The Role of GPT Models in Sentiment Analysis Tasks

Mashael M. Alsulami,
mashael.s@tu.edu.sa
Department of Information Technology, College of Computers and Information Technology,
Taif University, Taif, 21944, Saudi Arabia

Abstract
Sentiment analysis has become a pivotal component in understanding public opinion, market trends, and user experiences across various domains. The advent of GPT (Generative Pre-trained Transformer) models has revolutionized the landscape of natural language processing, introducing a new dimension to sentiment analysis. This comprehensive roadmap delves into the transformative impact of GPT models on sentiment analysis tasks, contrasting them with conventional methodologies. With an increasing need for nuanced and context-aware sentiment analysis, this study explores how GPT models, known for their ability to understand and generate human-like text, outperform traditional methods in capturing subtleties of sentiment expression. We scrutinize various case studies and benchmarks, highlighting GPT models' prowess in handling context, sarcasm, and idiomatic expressions. This roadmap not only underscores the superior performance of GPT models but also discusses challenges and future directions in this dynamic field, offering valuable insights for researchers, practitioners, and AI enthusiasts. The in-depth analysis provided in this paper serves as a testament to the transformational potential of GPT models in the realm of sentiment analysis.
Keywords:
Sentiment analysis, GPT models, ChatGPT, NLP

I. Introduction

Sentiment Analysis, or opinion mining, is a computational approach designed to extract and comprehend sentiments, emotions, or opinions embedded within textual data. With the explosion of user-generated content through digital platforms and social media, an intricate tapestry of sentiments has become accessible, offering invaluable insights into societal and individual emotional landscapes. The relevance of sentiment analysis transcends mere academic inquiry, playing an instrumental role in unraveling emotions, attitudes, and opinions towards varied topics, products, or services and providing organizations and entities with crucial insights into societal trends and user experiences.

The advent of Large Language Models (LLMs), particularly the GPT (Generative Pre-trained Transformer) models and their subsequent versions like GPT-3, GPT-3.5 Turbo, and GPT-4, has signified a remarkable evolution in the Natural Language Processing (NLP) domain [1]. These models, renowned for their colossal parameter sizes and extensive training datasets, exhibit an exceptional capacity to comprehend and generate text almost indistinguishable from human production. Their aptitude to navigate through complex linguistic constructs and contexts, accurately deciphering nuances, slang, emojis, and sarcasm, underlines their potency in tasks like sentiment analysis and places them at the forefront of executing advanced NLP tasks. Sentiment analysis tasks are categorized based on [2] into sentiment classification (SC), which aims to classify a given text into a sentiment polarity as positive, negative, or neutral, and Aspect-Based sentiment classification ABSC, which aims to categorize data into certain aspects based on the context of the text, then identify the sentiment polarity associated with each aspect. [2]

This roadmap navigates a pivotal research question: 'How effective are GPT-based models in sentiment analysis compared to traditional approaches, and what factors influence their performance?' Steering through this inquiry, the following sections will explore the capabilities, effectiveness, and limitations of GPT models in sentiment analysis, navigating through their history, methodologies, findings, implications, challenges, and applications across different domains. The aim is to furnish a comprehensive overview, evaluating the role, impact, and potential of GPT models in sentiment analysis, thus providing a robust review of their capabilities and potentials within this sphere.

II. Historical Evolution of Sentiment Analysis

A. Phase 1: Lexicon-Based Approach

The dawn of sentiment analysis was heavily dominated by a lexicon-based approach, which fundamentally utilized a vocabulary of words associated with positive and negative sentiments [3]. The principle behind this method involved tallying the frequency of positive and negative words in a text to deduce the overall sentiment. Words were ascribed a predetermined sentiment score, and algorithms were devised to calculate the cumulative score of the text, thus gauging the prevailing sentiment [4]. Although simple and computationally economical, this approach addressed context ignorance, sentiment dilution through neutral words, and an inability to discern sarcasm or implicit meanings.

Manuscript received September 5, 2024
Manuscript revised September 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.9.2

B. Phase 2: Machine Learning Approach

Sentiment analysis involved heavily in the machine learning (ML) age, as technical developments opened the door for more complex analytical techniques. This phase witnesses a paradigm shift from rule-based to model-based approaches, leveraging algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Random Forest to predict sentiments [5]. The machine learning era signaled a remarkable improvement in sentiment analysis, as models could be trained on vast datasets, enabling them to capture nuances and contextual meanings to a certain degree. While this era offered enhanced accuracy and capability to manage large datasets, it was often criticized for its dependency on hand-crafted features and lack of depth in understanding linguistic complications.

C. Phase 3: Transformer Models

The recent leap in the applicability and performance of sentiment analysis is substantially attributed to the advent of transformer models. Introduced by Vaswani et al. [6], transformer models revolutionized natural language processing (NLP) through attention mechanisms and scalability in training across diverse data. Regarding sentiment analysis, GPT models, especially in their later iterations like GPT-3, have proven adept at capturing contextual, semantic, and syntactical information in textual data, thereby providing a deeper understanding and nuanced analysis of sentiments [7]. Despite the impressive advancements, concerns regarding interpretability, ethical use, and computational expenses still need to be addressed in the transformer model era.

