The University of Hong Kong
Department of Computer Science

COMP4801 Final Year Project
Final Report

Automated Social Media Research and Sentiment Analysis using LLMs

Group: FYP23047

Students:
Hung Jasper (3035787671)
Ko Berrick Eldridge (3035784710)

Supervisor:
Dr. Kong Lingpeng

26 April 2024

I. Abstract

Social media provides organizations with valuable insights into user feedback and public sentiment. However, manual social media research can be impractical, and existing sentiment analysis approaches have demonstrated vulnerabilities. This project aims to overcome these challenges by developing an automated system that performs social media research and conducts aspect-based sentiment analysis using state-of-the-art LLMs. The system comprises five main procedures: data collection, aspect identification and classification, aspect-based sentiment analysis, aspect-based summarization, and result visualization. Successfully implemented using Reddit and Gemini, the system achieves high accuracy. This project not only presents a valuable tool for organizations but also contributes to a broader understanding of LLMs and their capabilities.

II. Acknowledgement

I would like to express my sincere gratitude to all those who have contributed to this project. I would like to thank my supervisor, Dr. Kong Lingpeng, for his guidance and support throughout the project. I would also like to thank my groupmate, Jasper, for his contributions and collaboration. Finally, I would like to thank my CAES9542 instructor, Mr. Matthew Anderson, for his guidance in technical writing and presentation.

III. Table of Contents

I. Abstract **1**

II. Acknowledgement **2**

III. Table of Contents **3**

IV. List of Figures **5**

V. List of Tables **6**

VI. Abbreviations **7**

1. Introduction **8**

1.1 Background 8

1.1.1 User Feedback 8

1.1.2 Social Media 8

1.1.3 Sentiment Analysis 9

1.1.4 Large Language Models 9

1.2 Project Motivations 9

1.2.1 Automated Social Media Research 9

1.2.2 Sentiment Analysis using LLMs 9

1.3 Objectives 10

1.4 Deliverables 10

1.5 Project Contributions 10

1.5.1 Invaluable Tool for Organizations 10

1.5.2 Exploring Potential of LLMs 10

1.6 Report Outline 11

2. Literature Review **12**

2.1 Related Work 12

2.1.1 Sentiment Analysis 12

2.1.2 Large Language Models 13

2.1.3 Sentiment Analysis with Large Language Models 14

2.2 Research Gap 15

3. Methodology **16**

3.1 System Design....16
3.2 Procedures....17
3.2.1 Data Collection....17
3.2.2 Data Analysis....21
3.2.3.1 Aspect Identifcation and Classification....21
3.2.3.2 Aspect-based Sentiment Analysis....23
3.2.3.3 Aspect-based Summarization....24
3.2.4 Result Visualization....26
3.3 Testing and Evaluation....26
4. Results and Discussion....27
4.1 Data Collection....27
4.2 Data Analysis....28
4.2.1 Aspect Identification and Classification....28
4.2.2 Aspect-based Sentiment Analysis....29
4.2.3 Aspect-based Summarization....30
4.3 Result Visualization....31
5. Future Works....33
5.1 Improved Data Collection....33
5.2 Use of Offline Models....34
5.3 Depth of analysis....35
6. Conclusion....36
VII. References....37

IV. List of Figures

Figure 1. An overview of the system workflow 16
Figure 2. API plans for accessing X's data 18
Figure 3. "Page Public Content Access" permission required to access Meta's data 18
Figure 4. Quota system for accessing YouTube's data 19
Figure 5. Workflow for aspect identification and classification 21
Figure 6. Workflow for aspect summarization 24
Figure 7. A sample CSV file with data collected from Reddit 27
Figure 8. A sample CSV file with data classified into identified aspects 28
Figure 9. A sample CSV file with data and assigned sentiment scores 29
Figure 10. A sample summary for an identified aspect 30
Figure 11. A sample overview page in a generated report 31
Figure 12. A sample summary page in a generated report 31
Figure 13. Posts sorted by "relevance" or "hot" in Reddit 33
Figure 14. An internal server error occurred when using Gemini 34
Figure 15. Collecting textual content and rating of relevant comments 35

V. List of Tables

Table 1. Targeted social media platforms for data collection 17

Table 2. Scores assigned to each sentiment 23

VI. Abbreviations

LLM	Large Language Model
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformers
MAST	Multifaceted Analysis of Subjective Texts
RoBERTa	Robustly optimized BERT approach
BART	Bidirectional and Auto-Regressive Transformer
T5	Text-To-Text Transfer Transformer
MMLU	Massive Multitask Language Understanding
API	Application Programming Interface
CSV	Comma-Separated Values
VPN	Virtual Private Network
PDF	Portable Document Format
HTML	HyperText Markup Language

1. Introduction

This section introduces the project. Section 1.1 provides the project's background; section 1.2 explains the project's motivations; section 1.3 outlines the project's objectives; section 1.4 details the project's deliverables; section 1.5 justifies the project's contributions; section 1.6 provides an overview of the remaining content of the report.

1.1 Background

1.1.1 User Feedback

User feedback is an invaluable resource for businesses, enabling them to gain insights into the needs and preferences of their users [1]. This understanding allows them to identify areas for improvement, and make informed decisions in the development of new products [1]. Effectively addressing user feedback also make users feel acknowledged and result in enhanced customer loyalty.

Traditionally, collecting user feedback involves the usage of questionnaires, interview, and focus groups [2]. However, these approaches come with several drawbacks. Firstly, they rely on limited sample sizes, leading to inadequate representation of certain user segments. Furthermore, designing and conducting these methods require significant time and effort. Additionally, they typically provide predefined questions and lack open-ended discussions, thereby limiting the depth of the feedback collected.

1.1.2 Social Media

In today's digital age, social media has emerged as an integral part of our lives, offering individuals a means to freely express opinions and engage in discussions. As of March 2024, it is estimated that over 63% of the world's population are active social media users [3]. The wide usage in social media has resulted in a substantial volume of user-generated content across such platforms, making them valuable resources for organizations to gain insight into user feedback and public sentiment.

Social media encompasses a wide range of platforms that cater to different content and communication styles. They can be broadly classified into several categories: social networking sites, such as Facebook, facilitate connections and socialization; discussion forums, like Reddit, allows users to engage in question-and-answer interactions; video-sharing platforms, like YouTube, focus on the sharing and viewing of videos [3].

1.1.3 Sentiment Analysis

Sentiment analysis is the process of analyzing emotions expressed in text and classifying them into positive, neutral, or negative sentiments [4]. It has found broad applications, especially for businesses to perform brand monitoring, conduct market research, or analyze user feedback [5].

1.1.4 Large Language Models

Large Language Models (LLM) are machine learning models designed for Natural Language Processing (NLP) tasks. Leveraging transformer architectures, these models demonstrate proficiency in understanding and generating human-like languages, making them valuable assets in various applications such as translation and text generation [6]. Prominent examples of state-of-the-art LLMs include Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformers (GPT) and Gemini.

1.2 Project Motivations

1.2.1 Automated Social Media Research

The vast amount of user-generated content on social media offers organizations a valuable resource for collecting user feedback and conducting sentiment analysis. However, the conventional manual approach of collating, analyzing, and synthesizing findings can be labor-intensive, time-consuming, and prone to errors, primarily due to the sheer volume of posts and comments across multiple platforms [1]. Therefore, this project aims to address these challenges by developing a tool to automate the research process.

1.2.2 Sentiment Analysis using LLMs

Existing approaches to sentiment analysis have shown limitations, particularly in understanding human language accurately [7]. Considering the capacity of LLMs to comprehend human language [6], this project aims to explore the utilization of LLMs as an improved approach to sentiment analysis.

