

Opinion Mining Unleashed: Advancing Public Comment Summarization with NLP

Divya Karthik, Co-Author: Jonathan Li, Co-Author: Daiguang Zhang

Abstract

In the digital age, the prevalence of public comments across online platforms presents both a rich resource for insights and a challenge due to the sheer volume and diversity of opinions. This complexity often overwhelms traditional approaches to text analysis, necessitating more refined strategies for processing and understanding user-generated content. Our study proposes an advanced methodology for preprocessing text data tailored to the unique characteristics of various online platforms, including forums, reviews, social media, and more.

Our approach emphasizes the customization of preprocessing steps to enhance the relevance and accuracy of subsequent Natural Language Processing (NLP) tasks, such as sentiment analysis, topic modeling, and text summarization. By addressing specific preprocessing needs—such as acronym expansion in social media, sentiment-laden language in reviews, and technical terminology in forums—we aim to improve the quality of data fed into NLP models, thereby enriching the insights derived from them.

Crucially, this research diverges from conventional methods by not merely seeking to neutralize biases or eliminate redundancy. Instead, it acknowledges the inherent value in the diverse perspectives and opinions expressed by users, organizing and processing them in a way that illuminates rather than obscures the richness of public discourse. Through a detailed examination of preprocessing techniques across different types of content, we facilitate a more nuanced understanding and navigation of public sentiment, enhancing user experience and decision-making in online environments.

Keywords

Comments, Summarization, Pre-processing, topic modeling, Sentiment Analysis

1. Introduction

The preprocessing of text constitutes a fundamental step towards achieving effective outcomes in text summarization tasks. While the conversion of text into a suitable file format represents the most basic form of preprocessing, achieving optimal summarization results typically necessitates a more nuanced approach. Essential preprocessing actions include the removal of whitespace, punctuation, and standardization of language, which serve as the foundational steps for cleaning and preparing text data for analysis (Widyassari et al. 2023).

In the realm of narrative texts such as stories, articles, and instructions, both extractive and abstractive summarization models have demonstrated proficiency. These models excel

largely because the source materials are coherent, sequentially organized, and centered around a singular thematic or narrative thread (Patel and Tabrizi 2022).

However, the domain of comment summarization introduces distinct challenges, primarily due to the nature of the input corpus comprising a collection of disparate statements. This fragmentation poses significant hurdles for extractive summarization approaches, which rely on identifying and compiling key segments of text. Similarly, abstractive models, which generate summaries by interpreting and paraphrasing the essence of the text (Cheung 2008), often struggle to capture the multifaceted sentiments expressed across diverse comments accurately.

The diversity of comments across platforms like Reddit, Quora, YouTube, AllRecipes, Amazon, and Yelp introduces unique challenges to text summarization. Tutorial and instructional content on platforms such as AllRecipes and YouTube often features a corpus dominated by repetitive comments expressing extreme satisfaction or dissatisfaction. Specifically, recipe comments frequently include outliers due to modifications and tweaks, alongside gratitude expressions that, while positive, may not directly contribute to summarization goals aimed at capturing core content insights.

For platforms focused on product and service reviews, such as Amazon and Yelp, comments provide valuable benchmarking opportunities. Amazon's use of filters and star ratings, for instance, offers predefined topics that serve as benchmarks for refining summarization techniques. Similarly, Yelp utilizes business categories and ratings as essential benchmarks, while YouTube employs the like-to-dislike ratio and video tags to indicate community engagement and preferences. These metrics guide the development of effective summarization models by highlighting key aspects of user feedback and interaction.

Moreover, platforms with a broad content spectrum, like Reddit and YouTube, often feature comments that lack focus, covering a wide range of topics with little relevance to one another. This includes jokes, external references, and personal anecdotes. In contrast, comments on the New York Times may reference historical figures and events, with latecomers to discussions introducing outlier opinions. Such variability demands a nuanced approach to text analysis and preprocessing, requiring strategies that can accommodate the wide-ranging and often diffuse nature of public discourse.

The ability to succinctly summarize comments in a manner that accurately reflects public sentiment holds profound implications. Such summarization not only facilitates a deeper understanding of collective perspectives but also empowers decision-making processes. From everyday choices, such as selecting a dining venue, to critical financial decisions like stock investments (Dunn 2023), the organized summarization of public comments can unveil insights and highlight considerations that may otherwise remain unexplored. In this context, readers increasingly depend on the aggregation of public discourse to inform their decisions, underscoring the importance of advanced summarization techniques that can navigate the complexities of comment-based texts effectively.

2. Related Works

Comment datasets are crucial for sentiment analysis, offering profound insights into human sentiments, especially through review-type comments. These insights are often quantified by star ratings, which serve as a vital tool for corroborating the findings of NLP sentiment analysis. The goal of review summarization is to determine whether a product or service consistently satisfies or disappoints, thereby classifying the feedback as positive or negative.

The introduction of review summarization by Yelp and Amazon highlights the increasing importance of effective comment summarization techniques (Mehta, 2024). Despite the progress made, these platform-specific approaches, especially Yelp's, are inclined to present businesses in a favorable light, and the operational costs associated with sustaining large language models (LLMs) are substantial (Musser, 2023).

The challenges presented by LLMs in summarizing comment datasets accentuate the need for advanced summarization methods, especially in the realm of public discourse. Compared to reviews, social media commentary is more complex due to its diverse nature and the wide range of topics discussed. Social media comments not only critique but also spark various conversations, reflecting the public's myriad viewpoints on issues like governmental impact or health, thus viewing content from multiple perspectives.

Efforts to refine social media comment summarization have included diverse approaches. A notable study on Reddit comments employed fine-tuned T5 models to assess the consistency between user-generated TLDR (too long; didn't read) summaries and the actual content of comment threads (Blekanov 2022). The study found that many comment pools yielded summaries that failed to accurately reflect the discussions, highlighting the difficulty of creating a one-size-fits-all summarization model that fulfills the varied demands of the marketplace and scholarly research.

Research focusing on YouTube coding tutorial comments pointed out the challenge of processing comments laden with irrelevant terms, a prevalent issue in social media data analysis (Poché et al. 2017). This difficulty is compounded by the dynamic and varied language prevalent on platforms such as YouTube, Twitter, and Facebook, characterized by the use of abbreviations, emoticons, phonetic spellings, and other forms of internet vernacular.

3. Methods

3.1 Datasets

To understand the complexities of proposing for comment summarization, a diverse dataset was needed. In our study, we incorporated a diverse range of datasets, encompassing platforms such as Reddit, New York Times, AllRecipes, Yelp, Amazon, and YouTube.

