Journal of Information Systems and Informatics
Vol. 6, No. 2, June 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i2.740 Published By DRPM-UBD

Analyzing an Interest in GPT 4o through Sentiment Analysis using CRISP-DM

Yerik Afrianto Singgalen[1*]

[1*] Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1*]yerik.afrianto@atmajaya.ac.id

Abstract

This study investigates the sentiment of viewers towards GPT-4o technology videos by analyzing 1538 English language posts using two sentiment analysis tools, VADER and TextBlob. The analysis reveals a fair level of agreement between the two tools, with 929 posts (60.40%) classified consistently, yielding a Cohen's kappa statistic of 0.388. The sentiment distribution among the posts is as follows: 182 posts (19.59%) exhibit negative sentiments, 390 posts (41.98%) are neutral, and 357 posts (38.43%) show positive sentiments. These findings highlight the importance of utilizing multiple tools for comprehensive sentiment analysis and underscore the complexity of interpreting public reactions to AI advancements. The study provides valuable insights into the nuanced responses of viewers, emphasizing the diverse perspectives towards the GPT-4o technology.

Keywords: Sentiment; Classification; VADER; TextBlob; GPT-4o

1. INTRODUCTION

The rapid advancements in artificial intelligence technology and the integration of AI tools across various sectors have sparked significant debate. AI systems are being deployed from healthcare to finance to enhance efficiency and accuracy, offering transformative potential [1]–[4]. However, concerns about ethical implications, data privacy, and job displacement have emerged as critical issues [5]–[7]. It is posited that while AI leads to unprecedented innovations, rigorous regulations and ethical guidelines are imperative to mitigate potential risks [8], [9]. In conclusion, the discourse surrounding AI underscores the necessity for a balanced approach that maximizes benefits while addressing societal challenges.

Public perception of advancements in AI technology reflects a diverse response to its growing influence. Opinions range from enthusiastic support for AI's potential benefits to apprehension about its societal impact [10], [11]. Thus, sentiment analysis is essential to gauge public acceptance of AI technology. This

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

approach provides valuable insights into the prevailing attitudes and concerns, enabling stakeholders to address issues proactively [12]–[15]. Understanding public sentiment through systematic analysis is crucial for fostering a balanced integration of AI into society.

This study aims to analyze the toxicity score and sentiment viewer of video content concerning GPT-4o published on YouTube. By evaluating the toxicity levels in comments and the overall sentiment of viewers, the research provides a comprehensive understanding of the public's reaction to this advanced AI model [16]. It is asserted that such analysis is pivotal for identifying potential issues in public discourse and guiding the development of more responsible AI communication strategies [17]–[19]. In conclusion, the findings will offer critical insights into the interaction between AI advancements and public perception, informing future engagements and policy decisions.

The method employed in this sentiment analysis is the Cross-Industry Standard Process for Data Mining (CRISP-DM). This robust framework guides the systematic execution of data mining projects through its six well-defined phases: business understanding, data understanding, modeling, evaluation, and deployment [20]. Utilizing CRISP-DM ensures a structured approach that enhances the reliability and validity of the sentiment analysis results [21]. Therefore, this methodology is highly effective for extracting meaningful insights from large datasets. In conclusion, CRISP-DM provides a comprehensive and adaptable framework for successfully implementing sentiment analysis projects.

The urgency of this research lies in its potential to address critical challenges posed by rapidly advancing AI technologies. In an era where AI systems increasingly influence daily life, understanding public sentiment and identifying toxicity in online discourse is paramount [22]–[24]. Prompt and thorough analysis not only informs the ethical development of AI but also helps mitigate societal risks associated with misinformation and negative public perception [25], [26]. Therefore, this research is indispensable for fostering a balanced and informed integration of AI technologies into society [27]. In conclusion, addressing these pressing issues through timely and rigorous analysis underscores the critical importance of this study.

This research's theoretical and practical implications extend across multiple dimensions of AI and public interaction. Theoretically, it contributes to a deeper understanding of sentiment analysis methodologies and their applicability in real-world scenarios, enriching the academic discourse on data mining and AI ethics. Practically, the findings offer actionable insights for policymakers and technology developers to enhance AI communication strategies and mitigate adverse public reactions. Such dual impact ensures that the research advances theoretical knowledge and provides tangible benefits for societal engagement with AI

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

technologies. In conclusion, the study's outcomes underscore its significance in bridging theoretical constructs with practical applications.

The limitations and comparison to similar studies provide a comprehensive context for this research. One fundamental limitation is the potential bias in data sources, which may not fully represent the diverse spectrum of public opinion. Additionally, the dynamic nature of online sentiment poses challenges in maintaining the timeliness and accuracy of the analysis. Similar studies, such as those examining sentiment in social media platforms or assessing public opinion on emerging technologies, offer valuable benchmarks but often lack the specificity required for focused AI-related discourse [28], [29]. Addressing these limitations through methodological refinements will enhance the robustness of future research. In conclusion, acknowledging these constraints and drawing parallels with related studies enriches the understanding and applicability of the findings.