D. Merits and Demerits across the Phases

1) Lexicon-Based Era:

- Merits: Simplicity, low computational cost, and easy interpretability.
- Demerits: Inability to understand context and manage complex linguistic attributes like sarcasm and ambiguity [4].

2) Machine Learning Era:

- Merits: Enhanced accuracy, ability to manage larger datasets and adaptation to various linguistic patterns through training.
- Demerits: Dependency on manual feature extraction and limited understanding of deep contextual meanings [5]

3) Transformer Model Era:

- Merits: Superior comprehension of linguistic nuances, scalability, and improved accuracy in sentiment predictions.
- Demerits: High computational costs, ethical considerations, and interpretability challenges [7].

In retrospect, the journey of sentiment analysis from simple lexicon-based models to the contemporary transformer models narrates a tale of continuous evolution and adaptation, persistently striving towards a more profound and holistic understanding of human sentiments expressed through text.

III. Large Language Models (LLMs) and Their Rise in NLP

A. Overview of LLMs

The momentum in the field of Natural Language Processing (NLP) witnessed a paradigmatic shift with the introduction of large language models (LLMs). LLMs, characterized by their enormous scale and capabilities, facilitate an advanced understanding and generation of human language. Typically trained on diverse and expansive corpora, these models have demonstrated unparalleled proficiency in comprehending context, syntax, and semantics [8]. Employing deep learning and exploiting the breadth of training data, LLMs have paved the way for numerous applications across various domains, such as translation, summarization, and notably, sentiment analysis.

The GPT models, architected on the transformer structure, have become flag-bearers of LLMs in NLP. Originating from GPT to GPT-2 and subsequently to GPT-3, each iteration of these models signifies a monumental step forward regarding scale, training data, and capabilities [9]. The foundational architecture, introduced by Vaswani et al. (2017) [6], utilizes attention mechanisms, enabling the model to focus on different words or phrases in the input text, thus empowering it to capture intricate patterns, relationships, and nuances in language.

B. Significance of GPT-3 and GPT-3.5 Turbo in Sentiment Analysis

GPT-3, with its 175 billion machine learning parameters, and its subsequent version, GPT-3.5 Turbo and GPT-4, have reshaped the landscape of sentiment analysis through their profound linguistic understanding and contextual comprehension [8], [10]. Their ability to analyze textual data and generate coherent, relevant, and contextually rich responses has driven advancements in creating more accurate and nuanced sentiment analysis tools. GPT-3 and its versions, by their training on a diverse range of internet text, understand various linguistic constructs idiomatic expressions and can generate responses or analyses that are coherent and contextually relevant, making them instrumental in deriving sentiment from textual data with increased in accuracy and reliability.

GPT models were used recently by Susnjak [11] as a validation technique of the sentiment analysis model. The article in [b] demonstrate the effectiveness of combining ChatGPT with other pre-trained language models such as SHAP in the validation stage for pre-liminary sentiment analysis results. ChatGPT to subjectivity interpretation of the textual data. [11]

Zhang et al. [12] conduct several experiments to investigate the impact of different prompt designs on the performance of different sentiment analysis tasks. They used GPT-4 with python to automatically generate different prompts. Their results indicate that the performance vary depending on type of tasks rather than the choice of prompts except for fine-grained output [12]. In addition, the paper [12] report several observations about the use of ChatGPT om sentiment classification tasks including that it shows a remarkable ability to improve binary classifications. However, the use of ChatGPT in sentiment analysis do not guarantee better performance and that would be dependent on the number of used parameters. One of the interesting findings reported by [12] is that ChatGPT shows very low performance in detecting hate and offensive language in text. [12]

C. Applications and Impact of GPT Models in Various Domains

The influence of GPT models transcends sentiment analysis, permeating various domains and revolutionizing functionalities. In healthcare, GPT models facilitate patient communication and information retrieval through chatbots and virtual assistants [13]. The legal domain leverages these models for contract analysis and legal research, while in education, they assist in automating feedback and creating educational content. The applications extend to creating art, composing music, and even aiding in scientific research, showcasing their versatile and transformative impact across various domains.