For datasets for generic comment-study purposes we exported data from Kaggle, with at least 1 million comments spanning over several content pieces. For experimentation and testing of our methods, we sourced comments for different articles, videos, and recipes by scraping the above platforms with API calls. To ensure standard and good testing, we only worked with data pieces over 100+ comments.

For our initial pre-processing, we concatenated comments into a single text file separated by line. For our topic extraction dataset, we made the text lowercase, stripped first person pronouns, removed stopwords and html tags, discarded non-english comments, and performed tokenization. We also removed comments that were less than 2 words or solely emojis. For our summarization model, we preserved the original sentences with the exception of comments that were less than 2 words or solely emojis.

3.2 Feature Extraction

Previous research has made it evident that for recommender systems, effective justification should include relevant and distinguishing traits of the items (Musto, 2019). We apply this idea in our analysis of the domains and identify consistent themes within comments across various content pieces within the same domain. Our investigation into the corpus features employs POS tagging to delineate its structural components effectively. Concurrently, we implement a word frequency model that encompasses unigrams to trigrams, with the frequency threshold meticulously calibrated according to the dataset's volume. The word frequency model identifies the most recurrent unigrams, bigrams, and trigrams across the datasets. However, not all n-grams hold equal value in semantic analysis (Kapusta 2021). By applying POS tagging to these n-grams, we attempt to filter out less informative grammatical structures, such as those predominantly consisting of conjunctions or prepositions, focusing instead on noun phrases and verb phrases that carry substantive information about the topics of discussion.

Table 1: Generic themes extracted from large comment datasets for each domain

Amazon	service, place, time, food, staff, experience, location, room
Yelp	price, quality, performance, packaging,
All Recipes	Substitute, time, oil, flavor, salt, ill, heat, family, leftover
Reddit	people, time, someone, bot, moderators
Youtube	video, game, think, want, need, love, work, world, life
New York Times	President, tax, world, country, money, government, America

Topic Modeling stood out as exceptionally effective in the Amazon and Yelp datasets, consistently uncovering an extensive selection of proper nouns and nouns. This finding emphasizes the value of Part-of-Speech (POS) tagging within our analysis, especially given the evaluative and critical nature of the "comments" sections on these platforms.

Leveraging these insights, we propose the introduction of a 'relevance' score for each topic extracted from a set of comments related to a specific piece of content. This score prioritizes topics that align with universally identified themes, assigning them higher relevance. The rationale for combining relevance with frequency is to compensate for the inability to integrate disparate elements into a unified overview of discourse (Tsai et al., 2020). This limitation often arises from insufficient discussion on certain aspects. By adopting this approach, we aim to produce summaries that, although potentially less comprehensive, more accurately reflect topics the public deems significant for discussion.

Table 2: Topic extraction for content based on relevance of generic themes and word frequency

Yelp: South Point Hotel	Place, hotel, hotel room, staff, everything, strip. Resort casino, hotel rate, buffet, mandalay bay hotel, pool, vegas, hotel room elevator, vegas hotel, hotel price, hotel guest, star hotel, food, hotel casino, south point hotel, big room, price, reservation desk, time, lot, hotel registration, casino, room, bar
YouTube: Easy origami dinosaur tutorial	Step, tutorial, origami, split, dino, foil color origami, origami claw, clip, origami artist, paper, origami paper, version, dinosaur
New York Times: Senate Republicans	Nuclear option, democrat, filibuster, even filibuster, garland, gop, government, justice, majority, McConnell,
Amazon: Red Heart Super Saver Yarn	Craft, simple easy craft, red heart yarn, quilting, price, local craft store, skein, amazon, rainbow art project, crochet project, crochet lesson, color, art blanket, afghan color, art, brand, kid's project, learner, patrick's day craft project, cmas decorating, product store

3.3 Enhanced Emoji Classification in Diverse Online Environments

The classification of emojis is a critical component in our advanced text preprocessing methodology, particularly due to their widespread use in digital communication. This process is essential for extracting deeper insights into sentiment, tone, and context from text data, a crucial aspect for accurate interpretation across diverse online platforms. Our enhanced emoji classification approach is particularly applicable in environments where textual communication is predominant, such as social media, customer feedback, chatbot interactions, and online forums. These platforms often feature a blend of text and emoji-based communication, offering a rich tapestry of human emotions and expressions that traditional text analysis methods might not fully capture.

Our research introduces a sophisticated emoji classification strategy, integrating a direct replacement mechanism within a carefully curated one-to-one semantic mapping archive. This method consists of two key components:

One-to-One Semantic Mapping Archive: Central to our methodology is a detailed archive where each emoji is precisely linked to a specific lexeme or phrase that captures its fundamental semantic value. This archive, developed through comprehensive curation, ensures each emoji aligns with its most universally recognized meaning, adapting to the nuances of different online platforms.

Semantic Replacement Algorithm: In practical application, our algorithm methodically scans text data across various platforms for emojis. Upon detection, it employs the semantic mapping archive to replace the emoji with its textual equivalent. This nuanced replacement strategy is informed by current research (Bai et al., 2019; Ferrari, 2023), emphasizing the complex linguistic behaviors and communicative functions of emojis in digital environments.

This methodology is particularly effective in contexts requiring clarity and depth in communication analysis, such as in interpreting customer feedback or monitoring social media interactions. By converting emojis to their direct textual interpretations, we preserve the integrity and enhance the clarity of the underlying sentiment and thematic content in messages, even when emojis are absent. This process not only simplifies data preprocessing for various NLP tasks but also adapts to the specific complexities introduced by emojis in different types of online communications.

Addressing the use of emojis in contexts like sarcasm, our method retains its effectiveness. Sarcasm often relies on the interplay between apparent meaning and actual intent. Here, emojis can either accentuate the sarcastic tone or signal a departure from literal interpretation. Our approach, through direct emoji replacement, provides clarity in these instances. For example, an emoji like ☹️, which typically indicates displeasure or skepticism, is mapped to its direct meaning to highlight a sarcastic tone in the message. This is particularly valuable in datasets from platforms like social media, where sarcasm is frequent but subtle.

Furthermore, our method sets a foundational interpretation for emojis, upon which additional context-specific analyses can be layered. This is vital in cases where the text and accompanying emoji might present a contradiction or irony, as often seen in sarcastic exchanges. The one-to-one mapping thus serves as a basis for more sophisticated semantic and sentiment analyses, enabling a deeper understanding of complex communication nuances like sarcasm in digital discourse.

Our enhanced emoji classification approach, aligned with the broader aims of our research, offers a refined and platform-specific understanding of emoji usage in digital communication. This method not only simplifies the initial interpretive process but also provides a framework for advanced NLP techniques to address the intricacies of digital communication across diverse online platforms.

Table 3: Emoji Descriptions Mappings used for comment text processing.