2. METHODS

2.1 Gap Analysis

The gap analysis is conducted to pinpoint underexplored topics regarding public responses to AI, facilitating a more comprehensive understanding of this complex interaction. This process highlights avenues for future investigation and theoretical development by identifying areas where research is scarce or non-existent. Furthermore, it enables researchers to uncover nuanced aspects of public sentiment that may have been overlooked in previous studies. Consequently, the gap analysis serves as a crucial starting point for advancing knowledge and addressing critical gaps in the discourse surrounding public perceptions of AI. In conclusion, systematically identifying research gaps lays the groundwork for a more nuanced and holistic understanding of public attitudes toward AI.

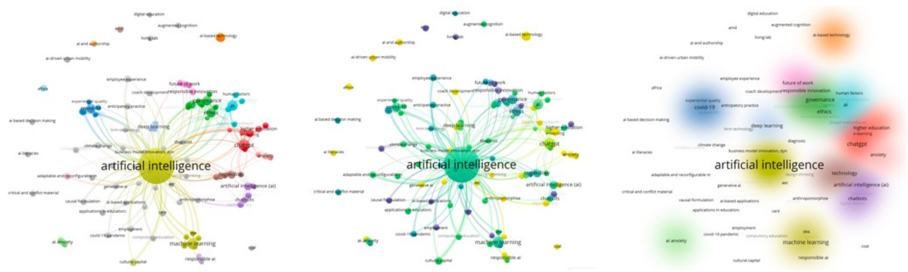


Figure 1. Gap Analysis of AI Research through Sentiment Classification

Figure 1 shows the gap analysis of AI research. Despite the burgeoning interest in artificial intelligence, comprehensive sentiment analysis remains relatively

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

unexplored, indicating a significant research gap. Current literature primarily addresses AI's technical advancements and ethical considerations, yet scant attention is paid to the public's emotional and perceptual responses [30]–[36]. Analyzing these sentiments is crucial as they influence policy-making, adoption rates, and societal acceptance of AI technologies [37]. Therefore, this study aims to bridge this gap by systematically examining sentiment trends related to AI, ultimately contributing to a more holistic understanding of its impact and informing future research directions.

Consequently, this study proposes using the CRISP-DM framework to systematically identify and analyze public sentiment toward GPT-4 video content. CRISP-DM, renowned for its structured approach to data mining, offers a comprehensive methodology for handling complex datasets and deriving meaningful insights [38]. Employing this framework enables a meticulous examination of sentiment dynamics, enhancing the accuracy and reliability of the findings. By adopting CRISP-DM, this research aims to provide a robust analytical foundation, facilitating a deeper understanding of public perceptions and contributing valuable knowledge to AI sentiment analysis.

2.2 Cross-Industry Standard Process for Data-Mining (CRISP-DM)

The framework employed in this study is CRISP-DM, a widely recognized methodology in data mining. CRISP-DM's structured phases, including business understanding, data preparation, modeling, evaluation, and deployment, ensure a systematic approach to data analysis [39]. Its flexibility and adaptability to various domains make it an ideal choice for examining complex datasets [40]. By leveraging CRISP-DM, this research aims to achieve comprehensive and reliable insights, ultimately enhancing the analytical rigor and contributing significantly to the field of data-driven research.

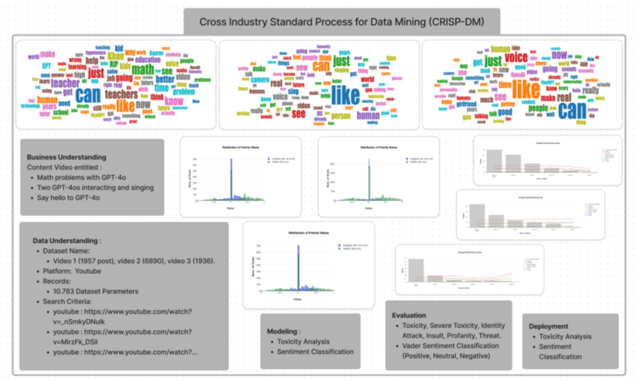


Figure 2. Implementation of Cross-Industry Standard Process for Data-Mining (CRISP-DM)

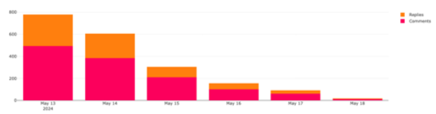
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 2 shows the implementation of CRISP-DM. CRISP-DM implementation encompasses several stages: business understanding, data understanding, modeling, evaluation, and deployment. This study specifically focuses on identifying toxicity scores and analyzing public sentiment regarding the AI technology known as GPT-4o. The business understanding phase ensures alignment with research objectives, while data understanding involves thorough exploration and preprocessing of relevant datasets. Subsequent modeling and evaluation phases are critical for developing robust analytical models and assessing their performance. Finally, deployment translates these insights into actionable outcomes, ultimately advancing knowledge in AI sentiment analysis and addressing the complexities of public perceptions towards emerging technologies.

