

Deciphering Public Sentiment on iPhone 14: A BERT-Based Analysis of Twitter Discourse

Tianxuan Sun^{1,a,*}, Tianji Sun^{1,b}, Longhao Tan^{2,c}

¹Department of Data Science, Duke Kunshan University, Suzhou, 215300, China

²Department of Chemistry, Donghua University, Shanghai, 200000, China

a. ts402@duke.edu, b. tianji.sun@dukekunshan.edu.cn, c. yueyuezitlh@gmail.com

*corresponding author

Abstract: The study delves into understanding the sentiment of Twitter users towards the iPhone 14 by employing sentiment analysis models. Recognizing the limitations of traditional sentiment analysis tools, the research utilizes the BERT model, known for its bidirectional understanding of textual context. Initially, the study observed a prevailing negative sentiment using general models. However, after refining the approach using BERT, a more balanced representation of sentiments was observed. The BERT model's results highlighted both positive and negative reactions towards the iPhone 14 on Twitter, with distinct peaks in sentiment scores. The research underscores the importance of leveraging advanced models like BERT for tasks requiring nuanced understanding, providing stakeholders in the technology domain with comprehensive insights into public sentiment.

Keywords: Public Sentiment, Twitter, iPhone 14

1. Introduction

The Apple iPhone brand has persistently left an indelible mark in the technology industry. Its symbolic designing style, innovation, and credibility have not only cultivated a robust consumer base but also facilitated heated discussions across various platforms, such as Twitter. In this digital age, where user feedback can considerably influence brand perception and consumer purchasing behaviors, sentiment analysis emerges as a cardinal tool. Kauffmann et al. [1] shed light on the burgeoning importance of “big data,” particularly the user-generated content (UGC) that floods the internet corridors. This digital footprint, spanning blogs, social media, and online reviews, gives considerable insights from all perspectives, instrumental for sculpting marketing strategies. Beyond mere positive or negative tags, a review can oscillate between varied sentiments corresponding to different product features. For brand custodians, discerning these subtle nuances holds paramount significance, guiding product improvements and fostering brand management.

The fundamental research question guiding this investigation is: "What is the overall sentiment of Twitter users towards the iPhone 14, and how can sentiment analysis models be further enhanced and optimized?" Delving into this inquiry is not merely an academic interest. As the digital world continues to rapidly evolve and consumer conversations become increasingly diverse, the need for refined sentiment analysis tools becomes even more pronounced. These models should deliver beyond just accuracy; they should be adept at discerning the changing details and layers of consumer sentiment. Such continuous improvement in sentiment analysis models is critical, as static models

may fail to capture the depth and dynamism of online user opinions. Furthermore, in an age where opinions can change overnight, largely influenced by social dynamics and digital influencers, it's essential to understand the shifts and trends in sentiment. This ongoing comprehension not only provides a real-time pulse of consumer perception but also empowers businesses to preemptively address problems, innovate, and design their strategies in alignment with the result sentiments. In essence, refining these models is of great significance for businesses wishing to maintain proactive and ensure their products remain in harmony with the shifts of consumer opinion.

To tackle this problem, we use BERT (Bidirectional Encoder Representations from Transformers) models. BERT is anticipated to be adept at deciphering the multifaceted sentiments embedded in tweets. Its ability to consider both the left and right context aligns with our effort to comprehend people's complex opinions about the iPhone 14. By amalgamating these insights, this study endeavors to present a comprehensive analysis of consumer sentiments, potentially guiding stakeholders in the technology domain towards informed decisions.

There are two aspects concerning the significance of sentiment analysis. Firstly, sentiment analysis has tangible ramifications for a brand's image. Discerning sentiments on product features can propel product management strategies, potentially unveiling avenues of innovation [1]. Secondly, in an era where sustainable purchasing behaviors are coveted, the combination of sentiment analysis and big data can change the way consumers approach products, fostering an ecosystem where both brands and consumers thrive in harmony.

As the Apple iPhone faces unprecedented competition in global market, understanding the symphony of sentiments it evokes can shape its novel direction. By rigorous methodology, this study aims to offer insights that reach beyond this research, with potentials for the possible design of future iPhones.