☹️	(Rolling On The Floor Laughing)
😐	(Unamused Face)
😡	(Angry Face)
🙄	(Face With Rolling Eyes)
😞	(Pensive Face)
😍	(Smiling Face With Heart-Eyes)
🤔	(Thinking Face)

3.4 Acronym Expansion

Acronym usage is becoming more and more common in emails, tweets, blog posts, etc. And with the increasing popularity of mobile devices, the use of acronyms on social platforms has increased even more because typing in these devices is difficult and acronyms provide a succinct way to express information. (Thakker et al. 2017). Understanding their full meanings is essential for accurate data analysis and interpretation. This is particularly true in diverse online environments, where acronyms vary widely in usage and interpretation. Our Acronym Expansion Methodology is specifically designed to address this challenge, enhancing the clarity and usability of text data across various digital platforms.

Our approach commences with the creation of an Acronym Identification and Mapping Library. This extensive library, developed through a combination of online resource aggregation and user engagement across social media platforms, maps common acronyms to their full forms. This dual-source method ensures a comprehensive collection of acronyms, from everyday language to specialized jargon. Notably, this library includes both widely recognized acronyms like "LOL" (Laugh Out Loud) and more specific ones such as "NASA" (National Aeronautics and Space Administration), reflecting the diversity of digital communication.

The text processing algorithm is a component of our methodology. It systematically identifies acronyms within text data using the mapping library and replaces them with their expanded forms. This process is vital for contextual clarity, as demonstrated by the transformation of a phrase like "I can't believe I missed the AMA session" into "I can't believe I missed the Ask Me Anything session." This kind of expansion is crucial for enhancing the readability and interpretability of online communication, particularly in data-driven analyses.

The acronym expansion methodology we propose is an integral part of our advanced text preprocessing framework, tailored for the complexities of digital communication. By efficiently decoding acronyms, we enhance the overall quality of data for our NLP tasks, aiding in a more nuanced understanding of digital discourse. This methodology not only streamlines data processing for various analytical purposes but also ensures that the richness and diversity of online communication are preserved and accurately interpreted, aligning with the overarching objectives of our research.

Table 4: Reddit Acronyms Database for translation

OP	Original Poster
TL;DR	Too long; Didn't read
DAE	Does Anyone Else
IIRC	In case I remember correctly
FTFY	Fixed that for you
AMA	Ask me Anything

3.5 Sentiment Analysis

In the digital landscape, where public commentary on online platforms is both a rich data source and a complex challenge, the extraction and understanding of diverse opinions are crucial. Traditional text analysis methods often struggle with the volume and polarity of these comments.

Our study, therefore, adopts a sophisticated approach, blending sentiment analysis with advanced NLP models to derive more meaningful insights from this data.

Our initial step involves the use of fuzzy-lattice methodologies, as highlighted by Jelodar et al. (2020), to identify the sentiments associated with different topics in public commentary. However, understanding that mere sentiment analysis might not capture the entire semantic depth of the discourse, we focus on preserving the full semantic context of each sentence. This is especially important in cases where public comments show polar opposite views, requiring a nuanced understanding that goes beyond basic sentiment labels.

To enhance the processing and summarization of this complex data, we have incorporated sentiment ranking. By identifying the sentiment behind each comment, we can order them based on their sentiment, ranging from positive to negative. This ordering is not just a means of organization but a strategic step to prepare the data for feeding into our BERT summarization model. The BERT model, renowned for its deep learning capabilities in understanding the context of words in a sentence, can thus produce more accurate and representative summaries of public discourse when working with sentiment-ranked data.

In the context of rapid sentiment analysis in our methodology, we leverage the capabilities of libraries like SpaCy and TextBlob. SpaCy, known for its efficiency and accuracy in linguistic annotation, is used for quick sentiment extraction from large volumes of text. TextBlob complements this by providing a simple API for common NLP tasks, including sentiment analysis, which allows for a quick assessment of the emotional tone of comments. The integration of these tools enables us to categorize comments into sentiment classes rapidly, which is crucial for real-time data processing.

This approach is informed by the advancements in NLP tools and techniques. For instance, the efficiency of libraries like SpaCy in text processing is well-documented in research by (Honnibal and Montani 2017), who discuss its design and implementation for fast and accurate linguistic annotation. Similarly, the application and effectiveness of TextBlob in sentiment analysis are highlighted in research by (Loria et al. 2023) presenting it as a valuable tool for quick and straightforward NLP tasks.

The use of these advanced, yet fast-processing tools ensures that our methodology not only captures the sentiment of public commentary accurately but also does so efficiently, catering to the demands of real-time data analysis prevalent in platforms like social media and customer feedback systems.

The significance of our approach is particularly evident when applied to datasets that feature a wide array of opinions, such as social media discussions, online forums, or customer reviews. In these environments, opinions are not only diverse but often deeply polarized, making the task of summarization and understanding challenging. Our approach, by combining sentiment analysis with BERT's contextually aware summarization, ensures that the summaries generated are not just aggregations of predominant sentiments but are reflective of the entire spectrum of opinions present.

3.6 Final Processing Pipeline

The order of our final processing pipeline is meticulously designed for optimal efficiency and effectiveness in text summarization. It begins with standard preprocessing (removal of punctuation, formatting normalization), a crucial foundational step that ensures uniformity and readability of the text. This uniformity is essential for the more sophisticated processes that follow.

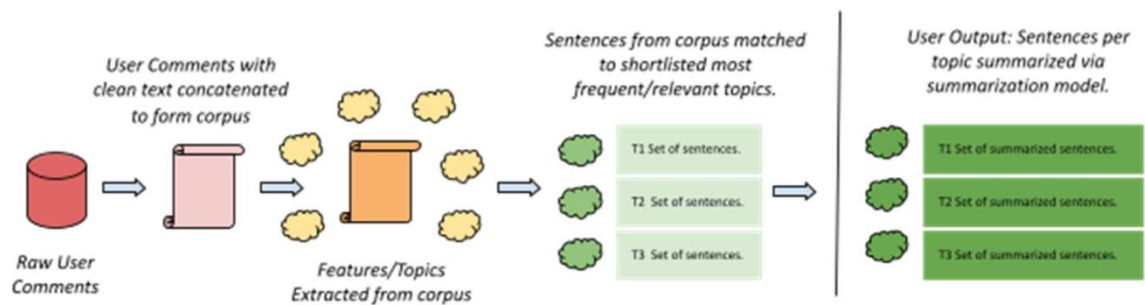
The integration of emoji preprocessing and acronym expansion, as detailed in previous sections, comes next. This sequence is deliberate; emojis and acronyms are integral parts of modern digital communication and need to be interpreted accurately before any further analysis. By converting emojis to their semantic equivalents and expanding acronyms, we bring clarity and context to the text, essential for accurate thematic and sentiment analysis.

