Research Article

International Journal of Data Science and Applications (JOINDATA) 8(2), 101-115, 2025

Received: September 20, 2024 Accepted: January 13, 2025

homepage: https://dergipark.org.tr/tr/pub/joindata

GPT-4o: Analysis of Natural Human-Computer Interaction and Social Effects of Generative Artificial Intelligence by Text Mining Method

Cemal Yüksel*

ABSTRACT

The rapid development of artificial intelligence technologies has brought significant innovations in the field of natural language processing. This study analyses both the technical features and the broader social implications of the GPT-4o model developed by OpenAI. Specifically, the research investigates the model's effects on human-computer interaction, societal acceptance of artificial intelligence, and emerging ethical concerns. A total of 90,016 user comments collected from the YouTube platform were analysed using the DistilBERT-based sentiment classification and BERTopic-based topic modelling methods. The sentiment analysis revealed that 40.5% of the comments were neutral, 33.7% were positive, and 25.8% were negative. Topic modelling indicated a strong focus on themes such as the future of AI, GPT integration, and its perceived social benefits. The model's performance was evaluated using topic coherence (0.429) and topic diversity (0.817) metrics. This study makes a novel contribution to the literature by combining large-scale user-generated content with advanced NLP techniques to empirically assess public perceptions of a cutting-edge AI model. Unlike previous studies, which primarily focus on technical capabilities or theoretical discussions, this research provides an integrated analysis of real-world user reactions, ethical discourse, and model interpretability. The findings highlight that while users show a generally positive attitude towards GPT-4o, concerns regarding misinformation and data privacy remain salient. These insights aim to inform future research and policy-making by offering a grounded perspective on the responsible and socially sensitive deployment of AI technologies.

Keywords: ChatGPT, GPT-4o, Natural Language Processing, Human-Computer Interaction, Text Mining.

1 Introduction

The rapid development of artificial intelligence technologies has led to significant transformations in social areas. This transformation has gained a new dimension, especially in the field of natural language processing (NLP), with Generative Pre-trained Transformer (GPT) models. Developed by OpenAI, these models offer revolutionary innovations in text generation, dialogue management, and language understanding. This study focuses on the GPT-4o model, which represents the latest stage of this evolution and is considered a major milestone in human-computer interaction. Unlike previous versions,

* Bandirma Onyedi Eylul University, Omer Seyfettin Faculty of Applied Sciences Corresponding Author's email: cyuksel@bandirma.edu.tr

GPT-4o has been trained on larger and more diverse data sets, enabling it to offer advanced language skills (Brown et al., 2020; Radford et al., 2018).

The original contribution of this study is its focus on the social acceptance and ethical challenges of the GPT-4o model. While the existing literature on GPT models has mostly addressed their technical capabilities, training methods, or application scenarios, this research aims to understand how GPT-4o is perceived on a social level and to examine the changing dynamics in the context of user experience. In this context, the following research questions are addressed: (1) How is the social acceptance of the GPT-4o model shaped? (2) What are the effects of the GPT-4o model on user experience, and how do these effects differ?

The GPT-1 model first demonstrated the importance of pre-trained language models with 117 million parameters; subsequent models, GPT-2 (1.5 billion parameters) and GPT-3 (175 billion parameters), achieved significant advancements in human-like text generation (Radford et al., 2019; Brown et al., 2020). GPT-4, although its parameter count is undisclosed, has surpassed previous models in terms of accuracy and language diversity. GPT-4o, derived from the Latin word “omni” (everything), is the final iteration of this evolution and aims to make human-computer interaction more natural with its multi-modal structure.

The GPT-4o model can accept a wide variety of data types as input, including text, audio, images, and video, and produce outputs of the same type. With a response time of 232 milliseconds, it can produce results close to human speech speed. While offering outputs equivalent to GPT-4 Turbo performance in English text and code generation, it also provides significant improvements in other languages. Additionally, it operates at 50% lower cost compared to previous models in terms of API usage and offers distinct advantages in understanding visual and auditory content (OpenAI, 2024).

The main objective of this study is to examine the technical innovations of the GPT-4o model and its contributions to human-computer interaction in detail. It also aims to examine the social benefits of this model and analyze user perceptions through YouTube comments. In this context, 90,016 comments collected from the YouTube platform were analyzed using the BERTopic model. This method, which combines DistilBERT and TF-IDF-based word importance measurements, enables the identification of meaningful topics in large-scale text data (Grootendorst, 2020). Thus, users' opinions about GPT-4o can be understood in greater depth.

The following sections of the study address various themes in detail. The second section explains the research methodology; the third section discusses the innovations of GPT-4o in the field of natural language processing and its differences from previous versions. The fourth section summarizes academic studies in different fields related to ChatGPT usage through a literature review. The fifth section evaluates YouTube users' perceptions of GPT-4o in light of BERTopic analyses. Applications in various fields such as education, healthcare, and business life, as well as their social contributions, are discussed. Subsequently, the future potential of GPT-4o and similar technologies is discussed, along with ethical and social issues that should be considered when using these technologies. The final section summarizes the main findings of the study and offers recommendations for further research. This study aims to contribute to future research by providing a guide for the responsible and effective use of artificial intelligence technologies.