2. Literature Review

2.1. Sentiment Analysis

Sentiment analysis is a crucial natural language processing (NLP) technique that plays a significant role in our digitalized world. As social media platforms like twitter, Facebook, and online communication channels have become indispensable means of giving reviews and expression, the rapid emergence of vast amounts of unprocessed data necessitates efficient analysis to understand human psychology and sentiments. The paper by Pansy Nandwani and Rupali Verma [2] emphasizes that sentiment analysis is a technique that identifies polarities in texts by human and determines whether the writer holds positive, negative, or neutral position towards subjects and locations. As one of the two important part of the Natural Language Processing (NLP), sentiment analysis focuses on polarity assessment and emotion detection delving into identifying specific human emotions like anger, happiness, or sadness [2]. Therefore, the use of sentiment analysis is vital for our research in analyzing the reviews of apple product.

2.2. Twitter

Twitter is one of the most prominent social media platforms, with over 450 million active users and 500 million daily tweets as of 2023 [3]. Users on this platform express their thoughts and engage with content by posting short texts known as "tweets." Twitter has unique features including retweets, mentions, and replies allow users to show their reactions and interactions with each other's tweets [4].

Twitter's specific characteristics make it an optimal resource for sentiment analysis. With its vast user number and the ability to share short messages, twitter serves as an irreplaceable platform for emotions and thoughts, providing considerable data for sentiment assessment. Furthermore, Twitter's open accessibility and Application Programming Interfaces (APIs) makes the data collection much

easier, allowing people to gather a more diverse range of real-time data with less effort. In various distinct fields, Twitter's dynamic characteristic and ease of access make it an optimal choice for sentiment analysis in various research domains. Users can apply for developer accounts and obtain authentication keys to access Twitter data, which includes tweets and user profiles [4]. The versatility of Twitter data collection methods, including API usage, manual data collection, and leveraging data from commercial sources, makes it a valuable asset for a wide range of research inquiries [4].

2.3. Bert Models

BERT stands for Bidirectional Encoder Representations from Transformers. The BERT model we used is designed to pre-train deep bidirectional representations from unlabeled text, making it unique compared to previous language models. Unlike its predecessors, BERT considers both left and right context in all layers, allowing it to capture more comprehensive language understanding. This bidirectional approach enables BERT to be an optimal choice in various natural language processing tasks with minimal task-specific modifications [5].

BERT's straightforward conceptual framework and effective performance have established it as a fundamental tool in natural language processing. It outperforms other models in various tasks like answering questions, deducing language relationships, and gauging sentiment. BERT's adaptability is a key feature, achieved by initially training on unlabeled data and then fine-tuning for specific tasks using labeled data. This adaptability makes BERT an excellent choice for analyzing sentiment in tweets about the iPhone 14, as it can effectively capture the subtleties involved in sentiment classification tasks [5].

3. Methodology

Social media has emerged as a pivotal platform for communication, fostering robust user engagement and boasting an extensive user base. Among its manifold applications, social media has proven to be an invaluable tool for the promotion of the automobile industry. In a particular research endeavor, Twitter was selected as the primary data source due to its pervasive influence and rich repository of user-generated content. Regrettably, free access to the Twitter API has been curtailed in response to concerns over bots and malicious activities. This restriction has further limited unverified users to a daily quota of 1000 tweets, impeding unfettered data retrieval.

To circumvent these constraints, our research turned to a Kaggle dataset generously provided by the user "TLEONEL". (iPhone 14 Tweets [July / Sept 2022 +144k English] | Kaggle). This dataset encapsulates a corpus of Twitter data pertaining to the highly anticipated iPhone 14 model. Comprising an impressive 144,000 tweets, this dataset furnishes a comprehensive panorama of the English-language discourse surrounding the iPhone 14. It offers a nuanced understanding of public sentiments, discussions, and speculations leading up to the device's launch, and extends into the aftermath of the event. Notably, the dataset was last updated on September 9th, encompassing post-launch deliberations that transpired after 1pm ET on September 7th. This meticulously curated dataset forms the cornerstone of our research, providing a robust foundation for comprehensive analysis and insights into the reception and perception of the iPhone 14 within the digital sphere.

