly, Tuckett and Nikolic (2017) shows that conviction narratives, defined as

---

[1] https://www.scopus.com/

coherent and emotionally compelling stories, enable investors to act with confidence despite uncertainty, reinforcing collective behaviors in markets. Building on this theoretical foundation, Shiller (2020) emphasizes that economic narratives function as widely shared interpretative stories about economic events, shaping collective beliefs and the propagation of expectations across agents, thereby providing a conceptual link between individual interpretation and aggregate market dynamics.

Recent advances in the literature further demonstrate that narratives are not only descriptive but can themselves actively shape observable macro-financial outcomes. Bybee et al. (2023) demonstrate that news-based narrative factors provide explanatory power for risk premia and future investment opportunities, aligning with predictions from intertemporal asset pricing theory, while Bybee et al. (2024) analyze news attention extracted from large-scale textual corpora and show that it closely tracks business cycles and helps forecast market returns. More recently, Flynn and Sastry (2024) developed a macroeconomic model in which narratives generate persistent, belief-driven fluctuations with empirical evidence from firm-level disclosures. Drawing from these perspectives, we define financial narratives as structured, transmissible interpretations or explanatory frameworks concerning financial markets or economic events, based on available information and possessing causal temporal structure.

Despite growing interest in the field, the concept of financial narratives remains underdefined, and the methodologies used to analyze them are diverse and often inconsistent. The purpose of this systematic literature review is to address this gap by offering a clearer picture of how narratives are conceptualized, modeled, and employed in the study of financial markets. It also provides a comprehensive overview of the textual analysis techniques employed, with particular attention to the evolution of methods following the introduction of transformer-based architectures. The review seeks to assess whether significant methodological advances have been made since early contributions, and to identify both conceptual gaps and best practices in this emerging research area. The study addresses the following research questions: (i) How can NLP and textual analysis techniques be used to quantify and model financial narratives? (ii) How is the concept of financial narrative defined and applied, and what methodological patterns emerge from the existing literature? These research questions focus on synthesizing how financial narratives are studied in the academic literature, rather than an empirical examination of narrative data itself. The results should therefore be read as an account of how the existing literature conceptualizes, operationalizes, and models narratives in financial markets.

The contribution of this research lies both in its synthesis of existing studies and in the methodology used to conduct the review. We propose a reproducible, algorithmic framework for systematic literature reviews that enhances the systematic discovery, screening, and synthesis of research using transformer-based embeddings, dimensionality reduction, and clustering. This approach contributes to the development of scalable, data-driven techniques for literature synthesis and supports more reproducible research in fast-evolving domains. The proposed framework enables a more systematic and transparent review process while improving selection quality and alignment with research objectives.

The review is structured as follows: Section 2 presents the algorithmic framework for systematic literature reviews, including the query design, filtering steps, clustering approach, and quality assessment procedures. Section 3 applies the methodology to the domain of financial narratives, analyzing research trends, modeling techniques, and conceptual challenges identified in the selected literature. Section **??** discusses the implications of the findings, the limitations of existing approaches, and potential directions for future research in financial narrative modeling.

2 Methodology

Our approach in this literature review is inspired by the work of Amato et al. (2024), who structured the review process into eight distinct steps: defining the research question, developing and implementing the review methodology, conducting literature exploration and analysis, applying selection criteria (inclusion/exclusion), assessing the quality of selected studies, extracting relevant data, synthesizing findings, and reporting insights. This structured framework, initially based on Varsha et al. (2024), starts by defining the research problem and formulating clear research questions. A structured review methodology is then developed to guide literature search and analysis. Thus, specific inclusion criteria are applied to filter studies and assess their relevance and quality. Once the final set of papers is selected, key information is extracted and synthesized to generate key insights, organized and presented in a structured report.

Table 1: Summary of article selection criteria.

Criteria	Decision
Inclusion of pre-defined keywords in title, abstract, or keyword list	Inclusion
Article not published in a scientific journal	Exclusion
Article in a language other than English	Exclusion
Article published in 2010 or before	Exclusion
Duplicates	Exclusion
Low relevance to research questions	Exclusion
Unavailability of the article online	Exclusion
Workshop Proceeds	Exclusion

However, because the core phase of the current methodology, the study selection, depends heavily on the researcher's judgment, the overall process remains subjective, difficult to replicate, and prone to bias. To address this, we extend the framework by integrating algorithmic rules to automate and refine the selection phase, improving the reproducibility and consistency of inclusion decisions and the assessment of study relevance. As shown in Table 1, our paper selection process remains fully systematic, following predefined rules. The idea of algorithmically structuring the literature selection process follows work proposed by de Oliveira et al. (2019), who emphasized automated mapping of research fields through keyword co-occurrence and citation networks. While these earlier frameworks demonstrated the value of systematic automation, they remained limited by their reliance on lexical similarity rather than semantic understanding. The present framework extends this aspect using transformer-based embeddings for semantic relevance assessment, interpretable dimensionality reduction for redundancy control, and quantitative evaluation of clustering quality. In this section, we present the overall review process summarized in Figure 2, from the initial research query definition to the data extraction and result analysis phase.

2.1 Initial Research Sourcing

The first phase involves defining the research context and the type of information we aim to extract from the literature. This review sources the publications from the Scopus database, specifically retrieving basic information from the results through the Application Programming Interface (API) and scraping the abstract, authors and keywords from Scopus URLs. We also automatically retrieved journals metadata (ranking and overall information) from SCImago

Journal Ranking website [2].

Scopus was selected as the primary data source because it offers reliable coverage of peer-reviewed research in economics, finance, and computer science, with metadata that is curated in a consistent and transparent manner. Its API provides immediate access for academic users, which allows the entire retrieval process to be automated and reproduced without institutional restrictions. Other repositories such as SSRN or arXiv contain valuable material but mix peer-reviewed and non-reviewed documents, making quality control and comparability more challenging. While recent open initiatives like OpenAlex (Priem et al., 2022) broaden access to bibliographic data, this study prioritizes the stability and curation standards of Scopus, which remains one of the most academically recognized databases for peer-reviewed research.

As this study focuses on narrative modeling using NLP techniques to better understand financial market, we are particularly interested in literature that addresses both the methodological development of narrative modeling from large textual corpora and the application of NLP (both traditional and recent approaches) in finance. This leads us to formulate two research questions: How can NLP and textual analysis techniques be used to quantify and model financial narratives? How is the concept of financial narrative defined and applied, and what methodological patterns emerge from the existing literature? Accordingly, we seek to identify rigorous and relevant academic work across three main themes: theoretical discussions of narratives in finance and economics (financial narratives theory), the use of NLP and text analysis to extract narratives in financial contexts (NLP for narratives processing), and more precisely, the quantification and tracking of narratives for financial applications (financial narratives modeling).