1.3 Objectives

The aim of this project is to develop an automated system that conducts social media research and performs sentiment analysis utilizing state-of-the-art LLMs. In line with this agenda, this project is designed to achieve the following objectives:

1. Collect textual data from social media that are relevant to user-specified keywords.
2. Conduct aspect-based sentiment analysis on the collected data using LLMs.
3. Generate reports with appropriate visualizations to present the analysis result.

1.4 Deliverables

This project aims to deliver a command-line program that accomplishes the objectives mentioned in section 1.3. The program should encompass the following procedures: data collection, aspect classification, aspect-based sentiment analysis, aspect-based summarization, and report generation.

1.5 Project Contributions

1.5.1 Invaluable Tool for Organizations

The system delivered by this project functions as a powerful tool for organizations to collect user feedback and analyze public sentiment. By inputting relevant keywords, such as product names, the system can generate comprehensive reports that a provide a clear understanding of the overall opinion surrounding the products. This enables them to readily identify areas for improvement and strategically plan their next steps. Moreover, the utilization of this system effectively addresses the labor-intensive and time-consuming aspects of manual analysis, while also offering the advantage of improved analysis accuracy.

1.5.2 Exploring Potential of LLMs

Several tools in this domain do already exist, such as Brandwatch and Brand24 [8, 9]. However, this project also explores the capabilities of state-of-the-art LLMs. These models have demonstrated remarkable performance in tasks such as question answering and text generation [6]. This project seeks to demonstrate the potential of these models in the context of aspect-based sentiment analysis.

1.6 Report Outline

The remainder of this report is organized into five sections. Section 2 presents work related to the project; section 3 explains and justifies the methodology employed in the project; section 4 discusses the results obtained from the project; section 5 proposes areas of improvement for future work; section 6 summarizes the essential components of this report.

2. Literature Review

This section presents the literature related to the project. Section 2.1 summarizes the work on related topics; section 2.2 identifies research gaps for this project.

2.1 Related Work

2.1.1 Sentiment Analysis

Sentiment analysis has emerged as an evolving research area in the field of NLP owing to two fundamental factors. First, sentiment analysis has witnessed an increase in its applications, particularly in light of the exponential growth pf user-generated content [10]. This includes business activities such as monitoring opinions and extracting market insights [11]. Second, as explained by Bubeck et al. [12], the ability to comprehend sentiments from text plays a crucial role not only in sentiment analysis itself but also in the broader context of human-level intelligence.

The analysis can be classified into various types based on scope and complexity. At its fundamental level, document-level sentiment analysis examines the sentiment of an entire text [13]. Building upon this foundation, sentiment analysis has progresses along two dimensions: depth and width [14]. In terms of depth, aspect-based sentiment analysis explores further by identifying specfic aspects within a text and analyzing the sentiment associated with each identified aspect [15]. On the other hand, the wide direction expands the analysis to encompoass multifaceted analysis of subjective texts (MAST), performing specialized tasks focusing on specific sentiment, such as hate speed detection and irony detection [16]. Nevertheless, these types of analysis involve determining sentiment in either binary format (postive, negative) or multi-class format (positive, neutral, negative) [17].

There are three main approaches to performing sentiment analysis. The first approach is the lexicon-based approach, which involves assigning sentiment scores to keywords based on a predefined dictionary, and aggregating these scores to determine the overal sentiment of the context [18]. While this method is straightforward and easy to interpret, it is susceptible to inaccuracies since the same word can convey both positive and negative meanings depending on the context [19]. The second approach is the traditional machine learning approach, which utilizes supervised learning techniques to predict sentiment in text, with algorithms such as Naive Bayes and Random Forests [20]. This method is more sophisticated and surpasses the lexicon-based approach in capturing contextual nuances of text [21]. However, it still falls

short of fully comprehending human language, particularly in areas like irony and sarcasm [7]. The third and most recent approach involves the use of transformer-based models. This method demonstrates superior contextual understanding compared to the previous approaches, and exhibits capability to handle negations, intensifiers, and implicit sentiments [22]. Experimental results indicate that this approach achieves 20% higher accuracy than the lexicon-based approach, and 10% higher accuracy than the traditional machine learning approach [23].

2.1.2 Large Language Models

LLMs are machine learning models specifically designed to comprehend and generate human-like language, enabling them to perform a wide range of NLP tasks. They are trained on extensive collections of unlabeled text corpora using self-supervised learning techniques, allowing them to acquire an understanding of the inherent context, grammar, and structure of language [24]. One notable feature of LLMs is their remarkable capabilities in zero-shot or few-shot learning settings [14]. Zero-shot learning refers to the model's ability to execute a task without any training examples. Instead, it relies on its own understanding of the underlying language. Similarly, few-shot learning refers to the ability of the model to learn quickly from a limited number of training examples. These capabilities contribute to the versatility and adaptability of LLMs, enabling them to handle NLP tasks even when there is limited or unavailable training data. The applications of LLMs span a wide range of areas, including language translation, text summarization, question-answering chatbots, and sentiment analysis [6].

LLMs can be categorized into three types based on their architectural design. The first type is known as encoder-based modesl, which solely utilize the econder of a transformer model to produce input representatives [6]. These models are commonly employed in tasks such as sentence classification and sentiment analysis [6]. Prominent examples include the Bidirectional Encoder Representations from Transformers (BERT) and its variants like the Robustly optimized BERT approach (RoBERTa) [25, 26]. The second type is referred to as decoder-based models, which solely utilize the decoder of a transformer model to generate output sequences given an input text [6]. They are particularly well-suited for text generation tasks [6]. The Generative Pre-trained Transformers (GPT) and Gemini are notable representatives of this group [27, 28]. The third type is the encoder-decoder model, which incorporates both the encoder and decoder components, thereby enabling the generation of

output sequences based on input representatives [6]. Such models are advantageous in text summarization tasks [6]. Examples of this type include the Bidirectional and Auto-Regressive Transformer (BART) and the Text-To-Text Transfer Transformer (T5) [29, 30].

Various strategies have been devised and implemented to improve the performance of LLMs. One prominent approach is prompt optimization. By designing well-optimized prompts, significant improvements can be achieved in the capabilities and output quality of LLMs [31]. Key elements of a well-optimized prompt include clear instructions, relevant context, example implementation, input data, and an output indicator [31]. In a recent study by Yao et al. [32], the concept of "Tree of Thoughts" was introduced as an extension of prompt optimization. This approach involves decomposing a task into smaller sub-tasks and solving them sequentially, guided by reasoning chains generated by LLMs themselves [32].

Gemini, introduced by Google on 6 December 2023, is a recent addition to the landscape of LLMs. It features a decoder-only architecture and is characterized by its multimodal capability to process text, images, audio, and videos [28]. Notably, it showcases exceptional performance in the domain of Massive Multitask Language Understanding (MMLU), achieving results comparable to those of the GPT-4 model [28].

2.1.3 Sentiment Analysis with Large Language Models

Mutlitple attempts have been made to leverage LLMs for sentiment analysis. Kheiri and Karimi explored the use of GPT-3.5 Turbo model for document-level sentiment analysis [31]. They investigated prompt-based GPT models, fine-tuned GPT models, and embedding-based GPT models. Among these approaches, their prompt-based GPT model implementation demonstrated superior performance, with accuracy rates over 90%. On the other hand, Hoang et al. [33] employed the BERT model to conduct aspect-based sentiment analysis. They developed three distinct models for aspect classification, sentiment classification, and combined aspect-sentiment classification. Their combined model achieved accuracy rates of more than 85%.