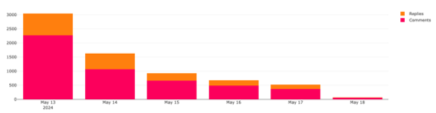
2.2.1 Business Understanding

In the business understanding phase, it is essential to comprehend the context of sentiment identification and analysis, specifically regarding public responses to the launch of AI technology GPT-4o. The data sources include YouTube video content with the following IDs: nSmkyDNulk (1,957 posts), MirzFk_DSiI (6,890 posts), and vgYi3Wr7v_g (1,936 posts). Understanding these responses aids in formulating precise objectives and aligning the analysis with the overall goals of the research. A thorough grasp of the context ensures the relevancy and accuracy of subsequent analytical stages. Ultimately, this approach lays a solid foundation for extracting meaningful insights from public sentiment toward GPT-4o.

Video 1



Video 2



Video 3

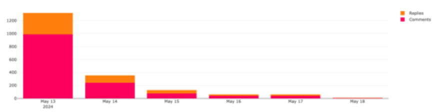


Figure 3. Post-Per-Day Statistic of Each Content

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 3 shows the statistics per day of the content video. Based on the post-per-day statistics from YouTube video content with IDs nSmkyDNulk (1,957 posts), MirzFk_DSiI (6,890 posts), and vgYi3Wr7v_g (1,936 posts), the volume of comments can be analyzed according to the dates following the launch of AI technology GPT-4o. This temporal distribution of posts provides valuable insights into the public's engagement and reaction trends over time. Examining these patterns makes it possible to discern peaks and fluctuations in public interest, which may correlate with specific events or announcements related to GPT-4o. Ultimately, understanding these dynamics aids in contextualizing the sentiment analysis within the broader timeline of the technology's reception.

2.2.2 Data Undesrtanding

In the data understanding phase, frequently used words in viewer comments are identified to gain insights into prevalent themes and topics. This analysis involves extracting and quantifying standard terms to reveal patterns in public discourse. Recognizing these frequently used words aids in understanding the audience's primary concerns, interests, and sentiments. Ultimately, this approach provides a foundational understanding of the data, guiding further analytical processes and enhancing the accuracy of sentiment analysis.

Video 1

Video 2

Video 3

Figure 4. Frequently used Words (Words Cloud)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 4 shows the frequently used words in the dataset. Identifying frequently used words reveals the key topics viewers highlight, providing a basis for in-depth analysis. This process uncovers recurring themes and focal points in the discourse, reflecting the audience's primary interests and concerns. Such insights are invaluable for a more detailed examination of sentiment and public opinion. Consequently, this understanding enables a nuanced exploration of viewer perspectives, enriching the overall analysis and contributing to a more comprehensive understanding of the public's reaction to the AI technology GPT-4o.

2.2.3 Modeling

The models employed in the sentiment extraction process are VADER and TextBlob, both renowned for their efficacy in natural language processing. VADER, an acronym for Valence Aware Dictionary and Sentiment Reasoner, excels in analyzing social media texts and understanding the intensity of sentiments. TextBlob, on the other hand, provides a versatile toolkit for sentiment analysis, offering both polarity and subjectivity metrics. Utilizing these models enhances the accuracy and depth of sentiment analysis, thereby ensuring robust and reliable results. Ultimately, this combination facilitates a comprehensive evaluation of public sentiment towards AI technology GPT-4.

Figure 5. Distribution of Polarity Value between TextBlob and Vader

Figure 5 shows the distribution of polarity value. Cohen's kappa statistic was determined by evaluating the distribution of polarity values, providing a measure of inter-rater reliability. The polarity values, representing the sentiment intensity of the text, are analyzed to assess consistency across different models or annotators. Cohen's kappa statistic, which adjusts for agreement occurring by chance, is crucial for validating the reliability of the sentiment analysis process. A high kappa value indicates strong agreement, reinforcing the robustness of the analytical methodology. Ultimately, this metric ensures the credibility and accuracy of the sentiment evaluation, contributing significantly to the overall research validity.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2.2.4 Evaluation

In the evaluation phase, sentiment classification results are categorized into negative, positive, and neutral based on Cohen's kappa statistic classification: fair, moderate, or poor. This classification process involves assessing the agreement between the predicted sentiment labels and the ground truth annotations. Cohen's kappa statistic provides a quantitative measure of the agreement beyond chance, thereby offering valuable insights into the reliability and accuracy of the sentiment classification model. By employing this classification framework, the evaluation phase ensures a rigorous assessment of sentiment classification performance, enhancing the credibility and trustworthiness of the research findings.

The "Fair" category in classification assessments denotes moderate agreement between two observers or models, albeit with notable room for improvement. Typically, a Cohen's kappa value ranging from 0.21 to 0.40 falls within this classification, indicating a reasonable but not entirely satisfactory level of concordance. While the agreement is discernible, there remains considerable variability in the assessments, suggesting potential avenues for refining the classification process. Consequently, the "Fair" category is a crucial indicator prompting further efforts to enhance agreement and reliability in classification tasks.

The "Moderate" category in classification assessments signifies a satisfactory agreement between two observers or models, albeit with notable judgment variations. Typically, Cohen's kappa values falling within the range of 0.41 to 0.60 are classified as "Moderate," indicating a discernible but not entirely consistent level of concordance. While the agreement is deemed adequate, there remains significant variability in assessments, suggesting areas for further refinement in classification processes. Consequently, the "Moderate" category is an important benchmark, prompting ongoing efforts to enhance agreement and reliability in classification tasks.