Topic modeling then categorizes the data into coherent themes, providing a structured framework that is vital for understanding the content's main ideas. This is followed by

sentiment ranking, which orders comments by their emotional tone, from positive to negative. This step is crucial for assessing the general sentiment of the text, which influences the final summarization.

Finally, this well-prepped data is fed into a BERT model for extractive summarization. BERT utilizes the structured and sentiment-ranked data to produce contextually rich and emotionally representative summaries. This ordered approach, from basic standardization to complex sentiment analysis and summarization, ensures that our pipeline is not only logical but also highly effective in capturing the intricacies of public discourse in digital communication.

Figure 1: Processing Pipeline



4. Results and Evaluation

4.1 Experiment Results

Table 5: Summaries of comments from the BERT Summarizer model from 6 public comment domains

Amazon	<p>The design of the earbud may cause pain after prolonged usage. In terms of sound this is way more advanced than basshead 103. Voice is so suppressing that you won't be able to enjoy the music properly. Youtube play pause, next previous video buttons wont go away when earphones are plugged in. I also wanted an alternative to in ear buds that are available in the market. However, I don't know what kind of a design boat is bragging of. Why even buy this if you cannot use it properly Volume buttons are not there. Fits just right, could hurt if you went to sleep wearing them and turned your head to the side as it's plastic. I had to manually lower the bass just to listen to normal yt videos and play game. Sturdy wires and the wire from jack to mic is braided and has a good texture and feel. Obviously they can't be use to hear music if you are hardcore music listener you won't find that good in this case. Its always the same problem no matter what brand you buy, one ear piece would stop working in less than a year. Go for it guys and fall in love with the old school most comfortable earphone style ever. Good and comfortable design, especially for whom bud type is uncomfortable. After the button, the matrrial is normal rubber and chances of break is there. However since its a wired earphone I won t recommend it for workouts.</p> <p>...</p> <p>Guys,Pls don't buy this product Automatically sound increasing and decreasing No controlling button at mic. Pretty good for the price Bad sound quality Getting some noise distubance and issues with pause button. Good product Wrost product I like base Good Broken and not a proper sound Wire quality is best. Amazing earphone i just loved it Disturbance coming from the mic. You wont feel much of a difference in Music, but while listening to voices, you may feel a bit uncomfortable. Mic and Sound quality is good. Bad service from Amazon and Boat. Delivery was in time and quality is also great with service they offer. But most earphones don't deviate too much from the mean, as regards that. other headphone are operating well with my mb however this product does not work with it. I bought this as a temporary replacement till I get back my sound magic premium earphones from service center. If you use In Ear earphones, your ears will get hurt in few hours. So good for an in ear earphone at this price and also the sound quality is perfect for what they had advertised. we can use this for talking purpose and listening peaceful music. Noise cancellation is good. I suggest buy it now Product is good ,but automaticly musical player and volume is getting on. good product and value for money Awesome and strongest earphone (*Parts redacted for brevity)</p>
--------	---

AllRecipes	<p>This soup didn't remind me of a cheeseburger in any way shape or form. The addendum is I use an immersion blender to make the soup creamy. Well, all I can say is look at the number of reviews this recipe has - you're outvoted LOL! Sour cream is not something that I keep on hand, so I rarely add it in. COULD CAULIFLOWER BE SUBSTITUTED FOR POTATOES TO MAKE MORE KETO FRIENDLY? Next time I'll use a full pound of beef and maybe cut back on the potatoes some. Also sauteed diced potatoes in separate pan before adding to soup. A sharp Cheddar in lieu of the Velveeta might be an improvement. I also added some shredded cheese as the Velveeta is a little bland for this Wisconsinite. The pickles made the soup taste like a cheeseburger in a bowl. We substitute real cheese for velveeta, and that just increases the flavor! We have been making this recipe for a few years now, but we make it a tad bit different. My daughter (who hates cheeseburgers) loved it and said that she wants it again and again. I did use a full pound of ground beef but didn't change anything else. It takes a little time to prepare, but it is so very worth every minute! I was really excited to make this soup after reading all the great reviews. I have to agree with the one review I read that said it was basically Hamburger Helper without noodles. I made a few changes because my husband likes a heartier soup. The only thing different was I used cheddar cheese and substituted noodles for the potatoes. Turns out great every time, no matter what type of cheese I use. Made this for our soup supper at church and everyone loved it! I switched velveeta for 1 cup sharp cheddar, 1 cup smoked cheddar, 1 cup smoked Gouda, and 1 cup gruyere. I don't know how the velveeta version tastes but this came out amazing! Well, it was -50c here in Canada and today felt like a soup day! This is the first time I've garnished it with pickle slices. I made this a couple of weeks ago minus the celery and it was wonderful... even the husband liked it. One person asked if anyone has used anything other than Velveeta. My husband, who is not a soup fan, raves over this recipe, and he and our 12-year-old will play fight for the leftovers! I always keep an extra pouch or box of mashed potatoes for this. I double the amount of cheese, onions, and hamburger... Adding the given 3/4 t of salt is crazy!!!! If there are any leftovers, which is rare, I put it in individual servings in the freezer for a quick lunch. I thought the basil was different, but it is excellent in the soup. I have been making this soup since it was first published in 1996. I used a full pound of ground beef and no Velveeta...cheddar cheese and 1/4 c. cream cheese instead. Looks great but unhealthy as all get out The best cheeseburger soup ever! I have to admit though, I did take a shortcut to make this a meal I could whip up after work. I substituted the prep work and cooking veggies with two cans of Veg-All. I've had another cheeseburger soup recipe that had them and it was DELICIOUS! It has everything in it my husband loves, and he requests it over and over. Sharp cheddar cheese with 8oz Velveeta instead of all 16oz Velveeta. Adding a bit of cream in with the milk gave it a nice, rich feel. and add one pint milk, check the salt and add pepper if you like. people that made this and say it has no flavor, put a bouillon cube in there. Of ground beef, 2 cartons of chicken broth, 1 cup of milk, and half cup of heavy cream. My kids are already asking when I'm going to make it again. half instead of milk (that's what I had on hand) used real cheese and did not use the sour cream. I had not made it before, but my family says that this needs to be a staple in our meal rotations! I made this soup but added 1 cup of cooked bacon and 1 Tablespoon bacon drippings. I will probably make it again and maybe add more ground beef. I doubled the meat and added a little crumbled bacon.</p>
23	<p><i>International Journal of Management, IT and Engineering</i> http://www.ijmra.us, Email: editorijmie@gmail.com</p>

