Designing an Investment Research System for Asset Management Based on Natural Language Processing

By

Yanzhang Chen

Submitted to the MIT Sloan School of Management in Partial Fulfillment of the Requirements of the Degree of

MASTER OF SCIENCE IN MANAGEMENT STUDIES

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Yanzhang Chen
MIT Sloan School of Management
May 10th, 2023

Certified by: Simon Johnson
Ronald A. Kurtz (1954) Professor of Entrepreneurship
Thesis Supervisor

Accepted by: Jacob Cohen
Senior Associate Dean for Undergraduate & Master's Program
MIT Sloan School of Management

Designing an Investment Research System for Asset Management Based on Natural Language Processing

By
Yanzhang Chen

Submitted to MIT Sloan School of Management
on May 7, 2023, in Partial Fulfillment of the
requirements for the Degree of Master of Science in
Management Studies.

ABSTRACT

In recent years, the asset management industry has experienced rapid growth, with the global asset management scale continuously increasing. Conventionally, investment research in asset management entails the acquisition of data and information from a myriad of sources, which is then manually processed and analyzed. However, in the face of macroeconomic volatility, fierce competition, and a deluge of fragmented information, this traditional approach to investment research increasingly struggles to manage the sheer volume of financial market data and information.

Natural Language Processing (NLP), an essential subset of artificial intelligence, has achieved significant breakthroughs in recent years. It facilitates automatic processing, analysis, and text generation to specific tasks, aiding investment institutions in swiftly assimilating and dissecting massive volumes of information, and consequently formulating investment research results. NLP can assist investment institutions in rapidly integrating and analyzing vast information and automatically generating investment reports. This paper aims to trace the evolution of NLP, evaluate its prospective positive impact on asset management, and deliberate on designing an investment research system grounded in NLP technology.

Thesis Supervisor: Simon Johnson
Title: Ronald A. Kurtz (1954) Professor of Entrepreneurship

Acknowledgment

I would like to begin by expressing my sincere gratitude to my advisor, Prof. Simon Johnson, for his invaluable guidance and mentorship throughout my graduate studies. Prof. Simon Johnson's extensive knowledge, rigorous scholarly attitude, and profound insights into the business have greatly influenced me, allowing me to learn and grow immensely. I truly appreciate the wisdom and support he has provided me during this journey.

Additionally, I wish to extend my heartfelt appreciation to the Massachusetts Institute of Technology (MIT) for the exceptional educational experience I have received during my memorable one-year tenure. The knowledge and skills I have acquired at this esteemed institution have broadened my perspectives and deepened my understanding of my chosen field. I am grateful for the outstanding opportunities MIT has afforded me.

Moreover, I am profoundly grateful to my family for their unwavering support and encouragement throughout my academic pursuits. Their love, understanding, and sacrifices have been the bedrock upon which my personal growth has been built. I dedicate this thesis to them, as a testament to their steadfast faith in me and my aspirations.

In conclusion, I would like to acknowledge and thank all the individuals who have contributed to my personal growth and academic journey, both directly and indirectly. Your support has made this milestone in my academic career a reality.

Content

1. Introduction 4
2. Research Background 5
2.1 Introduction of Artificial Intelligence 5
2.2 Overview of Natural Language Processing 6
2.3 Overview of the Asset Management Industry 10
2.4 The Beneficial Impact of Natural Language Processing on Asset Management 13
3. Introduction to Investment Research Systems Based on Natural Language Processing 16
3.1 Definition 16
3.2 System Workflow 16
3.3 Current Application 18
3.4 Current Research 18
4. Designing an Investment Research System Based on Natural Language Processing 20
4.1 System Architecture 20
4.2 The Data Platform 21
4.3 The Algorithm Platform 23
4.4 The Application Platform 33
4.5 The Management Platform 35
4.6 The Visualization Front-end 36
5. Development Trends and Challenges 38
5.1 Development Trends of Investment Research Systems 38
5.2 Challenges Faced by Investment Research Systems 38
6. Suggestions for Future Strategies 41
6.1 Introducing New NLP Algorithms 41
6.2 Enhancing Industry Data Mining 41
6.3 Strengthening Compliance Construction 41
6.4 Boosting System Security 42
6.5 Establishing Competent Manual Review Mechanisms 42
7. Conclusion and Outlook 43
References 45

Designing an Investment Research System for Asset Management Based on Natural Language Processing

1. Introduction

The asset management industry is an integral part of the financial market, and it has a significant impact on the stability of the global financial system. Ensuring the development of the asset management industry is of utmost importance. With the advancements and maturity of natural language processing (NLP), asset managers can leverage this technology to analyze vast amounts of data, enabling more efficient investment research and enhancing the operational efficiency of asset management.

Therefore, this thesis aims to analyze how to design an investment research system for asset management based on natural language processing. By exploring the current state of asset management and the characteristics of NLP, the thesis will design an investment research system, analyze relevant risks, and provide suggestions for future development strategies.

2. Research Background

2.1 Introduction of Artificial Intelligence

As technology has grown quickly, Artificial Intelligence (AI) has emerged as an important area of scholarly inquiry. AI, a subdivision of computer science, is devoted to equipping computers with the ability to mimic human intelligence. The genesis of this concept can be traced back to ancient Greece, when philosophers first deliberated on the potential of machines to exhibit human-like cognition. The pivotal turning point in the evolution of AI arrived in 1936, with British mathematician and computer science pioneer, Alan Turing, introducing the renowned "Turing Machine Model". Turing further elaborated on the principles of electronic digital computer design in 1945, and in 1950, proposed the "Turing Test". This test involved an individual conducting a series of Q&A sessions with another party, using a specified protocol, without any direct contact. If the individual, over a significant duration, cannot discern whether the respondent is a human or a computer, the computer is deemed to have achieved a level of intelligence comparable to human cognition. The crux of the Turing Test lies in its assessment of machine intelligence through human-like behavior, as opposed to evaluating the hardware or software of the machines themselves. This innovative testing approach has shifted the emphasis of AI research towards realizing intelligent behavior, thereby significantly propelling the progression of artificial intelligence.

The term "artificial intelligence" was officially introduced at the Dartmouth Conference in 1956[i], a meeting widely acknowledged as the formal inception point of AI research. Subsequently, the development of artificial intelligence went through ups and downs. During the period from 1970s to 1980s, the development of artificial intelligence ushered in its first golden age. Research during this period predominantly centered around logical reasoning, leading to the birth of Logic Theorist, the inaugural AI software. In the 1980s, expert systems began to emerge, with the focus shifting to the extraction of rules from expert knowledge and transposing these rules into computer programs. Coinciding with the rise of computers, AI encountered its second golden age of development. In 1997, the computer Deep Blue[ii] defeated the world chess champion, which became a milestone event in the history of artificial intelligence development, and the pursuit of artificial intelligence reached a new peak. In the 21st century, with Hinton proposing deep learning technology and the successful digital transformation of various fields, artificial intelligence entered an unprecedented high-speed development period. This era has seen considerable advancements in image recognition, speech recognition, and natural language processing. Nowadays, the current trajectory of AI research emphasizes its deep integration with diverse industries, aiming to enhance the application value and efficiency of AI technology further.

2.2 Overview of Natural Language Processing

2.2.1 Concept of Natural Language Processing

Natural Language Processing (NLP), a significant technology within the realm of artificial intelligence, aims to equip computers with the ability to comprehend, analyze, produce, and manipulate human language. NLP technologies typically include speech recognition, text classification, information extraction, automatic question-answering, machine translation, and speech synthesis. The implementation of natural language processing techniques mainly involves the interdisciplinary application of linguistics, computer science, and artificial intelligence.

