

StockGPT

Shreyas NM¹, Pradip Kumar Das², Sanath Prasanna Rajur³

^{1,3}Student, AI&ML, BNMIT

²Professor of Practice, AI&ML, BNMIT

Abstract

This project entails enhancing the accuracy of OpenAI Chat GPT – 3.5 Turbo by training it with a carefully selected dataset of Q&A related to the Indian stock market. By leveraging Flask, we develop a user-friendly interface that provides easy access to valuable insights for making informed decisions. Our goal is to equip users with a deep understanding of stock market dynamics, assisting both investors and enthusiasts in navigating the intricacies of financial markets. By blending advanced NLP methods with intuitive interface design, we strive to create a smooth interaction experience that boosts engagement and encourages informed financial choices.

Keywords: AI, Stocks, GPT

1. Introduction

In today's world, where financial literacy isn't universally available, many feel overwhelmed and unsure about investing in stocks. We're creating a platform to guide and support newcomers in the stock market, aiming to democratize investing and empower people from all backgrounds to participate confidently. Our project is driven by a commitment to promoting informed decision-making and responsible investing. We understand the importance of providing individuals with the tools and resources needed to navigate financial markets effectively.

By offering reliable information, educational content, and personalized guidance, our goal is to build confidence and encourage wise financial habits among users. We're motivated by the potential societal impact of this project. By empowering individuals to manage their finances, we not only enhance their personal well-being but also foster economic empowerment and resilience within communities. Through education and support, we aim to dismantle barriers to financial inclusion and promote economic equality and prosperity.

With the increasing popularity of stock market investment, there's a growing interest among newcomers. However, entering the market poses challenges, including understanding market dynamics, managing risks, and lack of financial literacy. Novice investors may struggle to assess and handle these risks effectively, exposing themselves to potential losses. A significant barrier for newcomers is the lack of financial knowledge and investment education.

Therefore, our project aims to address key issues faced by new entrants in the stock market, such as understanding market dynamics, managing risks, and accessing reliable information. We aim to assist newcomers in navigating investment complexities through an interactive platform. By covering a wide range of topics impacting the investment journey, our goal is to empower newcomers with the knowledge and resources for informed decision-making.

Our primary objective is to develop a user-friendly platform tailored to the needs of novice traders and newcomers to investing. This platform will provide a welcoming space for users to ask questions about trading, investing strategies, and market trends. Central to our goals is fostering open dialogue and inquiry, creating a supportive environment where users feel comfortable asking questions. We're also committed to providing users with accurate and comprehensive information, covering investment fundamentals, market analysis, risk management, and regulatory compliance.

2. System Design

The following section speaks about system design and the development process

2.1 Architectural Design

The architecture design outlines a system comprising a Large Language Model (LLM) for text generation, complemented by a News Block and an Analytics Page. The LLM serves as the core text generation engine, leveraging extensive training data to produce human-like text. The News Block provides users with live news updates, sourced from external APIs or sources, ensuring they stay informed about current events. Simultaneously, the Analytics Page offers insights into stock values and market trends through visual representations and data analysis tools, empowering users to make informed decisions regarding investments.

Application Interface: The application interface serves as the primary gateway for users to interact with the system, providing an intuitive and seamless experience. It features a user-friendly design with clear navigation elements, allowing users to effortlessly access the system's functionalities. The interface is divided into distinct sections, including text generation, news updates, and analytics, each accessible through intuitive menus or tabs.

Select Stock: This feature empowers users to focus on specific companies or industries of interest, allowing for targeted analysis and comparison of stock performance. Whether users are tracking their investments, conducting market research, or exploring potential opportunities, the ability to select stocks within the visualization interface enhances flexibility and efficiency.

Select Value: Integrating checkboxes to select specific data points such as open, close, high, and low within the visualization interface offers users enhanced control and customization over the displayed stock information. By incorporating checkboxes adjacent to each data point, users can easily toggle their preferences, choosing to view only the data points that are relevant to their analysis or decision-making process.

Visualization of Selected Stock: The visualization component of the system presents a dynamic and intuitive platform for users to analyze stock performance with ease and precision. Leveraging interactive line graphs, users can explore historical trends and patterns of selected stocks over customizable timeframes.