2 Method

In this study, the BERTopic and DistilBERT models were used to analyze user comments about GPT-4o on the YouTube platform. A total of 90,016 comments were collected using a web scraper developed with the Python programming language. During this process, the Selenium and BeautifulSoup libraries were used together to systematically obtain all comments under the relevant videos. A comprehensive preprocessing process was applied to the collected data. First, all text was converted to lowercase, and HTML tags and non-alphanumeric characters were removed. Then, semantically meaningless words were extracted using the NLTK library's English stop-word list (e.g., “the,” “is,” “not,” “with,” “you,” etc.).

The langdetect library was used for language detection, and only comments classified as English were included in the analysis. As a result of this filtering, the final analysis dataset consisted of 84,763 comments. The BERTopic model was used to extract meaningful topic headings from the comments.

BERTopic combines Sentence-BERT embedded representations with TF-IDF to provide high topic consistency and diversity, especially in short, scattered, and unstructured texts (Grootendorst, 2020). The model was selected based on its superior performance compared to other traditional methods. Comparatively, for the Latent Dirichlet Allocation (LDA) model, the topic consistency value was calculated as 0.261, topic diversity as 0.522, and the overall coherence score as 0.398. Similarly, for the Non-negative Matrix Factorization (NMF) model, these values were 0.301, 0.559, and 0.421, respectively. BERTopic, on the other hand, provided higher semantic consistency and topic diversity with values of topic consistency = 0.429, topic diversity = 0.817, and coherence = 0.548 on the same dataset. These numerical indicators reveal that LDA and NMF cannot adequately represent the context in short social media texts and produce semantically repetitive headings.

The DistilBERT model was preferred for the sentiment analysis phase of the study. DistilBERT is a version of BERT that is approximately 40% smaller and 60% faster, offering high accuracy rates in natural language processing tasks (Sanh et al., 2019). However, the model's performance in this study was not assumed based solely on general success rates in the literature; instead, it was tested through a validation process specific to this study. For this purpose, 1,000 randomly selected YouTube comments were labeled as positive, negative, and neutral by two independent human coders. The Cohen's Kappa coefficient between the coders was measured as 0.84, indicating a high level of consistency. The accuracy rate was calculated by comparing the human labels with the DistilBERT model's predictions; as a result, the model was found to work with 93.2% accuracy on these social media comments. This finding demonstrates that DistilBERT can be used as a reliable sentiment analysis tool for short, contextual, and multimodal social media texts.

In conclusion, the methods used in both topic modeling and sentiment analysis were carefully selected to suit the nature of social media data; the performance of the models was quantitatively compared and validated. This approach strengthens the methodological robustness of the study and enhances the reliability of the findings.

3 GPT-4o: A New Era in Human-Computer Interaction

On 13 May 2024, OpenAI announced one of the most important updates in the history of ChatGPT. Contrary to expectations, this update is named ‘GPT-4o’, not ‘ChatGPT 4.5’ or ‘ChatGPT 5’. The letter

‘o’ in ‘GPT-4o’ comes from the Latin word ‘omni’ and means ‘everything’. ‘Omni,’ It is a Latin prefix meaning ‘all’ or ‘every’ (Geach, 1973). This new version represents a significant advance in human-machine interaction and offers users a more natural and fluid experience.

Initially, different versions of GPT-4o were secretly introduced in the Chatbot Arena of the Large Model Systems Organisation (LMSYS) under different names. These versions were called ‘gpt2-chatbot’, ‘im-a-good-gpt2-chatbot’ and ‘im-also-a-good-gpt2-chatbot’ (Edwards, 2024). On 7 May 2024, Sam Altman tweeted ‘im-a-good-gpt2-chatbot’, which is often taken as an indication that these models are new OpenAI versions being A/B tested (Altman, 2024; Zeff, 2024).

GPT-4o is capable of processing various types of information, such as text, audio and video, in near real-time and presents this information in a natural interaction with the user. This model makes a significant progress in human-machine interaction, providing a natural experience that makes users forget that the entity in front of them is artificial intelligence. The experiences obtained with GPT-4o differ significantly from previous artificial intelligence models and offer a more natural interaction with humans.

On 13 May 2024, OpenAI's CTO Mira Murati introduced GPT-4o's new features during a live demo. For example, you are watching some videos while preparing for a job interview. But just passively watching these videos may not ensure that the information is learnt effectively enough.

In interview simulations with GPT-4o, the AI often starts the conversation by asking how you are, which creates a friendly atmosphere. The AI approaches it as if you were chatting with a friend and responds with ‘Tell me about it’. For example, ‘OpenAI? Sounds vaguely familiar. Hahaha, just kidding!’ and then laughs. These dialogues demonstrate GPT-4o's capacity for human-like responses (Hietalahti, 2021).

Using its multimodal features, GPT-4o can add images to voice inputs and show itself by switching on its camera. When he asks ‘How do I look?’, the response is ‘You certainly look like you've been up all night coding; that might actually work in your favour, but maybe... You could run your hands through your hair and tidy yourself up'. This response demonstrates the AI's ability to speak with human-like pauses and nuances (Chandra et al., 2022).

