

Taylor Swift Media Analysis: Focus and Sentiment in North America

Daniel Wassef 261009308

Jiahang Wang 261011319

Jiwon Chae 261018139

Yawen Guo 261048199

daniel.wassef@mail.mcgill.ca

jiahang.wang@mail.mcgill.ca

jiwon.chae@mail.mcgill.ca

yawen.guo2@mail.mcgill.ca

Introduction

In an era where media narratives can have a significant effect on public perception, understanding the different nuances of media coverage becomes critical, especially for prominent figures like Taylor Swift. This report presents an in-depth analysis of worldwide Icon, Taylor Swift's portrayal in North American news media. Our objective was to discern not only the sentiment of the coverage—whether positive, negative, or neutral—but also to identify the central themes prevalent in these narratives.

To achieve this, we carefully curated a dataset of around 500 news articles from various North American news outlets, selected through a thoughtful process ensuring a diverse range of sources. Utilizing the principles of open coding, we first categorized 200 articles into different topics based on their titles and opening segments, aiming to identify 3 to 8 significant themes.

Subsequently, the remaining articles were annotated with these defined topics from our code guide. Given the project's time constraints, we opted for the single annotation approach instead of the more robust double annotation method. Each topic was further analyzed by calculating the ten words with the highest Term Frequency-Inverse Document Frequency (TF-IDF) scores, providing insights into the unique characteristics of each category.

Finally, we conducted a second round of coding to assess the sentiment of the coverage. This step involved developing a defensible methodology for categorizing each article as positive, negative, or neutral towards Taylor Swift. This report details our findings, offering a nuanced understanding of how Taylor Swift's narrative is being shaped in the North American media landscape.

Our analysis of a large number of news articles indicates a predominance of neutral coverage, with a significant portion being positive. Swift's portrayal largely focuses on her career achievements, challenges, and reflections, which constitutes a notable part of the coverage. Her personal life and love interests also attract considerable attention, comprising a significant percentage of the articles. Additionally, the study uncovers a nuanced media narrative that varies from positive depictions of her career and cultural impact to more negative sentiments in reports of controversies.

Data

In the process of gathering news for our analysis, we decided against using news.api (new 2023). Our concern was that its requirement for users to specify a range of news agencies might lead to an incomplete collection and introduce potential bias. In search of a more suitable solution, we turned to the GNews.api (gne 2023). This alternative offered greater flexibility, allowing us to filter articles by language and country, a feature that aligned well with our goal of collecting a comprehensive set of North American news articles related to Taylor Swift.

We then assembled a dataset of 539 news articles, each featuring 'Taylor Swift' either in the title or the opening section. These articles were specifically sourced from the United States (US) and Canada (CA), focusing our analysis on North American media. Our collection period spanned from October 27 to November 27, ensuring that the data was both recent and relevant. Initially, we identified 996 articles, but to ensure a balanced representation and avoid data skewing, we limited the number of articles to a maximum of 20 per day by a random sample. This refinement process led us to the final count of 539. We believe this method of sampling will introduce the least amount of biases compared to the alternatives, like sampling first 20 articles of the day. For more information about biases of this study, kindly refer to the Results section of this report.

During data extraction, we utilized "GNews.api", which categorizes articles based on their country of publication. During this phase, we encountered a number of articles with URLs that suggested they were published outside of North America, primarily in Australia, Ireland, and Pakistan. However, upon closer inspection, we found that these articles were also published in North American magazines, explaining their inclusion as North American content by the API. We conducted a thorough verification of our dataset to ensure the absence of repeated articles, thereby maintaining the integrity and uniqueness of our collected data.

Methods

Open coding

For the development of topics, a crucial and often a lengthy journey, to ensure the typology we come up with is solid -

comprehensive, well-defined and objective as possible - we started with an open coding on 200 articles. This process involved reading the title and opening(or description) of each article, without delving into the entire content. Each article was assigned to exactly one topic, based on its predominant theme. This approach was chosen to maintain clarity, focus in our analysis, and facilitate future steps. Although the open coding part of our project was primarily exploratory, we aimed to categorize the articles into 3-8 distinct topics, trying as much as possible, to ensure that we have a comprehensive coverage of the different aspects of Taylor Swift's media portrayal, incorporating regular team discussions to aid in the process.