To commence our analysis on the sentiment towards the iPhone 14 on Twitter, we initiated the process by importing essential libraries such as pandas, numpy, and transformers. Our dataset, sourced from `iphone14-query-tweets.csv`, primarily consisted of tweets that mentioned the iPhone 14. Before any analysis, we loaded this data into a Pandas DataFrame for a preliminary inspection. Acknowledging the significance of sentiment in textual content, especially in user-generated tweets, we leveraged the transformers library to undertake a comprehensive sentiment analysis. Using a pre-trained sentiment analysis pipeline from this library, each tweet from the dataset was processed to

derive its sentiment label (e.g., positive, negative) and a corresponding sentiment score. This enriched data, comprising the original tweets coupled with their sentiment labels and scores, was then consolidated into a new DataFrame. To facilitate further analysis and potential external evaluations, we stored this sentiment-augmented dataset into an Excel file titled sentiment score.xlsx.

3.1. Data Visualization and Analysis

In our pursuit to uncover insights from the Apple-related Twitter data, we employed a series of data visualization techniques. To begin with, we depicted the sentiment score distribution in two distinct formats: a regular frequency and a log frequency representation. This allowed us to discern patterns and outliers more effectively, especially in a dataset with varying magnitude. The daily average sentiment score trend was another pivotal visualization, offering a temporal view of sentiment evolution. This was instrumental in identifying specific days or events that might have influenced public sentiment. Furthermore, to gain a granular understanding of the topics and themes dominating the discourse, we employed text mining techniques and subsequently constructed a word cloud. This visualization surfaced the most prevalent words, hinting at dominant themes or recurrent topics in the dataset.

3.2. Model Limitations and Need for Alternatives

After scrutinizing the sentiment scores, a noticeable skewness emerged, predominantly leaning towards a negative sentiment trend. As depicted in Figure 1, the distribution of absolute sentiment scores reveals a convergence predominantly within the range of 0.9 to 1.0. This indicates that the sentiment scores tend to be quite extreme.

In Figure 2, the distribution of sentiment scores is further illustrated, with percentages allocated to both negative and positive sentiments.

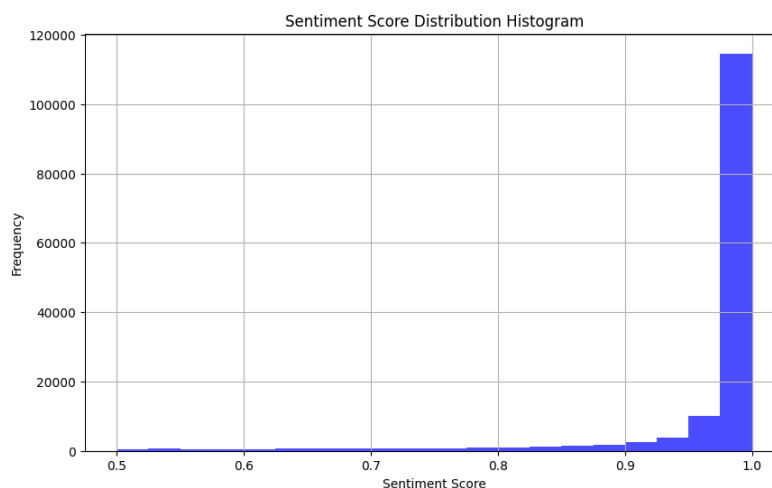


Figure 1: Absolute Sentiment Score Distribution Using Pipeline

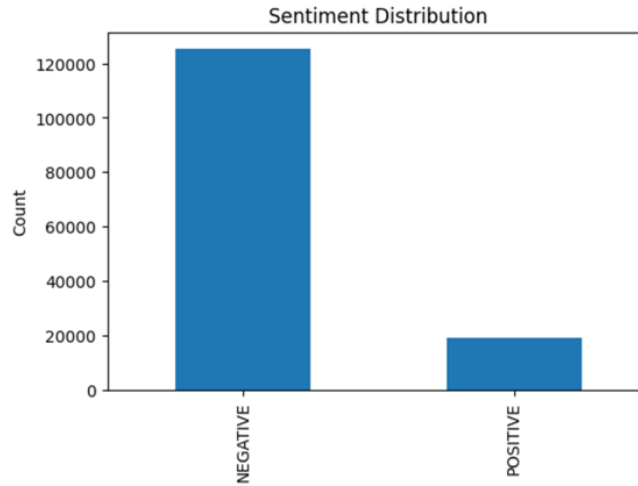


Figure 2: Distribution of sentiment scores

Upon closer examination, the average negative score hovers around 0.974, while the average positive score stands at 0.909. In order to provide an overall sentiment assessment, a negative factor was applied to all negative scores, resulting in an overall average score of -0.835840.