2.2 Research Gap

As shown in section 2.1.3, it is observed that there are existing research closely related to our project. However, their work primarily focused on few-shot learning, which involves training LLM with a number of samples. This presents a research gap regarding the application of LLMs for sentiment analysis within the context of zero-shot learning.

To bridge this gap, this project aims to develop a system that leverages LLM to perform aspect-based sentiment analysis, under zero-shot learning conditions. The successful implementation of the project would deliver a system that is capable of adapting to new and emerging topics in real-time, without the need for extensive pre-training on specific domains.

3. Methodology

This section explains and justifies the methodology employed in the project. Section 3.1 highlights the overall design of the system; section 3.2 elaborates the procedures involved in conducting social media research and sentiment analysis; section 3.3 introduces the dataset used for testing and evaluation.

3.1 System Design

The system was developed as a command-line program. This approach allowed for a more focused allocation of efforts, prioritizing the enhancement of core capabilities for social media research and sentiment analysis, and avoiding the complexities associated with designing and developing dedicated frontends or backends. Additionally, the use of command-line programs provides organizations with flexibility in integrating the system with their existing applications. Furthermore, it offers room for future expansions, for example enabling support for additional social media platforms as needed.

Python was chosen as the programming language for the system based on two reasons. First, Python provides access to a wide range of libraries, such as csv and pandas, that align well with the specific requirements of the project. Second, Python is widely recognized as the preferred language for data analysis and machine learning, ensuring access to a wealth of online resources and documentation.

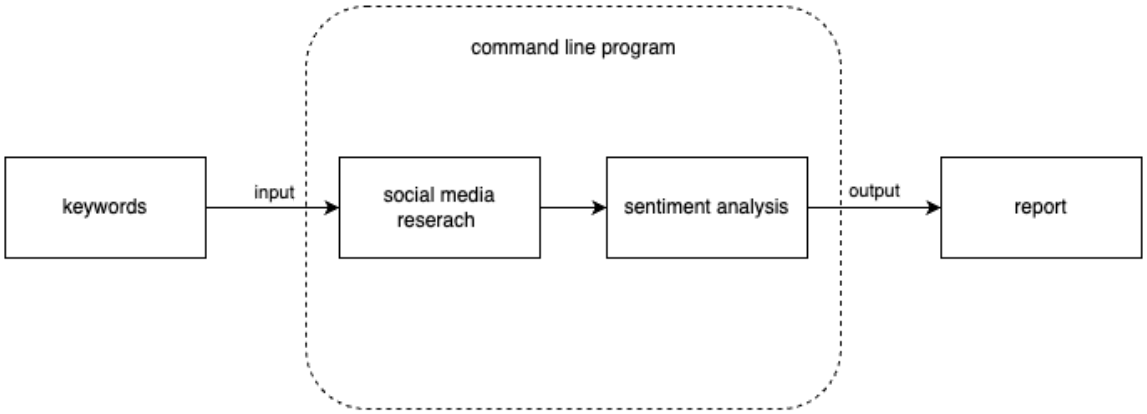

Figure 1. An overview of the system workflow

Figure 1 illustrates the overall workflow of the system. First, users were prompted to input keywords of interest, which could be specific product names or other relevant terms. Then, the system would initiate social media research by collecting data related to the provided keywords. Subsequently, sentiment analysis would be performed on the collected data. Finally, the system would generate a report to present the analysis results.

3.2 Procedures

The system consists of three primary procedures for performing social media research and sentiment analysis: data collection, data analysis, and result visualization.

3.2.1 Data Collection

	X	**Reddit**	**YouTube**	**Facebook**	**Instagram**
rank [3]	10th	15th	3rd	1st	4th
type	discussion forums		video-sharing platform	social networking sites	
official API	yes	yes	yes	yes	yes

Table 1. Targeted social media platforms for data collection

Table 1 presents the social media platforms initially identified for data collection, namely X, Reddit, YouTube, Facebook, and Instagram. These platforms were chosen for several reasons. First, they are among the most widely used social media platforms, ensuring availability of extensive user-generated content. Second, the selection encompasses major types of social media, allowing the collection of diverse content from different platforms. Third, each of the chosen platforms provides official APIs for reliable and efficient access to their data.

Find the right access for you

Free

For write-only use cases and testing the X API

- Rate limited access to v2 post posting and media upload endpoints
- 1,500 Posts per month - posting limit at the app level
- 1 app ID
- Login with X
- Free

Get started

Basic

For hobbyists or prototypes

- Rate limited access to suite of v2 endpoints
- 3,000 Posts per month - posting limit at the user level
- 50,000 Posts per month - posting limit at the app level
- 10,000 Posts per month - read-limit rate cap
- 2 app IDs
- Login with X
- $100 per month

Subscribe now

Pro

For startups scaling their business

- Rate-limited access to suite of v2 endpoints, including search and filtered stream
- 1,000,000 Posts per month - GET at the app level
- 300,000 Posts per month - posting limit at the app level
- 3 app IDs
- Login with X
- $5,000 per month

Subscribe now

Enterprise

For businesses and scaled commercial projects

- Commercial-level access that meets your and your customer's specific needs
- Managed services by a dedicated account team
- Complete streams: replay, engagement metrics, backfill, and more features
- Monthly subscription tiers

Apply now

Figure 2. API plans for accessing X's data [33]

Page Public Content Access

The **Page Public Content Access** feature allows an app access to the Pages Search API and to read public data for Pages for which you lack the **pages_read_engagement** permission and the **pages_read_user_content** permission. Readable data includes business metadata, public comments and posts. The allowed usage for this feature is to analyze and/or display posts and engagement on Pages.

Allowed Usage

- Analyze and/or display posts and engagement on Pages.

Common Endpoints

/page/feed
/page-post
/page-post/comments

Additional Details

- This permission or feature requires successful completion of the App Review process before your app can access live data. Learn More
- This permission or feature is only available with business verification. You may also need to sign additional contracts before your app can access data. Learn More Here
- While you are testing your app and before you submit it for review, your app can only access content on a Page for which the following is true: The person who holds the admin role for the Page also holds an admin, developer, or tester role on the app. If you want the app to be able to access public content on other Pages, you must submit this feature for review. Once you set your app to live mode, it will not be able to see any Page public content without this feature.

Figure 3. "Page Public Content Access" permission required to access Meta's data [34]

Home > Products > YouTube > Data API > Guides

Was this helpful?

YouTube Data API - Quota and Compliance Audits

The YouTube Data API uses a quota system to ensure that developers use the service as intended and do not create API clients that unfairly reduce service quality or limit access for others.

Projects that enable the YouTube Data API have a default quota allocation of 10,000 units per day, an amount sufficient for the majority of our API users. You can see your quota usage on the Quotas page in the API Console.

If you would like to request additional quota beyond the default allocation, you must first complete an audit to show that your project is in compliance with the YouTube API Services Terms of Service. This gives YouTube visibility into the intended use cases of large projects and ensures that YouTube's API services are being used in a manner that is free from abuse. Visit this link for additional details on complying with YouTube's Developer Policies.

To begin an audit of your project, please fill out and submit the YouTube API Services - Audit and Quota Extension Form. A member of YouTube's API Services team will get back to you as soon as possible.

If you have completed an API Compliance Audit within the last 12 months but require an additional quota extension, please fill out and submit the Audited Developer Requests Form.

If you have recently failed an API Compliance Audit and would like to appeal that decision, please fill out and submit the Appeals Form.