When the participant asks GPT-4o to evaluate its voice like a game show host, the AI successfully fulfils this request, leaving the participants unable to hide their surprise. These dialogues show how far AI can go in human-like interactions and the features that increase the realism of these interactions (Nehaniv & Dautenhahn, 2001).

During an AI-supported game, the AI monitors the hand gestures of two people in real time and tries to determine who wins in each round. In the first round, both players make ‘scissors’ and the result is a draw. In the second round, the same situation occurs and they draw again. However, in the third round, the artificial intelligence does not just say ‘And now the third round’. Instead, it adds a characteristic and humorous interpretation to the situation with expressions such as ‘Three has magic’ or ‘Three is the magic number’. Such phrases demonstrate the AI's ability to react in a human-like manner. In the third round, one player ‘scissors’ while the other player ‘cards’ and a winner is determined. The AI summarises the situation and announces the winner and loser by name. This is a consequence of the players being introduced by name at the beginning of the session and the AI's capacity to memorise this

information. Such interactions demonstrate the memory and personalised interaction capabilities of AI (Bai et al., 2023).

The interaction of artificial intelligence systems with each other is one of the most important steps in the development of this technology. Greg Brockman, one of the founders of OpenAI, has demonstrated how two AIs communicate in real time with GPT-4o. The first AI starts the interaction by passing information to the other and describing its environment. The second AI uses the information it receives to make detailed observations, identifying objects in the room and people's facial expressions. In this process, the AIs make their interactions more human-like by using natural and humorous language. These interactions demonstrate the potential of AIs to pass the Turing test and their ability to co-operate in creative processes. The fact that two AIs write and sing songs together also reveals the creative abilities of AIs (Lund, 2004). Recent experimental studies have provided clearer findings regarding the level of success achieved by GPT-4o in such human-like interactions. In a multi-center study conducted in 2025, GPT-4o's performance in the Turing test was compared with that of other large language models such as ChatGPT-4.5 and LLaMA-3.1-405B. In this experiment, participants were asked to guess whether the person they were chatting with was human or AI in randomly assigned chat sessions. According to the results, GPT-4o was only recognized as human 21% of the time, which is significantly lower than ChatGPT-4.5's 73% and LLaMA-3.1-405B's 56% rate. This result indicates that, despite its multimodal capabilities, GPT-4o cannot achieve a level of human-like similarity sufficient to pass the Turing test.

ChatGPT-4o promises to offer important innovations in education and service sector. The model exhibits a pedagogical approach by not only solving maths and geometry problems, but also showing step-by-step solution processes. However, how this approach actually contributes to students should be analysed in detail by educational experts. It is claimed that such applications of ChatGPT-4o can enrich individual learning experiences in education, but this effect needs to be thoroughly validated (Supriyadi & Kuncoro, 2023).

The use of ChatGPT-4o in call centres can potentially answer complex customer questions in detail and increase customer satisfaction. However, more research is needed on the extent to which this potential is verified in real-world applications (George and George, 2023).

The potential benefits of ChatGPT-4o for the visually impaired are considerable. The model can identify and guide what the camera sees in real time; however, the effectiveness of these features should be tested through extensive field studies. In addition, future studies should examine how this technology will perform in other applications such as tourist guidance. Such applications, which have already been discussed in the literature, demonstrate a wide range of societal benefits of ChatGPT-4o (Kuzdeuov et al., 2023; Sudirjo et al., 2023).

4 Literature Review

Since the introduction of ChatGPT, many studies have addressed the impacts of this technology in social and professional domains. In particular, social impacts, such as misinformation and disinformation, have been extensively studied (Abdullah et al., 2022). However, studies of its potential uses in education and healthcare do not adequately address the limitations and potential risks faced in these fields (Biswas, 2023; Shoufan, 2023). Existing studies generally assess the impacts of ChatGPT using social science methods such as surveys, usage experience analyses and expert opinions. However, these methods are

far from providing the in-depth insights provided by studies based on big data analysis. Studies analysing large-scale social media data fill an important gap in this field and allow for more comprehensive results.

Studies analysing public debates about ChatGPT on Twitter have generally focused on the first few months of its release. While these early studies provide some important findings on public acceptance and perception, these findings are based on a limited dataset (Tounsi et al., 2023; Leiter et al., 2023; Haque et al., 2023; Korkmaz, 2023). This makes it difficult to assess long-term effects and understand the dynamics of social acceptance. Such research, especially on social media platforms, provides valuable information on users' first impressions and general attitudes towards technology.