NYT

Whistleblowers afraid to formally complain may come forward after January 20th. I shall ask our ten grandchildren and five adult children to read every word. But so much needless pain and suffering already has been inflicted by this awful person. May Mr. Trump be prosecuted only for those violations that are abundantly clear. ☹️ The precedent was cast by G.W. Bush when he pardoned Casper Weinberger in the Iran Contra scandal. the crime of giving aid or comfort to the enemies of one's government. Keep it simple... If all men are created equal, then Justice must be doled out equally as each person deserves. It is imperative he is held accountable and those who enabled him . Wasn't Oliver Wendell Holmes who spoke that the foundation of the law is vengeance? Or, as has been said by Groucho Marx and others, "Time wounds all heels." The last thing our country needs is for this man to be free and to have access to any further political influence. He already has a host of indictable offences for which he remains unaccountable, shielded as he is a sitting president. He is guilty of manslaughter, womanslaughter and childslaughter. The mainstream media including NYT have written extensively about this. No, any opportunity to figuratively decapitate this incipient authoritarian movement needs to be taken ASAP. If allowed to just get a pass, Trump, will serve as a template for further criminal presidencies. We the People are the Sovereign Authority behind the Constitution. I shudder to think... We can't perpetuate a sense of no consequences. The messaging lately has been, "if you're going to be a criminal, be a billion dollar criminal. We would only grease the skids to tyranny if Trump is not prosecuted and, if found guilty, not imprisoned. They're not going anywhere whether Trump is charged or left to hang out at Mar-a-lago. Is there really any reason that 57 years later documents related to the Kennedy assassination are still classified? Including potentially classified materials that may directly impact foreign policy? Or-- (2)--is the monarch himself PROTECTED by those very laws? he shouted (as his appalled ministers dropped to their knees. He should be prosecuted to demonstrate that in a democracy we live by the rule of law. But he cannot get a free pass, otherwise the rule of law breaks down. Have a press conference and detail what's in the indictments. Trump fired that National Security guy today because he stated the elections were fraud free. Or does the entire population shrug its collective shoulders and "move on." Folks, we can spend the next four years trying to prosecute Mr Trump for obvious crimes while in office. Alternatively, forget Trump - let the NY prosecutors feast on him! If we wish to preserve and protect our republic, a full reset, as mentioned in this article, looks like our only bet. International cases suggest that what is needed is a Truth & Reconciliation Commission. Prosecution of Trump is criminalization of policy differences? This article covers many of Trump's alleged crimes, reminding us that his potential legal liability is "yuge." (The only question is which prosecutors will lead and which will follow. It would also likely include any argument that he lied or encouraged anyone else to lie to Congress. Trump is not above the law so he must be prosecuted for the heinous crimes he committed as president. ...

This is supposedly a country based on the rule of law, that no one is above the law. Do we really want to empower those who would rather have a fascist state? They have willingly and knowingly done unprecedented damage to our democracy. They are very close to being Soviet communists but not yet that much like them. Trump and his enablers may find smelly wiggle room today, but after January 20 Trump is a private citizen. The

Reddit	<p>Honestly pretty happy with this - getting a prime living space should be a reward in and of itself. What an improvement in housing policy from the previous minister (Lawrence Wong). Chances are it should, otherwise a lot of civil serpents will hentak kaki for this. We cannot allow people to privatise profits while socialising costs using public housing. This is terrible and it is in the national interest for a higher fertility rate. Many Singaporeans don't leave their nest till they are in their 30s or even 40s. Even the kids who move out early and rent a room don't completely solve this problem. Thank you Desmond Lee Only gripe is that people with non-nuclear families are not eligible to purchase these flats. Public housing rules that are focused on providing housing not an investment vehicle! You're right but I'm not too worried about rental income's effects. Most of the price increase we see is largely owed to people wanting to buy to stay, not rent out. And less wealthy people can access the location they can't last time with this new rule, so the law is still good. The price rise won't affect the less wealthy the government is helping but only affect the wealthy. And the flat will become worthless when the lease is up, so it is just the rich holding each other's bag. Last I check Public Housing is and shouldn't be the same as how we approach Private Properties. Wouldn't this mean surrounding landlords benefit due to tenants having lesser rental options? The ground sentiment seems to be ready for this after the many public engagements they did recently though. That's not supposed to happen because not all new BTOs will be PLH. i think it makes sense to prioritise people who wants to build a family Or get a studio apartment. > what ppl has been asking for is disallowing selling on open market, only can sell back to HDB. many on /r/singapore already kpkb so much about being left out from economic growth of Singapore. Which lets lower to median income people to be able to enjoy the good locations without breaking banks. 14k per household, so must include your wife's salary if she is working Yup agreed. Singles who BTO can only purchase 2-room flexi flats in non-mature estates according to the current policies. They cannot buy 3 or 4-rooms to begin with, unless it is from the resale market. But, how does it make sense banning singles outright from even buying a small resale flat in "prime location". IMO, it only limits the pool of buyers but that doesn't solve the issue of overly priced resale flats. Because it's not a 10 year MOP for all new BTO, only for PLH BTO and that would be about 1 a year. End of day everyone is holding onto depreciating assets that will zero out after 99y. [what i know is there rarely are any policies that will satisfy everyone's agenda so yea, my sympathies. Couples have the best chance to achieve this as they can procreate. Married couples have higher chance of procreating than non-married couples. deleted] Some people here will very soon come out to say that these measures are not harsh enough. That's precisely why most American billionaires don't pay tax, they just borrow against their shares. I am referring specifically to how only married people are able to purchase these upcoming PLH flats at all. I also think some things should be done to make housing more affordable for singles. But ceteris paribus I think just opening the prime estate resales to singles is also not the solution. lmao bro/sis, you might as well say if the buyer has \$1mil in cash the income ceiling doesn't matter. ya so it basically won't reduce prices for the vast majority HDB buyers? Once they own a home they want their home to appreciate in value. Engineering is pretty mundane, where you just defy the rules of physics. It's so much more... creative and there's so much free real estate to mess around with. There are no limitations regarding tenure of loan or interest rate. PE MSR can be depressed. This necessarily means that a seller trying to sell the flat for a million will need to reduce prices.</p>
25	<p><i>International Journal of Management, IT and Engineering</i> http://www.ijmra.us Email: editorijmie@gmail.com</p>