Natural language processing technology is characterized by high complexity and diversity. Different NLP tasks require different technical methods and algorithmic models, often requiring a combination of multiple techniques for comprehensive analysis and processing. For example, text classification requires the use of machine learning and deep learning technologies for text feature extraction and classification; automatic question-answering requires natural language generation and reasoning technologies for question analysis and answer generation; machine translation requires the integration of language models, translation memories, and neural machine translation, among other techniques.

2.2.2 Development of Natural Language Technologies

From 1950 to 1970, scholarly inquiry predominantly used rule-based methods, with primary emphasis on areas like machine translation and natural language understanding. In 1956, the theory of generative grammar emerged, formalizing language description and elucidating the syntactic and semantic structures inherent in natural languages. This theory presupposes an objective generative rule for natural languages, the extraction of which can enable humans to decode the intricacies of natural language. Subsequently, Symbolism based on linguistics became the mainstream, with extensive research focusing on analyzing the lexical and syntactic structure of natural language from a linguistic perspective and summarizing relevant rules to understand natural language. Generative grammar served as a pivotal theoretical underpinning in the field of natural language processing, yet it was not without its limitations. Generative grammar can only describe formalized language structures and cannot address natural language complexities such as ambiguity and discrimination. Furthermore, since the derivation process of generative grammar is based on rule-based recursive derivation, its derivation efficiency is greatly affected when dealing with more complex long sentences. Therefore, this method has gradually been replaced as technology has advanced. By the late 1960s, statistical learning-based methods began to be widely used in natural language processing. Early statistical methods mainly included n-grams[iii], Bayesian classification[iv], and hidden Markov models[v], which transformed natural language processing tasks into a probability inference problem and trained model parameters using large corpora. This approach markedly elevated the accuracy and efficiency of natural language processing, instigating substantial advancements in areas such as machine translation, speech recognition, and information retrieval. In the 1980s, knowledge representation and reasoning-based methods emerged in natural language processing. Knowledge-based methodologies can process language structure and meaning by codifying language knowledge in the guise of rules, semantic networks, or ontologies. Semantic networks used graphical models to depict language structure and semantic information as nodes and edges, while ontologies modeled the interrelationships between domain-specific concepts. Despite their considerable strides in knowledge representation and reasoning, these methodologies did not achieve widespread practical application due to the complexities inherent in knowledge representation and acquisition. At the dawn of the 21st century, the exponential surge in computational power catalyzed the emergence of deep learning methodologies predicated on neural networks in natural language processing. Deep learning methodologies automatically decipher language structure and semantic information by constructing deep neural network models, achieving significant breakthroughs in tasks such as text categorization, sentiment analysis, and semantic similarity computation. Moreover, the steady growth in language data and computational power has positioned deep learning methodologies as one of the mainstream approaches in natural language processing. The word2vec[vi] The technology proposed in 2013 generates word vectors containing rich semantic information through unsupervised training on large amounts of text. Since then, the evolution of natural language processing has experienced periods of sequence-based models (mainly RNNs[vii] and LSTMs[viii]), attention-based Transformer models, and pre-trained language models based on large-scale corpora, such as ELMo[ix], BERT[x], and GPT[xi].

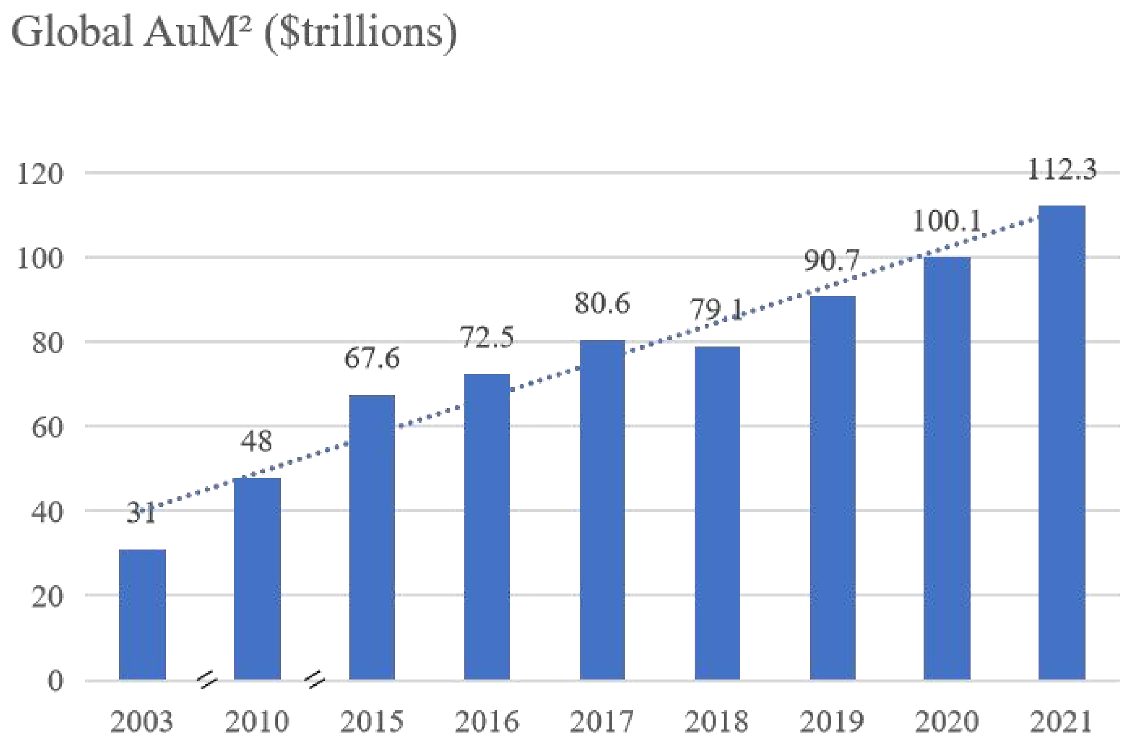
Fast-forward to 2023, the large-scale language model GPT 3.5, built on GPT framework, made a remarkable entry into natural language processing. GPT 3.5 has demonstrated astonishing competencies in natural language generation tasks, yielding fluent and natural text. Concurrently, GPT has exhibited superior performance in tasks such as text classification, machine translation, and sentiment analysis. This has reshaped our perception of machine intelligence, rapidly capturing global attention and propelling natural language processing back into the spotlight.