In education, Authors of [14] investigate how ChatGPT may affect student assessments in higher education. They address ChatGPT's consequences for higher education and examines its features, advantages, and disadvantages. The authors look at the potential benefits and drawbacks of AI chatbots in the classroom and offers suggestions for educators, students, and organizations, with a focus on assessment. [14]

In the software engineering domain, Authors of [15] investigates the application of ChatGPT for resolving programming errors. They look at ChatGPT's features and how to use them to help with debugging, bug forecasting, and bug explanation to help with programming issues. The limits of ChatGPT in resolving programming errors are also covered in their study, and it is emphasized how crucial it is to verify ChatGPT's predictions and explanations using additional debugging tools and methodologies. In order to find and solve problems more efficiently, the paper's [15] conclusion emphasizes ChatGPT's potential as a component of a complete debugging toolset as well as the advantages of combining its advantages with those of other debugging tools. [15]

In healthcare domain, Authors of [11] provides a study reporting that by offering information on infectious diseases, chronic illnesses, environmental health risks, and methods for illness prevention and promotion, ChatGPT can help to promote public health. ChatGPT can furnish details regarding tactics to encourage salubrious lifestyle decisions, immunization, screening and prompt identification, and mitigation of risk factors for long-term illnesses, like stress management, blood pressure and cholesterol regulation, abstinence from tobacco and excessive alcohol intake. [11]

In summary, LLMs, epitomized by the GPT models, have not only elevated the capabilities of NLP applications such as sentiment analysis but have also permeated through various domains, substantiating their applicability, and catalyzing technological advancements. Their evolution, especially through iterations like GPT-3 and GPT-3.5 Turbo, embodies the continuous strive towards understanding and generating human language with precision and contextual relevance, thus forging new frontiers in artificial intelligence.

IV. Methodology in Evaluating GPT Models for Sentiment Analysis

A. Introduction to Methodological Approaches

In developing and evaluating models for sentiment analysis, establishing a rigorous methodology is imperative to ensure the reliability and validity of the results. Approaches typically involve a combination of various strategies and metrics tailored to the specific challenges and nature of natural language understanding and sentiment prediction [16]. An adept methodology encompasses model configuration and training and a meticulous assessment through comparative and metrics-driven analyses.

Authors in [17] employ GPT models for sentiment analysis, they integrate strategies like prompt engineering, fine-tuning, and embedding classification to adapt these generalized models to specific tasks. Prompt engineering involves crafting inputs that guide the model to produce desired outputs, leveraging its innate knowledge and abilities [17]. This approach facilitates exploiting the pre-trained capabilities of models like GPT-3.5 Turbo without necessitating fine-tuning.

Fine-tuning, conversely, involves adapting the pre-trained GPT models on a smaller, task-specific dataset to refine its predictions and align them with the domain-specific requirements [17]. This strategy aids in tailoring the broad capabilities of GPT models to the nuanced and specific demands of sentiment analysis.

Embedding classification employs the vector representations generated by the GPT models. These embeddings, capturing semantic and syntactic information of the input text, are utilized to classify sentiments through additional classification layers or models, providing a robust mechanism to exploit the rich representations curated by GPT models for sentiment prediction.

B. Metrics for Evaluating Sentiment Analysis

Various metrics are employed to evaluate the models' performance in sentiment analysis meticulously. Accuracy, representing the proportion of correctly predicted sentiments against the total number of predictions, provides a baseline metric for assessment. Precision, recall, and F1-score offer deeper insights into the model's performance, providing a balanced view of its predictive capabilities across different sentiment classes.

Precision considers the ratio of true positive results to all positive results predicted by the classifier as shown by formula A, while recall (or sensitivity) considers the ratio of true positive results to all actual positives within the dataset as shown by formula B. The score harmonizes precision and recall into a single metric, providing a balanced perspective, especially in scenarios with imbalanced data distributions [18].

$$\text{Precision} = TP/TP + FP \tag{A}$$

$$\text{Recall} = TP/TP + FN \tag{B}$$

In essence, the methodologies in evaluating GPT models for sentiment analysis encompass strategic application through prompt engineering, fine-tuning, and embedding classification.

V. Usage Aanalysis and Findings

On navigating through the hierarchy of GPT models, from GPT-2 to GPT-3.5 Turbo, Bhattarai et. al [10] witness by their research a palpable enhancement in the models' understanding and generating capabilities, attributed to the incremented model sizes and improved training strategies [9]. GPT-3.5 Turbo, with its enriched training data and refined architecture, ostensibly offers a more nuanced understanding of sentiments compared to its predecessors. The intrinsic capability to discern and generate varied linguistic patterns and styles provides GPT-3.5 Turbo with remarkable efficacy in handling diverse and intricate sentiment analysis tasks.

When juxtaposed against traditional models like Support Vector Machines (SVM) and Naïve Bayes classifiers, GPT models manifest a discernibly superior performance in accuracy and adaptability [18]. The pre-training on extensive corpora imparts GPT models with the ability to comprehend and generate intricate and contextually rich linguistic constructs, which starkly contrasts with traditional models' limited, feature-dependent learning. This discrepancy in understanding and generating language becomes particularly pronounced in handling implicit sentiments, sarcasm, and contextually embedded emotions, rendering the GPT models more adept at nuanced sentiment analysis.