Yelp

Soooo much hype about the oolong soft serve and boba topping, but of course we got drinks instead. I ordered the brown sugar milk tea, which was kind of bitter because I ordered black tea with it (my mistake though). I like how the tea flavor was prominent in flavor but not too overpowering. Aside from tea soft serves, they also serve tea lattes and ice cream drinks. They have a rotation of seasonal items so be sure to check their Instagram for updates. I'm surprised I haven't written a review yet considering how much I love this place. It's not like boba milk, you're really getting milk tea with these drinks. I usually get the oolong soft serve because it's a light but you can definitely taste the tea flavor. They have cute wall decorations like a neon sign and some comic wall picture. Definitely try the oolong soft serve and try one of their drinks when you come here. It was super refreshing and a fun take on bubble tea flavors! Let me start by saying I grew up in an area of Los Angeles that is known for everything Asian - SGV. You can feel the tea flavor than the milk unlike other milk teas. They all sound good, but I always end up trying their soft serve. he CoCo here also called it bubbles) for 50 cents, which were nice and chewy -- a pleasant textural component. I was worried that it would have too much of a tea taste to it, but it was delicious. I could taste the oolong really well bc the milk wasn't very overpowering at all. I am a HUGE lover of bubble tea and let me tell you I was in heaven at Bar Pa Tea. The space itself is a cute modern/minimal almost french themed kind of set up. I honestly don't know what else I can say to get you to try this stuff. This place was surprisingly packed on Sunday night around 7pm. The large bubbles are standard size bubbles we typically get, and the mini ones are just smaller size. Next time I go, I definitely want to try the oolong tea soft serve. Too bad I'm not from NYC or else I would definitely be come back here frequently. The oolong tea is very fragrant, and the pairing of bubbles is soo delicious. I was able to sample both flavors of ice cream before I made the choice. You also got to choose from small bubbles or big bubbles and if you wanted it to be on top or bottom. And I have to say, it's quite delicious if you're a fan of oolong milk tea. I went here on a Thursday after work at around 6 and there was no wait. Once we ordered we waited for another 5 minutes before we got our soft serve. How is that I haven't really seen any other places that have done this? It's about \$4.25 for a cup of the soft serve and then 50 cents for the extra toppings. The flavor was really good, it reminded me of jasmine tea flavor ice cream from afters. Overall, good tasting soft serve, not unreasonably priced compared to the other fad desserts that have been popping up. First time trying bubble tea flavored soft serve and it was amazing . I got the Oolong ice cream with bobas and oreos toppings and the Oolong Tea Latte. The tea flavored ice cream was really yummy since it had a strong tea flavor and sweet. However, the tea latte was somewhat bland for my taste and not as good. I was late for a meeting and told her I wanted to cancel the order as I had to go. Found this place randomly though watching the FoodInsider video. I really liked my ice cream and was digging the black tea flavor more than the oolong tea flavor. I can honestly say, their soft serve is one of the best I've ever had! Obviously, I will be back for more and now they have a rewards program. I came across Bar Pa Tea as it was literally across the street from sweetgreens! It does take a bit before you get your drink, but that's because they make your drink super fresh. Few places in NY have the option of serving bubble tea with small tapioca pearls and I was so happy to find them here. Of course, they still offer the standard regular sized tapioca, but the mini pearls are where it's at! It probably looks more impressive than it tastes, but it was fine, I enjoyed the drink anyway. I also tried the oolong soft serve and it had a sweet but light oolong tea flavor. I wouldn't go out of my way for this, but it was still a really nice treat on a hot day. Was always wanting

YouTube	<p>Like I said in the video, subscribe if you haven't already and you could win \$10,000! Would love to know more about the team that designed and built them! So much work and detail went into making this happen ☐ This is such a better version than the Netflix game show. How wonderful you are, Mr. We hope to participate with you Solid high quality content Jimmy & team. This dude is a genius Can't believe he actually spent 3.5 million dollars for giving us 25 mins of pure entertainment. This is literally better than the new series haha This man is literally born to make History. Hatsoff to Jimmy and team its amazing that Mr. Beast did it better then Netflix's "Squid Game: The fact that Jimmy is able to spend and give so much money to his videos is astounding. This aged well ♥ Your videos always bring a creative and fresh perspective. The fact that he got all these people and then recreated this show was just so good. Props to you Jimmy just give jimmy his oscar he really deserves it☐☐☐ This was SENSATIONAL!!! Mr beast never fails to make good content hace 2 años salio esta joya he really did the series setup 102 was a legend. Love the glass bridge set Way better than the netflix one☐ Rewatching this while watching squid game: Lets give a huge round of applause for jimmy who spent millions of dollars on it ☐ his videos are so great☐ I've lowkey, never thought Jimmy would millions of dollars to give us entertainment. ♥the way he gets 3 men to play musical chairs ☐☐Insane how realistic these sets are, honestly hats off to you Jimmy This has more views than the actual squid game. Why didn't they use Mr Beast's method for dalgona game in Squid Game : Hats off to all the people to pull off this realistic set, especially to Jimmy! That guy who won is honestly awesome Omg did you guys dub the voice in multiple languages including Korean? When Jimmy is dedicated to something literally nothing can stop him! What a joke of a season 2 on Netflix His videos are so entertaining! Impossible, this was INSANE Jimmy is the most hard working person to just make his fans entertained. When you realize that Mr Beast makes something better than Netflix. Congrats to the winner and congrats to Jimmy for reaching this level of success. Where has the time gone ☐ just finished it, enjoyed every single second. ☐ It's the fact that hes generous enough to create these amazing concepts and sets, for RANDOM people. Hats off, i am completely astounded by this Very well recreated. We need a follow up with what the winner did with the money!! Jimmy needs his own show where he gives people money or challenge them. Just finished it, enjoyed every single second. Come on, we are halfway to a billion! The 330,000,000\$ was the best one for sure☐ The joy in my body when 368 went out near the end. The show was insanely good and reliving the games this way is epic. The views this has received in such a short time is INSANE lol. Hats off to you and the set designers, you all did an amazing job pulling this off. ☐☐☐ It's absolutely insane to me how accurate the sets are man, well worth the wait! Also crazy how far 456 was able to make it considering he's the "main character" and it wasn't scripted. Season 2 ♥huge props on still making the marbles game a betrayel/twist scenario. It's absolutely insane to me how accurate the sets are man, well worth the wait!</p>
---------	--

Table 6: Summaries from the pipeline processing method from 6 public comment domains

	Topics Extracted	Topic Summaries
Amazon	<p>really good brand sound wire quality medium music ear Price boat calls meeting sound effect value mic earphone condition sound quality headphone package regular use base wire bass</p>	<p>Mic (-) The mic is not useful for attending calls. It's such a disappointment they advertise this headphones have mic and doesn't even work it's such a basic thing. Mic does not work properly, you have to keep it very close to your mouth so that the other person can hear you. Only phone mic is to be used</p> <p>Quality (+) Quality as per boat brand is best sturdy. must buy if you still prefer wired headphones not expensive with good quality. So far the quality of wires is good and doesn't get tangled up audio quality is good too. Very good quality in this price and the most important is bass treble is really very good.</p> <p>Ear (-) Your ear will start painting after sometime not worth a penny or even a half star. your ears will get hurt in a few hours. it's sorta painful in my ears only while adjusting. It will cause pain in ears for using like 3 4 hrs sound quality is good pain in ears.</p>