To initiate the selection process, we designed a search query using the advanced search function of the Scopus database. We experimented with various combinations of keywords in an effort to construct a query that is both broad enough to capture relevant research and narrow enough to exclude unrelated domains. The chosen query captures a wide range of pertinent literature, though it also retrieved some irrelevant papers (e.g. narratives in health or psychology, domains where narratives are also a major topic of study). The query was constructed to retrieve publications related to narratives, NLP, and financial markets:

- **Economic/Financial Narratives in title, abstract or keywords:**
  TITLE-ABS-KEY("financial narrative" OR "financial narratives" OR "economic narrative" OR "economic narratives")

- **Narratives, NLP/text analysis and financial terms in title:**
  OR TITLE(narrative* AND ("language processing" OR nlp OR "language understanding" OR nlu OR "text mining" OR "textual analysis" OR "text analysis" OR "text processing" OR lexicon OR sentiment OR "topic modeling" OR "topic extraction" OR "word embedding" OR "word embeddings" OR "entity recognition" OR "narrative processing" OR "narrative modeling") AND (economic* OR macroeconomic* OR "financial market" OR "financial markets" OR "market dynamics" OR "market movements" OR "price dynamics" OR "price movements" OR "financial forecasting" OR "stock markets" OR "equity markets" OR "foreign exchange" OR commodit* OR bond* OR "asset prices" OR "price returns" OR "asset returns" OR "market returns" OR volatility OR "risk management" OR "portfolio management"))

- **Narratives, NLP/text analysis and financial terms in keywords:**
  OR KEY(narrative* AND ("language processing" OR nlp OR "language understanding" OR nlu OR "text mining" OR "textual analysis" OR "text analysis" OR "text processing" OR lexicon OR sentiment OR "topic modeling" OR "topic extraction" OR "word embedding"

OR "word embeddings" OR "entity recognition" OR "narrative processing" OR "narrative modeling") AND (economic* OR macroeconomic* OR "financial market" OR "financial markets" OR "market dynamics" OR "market movements" OR "price dynamics" OR "price movements" OR "financial forecasting" OR "stock markets" OR "equity markets" OR "foreign exchange" OR commodit* OR bond* OR "asset prices" OR "price returns" OR "asset returns" OR "market returns" OR volatility OR "risk management" OR "portfolio management"))

The inclusion of the term "narrative(s)" is central, as it ensures thematic relevance to our research focus. We prioritized papers that explicitly mention "financial narrative(s)" or "economic narrative(s)" in the title, abstract, or keywords. Additionally, to broaden the scope while maintaining relevance, we formulated a more inclusive search strategy targeting titles and keywords (but not abstracts because they tend to be more descriptive and often mention broader methodological or contextual terms that may not reflect the core focus of the paper), combining "narrative(s)" with terms related to text analysis or NLP, and with terms referring to financial markets.

The second phase involves applying additional filters to refine the search, as well as removing duplicates entries. In our case, we extended the initial query by restricting the publication year to 2011 or later, limiting the results to peer-reviewed journal articles, and including only English-language publications:

- **Economic/Financial Narratives in title, abstract or keywords:**
  Identical to phase 1

- **Narratives, NLP/text analysis and financial terms in title:**
  Identical to phase 1

- **Narratives, NLP/text analysis and financial terms in keywords:**
  Identical to phase 1

- **Time constraint:**
  AND PUBYEAR > 2010

- **Language constraint:**
  AND LANGUAGE(english)

- **Article constraint:**
  AND DOCTYPE(ar)

At this stage of the selection process, an additional filtering criterion was introduced to enhance thematic alignment. Journal subject areas were cross-referenced with disciplinary classifications from the SCI journal ranking database to identify and remove publications originating from fields clearly unrelated to the research context. Journals primarily associated with disciplines outside the intended scope were excluded. Conversely, journals whose subject areas could plausibly intersect with the research objectives, or for which thematic relevance could not be reliably assessed at this stage, were retained for further evaluation.

2.2   Algorithmic Selection Framework

This section presents the third phase, consisting in an algorithmic process used to filter research papers based on their relevance. The selection framework leverages NLP and machine learning techniques to classify papers into three categories: high, medium, and low relevance. The process consists of three main stages: textual analysis, dimensionality reduction, and clustering.

2.2.1 Textual Analysis: Research Properties Statements

The selection methodology relies on a zero-shot latent embedding approach, in which the user defines key statements which describe the criteria that selected papers should meet. In our case, the statements respectively reflect the research focus, research type, context, methodology, data sources, and research questions:

- "The research discusses Financial Narrative Processing, Financial Narrative Modeling or the use of textual data to understand financial markets";

- "The research is highly relevant in the context of financial markets, including: equities, foreign exchange, cryptocurrencies, bonds, commodities, or real estate";

- "The research is an empirical study showcasing the use of textual data to model narratives and understand financial markets dynamics";

- "The research methodologies include: textual analysis, text mining or Natural Language Processing techniques such as topic modeling, emotion analysis, sentiment analysis, word embeddings or transformer-based models";

- "The research leverages the following data: large textual datasets, including financial reports, news articles, social media posts, audio or video transcripts, or any other form of financial textual data";

- "The research helps answering the research question(s): How can NLP and textual analysis techniques be used to quantify and model financial narratives? Can financial narratives modeling enhance financial market dynamics understanding?"

To limit the sensitivity of the selection procedure to the exact wording of the user, each statement was expanded into several paraphrased variants using OpenAI's chat-completion API (`gpt-3.5-turbo` model), and embeddings were averaged. This reduces the influence of any single formulation on the similarity scores and provides a more stable representation of the underlying research criteria.

For text embedding tasks (statements and candidate research papers), we opted for OpenAI's `text-embedding-3-small` because it provides strong semantic alignment, low embedding distortion, and consistent performance across diverse academic and technical domains. Among the available models, it offers an optimal trade-off between accuracy, computational efficiency, and contextual sensitivity. It is worth mentioning that, as the model outputs unit-normalized vectors, cosine similarity reduces to a simple dot product, which simplifies computation at scale. For each potential paper, the title, abstract, and keywords are concatenated and embedded. Cosine similarities between each paper embedding and each of the six statement embeddings are then computed, yielding six similarity scores per paper, which are then averaged to obtain an overall relevance score.

2.2.2 Data Preparation

Before proceeding with the clustering step, the matrix of similarity scores is standardized (Z-score) to ensure each dimension contributes equally to the clustering algorithm. The next step is to reduce the dimensionality of the data while preserving its most informative components, in the case where we observe high correlation between relevance features. This is achieved using Principal Component Analysis (PCA), which transforms the six-dimensional similarity space into a lower-dimensional representation by identifying the most important variance-explaining components.