Annotation Process

After multiple discussions, we established the topics for our analysis (please refer to the Results section for the detailed typology). Moving forward, we continued our process by looking at the remaining articles in our dataset, and annotated the articles manually, adhering to the topics we defined. However, due to the time constraints, we only employed single annotation rather than the usual double annotation, with each team member being responsible for a subset of the articles. Nevertheless, to ensure that our annotation is as accurate as possible, we engaged in regular team discussions collaboratively reviewing the news articles to maintain consistency and objectivity in our annotations.

Sentiment Coding

Additionally, as we annotated the articles with the established topics, we added an additional step to our annotation process. A dedicated column was added for the classification of each article's sentiment as positive, negative, or neutral, facilitating further analysis.

Analysis

Following the completion of the first coding and later coding phases, we conducted a comprehensive analysis of approximately 532 news articles about Taylor Swift. These articles were classified into seven categories based on a typology we developed: Irrelevant, Ambiguous, Love Life, Cultural, Social and Commercial Impact, Career Achievements, Challenges, and Reflections, Personal life excluding love life and Controversy. Additionally, each news article was evaluated for its sentiment towards Taylor Swift, categorized as Neutral, Positive, Negative, or Irrelevant. News labelled as Ambiguous would be categorized as Irrelevant in the sentiment coding.

Quantification and Visualization of Categories

- We quantified the frequency of each category within our dataset and calculated their respective counts and proportions.
- Visualization techniques were employed to represent these distributions, providing an intuitive understanding of the categorical breakdown. As shown in Figure 1.

Sentiment Analysis and Visualization

- Similar like categories, we quantified the frequency of type of sentiments within our dataset and calculated their respective counts and proportions.
- This step provided insights into the prevailing sentiments about Taylor Swift in the media. Visualization techniques were employed to represent these distributions as shown in Figure 2

Data Preprocessing for TF-IDF Analysis

- Data preprocessing was initiated for the computation of the top 10 TF-IDF scores for each category.
- The preprocessing involved several crucial steps:
 - **Removing Punctuation and Numbers:** Simplifies the text, reducing noise and focusing on key words.
 - **Eliminating Common Stopwords:** Excludes frequently used but non-informative words, enabling a clearer analysis of significant terms.
 - **Excluding Taylor Swift's Name:** Removes the subject's name, which is expectedly frequent but non-revealing, to allow for a deeper exploration of thematic content.
- The stopwords list we use comes from (Yencken 2011)

Word Count Analysis

- We calculated the frequency of every word in the titles and descriptions of the articles, as well as a combined count.
- This step was crucial for preparing the data for TF-IDF analysis.

TF-IDF Scoring and Visualization

- Based on the frequency data, the top ten TF-IDF scores for each category across titles, descriptions, and a combination of both were computed.
- The TF-IDF (Term Frequency-Inverse Document Frequency) scores for each category were computed using the following formula:

$$\text{tf-idf}(w, \text{category}, \text{doc}) = \text{tf}(w, \text{category}) \times \text{idf}(w, \text{doc}) \quad (1)$$

Where:

$$\text{tf}(w, \text{category}) = \text{the number of times category uses the word } w \quad (2)$$

And:

$$\text{idf}(w, \text{doc}) = \log \left(\frac{\text{total number of categories}}{\text{number of categories that use the word } w} \right) \quad (3)$$

- These scores were then visualized to highlight the most significant words associated with each category.

Results

After our group discussion of typology based on result of open coding, we came up with the following coding guide that was used to perform later coding:

Motivation and context

We would like to study how Taylor Swift is covered in the media. We are especially concerned with North American coverage, and we specifically want to investigate on whether the coverage is positive or negative, and what topics are covered.

List of types in our typology with their definitions and examples:

- 1. Irrelevant:** This category covers the articles that are not related to Taylor Swift, although mentioning “Taylor Swift” in their title and/or description.
Example: In this news article’s title and opening, it talks about Canada’s small-town movie theatres’ current situation. Although “Taylor Swift” appeared in the title, the news itself is definitely not around Taylor Swift.
Title: “They’re going to lose us’: Barbenheimer and Taylor Swift alone can’t save Canada’s small-town movie theatres.”
Description: “Small-market cinemas are facing a conflagration of challenges that could force many to disappear.”
- 2. Ambiguous:** This category covers the news where although it is about Taylor Swift, the title and/or description is not expressive enough for us to come to a conclusion about its label/category.
Example: In this news, it is hard to tell if this news is about Taylor Swift’s love life, or irrelevant with Taylor Swift since in the description, Travis Kelce is clearly the main character of this news.
Title: “Travis Kelce apologizes to Taylor Swift’s dad”
Description: “Remember when Travis Kelce said he was going to stick to talking about football and not his private life? Good thing for the Swifties it looks like he didn’t mean it.”
- 3. Love Life:** This category covers the articles about her romantic relationships. This category may be confused with “Personal life excluding love life” category, even though it is possible to combine these two types, and get one called Personal life, we felt that from the point of view of stakeholders, her love life often gets more attention from consumers, therefore, we decided to keep them separate and study the coverage separately.
Example: This news talks about Taylor Swift’s love life, specifically focusing on Travis Kelce watching her concert.
Title: “Travis Kelce Is Taylor Swift’s Biggest Fan at Her Argentina Concert.”
Description: “Travis Kelce was spotted watching Taylor Swift perform in Buenos Aires during the international leg of her sold-out Eras tour, a day after they had dinner in the city.”
Edge case: We feel like this category may be confused with Career, so we would like to provide a corner case to illustrate this.
Title: “Taylor Swift brands ex-boyfriend ‘lying traitor’ in new song - but fans left wondering who she’s singing about.”
This seemed to be in the Career Achievements, Challenges, and Reflections category at first, but since this covers, among her ex-boyfriends, who this new song is about, we concluded that this is in the Love Life category.
- 4. Cultural, Social and Commercial Impact:** This category covers the articles about the effect Taylor Swift has had on society.
Example: This news illustrates Taylor Swift’s social impact, specifically targeting on political influences.
Title: “Taylor Swift’s Argentina concert turns political as presidential election nears.”
Description: “Fans waiting to watch the singer’s concert read posters encouraging them not to vote for one of the candidates.”
- 5. Career Achievements, Challenges, and Reflections:** This category covers the articles about both her music and non music related career activities, excluding her social and commercial impact.
Example: The below news talks about Taylor Swift’s music sales, which highlights her career achievement.
Title: “Taylor Swift’s 1989 Re-Do Tops Original’s First-Week Sales”
Description: “A 1989 (Taylor’s Version) sold 1.65 million copies across formats in pop star’s biggest debut ever”
- 6. Controversy:** This category covers the articles about scandals. This includes public disagreements or debates marked by differing sentiments and strong emotions. The fundamental criteria to be labeled as controversy is whether the news is controversial or not. We feel this category is of particular interest to media companies, since this is what would draw consumers’ attention a lot.
Example: This news shows controversy around Taylor Swift.
Title: “Taylor Swift Under Attack by MAGA Christians”
Description: “A self-proclaimed prophet and Donald Trump fan claims the pop star has been doing the devil’s work, along with Pfizer and rumored boyfriend Travis Kelce.”
- 7. Personal life excluding love life:** This category covers the articles about her non-romantic relationships, personal purchases, etc.
Example: Below news describes Taylor Swift’s friendship with Brittany Mahomes.
Title: “Taylor Swift’s New Friend Brittany Mahomes Wants to Play Cupid for Sophie Turner”
Description: “Mahomes and Turner’s friendship began in September when Swift invited them to dinner in

New York City before a Chiefs game.”

Sentiment:

1. **Positive:** News is considered positive if it portrays Taylor Swift in a favorable manner. This could include her achievements, successes, positive personal developments etc.
Example: The below news discusses Taylor Swift’s career success.
Title: “Taylor Swift’s 1.6M Bow for ’1989 (Taylor’s Version)’ : A Personal Best”
Description: “Taylor Swift has debuted at No. 1 on the Billboard 200 with her best first-week numbers ever, topping her previous personal best with ’Midnights.’”
2. **Negative:** Negative news delivers unfavorable news about her. This could include for example, controversies, criticisms, or issues etc.
Example: The example below covers an accident happened at her concert.
Title: “Taylor Swift postpones Brazil show after fan dies in extreme heat”
Description: “Taylor Swift canceled the second of three concerts in Rio de Janeiro a day after a 23-year-old fan died amid sweltering heat.”
3. **Neutral:** Neutral news is based on an absence of overtly positive or negative sentiments. It is unbiased and does not lean towards either praise or criticism. Sometimes it simply delivers the information
Example: This news, although relevant around Taylor Swift, does not give a sentiment about her.
Title: “We’ve Been Talking About How Taylor Swift Changed Her Karma Lyrics For Travis Kelce, But There Was Actually A Second Big Eras Tour Moment She Created For Him”
Description: “While Taylor Swift changing the lyrics for “Karma” has gone viral, she also made another big gesture for Travis Kelce during that Eras Tour show.”
4. **Irrelevant:** News unrelated to Taylor Swift, their sentiments are categorized under this class.
Example: This news is not relevant with Taylor Swift, although mentioning her name.
Title: “Ohio Woman Turns 399-Pound Pumpkin Into Taylor Swift for Halloween”
Description: “Jeanette Paras has painted pumpkins in the likeness of Kanye West, Donald Trump, and even Baby Yoda before. This year, it’s Taylor Swift’s turn.”