AllRecipes	<p>family favorite time sour cream potato soup sodium content unsalted butter velveeta regular chicken broth flavor taste added salt ingredient preparation hearty bread cornbread recipe</p>	<p>Potatoes (+) Smashed some of the potatoes with a potato masher to thicken. Potatoes make it a very filling soup. Cauliflower can be substituted for potatoes to make it more keto friendly.</p> <p>Potatoes (-) Who puts potatoes and carrots on their hamburger. The potatoes just didn't seem to belong. Leave the potatoes out next time. Hamburgers do not have potatoes in them.</p> <p>Velveeta (+/-) Has anyone used another cheese other than velveeta? A whole pound of velveeta? I'd cut that to about 1/4 of that. Velveeta is for those who have no idea what real cheese.. Used cheddar instead of velveeta. Used shredded cheddar and shredded mexican cheese instead of velveeta.</p> <p>Sodium Content (-) The recipe was a bit high in sodium. For those concerned about the sodium content you can use homemade chicken broth. Added salt at the end of the recipe in order to keep sodium to a minimum.</p>
------------	---	---

<p>NYT</p>	<p>Justice supporter imperial presidency trump bidens conumdrum power trumps behavior president court nation legal transgression global community tradition charge independent justice department baptism prosecution current future president democracy legitimate position complete lawlessness broadcast channel media type crime action strong legislation consequence money country law office division america biden time houses judiciary committee</p>	<p>Justice (+/-) Treason, however manifested, should be prosecuted, as should conspiracy to obstruct justice. He [the president] and anyone else suspected of crimes must be investigated and, where evidence is sufficient, prosecuted. Let justice be done though the heavens fall; otherwise, there is no justice. May justice be served; you are the punchline of jokes, the ultimate bad example." "If all men are created equal, then justice must be doled out equally as each person deserves. There is no civilization without the rule of law; healing can't take place until justice is served.</p> <p>Country (-) Right now talk of healing sounds soothing, but it's no answer for a country caught between cult-crazed adulation and conflicted cop-out. This country will not remain a democracy if the most flagrant and lawless would-be autocrat in its history is not indicted and prosecuted for all the crimes he may have committed. Our country is on a precipice. Either prosecute him to the fullest extent of the law and all his minions including Barr and Pompeo, or be prepared for another dictator to come along in 2024. The country needs healing. The country will not heal if Trump is not held accountable. If sufficient evidence is found, he should be indicted; that goes for his whole grifter family.</p>
------------	---	--

Reddit	<p>family married couple 10 year people time single policy case best chance couple house singaporean population age flat bto single male precursor couple hdb housing company significant precursor electorate child child bearing singapore agenda application poor relationship single hdb buyer news indication sex change price hdb</p>	<p>Hdb (housing development board) The difference between being rich and living in an HDB and being rich and wanting to buy an HDB to live in is significant. Just let people sell or rent whenever they want, but 90% of the difference between the selling price and the subsidized price goes back to HDB if renting; 90% of the profits go to HDB too, making HDB more affordable. What people say they want can be a lot different from what they actually want; singles should be glad they can even get an HDB, but this doesn't screw over people who use HDB as a retirement asset like a compulsory buyback scheme would.</p> <p>BTO (build to order) Many exploit the BTO system for additional income, renting out their BTO while living with parents. Singles face restrictions in the BTO market, limited to 2-room flexi flats in non-mature estates unless they enter the resale market. Complaints arise from married couples waiting for their BTO; there's no simple solution to satisfy everyone. There's confusion about the effectiveness of queuing for a BTO given the long wait times and potential financial gains.</p> <p>Price Strict regulations may disrupt prices, lowering speculation-driven high prices but not those driven by high demand, possibly allowing transactions at higher prices due to the "lottery effect." Incremental policy adjustments suggested to prevent future price surges due to market shortages, highlighting the complexity of balancing supply, demand, and affordability. Experts argue new policies don't make public housing more affordable or address societal divides, emphasizing the challenge of managing high prices through demand reduction. Restrictions on resale eligibility and high pricing strategies may naturally limit the buyer pool, potentially stabilizing or lowering prices for eligible buyers.</p>
31	<p style="text-align: right;"><i>International Journal of Management, IT and Engineering</i> http://www.ijmra.us, Email: editorijmie@gmail.com</p>	

Yelp	place price line staff credit card bubble tea typical swirl tea milk tea tea flavor milk tea boba store oolong topping soft serve texture swirl bar pa tea black tea bubble tea store tapioca boba refreshing drink super hot day mini bubble flavor ice cream great option better milk teas place caramel soft swirl oolong mixed soft serve boba	<p>Line (+) There was no line and the staff was friendly best boba in NYC. There wasn't much of a line Sunday afternoon. The line moved quick enough since it doesn't take too long since there's a lot of staff. There is never a line which is awesome because there are seats open. No line on a weekday around 4 pm. And there was no line.</p> <p>Line(-) The line was out the door. It takes a while to get to the front of the line. Expect to wait in line for a bit of time. If there's a line it'll feel pretty crowded. The lines are extremely long and there is no space at all to sit down. The line moves quickly though the tea does take longer to make because it's made from scratch.</p> <p>Oolong (+) The oolong soft serve ice cream there's basically two flavors to try here: oolong and black tea. The soft serve with boba is a must try at this place. Love the soft serve tea flavor with boba concept. Avail flavors are oolong and black tea. The ice cream is good and with the bubbles is perfect. The oolong tasted simple like unsweetened tea and brown sugar... it was an amazing blend would definitely come back. Definitely come here and try their oolong soft serve with bubble tea, the oolong is pretty strong.</p>
------	---	---

Youtube	budget set video game time content set designer money mr beast jimmy squid games	set designer(+) The sets are unbelievable. Huge props to the set designers. The sets in this are incredible, would love to know more about the team that designed and built them. Game (+) Love this game. It was so much fun to watch. The show was insanely good and reliving the games this way is epic Reliving the games this way this version of the game is pretty much what the contestants of the show were probably expecting Jimmy(+) Congrats to the winner and congrats to Jimmy for reaching this level of success. Crazy to think how far your hard work has brought you to keep it up. Jimmy is able to spend and give so much money to his videos. It is astounding gotta give some respect. This man is literally born to make history hats off to Jimmy and the team. jimmy needs his own show where he gives people money or challenge them this is absolutely so awesome
---------	--	--

4.2 Evaluation

In our study, we evaluated summarization performance using a content piece from each domain from the BERT-extractive-summarizer library to those generated by our custom comment summarizer pipeline. Our assessment, through a focus group survey of the Facebook Davis Computer Science Club and Slack EEC193B Group (25 respondents), revealed a strong preference (23 out of 25) for the pipeline method over the BERT model, especially when pre-processing was applied. Participants appreciated the topic-organized summaries for highlighting key discussion points. Notably, the pipeline's ability to clarify acronyms in Reddit posts was commended for enhancing context comprehension. Additionally, the preference for topic-filtered summaries over lengthy paragraph formats was evident. A major challenge identified with direct comment data input into summarizers was the lack of structured organization, causing users to navigate through randomly presented topics.