Encompassing insight into the models' efficacy in sentiment prediction is obtained in Saroufim et. al [16] by examining the results through metrics like accuracy, F1 score, and recall. While accuracy provides a general overview of the performance, F1-score, and recall, by considering false positives and negatives, impart a more nuanced understanding, which is particularly vital for imbalanced datasets [16]. GPT models, with their profound language understanding, tend to exhibit enhanced recall and precision, particularly in discerning subtle and contextually embedded sentiments, which can often elude traditional models. However, it is crucial to contextualize these results concerning computational and resource expenses, considering the significantly larger and computationally intensive nature of GPT models.

Saroufim et. al [16] show in their study that pivotal facet where GPT models notably excel is their adeptness at grappling with linguistic nuances and complexities, attributable to their training on extensive and varied data [16]. The inherent capability to parse and generate syntactically and semantically rich language allows GPT models to discern sentiments embedded within complex constructs, implicit expressions, and domain-specific jargon with higher accuracy and reliability. However, it is also imperative to acknowledge potential pitfalls, as biases and misinterpretations can stem from the training data, necessitating a vigilant and context-aware application of these models in sentiment analysis tasks.

Authors of [19] investigates how ChatGPT might help businesses do better consumer sentiment analysis. According to the study, ChatGPT offers a great deal of promise for comprehending and addressing the needs, preferences, and satisfaction levels of customers. It can offer businesses insightful viewpoints to enhance their decision-making process. The primary goal of the study was to better understand and explore the use of ChatGPT for customer sentiment analysis in organizations. The goal of the study is to give company professionals and decision-makers insightful information that will help them boost product quality, optimize marketing strategies, and improve customer service. [19]

Authors of [11] offers a step-by-step guide on applying Natural Language Processing (NLP) methods to Sentiment Analysis in the context of text related to tick-borne diseases. The goal is to illustrate the methodology for assessing the existence of bias in the discourse related to long-term Lyme disease symptoms. The chapter outlines how to use pre-trained language models for sentiment analysis and validate the results with big language models like ChatGPT and interpretable machine learning tools. [11]

Authors of [20] addresses in their paper the application of ChatGPT, as a large-scale language model, to the analysis of mental health. The paper identifies the shortcomings of previous research in this field and suggests a thorough assessment of ChatGPT's emotional reasoning and mental health analytic capabilities. Analysis is carried out on the efficacy of various prompting techniques, such as emotion-enhanced and zero-shot prompting. The paper also examines how ChatGPT generates explanations and provides human assessments to gauge how well these explanations are. The results show that ChatGPT demonstrates its promise in mental health analysis and emotional reasoning in conversation, outperforming conventional neural networks like CNN and GRU. [20]

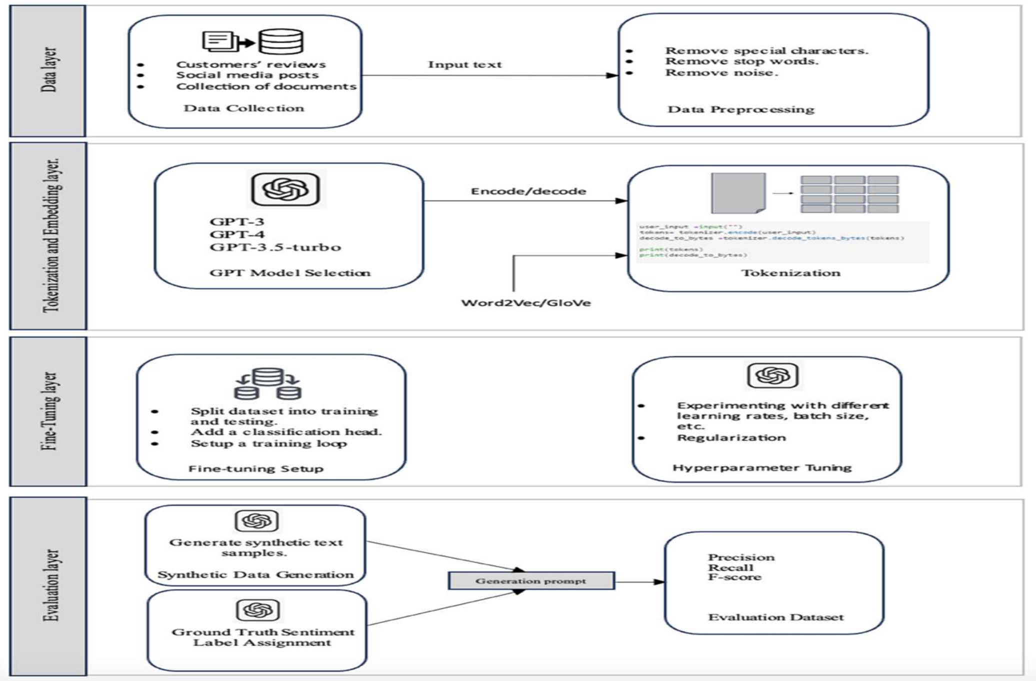

Figure 1. Overview of GPT models' usage in sentiment analysis process.

Figure 1 shows an overview of the usage of GPT models in sentiment analysis process. The usage of GPT models divided into four layers. Each layer encapsulates the main activities that demonstrate a phase in sentiment analysis process. Data layer represents the data collection phase which is usually the first phase when a sentiment analysis model is constructed.