We use the Kaiser-Meyer-Olkin (KMO) score and the Condition Number (CN) as diagnostic indicators to assess whether applying Principal Component Analysis (PCA) is appropriate. The KMO score, while traditionally used in factor analysis, serves here as a heuristic measure of inter-variable correlation strength. A value below 0.5 suggests that variables share little common variance, making PCA uninformative, whereas values above 0.7 indicate sufficient correlation structure for meaningful components. When KMO falls in the intermediate range (0.5-0.7), we additionally consider the CN, which quantifies multicollinearity. If $CN > 100$, implying near-linear dependencies among features, PCA is applied to stabilize the representation; otherwise, it is skipped.

When PCA is applied, we retain enough components to explain 99% of total variance. Unlike predictive or visualization-oriented uses of PCA, this study employs PCA primarily for diagnostic and interpretive purposes. The small number of input variables mitigates overfitting risk, and retaining nearly all variance allows transparent inspection of factor loadings to identify redundant or semantically overlapping features. We show later that this parameter has no effect on the selected papers. Papers that initially shared similar similarity profiles across multiple research statements are now grouped based on their principal components, which encapsulate the most distinguishing features of their textual content. This approach contrasts with non-linear dimensionality-reduction methods, which, while effective for visualization, distort variance structure and are less suited for analytical interpretation.

2.2.3 Clustering Research Selection

The next step in the selection process involves clustering the papers into distinct relevance groups using K-means. This method provides stable partitions in low-dimensional spaces, and its objective function of minimizing within-cluster variance offers a simple and transparent criterion that aligns with the reproducibility goals of this study. Its interpretability also facilitates downstream validation, which is important for a systematic review framework.

Given the diversity in methodologies, objectives, and data sources within the dataset, we adopted a three-cluster optimization. After applying K-means, we computed the mean relevance score within each cluster and assigned relevance tiers by ranking these means from highest to lowest. This procedure resulted in three groups that we refer to as high, medium, and low relevance. Manual inspection of "frontier" papers located near the decision boundary between the high- and medium-relevance clusters revealed that most borderline cases did not fall within the scope of this review. Because these borderline studies added conceptual noise without improving coverage, the medium-relevance cluster was not retained in the final selection. To evaluate the quality of the produced clusters, three metrics were computed on the resulting high-relevance cluster: the average relevance score, the Silhouette score (capturing both cohesion within the cluster and separation from the other clusters), and the number of papers retained (which we aim to be high).

2.3 Final Validation and Data Extraction

Once the algorithmic selection process was finalized, a last manual validation phase was conducted to refine the dataset. This stage aims to address limitations that algorithmic methods alone could not fully resolve. More precisely, publications whose full text was not accessible, either because of paywall restrictions or the lack of institutional access, were excluded from the corpus. In addition, unconventional documents such as workshop proceedings and compilation volumes were identified and excluded. These types of publications often aggregate a large number of individual studies under a general theme, without guaranteeing a consistent focus aligned with the research objectives. Moreover, the scale and heterogeneity of such documents made them impractical to process within the structured framework of this study.

To structure the extraction process, a reporting framework was designed to capture key aspects of each study. This framework includes details on the research purpose, methodology, data sources, dataset characteristics, main findings, and practical implications, the result of which is summarized in appendix (Table 3). While the current extraction process remains manual, future work will examine the use of unsupervised clustering approaches to group similar studies into coherent groups and limit subjective judgments during synthesis. Combining contextual clustering with automated content descriptors (TF-IDF representations or Large Language Models (LLM) summarization), would make it possible to identify, both quantitatively and qualitatively, the common characteristics shared across studies in a transparent and replicable manner.

3 Results

This section presents the key results of our systematic review, categorizing the academic literature into two categories: studies focused on detecting narratives in financial markets, and studies focused on modeling them empirically. The distinction lies in the epistemic and methodological aim. Narrative understanding papers aim to conceptualize or interpret the nature and role of narratives in economic contexts. Narrative modeling papers, on the other hand, propose empirical or algorithmic techniques to quantify, extract, or use narratives for forecasting or explanatory purposes.

3.1 Selection Phase Results

The paper selection process followed the four-phase pipeline presented earlier. In the first phase, the initial query was applied to retrieve 288 publications without preliminary restrictions on date or subject area, maximizing the initial recall of potentially relevant studies. In the second phase, a series of metadata-based filters were applied. Publications were restricted to journal articles written in English and dated after 2010. To eliminate papers clearly outside the financial or economic domains, we automatically excluded publications that appeared in journals associated with fields unrelated to the topic. These were manually identified by subject areas such as 'Arts and Humanities', 'Medicine', 'Health Professions', 'Earth and Planetary Sciences', 'Environmental Science', 'Agricultural and Biological Sciences', 'Biochemistry, Genetics and Molecular Biology', and 'Energy'. However, we chose not to exclude papers from journals listed as 'Engineering', 'Social Sciences', 'Multidisciplinary', or with no assigned category, as their relevance is evaluated during later stages of the selection process. Duplicate entries were also identified and removed during this phase, reducing the corpus to 125 publications.

The third phase introduced the algorithmic selection framework. Text embeddings were computed for the title, abstract, and keywords of each paper. In our use-case, the KMO score is 0.815, indicating strong sampling adequacy, and the CN is 370, which further confirms significant multicollinearity. Both indicators suggest that PCA is well suited for this dataset and we therefore obtained four components for use in subsequent analysis. The clustering selection resulted in 24 papers, with an average relevance and Silhouette scores of respectively 0.507 and 0.352.

In the final phase, the manual validation process, two papers were found to be inaccessible, and several workshop proceedings were identified among the selected documents. Four workshop papers were detected, each containing more than one hundred individual studies. Given the uncertainty regarding their thematic alignment with the research objective, these documents were excluded from the final dataset. After this refinement step, a total of 16 papers remained for detailed analysis. The evolution of the paper count through each phase of the process is summarized in appendix (Figure 2).
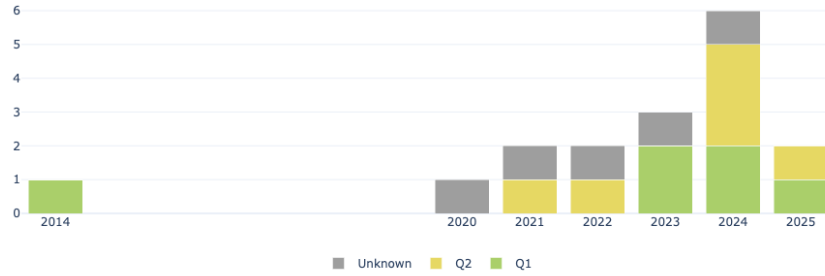
Figure 1: Temporal and rank distributions of selected research papers.