The aggregate sentiment analysis reveals a prevailing negative sentiment, with a positive count of 19,043 (13.2%) and a negative count of 125,202 (86.8%). This phenomenon can be attributed to various factors. Firstly, social media discourse, particularly on platforms like Twitter, often tends to be polarized, with users expressing strong opinions, whether positive or negative. Additionally, the employed model, although robust, may not always capture the subtleties of tweets, especially nuances like slang, emojis, or context-specific references. Notably, sarcasm, known for its complexity in detection, poses a challenge for algorithms. In light of these limitations, it becomes crucial to explore alternative models. While foundational insights were gained from the Naive Bayes and LSTM models, delving into approaches such as Support Vector Machines (SVM) or more advanced neural architectures may yield a sentiment representation that is both balanced and finely attuned to real-world perceptions.

Several reasons can underpin this phenomenon. First, the nature of social media discourse, especially on platforms like Twitter, can be polarized, with users often voicing strong opinions, be it positive or negative. Second, the model employed—while robust—might not always be attuned to the nuances of tweets, particularly slang, emojis, or context-specific references. For instance, sarcasm is notoriously challenging for algorithms to detect. Recognizing these limitations, we believe it is imperative to explore alternative models. While Naive Bayes and LSTM provided foundational insights, delving into models like SVM or even more advanced neural architectures might yield a more balanced and nuanced sentiment representation, better aligned with real-world perceptions.

3.3. Leveraging BERT for Enhanced Sentiment Analysis

Sentiment analysis, especially on platforms like Twitter, is a nuanced task. In our initial approach, we relied on a general sentiment analysis library to evaluate tweets related to the iPhone 14. However, the results exhibited noticeable skewness, predominantly leaning towards negative sentiments. This observation, combined with the inherent limitations of general-purpose sentiment models and the subjective nature of sentiment itself, prompted us to consider alternative, more sophisticated models.

3.4. Enter BERT (Bidirectional Encoder Representations from Transformers).

Why BERT? BERT stands out from traditional language processing models with its unique approach to understanding context. Traditional models process words in an input one by one in order. BERT, however, considers both the previous and next words simultaneously, allowing it to grasp the full context of words in a sentence. This bidirectional understanding makes it especially apt for tasks like sentiment analysis, where the meaning or sentiment of a word can change based on surrounding words.

Fine-tuning a pre-trained model like BERT on specific tasks is akin to imparting specialized knowledge to a generalist. The model has already learned the intricacies of the language from vast amounts of text during its pre-training phase. During fine-tuning, we teach it the specifics of our task, allowing it to adapt its generalized knowledge to be more effective at, in our case, sentiment analysis.

Fine-tuning BERT: Our journey with BERT began with loading a sample dataset, which underwent stratified sampling from a massive dataset of 144k tweets. This dataset was tokenized using BERT's tokenizer, ensuring that our textual data was transformed into a format that BERT understands. With data prepared, we initiated the training phase, utilizing the BERT model for sequence classification. Through iterative processes, the model was exposed to batches of our data, adjusting itself at each step to minimize prediction errors.

Sampling, Scoring, and Subjectivity: Sentiment is inherently subjective. What one person perceives as neutral, another might see as slightly positive or negative. Recognizing this, we employed a collaborative scoring approach for our sample dataset. Project members individually evaluated the sentiment of tweets, after which an average score was computed to provide a balanced view. This methodology not only mitigates individual biases but also offers a more holistic sentiment score, accounting for the varied interpretations that are inherent when gauging sentiment.

Reaping the Benefits: Armed with our fine-tuned BERT model, we ventured back to our vast dataset of 144k tweets. Applying our model, we recalculated sentiment scores, hoping that our specialized BERT model would provide a more accurate, context-aware evaluation of sentiments. The results were promising, underscoring the benefits of using a sophisticated model tailored to our specific needs.