2.3 Overview of the Asset Management Industry

2.3.1 The development of the Asset Management Industry

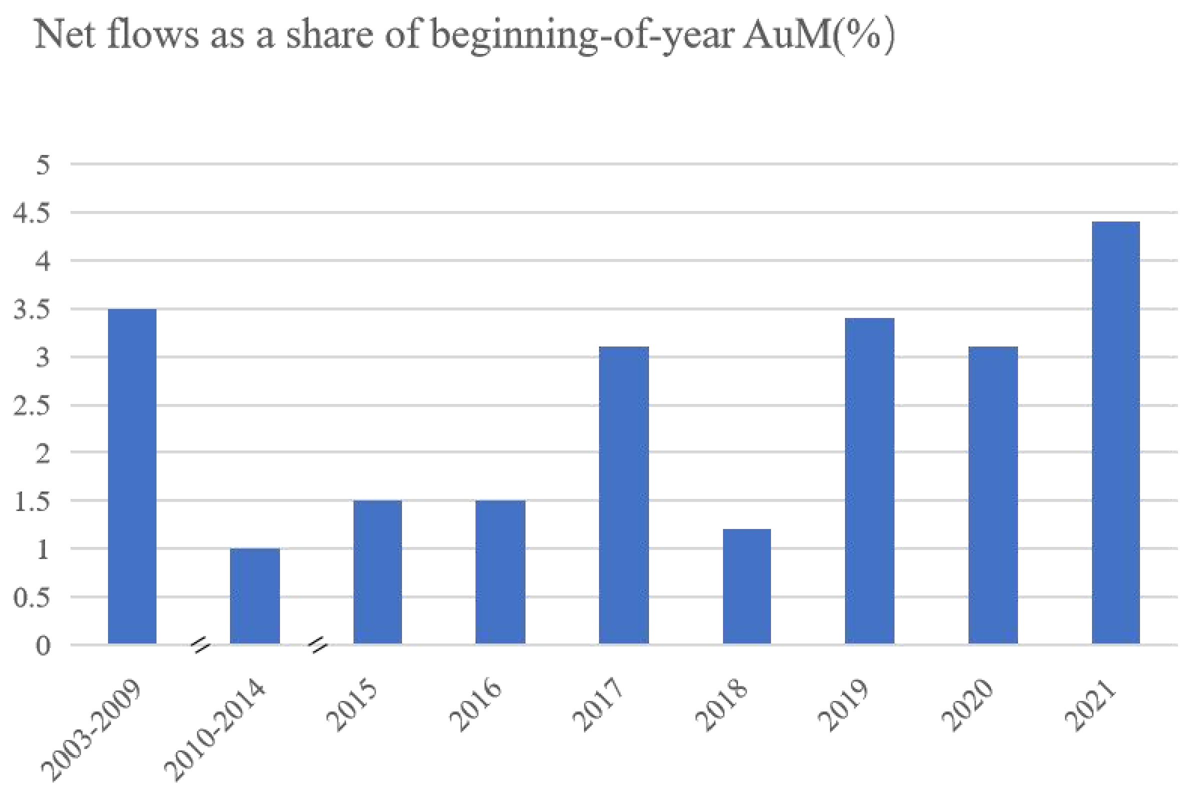
Asset management pertains to the systematic operation and administration of clients' assets by professional organizations, extending investment management services for an array of financial products to clients. The asset management industry offers a wide range of products and services, including stocks, bonds, money market funds, exchange-traded funds (ETFs), index funds, hedge funds, private equity, real estate, and infrastructure. Its customer base is diverse, spanning individual investors, family units, corporations, charitable entities, pension funds, insurance companies, and other investment institutions.

Over an extended timeline, the asset management industry has emerged as one of the most rapidly expanding sectors in terms of capital scale within the global financial services landscape. It plays a pivotal role in optimizing the allocation of financial resources, bolstering the efficiency of financial markets, and fostering innovation in financial instruments. From 2001 to 2021, the asset management industry underwent sustained growth. Positive net inflows into the asset management industry have remained consistent generally since 2001. In 2020, the global asset management market eclipsed the $100 trillion threshold, with 2021 witnessing even more robust growth as global assets under management surged by 12% to surpass $112 trillion. The asset management industry's growth trajectory persists, propelled by the ascent of emerging markets, demographic changes resulting in heightened demand for retirement funds, and an escalating focus on sustainable investment.



Graph1: Global AuM ($trillions)

Data Source：BCG 2022 Global Asset Management Report



Graph2: Net flows as a share of beginning-of-year AuM (%)

Data Source：BCG 2022 Global Asset Management Report

2.3.2 Challenges Faced by the Asset Management Industry

Despite the rapid development and expansion of the asset management industry on a global scale, and its ability to sustain a high-speed growth trajectory over the past two decades, it faces increasingly more challenges due to the industry's rapid development, macro-environmental changes, and deepening digital transformation.

2.3.2.1 Intensified Industry Competition

As a large amount of capital flows into the asset management industry, the number of asset managers also grows, leading to increasingly fierce competition within the industry. The overall management fee rate in the asset management industry continues to decline, resulting in reduced income for asset managers, decreased profitability, and increased operational pressure. To achieve alpha returns[xii] in this hyper-competitive environment, asset managers need to persistently optimize their services, enhance investment returns, and curtail costs.

2.3.2.2 Macro-environmental Turbulence

The asset management industry is affected by macroeconomic and political environments. Instabilities in the global economic situation and political landscape can severely impact the industry. Factors such as trade wars, geopolitical risks, and changes in monetary policy can all trigger market fluctuations. Moreover, the volatility of the macro-environment can lead to fluctuating investor sentiment, heightening the risks associated with asset management investment decisions. This places more stringent requirements on the investment strategies and risk management capabilities of asset managers.

2.3.2.3 Escalating Market Noise

During the asset management process, investment managers must gather extensive information and data and undertake profound analysis to generate proprietary insights of higher informational value. As the global digital transformation deepens, data accumulated across various information platforms increases, and correspondingly, the information noise in the investment market intensifies. This brings a series of challenges to the asset management industry. Much of the data are unstructured or fragmented and, without processing, cannot facilitate investment decision-making. This necessitates investment managers to devote more time and energy to reading and analyzing information to filter out valuable insights, thereby placing significant pressure on asset managers.

2.4 The Beneficial Impact of Natural Language Processing on Asset Management

As an artificial intelligence technology capable of processing information on a large scale, capturing features, and generating text, natural language processing (NLP)has the potential to assist asset managers in mining investment opportunities from copious amounts of unstructured information, monitoring real-time financial events, and automatically generating investment research recommendations. There is a growing trend of asset managers incorporating NLP into their investment research practices.

2.4.1 Enhancement of Investment Research Quality

NLP technology can help asset managers improve the quality of investment Research. By analyzing extensive textual data, including news, research reports, and social media, NLP can offer investment professionals profound insights into companies, industries, and macroeconomics. This assists investment managers in gaining a more comprehensive understanding of market dynamics and potential risks, enabling them to devise more precise investment strategies and enhance investment returns.

2.4.2 Augmentation of Research Efficiency in Asset Managers

NLP technology can substantially augment the research efficiency of asset managers. When confronted with an overwhelming amount of textual data, manual screening and analysis of information prove to be limited in speed and effectiveness. However, NLP can swiftly process these data, capturing key information and trends, thereby providing invaluable support for investment research. This not only curtails labor costs but also elevates research quality and accuracy.

2.4.3 Expansion of Financial Event Monitoring Scope

Natural language processing can extend the monitoring scope of financial events that bear an impact on investment portfolios. With NLP technology, asset managers can widen their monitoring range. For instance, in the event of a risk occurring at a national or regional level, asset managers can leverage NLP to promptly identify the event's impact on all companies within their investment portfolios.

As natural language processing technology continues to evolve, it will progressively integrate with the asset management industry, fostering the robust development of asset managers.

3. Introduction to Investment Research Systems Based on Natural Language Processing

3.1 Definition

An investment research system based on natural language processing technology is an instrumental tool that leverages NLP to scrutinize, extract, and structure unstructured textual data, ultimately providing valuable insights to asset managers in support of their investment decisions. This investment research system can process various types of textual data, such as financial reports, stock prices, and economic indicators, extracting key information and trends, and analyzing and predicting them, thereby helping investment managers gain a comprehensive understanding of market dynamics, corporate information, and investment opportunities. Additionally, the system can analyze news reports and social media discourse using NLP technology to gauge public opinion and sentiment about a particular company or industry, thus providing a reference for investment decisions.