As illustrated on Figure 1, the temporal distribution of the final selected research spans from 2014 to 2025, with a majority of studies published in recent years. This distribution suggests a growing interest in the application of NLP techniques to financial narrative analysis, particularly in the last five years. The papers originate from journals covering multiple academic disciplines, including 'Economics, Econometrics and Finance', 'Social Sciences', 'Business, Management and Accounting', 'Computer Science', 'Mathematics', 'Psychology', and 'Decision Sciences'.

3.2 Narrative Understanding Papers

Across the reviewed literature, the exploration of narrative understanding in financial contexts has grown significantly, particularly as researchers noticed that financial narratives go beyond sentiment or topic extraction. Early works laid a foundation for understanding how narratives shape belief systems, institutional expectations, and investor behavior.

Hu et al. (2021) developed a detailed annotation model to detect opinion and emotion expressions in economic texts. Their methodology, which integrates intra-sentential labeling with economic terminology and rhetorical patterns, emphasizes the importance of understanding how subjective language conveys implicit judgments and anticipatory assessments in financial contexts. The resulting corpus, drawn from central bank speeches, 10-k fillings, news, and social media, allows for fine-grained linguistic modeling of investor and institutional sentiment, offering a window into how narratives materialize from text. From a multilingual and applied perspective, Zmandar et al. (2022) introduced CoFiF Plus, the first large-scale corpus of French financial narrative summaries. Though the paper's primary aim was dataset construction, the authors contextualized financial narratives as communicative devices used by firms to shape investor perceptions. Their discussion reinforces the idea that narratives play a central role in conveying trust and persuasion within institutional and linguistic conventions. Sy et al. (2023) also proposed a fine-grained argument mining approach applied to financial earnings calls, using BERT ensembles to classify and relate argumentative units. Their focus on logic structures and discourse coherence highlights the interpretative structure of financial narratives, especially in how sentiment and logic intertwine in decision-relevant communication.

Liu et al. (2024) addressed the subtlety of year-over-year semantic drift in financial reports. They introduced the Financial-STS task to quantify nuanced differences in narratives from company disclosures. Their work goes beyond textual similarity, proposing a categorization of semantic shifts (e.g., intensified sentiment, emerging situations) that reflect how managerial narratives evolve over time. This conceptual framework offers a way to systematically detect narrative shifts relevant for investors.

Finally, Roos and Reccius (2024) took a more theoretical stance, arguing that much of the

empirical literature confuses topics with narratives. They provided a rigorous definition of collective economic narratives as socially shared, action-oriented stories that emerge in context and suggest coordinated beliefs or behaviors. Their contribution lies in reconciling insights from institutional economics, literary theory, and complexity economics to establish conceptual clarity. Importantly, they emphasize that narratives are distinct from mere themes or sentiment, and must be sense-making and socially transmitted.

Together, these studies show that the literature conceptualizes financial narratives not as simple collections of words or topics, but as structured, evaluative, and often persuasive expressions shaped by economic language, argumentation, and context. Understanding narratives therefore requires both conceptual clarity and methodological precision, from theoretical definitions to corpus-level annotation and semantic modeling.

3.3 Narrative Modeling Papers

The second group of studies focuses on modeling narratives through quantitative methods, aiming to extract, represent, and utilize narrative signals in financial prediction or macroeconomic analysis. These works share a common methodological objective: translating textual narratives into structured, predictive signals for the economy or financial markets.

Among the earliest contributions, Tuckett et al. (2014) proposed a social-psychological framework based on conviction narrative theory, analyzing the dynamics of emotionally charged narratives (named "phantastic objects") and their impact on financial decision-making. Using sentiment shift detection on Reuters articles and the Enron email archive, they developed a method to capture rising emotional conviction and eventual disillusionment in narrative content, offering one of the first systematic approaches to narrative regime detection in financial text.

Hsu et al. (2021) explored how narrative salience in historical Chinese news sources related to the U.S. Silver Purchase Act could forecast economic indicators such as price levels. This study applied textual frequency analysis combined with regularized regression, highlighting the value of narrative-based proxies even in early 20th-century macroeconomic settings.

In contemporary financial markets, Chen et al. (2022) extracted narrative features related to the COVID-19 pandemic and studied their influence on financial volatility. They showed that high narrative virality and pessimism coincided with extreme market movements, demonstrating the real-time utility of narrative indicators in turbulent contexts.

Zhu et al. (2023) applied deep learning and natural language processing on social media data to model housing market sentiment in China. They built forward- and backward-looking indices using LSTM-based models trained on real estate-related posts, showing that narrative sentiment predicted price dynamics and investor expectations in housing.

Borup et al. (2023) collected free-text investor expectations from U.S. households and applied Latent Dirichlet Allocation (LDA) topic modeling to extract narratives. Their results revealed that these subjective narrative features significantly predicted excess asset returns, outperforming traditional sentiment indices. Ma et al. (2024) also used LDA on a Narrative-based Energy General Index (NEG) constructed with news from Wall Street Journal articles, specifically targeting energy-related topics. Their results showed that NEG predicted returns in the energy sector and outperformed both macroeconomic and sentiment-based predictors. Moreover, NEG showed predictive power across multiple sectors and the aggregate market, offering a robust narrative signal for asset allocation.

From an emotional angle, Agarwal et al. (2024) used dictionary-based emotion metrics to study Chinese stock bubbles. They found that emotions such as excitement and anxiety embedded in

media narratives had strong explanatory power for stock returns, trading volume, and volatility, suggesting that financial cycles are heavily narrative-driven.

Miori and Petrov (2023) applied large language models (GPT-3.5) and graph theory to construct weekly networks of news narratives. Their findings indicated that fragmentation in these narrative graphs often preceded market dislocations, proposing a novel structural indicator of systemic risk. Stander (2024) also leveraged Transformers architecture to developed a sentiment index from South African financial news, specifically using FinBERT with the aim of improving credit risk assessment under International Financial Reporting Standard (IFRS) 9 impairments. The study showed that such a narrative-based index could serve as financial signal as well as a regulatory input in volatile environments.

Later, Taffler et al. (2024) examined investor emotions during financial crises and developed crisis-specific emotion dictionaries. They demonstrated that narrative affect extracted from news explained over half of market returns during extreme events, pointing to the central role of emotion in price formation during crises.

The most recent study adopts a more structured approach: Hong et al. (2025) used over 880,000 WSJ articles to forecast U.S. inflation. Using topic decomposition and machine learning models, they showed that narrative predictors significantly outperformed standard macroeconomic models, especially during recessions.