It involves collecting data from different sources to infer their sentiment polarity. This layer includes all the preprocessing steps that should be applied to a collected textual data. The second layer is the tokenization and embedding layer. During ChatGPT's tokenization process, input text is divided into smaller pieces known as tokens. These symbols might be characters, words, or sub-words. After that, the model gives each token a numerical value that is utilized for additional processing. In terms of embedding, ChatGPT uses embeddings to numerically represent these tokens. In essence, embeddings are multidimensional vector representations of words or tokens that capture semantic relationships. Because the model can learn contextual information from the surrounding tokens, these embeddings facilitate the model's comprehension and processing of the input text. The third layer focused on applying GPT models in fine tuning sentiment classification task.

Figure 2. Overview of GPT models' usage in sentiment analysis process

Although the fine-tuning process's specifics are unknown, ChatGPT might be improved for sentiment analysis if Open AI has refined its models by training on datasets created especially for sentiment classification.

The final layer represents the evaluation layer, which represents how GPT models are used to enhance the evaluation of sentiment analysis models. GPT models can be used to evaluate sentiment analysis models through several approaches including generating text as input, and data augmentation by generating more examples to improve the robustness and generalization of the sentiment analysis model.

VI. Challenges and Limitations

A. Privacy and Ethical Concerns in Using GPT Models

In the excitement stemming from GPT models' profound capabilities, it is imperative to scrutinize the associated privacy and ethical concerns. Notably, the models can inadvertently generate outputs that reveal sensitive or private information about the individuals in their training data [21]. Though the training data is broadly extracted and not linked to specific individuals, the expansive knowledge and intricate mimicking capabilities of GPT models often give rise to ethical dilemmas regarding user privacy and data protection. Further, generating persuasive and contextually accurate textual outputs could be exploited to produce misleading or harmful content, underscoring the need for establishing ethical guidelines and utilization norms.

B. Addressing Social and Data Bias and Financial Impediments

The GPT models, though exceptional in language comprehension and generation, are not devoid of biases. They inherit biases present in the training data, which encompass an assortment of texts from various sources, potentially reflecting the social and cultural biases prevalent during the creation of those texts [21]. Mitigating such biases involves comprehensive and conscious efforts in curating training data and developing techniques that neutralize biased weights in model parameters. Active research and deployment of de-biasing strategies and ethical guidelines ensure that sentiment analysis outputs are equitable and non-discriminatory across diverse demographic and socio-cultural segments.

The computational prowess of GPT models comes with substantial financial and resource implications. The training of such models demands colossal computational resources and energy, making it financially restrictive and environmentally impactful [22]. The accessibility and democratization of AI technologies are, therefore, potentially hampered by these factors, as independent researchers and smaller organizations might find it challenging to bear the financial burden associated with utilizing or fine-tuning these large models. Consequently, developing and adapting more computationally efficient models and training methodologies present a formidable challenge and an essential pursuit in the NLP domain.

C. Generalization and Applicability in Varied Contexts

Despite their remarkable performance in numerous applications, the generalization and applicability of GPT models across varied contexts and domains pose a significant challenge. Certain domains, especially those with specialized jargon or limited available data for training and tuning, might experience suboptimal performance [23]. Ensuring that the models understand and generate text effectively and accurately across various domains, linguistic styles, and contexts necessitates specialized fine-tuning and adaptation strategies. Developing methodologies that enhance the models' adaptability and performance across varied sentiment analysis contexts becomes crucial in realizing their full potential.

VII. Implications and Contributions to Sentiment Analysis

A. Interpretation of Key Findings and Their Significance

The profound capabilities of Generative Pretrained Transformer (GPT) models in sentiment analysis cannot be understated, with notable efficacy in contextual understanding and linguistic generation. The intricate design and expansive training data allow these models to comprehend and produce human-like textual responses, thus yielding significant insights into sentiment analysis [24]. Their ability to discern underlying sentiment, especially when fine-tuned for specific domains, has been considerably impactful in understanding user and customer behaviors, paving the way for more nuanced and informed decision-making processes.

B. Impacts on Industry and Academic Research

The integration of GPT models into sentiment analysis has substantial implications for industry and academia. In industries from customer service bots to personalized marketing strategies, GPT models have enabled a more nuanced understanding and prediction of customer sentiments and preferences [25]. In academia, the models facilitate researchers in delving deeper into societal and psychological dynamics, providing tools capable of analyzing large datasets of textual information to glean valuable insights into social, cultural, and individual sentiment trends. Furthermore, it has sparked discourse and research into ethical AI use, bias mitigation, and developing more energy-efficient models.