3.2 System Workflow

An investment research system unfolds through a methodical sequence of four distinct steps: data collection, data cleaning, analysis and research, and result output. The first step is data collection, following this initial steps, technologies such as Natural Language Processing (NLP) are used to process and analyze the collected data, enhancing its utility and comprehensibility. The final step is result output, summarizing investment research findings in an intelligible and accessible format.

3.2.1 Data Collection

Data collection represents the most fundamental step. Upon determining the appropriate data scope, various strategies must be devised to collect and store data in a structured manner on a regular and irregular basis, providing the raw materials for other operational segments.

3.2.2 Data Cleaning

In the preceding step, the unstructured raw data necessary for investment research was prepared. This raw data requires processing according to specific operational logic, which includes cleaning, filtering, extraction, and computation. Data processed in accordance with this logic can subsequently be analyzed to uncover patterns and shifts in data or operations.

3.2.3 Analysis and Research

This phase constitutes the critical core of the entire system. During this stage, the system employs NLP technology, underpinned by deep learning and machine learning foundations, to perform trend analysis, factor analysis, and event analysis based on the processed data.

3.2.4 Result Output

The final output will be presented in a visually comprehensible format, providing business and decision support for financial investment and research. This phase is the output stage and can take various forms, such as web pages, TXT, Word, charts, and public account graphics, assisting asset managers in efficiently understanding the investment advice generated by the system.

3.3 Current Application

The application of NLP within investment research is still in its nascent stages. The United States pioneered exploration in this domain; as far back as 2000, BlackRock initiated the development of the Aladdin system. The system employed NLP technology to monitor risks, intelligently interpret documents, and furnish risk management and investment advisory information services. Concurrently, Europe and Asia are also making strides, cultivating a burgeoning crop of startups in the process. After more than two decades of evolution, several notable companies have emerged within this field, such as Kensho, AlphaSense, Visible Alpha, Dataminr, and Econob.

Company	**Focus**
Kensho	Finding the relationship between events and assets, and predicting asset prices
Trefis	Breaking down various products/businesses of companies and predicting their revenue
AlphaSense	An intelligent search engine designed for professional investors
Visible Alpha	Established a proprietary new dataset and tool suite to enhance institutional investors' ability to quantify fundamental insights into companies' futures. Acquired ONEaccess in 2017 and integrated it into the Visible Alpha platform to improve the workflow of discovering, tracking, and evaluating all sell-side interactions
Dataminr	Serving financial and government institutions by converting real-time data obtained from public sources such as Twitter into actionable signals.
Econob	Allows for trading based on real-time news. ATRAP system checks facts whenever news is released. If certain conditions are met, it alerts traders or automatically places orders in the market.

Graph3: Companies using NLP in asset management
Data Source: Company Websites

3.4 Current Research

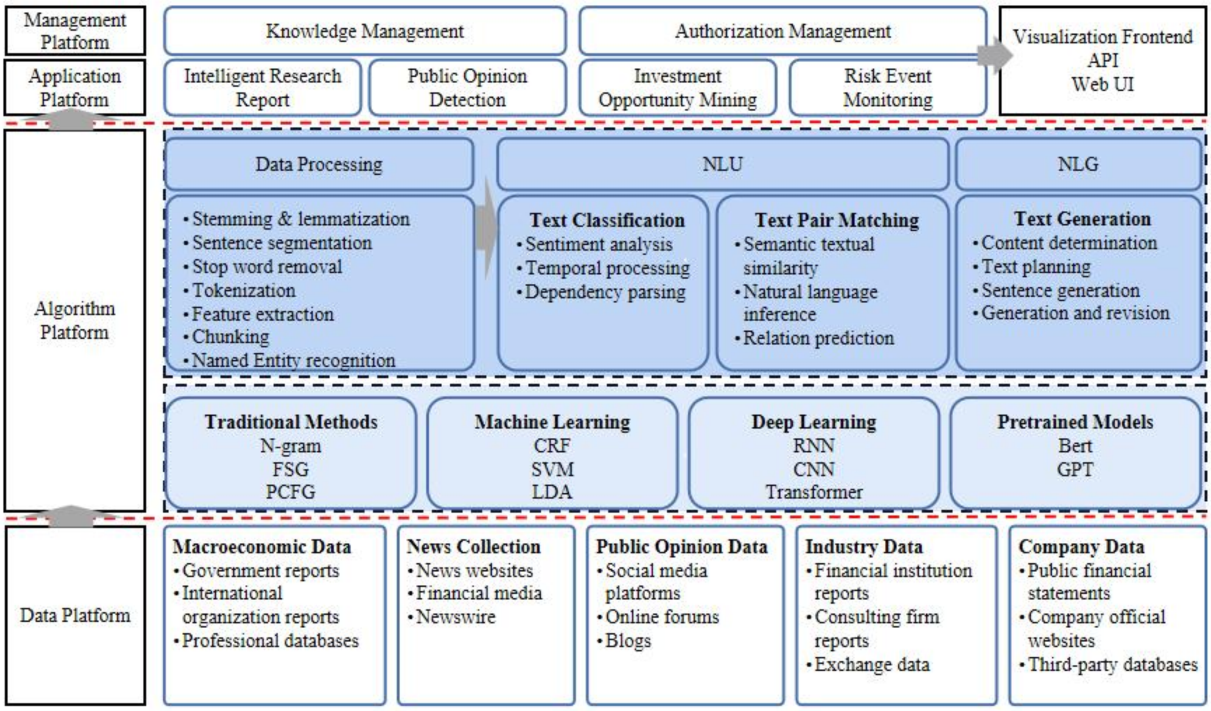
Presently, research on the application of Natural Language Processing (NLP) in the field of asset management predominantly concentrates on how to utilize NLP technology for textual analysis in the financial domain. The bulk of these studies focus on the algorithmic aspects of NLP, including topics such as data cleaning, semantic understanding, and portfolio construction. These investigations delve into the developmental trajectory and technical features of NLP, and how it aids investors in extracting useful information from vast quantities of unstructured data to form investment portfolios. They also explore how NLP can be applied through algorithms for tasks such as investment information filtering and financial sentiment monitoring.

However, in comparison, there is a relative dearth of research on the construction of investment research systems. While existing studies have demonstrated the immense potential of NLP technology in investment research, the integration of these techniques into a comprehensive investment research system for the realization of more efficient and accurate investment decisions remains an area that is not fully explored. Therefore, future research might benefit from placing more emphasis on the architectural design of investment research systems, in order to better harness the potential of advanced technologies like NLP in improving the quality and efficiency of investment decisions.

4. Designing an Investment Research System Based on Natural Language Processing

4.1 System Architecture

An Investment Research System primarily comprises the data platform, algorithm platform, application platform, management platform, and visualization front-end. The data platform is tasked with the collection and storage of diverse financial data, including macroeconomic data, corporate financial statements, financial market data, and public opinions, thereby providing a robust data foundation for the system. The algorithm platform leverages an array of machine learning and natural language processing techniques to cleanse,



Graph4: The structure of the investment research system

4.2 The Data Platform

The data platform serves as the foundation of an investment research system, primarily collecting data related to investment research through various sources. These data mainly include macroeconomic data, news information data, public media data, industry data, and corporate data. The key objective of the data platform is to ensure the accuracy, completeness, and timeliness of the data, providing a reliable information foundation for the subsequent analysis conducted by the investment research system.