Haque et al. (2022), using Latent Dirichlet Allocation (LDA) technique, tried to identify popular topics in ChatGPT-related tweets created by individuals. However, the limited capacity of LDA to provide contextual information limits the depth of the topics obtained. Leiter et al. (2023) classified tweets into identified topics with a roBERTa-based model, but the effect of biases in this approach was not fully evaluated. Korkmaz et al. (2023) used dictionary and rule-based models to obtain sentiment scores, but it was stated that these models could not adequately capture complex emotional expressions. In this study, dictionary and rule-based models were applied to obtain sentiment scores. Dictionary-based models analyse texts using a dictionary containing sentiment values of specific words and expressions; however, these models may fail to capture contextual meaning. Rule-based models, on the other hand, categorise texts based on certain rules, but may lack flexibility. Leiter et al. (2023) used a fine-tuned Transformer-based model for sentiment analysis to classify tweets as positive, negative or neutral; this model offers high accuracy rates, but may show limitations in assessing complex expressions of sentiment. Koonchanok et al. (2024) used the VADER, roBERTa and XLM-T models for sentiment analysis and the BERTopic model for topic modelling to analyse public attitudes towards ChatGPT. The choice of these models was based on their ability to capture contextual meaning and their performance in short texts; however, each of these models has certain limitations. Table 1 summarises the social media analyses conducted on ChatGPT:

Table 1. Some of the studies that include analyses using social media data within the scope of ChatGPT

Author(s)	Year	Technique Used	Focus Topic
Haque et al.	2022	Latent Dirichlet Allocation (LDA)	Popular topics in tweets related to ChatGPT
Leiter et al.	2023	roBERTa based model	Classification into predetermined subjects
Korkmaz et al.	2023	Dictionary or rule-based models	User sensitivities towards ChatGPT
Koonchanok et al.	2024	VADER, roBERTa, XLM-T; BERTopic	Public attitudes towards ChatGPT

Koonchanok et al. (2024) used the VADER, roBERTa and XLM-T models for sentiment analysis and the BERTopic model for topic modelling to analyse public attitudes towards ChatGPT. Although these models offer advantages in capturing contextual meaning and analysing short texts, they have limitations

such as missing contextual nuances and overgeneral classification. The lack of methodological consistency among the social media analyses summarised in Table 1 makes it difficult to compare the findings of the studies, thus preventing a clear identification of general trends in this literature.

5 Study Findings

This study aims to evaluate the opportunities offered by GPT-4o and to analyse users' opinions and perceptions about GPT-4o in line with the data obtained from YouTube comments. In this context, word cloud analysis was used to identify general trends by visualising the most frequently used words in the comments. However, word cloud analysis is limited in reflecting the context or depth of meaning of words. As shown in Figure 1, words such as ‘chatgpt’, ‘ai’, ‘gonna’ and ‘khan’ are frequently used; these words reflect the general themes of users' interest in GPT-4o. However, a more detailed analysis of the context in which these words are used in the comments and how they shape user perception is needed.

Prefixes such as “non”, “sub”, “micro”, “multi”, and “ultra” are not independent words; they should be joined to the words they modify, usually without a hyphen. There is no period after the “et” in the Latin abbreviation “*et al.*” (it is also italicized). The abbreviation “i.e.,” means “that is,” and the abbreviation “e.g.,” means “for example” (these abbreviations are not italicized).

Figure 1. Word Cloud

Word cloud analysis is a method that helps to identify the main topics and themes that users are interested in by visually presenting the frequency of prominent words in text data. For example, the frequent use of words such as ‘chatgpt’ and ‘ai’ reveals a strong interest in these areas. This analysis provides valuable insights for understanding the impacts of GPT-4o on society and identifying which features users pay attention to.

The word cloud in Figure 1 visualises the major themes prominent in YouTube comments, providing important insights into users' perceptions and attitudes towards technology. Such data allows researchers and developers to better understand user expectations and concerns.

In this study, sentiment analysis of YouTube comments was performed. The sentiment distribution of user comments reveals general trends between positive, negative and neutral reactions. Figure 2 shows the emotional distribution of user comments:

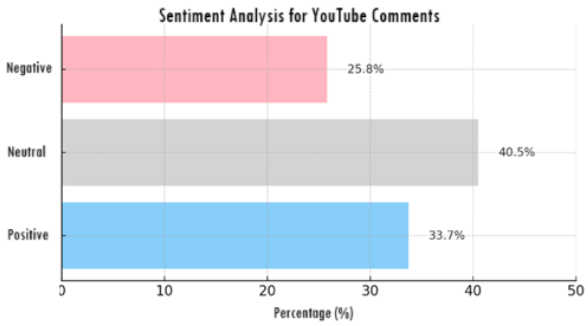

Figure 2. YouTube Comments Sentiment Analysis

The results of the analysis show that 40.5% of the comments are neutral, 33.7% are positive and 25.8% are negative. These findings reveal that user comments on the YouTube platform mostly express neutral or positive sentiments, but a significant portion of them contain negative opinions. Such analyses are critical to understanding users' overall attitudes towards a product, service or topic. By visualising users' emotional reactions, the figure contributes to the assessment of general public perception and helps interested parties to make strategic decisions by better understanding user feedback.

The reason why DistilBERT is preferred in sentiment analysis is the advantages it offers in terms of speed and efficiency. DistilBERT works 60% faster and requires fewer computational resources while maintaining 95% of the accuracy of the BERT model. These features provide a great advantage, especially in situations where large datasets need to be processed quickly. In large-scale data analyses on platforms such as YouTube, these performance advantages of DistilBERT make it possible to achieve efficient and fast results in sentiment analysis tasks (Bárcena and Jesús, 2024).