C. Application and Utility in Different Domains

Diverse domains have found utility in employing GPT models for sentiment analysis, each revealing distinct implications. In healthcare, for instance, patient feedback and experiences can be analyzed to enhance service delivery and patient care. The models can also potentially understand and interpret the sentiments expressed in patient communications, assisting healthcare professionals in providing tailored responses or interventions [26]. In the political arena, sentiment analysis facilitated by GPT models can analyze public opinion on policies and politicians, influencing political strategies and communications. The robustness of GPT models in handling various linguistic styles and jargon significantly enhances their utility across multiple domains, each with its unique linguistic characteristics and challenges.

D. Future Prospects and Developments in Sentiment Analysis

The future trajectory in sentiment analysis, emboldened by the advancements of GPT models, tilt towards even more accurate and context-aware sentiment understanding. The evolution could potentially witness models that understand textual sentiment and integrate multimodal inputs, such as vocal intonations and facial expressions, for a more comprehensive sentiment analysis [27]. Furthermore, there is an emerging need for models capable of understanding and generating responses in many languages and dialects, ensuring that the advantages of advanced sentiment analysis are accessible and applicable globally.

E. Future Prospects and Developments in Sentiment Analysis

The future trajectory in sentiment analysis, emboldened by the advancements of GPT models, tilt towards even more accurate and context-aware sentiment understanding. The evolution could potentially witness models that understand textual sentiment and integrate multimodal inputs, such as vocal intonations and facial expressions, for a more comprehensive sentiment analysis [27]. Furthermore, there is an emerging need for models capable of understanding and generating responses in many languages and dialects, ensuring that the advantages of advanced sentiment analysis are accessible and applicable globally.

Table 1 summarizes the results of comparative analysis of ChatGPT's application in sentiment analysis, highlighting both its strengths and weaknesses across various aspects.

TABLE 1 [ASPECTS OF GPT MODELS IN SENTIMENT ANALYSIS TASKS]

Aspect	Strengths of ChatGPT	Weaknesses of ChatGPT	References
Performance on Sentiment Classification	- High accuracy in basic sentiment classification tasks. - Contextual understanding outperforms traditional models in many cases	- Struggles with specialized tasks like ABSA and ECE without fine-tuning. - May not always surpass specialized models	[2] [28]
Handling Complex Sentiment Nuances	- Capable of understanding nuanced sentiments better than older models	- Difficulty with sarcasm, irony, and fine-grained sentiment intensities without domain-specific tuning.	[29]
Cross-Domain and Multilingual Capabilities	- Strong generalization across different domains. - Robust in multilingual sentiment analysis	- Performance varies across languages and dialects. - Not always as effective as fine-tuned domain-specific models	[29] [30]
Adaptability and Fine-Tuning	- Significant performance improvement with fine-tuning. - Tools like W&B enhance domain-specific accuracy	- Requires substantial computational resources for fine-tuning - Dependent on high-quality training data	[31]
Limitations and Ethical Considerations	- Capable and adaptable for use in various situations	- Inherits biases from training data. - Potential for biased sentiment results in sensitive applications	[8] [32]

VIII. Future Perspectives and Directions

A. Multilingual Sentiment Analysis and Cultural Context

The globalized digital landscape necessitates multilingual sentiment analysis to ensure inclusivity and wider applicability. Multilingual sentiment analysis is pivotal in comprehending sentiments across various languages and dialects, incorporating cultural and contextual nuances to offer more authentic and valuable insights [33]. The endeavor towards developing GPT models that adeptly navigate the complexities of various languages while respecting cultural context represents a significant stride in establishing more universally applicable sentiment analysis tools.

B. Extension to Emotion Recognition and Other Advanced NLP Tasks

Emotion recognition extends beyond sentiment analysis by delving into the specific emotions conveyed in the text. GPT models can be leveraged to enhance the understanding and recognition of a spectrum of emotions in textual data, offering more nuanced insights [34]. Moreover, integrating GPT models into advanced NLP tasks like sarcasm detection and irony understanding, which involve deciphering intricate linguistic patterns and deep contextual understanding, can further refine sentiment analysis methodologies, especially in domains like social media, where such expressions are prevalent.

C. Addressing Challenges and Limitations: Suggested Approaches

Addressing the limitations and challenges associated with GPT models, especially concerning bias, ethical use, and computational costs, is paramount. Strategies like employing more inclusive and diverse training data, implementing robust bias-mitigation algorithms, and developing energy-efficient model architectures can counteract some inherent limitations [34]. Furthermore, ethical guidelines and regulatory frameworks should be established and adhered to, ensuring responsible and equitable use of these powerful models across all domains.

D. Potential Innovations and Evolutions in Sentiment Analysis Methodologies

Innovations in sentiment analysis methodologies are anticipated to evolve towards more integrative and holistic approaches. This may involve developing models that comprehend text and other data types, such as audio and visual inputs, ensuring a comprehensive understanding of communicated sentiments. Moreover, methodologies that harness the amalgamation of various AI technologies, like reinforcement learning and GANs, could offer advanced capabilities in sentiment analysis, producing models capable of learning and adapting dynamically to evolving linguistic landscapes and styles [35].