Figure 4. Quota system for accessing YouTube's data [35]

However, in the final implementation of the system, only Reddit was utilized as the source for data collection. The implementation for the other platforms was prevented by certain obstacles. In the case of X, although an official API is available, it comes with a significant cost of US$100 per month for read access to its data, as depicted in Figure 3. Considering the project's budget of HK$2,000, utilizing the API was not feasible due to financial constraints. Regarding Facebook and Instagram, Meta provides official APIs for both platforms, but accessing their public data requires obtaining the "Page Public Content Access" permission, as shown in Figure 3. Unfortunately, obtaining this permission had proven to be challenging. As for YouTube, an affordable and accessible official API does exist. However, its usage is constrained by a quota cost of 10,000 units, as shown in Figure 4. For instance, searching for relevant videos incurs a cost of 100 units [35]. Such limitation restricts the amount of data that can be collected from the platform. While web scraping could potentially address the challenges posed by these platforms, it is explicitly stated in their terms of service that such a practice is prohibited [33, 34, 35].

Nevertheless, by focusing solely on Reddit, the implementation was streamlined and ensured consistency throughout procedures. Furthermore, in addition to the keywords of interest, the final implementation allowed users to specify a particular subreddit page from which the

system would collect data from. Consequently, the collected data became more relevant and directly applicable to the desired subject matter.

During data collection, the primary focus was on extracting the textual content of relevant posts or comments, regardless of the availability of other data types such as images and videos. This is to align with the subsequent analysis, which primary revolves around text-based content. However, due to the vast amount of content available on Reddit, it would be impractical to collect all available data relevant to the provided keywords. To strike a balance between computational feasibility and data comprehensiveness, the system limited the number of data instances collected to a maximum of 5,000.

Following the data collection process, the collected data was stored in Comma-Separated Values (CSV) files. This file format was chosen for its ease of retrieval and compatibility with subsequent procedures.

3.2.2 Data Analysis

The collected data then underwent analysis using LLMs. Following the approach proposed by Hoang et al. [33], this analysis procedure was divided into three distinct stages: aspect identification and classification, aspect-based sentiment analysis, and aspect-based summarization. CSV files were utilized as a means to transfer results between stages.

3.2.3.1 Aspect Identifcation and Classification

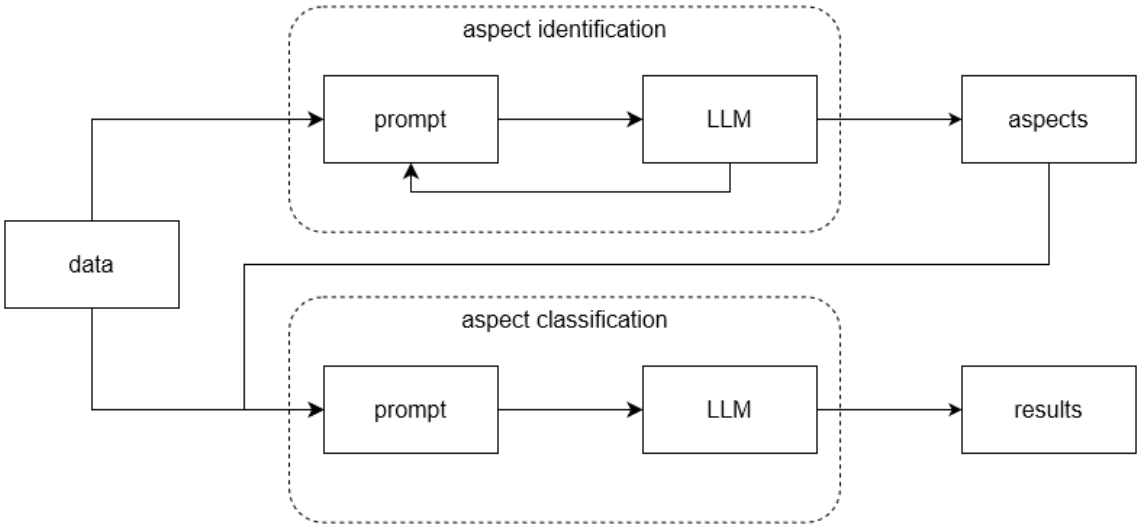

Figure 5. Workflow for aspect identification and classification

In this stage, aspects within the pre-processed data was identified, and the data was classified based on these identified aspects. The overall flow for this stage is illustrated in Figure 5. First, the data was structured into a suitable prompt and input into a LLM for aspect identification. The identified aspects were then utilized to structure another prompt, subsequently fed into a separate instance of the LLM for aspect classification.

Due to the limitations on input prompt length for LLMs, the described process was executed in batches of 50 data instances. In each batch, all 50 data instances were used to structure the prompt for aspect identification. However, for the prompt used in aspect classification, the batch was further divided into groups of 5. This setup resulted in the best accuracies during testing, as it allowed the LLM to conduct classification with greater focus and manageability. Moreover, to ensure consistency and avoid variations in the wording of the same aspect

across different batches of data, the identified aspects from previous batches were included as references in the prompt for aspect identification.

Following the guide provided by Li and Liang [31], the prompts were optimized to include clear instructions, relevant context, example implementation, input data, and an output indicator. Below are the prompts used for aspect identification and classification respectively:

"You are analysing comments collected regarding '{keywords}' from the subreddit page '{subreddit}'. Given the following comments, identify the key aspects that are most frequently mentioned. The aspects can and are preferred to be general categories. The output should be a list of strings, each enclosed in quotation marks. For example, given the comments ['pasta needs more flavour', 'pasta is too soft', 'pasta is too expensive', 'floor is dirty'], the output should be ['food', 'pasta', 'price', 'environment']. Please ensure the output is in this format only. Here are some previously identified aspects for your reference: {identified_aspects}. It is preferred to reuse these aspects, but feel free to introduce new ones if necessary. The comments for analysis are as follows: {comments}."

"You are analysing comments collected regarding '{keywords}' from the subreddit page '{subreddit}'. Given the following comments and a list of identified aspects, assign the most relevant aspect(s) to each comment. A comment can be associated with multiple aspects if it covers more than one aspect. The output should be a list of tuples, where each tuple contains the comment's corresponding aspect(s). If there is no aspects to be assigned, leave the aspects part as an empty list. For example, given the identified aspects ['food', 'price', 'environment', 'water'] and the comments ['pasta needs more flavour', 'too expensive', 'I love the ambiance', 'the pasta needs more flavour and is too expensive', 'nice music'], the output could be [('pasta needs more flavour', ['food']), ('pasta is too expensive', ['price']), ('I love the ambiance', ['environment']), ('the pasta needs more flavour and is too expensive', ['food', 'price']), ('nice music', [ ])]. Please ensure the output is in this format only. The identified aspects are: {aspects}. The comments for analysis are as follows: {comments}."

Based on the research by Kheiri and Karimi [31], the GPT model was initially selected as the LLM of choice to extend their implementation from document-level sentiment analysis to aspect-based sentiment analysis. However, it was subsequently discovered that direct access to GPT services from Hong Kong was not officially supported. While this was temporarily

addressed with the use of a virtual private network (VPN), the access to GPT services was once blocked again when it became necessary to provide payment information, specifically a foreign credit card, after the expiration of the free trial period. Consequently, Gemini was utilized as the chosen LLM in the final implementation. Although a VPN was still necessary to establish access to Gemini services, the services were free of charge and did not require a foreign credit card for usage.