In this study, topic modelling was also performed using the BERTopic technique to identify popular topics in YouTube comments. Table 2 shows the keywords of each topic:

Table 2. Topics and keywords

No	Topics	Keywords
1	Future of AI	years, ai, future, gaurav, sir, bro, garg, gpt, people, help
2	How Artificial Intelligences Communicate with Each Other	interaction, robots, ai, chatgpt, collaboration, communication
3	GPT Integration	ai, gpt, like, app, chat, open, chatgpt, camera, video, live
4	ChatGPT and Human-Like Interaction	like, chatgpt, ai, openai, voice, look, good, scarlett, human, really, her, movie
5	Improvements to Code Writing and Maths Question Solving	guys, free, new, ai, math, gpt, feature, study, good, code
6	Social Benefits	social benefit, accessibility, assistive technology, visually impaired, real-time assistance, audio guidance, inclusive technology, tourism guide,
7	Human-Like Audio and Video Feedback	ai, like, human, voice, video, insane, feedback, time, audio, visual
8	Realistic Artificial Intelligence Experience	real, ai, wow, time, know, amazing, cooked, chatgpt, like, life
9	Innovation in Education	available, rip, voice, make, like, awesome, better, world, impressive, teachers, time, gpt, khan, people, using, voice, help, nice, video, going
10	Productivity in Business Life	job, gpt, great, mode, work, phone, teacher, chat, help, try

These analyses provide us with important data to understand which topics users focus on about GPT 4o on YouTube. The keywords listed under specific topics allow us to gain a deeper understanding of the scope of these topics and the areas that users are particularly focusing on. Analyses of YouTube comments provide a detailed insight into what users think about GPT 4o and how they approach these topics. Topics such as 'The Future of Artificial Intelligence' and 'AIs Communicating with Each Other' reveal that users intensely discuss the long-term implications of AI technologies and their role in communication. The frequent use of keywords such as 'years', 'ai', 'future', 'interaction' and 'communication' shows that users are thinking deeply about the future of AI and the interaction capabilities of these technologies.

Similarly, the topics of 'GPT Integration' and 'ChatGPT and Human-like Interaction' reveal that users focus on the integration of ChatGPT with different applications and its human-like interaction capabilities. The frequent use of keywords such as 'app', 'chatgpt', 'voice' and 'human' in these topics shows how much importance users attach to innovations and interactions in these areas. Topics such as 'Improvements in Code Writing and Maths Question Solving', 'Social Benefits' and 'Innovation in Education' emphasise the use of artificial intelligence in practical applications and the benefits of these technologies to society. Keywords such as 'math', 'code', 'free' and 'teachers' reflect innovations in education and how users use ChatGPT-4o to tackle everyday challenges such as solving maths questions or writing code.

Keywords such as 'voice', 'video', 'real' and 'amazing' in 'Human-like Audio and Video Feedback' and 'Realistic AI Experience' reflect users' perceptions of the realism and impressiveness of AI. Finally, the keywords 'job', 'work' and 'productivity' under the heading 'Productivity in Business Life' show how artificial intelligence improves business processes and how users evaluate this technology in business life. These findings provide a valuable basis for better understanding the wider impacts of ChatGPT 4o on society and users' perceptions of this technology.

In order to analyse the emotional reactions in YouTube comments in more detail, sentiment analysis for specific topics is of great importance. Figure 3 visualises how emotional reactions in YouTube comments are distributed according to different topics. This analysis allows us to delve deeper into the positive, negative and neutral emotional reactions of users to specific topics. The proportions of positive (green), negative (red) and neutral (blue) emotions for each topic are presented in horizontal bar graphs.

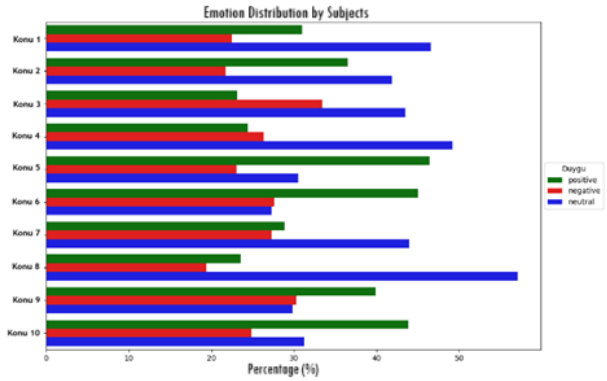

Figure 3. Emotion Distribution by Subjects

Figure 3, which shows the emotional distribution of YouTube comments by topic, allows us to analyse in detail the positive, negative and neutral reactions of users towards various AI topics. The figure visualises the positive (green), negative (red) and neutral (blue) emotional ratios for each topic with horizontal bar charts, helping us to understand users' general tendencies towards specific AI topics and their attitudes towards these technologies.

The topics featured in the figure, such as 'The Future of AI' and 'AIs Communicating with Each Other', are generally met with neutral and positive sentiments, indicating that users approach these topics with curiosity and hope. Similarly, the topics 'GPT Integration' and 'ChatGPT and Human-like Interaction' are also dominated by positive and neutral sentiments. This trend reveals that the integration of GPT technology into different applications and human-like interaction capabilities are viewed positively by users and that there are high expectations for these technologies.