IX. Conclusion

Conclusively, the thorough examination presented throughout this roadmap underscores the pivotal role and advanced capabilities of GPT models in sentiment analysis, tracing the evolution from simpler, lexicon-based methods to the sophisticated and nuanced processing of the latest transformer models. As delineated, GPT models have exhibited a robust ability to understand and generate text, managing complex natural language processing tasks like sentiment analysis with appreciable accuracy and depth. By delving into different domains and applications, this roadmap has sought to provide a comprehensive overview of the role and impact of GPT models in sentiment analysis, weighing their effectiveness and limitations against traditional models and methods.

While GPT models have opened up new possibilities and efficiencies in sentiment analysis, they bring forth challenges and ethical considerations that must be addressed diligently in future research and applications. Moreover, as technology advances, the prospect of enhancing sentiment analysis through innovations in methodology and addressing the current limitations of GPT models becomes apparent. Ensuring that these models are utilized in a manner that is ethically responsible and universally accessible will be paramount, thus safeguarding the potential for these technologies to be utilized broadly and beneficially across varied domains and applications.

References

[1] J. L. Espejel, E. H. Ettifouri, M. S. Y. Alassan, E. M. Chouham, and W. Dahhane, “GPT-3.5, GPT-4, or BARD? Evaluating LLMs Reasoning Ability in Zero-Shot Setting and Performance Boosting Through Prompts,” *Nat. Lang. Process. J.,* vol. 5, no. 100032, May 2023.

[2] Z. Wang, Q. Xie, Z. Ding, Y. Feng, and R. Xia, “Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study,” *arXiv,* Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.04339

[3] B. Liu, “Sentiment Lexicon Generation,” in *Sentiment Analysis and Opinion Mining,* pp. 79–89, 2012, B. Liu, Ed. doi: 10.1007/978-3-031-02145-9_6.

[4] H. Hamdan, P. Bellot, and F. Bechet, “Sentiment Lexicon-Based Features for Sentiment Analysis in Short Text,” *Res. Comput. Sci.,* vol. 90, pp. 217–226, 2015.

[5] U. Hassan, “Sentiment Analysis Using Machine Learning Classification Models,” *TechRxiv,* May 2022. doi: 10.36227/techrxiv.19783384.v1.

[6] C. Wu, F. Wu, T. Qi, Y. Huang, and X. Xie, “Fastformer: Additive Attention Can Be All You Need,” *ArXiv,* arXiv:2108.09084, Aug. 2021.

[7] T. de Kok, “Generative LLMs and Textual Analysis in Accounting: (Chat) GPT as Research Assistant?,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4429658.

[8] P. Kumar and S. Kathuria, “Large Language Models (LLMs) for Natural Language Processing (NLP) of Oil and Gas Drilling Data,” in SPE, Day 2 Tue, October 17, 2023. doi: 10.2118/215167-MS.

[9] E. Sezgin, J. Sirrianni, and S. L. Linwood, “Operationalizing and Implementing Pretrained, Large Artificial Intelligence Linguistic Models in the US Health Care System: Outlook of Generative Pretrained Transformer 3 (GPT-3) as a Service Model,” *JMIR Med. Inform.*, vol. 10, no. 2, p. e32875, Feb. 2022, doi: 10.2196/32875.

[10] K. Bhattarai, I. Y. Oh, J. M. Sierra, P. R. O. Payne, Z. B. Abrams, and A. M. Lai, “Leveraging GPT-4 for Identifying Clinical Phenotypes in Electronic Health Records: A Performance Comparison between GPT-4, GPT-3.5-turbo and spaCy’s Rule-based & Machine Learning-based methods”, *bioRxiv,* Apr. 2024. doi: 10.1101/2023.09.27.559788.

[11] T. Susnjak, “Applying BERT and ChatGPT for Sentiment Analysis of Lyme Disease in Scientific Literature,” *arXiv,* Feb. 2023, [Online]. http://arxiv.org/abs/2302.06474

[12] W. Zhang, Y. Deng, B. Liu, S. J. Pan, and L. Bing, “Sentiment Analysis in the Era of Large Language Models: A Reality Check,” *arXiv,* May 2023, [Online]. Available: http://arxiv.org/abs/2305.15005

[13] L. Fröhling and A. Zubiaga, “Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover,” *PeerJ Comput. Sci.*, vol. 7, p. e443, Apr. 2021. doi: 10.7717/peerj-cs.443.

[14] J. Rudolph, S. Tan, and S. Tan, “ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?,” *J. Appl. Learn. Teach.*, vol. 6, no. 1, pp. 342–363, Jan. 2023. doi: 10.37074/jalt.2023.6.1.9.

[15] N. M. S. Surameery and M. Y. Shakor, “Use Chat GPT to Solve Programming Bugs,” *Int. J. Inf. Technol. Comput. Eng.*, no. 31, pp. 17–22, Jan. 2023. doi: 10.55529/ijitc.31.17.22.