Conversational AI: Leveraging advanced natural language processing (NLP) capabilities, conversational AI systems can understand and respond to user queries regarding stock market trends, company performance, investment strategies, and more. By analyzing vast amounts of data from diverse sources such as financial news articles, market reports, and historical stock data, these systems offer users timely and relevant information to support their investment decisions.

Live News Article: Application offers a comprehensive feature to access news articles related to companies and finance, providing users with valuable insights into factors that may impact stock prices.

Through an integrated news feed, users can stay updated on the latest developments in the financial world, including company announcements, market trends, economic indicators, and geopolitical events.

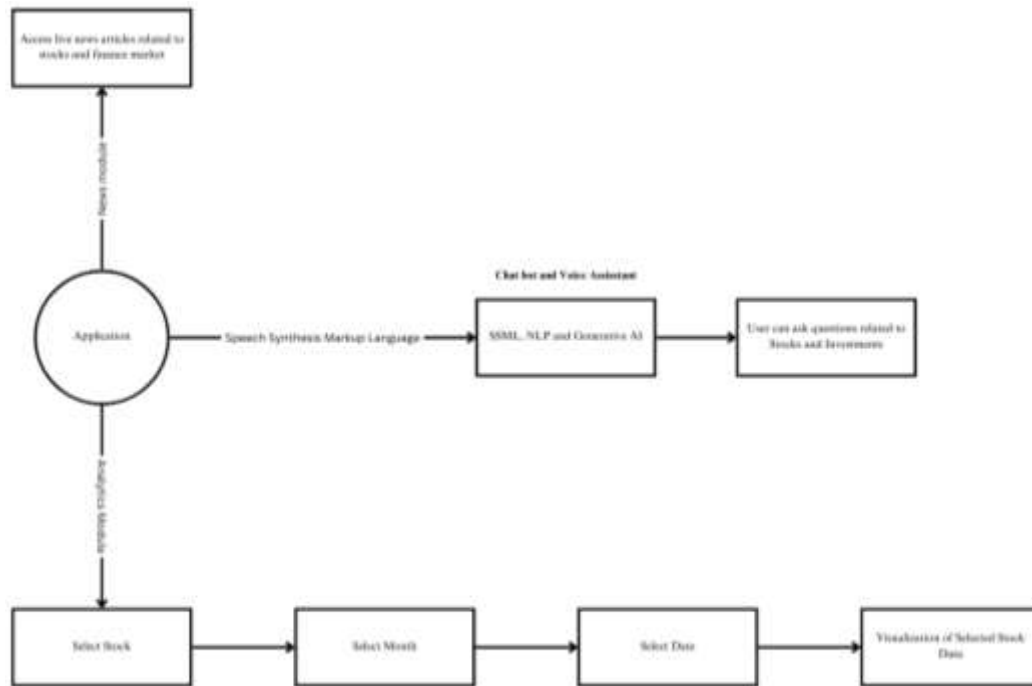


Figure 2.1 LLM Architecture

2.2 Data Flow diagram

A data flow diagram is a visual representation that illustrates the flow of data within a system. It consists of various components, including processes, data stores, data flows, and external entities, connected by arrows to depict the movement of data.

User Input: User input serves as the starting point of the data flow within the system. It represents any data or information provided by the user to the system, such as text prompts, commands, or queries. This input can originate from various sources, including text input fields, voice commands, or user interactions with graphical user interfaces (GUIs).

Text Processing by performing tokenization: After receiving user input, the system engages in text processing, which involves various operations to analyze and manipulate the input text. One crucial step in text processing is tokenization, where the input text is broken down into smaller units called tokens. These tokens can be words, phrases, or symbols, depending on the specific requirements of the system.

LLM module to retrieve text response: Once the input text has been processed and tokenized, it is passed on to the Large Language Model (LLM) module for further analysis and generation of a text response. The LLM module, which comprises a sophisticated AI model trained on vast amounts of textual data, leverages its language understanding capabilities to comprehend the input text and generate a coherent and contextually relevant response.

Response in the form of text: Once the LLM module has processed the input text and generated a response, the system delivers the output to the user in the form of text. This response typically comprises a coherent and contextually relevant message that addresses the user's query or input.