Regarding 'Improvements to Code Writing and Maths Question Solving', users generally express positive sentiments and emphasise the benefits of these technologies in practical applications. Similarly, the topics 'Innovation in Education' and 'Social Benefits' are also met with a high level of positive sentiment, with users appreciating the innovative contributions of AI in education and social areas. In particular, keywords such as 'math', 'code', 'free' and 'teachers' reflect users' positive views that they can overcome the challenges they face in daily life with ChatGPT-4o.

Most of the users express positive and neutral feelings about 'Human-like Voice and Image Feedback' and 'Realistic Artificial Intelligence Experience'. Keywords such as 'voice', 'video', 'real' and 'amazing', which stand out in these topics, reveal positive thoughts about how realistic and impressive AI is. This indicates a strong perception of AI's potential to enrich the user experience.

Finally, the topic of 'Efficiency in Business Life' stands out with user comments emphasising the potential of AI to improve business processes and increase efficiency. The frequent use of the keywords 'job', 'work' and 'productivity' shows how artificial intelligence is evaluated in business life applications and the positive impact of this technology on business processes. These findings provide an important basis for better understanding the wider impacts of ChatGPT-4o on society and users' perceptions of these technologies.

The BERTopic method has achieved very successful results in studies on different datasets. For example, in a study on datasets consisting of 20 Newsgroups, BBC News and Trump's tweets, the BERTopic method achieved higher topic coherence and topic diversity scores compared to the classical LDA (Latent Dirichlet Allocation) method. In this study by Bárcena and Jesús (2024), the topic coherence score obtained with BERTopic was reported as 0.166 and the topic diversity score as 0.851. In contrast, the consistency score of LDA was 0.058 and the diversity score was 0.749. Furthermore, in another study, the BERTopic method was evaluated on social media data and showed superior performance compared to other methods, especially on short texts (e.g. Twitter data). In this study, it was stated that the contextual information encoding capacity of BERTopic was particularly prominent in short and unstructured texts (Egger & Yu, 2022).

In this study using the BERTopic model, the performance of the model was evaluated on two main metrics: topic coherence and topic diversity. The results obtained provide important clues to understand the overall performance of the model. The topic coherence score was determined as 0.429. This score indicates that the capacity of the model to establish a meaningful link between topics is at a reasonable

level. However, it also indicates that there is room for improvement to achieve a higher level of coherence. While this score is acceptable by academic standards, a higher coherence score would indicate that the model is capable of producing more refined and meaningful topics.

The topic diversity score was measured as 0.817. This score emphasises the model's ability to generate a diverse and wide range of topics. The scarcity of repetitive words among the topics indicates that the model is able to generate different and unique topics. This diversity is particularly valuable in large and heterogeneous datasets and demonstrates that the model can be used effectively in different application areas.

These performance metrics reveal that the BERTopic model has achieved above average success and can be reliably used in a variety of academic and practical applications. The model's balance between topic consistency and diversity shows a promising performance on different datasets, indicating that it may require fine-tuning for more advanced applications.

Overall, the figure and related analyses help us understand the general attitudes and expectations towards AI technologies and provide a solid basis for evaluating the societal impacts of these technologies. The positive and neutral reactions of users towards specific topics indicate that AI technologies tend to be generally accepted and that trust in these technologies is increasing. These findings constitute an important reference point for future research and development studies.

6 Conclusion, Discussion and Suggestions

Since the introduction of ChatGPT, many studies have addressed the impacts of this technology in social and professional domains. In particular, social concerns such as misinformation and disinformation have been extensively studied (Abdullah et al., 2022). However, the potential applications of ChatGPT in education and healthcare have not been sufficiently examined in terms of their limitations and associated risks, which constitute a critical research gap (Biswas, 2023; Shoufan, 2023).

In the field of education, while ChatGPT is praised for providing quick explanations, personalised feedback and support for language learning, studies have highlighted serious concerns about its reliability, pedagogical soundness, and risk of fostering academic dishonesty. For instance, Biswas (2023) emphasizes that students may become over-reliant on ChatGPT, leading to surface-level learning and a decline in critical thinking skills. Moreover, when prompted with academic essay questions, ChatGPT may produce convincing but factually incorrect or overly generalised content—raising concerns about "AI-generated misinformation" within learning environments. Similarly, Shoufan (2023) notes that ChatGPT’s explanations of mathematical or scientific concepts, while fluent in language, often lack conceptual depth, which could mislead learners, especially in STEM fields.

Furthermore, some empirical studies have begun testing ChatGPT's educational accuracy. For example, King and Chatfield (2023) evaluated ChatGPT’s responses to high-school-level exam questions in biology and found that the model gave partially incorrect answers in 27% of the cases, often due to subtle conceptual misunderstandings that may not be easily spotted by learners. Such errors, when uncritically accepted, may foster misconceptions that are harder to correct later. These findings suggest that ChatGPT should be integrated into educational settings with caution and clear pedagogical scaffolding.

In the field of healthcare, ChatGPT's potential for assisting in medical education, mental health support, and symptom checking has been explored. However, these use cases raise serious ethical and operational concerns. Shoufan (2023) and Jeblick et al. (2023) have shown that ChatGPT may produce medically plausible responses that are, nonetheless, inaccurate, outdated, or inappropriate without human oversight. In one evaluation, ChatGPT was asked to provide information on cancer treatment options; while its response was linguistically coherent, it failed to include current guidelines and offered advice that, if followed, could delay critical medical interventions.