Topic characterization

For the topic characterization, we computed the 10 words with the highest tf-idf scores in title and description for each category. We can say that these 10 words are characteristics of each category.

- Career Achievements, Challenges, and Reflections : version, album, billboard, awards, dates, music, grammy, vancouver, artist, nominations
- Personal life : brittany, york, nyc, mahomes, dinner, instagram, wags, photos, sends, legs

- Controversy : death, died, rio, janeiro, heat, dies, ana, benevides, clara, concert
- Irrelevant : halloween, broncos, dressed, sexiest, alive, denver, play, movie, harry, win
- Love Life : relationship, romance, kelce, travis, argentina, plans, boyfriend, thanksgiving, eagles, chiefs
- Ambiguous : biden, britney, spears, white, house, birthday, turkey, president, event, book
- Cultural, Social and Commercial Impact : christ, law, shop, iconic, deals, black, tickets, camping, months, inspired

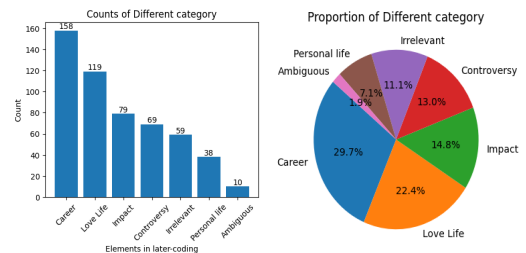


Figure 1: Counts and Proportion of Different Categories

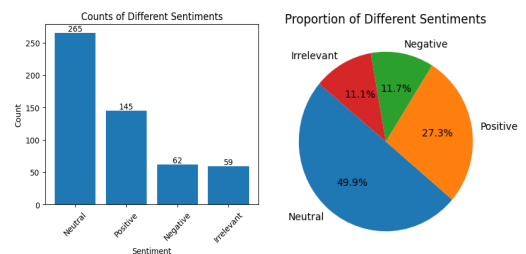


Figure 2: Counts and Proportion of Different Sentiments

Topic engagement

Articles were mostly neutral (49.9 %), and 27 % of them were positive. Only 11 % were negative. We didn’t mark whether it is positive, negative, or neutral for irrelevant articles.

Articles concerning Career Achievements, Challenges, and Reflections were the most as 29.7%. Articles about her love life were the next (22.4%). Next, articles regarding her impact on society, culture, and fashion counted for 14.8%. Articles about controversial issues and irrelevant articles counted for 13% and 11.1%, respectively. Articles about her personal life appeared as 7.1%. And 1.9% of the articles were ambiguous.

Comprehensiveness

The typology we made was well-defined, so in our later coding, there was no difficulty of classifying them into each category. Therefore, our typology is comprehensive. We believe our definitions and examples also made the code guide containing minimal subjectivity.

Discussion

Categories

- **Career Achievements, Challenges, and Reflections (158 mentions):** The prominence of words like “version” and “album” aligns with Swift’s music releases, while “billboard,” “awards,” and “nominations” emphasize her commercial success and industry recognition.
- **Love Life (119 mentions):** Frequent words “relationship” and “romance” highlight the media’s attention on her intimate relationships, a common theme in celebrity reporting.
- **Cultural, Social and Commercial Impact (79 mentions):** With words like “Christ” and “iconic,” the coverage seems to delve into her broader cultural influence, beyond just entertainment.
- **Controversy (69 mentions):** Words like “death” and “concert” relate to specific, possibly sensational events, showing media’s focus on the more dramatic aspects of her life.
- **Personal Life (59 mentions):** Lower frequency terms such as “New York” and “dinner” suggest selective interest in her personal activities and locations, possibly reflecting a balance between her public image and privacy.
- **Irrelevant (38 mentions):** The use of terms such as “Halloween” and “movie” indicates sporadic coverage where Swift’s name might be incidentally mentioned, rather than being the core focus.
- **Ambiguous (10 mentions):** Fewer mentions with diverse terms like “Biden” and “White House” suggest occasional associations with political events or figures, albeit tangentially.