The literature has moved well beyond dictionary sentiment measures or topic model. Modern approaches rely on topic models, neural networks, transformer-based representations or LLMs to capture how narratives develop and interact with market conditions, allowing them to be treated as measurable inputs in empirical analysis.

3.4 Stability Checks

We propose to examine whether the selection of high-relevance papers is stable under reasonable modifications of the dimensionality reduction and clustering steps. The objective is to verify that the final corpus is not the result of specific parameter choices. Robustness is assessed using the three control metrics introduced earlier: average relevance score, Silhouette score, and retained paper count. We also report the number of papers added or removed relative to the baseline, defined as PCA retaining 99% of explained variance followed by K-means clustering. Sensitivity to word choice in the statements was mitigated by generating multiple paraphrased versions of each statement and averaging their embeddings. This approach reduces the dependence of similarity scores on specific formulations and lowers the risk of word-sensitivity effects. Because this mitigation is built directly into the pipeline, no additional robustness checks were performed on the wording of the statements. Similarly, model-selection robustness was not performed, since high-performance closed-source embedding models tend to yield comparable semantic behaviour. Open-source testing is left for future work on portability and domain adaptation.

We evaluated PCA robustness by varying the retained explained variance within 80%, 85%, 90%, 95% and 98%. Across all configurations, K-means consistently recovered the same set of 26 high-relevance papers. The average relevance score remained fixed at 0.504, the Silhouette value stayed at 0.347, and no papers were added or removed relative to the baseline. These results show that the selection step is essentially insensitive to the choice of PCA variance threshold within a wide and standard range. PCA improves interpretability, but the exact level of dimensionality reduction does not affect the composition or the quality of the high-relevance cluster.

We also compared the baseline K-means results to Gaussian Mixture Models (GMM) and Agglomerative Clustering (AC), applied to the same PCA space and identical number of clusters. The comparison shows that alternative clustering strategies modify the size and composition

Table 2: Comparison of clustering configurations applied to the PCA-transformed similarity space.

Method	#Papers	Avg. relevance	Silhouette
K-means (baseline)	26	0.504	0.347
Gaussian Mixture Model (GMM)	19	0.509	0.270
Agglomerative Clustering (AC)	50	0.480	0.400

of the high-relevance group. GMM yields the highest average relevance score but removes seven papers that appear in the baseline and retains fewer papers overall. AC produces the highest Silhouette score but does so by expanding the cluster to fifty papers, many of which have substantially lower relevance scores. Both alternatives therefore weaken at least one of the quality criteria: GMM reduces overlap and retains fewer relevant papers, while AC incorporates a large number of marginal studies. Relative to these configurations, K-means maintains a stable cluster size, balanced internal cohesion, and a higher concentration of highly relevant papers, which makes it the most reliable choice for the final selection.

4 Discussion

This section synthesizes the key insights of the review in relation to the two guiding research questions and the broader literature on financial narratives. We examine how existing studies define and model narratives, how NLP techniques are used to model them, and what this reveals about the current state and direction of the field. The discussion also outlines the methodological contribution of this work, its limitations, and the implications for future research on both financial narratives and systematic literature analysis.

4.1 Adressing Research Questions

(RQ1) How can NLP and textual analysis techniques be used to quantify and model financial narratives? The reviewed literature reveals a clear evolution from static, sentiment-based proxies toward models that account for the structure, tone, and temporal dynamics of financial discourse. Early studies typically relied on lexicon-based sentiment scores or topic frequencies, which offered limited contextual understanding and failed to capture how meaning changes with context. More recent works apply transformer-based architectures (contextual embeddings or LLMs) that enable a more nuanced understanding of narrative formation and evolution. These approaches represent a shift from simply identifying textual sentiment to modeling the relationships between ideas, actors, and events as they unfold across time.

Another stream of research also seeks to represent narratives as continuous trajectories in semantic space, where their position and movement capture diffusion, persistence, and transformation. Embedding-based temporal models and context-aware clustering allow for detecting the emergence of new narratives while tracking existing ones as they evolve. Such methods move narrative modeling closer to dynamic systems analysis (capable of representing reinforcement, decay, and convergence between narratives) rather than treating them as fixed textual objects.

Despite these advances, the field remains fragmented. Few studies provide methodological benchmarks or reproducible validation protocols, and differences in data sources and preprocessing often limit comparability. Transformer models improve contextual understanding but also introduce challenges such as interpretability, computational cost, and out-of-sample testing. Overall, NLP techniques now make it possible to represent narratives as evolving, probabilistic

structures detected in financial texts, but consistent evaluation frameworks and methodological transparency are still missing.

(RQ2) How is the concept of financial narrative defined and applied, and what methodological patterns emerge? Across the literature, financial narratives are defined with varying nuances, ranging from broad communicative structures to more specific social mechanisms of belief formation. Despite these differences, a clear convergence has emerged: narratives are now viewed as structured interpretative frameworks that link facts, emotions, and expectations, influencing how information spreads, thus impacting financial markets.

Methodologically, most empirical researches operationalize narratives as clusters of semantically related texts or as latent features extracted from large corpora. The consensus is that narratives are not micro-level topics or isolated opinions but recurrent, large-scale interpretive patterns that structure discourse over time. This perspective implies that purely sentiment-based approaches, while informative about emotional tone, are insufficient for advanced narrative modeling. Sentiment captures a snapshot of mood, whereas narratives encompass the attention share, evolution and linkage of ideas that give direction to financial markets.

Consequently, progress in this field depends on treating narratives as evolving informational processes rather than static text aggregates. Future approaches should aim to capture how narratives emerge, persist, and recur across time, assigning latent intensity scores to narrative structures while allowing to detect new ones as discourse evolves. Models that can adapt to changing contexts, learn from new data, and represent the temporal continuity of narratives would make it possible to study not only their semantic content but also their life cycle: how they grow, transform, and fade.

4.2 Methodological Contribution

This paper contributes methodologically by introducing an algorithmic framework for systematic literature reviews that improves reproducibility, efficiency, and selection quality assessment. Unlike traditional SLRs based solely on manual screening, the proposed framework combines semantic embeddings, dimensionality reduction, and clustering to automate and document inclusion decisions.

The framework differs from prior bibliometric approaches in two main ways. First, it evaluates study relevance semantically rather than lexically, enabling identification of conceptually related works even when terminology and key words evolve. Second, it quantifies selection quality using the number of papers retained (larger sets indicate fewer false exclusions), the average relevance and Silhouette scores, instead of relying solely on researcher judgment. These components enhance methodological transparency, reduce subjectivity, and make the process more scalable.