Moreover, Nori et al. (2023) examined ChatGPT's performance on US Medical Licensing Examination (USMLE)-style questions and found a mixed performance, with the model achieving pass-level accuracy only in selected domains like pharmacology or pathology, but struggling with case-based reasoning tasks that require nuanced judgment. Such findings point to the risks of using LLMs in clinical decision-making or patient-facing applications without rigorous validation and ethical safeguards.

Overall, the existing body of research in both domains underscores the need for caution, transparency, and human oversight in the deployment of ChatGPT in education and healthcare contexts. Although the technology offers promising applications, its integration must be accompanied by critical evaluation frameworks to avoid unintended harms, particularly in high-stakes settings.

In light of these findings, this study proposes the implementation of a Validation Module to ensure that outputs generated by ChatGPT—especially in critical domains such as education and healthcare—undergo a layer of real-time quality control before reaching end-users. This module aims to mitigate risks related to factual inaccuracies, misleading explanations, and inappropriate recommendations.

The proposed Validation Module would operate as a post-generation verification layer integrated into AI deployment pipelines. It consists of three main components:

- Factual Consistency Checker: This component automatically compares AI-generated outputs with verified external knowledge bases (e.g., PubMed, WHO guidelines, national curricula). For example, if the model generates a response about cancer treatment or mitosis, the content is validated against current clinical or academic standards to detect any contradictions or outdated information.

- Contextual Risk Evaluator: Outputs are scanned for high-risk content indicators—such as overconfident claims, lack of nuance in complex topics, or emotionally sensitive phrasing. In educational scenarios, this could include detecting oversimplified scientific explanations; in healthcare, it would flag responses that omit risk disclaimers.

- User-Facing Transparency Layer: If an output fails validation, the system alerts the user through confidence scoring or flags, e.g., "This information may be incomplete or outdated. Please consult an expert or authoritative source before acting." This component encourages critical engagement rather than blind trust.

By embedding such a module into LLM systems, developers and institutions can uphold higher standards of accuracy, safety, and ethical accountability—in line with the concerns raised by Biswas (2023), Shoufan (2023), and Jeblick et al. (2023). Future implementations may also integrate

explainability metrics and user feedback loops to continuously improve validation performance over time.

Existing studies generally assess the impacts of ChatGPT using social science methods such as surveys, usage experience analyses and expert opinions. However, these methods are far from providing the in-depth insights enabled by studies based on big data analysis. Studies analysing large-scale social media data fill an important gap in this field and allow for more comprehensive results.

6.1 Study Limitations

The fact that the study is based on a specific data set may limit the applicability of the findings in a broader context. However, these limitations may contribute to a more in-depth examination of the topic by providing an essential basis for future research.

6.2 Acknowledgments

This paper is derived from the first author's doctoral thesis.

6.3 Funding source

There is no funding source.

6.4 Competing Interests

There is no conflict of interest in this study.

6.5 Authors' Contributions

Study conception and design: Akın, Terzi. Data acquisition: Terzi. Data analysis and interpretation: Akın. Article drafting and revising: Akın, Terzi. All authors approved the final manuscript.2

References

[1] Abdullah, M., Madain, A., & Jararweh, Y. (2022). Chatgpt: Fundamentals, applications and social impacts. In 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–8). IEEE.
[2] Altman, S. (2024, May 7). im-a-good-gpt2-chatbot [Tweet]. Twitter. Retrieved May 14, 2024, from https://twitter.com/sama/status/1787222050589028528
[3] Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. Brain-X, 1(3), e30.
[4] Biswas, S. S. (2023). Role of ChatGPT in public health. Annals of Biomedical Engineering, 1–2.
[5] Bárcena Ruiz, G., de Jesús Gil, R. (2024). BERT Transformers Performance Comparison for Sentiment Analysis: A Case Study in Spanish. In: Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F., Poniszewska-Marańda, A. (eds) Good Practices and New Perspectives in Information Systems and Technologies. WorldCIST 2024. Lecture Notes in Networks and Systems, vol 989. Springer, Cham. https://doi.org/10.1007/978-3-031-60227-6_13
[6] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