3.2.3.2 Aspect-based Sentiment Analysis

Sentiment	Score
Positive	0
Neutral	1
Negative	2

Table 2. Scores assigned to each sentiment

Once the data was categorized into the identified aspects, the next stage involves classifying them into specific sentiments related to their respective aspects. This was accomplished by assigning scores to data based on their underlying sentiment, determined by the LLM's understanding of the context, language, and structure of the data. This approach takes inspiration from encoder-based models that are proficient in performing sentiment analysis [6]. The multiclass format was employed to differentiate between positive, neutral, and negative sentiments. Table 2 provides an overview of the scores assigned to each sentiment.

The following prompt was formulated and input into the LLM to perform the above task:

"You are analysing comments collected regarding '{keywords}' from the subreddit page '{subreddit}'. Given a list of comments and a specific aspect, perform sentiment analysis on the comments with respect to the given aspect. The sentiment should be classified as positive, neutral, or negative. Assign a numerical label to each sentiment: 2 for positive, 1 for neutral, and 0 for negative. The output should be a list of these numerical labels, in the same order as the provided comments. For example, given the aspect 'pasta' and the comments ['pasta

needs more flavour', 'pasta is too expensive', 'I love the ambiance', 'pasta needs more flavour and is too expensive', 'nice music', 'taste delicious'], the output should be [0, 1, 1, 0, 1, 2]. Note that comments not relevant to the aspect should be labeled as 'neutral' (1). Comments that are satisfactory towards the current state shall be classified as positive. Beware of the use of sarcasm within the comments too. The given aspect is: {aspect}. The comments for analysis are as follows: {comments}."

For the same reasons elaborated in section 3.2.3.1, Gemini was employed as the LLM of choice for this stage.

3.2.3.3 Aspect-based Summarization

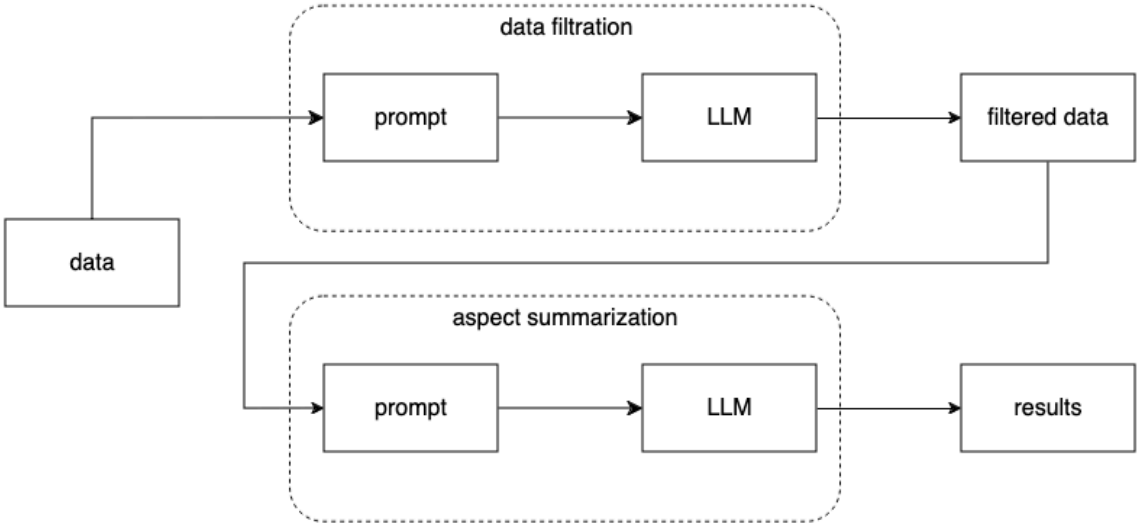

Figure 6. Workflow for aspect summarization

Following the sentiment analysis performed on individual data, the results were aggregated and summarized to provide aspect-based insights. The overall flow of this state is depicted in Figure 6. The first step involved filtering out non-essential data by utilizing an appropriate prompt and an instance of the LLM. This process facilitates the extraction of aspect-specific sentiments. Subsequently, another prompt and a separate instance of the LLM were employed to generate a summary of the sentiment associated with each aspect.

In addition to the sentiment analytics, the system also extracted key points from the data during the summarization procedure to provide a comprehensive result. This involved

identifying prevalent positive and negative viewpoints expressed within the data, as well as any suggested advice for improvement. The summarization capabilities of the LLMs were leveraged to generate concise and informative summaries. The elements were formatted in bullet points instead of paragraphs to enhance readability.

The following prompts were utilized for data filtration and aspect summarization respectively:

"You are analysing comments collected regarding '{keywords}' from the subreddit page '{subreddit}'. You are preparing to include certain comments in a subsequent market report. The collected comments have been categorized into various aspects, and you are currently examining the aspect {aspect}. From the provided comments, you need to select the ones that you deem important and valid for further analysis and inclusion in the report. Label the comments you wish to analyze later with 1, and the rest with 0. Return the results as a list of labels. For example, if the comments were ['pasta needs more flavour', 'pasta is too expensive', 'I love the ambiance', ' pasta needs more flavour and is too expensive'] and you consider the first and fourth comments as important, return [1, 0, 0, 1]. You must only provide the required format (the list) without any explanations. The comments for analysis are as follows: {comments}."

"You are analysing comments collected regarding '{keywords}' from the subreddit page '{subreddit}'. Focus specifically on the aspect: '{aspect}'. When comments contain multiple aspects, extract and include only the information pertinent to the aspect in focus. In other words, ignore other aspects as they will be covered in other paragraphs written later. The comments will be presented in a list format. Your objective is to synthesize a comprehensive report based on these comments, highlighting the predominant sentiments, suggestions, and requests articulated by the users. Where comments reflect both positive and negative sentiments, delineate and summarize these viewpoints distinctly. Delve into the context and subtleties within the comments to unearth valuable insights that may inform future strategies or decisions. The comments for analysis are as follows: {comments}."

3.2.4 Result Visualization

The results obtained from the data analysis procedure were presented using suitable visualization and exports as a Portable Document Format (PDF) file.

For textual information, such as prevailing viewpoints, suggestions, and other insights, they were presented as obtained from the previous procedure. As for statistical data, including the frequency distributions of aspects and the sentiment ratios, they were visualized using Matplotlib to create appropriate diagrams and charts.

To combine all the elements together and export them as a report, FPDF was utlized. The library was chosen for its versatility and its ability to support Mardown and HyperText Markup Language (HTML), enabling customization options.

3.3 Testing and Evaluation

To evaluate the performance of the utlized prompts as well as LLM, data collected from Reddit relevant to World of Warships was utilized for testing. It is a player-versus-player game that has generated substantial debates concerning the concept of “game balance”, which refers to the fairsness and equilibrium of power among different warships in the game. Discussions found on Reddit provided clear definitions of the aspects, corresponding to the specific warships under discussion, as well as the stance, indicating whether the warship is perceived as too strong or weak. This clarity in the dataset allows for a straightforward quantification and assessment of the system’s performance. In each test, 100 data instances were randomly selected from the resulting CSV files and subjected for evaluation.

4. Results and Discussion

This section presents and dicusses the results of the project. Section 4.1 dicusses those regarding data collection; section 4.2 discusses those regarding data analysis; section 4.3 presents those regarding result visualization.