Within the realm of classification assessments, the "Poor" category indicates a low level of agreement between two observers or models, accompanied by significant variations in judgments. Cohen's kappa values below 0.20 are typically classified as "Poor," signifying a notable lack of consensus in assessments. Despite efforts to reach an agreement, the observed variability suggests fundamental challenges in the classification process, necessitating thorough review and potential revisions. Consequently, identifying a classification as "Poor" underscores the importance of addressing underlying issues to enhance agreement and reliability in future classification endeavors.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.2.5 Deployment

During the deployment phase, the sentiment classification results portray viewers' responses to the advancements in AI technology, specifically GPT-4. This phase involves translating the analytical insights obtained from sentiment classification into actionable outcomes or recommendations. By analyzing the sentiment of viewer responses, stakeholders can gauge public perception, identify areas of concern or enthusiasm, and tailor their strategies accordingly. Consequently, the deployment phase is pivotal in informing decision-making processes and shaping future developments in AI technology, ensuring alignment with public sentiments and preferences.

Through the toxicity score, classifications such as Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat can be discerned regarding video content of AI technology, specifically GPT-4o. This score is a quantitative measure of the level of harmful or offensive language present within the content, enabling a nuanced evaluation of its impact on viewers. By categorizing toxicity into distinct types, stakeholders can gain a comprehensive understanding of the nature and severity of potentially harmful content, facilitating targeted interventions or content moderation strategies. Consequently, leveraging the toxicity score enhances the ability to safeguard user experiences and maintain a constructive discourse surrounding AI technologies like GPT-4o.

3. RESULTS AND DISCUSSION

3.1 Toxicity Score

The toxicity score serves as a representative measure of viewer responses to GPT-4o video content, prompting this research to identify the toxicity score across three videos related to GPT-4o. This metric quantifies the degree of harmful or offensive language in the videos, providing valuable insights into viewer perceptions and reactions. By systematically assessing toxicity scores, researchers can gain a nuanced understanding of the content's impact on viewers and identify potential areas of concern or improvement. Consequently, this approach facilitates informed decision-making and content moderation strategies to ensure a positive and constructive viewer experience surrounding AI technologies like GPT-4o.

Figure 6 shows the toxicity score of the first video. Several vital insights emerge from the identification results of toxicity scores from the first video, where 1795 posts out of 1957 were analyzed. The toxicity score, representing the proportion of harmful or offensive language, indicates relatively low levels across various categories, including Toxicity (0.12131), Severe Toxicity (0.00860), Identity Attack (0.01691), Insult (0.06785), Profanity (0.07211), and Threat (0.01533). These values suggest that the overall level of toxicity within the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

comments is relatively modest, with the majority falling below 0.1, indicating a minimal presence of harmful content. However, it is noteworthy that some categories, such as Profanity and Insult, exhibit slightly higher scores, warranting further investigation into the nature and context of these expressions.