From the findings of each category listed above, the data suggests a media portrayal that is heavily skewed towards Swift’s career achievements, with a notable interest in her personal life and relationships, punctuated by coverage of controversies and her cultural impact.

Aslo, as shown in fig. 3, The “Career” category shows a balanced sentiment with a slight lean towards positive. “Controversy” has a notable negative sentiment. “Love Life” and “Personal Life” show more neutral and positive sentiments, suggesting a less controversial portrayal in these areas. “Impact” has a positive sentiment, indicating favorable coverage on her influence. “Irrelevant” has very little negative sentiment. “Ambiguous” shows a mix, with neutral sentiment being the most prominent.

This suggests a nuanced media portrayal with a tendency towards positive coverage in career and impact, while controversies are reported with more negativity.

Impact of Cultural and Social Factors

- **American culture and fashion:** Halloween is an American culture. And since Halloween (Oct 31) was included in the period when we collected data, there

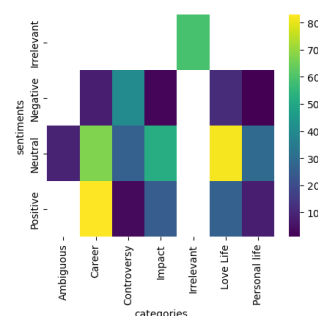


Figure 3: Relation Between Sentiments and Categories

were many articles that mentioned Taylor Swift’s Halloween costume. Many other celebrities dressed like her inspired by her costume. However, since these articles didn’t cover her other aspects deep, we classified them into ‘irrelevant’ category. The public might perceive her as a fashion icon influenced by them. (halloween, dressed in Irrelevant category)

- **Gender roles:** Women have traditionally had a role of raising children, taking care of a house, and it has been a men’s role to earn money. However, Taylor Swift is a professional pop star, and even has recently become a billionaire. Even she has a net worth more than her boyfriend, so in the articles covering her romance, her boyfriend’s wealth or status was not mentioned. Just their romance or dates were covered. This could influence people to focus on their pure romantic relationship. (relationship, romance, kelce, travis, boyfriend in Love Life category)
- **Security:** At her recent concert in Brazil, one fan died because of the extreme heat. Security is an important factor in society, but her fan died at the concert where security must be guaranteed. These articles were mostly marked as negative in our open coding and later coding. People might think that it is her responsibility that she didn’t take care of a security of a concert hall. (death, died, rio, janeiro, heat, dies, concert in Controversy category)
- **Seeking experience and youth culture:** Young generation value a shared experience that they can enjoy together even if they have to spend a lot of time and money to attend a concert. Many fans and young people have been waiting for even few months sharing tents and camping to get a Taylor Swift concert ticket. This shows how much influence Taylor Swift’s music and concert have on youth and musical culture. These articles were mostly positive on her and influence the public’s perception on Taylor Swift’s fandom and live performances at her concert. The public would perceive her fandom supportive, devoted and passionate. They could think that her live performances worth to attend and special to attract many people. (tickets, camping, months in Cultural, Social and Commercial Impact category)

Bias, Challenges and Limitations

- **Bias:** Since we limit our news for the period of one month, specifically, our data consists of news published between October 27 and November 27, which is exactly the period of time when Taylor Swift released a new album. We see that "album" is one of the most common words in our tf-idf analysis of career as well. Therefore, among all news discussing her career related activities, we see that most of them is around this activity. Due to the time frame, in case of her career, we are biasing towards her album release, and against other aspects of career. This may not be an issue if the media company wants to study Taylor Swift's coverage specifically for this period of time, but may give biased result if the goal is to study Taylor Swift's coverage in general. Therefore, we would like to point this out. In addition, in our data collection phase, since we were using G-News API, we relied on the attribute of country to filter out news published in Canada or US, therefore, we are biasing towards these two countries, and excluding news from other countries of North America. We are fully aware of these biases, however, we believe that this still allows this study to deliver meaningful insights, since we tried our best to collect data with the least amount of problematic biases.
- **Challenges:** We initially faced challenges in data collection phase. Our primarily tool was News API, however there wasn't a straightforward way to filter out news published in North America, since the only attribute in news API around region is the publisher. We first filtered out all news in the specified time period, then extracted all news publishers, did some research to try to correctly filter out all the North American news media outlets, and finally filtered all news with "Taylor Swift" that was published by those North American publishers. We encountered several issues: some news did not contain "Taylor Swift" in their title nor their opening, we re-checked our algorithm and did not find a bug in it. We checked several articles, and discovered that the reason was that some articles mentioned "Taylor Swift" in their bodies but not in the title nor the opening. However, the API would still return us those entries. We also noticed the issue of redundant news entries in our collected data. After a second filtering of all redundant articles, we were left with a smaller number of articles that was a lot less than 500. Faced with these challenges, we decided to switch to another API. One of our teammates discovered G-news API, which has an attribute for the country of publication. We filtered all news that was published in Canada or the US, and this is how we ended up gathering our data.
We definitely had a difficult time deciding our typology because of the different opinions within the group about which categories should be in the code guide. To be more specific, we were particularly discussing the possible overlap between categories. Luckily, we were