4.2.1 Macroeconomic Data

Macroeconomic data, encapsulating the overall economic health of a nation or region, is primarily utilized in investment research to discern market trends, aiding investment managers in understanding economic cycles and policy environments. The primary sources of this data include government agency reports (e.g., national statistical bureaus, central banks), international organization reports (e.g., World Bank, International Monetary Fund), and professional databases (e.g., CEIC, Wind).

4.2.2 News Information Data

News information data, comprising financial news, market dynamics, policy regulations, etc., assist investment managers in keeping abreast of the latest market developments, thus enabling swift judgments. The sources for this data include news websites, financial media (e.g., financial magazines, economic observer newspapers), and news agencies.

4.2.3 Publicly Media Data

Public opinion data captures user views and sentiments on social media, online forums, and blogs. In investment research, public opinion data can illuminate market perspectives on specific assets or events, thereby enabling analysis of market expectations and sentiment. Sources for this data include social media platforms (e.g., Facebook, Twitter), online forums, and blogs (e.g., personal investment blogs, institutional research blogs).

4.2.4 Industry Data

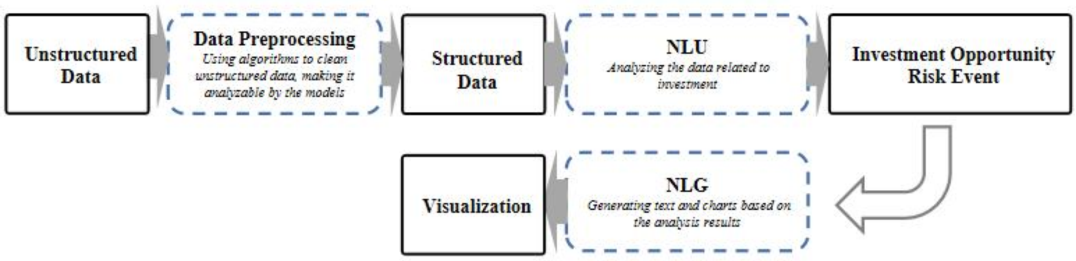
Industry data primarily encompasses information about market size, competitive landscape, and developmental trends for various industries. In investment research, industry data can facilitate investor understanding of the developmental status and opportunities within specific industries, thus promoting more targeted investments. Data sources include financial institution reports (e.g., investment banks, securities companies), consulting firm reports (e.g., McKinsey, Boston Consulting Group), and exchange data (e.g., stock exchanges, futures exchanges).

4.2.5 Corporate Data

Corporate data covers aspects such as a company's financial statements, operational status, and management team. Corporate data is instrumental in assessing company value and investment potential, aiding investment managers in acquiring a profound understanding of a target company's business performance and profitability. Data sources include publicly available financial statements (e.g., balance sheets, income statements, cash flow statements), company websites, securities exchange websites, and other financial information service platforms. Third-party databases (e.g., Bloomberg, Reuters) also serve as significant channels for obtaining corporate data, providing investment managers with a wealth of company information and research reports.

4.3 The Algorithm Platform

The algorithm platform is the core of an investment research system, which processes massive data from databases, analyzes unstructured data using NLP technologies, and combines different algorithms to achieve the functionalities of the core modules, ultimately realizing investment research. The algorithm module consists of a foundational algorithm module, data preprocessing module, and model processing module which can be further divided into NLU and NLG modules. Natural Language Understanding (NLU) focuses on text analysis and comprehension, while Natural Language Generation (NLG) focuses on generating machine-generated text to form readable reports.



Graph5: An example of the algorithm platform process pipeline

4.3.1 Foundational Algorithm

Natural Language Processing (NLP) has developed a wealth of algorithms and frameworks, which can be categorized into four groups: traditional methods, machine learning methods, deep learning methods, and pre-trained language models. These algorithms provide a robust infrastructure for processing various natural language tasks. In practice, different models should be employed flexibly according to the characteristics of the task.

4.3.1.1 Traditional Methods

a. Rule-based methods: These methods use predefined grammar rules, vocabulary, and syntactic structures to parse and generate text. Examples include context-free grammar (CFG) and feature structure grammar (FSG).

b. Statistical methods: These methods utilize statistical features from large volumes of text data for language modeling and other NLP tasks. Examples include n-gram models, hidden Markov models (HMM), and probabilistic context-free grammars (PCFG).

4.3.1.2 Machine Learning Methods

a. Supervised learning: These methods use labeled training data to train models, such as Naive Bayes classifiers, support vector machines (SVM), and decision trees.

b. Unsupervised learning: These methods learn the latent structure of text without labeled data, such as K-means clustering, self-organizing maps (SOM), and topic models (e.g., LDA).

4.3.1.3 Deep Learning Methods

a. Recurrent Neural Networks (RNN): RNNs are sequence models suitable for processing data with temporal or sequential relationships. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are two improved versions of RNN that can effectively address long-distance dependencies.

b. Convolutional Neural Networks (CNN): Although initially designed for image recognition tasks, CNNs can also be applied to text data processing, particularly in tasks like text classification and semantic relation extraction.

c. Transformer framework: Transformers are models based on self-attention mechanisms that can process text sequences in parallel, effectively solving long-distance dependency issues. BERT, GPT, and T5 are models based on the Transformer framework.

4.3.1.4 Pre-trained Language Models[xiii]

a. BERT (Bidirectional Encoder Representations from Transformers): BERT is a bidirectional pre-trained model based on Transformers. It can examine the context from both directions which helps in better understanding the semantics of each word. It is applicable to various NLP tasks such as text classification, named entity recognition, and question answering.

b. GPT (Generative Pre-trained Transformer): GPT is an autoregressive pre-trained model based on Transformers that focuses on generative tasks such as text generation and summarization. The GPT4.0[xiv] Version performs exceptionally well in various natural language tasks, accurately understanding different types of text and demonstrating strong text generation capabilities.

c. T5 (Text-to-Text Transfer Transformer): T5 is a pre-trained model based on Transformers that transform all NLP tasks into text-to-text formats. T5 achieves impressive results in many NLP tasks such as text classification, translation, and question answering.

d. RoBERTa (A Robustly Optimized BERT Pretraining Approach): RoBERTa is a variant of BERT. It was developed with the aim to improve upon BERT by modifying key hyperparameters, extending the training time, and training on a much larger amount of data. RoBERTa removes the next sentence prediction task during training and uses dynamic masking rather than static masking.The modifications made in RoBERTa have allowed it to achieve even better performance than BERT on several NLP tasks, including sentence classification, question answering, and others, leading to its widespread use in the field of NLP.

e. ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately): ELECTRA is a Transformer-based pre-trained model that employs a novel pre-training task called replacement token classification. ELECTRA performs exceptionally well in numerous NLP tasks while maintaining lower computational requirements.

f. ALBERT (A Lite BERT): ALBERT is a lightweight BERT variant optimized through parameter sharing and factorization techniques, reducing model size and computational requirements. ALBERT maintains excellent performance across various NLP tasks.

4.3.2 Data Processing Module

In the investment process of asset managers, it is essential to analyze vast amounts of unstructured data to extract key information for investment decision-making. Data preprocessing is the step that transforms unstructured data into structured data amenable to analysis, forming the bedrock for text understanding and generation. It plays a pivotal role in the entire natural language processing pipeline.

4.3.2.1 Stemming & Lemmatization

Stemming and lemmatization are processes that reduce words to their basic forms. Stemming removes affixes to obtain the stem, while lemmatization identifies the base form by analyzing the word's morphological changes and parts of speech. These methods help reduce data complexity and eliminate errors caused by word form variations.