4.1 Data Collection

Text
Great and helpful contribution ! What would you say about the Soviet sub (tier VI
One caveat is that for DDs, especially hydro-equipped DDs, subs definitely have co
I just got hit for 42K when the sub was less than 3km away. I'm not sure if the shot
It's torp distance traveled, not distance to the sub, so if you were sailing away from
I wasn't pinged, so I'm pretty sure they weren't homing torpedoes, plus, it was a T
I was just mentioning the homing torps since it's something most people probably
Gato's unguided torps do over 20k per, and the homing torps are 10k per, so 2-4
I've used it twice so far in ranked to cowboy tank an attempted shotgun and sink t
Except that now they can and will just surface into you thanks to the turning buff.
Name a single sub that isn't the I-56 that can ram a DD and actually live.
Thrasher did it to me and ate 4 depth charges and lived. Any sub with more HP th
First of all how did you manage to get rammed by a thrasher??? That thing is genu
Ended up nose to nose in my Lushun, turned and he matched me started moving
Huh, that's actually hilariously unlucky. I personally would have laughed and move
Honestly I think my entire chain here is spawned from that day being awful for me
if someone gets a submerged sub (other than U-4501) on hydro, then i would thin
The problem is the closing speed if you end up nose to nose or close to that. Subs
dds dont need to get that close anymore. ship based asw has 3km range now, if

Figure 7. A sample CSV file with data collected from Reddit

Regarding data collection, the system was able to collect relevant posts and respective comments from Reddit, based on the user-provided keywords and, if applicable, a specific subreddit. Figure 7 shows an example of the resulting CSV files. At this stage, the CSV files only consist of the “Text” column, which represents the textual content of the relevant posts and comments.

The data collection process was efficient, typically requiring less than three minutes to gather thousands of data instances. This promptness could be attributed to the utilization of the official Reddit API, along with the Python Reddit API Wrapper (PRAW) library that streamlined the interaction between the system and the API.

4.2 Data Analysis

4.2.1 Aspect Identification and Classification

Text	Aspects
could instantly materialize out of thin air to drop charges.	['historical accuracy']
ming pulling out the historical accuracy card	['historical accuracy']
y nothing about the game is realistic or historical to begin with.	['historical accuracy']
) are good enough at hitting targets past 8km without calculus t	['submarines', 'game mechanics', 'counterplay']
ng I dislike are CVs when I play torp DD.	['submarines', 'CVs']
e standard, U4501 is only barely better than the non-premium B	['submarines']
en use the ping to set their depth and last minute corrections wi	['homing torpedoes', 'shotgunning']
u in response, not to mention the U4501 has an odd launcher co	['submarines', 'shotgunning']
ealment and mobility. Besides, she has almost the same homing	['submarines']
efore they are overrun. What exactly? Certainly not the sub hydr	['submarines', 'shotgunning']
s being searched by someone using hydro thanks to the comma	['submarines', 'hydroacoustic search', 'shotgunning']
what makes shotgunning harder? Yeah that makes sense.	['shotgunning']
me cruisers may have sub radar, and that several DC cruisers are	['shotgunning', 'submarines', 'radar']
, just more difficult against some specific target due to subs ove	['shotgunning', 'submarines']
uch less damage within 3km	['torpedoes', 'damage']
ut the shotgunning is claimed to be nerfed in the earlier sectior	['shotgunning']
ely overhaul all dumbfires for all submarines in addition to maki	['shotgunning']
l make it way harder to randomly surface next to one and oblite	['detection']
deal low damage until about 3 km distance travelled. But it cou	['torpedoes']

Figure 8. A sample CSV file with data classified into identified aspects

Using the collected data, the system was able to first identify aspects within the data, and then classify the data into the identified aspects. Figure 8 provides an example of the resulting CSV files, with data collected from the previous procedure and classified into the identified aspects. These files consist of an additional column “Aspects’, which lists the aspects identified from the corresponding post or comment.

The system required around 30 minutes for every 1000 comments during the procedure, and achieved an accuray rate of 0.8649 for this task. It is worth noting that most inaccuracies occurred due to the LLM generating generic outputs. One particular scenario involved comments mentioning multiple ships in the game. In such cases, the LLM tended to classify the comment under the broad aspect of 'warships' instead of recognizing and categorizing each individually mentioned warship.

The accuracy was tested against an alternative using the facebook/bart-large-mnli model, which is an encoder-based model specifically designed for zero-shot classification. However, the accuracy achieved by this alternative model was only 0.3514, significantly lower than that by Gemini. This discrepancy in performance can be attributed to several factors. Firstly, the encoder-based model demonstrated a strong bias towards exact wordings, resulting in

numerous false classifications. Additionally, the alternative model struggled with capturing some abbreviations, leading to the omission of many aspects during the classification process. While it is possible that these issues could be mitigated through further fine-tuning of the encoder-based model, it is important to highlight the advantages of LLMs in this context, in particular their ability to generalise information and comprehend text beyond exact wordings.

4.2.2 Aspect-based Sentiment Analysis

Text	Sentiment
Great and helpful contribution ! What would you say about the Soviet sub (tier VI	{'submarines': 1}
One caveat is that for DDs, especially hydro-equipped DDs, subs definitely have cc	{'submarines': 0}
I just got hit for 42K when the sub was less than 3km away. I'm not sure if the shot	{'submarines': 0}
It's torp distance traveled, not distance to the sub, so if you were sailing away from	{'torpedoes': 1}
I wasn't pinged, so I'm pretty sure they weren't homing torpedoes, plus, it was a T	{'subs': 1, 'torpedoes': 1}
I was just mentioning the homing torps since it's something most people probably	{'subs': 1, 'torpedoes': 1}
Gato's unguided torps do over 20k per, and the homing torps are 10k per, so 2-4	{'subs': 1, 'torpedoes': 1}
I've used it twice so far in ranked to cowboy tank an attempted shotgun and sink t	{'subs': 2, 'torpedoes': 1}
Except that now they can and will just surface into you thanks to the turning buff. I	{'subs': 0}
Name a single sub that isn't the I-56 that can ram a DD and actually live.	{'submarines': 0}
Thrasher did it to me and ate 4 depth charges and lived. Any sub with more HP tha	{'submarines': 2}
First of all how did you manage to get rammed by a thrasher??? That thing is genu	{'submarines': 1}
Ended up nose to nose in my Lushun, turned and he matched me started moving	{'submarines': 0}
Huh, that's actually hilariously unlucky. I personally would have laughed and move	{'submarines': 1}
Honestly I think my entire chain here is spawned from that day being awful for me	{'subs': 0}
if someone gets a submerged sub (other than U-4501) on hydro, then i would thin	{'concealment': 1, 'subs': 1}
The problem is the closing speed if you end up nose to nose or close to that. Subs	{'concealment': 1, 'subs': 0}
dds dont need to get that close anymore. ship based asw has 3km range now, if	{'concealment': 1, 'subs': 0}

Figure 9. A sample CSV file with data and assigned sentiment scores

With the data classified into the identified aspects, the system was able to assign a sentiment score to each aspect of the data. Figure 9 illustrates a resulting CSV file from this procedure. These files include an additional column labelled “Setiments”, which lists the sentiment scores assigned to each classified aspects of the data.

During manual evaluation, it was observed that the performance of the sentiment analysis results varied depending on the nature of the aspect being analyzed. On controversial topics characterized by a high volume of arguments and complaints, the system achieved a higher accuracy of 0.8461. However, on less controversial topics where comments tended to be more neutral or the sentiments expressed are less polarized, the accuracy of the system degraded to a range of 0.5 to 0.7.