able to agree on the final code guide that was presented above, and we worked hard on defining each category with as much as details as possible, so that there is least amount of confusion in between.

- **Limitations:** This study has several limitations that should be considered when interpreting its findings. Firstly, studying around 500 news might not be as comprehensive as we want for our study. Additionally, to get a more accurate report about the Taylor Swift's coverage in the North American media, it may be more beneficial to not limit ourselves to only the articles of the past month. Finally, this study only takes into consideration news articles, and not other forms of media like TV programs, podcasts, TikTok videos, X posts, etc. Encompassing those other forms of media, might offer a better understanding of the coverage of Taylor Swift.

Sentiment Tendencies

The sentiment analysis we carried out in the previous section, indicates a predominant neutrality in the media coverage about Taylor Swift, with 49.9% of the articles maintaining a neutral tone. This suggests that the approach taken by North American media outlets to report on Taylor Swift, focuses more on facts and events without an explicit positive or negative spin. However, we can see a presence of 27% positive and 11% negative articles, possibly pointing to a divide in public perception, reflected through the media. The positive media coverage likely resonates with her successes and her public image as a beloved artist by many. On the contrary, the negative sentiment could be attributed to controversies, personal life scrutiny, and the critical eye often cast on celebrities whose lives are very public. Due to our bias, we could assume that Taylor Swift is usually covered less negatively in the media which was heavily impacted by the unfortunate event that took place last month, where one of her fans died during one of her concerts in Brazil.

Deciphering the Narrative

Thanks to the thematic breakdown of our typology, we can very clearly see the key areas of focus in Swift's media coverage. The most prominent themes revolve around her 'Career Achievements, Challenges and Reflections', which constitutes the biggest proportion of the articles, with 29.7%. This confirms the public's recognition of Taylor Swift as one of the most influential figures in the music industry. Articles about her personal life (excluding her love life) are less prevalent with only 7.1%, which contrasts with the proportion of articles about her love life (22.4%), suggesting a fascination by the public with her personal life, always eager to know about her new lover since Taylor Swift's music primarily focuses on her love life, and romances.

The 'Cultural, Social and Commercial Impact' theme underscores the link between Taylor Swift's career and broader societal issues. It shows her influence that goes

beyond the music industry, to certain aspects like social trends.

Interestingly, we have a considerable amount of articles that we deemed as "Irrelevant" or "Ambiguous". This implies that there isn't always a thematic focus, and that some media outlets utilize Swift's name in the media as a draw for readership, irrespective of the content's relevance.

Group Member Contributions

We all equally contributed to the project.

- **Daniel:** Data collection, Open coding, Discussion of Code guide, Later coding, Analysis, Writing of Report.
- **Jiahang:** Data collection, Open coding, Discussion of Code guide, Later coding, Analysis, Writing of Report, Main coder.
- **Jiwon:** Data collection, Open coding, Discussion of Code guide, Later coding, Analysis, Writing of Report.
- **Yawen:** Data collection, Open coding, Discussion of Code guide, Later coding, Analysis, Writing of Report.

References

2023. GNews API. <https://gnews.io/>. Accessed: 2023-12-05.
2023. News API. <https://newsapi.org/>. Accessed: 2023-12-05.
- Yencken, L. 2011. Stopwords List. <https://gist.github.com/larsyencken/1440509/raw/53273c6c202b35ef00194d06751d8ef630e53df2/stopwords.txt>. Accessed: 2023-12-10.