In practice, these improvements affect efficiency and reproducibility through several channels. Automation reduces manual screening time and ensures that selection decisions follow consistent semantic criteria. The standardized embedding and clustering pipeline makes the review replicable across domains, while quantitative quality metrics provide objective validation of the selection process. The approach also supports iterative refinement: new literature can be incorporated by re-embedding and re-clustering, allowing systematic reviews to evolve rather than remain fixed snapshots.

4.3 Limitations

Despite the robustness of the framework, several limitations remain. The first concerns database coverage: Scopus was selected for its reliability, comprehensive metadata, and accessible API. However, its focus on peer-reviewed journals inevitably narrows the scope of analysis and may

omit early-stage or gray literature, particularly in fast-moving domains such as NLP and financial AI. A pragmatic solution is to combine Scopus with open bibliometric sources such as OpenAlex, which aggregates content from repositories. This hybrid approach could capture both established and emerging research, reducing potential source bias. A second area for improvement relates to data accessibility and document selection. Excluding publications without full-text access ensured consistent processing but introduced an unavoidable bias toward open-access or recently published papers.

Another consideration concerns the embedding and clustering pipeline. Transformer models provide strong semantic alignment but reflect the biases and temporal characteristics of their pre-training data, and the model choice is crucial. Bidirectional encoders adapt more smoothly to shifts in vocabulary and phrasing than autoregressive GPT-style models, which makes them somewhat more robust when new terms or domain-specific expressions enter the corpus. Additionally, fine-tuning using domain-specific corpora or contrastive learning could mitigate these issues. The current exclusion of medium-relevance clusters, while enhancing precision, may also remove studies at the edge of the topic that still contribute valuable perspectives. A semi-automatic variant where medium clusters are manually reviewed under transparent inclusion rules (e.g. words presences in abstract or keywords, similarity score with one or subset of scores, etc.) would balance reproducibility with interpretative flexibility, strengthening the overall robustness of the review.

Finally, the practical implementation of algorithmic literature reviews raises trade-offs between transparency, efficiency, and scalability. The choice between open- and closed-source language models could eventually affect not only accuracy but also cost, reproducibility, and control over the pipeline. Comparable open-source models allow greater transparency and fine-tuning to specific research domains but require significant computational resources. Closed-source models, in contrast, provide higher semantic precision and lower setup complexity but at the expense of interpretability, long-term accessibility, and dependency on proprietary infrastructures. Beyond the model choice, scalability also remains a challenge: embedding and clustering a few hundred abstracts is manageable on a standard workstation, but extending the process to tens of thousands of papers would require advanced computation resources.

4.4   Implications and Future Research Directions

The findings of this review carry implications for both the study of financial narratives and the methodology of literature synthesis.

From a conceptual angle, financial narrative research would benefit from unified framework and stronger theoretical integration with behavioral finance and information economics. Future studies should establish consistent typologies and empirically test how these categories map to market phenomena such as volatility, liquidity, or structural breaks.

From a methodological perspective, several strands of work appear particularly valuable. Comparative studies across narrative scales, from firm-specific disclosures to sectoral texts and broad macroeconomic discourse, would help clarify which NLP techniques perform best in each setting. Interpretability remains essential, since the increasing reliance on deep models requires an understanding of how textual features translate into narrative signals. A further step is to link narrative measures with quantitative market data in a coherent framework, which would make it possible to study how changes in discourse relate to price formation and other financial outcomes.

From a practical perspective, narrative insights are beginning to influence financial practice, from risk monitoring to regulatory stress testing. However, the reviewed literature still lacks standardized evaluation frameworks or open benchmarks to assess the robustness of such indicators.

Collaborative efforts between academia, regulators, and industry could accelerate the development of shared datasets and transparent validation protocols.

Overall, this review shows that the field of financial narrative modeling is advancing rapidly but unevenly. Conceptual fragmentation, data limitations, and methodological opacity remain key barriers. Addressing these issues will require stronger theoretical integration, standardized empirical protocols, and reproducible computational tools.

5   Conclusion

This paper introduces a systematic literature review framework that combines algorithmic paper selection with structured data extraction to investigate how financial narratives are conceptualized, quantified, and applied in financial market analysis. Leveraging NLP and machine learning tools, the review process refined an initial dataset of 288 papers down to 16 core contributions, classified into two main categories: narrative understanding and narrative modeling. The effectiveness of the selection framework is supported by the relevance and depth of the insights obtained from the final corpus. The selected research provides key contributions toward addressing the guiding research questions, highlighting the evolving conceptualization of narratives, the methodological advances enabling their modeling, and the potential applications of narrative-based signals in understanding and forecasting financial market behavior.

Studies under narrative understanding mainly seek to define the theoretical, semantic, and rhetorical boundaries of financial narratives. They offer annotation schemes, linguistic typologies, and structural representations that frame narratives as evaluative, temporal, and action-oriented discourses. These works contribute essential groundwork for developing more robust computational models. Secondly, the narrative modeling stream treats narratives as predictive signals. These studies use narrative-derived features to forecast inflation, returns, systemic risk, and investor behavior. They demonstrate that narratives can improve both the explanatory and predictive power of econometric and machine learning models, especially during periods of market dislocation or heightened uncertainty. Importantly, a few of these contributions go beyond traditional sentiment or topic analysis and incorporate structural and behavioral dimensions, such as narrative virality, emotional polarity, and semantic shift.

The review highlights both the richness and the fragmentation of the current literature. On the one hand, there is increasing recognition that financial narratives are more than expressions of sentiment or thematic frequency: they are structured, evolving, and socially embedded interpretations of economic phenomena. On the other hand, approaches to operationalizing and modeling these narratives remain diverse, with varying assumptions about what constitutes a meaningful narrative signal.

From a methodological standpoint, the field has progressed rapidly, incorporating unsupervised topic modeling, deep learning, and large language models that enabled the construction of narrative indices that reflect attention, sentiment, emotion, and network structure. Foundational work continues to refine the linguistic, rhetorical, and epistemic dimensions of financial narratives, suggesting the need for continuous improvements in technical approaches. Consequently, this review finds that NLP and textual analysis techniques offer a broad set of tools for quantifying financial narratives in a systematic manner. The evidence strongly suggests that incorporating narrative modeling enhances our understanding of market dynamics, as narrative-based indicators consistently capture behavioral signals that conventional financial and macroeconomic variables often overlook. These indicators not only contribute to improved forecasting performance but also enhance the interpretability of those predictions, providing a structured lens through which to analyze market sentiment, contagion effects, and regime shifts—phenomena that are otherwise difficult to quantify.