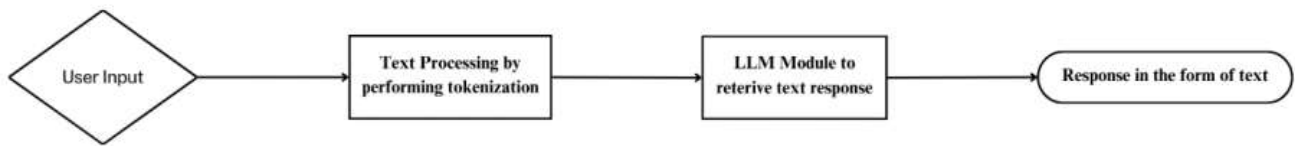


Figure 2.2 Data Flow diagram

2.3 Algorithm

Collect and preprocess a large corpus of text data relevant to the domain or task the LLM is intended for. Tokenize the raw text data into sequences of tokens using a tokenizer module, ensuring consistency with the tokenizer used during model pretraining.

Initialize a ChatGPT 3.5 Turbo model using the ChatGPT API. This will create an instance of the model that you can further train on your dataset.

Use the fine-tune endpoint of the ChatGPT API to fine-tune the model on your dataset. This involves providing the preprocessed training data to the API and specifying parameters such as the number of training epochs and the learning rate. • Monitor the training process and evaluate the model's performance on a validation dataset periodically to ensure it is learning effectively and not overfitting to the training data

Once the fine-tuning process is complete, evaluate the performance of the fine-tuned model on a separate test dataset. This will give you an indication of how well the model has learned to perform the task or understand the domain-specific language

Once you are satisfied with the performance of the fine-tuned model, deploy it for use in your application or service. This may involve integrating the model into your existing infrastructure or deploying it as a standalone service using the ChatGPT API.

Continuously monitor the performance of the deployed model in production to ensure it continues to meet your requirements.

Periodically retrain the model using new data to keep it up-to-date and improve its performance over time.

3. Implementation

Given below is the implementation steps of the research

3.1 Dataset

The dataset is acquired by using yfinance function which web scrapes the yahoo finance website to The dataset is obtained using the yfinance function, which scrapes data from the Yahoo Finance website. Each entry in the dataset includes fields for Date, Open, High, Low, and Close prices. The Date field indicates the timestamp of the stock data, while the Open, High, Low, and Close fields contain the corresponding price values for the opening, highest, lowest, and closing prices of the stock on a given day. This dataset is an important resource for analyzing historical trends and performance in the stock market, providing insights into price fluctuations, volatility, and potential investment prospects across various stocks.

3.2 Module Description

- **Analytics Module:** The Analytics Module within the system serves as a critical component for analyzing and interpreting data related to stock market trends and performance. Leveraging

advanced data visualization techniques, the module enables users to gain actionable insights into various aspects of the financial markets. Through interactive dashboards, charts, and graphs, users can visualize historical data.

- **LLM Module:** The LLM (Large Language Model) Module represents the core engine of the system, embodying sophisticated artificial intelligence capabilities for natural language processing and text generation. Built upon advanced deep learning architectures like transformers, the LLM Module is trained on vast corpora of textual data.
- **Text Preprocessing Layer:** The Text Preprocessing Module is a crucial component of the system, responsible for preparing textual user inputs for further analysis and processing. Its primary function is to perform tokenization, a fundamental text processing technique that breaks down the input text into smaller, analyzable components called tokens.
- **User Interface and GUI:** The UI features a minimalist and user-friendly design, comprising essential elements that facilitate seamless interaction with the system's functionalities. At the forefront is a text box, prominently displayed, where users can engage in natural language conversation with the AI. Adjacent to the text box are tabs or buttons providing access to additional features of the application. One such tab leads to a dedicated News Page, where users can stay informed with live news articles related to companies, finance, and other topics impacting the stock market. Another prominent feature of the UI is the Analytics Page, accessible through a separate tab or menu option. Here, users can delve into the visualization of historical data pertaining to stocks.
- **News Module:** The News Module within the system serves as a vital component for delivering realtime news updates and articles relevant to the financial market and stock trading. This module is designed to aggregate news content from various reputable sources, including financial news outlets, business journals, and industry publications. Through APIs or web scraping, the module continuously fetches and updates a curated feed of news articles, ensuring users have access to the latest information that may impact stock prices and market trends.
- **Server Module:** The Server Module serves as the intermediary connecting the UI to the Classification and LLM Modules. It oversees API calls, streamlining the flow of data between the user interface and the backend processing components. This module guarantees the system's performance is dependable and adaptable, equipped to manage concurrent requests efficiently. By facilitating seamless communication between the user-facing interface and the backend systems, the Server Module optimizes system performance and responsiveness.