After data processing we have got some results, with figure 3 showing the distribution of sentiment scores after fine tuning

The dataset, assumed to be titled 'iphone14_new.csv', is first read into a DataFrame. Sentiments, presumably pre-analyzed and scaled between 0 (most negative) and 1 (most positive), are categorized into three: negative (0 to 0.5, excluding 0.5), neutral (exactly 0.5), and positive (above 0.5 to 1). The distribution of these sentiments is then visually depicted using a bar chart.

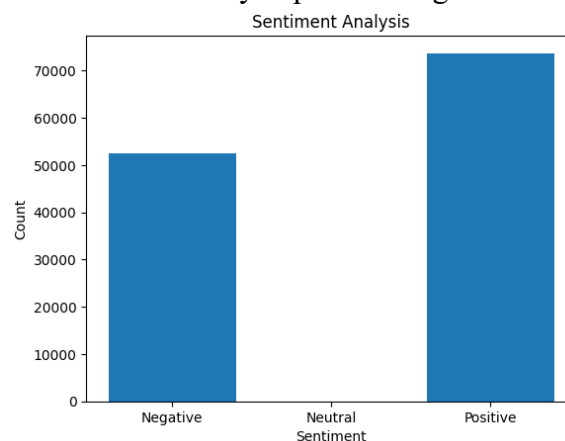


Figure 3: Distribution of sentiment scores after fine tuning BERT model

Using the new model, we conducted an in-depth analysis of the sentiment scores derived from our refined approach. The results were intriguing and provided valuable insights into the emotional landscape surrounding discussions about the iPhone 14 on Twitter. Among the collected tweets, a substantial portion exhibited negative sentiment, totaling 52,394 instances. Conversely, we observed a robust positive sentiment, with 73,686 tweets expressing satisfaction, excitement, or positive feedback about the iPhone 14.

Following this initial assessment, we delved deeper by scrutinizing the distribution of sentiment scores across the dataset. Figure 4 provides a visual representation of this distribution. Notably, we identified a striking symmetry in the spike of sentiment scores. This symmetrical pattern implies a balanced spread of emotions, encompassing both positive and negative sentiments. This observation aligns with our goal of achieving a more comprehensive understanding of the diverse range of opinions and reactions circulating on Twitter.

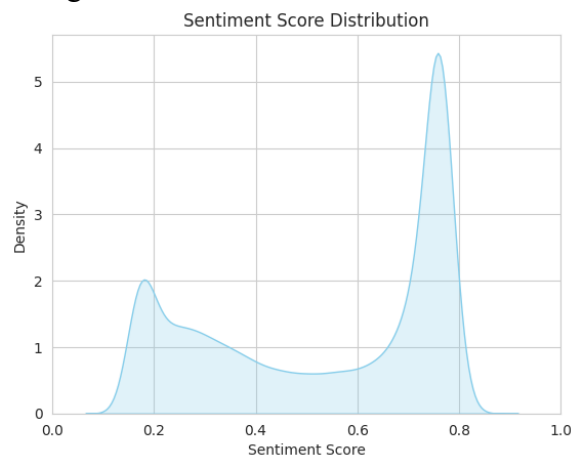


Figure 4: sentiment score distribution using BERT model

However, it's crucial to acknowledge the influence of our specific scoring methodology on the neutrality category. By designating a sentiment score of exactly 0.5 as the threshold for neutrality, we effectively created a stricter classification system. As a result, tweets that might be perceived as borderline neutral were more likely to be categorized as either slightly positive or slightly negative. This decision, while enhancing clarity in classification, inevitably led to a lower frequency of neutral sentiment instances.

Nonetheless, this refinement in categorization yielded a clearer and more nuanced view of sentiment distribution. The BERT model's advanced understanding of context allowed for a more accurate interpretation of the emotional nuances within the dataset. This precision was crucial in capturing the subtleties of sentiment expression, enabling a more refined analysis of public sentiment towards the iPhone 14.

In summary, our utilization of the BERT model has significantly elevated the quality and granularity of our sentiment analysis. The distinct peaks in sentiment scores, the refined categorization approach, and the enhanced clarity in sentiment distribution collectively contribute to a more comprehensive understanding of public sentiment on Twitter in relation to the iPhone 14. This refined methodology not only improves the accuracy of our analyses but also opens the door to a deeper exploration of the diverse range of opinions and emotions within the dataset.