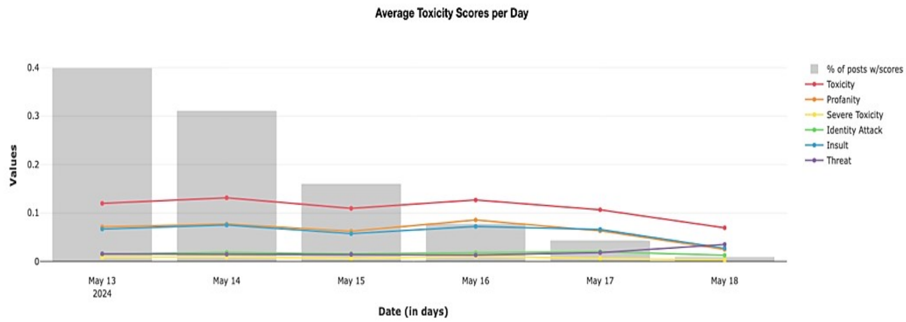


Figure 6. Toxicity Score of First Video

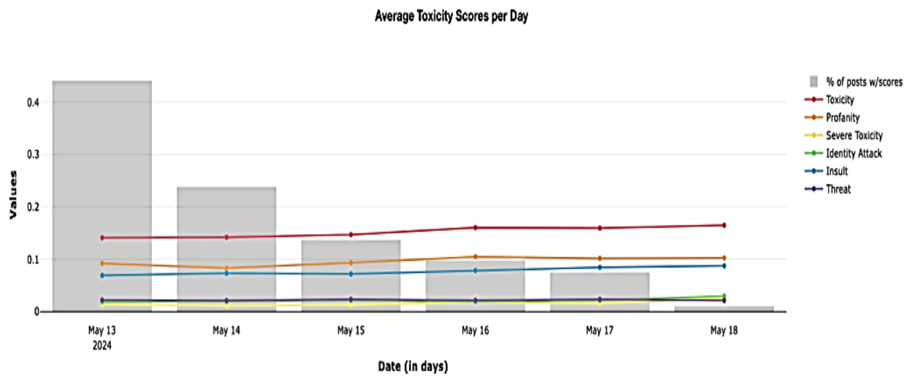


Figure 7. Toxicity Score of Second Video

Figure 7 shows the toxicity score of the second video. Based on the identification results of toxicity scores from the second video, where 6206 posts out of 6890 were analyzed, several vital insights emerge. The toxicity score, which measures the prevalence of harmful or offensive language, reveals moderate levels across various categories, including Toxicity (0.14589), Severe Toxicity (0.01434), Identity Attack (0.01973), Insult (0.07315), Profanity (0.09247), and Threat (0.02236). These values indicate a slightly higher toxic content than the first video analysis. While the overall toxicity levels remain relatively modest, the increased scores in categories like Profanity and Insult warrant attention and further scrutiny to understand the nature and context of these expressions.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

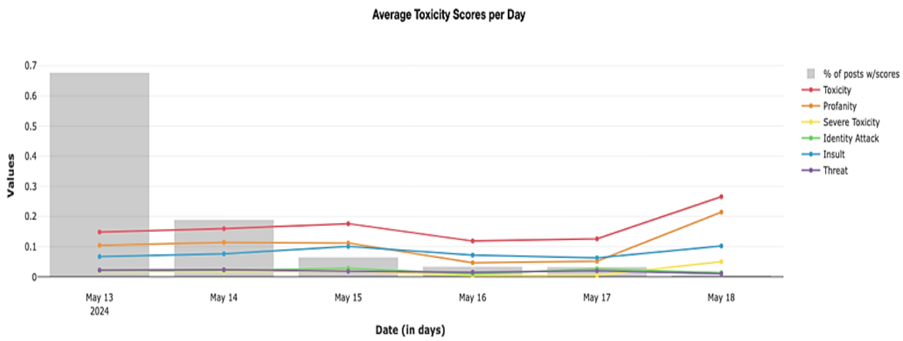


Figure 8. Toxicity Score of Third Video

Figure 8 shows the toxicity score of the third video. Several vital insights emerge from the identification results of toxicity scores from the third video, where 1711 posts out of 1936 were analyzed. The toxicity score, which measures the prevalence of harmful or offensive language, indicates moderate to high levels across various categories, including Toxicity (0.15114), Severe Toxicity (0.01793), Identity Attack (0.02161), Insult (0.07086), Profanity (0.10346), and Threat (0.02178). These values suggest a higher toxic content presence than the previous video analyses. The elevated scores in the Profanity and Identity Attack categories are particularly notable, indicating a potential escalation in harmful expressions.

3.2 Sentiment Classification

Implementing the Vader and TextBlob models reveals divergent outputs in sentiment classification across negative, neutral, and positive classes. The varying distribution of sentiment classifications underscores the nuanced nature of sentiment analysis and the influence of model choice on interpretation. Additionally, the fluctuating volume of data processed with Cohen's kappa statistic highlights the variability in agreement assessment between different classification scenarios. These disparities underscore the complexity of sentiment analysis tasks and emphasize the importance of comprehensive evaluation methodologies to accurately capture the nuances of sentiment expression.

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	1688	368 (21.80%)	537 (31.81%)	783 (46.39%)
TextBlob (English/EN)	1688	275 (16.29%)	723 (42.83%)	690 (40.88%)
TextBlob (French/FR)	12	0 (0.00%)	8 (66.67%)	4 (33.33%)
TextBlob (German/DE)	13	0 (0.00%)	11 (84.62%)	2 (15.38%)

Figure 9. Sentiment Classification of First Video

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 9 shows the toxicity score of the first video. Several key insights can be gleaned based on the sentiment classification results of the first video, comprising 1716 out of 1957 posts. The analysis, conducted using VADER and TextBlob models across multiple languages, reveals the distribution of negative, neutral, and positive sentiments among viewer comments. Notably, VADER identifies a higher proportion of posts with positive sentiment (46.39%) compared to negative (21.80%) and neutral (31.81%) sentiments. Conversely, TextBlob, particularly in English, demonstrates a more balanced distribution, with a slightly higher percentage of posts classified as neutral (42.83%) followed closely by positive (40.88%) and negative (16.29%) sentiments. Furthermore, the sentiment analysis across different languages, such as French and German, highlights variations in sentiment expression among diverse linguistic communities.

In the analysis of English language posts, VADER and TextBlob demonstrate agreement in categorizing 1040 (62.24%) out of 1671 posts, indicating a moderate level of concordance with a Cohen's kappa statistic of 0.409. The remaining posts exhibit varying levels of sentiment polarity, with 158 (15.19%) classified as negative, 383 (36.83%) as neutral, and 499 (47.98%) as positive. These findings highlight the utility of employing multiple sentiment analysis tools to understand sentiment distributions within text data comprehensively. While the agreement level between VADER and TextBlob is moderate, the nuanced distribution of sentiment across posts underscores the complexity inherent in analyzing viewer sentiments toward GPT-4o videos.