The performance degradation can be attributed to the inherent complexity of aspect-based sentiment analysis. For instance, comments that exhibit mixed sentiments can pose challenges. An example comment like "Glad that I bought an RTX4090 before its price skyrocketed" contains both positive sentiment, expressed by “glad”, and a hint of dissatisfaction related to the price increase. Such mixed sentiments can make aspect-based sentiment analysis difficult, even for human evaluators. It was observed that these comments with mixed sentiments were more prevalent in less controversial topics, leading to a lower accuracy in sentiment analysis for these topics.

4.2.3 Aspect-based Summarization

```
**Gameplay**

**Positive Sentiments:**

* Players appreciate the challenge of adapting to new gameplay elements, such as submarines.
* Some players enjoy the slow gameplay of battleships.
* Players appreciate the recent adjustments to AA, which make it more effective against aircraft.

**Negative Sentiments:**

* Players express frustration with gameplay elements that limit their ability to engage with enemies, such as submar
* Players find the interaction with submarines to be unenjoyable and not in line with the rest of the game.
* Players criticize the lack of counterplay options against aircraft carriers, particularly in terms of fighter plan
* Players dislike the passive playstyle adopted by some players in brawls, which they believe detracts from the enjo
* Players express concern about the impact of matchmaking on gameplay, particularly when facing higher-tier opponent
* Players criticize the lack of permanent sub detection for destroyers, making it difficult to hunt and kill submari
* Players express frustration with poor gameplay by destroyers, such as those that trade their ships for cap circles

**Suggestions and Requests:**

* Players suggest increasing the permanent sub detection range for destroyers.
* Players request the addition of various ASW launchers to the game, especially for destroyers.
* Players advocate for a nerf or removal of aircraft carriers due to their perceived lack of counterplay options.
* Players suggest that matchmaking should be improved to reduce the frequency of uptiering and downtiering.
* Players recommend that destroyers should be more vigilant for cruisers with sub surveillance and use hydrophone to
```

Figure 10. A sample summary for an identified aspect

In the final stage of data analysis, the system was able to generate summaries for the identified aspects. These summaries included prevailing positive and negative viewpoints, suggestions and requests, as well as any additional insights generated by Gemini. Figure 10 provides an example of an aspect summary generated by the system.

4.3 Result Visualization

In the result visualization procedure, the system was able to generate appropriate visualizations to present the required statistical data. These visualizations included a pie chart illustrating the frequency distribution of the identified aspects, as well as bar charts illustrating the sentiment ratios associated with each aspect.

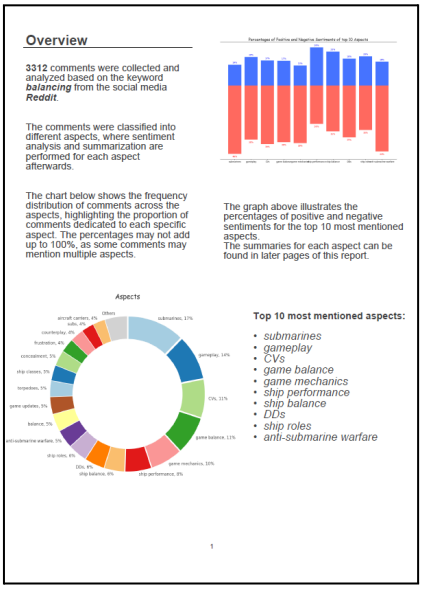

Figure 11. A sample overview page in a generated report

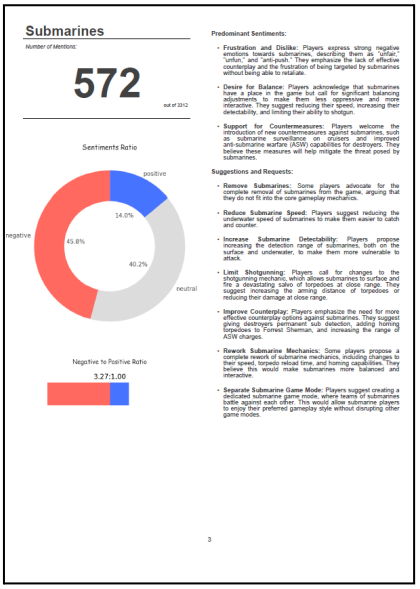

Figure 12. A sample summary page in a generated report

Subsequently, these visualizations were combined with other textual results from the data analysis procedure, creating a report in PDF format. A report consists of the following pages:

1. Cover page
2. Overview
3. Table of Contents
4. Summary pages for the 10 most mentioned aspects

5. Future Works

This section presents areas for improvement of the system in future work. Section 5.1 focuses on enhancing data collection; section 5.2 discusses improvements regarding the use of offline models; section 5.3 addresses enhancing the depth of analysis.

5.1 Improved Data Collection

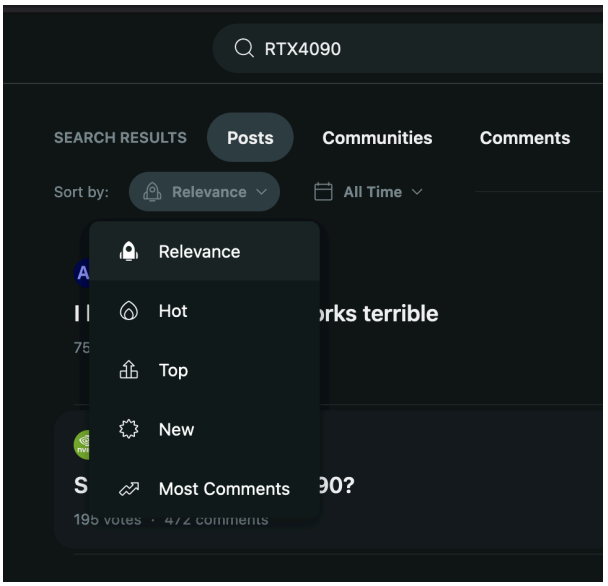

Figure 13. Posts sorted by "relevance" or "hot" in Reddit

In the current implementation, users are prompted to input keywords of interest, and optionally specify a subreddit from which the system collects data. If a specific subreddit is provided, the system collects data in the order of "hot", which is suitable for the data analysis procedure. However, when data is collected from all subreddits, it is ordered by "relevance". This ordering by relevance may result in collected data that could be outdated, which is undesirable for insightful data analysis.

To address this limitation, a potential enhancement would be to dynamically identify the most relevant subreddit based on the provided keywords. One approach is to analyze the top posts and identify the subreddit that appears most among these posts. Another approach is to utlilize a LLM and determine the most relevant subreddit. With this improvement, the system can ensure that the collected data remains both relevant and up to date.

5.2 Use of Offline Models

```
google.api_core.exceptions.InternalServerError:
500 An internal error has occurred. Please retry
 or report in https://developers.generativeai.go
ogle/guide/troubleshooting
```

Figure 14. An internal server error occurred when using Gemini

During the development of the system, the use of Gemini through its online services revealed instability, with frequent interruptions and timeouts. Figure 7 shows one such example, where Gemini's services experienced timeouts during peak usage hours. These interruptions raised significant concerns regarding the realiability and performance of the system.

To overcome these challenges, incorporating offline models is a logical solution. In addition to the freedom from connectivity issues, this approach also brings the benefit of avoiding potential future charges associated with online services. Moreover, offline models can be easily fine-tuned to better align with the system's requirements and offer improved performance. Furthermore, the synthesis of different models can be explored to further enhance the system's accuracy and efficiency.