In conclusion, while generic sentiment analysis tools offer a good starting point, tasks requiring nuanced understanding and precision, like discerning sentiment from tweets, benefit immensely from specialized models like BERT. Our journey from recognizing the limitations of our initial approach

to harnessing the power of BERT underscores the importance of choosing the right tool for the task and adapting it to our specific requirements.

4. Conclusion

4.1. Discussion

The results of our sentiment analysis on iPhone 14 tweets reveal mostly positive sentiment, as indicated by the distribution of sentiment scores. This finding raises several important points for discussion.

Firstly, the mostly positive sentiment observed in the Twitter data highlights the unique nature of social media discourse, particularly on platforms like Twitter. Users often express strong opinions, and the brevity of tweets can lead to polarized expressions. This phenomenon underscores the importance of context and the challenges of accurately interpreting sentiment in short, contextually diverse messages.

Furthermore, it is important to acknowledge that our sentiment analysis model, while refined during the experiments, may not fully capture the nuances present in tweets. Things like slang and context-specific references could pose challenges for the sentiment analysis. These limitations engender the necessity to explore alternative sentiment analysis models.

The bias in Twitter reviews towards negativity may also affect the results. Twitter, as a platform, tends to amplify and magnify negative sentiments. Users on Twitter are more likely to voice complaints or criticisms than praise, which can skew sentiment analysis results. On the other hand, the promotion on Twitter on Apple would also bias the result. This inherent bias in social media platforms should be taken into consideration when interpreting sentiment analysis findings in the context of brand perception.

Our decision to employ the BERT model proved valuable in addressing some of the limitations of generic sentiment analysis tools. BERT's bidirectional understanding of context allowed it to better capture the complexities of sentiment in tweets. Fine-tuning the model for our specific task of sentiment analysis yielded more context-aware results, as evidenced by the distribution of sentiment scores after fine-tuning.

Our study underscores the importance of using models like BERT for tasks that require nuanced sentiment analysis, especially in social media data. The prevalence of positive sentiment in iPhone 14 tweets highlights the challenges of interpreting sentiment in short, polarized messages. As brands like Apple continue to navigate the digital landscape, understanding and addressing sentiment on platforms like Twitter remains crucial for shaping brand perception and guiding product management strategies. Future research should explore alternative models and methodologies to further refine sentiment analysis in the dynamic area of social media.

4.2. Limitation

While this research furnishes valuable insights into the sentiment of Twitter users regarding the iPhone 14, it is essential to recognize some inherent limitations in our research. Firstly, the dataset size, comprising 144 thousand tweets and a relatively small subset of 400 tweets for training the BERT model, may not be representative of the broader public opinion. And the data after steps of data cleaning were reduced to 126 thousand tweets, halving the dataset. The sheer volume of Twitter's daily tweets suggests that this is a relatively small fraction of the potential data available on the topic. Also, Additionally, the sentiment score calibration based on the averaging of scores from a three-person team may inadvertently introduce biases or subjective interpretations. This methodological choice, while pragmatic, deviates from the more robust sentiment scoring systems that utilize larger pools of evaluators or predefined scoring algorithms. Finally, the exclusion of neutral sentiments from

our dataset can be constraining. A significant portion of user reviews might express neutral feelings, and neglecting neutral sentiment could mean missing a comprehensive view of the sentiment range, and also categorizing the neutral tweets into positive or negative may bias the result. Addressing these limitations in future research could further refine our understanding of consumer sentiment and contribute to a more nuanced analysis.

References

- [1] Kauffmann, Peral, Gil, Ferrández, Sellers, & Mora. (2019). *Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining*. *Sustainability*, 11(15), 4235. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/su11154235>
- [2] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>
- [3] Shewale, R. (2023, August 10). *Twitter statistics in 2023 - (facts after "X" rebranding)*. DemandSage. <https://www.demandsage.com/twitter-statistics/>
- [4] Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). *Twitter and Research: A Systematic Literature Review Through Text Mining*. *IEEE Access*, 8, 67698-67717. <https://doi.org/10.1109/ACCESS.2020.2983656>
- [5] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. *ArXiv*, 1810.04805.