4. Testing and Validation

Testing the chat model involves a multifaceted approach to ensure its effectiveness and reliability. One aspect of testing involves evaluating the model's responses to a variety of questions asked by users, assessing whether it provides accurate and relevant answers.

4.1 Testing Methods

This process helps verify the model's comprehension and knowledge across different topics. Additionally, testing includes examining how the model handles non-related topics, ensuring it appropriately responds or gracefully handles queries outside its scope. Furthermore, evaluating whether the model suggests buying any particular stock is essential, especially in financial contexts. The model should provide informed and unbiased responses, avoiding any explicit recommendations or biases towards specific stocks. By rigorously testing these aspects, we can gauge the model's performance,

identify areas for improvement, and ensure it delivers a high-quality conversational experience for users. Test the model's responses to various questions asked by users to ensure accuracy and relevance.

- Evaluate how the model handles non-related topics, ensuring it provides appropriate responses or gracefully handles queries outside its scope.
- Examine whether the model suggests buying any particular stock, ensuring it avoids explicit recommendations or biases towards specific stocks.
- Assess the overall performance of the chat model by examining its ability to provide informed, unbiased, and coherent responses across a diverse range of topics.
- Use testing results to identify areas for improvement and refine the model to deliver a high-quality conversational experience for users.

4.2 Test Cases

The above test cases show the working of the model with the prompt along with the result generated but the LLM. It shows answering accurately to the various questions and ignoring the questions that are beyond its scope. Also, below we see that the bot is not making any specific recommendation for any particular stock or any form of preferential treatment.