[7] Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To be or not to be… human? Theorizing the role of human-like competencies in conversational artificial intelligence agents. Journal of Management Information Systems, 39(4), 969-1005.
[8] Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology, 7, 886498.
[9] Edwards, B. (2024, May 13). Before launching, GPT-4o broke records on chatbot leaderboard under a secret name. Ars Technica. Retrieved May 17, 2024, from https://arstechnica.com/information-technology/2024/05/gpt-4o-broke-records-on-chatbot-leaderboard/
[10] Geach, P. T. (1973). Omnipotence. Philosophy, 48(183), 7-20.
[11] George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. Partners Universal International Innovation Journal, 1(1), 9-23.
[12] Grootendorst, M. (2020). BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics. arXiv preprint arXiv:2008.03979.
[13] Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). "I think this is the most disruptive technology": Exploring sentiments of ChatGPT early adopters using Twitter data. arXiv preprint arXiv:2212.05856.
[14] Hietalahti, J. (2021). Laughing with machines. The European Journal of Humour Research, 9(2), 154-171.
[15] Koonchanok, R., Pan, Y., & Jang, H. (2024). Public attitudes toward ChatGPT on Twitter: sentiments, topics, and occupations. Social Network Analysis and Mining, 14(1), 106.
[16] Korkmaz, A., Aktürk, C., & Talan, T. (2023). Analyzing the user's sentiments of ChatGPT using Twitter data. Iraqi Journal for Computer Science and Mathematics, 4(2), 202–214.
[17] Kuzdeuov, A., Mukayev, O., Nurgaliyev, S., Kunbolsyn, A., & Varol, H. A. (2024, February). ChatGPT for visually impaired and blind. In 2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 722-727). IEE
[18] Leiter, C., Zhang, R., Chen, Y., Belouadi, J., Larionov, D., Fresen, V., & Eger, S. (2023). ChatGPT: A meta-analysis after 2.5 months. arXiv preprint arXiv:2302.13795.
[19] Lund, H. H. (2004). Modern artificial intelligence for human-robot interaction. Proceedings of the IEEE, 92(11), 1821-1838.
[20] Nehaniv, C. L., & Dautenhahn, K. (2001). Imitation in natural and artificial systems. Cybernetics & Systems, 32(1-2), 1-10.
[21] OpenAI. (2024). Hello GPT-4o. [Online] Available at: https://openai.com/index/hello-gpt-4o/
[22] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI.
[23] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., ve diğerleri. (2019). Language models are unsupervised multitask learners. OpenAI.
[24] Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. IEEE Access.
[25] Sudirjo, F., Diawati, P., Riady, Y., Ausat, A. M. A., & Suherlan, S. (2023). The role of chatgpt in enhancing the information search and decision-making process of travellers. Jurnal Minfo Polgan, 12(1), 500-507.
[26] Supriyadi, E., & Kuncoro, K. S. (2023). Exploring the future of mathematics teaching: Insight with ChatGPT. Union: Jurnal Ilmiah Pendidikan Matematika, 11(2), 305-316.
[27] Tounsi, A., Elkefi, S., & Bhar, S. L. (2023). Exploring the reactions of early users of ChatGPT to the tool using Twitter data: Sentiment and topic analyses. In 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC ASET) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/10150870
[28] Zeff, M. (2024, May 7). Powerful New Chatbot Mysteriously Returns in the Middle of the Night. Gizmodo. Retrieved May 17, 2024, from https://gizmodo.com/powerful-new-chatbot-mysteriously-returns-1849968290new mechanics of change. Idea.

[29] Karimi, R., Baghalzadeh Shishehgarkhaneh, M., Moehler, R. C., & Fang, Y. (2024). Exploring the Impact of Social Media Use on Team Feedback and Team Performance in Construction Projects: A Systematic Literature Review. Buildings, 14(2), 528.
[30] Leong, C., Pan, S. L., Bahri, S., & Fauzi, A. (2019). Social media empowerment in social movements: power activation and power accrual in digital activism. European Journal of Information Systems, 28(2), 173-204.
[31] Lobera, J., & Portos, M. (2021). Decentralizing electoral campaigns? New-old parties, grassroots and digital activism. Information, Communication & Society, 24(10), 1419-1440.
[32] Minocher, X. (2019). Online consumer activism: Challenging companies with Change. org. New Media & Society, 21(3), 620-638.
[33] Mousavi, R., Johar, M., & Mookerjee, V. S. (2020). The voice of the customer: Managing customer care in Twitter. Information Systems Research, 31(2), 340-360.
[34] Mutlu, E. (1991). Televizyonu Anlamak. İstanbul: Gündoğan Yayınları.
[35] Napoli, P. M. (2011). Audience evolution: New technologies and the transformation of media audiences. Columbia University Press.
[36] Nielsen. (2023). 2023 Nielsen Annual Marketing Report. Nielsen. https://sproutsocial.com/insights/data/harris-insights-report-2023/ (Erişim Tarihi: 30 Haziran 2024)
[37] Ouma, N. C. (2013). Effects of Social Media on Content of Local Television Programs in Kenya: a Case Study of Citizen Tv's Gospel Sunday Show (Doctoral dissertation, University of Nairobi).
[38] Özkula, S. M., Reilly, P. J., & Hayes, J. (2023). Easy data, same old platforms? A systematic review of digital activism methodologies. Information, Communication & Society, 26(7), 1470-1489.
[39] Pérez-Torres, V. (2024). Social media: a digital social mirror for identity development during adolescence. Current Psychology, 1-11.
[40] Waddell, T. F., & Sundar, S. S. (2017). # thisshowsucks! The overpowering influence of negative social media comments on television viewers. Journal of Broadcasting & Electronic Media, 61(2), 393-409.
[41] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731-5780.
[42] Zaheroghli, M. N. (2023). Türk dizi ihracatının Türk ürünlerine yönelik satın alma niyeti açısından incelenmesi= Examination of Turkish TV series exports in terms of purchasing intention for Turkish products (Master's thesis, Sakarya Üniversitesi).