4.3.2.2 Sentence Segmentation

Sentence segmentation is the practice of fragmenting a text into standalone sentences. Many NLP tasks necessitate sentence-level analysis. Sentence segmentation aids in extracting sentence structure information and provides more lucid input for subsequent tasks.

4.3.2.3 Stop Word Removal

Stop words are frequently occurring words in a language that contribute little to text analysis tasks, such as "a," "an," and "the." Removing stop words can reduce data noise and improve the accuracy and speed of subsequent tasks.

4.3.2.4 Tokenization

Tokenization is the process of breaking text into independent elements, such as words and punctuation marks. Tokenization is a fundamental step in NLP tasks, providing basic units for text analysis and understanding.

4.3.2.5 Feature Extraction

Feature extraction is the process of converting text data into numerical forms that computers can understand. Common feature extraction methods in NLP include the bag-of-words model (BoW), TF-IDF, word embeddings, and contextualized word representations. These methods capture key information in text and transform it into feature vectors suitable for machine learning and deep learning models.

4.3.2.6 Chunking

Chunking is the process of identifying phrase structures (e.g., noun phrases, verb phrases) in text. By chunking, we can capture local structural information in the text, providing useful information for subsequent semantic analysis and relation extraction tasks.

4.3.2.7 Named Entity Recognition (NER)

Named entity recognition is the process of identifying entities (e.g., names of people, places, organizations) in text and classifying them into predefined categories. NER is essential for information extraction, question-answering systems, and knowledge graph construction tasks.

4.3.3 Natural Language Understanding (NLU)

Natural Language Understanding (NLU) is the process by which computers acquire meaning and knowledge from natural language text through comprehension and parsing. The primary tasks of NLU include text classification and text pair matching.

4.3.3.1 Text Classification

a. Sentiment Analysis: Sentiment analysis involves the automatic identification and extraction of subjective information from text by computers to determine the sentiment expressed in a given text (e.g., positive, negative, or neutral). Sentiment analysis is used to analyze opinions, emotions, and attitudes in text data from social media, reviews, and forums, helping businesses, organizations, or individuals understand user perspectives on products, services, or events. Applications of sentiment analysis include market research, brand monitoring, and political opinion analysis.

b.Temporal Processing: Temporal processing focuses on analyzing and processing time-related information in text. This includes identifying and parsing time expressions (e.g., dates, times, and time periods), determining the sequence and temporal relations of events, and tracking timelines of entities and events mentioned in the text. Temporal processing has significant application value in NLP since time information is one of the key elements for understanding text meaning and achieving in-depth semantic understanding.

c. Dependency Parsing: Dependency parsing is a subtask in natural language processing (NLP) that identifies syntactic relationships between words in a sentence and constructs the sentence's dependency structure. The dependency structure is a directed graph where nodes represent words in the sentence and edges represent dependencies between words. Dependency relationships can express various grammatical relationships, such as subject-verb, modifier-head, and verb-object relationships.

4.3.3.2 Text Pair Matching

Text pair matching is the task of evaluating the relationship between two text segments. The following are some typical text pair matching tasks:

a. Semantic Textual Similarity: Semantic textual similarity has important value in multiple NLP tasks. For example, in information retrieval, relevant results can be returned by evaluating the semantic similarity between user queries and documents; in text summarization, similar content can be identified and merged; in machine translation evaluation, the semantic similarity between the source text and translation can be compared to assess translation quality; in question-answering systems, it can help match questions and answers, especially when question phrasing differs.

b. Natural Language Inference: Natural language inference is an important task for measuring a model's text understanding capabilities, as it requires the model to make inferences about another statement based on a given premise. This has significant implications for many NLP tasks, such as text summarization, question-answering systems, and intelligent dialogue. Through natural language inference, we can better understand whether models can capture logical relationships in text, thereby improving the accuracy of semantic understanding.

c. Relation Prediction: Relation prediction is crucial for identifying and extracting structured information from text. For example, in knowledge graph construction, unstructured text data can be transformed into structured knowledge representations by identifying relationships between entities (e.g., causal relationships, spatial relationships, etc.). Relation prediction can also be applied to named entity recognition, event extraction, sentiment analysis, and other tasks, helping models understand complex semantic information in texts.

4.3.4 Natural Language Generation (NLG)

Natural Language Generation (NLG) is the process of converting knowledge represented internally by a computer into natural language text. The main task of NLG is text generation, which includes the four steps: content determination, text planning, Sentence Generation, Generation and Revision.

4.3.4.1 Content Determination

This stage involves deciding which information to include in the generated text. Algorithmic modules can include heuristic search, rule engines, or knowledge graph queries to extract and select relevant information from input data.

4.3.4.2 Text Planning

At this stage, the system organizes the selected information into a logical, coherent structure. Algorithmic modules can include greedy algorithms, graph search algorithms, or constraint satisfaction problem solvers to determine the order, hierarchy, and relationships of the information.

4.3.4.3 Sentence Generation

This stage converts the structure obtained in the text planning stage into natural language sentences:

a. Lexical Selection: Rule-based methods (e.g., dictionary lookups), template-based methods (e.g., predefined phrase libraries), or statistical methods (e.g., word frequency statistics, TF-IDF, etc.).

b. Syntactic Generation: Rule-based methods (e.g., context-free grammars, feature structure grammars, etc.), template-based methods (e.g., predefined sentence templates), or statistical methods (e.g., n-gram models, PCFG, etc.).

c. Rhetorical Generation: Rule-based methods (e.g., rhetorical rule libraries), template-based methods (e.g., predefined connector libraries), or statistical methods (e.g., rhetorical relation classifiers, etc.).

4.3.4.4 Generation and Revision

This step involves combining generated sentences into a complete text and revising it to make sure the generated sentences are fluent.

a. Spelling, grammar, and semantic error checking: Rule-based methods (e.g., spell checkers, grammar checkers, etc.) or statistical methods (e.g., n-gram-based error detection models, etc.).

b. Text Optimization: Rule-based methods (e.g., text style and tone rule libraries), template-based methods (e.g., predefined substitution phrase libraries), or statistical methods (e.g., machine-learning-based text scoring models, etc.).

4.4 The Application Platform

The platform's core functions include intelligent research reports, sentiment monitoring, investment opportunity discovery, and risk event monitoring. These functions utilize the results of unstructured data processing from the algorithmic module, helping asset managers enhance efficiency in investment research.

4.4.1 Intelligent Research Reports

The intelligent research report module generates research reports on industries, markets, and companies using natural language processing (NLP) technology. These reports provide asset managers with timely, accurate information, allowing them to gain in-depth insights into market dynamics and competitive landscapes. Compared to traditional manually written reports, intelligent research reports offer faster generation and more frequent updates, significantly improving research efficiency.

4.4.2 Sentiment Monitoring

The sentiment monitoring module captures and analyzes real-time sentiment information from news, social media, forums, and other online channels. This information is valuable for asset managers as sentiment changes often correlate closely with market trends and investment opportunities. Continuous monitoring of sentiment enables asset managers to detect market signals faster, anticipating potential risks and opportunities.

4.4.3 Investment Opportunity Discovery

The investment opportunity discovery module identifies influential features for predicting investment opportunities from financial information data using feature selection techniques. Based on the extracted features, it constructs investment opportunity prediction models to discover investment opportunities in real-time or historical financial data. These investment opportunities provide strong decision support for asset managers, helping them optimize investment portfolios and increase returns.

4.4.4 Risk Event Monitoring

The risk event monitoring module identifies real-time risk factors in financial markets, such as abnormal fluctuations, unexpected events, and policy changes. Timely detection and response to risk events are vital for asset managers, as these events can significantly affect investment portfolios. Through risk event monitoring, asset managers can more flexibly adjust investment strategies, thus mitigating potential losses.