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	5791	1496 (25.83%)	1931 (33.34%)	2364 (40.82%)
TextBlob (English/EN)	5791	1191 (20.57%)	2415 (41.70%)	2185 (37.73%)
TextBlob (French/FR)	50	2 (4.00%)	46 (92.00%)	2 (4.00%)
TextBlob (German/DE)	60	0 (0.00%)	57 (95.00%)	3 (5.00%)

Figure 10. Sentiment Classification of Second Video

Figure 10 shows the toxicity score of the second video. Based on the sentiment classification results of the second video, where 5917 out of 6890 posts were analyzed, several notable findings emerge. The sentiment analysis using VADER and TextBlob reveals varying distributions of negative, neutral, and positive sentiments across different languages. Specifically, in English language posts, both VADER and TextBlob demonstrate similar patterns in sentiment categorization, albeit with slight variations. VADER identifies 25.83% of posts as negative, 33.34% as neutral, and 40.82% as positive, while TextBlob assigns 20.57% as negative, 41.70% as neutral, and 37.73% as positive.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In the analysis of English language posts, VADER and TextBlob demonstrate agreement in categorizing 3418 (60.08%) out of 5689 posts, indicating a fair level of concordance with a Cohen's kappa statistic of 0.389. The remaining posts exhibit varying levels of sentiment polarity, with 696 (20.36%) classified as negative, 1271 (37.19%) as neutral, and 1451 (42.45%) as positive. These findings underscore the importance of employing multiple sentiment analysis tools to understand sentiment distributions within text data comprehensively. While the agreement level between VADER and TextBlob is considered fair, the nuanced distribution of sentiment across posts highlights the complexity inherent in analyzing viewer sentiments towards GPT-4o videos.

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	1553	380 (24.47%)	573 (36.90%)	600 (38.63%)
TextBlob (English/EN)	1553	277 (17.84%)	701 (45.14%)	575 (37.03%)
TextBlob (French/FR)	12	0 (0.00%)	12 (100.00%)	0 (0.00%)
TextBlob (German/DE)	25	0 (0.00%)	24 (96.00%)	1 (4.00%)

Figure 11. Sentiment Classification of Third Video

Figure 11 shows the toxicity score of the third video. Several vital insights emerge based on the sentiment classification results of the third video, where 1593 out of 1936 posts were analyzed. The sentiment analysis using VADER and TextBlob reveals different distributions of negative, neutral, and positive sentiments across various languages. Specifically, in English language posts, VADER identifies 24.47% of posts as negative, 36.90% as neutral, and 38.63% as positive, while TextBlob categorizes 17.84% of posts as negative, 45.14% as neutral, and 37.03% as positive. Additionally, the analysis of posts in French and German shows that all French posts are neutral, and most German posts are neutral, with a small percentage being positive.

VADER and TextBlob agree on categorizing 929 (60.40%) out of 1538 English language posts, indicating a fair level of concordance with a Cohen's kappa statistic of 0.388. Within the analyzed posts, 182 (19.59%) exhibit negative sentiments with polarity scores of -0.05 or lower, 390 (41.98%) are neutral with polarity scores between -0.05 and 0.05, and 357 (38.43%) demonstrate positive sentiments with polarity scores of 0.05 or higher. This level of agreement highlights the importance of using multiple sentiment analysis tools to obtain a more comprehensive understanding of sentiment distributions. While the agreement between VADER and TextBlob is considered fair, the detailed distribution of sentiments across the posts provides valuable insights into the nuanced reactions of viewers to GPT-4o videos.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

4. CONCLUSION

The analysis of sentiment towards GPT-4o videos, utilizing VADER and TextBlob, reveals significant insights into viewer reactions. VADER and TextBlob demonstrated fair agreement in categorizing sentiments, with 929 (60.40%) out of 1538 English language posts classified consistently, reflected by a Cohen's kappa statistic of 0.388. Specifically, 182 (19.59%) of the posts were identified as having negative sentiments, 390 (41.98%) as neutral, and 357 (38.43%) as positive. These findings underscore the importance of employing multiple sentiment analysis tools to capture viewers' diverse and nuanced sentiments. The agreement level, although fair, highlights the inherent complexity in sentiment analysis, providing a comprehensive understanding of the public's response to the advancements in AI technology represented by GPT-4o.

REFERENCES

[1] M. S. Kannelønning, "Contesting futures of Artificial Intelligence (AI) in healthcare: formal expectations meet informal anticipations," *Technol. Anal. Strateg. Manag.*, pp. 1–12, 2023, doi: 10.1080/09537325.2023.2226243.

[2] P. Apell and H. Eriksson, "Artificial intelligence (AI) healthcare technology innovations: the current state and challenges from a life science industry perspective," *Technol. Anal. Strateg. Manag.*, vol. 35, no. 2, pp. 179–193, 2023, doi: 10.1080/09537325.2021.1971188.

[3] A. Adikari, N. Hernandez, D. Alahakoon, M. L. Rose, and J. E. Pierce, "From concept to practice: a scoping review of the application of AI to aphasia diagnosis and management," *Disabil. Rehabil.*, vol. 46, no. 7, pp. 1288–1297, 2024, doi: 10.1080/09638288.2023.2199463.