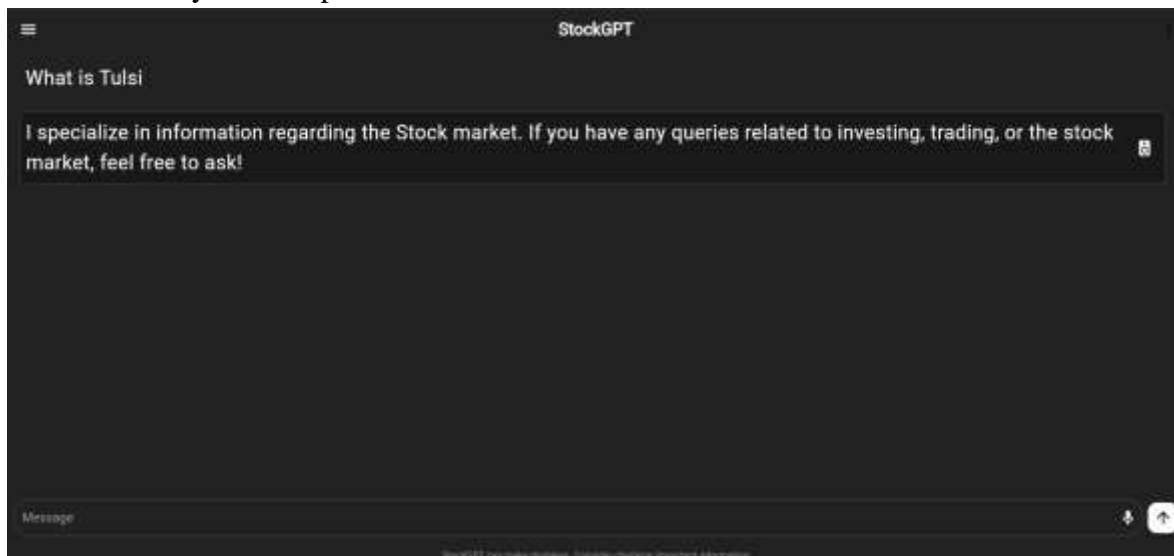


Figure 4.1 Test Case 1

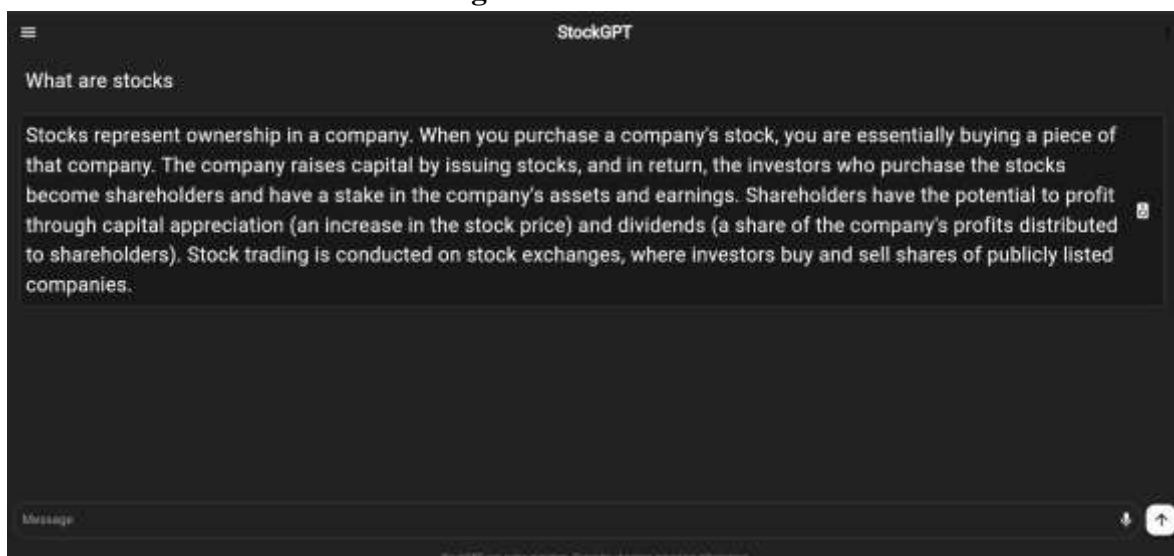


Figure 4.2 Test Case 2

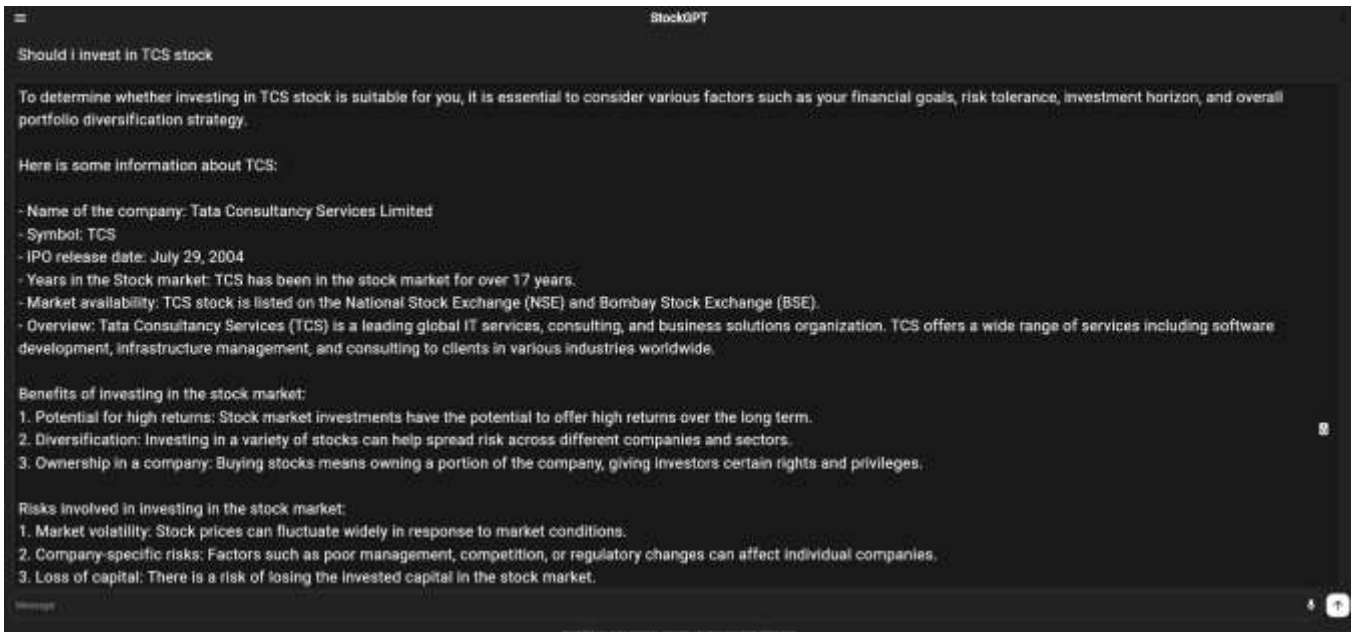


Figure 4.3 Test Case 3

5. Results and Discussion

BERTScore is a metric used to evaluate the quality of generated text, especially in the context of natural language processing tasks like machine translation, text summarization, and text generation. It measures the similarity between the reference (ground truth) text and the generated text based on contextual embeddings obtained from BERT (Bidirectional Encoder Representations from Transformers) models. Here's a breakdown of the test results:

- **Precision:** The precision scores represent how accurately the model identifies relevant items among the retrieved items. A perfect precision score of 1.0 indicates that all items retrieved by the model are relevant to the reference text. In this case, the precision score for the reference text is 1.0, indicating perfect precision, while the precision score for the generated text is approximately 0.884, suggesting that around 88.4% of the retrieved items are relevant to the reference text.
- **Recall:** The recall scores indicate the proportion of relevant items from the reference text that have been correctly retrieved by the model. A recall score of 1.0 means that all relevant items from the reference text have been retrieved. The recall score for the reference text is 1.0, indicating perfect recall, while the recall score for the generated text is approximately 0.927, suggesting that around 92.7% of the relevant items from the reference text have been retrieved.
- **F1 Score:** The F1 score is the harmonic mean of precision and recall, providing a single metric to evaluate the model's performance, balancing both precision and recall. A higher F1 score indicates better overall performance. The F1 score for the reference text is 1.0, indicating perfect balance between precision and recall, while the F1 score for the generated text is approximately 0.905, suggesting that the model achieves a good balance between precision and recall, but there is still some room for improvement.

```
{'precision': [1.0000001192092896, 0.8840766549110413], 'recall': [1.0000001192092896, 0.9271026849746704]}
```

Figure 5.1 Precision and Recall values

```
"f1": [1.0000001192092896, 0.9050785899162292], "hashcode": "roberta-large_L17_no-idf_version=0.3.12(hug_trans=4.38.2)"]
```

Figure 5.2 F1 Score

6. Conclusion