[16] A. C. Gyllensten and M. Sahlgren, “Measuring Issue Ownership using Word Embeddings,” in *WASSA 2018 - 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Proceedings of the Workshop*, Association for Computational Linguistics (ACL), 2018, pp. 149–155. doi: 10.18653/v1/P17.

[17] H. Wen and Z. Zhang, “SAKP: A Korean Sentiment Analysis Model via Knowledge Base and Prompt Tuning,” in *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)*, IEEE, May 2023, pp. 147–152. doi: 10.1109/CCAI57533.2023.10201257.

[18] E. Azeraf, E. Monfrini, and W. Pieczynski, “Improving usual Naive Bayes classifier performances with Neural Naive Bayes based

models," in *International Conference on Pattern Recognition Applications and Methods,* Nov. 2021.

[19] F. Sudirjo, K. Diantoro, J. A. Al-Gasawneh, H. Khootimah Azzaakiyyah, and A. M. Almaududi Ausat, "Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 3, pp. 283–288, Jul. 2023. doi: 10.47233/jteksis.v5i3.871.

[20] K. Yang, S. Ji, T. Zhang, Q. Xie, Z. Kuang, and S. Ananiadou, "Towards Interpretable Mental Health Analysis with Large Language Models," *arXiv,* Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.03347

[21] M. Leippold, "Swiss Finance Institute Research Paper Series N°23-11 Sentiment Spin: Attacking Financial Sentiment with GPT-3 Sentiment Spin: Attacking Financial Sentiment with GPT-3," *Swiss Finance Institute Research Paper No. 23-11,* Jan. 2023. [Online]. Available: https://ssrn.com/abstract=4337182

[22] C. Sarkar, B. Das, V.S. Rawat, J.B. Wahlang, A. Nongpiur, I. Tiewsoh, *et al.*, "Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development," *Int. J. Mol. Sci.*, vol. 24, no. 3, p. 2026, Jan. 2023, doi: 10.3390/ijms24032026.

[23] M. M. Fischer, "Neural Spatial Interaction Models: Network Training, Model Complexity and Generalization Performance," In *Computational Science and Its Applications–ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24-27, 2013, Proceedings, Part IV 13* (pp. 1-16). Springer Berlin Heidelberg.

[24] A. Ioste, "Transforming the Output of Generative Pre-trained Transformer: The Influence of the PGI Framework on Attention Dynamics," *arXiv,* Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.13317

[25] M. Alawida, S. Mejri, A. Mehmood, B. Chikhaoui, and O. Isaac Abiodun, "A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity," *Information*, vol. 14, no. 8, p. 462, Aug. 2023. doi: 10.3390/info14080462.

[26] O. A. Garcia Valencia, C. Thongprayoon, C. C. Jadlowiec, S. A. Mao, J. Miao, and W. Cheungpasitporn, "Enhancing Kidney Transplant Care through the Integration of Chatbot," *Healthcare*, vol. 11, no. 18, p. 2518, Sep. 2023. doi: 10.3390/healthcare11182518.

[27] H. N. Tran and E. Cambria, "Ensemble Application of ELM and GPU for Real-Time Multimodal Sentiment Analysis," *Memet. Comput.*, vol. 10, no. 1, pp. 3–13, Mar. 2018. doi: 10.1007/s12293-017-0228-3.

[28] Q. Zhong, L. Ding, J. Liu, B. Du, and D. Tao, "Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT," *arXiv,* Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.10198

[29] A. Buscemi and D. Proverbio, "ChatGPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis," *arXiv,* Jan. 2024, [Online]. Available: http://arxiv.org/abs/2402.01715

[30] V. Hudeček and O. Dušek, "Are Large Language Models All You Need for Task-Oriented Dialogue?" [Online]. Available: https://openai.com/blog/chatgpt

[31] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, *et al.*, "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond," *ACM Trans. Knowl. Discov. Data.*, vol. 18, no. 6, pp. 1–32, Jul. 2024. doi: 10.1145/3649506.

[32] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, *et al.*, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learn. Individ. Diff.*, vol. 103. Elsevier Ltd, Apr. 01, 2023. doi: 10.1016/j.lindif.2023.102274.

[33] K. Kheiri and H. Karimi, "SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning," *arXiv,* Jul. 2023.

[34] R. Anderson, C. Scala, J. Samuel, V. Kumar, and P. Jain, "Are Emotions Conveyed Across Machine Translations? Establishing an Analytical Process for the Effectiveness of Multilingual Sentiment Analysis with Italian Text," *J. Big Data Artif. Intell.*, vol. 2, no. 1, Aug. 2023. doi: 10.20944/preprints202308.1003.v1.

[35] S. Cummins, L. Burd, and A. Hatch, "Using Feedback Tags and Sentiment Analysis to Generate Sharable Learning Resources Investigating Automated Sentiment Analysis of Feedback Tags in a Programming Course," in *2010 10th IEEE International Conference on Advanced Learning Technologies*, IEEE, pp. 653–657, Jul. 2010. doi: 10.1109/ICALT.2010.186.