Although our topic-based summarization approach received positive feedback, users suggested improvements due to repetitive sentences in the extractive summaries. Common

feedback highlighted that sentences with similar meanings were not combined effectively, pointing to a limitation in summarizer models' ability to understand context and dynamically organize information on a specific topic. This feedback underscores the need for further development in summarization technologies to more efficiently condense and synthesize information. This problem could potentially be addressed by leveraging advanced Natural Language Processing (NLP) text clustering methods. Techniques such as Latent Dirichlet Allocation (LDA) and K-means clustering could be explored to enhance the ability of summarization models to identify and group similar sentences or topics. By employing these methods, it may be possible to reduce redundancy and create more succinct, context-aware summaries that better capture the essence of the text in a coherent manner.

5. Conclusion

In this comprehensive research paper, we rigorously define and explore the complexities of summarizing topic-driven reader comments across varied internet platforms. Our primary focus was to understand and articulate the unique challenges involved in processing and interpreting the multifaceted and dynamic nature of public comments from diverse sources like Reddit, the New York Times, AllRecipes, Yelp, Amazon, and YouTube. We developed and utilized a specially processed dataset, a novel contribution in the field, as a foundation for extensive experimentation with different text summarization models, particularly BERT.

Throughout our study, we delved deep into the subtleties of comment corpuses, acknowledging the distinctiveness of each platform and the varied interactions of users. This exploration was not just about identifying the problems in comment summarization but also about innovating and proposing effective solutions. Our experimental phase, where we applied advanced text summarization models to our unique dataset, was crucial in validating the effectiveness of our methodologies.

This research paper significantly underscores the necessity of tailored approaches to manage the intricacies of public comments across diverse digital landscapes. As the digital world continually transforms, the methodologies and strategies we've proposed set a solid groundwork for future explorations and enhancements in the field of comment summarization. Our endeavors aim to push the boundaries of understanding in natural language processing, specifically in the challenging arena of topic-driven reader comment summarization, contributing significantly to ongoing dialogues and advancements in the domain.

6. Acknowledgments

Thank you to the people who contributed responses to our survey and provided us valuable insights and feedback about the applications of our text summarization by topic project idea. A big thank you to TA Weitai Qian, who guided us through aspects of the project and written reports. Also, a big thank you to Professor Venkatesh Akella for the help and guidance on areas of the project to focus on.

We also thank Google Collaboration Notebooks for providing us GPUs on which to run our text summarization models and thank UC Davis for bringing together the essential resources and people to our project.

7. References

Aditya Thakker, Suhail Barot, and Sudhir Bagul. 2017. Acronym Expansion: A General Approach Using Deep Learning. In Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017), pages 322–327, Kolkata, India. NLP Association of India.

Bai, Q., Dan, Q., Mu, Z., & Yang, M. (2019). A Systematic Review of Emoji: Current Research and Future Perspectives. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02221>

Blekanov, Ivan S., Nikita Tarasov, and Svetlana S. Bodrunova. 2022. "Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages" *Future Internet* 14, no. 3: 69. <https://doi.org/10.3390/fi14030069>

Cheung, J.C. (2008). Comparing Abstractive and Extractive Summarization of Evaluative Text: Controversiality and Content Selection.

Dunn, W. (2023, July 15). WallStreetBets founder Jamie Rogozinski: "Disruptive finance". *New Statesman*. <https://www.newstatesman.com/the-weekend-interview/2023/07/wallstreetbets-founder-jamie-rogozinski-disruptive-finance>

Ferrari, Enrique. (2023). The study of emoji linguistic behaviour: an examination of the theses raised (and not raised) in the academic literature. *Communication & Society*. 36. 115-128. 10.15581/003.36.2.115-128.

Honnibal, M., & Montani, I. (2017). SpaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. <https://spacy.io/usage/spacy-101>

H. Jelodar, Y. Wang, M. Rabbani, S. B. B. Ahmadi, L. Boukela, R. Zhao, and R. S. A. Larik, "A NLP framework based on meaningful latent-topic detection and sentiment

analysis via fuzzy lattice reasoning on YouTube comments,” *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 4155–4181, 2020.

Kapusta J, Drlik M, Munk M. Using of n-grams from morphological tags for fake news classification. *PeerJ Comput Sci.* 2021 Jul 19;7:e624. doi: 10.7717/peerj-cs.624. PMID: 34395862; PMCID: PMC8323729.

Loria, S. (n.d.). *TextBlob: Simplified Text Processing*. <https://textblob.readthedocs.io/en/dev/> 2023.

Mehta, Ivan. “Yelp Launches Revamped Feed with Ai-Powered Business Summaries.” *TechCrunch*, 30 Jan. 2024, [tcrn.ch/3HEFHDC](https://techcrunch.com/2024/01/30/yelp-launches-revamped-feed-with-ai-powered-business-summaries/).

C. Musto, G. Rossiello, M. D. Gemmis, P. Lops, and G. Semeraro, “Combining text summarization and aspect-based sentiment analysis of users reviews to justify recommendations,” *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019

Musser, Micah. (2023). *A Cost Analysis of Generative Language Models and Influence Operations*. .

E. Poché, N. Jha, G. Williams, J. Staten, M. Vesper and A. Mahmoud, "Analyzing User Comments on YouTube Coding Tutorial Videos," 2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC), Buenos Aires, Argentina, 2017, pp. 196-206, doi: 10.1109/ICPC.2017.26.

Patel, Vishwa & Tabrizi, Nasseh. (2022). *An Automatic Text Summarization: A Systematic Review*. *Computación y Sistemas*. 26. 10.13053/cys-26-3-4347.

R. Atanda, “Text Summarization of Customer Reviews Using Natural Language Processing,” thesis, 2020.

C.-F. Tsai, K. Chen, Y.-H. Hu, and W.- K. Chen, “Improving text summarization of online hotel reviews with review helpfulness and sentiment,” *Tourism Management*, vol. 80, p. 104122, 2020.

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University - Computer and Information Sciences*, 34(4), 1029-1046. <https://doi.org/10.1016/j.jksuci>.