[4] E. Monod, R. Lissillour, A. Köster, and Q. Jiayin, "Does AI control or support? Power shifts after AI system implementation in customer relationship management," *J. Decis. Syst.*, vol. 32, no. 3, pp. 542–565, 2023, doi: 10.1080/12460125.2022.2066051.

[5] P. Kovalishin *et al.*, "Using Artificial Intelligence (AI) methods for effectively responding to climate change at marine ports," *J. Int. Marit. Safety, Environ. Aff. Shipp.*, vol. 7, no. 1, 2023, doi: 10.1080/25725084.2023.2186589.

[6] M. U. Modén, M. Ponti, J. Lundin, and M. Tallvid, "When fairness is an abstraction: Equity and AI in Swedish compulsory education," *arXiv Prepr. arXiv2311.01838*, no. May, pp. 0–2, 2023, doi: 10.1080/00313831.2024.2349908.

[7] E. H. Park, K. Werder, L. Cao, and B. Ramesh, "Why do Family Members Reject AI in Health Care? Competing Effects of Emotions," *J. Manag. Inf. Syst.*, vol. 39, no. 3, pp. 765–792, 2022, doi: 10.1080/07421222.2022.2096550.

[8] M. M. Bishnoi, S. Ramakrishnan, S. Suraj, and A. Dwivedi, "Impact of AI

and COVID-19 on manufacturing systems: An Asia Pacific Perspective on the two Competing exigencies," *Prod. Manuf. Res.*, vol. 11, no. 1, 2023, doi: 10.1080/21693277.2023.2236684.
[9] D. Trusilo, "Autonomous AI Systems in Conflict: Emergent Behavior and Its Impact on Predictability and Reliability," *J. Mil. Ethics*, vol. 22, no. 1, pp. 2–17, 2023, doi: 10.1080/15027570.2023.2213985.
[10] H. Sayyed, "Artificial intelligence and criminal liability in India : exploring legal implications and challenges," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2343195.
[11] S. Matilda Bez and H. Chesbrough, "Competitor Collaboration Before a Crisis: What the AI Industry Can LearnThe Partnership on AI can use the Dynamic Capabilities Framework and lessons from other industries to proactively identify AI risks and create solutions.," *Res. Technol. Manag.*, vol. 63, no. 3, pp. 42–48, 2020, doi: 10.1080/08956308.2020.1733889.
[12] J. A. Moldt, T. Festl-Wietek, A. Madany Mamlouk, K. Nieselt, W. Fuhl, and A. Herrmann-Werner, "Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots," *Med. Educ. Online*, vol. 28, no. 1, 2023, doi: 10.1080/10872981.2023.2182659.
[13] Y. Ikkatai, T. Hartwig, N. Takanashi, and H. M. Yokoyama, "Octagon Measurement: Public Attitudes toward AI Ethics," *Int. J. Hum. Comput. Interact.*, vol. 38, no. 17, pp. 1589–1606, 2022, doi: 10.1080/10447318.2021.2009669.
[14] R. O'Connor, M. Bolton, A. K. Saeri, T. Chan, and R. Pearson, "Artificial intelligence and complex sustainability policy problems: translating promise into practice," *Policy Des. Pract.*, vol. 0, no. 0, pp. 1–16, 2024, doi: 10.1080/25741292.2024.2348834.
[15] Y. M. Wang, C. L. Wei, H. H. Lin, S. C. Wang, and Y. S. Wang, "What drives students' AI learning behavior: a perspective of AI anxiety," *Interact. Learn. Environ.*, vol. 0, no. 0, pp. 1–17, 2022, doi: 10.1080/10494820.2022.2153147.
[16] E. Häglund and J. Björklund, "AI-Driven Contextual Advertising: Toward Relevant Messaging Without Personal Data," *J. Curr. Issues Res. Advert.*, vol. 0, no. 0, pp. 1–19, 2024, doi: 10.1080/10641734.2024.2334939.
[17] C. Gahnberg, "What rules? Framing the governance of artificial agency," *Policy Soc.*, vol. 40, no. 2, pp. 194–210, 2021, doi: 10.1080/14494035.2021.1929729.
[18] P. Alilunas, "What we must be: AI and the future of porn studies," *Porn Stud.*, vol. 11, no. 1, pp. 99–112, 2024, doi: 10.1080/23268743.2024.2312181.
[19] J. M. White and R. Lidskog, "Ignorance and the regulation of artificial intelligence," *J. Risk Res.*, vol. 25, no. 4, pp. 488–500, 2022, doi: 10.1080/13669877.2021.1957985.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[20] I. Z. P. Hamdan and M. Othman, "Predicting Customer Loyalty Using Machine Learning for Hotel Industry," *J. Soft Comput. Data Min.*, vol. 3, no. 2, pp. 31–42, 2022.
[21] A. Khumaidi, "Data Mining for Predicting the Amount of Coffee Production Using Crisp-Dm Method," *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 1–8, 2020, doi: 10.33480/techno.v17i1.1240.
[22] N. Saleem, T. Mufti, S. S. Sohail, and D. Ø. Madsen, "ChatGPT as an innovative heutagogical tool in medical education," *Cogent Educ.*, vol. 11, no. 1, p., 2024, doi: 10.1080/2331186X.2024.2332850.
[23] B. M. Colosimo, E. del Castillo, L. A. Jones-Farmer, and K. Paynabar, "Artificial intelligence and statistics for quality technology: an introduction to the special issue," *J. Qual. Technol.*, vol. 53, no. 5, pp. 443–453, 2021, doi: 10.1080/00224065.2021.1987806.
[24] C. Connolly *et al.*, "Artificial Intelligence in Interprofessional Healthcare Practice Education–Insights from the Home Health Project, an Exemplar for Change," *Comput. Sch.*, vol. 40, no. 4, pp. 412–429, 2023, doi: 10.1080/07380569.2023.2247393.
[25] D. Belanche, R. W. Belk, L. V. Casaló, and C. Flavián, "The dark side of artificial intelligence in services," *Serv. Ind. J.*, vol. 44, no. 3–4, pp. 149–172, 2024, doi: 10.1080/02642069.2024.2305451.
[26] G. Papyshev and M. Yarime, "The state's role in governing artificial intelligence: development, control, and promotion through national strategies," *Policy Des. Pract.*, vol. 6, no. 1, pp. 79–102, 2023, doi: 10.1080/25741292.2022.2162252.
[27] J. Hautala and T. Ahlqvist, "Integrating futures imaginaries, expectations and anticipatory practices: practitioners of artificial intelligence between now and future," *Technol. Anal. Strateg. Manag.*, pp. 1–13, 2022, doi: 10.1080/09537325.2022.2130041.
[28] M. Treve, "What COVID-19 has introduced into education: challenges Facing Higher Education Institutions (HEIs)," *High. Educ. Pedagog.*, vol. 6, no. 1, pp. 212–227, 2021, doi: 10.1080/23752696.2021.1951616.
[29] B. Wise, L. Emerson, A. Van Luyn, B. Dyson, C. Bjork, and S. E. Thomas, "A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI," *High. Educ. Res. Dev.*, vol. 43, no. 3, pp. 578–590, 2024, doi: 10.1080/07294360.2023.2280195.
[30] P. Sepehr, "Mundane Urban Governance and AI Oversight: The Case of Vienna's Intelligent Pedestrian Traffic Lights," *J. Urban Technol.*, vol. 0, no. 0, pp. 1–18, 2024, doi: 10.1080/10630732.2024.2302280.
[31] F. Naz, A. Kumar, R. Agrawal, J. A. Garza-Reyes, A. Majumdar, and H. Chokshi, "Artificial intelligence as an enabler of quick and effective production repurposing: an exploratory review and future research propositions," *Prod. Plan. Control*, vol. 0, no. 0, pp. 1–24, 2023, doi: 10.1080/09537287.2023.2248947.
[32] Y. Pan, F. Froese, N. Liu, Y. Hu, and M. Ye, "The adoption of artificial