4.5 The Management Platform

The management system modules serve as support mechanisms for the operation, updating, and iteration of the entire system, in addition to managing the system's users. They are integral to ensuring the seamless operation of the system's business processes and the security of its data, facilitating efficient system utilization. Through the design and implementation of knowledge management modules and access control modules, an investment research system can provide continuously optimized investment advice while ensuring data security.

4.5.1 Knowledge Management

The knowledge management module primarily maintains and optimizes the platform's data and algorithm resources. It includes data quality monitoring, model updates, and maintenance. This module periodically verifies the accuracy, completeness, and consistency of the platform's data, guaranteeing it aligns with the prerequisites for investment research. Furthermore, it supervises the performance of the models, monitoring metrics like predictive accuracy, and recall, among others. For underperforming models, it provides timely alerts, updates, and optimizations. This module also amalgamates data, models, and analysis results within the platform to construct a knowledge base, thereby enabling users to swiftly locate relevant information and enhance investment research efficiency.

4.5.2 Access Control Management

The access control management module ensures data security and confidentiality. In asset managers, data often possess sensitivity and confidentiality, making access and operation control crucial. The access control module facilitates user management by categorizing users and assigning different roles, such as administrators, analysts, and investment managers. Based on user roles, it assigns corresponding access and operation permissions. The module also enforces data access control by setting data access permissions based on user roles and data sensitivity. For example, only certain user roles can access data involving internal company secrets. Moreover, the access control module conducts operation audits, logging user activities on the platform, such as data queries, model usage, and report generation. By auditing user activities, it traces potential data leakage risks and promptly executes suitable security measures.

4.6 The Visualization Front-end

The visualization front-end is essential for an investment research system, serving as the showcase for the system's outputs. All model investment results are presented through this module.

4.6.1 API Integration

The front-end visualization display module provides API interfaces, allowing system results to be exported to external platforms for integration with other applications and services. The availability of API interfaces brings more application scenarios and development potential to the system.

4.6.2 Data Visualization

The data visualization module uses a variety of visualization elements, including charts, maps, and dashboards, to present model investment research results in a lucid and digestible fashion. This enables users to quickly comprehend key data and trends. Moreover, it offers an intuitive interface and streamlined interactive operations, enhancing efficiency. High-quality visualizations foster increased user trust and satisfaction with the platform.

5. Development Trends and Challenges

5.1 Development Trends of Investment Research Systems

In the short term, the application of intelligent investment systems will not compete with investment researchers but rather complement each other. Investment researchers rely on their knowledge and logic to interpret the issues identified by investment research systems, while the latter can quickly and accurately perform more comprehensive and extensive data processing, preventing researchers from getting bogged down in repetitive tasks and helping them stay ahead of market trends.

From a technical perspective, the application of large-scale automatic data extraction and association technologies is not limited to investment research areas such as negative public sentiment event extraction, association, and early warning but also has significant potential in risk management, compliance, legal affairs, and regulatory domains, which are important research topics.

5.2 Challenges Faced by Investment Research Systems

Although the development of investment research systems is thriving, it still faces potential challenges and risks. These risks stem from the limitations of NLP technology development, data accuracy and security, homogeneous competition, uncontrollable external risks, and legal compliance risks.

5.2.1 NLP Technology Bottlenecks Risk

Overall, the evolution of NLP has transpired through several significant stages of historical development, with considerable progress made through interdisciplinary collaboration. However, current deep learning methodologies are not without their limitations. For instance, they often achieve mere "curve fitting" through extensive neural networks and dedicated chips, rather than genuinely "understanding" the data. Therefore, equipping NLP technology with human-like capabilities in areas such as small sample learning, transfer learning, associative reasoning, and real-time error correction remains an ongoing challenge.

5.2.2 Data Risk

The effectiveness of investment research systems largely depends on the quality of the data used. Inaccurate, incomplete, or outdated data may lead to erroneous investment advice. Ensuring data quality is an ongoing challenge, as it requires regular data validation, cleaning, and updating.

Moreover, distinguishing genuine data in financial sentiment detection is a difficult problem. In our contemporary, highly developed online media landscape, the market is inundated with counterfeit data. For instance, malicious individuals may disseminate false information via social media platforms. If an investment research system fails to identify such misinformation and incorporates it into its algorithms, the reliability of the system could be compromised.

5.2.3 Homogenization Competition Risk

With the development of AI technology, more and more asset managers are developing investment research systems, leading to a new round of competition. Failing to update data or develop new strategy models promptly may result in falling behind, rendering initial research investments fruitless and eventually losing market position.

5.2.4 External Risk

Investment research systems represent the future direction of the asset management industry and will greatly enhance investment decision-making efficiency. It is foreseeable that asset managers will become increasingly reliant on these systems. However, they may still encounter unforeseen external risks. For example, investment research systems often store massive amounts of data, including various corporate and personal information, making them potential targets for cyberattacks. Hacking incidents targeting hardware servers could cause significant problems for asset managers.

5.2.5 Legal and Compliance Risk

Investment research systems require access to vast amounts of data, and the legality and permission to use this data is a sensitive issue. As the AI industry develops, government oversight of investment research systems will become increasingly stringent. Asset managers must exercise caution in establishing databases to avoid legal and compliance risks.

6. Suggestions for Future Strategies

6.1 Introducing New NLP Algorithms

The algorithms of NLP have been rapidly developing in recent years, with new models constantly emerging and pushing the capabilities of natural language processing to new heights. Especially since 2023, the Transformer-based ChatGPT model developed by OpenAI has gained significant attention. To maintain technical superiority, asset managers need to keep a keen eye on developments in the AI industry and regularly update the underlying algorithm modules of their investment research systems.

6.2 Enhancing Industry Data Mining

Investment research systems rely on vast amounts of data for training and analysis. To improve the system's predictive accuracy and investment returns, efforts must be made to enhance the mining of industry data. Particularly for emerging industries and technology sectors, it is crucial to promptly collect and analyze relevant data to better identify investment opportunities and risks.

6.3 Strengthening Compliance Construction

Investment research systems must adhere to relevant laws, regulations, and ethical norms. Thus, a robust compliance management system should be in place, complete with enhanced internal management and oversight. Additionally, conducting regular risk assessments and vulnerability tests can assure data security and regulatory compliance.

6.4 Boosting System Security

Given that investment research systems are crucial tools for protecting investment managers' assets and information, system security needs to be prioritized. This can be achieved by strengthening network and information security measures, including data encryption, access control, and firewalls. These measures will safeguard the system from hacking attempts and potential information leaks.

6.5 Establishing Competent Manual Review Mechanisms

While investment research systems offer high levels of intelligence and automation, they are still bound by existing limitations in AI technology and cannot fully replace human judgment. For the foreseeable future, the system will primarily serve as an investment support tool assisting asset managers in sifting through extensive financial market data to identify valuable insights and enhance operational efficiency. Hence, manual review mechanisms are indispensable. A complementary manual review mechanism should be set up to audit and verify the system's analytical results. Furthermore, regular system improvements and optimizations should be implemented to ensure increasingly accurate and reliable results.

7. Conclusion and Outlook

In recent years, natural language processing (NLP) has achieved significant breakthroughs and has been progressively employed in investment research within the field of financial asset management. Nonetheless, challenges remain in the NLP realm, including limited understanding and reasoning abilities, a deficit of specific industry background knowledge, and a reliance on extensive labeled data. Multiple research initiatives are underway to address these challenges. It is anticipated that future NLP models will tackle more specialized and domain-specific language tasks and data, thereby providing more accurate and practical language applications and services.