The project tackles the widespread issue of financial illiteracy by creating a platform designed for new investors venturing into the stock market. Recognizing the significant societal impact of enabling individuals to make well-informed financial choices, this initiative aims to fill the gap in knowledge and support for newcomers. By acknowledging the complexities involved in stock market investment, such as market dynamics, risk management, and psychological biases, the project aims to establish a welcoming space where users can seek guidance and access curated educational materials. Through fostering open dialogue and collaboration, the platform encourages a culture of curiosity and empowerment, empowering users to navigate investment intricacies with confidence.

The proposed model architecture underscores the importance of tailoring language models to specific tasks, like text generation or sentiment analysis. This involves preprocessing data, initializing the model, defining fine-tuning objectives, and iteratively training and assessing the model's performance. Additionally, the evaluation process using BERTScore offers valuable insights into the quality of generated text, pinpointing areas for enhancement in precision, recall, and overall performance. By integrating these elements, the project aims to democratize financial education access and advocate for economic equality, ultimately empowering individuals to construct secure financial futures for themselves and their communities.

7. Future Work

For future enhancements, several key areas offer opportunities for further development and refinement. Firstly, integrating live data feeds of stocks would provide users with real-time market information, enhancing the platform's relevance and utility. This feature would enable users to make informed decisions based on up-to-date market trends and fluctuations. Secondly, improving the speed of text generation is crucial for enhancing user experience and responsiveness. Optimizing the underlying algorithms and leveraging efficient hardware resources can help accelerate the generation process without compromising quality. Additionally, optimizing the codebase through refactoring and performance tuning can streamline execution and reduce computational overhead, ultimately leading to a more efficient and scalable platform. Lastly, enabling data analytics capabilities would empower users to perform deeper analysis of market trends and patterns, facilitating more informed investment decisions. By incorporating these enhancements, the platform can evolve into a comprehensive tool that not only educates users but also empowers them to navigate the complexities of the stock market with confidence and efficiency.