In conclusion, the literature reviewed here supports the growing view that narratives are not peripheral to financial markets but are central to how agents form beliefs, process information, and make decisions. As tools for extracting and modeling narratives become more sophisticated, their role in empirical finance, policy analysis, and risk management is likely to expand. Nonetheless, current approaches face several limitations: the representativeness of textual data remains uncertain; interpretability of complex NLP models is still a concern; and the validation of narrative indicators across market regimes is limited. Addressing these challenges will be critical to establishing narratives as robust, transparent, and actionable components of financial analysis.

List of Abbreviations

- **AC** – Agglomerative Clustering

- **API** – Application Programming Interface

- **CN** – Condition Number

- **GMM** – Gaussian Mixture Model

- **IFRS** – International Financial Reporting Standards

- **KMO** – Kaiser–Meyer–Olkin Test

- **LDA** – Latent Dirichlet Allocation

- **ML** – Machine Learning

- **NEG** – Negative Emotion Score

- **NLP** – Natural Language Processing

- **PCA** – Principal Component Analysis

- **SLR(s)** – Systematic Literature Review(s)

References

Vineet Agarwal, Richard J. Taffler, and Chenyang Wang. 2024. Investor emotions and market bubbles. *Review of Quantitative Finance and Accounting*.

Alessandra Amato, Joerg R. Osterrieder, and Marcos R. Machado. 2024. How can artificial intelligence help customer intelligence for credit portfolio management? A systematic literature review. *International Journal of Information Management Data Insights*, 4(2):100234.

Daniel Borup, Jorge Wolfgang Hansen, Benjamin Dybro Liengaard, and Erik Christian Montes Schütte. 2023. Quantifying investor narratives and their role during COVID-19. *Journal of Applied Econometrics*, 38(4):512–532.

Leland Bybee, Bryan Kelly, Asaf Manela, and Dacheng Xiu. 2024. Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147.

Leland Bybee, Bryan Kelly, and Yinan Su. 2023. Narrative asset pricing: Interpretable systematic risk factors from news text. *The Review of Financial Studies*, 36(12):4759–4787.

Yuting Chen, Don Bredin, Valerio Potì, and Roman Matkovskyy. 2022. COVID risk narratives: a computational linguistic approach to the econometric identification of narrative risk during a pandemic. *Digital Finance*, 4(1):17–61.

Otávio José de Oliveira, Fabio Francisco da Silva, Fernando Juliani, Luis César Ferreira Motta Barbosa, and Thaís Vieira Nunhes. 2019. Bibliometric method for mapping the state-of-the-art and identifying research gaps and trends in literature: An essential instrument to support the development of scientific projects. In Suad Kunosic and Enver Zerem, editors, *Scientometrics Recent Advances*, chapter 3. IntechOpen, London.

Joel P. Flynn and Karthik Sastry. 2024. The macroeconomics of narratives. *NBER Working Paper No. w32602*.

Sanford J. Grossman. 1980. On the Impossibility of Informationally Efficient Markets.

Yongmiao Hong, Jiang Fuwei, Meng Lingchao, and Bowen Xue. 2025. Forecasting Inflation Using Economic Narratives. *Journal of Business & Economic Statistics*, 43(1):216–231.

Ching Hsu, Tina Yu, and Shu-Heng Chen. 2021. Narrative economics using textual analysis of newspaper data: new insights into the U.S. Silver Purchase Act and Chinese price level in 1928–1936. *Journal of Computational Social Science*, 4(2):761–785.

Jiahui Hu, Patrick Paroubek, and Dirk Schumacher. 2021. Annotation model and corpus for opinionated economy and finance narrative detection. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 61–66, Lancaster, United Kingdom. Association for Computational Linguistics.

Jiaxin Liu, Yi Yang, and Kar Yan Tam. 2024. Beyond Surface Similarity: Detecting Subtle Semantic Shifts in Financial Narratives.

Tian Ma, Ganghui Li, and Huajing Zhang. 2024. Stock return predictability using economic narrative: Evidence from energy sectors. *Journal of Commodity Markets*, 35:100418.

Deborah Miori and Constantin Petrov. 2023. Narratives from GPT-derived Networks of News, and a link to Financial Markets Dislocations.

Jason Priem, Heather Piwowar, and Richard Orr. 2022. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts.

Michael Roos and Matthias Reccius. 2024. Narratives in economics. *Journal of Economic Surveys*, 38(2):303–341.

Robert J. Shiller. 2017. Narrative economics. *American Economic Review*, 107(4):967–1004.

Robert J. Shiller. 2020. Popular economic narratives advancing the longest U.S. expansion 2009–2019. *Journal of Policy Modeling*, 42(4):791–798.

Robert James Shiller. 2019. *Narrative economics: how stories go viral & drive major economic events*. Book collections on project MUSE. Princeton University press, Princeton.

Margaret R. Somers. 1994. The narrative constitution of identity: A relational and network approach. *Theory and Society*, 23(5):605–649.

Yolanda S. Stander. 2024. A News Sentiment Index to Inform International Financial Reporting Standard 9 Impairments. *Journal of Risk and Financial Management*, 17(7):282.

Eugene Sy, Tzu-Cheng Peng, Shih-Hsuan Huang, Heng-Yu Lin, and Yung-Chun Chang. 2023. Fine-Grained Argument Understanding with BERT Ensemble Techniques: A Deep Dive into Financial Sentiment Analysis. In *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, pages 242–249, Taipei City, Taiwan. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

Richard J. Taffler, Vineet Agarwal, and Maximilian Obring. 2024. Narrative Emotions and Market Crises. *Journal of Behavioral Finance*, 0(0):1–21.

David Tuckett and Milena Nikolic. 2017. The role of conviction and narrative in decision-making under radical uncertainty. *Theory & Psychology*, 27(4):501–523.

David Tuckett, Robert Elliot Smith, and Rickard Nyman. 2014. Tracking phantastic objects: A computer algorithmic investigation of narrative evolution in unstructured data sources. *Social Networks*, 38:121–133.

P S Varsha, Amrita Chakraborty, and Arpan Kumar Kar. 2024. How to undertake an impactful literature review: Understanding review approaches and guidelines for high-impact systematic literature reviews. *South Asian Journal of Business and Management Cases*, 13(1):18–35.

Enwei Zhu, Jing Wu, Hongyu Liu, and Keyang Li. 2023. A Sentiment Index of the Housing Market in China: Text Mining of Narratives on Social Media. *The Journal of Real Estate Finance and Economics*, 66(1):77–118.