In the asset management field, NLP has the potential to become a game-changer by offering innovative methods for analyzing and interpreting financial information. From processing massive unstructured data sets and automatically extracting key points to form investment research reports, to providing a more nuanced understanding of public sentiment, investment research systems built upon NLP have demonstrated their indispensability in the asset management industry. Nevertheless, asset managers should also be aware of the associated risks, such as technological development bottlenecks, data risks, homogenized competition risks, external risks, and legal compliance risks. Asset managers need to prepare for these risks in advance to better leverage NLP technology.

Overall, the application of NLP technology in investment research is highly promising. Asset managers will continually improve their investment efficiency and quality with the help of NLP technology. In the future, NLP-based investment research systems will play an increasingly important role in asset management, ushering the asset management industry into a new era.

---

[i] In the conference proposal, John McCarthy used the term "Artificial Intelligence" for the first time to distinguish it from other areas of research that were referred to as "automatic computing" at the time.

[ii] Deep Blue was a chess-playing computer developed by IBM

[iii] An n-gram is a contiguous sequence of n items from a given sample of text or speech.

[iv] Bayesian classification is a statistical technique used in machine learning and pattern recognition.

[v] Hidden Markov Models are statistical models that can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable.

[vi] The Word2Vec model was developed by a team of researchers led by Tomas Mikolov at Google.

[vii] Recurrent Neural Networks (RNNs) are a type of artificial neural network designed to recognize patterns in sequences of data, such as text, genomes, handwriting, or spoken word.

[viii] LSTM networks are a type of RNN introduced by Hochreiter & Schmidhuber in 1997. They are designed to avoid the long-term dependency problem in RNNs.

[ix] ELMo (Embeddings from Language Models) is a deep learning-based model for generating high-quality word embeddings (a way to represent words as vectors). It was developed by researchers at the Allen Institute for AI.

[x] BERT (Bidirectional Encoder Representations from Transformers) is another deep learning-based model for natural language processing, developed by Google.

[xi] GPT (Generative Pretrained Transformer) developed by OpenAI, is a model that uses transformer architecture to generate long sequences of text.

[xii] Alpha is the excess return of an investment relative to the return of a benchmark index.

[xiii] There are a plethora of pre-training models available. However, we will focus on introducing the most representative pre-training models here.

[xiv] GPT-4 is a new language model created by OpenAI that can generate text that is similar to human speech.

References

Apostolov, Alexander. "AI-BASED TOOLS IN EUROPEAN SECURITIES MARKETS: AN OVERVIEW OF ADOPTION, USE CASES, FINDINGS, AND IMPLICATIONS." *Endless light in science* апрель (2023): 201-210.

Benko, Attila, and Cecília Sik Lányi. "History of artificial intelligence." *Encyclopedia of Information Science and Technology, Second Edition* (2009): 1759-1762.

Braun, Daniel, et al. "Evaluating natural language understanding services for conversational question answering systems." *Proceedings of the 18th annual SIGdial meeting on discourse and dialogue*. 2017.

Buchanan, Bruce G. "A (very) brief history of artificial intelligence." *Ai Magazine* 26.4 (2005): 53-53.

Canonico, Massimo, and Luigi De Russis. "A comparison and critique of natural language understanding tools." *Cloud Computing* 2018 (2018): 120.

Costantino, Marco, Russell J. Collingham, and Richard G. Morgan. "Qualitative Information in Finance: Natural Language Processing and Information Extraction." *Neuro Ve $ t Journal* 4 (1996).

Dang, Nhan Cach, María N. Moreno-García, and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." *Electronics* 9.3 (2020): 483.

Deryck, Marjolein, et al. "Combining logic and natural language processing to support investment management." *Proceedings of the international conference on principles of knowledge representation and reasoning*. Vol. 18. No. 1. 2021.

Dima, Alden, et al. "Adapting natural language processing for technical text." *Applied AI Letters* 2.3 (2021): e33.

Fisher, Ingrid E., Margaret R. Garnsey, and Mark E. Hughes. "Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research." *Intelligent Systems in Accounting, Finance and Management* 23.3 (2016): 157-214.

Gopali, Nilesh, and Hasik Shetty. "AI in Indian Investment and Asset Management: Global Perspective." *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries* (2020): 132-133.

Hasan, Md Rakibul, Maisha Maliha, and M. Arifuzzaman. "Sentiment analysis with NLP on Twitter data." *2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)*. IEEE, 2019.

Huseynova, Nigar F. "Investment Decision Making by Using Natural Language Processing." *15th International Conference on Applications of Fuzzy Systems, Soft Computing and Artificial Intelligence Tools–ICAFS-2022*. Springer Nature Switzerland, 2023.

Johri, Prashant, et al. "Natural language processing: History, evolution, application, and future work." *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020*. Springer Singapore, 2021.

Jones, Karen Sparck. "Natural language processing: a historical review." *Current issues in computational linguistics: in honour of Don Walker* (1994): 3-16.

Kempen, G. A., ed. "Natural language generation: New results in artificial intelligence, psychology and linguistics." (2012).

Lezmi, Edmond, and Jiali Xu. "Time Series Forecasting with Transformer Models and Application to Asset Management." *Available at SSRN 4375798* (2023).

Lo, Andrew W., and Manish Singh. "From ELIZA to ChatGPT: The Evolution of NLP and Financial Applications." (2023).

Mazidi, Karen, and Paul Tarau. "Automatic question generation: From NLU to NLG." *Intelligent Tutoring Systems: 13th International Conference, ITS 2016, Zagreb, Croatia, June 7-10, 2016. Proceedings 13*. Springer International Publishing, 2016.

Miziołek, Tomasz. "Employing artificial intelligence in investment management." *The Digitalization of Financial Markets* (2021): 161-174.

Paris, Cécile L., William R. Swartout, and William C. Mann, eds. "Natural language generation in artificial intelligence and computational linguistics." (2013).

Patel, Ramkrishna, et al. "Lstm and nlp based forecasting model for stock market analysis." *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*. IEEE, 2021.

Payette, Mathieu, et al. "Improving Maintenance Data Quality: Application of Natural Language Processing to Asset Management." *16th WCEAM Proceedings* (2023): 582-589.

Rusu, Ștefan. "ARTIFICIAL INTELLIGENCE IN FINANCE AND ASSET MANAGEMENT." *Emerging Markets Economics and Business. Contributions of Young Researchers*: 340.

Sadlo, Sven. "Big Data and Machine Learning in Asset Management." (2021).

Santos, Pedro Costa. "Machine learning applied to asset management." (2022): 1-94.

Schuddebeurs, J. D., et al. "Text data mining as a tool for the asset management decision support process." (2016): 26-6.

Sumathy, K. L., and M. Chidambaram. "Text mining: concepts, applications, tools and issues-an overview." *International Journal of Computer Applications* 80.4 (2013).

Tervo, Jesse, Helena Kortelainen, and Anu Purhonen. "Utilizing Evidence in Asset Management in the Era of Industry 4.0 and Artificial Intelligence." *Advances in Systems Engineering: Proceedings of the 28th International Conference on Systems Engineering, ICSEng 2021, December 14-16, Wrocław, Poland*. Springer International Publishing, 2021.

Xing, Frank Z., Erik Cambria, and Roy E. Welsch. "Natural language based financial forecasting: a survey." *Artificial Intelligence Review* 50.1 (2018): 49-73.