5.3 Depth of analysis

	text	upvotes
0	Singles Inferno Season 3, \[Maybe spoilers\] ...	23
1	They aren't here for love that's for sure.	22
2	I'm pretty sure it's just for publicity.	7
3	The show and Netflix have gotten too big. I'm ...	5
4	He is rude he probably won't get much love fro...	15
...	...	...
2982	An Min-young - Single's Inferno 3 Ep1,	477
2983	[**Want to Become Moderator?**](https://www.re...	1
2984	Hot af	6
2985	Yum yum	2
2986	What will they do in this tv show?	2

2987 rows × 2 columns

Figure 15. Collecting textual content and rating of relevant comments

In the current implementation, only the textual content of the comments are collected and used for analysis. However, it is crucial to recognize that other information, such as the number of likes, can also be easily collected, as demonstrated in Figure 15. Incorporating this additional information is essential for uncovering more accurate sentiments. For instance, a comment receiving a substantial number of likes indicates widespread agreement with the expressed opinion. By incorporating these supplementary details, a more comprehensive and accurate analysis can be achieved.

6. Conclusion

The vast amount of user-generated content on social media offers organizations a valuable resource for collecting user feedback and public sentiment. However, the conventional manual approach can be impractical, and existing approaches to sentiment analysis have shown limitations, particularly in understanding human language accurately. Hence, this project aims to address these challenges, by developing a system to automate the process of social media research and to perform aspect-based sentiment analysis using state-of-the-art LLMs.

The system was completed with the utilization of Reddit and Gemini, and it was capable of great performances. The successful implementation of the system not only offers organizations with a valuable tool to conduct social media research and sentiment analysis, but also contributes to the broader understanding of LLMs and their capabilities.

VII. References

[1] Kapoor, K.K., Tamilmani, K., Rana, N.P. et al. Advances in Social Media Research: Past, Present and Future. Inf Syst Front 20, 531–558 (2018). https://doi.org/10.1007/s10796-017-9810-y

[2] M. Umair, S. Aun Irtaza and S. Salim, "User Feedback Severity Level Identification and Classification through Deeper Analysis of Text," 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2023, pp. 1-7, doi: 10.1109/iCoMET57998.2023.10099177.

[3] R. Shewale. (4 Mar. 2023). Social Media Users 2024 (Global Data & Statistics) [Online]. Available: https://www.demandsage.com/social-media-users. [Accessed: 30 Mar. 2024].

[4] A. Hassan and H. Korashy. “Sentiment analysis algorithms and applications: A survey,” Ain Shams Engineering Journal, vol. 5, pp. 1093-1113, Dec. 2014.

[5] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1

[6] W. X. Zhao, K. Zhou, and J. Li, “A Survey of Large Language Models,” 24 Nov. 2023. Available: https://arxiv.org/pdf/2303.12712.pdf

[7] B. Guthier, K. Ho and A. El Saddik, "Language-independent data set annotation for machine learning-based sentiment analysis," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 2017, pp. 2105-2110, doi: 10.1109/SMC.2017.8122930.

[8] Brandwatch. (n.d.). Brandwatch. [Online]. Available: https://www.brandwatch.com/ [Accessed: 20 Jan. 2024]

[9] Brand24 - #1 AI Social Listening Tool. (n.d.). Brand24 [Online]. Available: https://brand24.com/ [Accessed: 20 Jan. 2024]

[10] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” Knowledge and Information Systems, vol. 60, no. 2, pp. 617–663, Jul. 2018, doi: https://doi.org/10.1007/s10115-018-1236-4.

[11] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” Marketing Science, vol. 31, no. 3, pp. 521–543, May 2012, doi: https://doi.org/10.1287/mksc.1120.0713.

[12] S. Bubeck et al., “Sparks of Artificial General Intelligence: Early experiments with GPT-4,” Apr. 2023. Available: https://arxiv.org/pdf/2303.12712.pdf

[13] A. Nazir, Y. Rao, L. Wu and L. Sun. "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey," IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 845-863, Jun. 2022.

[14] W. Zhang, Y. Deng, B. Liu, S. Jialin Pan, and L. Bing, “Sentiment Analysis in the Era of Large Language Models: A Reality Check.” Available: https://arxiv.org/pdf/2305.15005.pdf

[15] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges,” arXiv:2203.01054 [cs], Nov. 2022, Available: https://arxiv.org/abs/2203.01054

[16] B. Liu, *Sentiment Analysis - Mining Opinions, Sentiments, and Emotions*, 2nd ed. Cambridge, 2020.

[17] P. Keung, Y. Lu, G. Szarvas, and N. A. Smith, “The Multilingual Amazon Reviews Corpus,” ACLWeb, Nov. 01, 2020. https://aclanthology.org/2020.emnlp-main.369/

[18] Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A High-Coverage Lexical Resource for Opinion Mining.

[19] K. Kenyon-Dean et al., “Sentiment Analysis: It’s Complicated!” Association for Computational Linguistics, 2018. Available: https://aclanthology.org/N18-1171.pdf

[20] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” Knowledge-Based Systems, vol. 226, no. 1, p. 107134, Aug. 2021, doi: https://doi.org/10.1016/j.knosys.2021.107134.

[21] B. Pang and L. Lee, “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.” Available: https://aclanthology.org/P04-1035.pdf

[22] A. Bello, S.-C. Ng, and M.-F. Leung, “A BERT Framework to Sentiment Analysis of Tweets,” Sensors, vol. 23, no. 1, p. 506, Jan. 2023, doi: https://doi.org/10.3390/s23010506.

[23] Author links open overlay panelJochen Hartmann a 1 et al., “More than a feeling: Accuracy and application of sentiment analysis,” International Journal of Research in Marketing, https://www.sciencedirect.com/science/article/pii/S0167811622000477?via%3Dihub (accessed 20 Apr. 2024).

[24] H. Wang, J. Li, H. Wu, E. Hovy, and Y. Sun, “Pre-Trained Language Models and Their Applications,” Engineering, Sep. 2022, doi: https://doi.org/10.1016/j.eng.2022.04.024.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv.org, Oct. 11, 2018. https://arxiv.org/abs/1810.04805

[26] Liu, Yinhan & Ott, Myle & Goyal, Naman & Du, Jingfei & Joshi, Mandar & Chen, Danqi & Levy, Omer & Lewis, Mike & Zettlemoyer, Luke & Stoyanov, Veselin. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

[27] Better language models and their implications. (14 Feb. 2019). OpenAI. [Online] Available: https://openai.com/research/better-language-models [Accessed: 20 Jan. 2024]

[28] G. Team and Google, “Gemini: A Family of Highly Capable Multimodal Models.” Available: https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf

[29] BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension

[30] C. Raffel et al., “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer,” arXiv.org, 2019. https://arxiv.org/abs/1910.10683

[31] White, Jules & Fu, Quchen & Hays, Sam & Sandborn, Michael & Olea, Carlos & Gilbert, Henry & Elnashar, Ashraf & Spencer-Smith, Jesse & Schmidt, Douglas. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.

[32] S. Yao et al., “Tree of Thoughts: Deliberate Problem Solving with Large Language Models,” arXiv.org, May 17, 2023. https://arxiv.org/abs/2305.10601

[33] Twitter API | Products | Twitter Developer Platform. (n.d.). X. [Online]. Available: https://developer.twitter.com/en/products/twitter-api. [Accessed: 20 Jan. 2024].

[34] Page Public Content Access. (n.d.). Meta. [Online]. Available: https://developers.facebook.com/docs/features-reference/page-public-content-access/. [Accessed: 20 Jan. 2024].

[35] YouTube Data API - Quota and Compliance Audits. (n.d.). Google. [Online]. Available: https://developers.google.com/youtube/v3/guides/quota_and_compliance_audits. [Accessed: 20 Jan. 2024].