Nadhem Zmandar, Tobias Daudert, Sina Ahmadi, Mahmoud El-Haj, and Paul Rayson. 2022. CoFiF Plus: A French Financial Narrative Summarisation Corpus. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1622–1639, Marseille, France. European Language Resources Association.
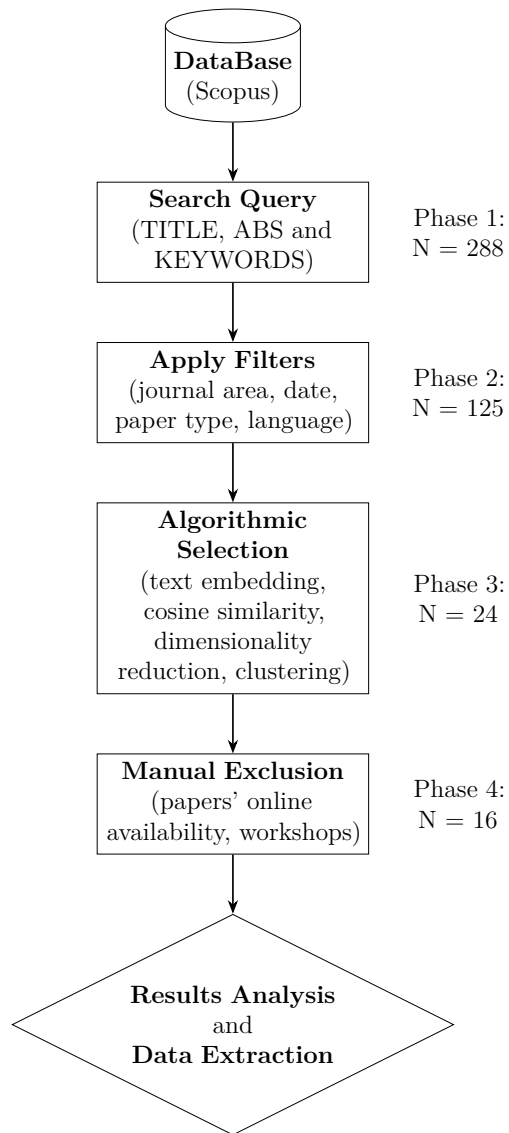
Appendix

Figure 2: Schematic representation of the paper selection process. The process includes database querying, filtering by inclusion/exclusion criteria, algorithmic selection via NLP and clustering, and manual exclusion of inaccessible studies or workshop proceeds.

Table 3: Summary of the data extraction phase.

Paper	Label	Purpose	Method	Narrative
Tuckett et al. (2014)	narrative modeling	Track evolution of emotionally charged financial narratives (phantastic objects) before crises.	Emotion keyword dictionaries, sentiment shift scoring, network analysis of sender-receiver patterns.	Excitement vs. anxiety sentiment index + social network clustering reveal narrative divergence before collapse.
Hu et al. (2021)	narrative understanding	Develop a corpus and model for detecting opinion and emotion in financial narratives.	Manual + SpaCy annotation using appraisal theory; dependency parsing, syntactic tagging.	Appraisal-based annotations, intra-sentence pairings of opinion and targets.
Hsu et al. (2021)	narrative modeling	Assess how narrative topics in newspaper articles relate to macroeconomic indicators in 1930s China.	Keyword frequency tracking, Ridge/LASSO/Elastic Net regressions, VAR, IRF, Granger causality.	Manual keyword selection and normalized frequency analysis linked to time series regression models.
Zmandar et al. (2022)	narrative understanding	Create a large-scale French corpus for summarizing financial reports.	Heuristic + CamemBERT + manual summary extraction.	Summary-level mappings from report sections; NER for entity highlights.
Chen et al. (2022)	narrative modeling	Assess narrative influence during COVID on market variables using causal testing.	LDA, LM sentiment, Word2Vec, SIR virality, VAR and Granger causality.	Narratives via LDA, semantic shift, and virality scoring, linked to econometric causality.
Zhu et al. (2023)	narrative modeling	Build a future-oriented sentiment index from Weibo posts on housing.	LSTM sentence classification into temporal/sentiment classes, using Word2Vec.	Narrative sentiment split by temporal framing and learned via deep LSTM classifier.

Paper	Label	Purpose	Method	Narrative
Sy et al. (2023)	narrative understanding	Improve financial sentiment analysis via argumentative unit detection.	BERT ensemble for argument classification and relation detection.	Pairwise relation classification and claim-premise detection with BERT ensemble.
Miori and Petrov (2023)	narrative modeling	Use GPT + graph theory to extract narrative structure from news and link to market dislocations.	Entity extraction via GPT; co-occurrence graphs; community detection; regress network features on volatility shocks.	GPT-ranked entities + graph metrics (modularity, entropy); topic communities track narrative complexity.
Borup et al. (2023)	narrative modeling	Analyze investor narratives via open-ended surveys during COVID and assess predictive value for markets.	LDA on survey texts; Elastic Net VARs; comparison with media narratives.	LDA-derived narrative topics tracked daily and linked to market behavior via econometric models.
Stander (2024)	narrative modeling	Construct a news sentiment index to act as an early warning for systemic risk and credit impairments.	FinBERT sentiment scoring, PCA on macro indicators, rolling regressions, aspect-based sentiment.	FinBERT sentiment, topic-based aspect analysis, PCA and regressions to link to macro risk.
Liu et al. (2024)	narrative understanding	Detect subtle semantic shifts in firm narratives using new financial-STS task.	Triplet-based contrastive learning with GPT-labeled sentence triplets.	Triplet embedding training for narrative similarity under financial framing.
Agarwal et al. (2024)	narrative modeling	Measure investor emotions from media during market bubbles and link them to returns, volatility, and volume.	Emotion keyword dictionaries; regression on market variables during two Chinese stock bubbles.	Emotion frequency indices (8 emotions), strong vs. weak emotion separation, applied in regression models.

Paper	Label	Purpose	Method	Narrative
Taffler et al. (2024)	narrative modeling	Analyze emotional content of narratives during market crises and its explanatory power over returns and uncertainty.	Keyword-based emotion indices; regression with returns, VIX, volume, EPU during three crises.	Context-specific emotion dictionary applied to financial articles; emotion score time series regressed on market indicators.
Roos and Reccius (2024)	narrative understanding	Review and define the concept of collective economic narratives in economics.	Theoretical review of 436 papers; derive five defining narrative features.	Defines narratives as temporal, socially emergent, sense-making, and action-suggesting story structures.
Ma et al. (2024)	narrative modeling	Use WSJ-derived narrative index (NEG) to forecast stock returns in the energy sector and beyond.	LDA on WSJ energy topics; predictive regressions; Sharpe/CE utility evaluation.	NEG index from topic attention on WSJ news, tracked as monthly time series and forecast factor.
Hong et al. (2025)	narrative modeling	Forecast U.S. inflation using WSJ-derived narrative features.	LDA for topic modeling, ML regressors (LASSO, ENet, RF, PLS).	LDA narrative topics as ML features in out-of-sample inflation prediction.