References

1. F. Achakir, N. Mohtaram and A. Escartin, "An automated AI-based Solution for Out-Of-Stock Detection in Retail Environments," IEEE Xplore, 2023, pp. 1-6.
2. A. L. Eisfeldt, G. Schubert, M. B. Zhang, and B. Taska, "Generative AI and Firm Values," SSRN May 2, 2023. "<https://ssrn.com/abstract=4436627> or <http://dx.doi.org/10.2139/ssrn.4436627>"

3. Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond, "Generative AI at Work", NBERG, April 2023. <https://www.nber.org/papers/w31161>
4. Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. Yu, and L. Sun, "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT," 2023.
5. S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI," SSRN, May 9, 2023. <http://dx.doi.org/10.2139/ssrn.4443189>
6. Zhihan Lv, "Generative artificial intelligence in the metaverse era", Science Direct, 2023. <https://www.sciencedirect.com/science/article/pii/S2667241323000198?via%3Dihub>
7. Kenneth N. Reid, Ph.D. "A Brief Introduction to GenAI". 2023 "https://docs.google.com/presentation/d/1X3VpadTOsUe2neFts24pURY3nNQ2k64k4d3MEqHIEgk/edit#slide=id.g25b6aed46c6_0_492 "
8. G. Franceschelli and M. Musolesi, "Reinforcement Learning for Generative AI: State of the Art, Opportunities and Open Research Challenges," 2023.
9. F. Han and Y. Ma, "Co-Creating the Future with Generative AI: Exploring AI's Capacities and Potential Incorporation in Statistical Education," SSRN Electronic Journal, 2023, doi: 10.2139/ssrn.4637265.
10. A. Cooper et al., "Report of the 1st Workshop on Generative AI and Law," SSRN Electronic Journal, 2023, doi: 10.2139/ssrn.4634513.
11. W. Huck, "Legal Prompts bei ChatGPT-4: Tipps und Techniken für Lehre, Forschung und Praxis," SSRN, November 12, 2023. <http://dx.doi.org/10.2139/ssrn.4630455>
12. M. Alshater, "Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT," SSRN, December 26, 2022. <http://dx.doi.org/10.2139/ssrn.4312358>
13. Ö. Aydın and E. Karaarslan, "OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare," 2022, doi: 10.2139/ssrn.4308687.
14. Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, Jenia Jitsev, "LAION-5B: An open large-scale dataset for training next generation image-text models", 16 Oct 2022. <https://arxiv.org/abs/2210.08402>
15. T. van der Zant, M. Kouw, and L. Schomaker, "Generative Artificial Intelligence," in Philosophy and Theory of Artificial Intelligence, V. Müller (Ed.), Studies in Applied Philosophy, Epistemology and Rational Ethics, vol. 5. Springer, Berlin, Heidelberg, 2013, doi: 10.1007/978-3-642-31674-6_8.
16. S. Mokhtari, K. Yen, and J. Liu, "Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning," 2021. doi: 10.1007/978-3-642-31674-6_8.
17. M. -S. Cheong, M. -C. Wu and S. -H. Huang, "Interpretable Stock Anomaly Detection Based on SpatioTemporal Relation Networks With Genetic Algorithm," in IEEE Access, 2021, pp. 68302-68319.
18. Reddy, Y.V, Parab, N, "Literature on stock returns: A content analysis". 2016. <http://irgu.unigoa.ac.in/drs/handle/unigoa/4601>
19. R. Morck et al., "The Stock Market and Investment: Is the Market a Sideshow?" Brookings Papers on Economic Activity, 1990, pp. 157–215. <https://doi.org/10.2307/2534506>
20. D. J. Milner, F. Teel, and G. Winans, "A Genetic Method of Stock Identification in Mixed Populations of Pacific Salmon, *Oncorhynchus* spp.," Marine Fisheries Review, pp. 1-8, 1985.