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

intelligence in employee recruitment: The influence of contextual factors," *Int. J. Hum. Resour. Manag.*, vol. 33, no. 6, pp. 1125–1147, 2022, doi: 10.1080/09585192.2021.1879206.

[33] A. Jaiswal, C. J. Arun, and A. Varma, "Rebooting employees: upskilling for artificial intelligence in multinational corporations," *Int. J. Hum. Resour. Manag.*, vol. 33, no. 6, pp. 1179–1208, 2022, doi: 10.1080/09585192.2021.1891114.

[34] G. Lai, C. Dunlap, A. Gluskin, W. B. Nehme, and A. A. Azim, "Artificial Intelligence in Endodontics," *J. Calif. Dent. Assoc.*, vol. 51, no. 1, 2023, doi: 10.1080/19424396.2023.2199933.

[35] A. Calderaro and S. Blumfelde, "Artificial intelligence and EU security: the false promise of digital sovereignty?," *Eur. Secur.*, vol. 31, no. 3, pp. 415–434, 2022, doi: 10.1080/09662839.2022.2101885.

[36] Y. Liu, J. Sun, Z. Zhang, M. Wu, H. Sima, and Y. M. Ooi, "How AI Impacts Companies' Dynamic Capabilities: Lessons from Six Chinese Construction Firms," *Res. Technol. Manag.*, vol. 67, no. 3, pp. 64–76, 2024, doi: 10.1080/08956308.2024.2324407.

[37] S. Kelly, S. A. Kaye, K. M. White, and O. Oviedo-Trespalacios, "Clearing the way for participatory data stewardship in artificial intelligence development: a mixed methods approach," *Ergonomics*, vol. 66, no. 11, pp. 1782–1799, 2023, doi: 10.1080/00140139.2023.2289864.

[38] I. Maskanah, A. Primajaya, and A. Rizal, "Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran," *J. INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, pp. 218–228, 2020, doi: 10.35314/isi.v5i2.1443.

[39] Y. A. Singgalen, "Social Network Analysis and Sentiment Classification of Extended Reality Product Content," *Klik Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2197–2208, 2024, doi: 10.30865/klik.v4i4.1710.

[40] Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 237–248, 2023, doi: 10.30865/json.v5i2.7081.