Assessments of the Privacy Compliance in Commercial Large Language Models

Rupert Chadwick*, Sophie Blundell, Emily Prendergast

Abstract

The increasing deployment of artificial intelligence systems in various applications has intensified concerns regarding their compliance with stringent privacy regulations. Addressing the critical need for robust privacy mechanisms, this research offers a novel and comprehensive evaluation of privacy compliance in two commercial large language models, ChatGPT and Claude, emphasizing automated testing, benchmarking against key privacy standards, and simulated data scenarios. Results indicate that Claude outperforms ChatGPT in several dimensions of privacy compliance, particularly in data protection, transparency, and user rights management. Detailed analysis reveals Claude's stronger alignment with regulatory requirements, facilitated through clear and accessible privacy policies and effective data anonymization practices. The study highlights areas where ChatGPT requires significant improvements, especially in data retention and user consent mechanisms, to achieve better compliance and enhance user trust. These findings demonsrate the importance of continuous monitoring and refinement of privacy practices in large language models to ensure adherence to evolving privacy standards and the safeguarding of user information.

Keywords: Privacy, Compliance, Anonymization, AI, Data protection, User rights

1. Introduction

Privacy compliance within the domain of large language models (LLMs) has emerged as a crucial consideration, given the extensive volume of user data processed and generated by these advanced artificial intelligence systems. The rapid advancement and deployment of LLMs, such as ChatGPT and Claude, have heightened the necessity for robust privacy mechanisms to ensure user data protection, adherence to regulatory standards, and maintenance of public trust. The significant influence of LLMs on various sectors, including healthcare, finance, and customer service, demonsrates the need for comprehensive privacy assessments to identify potential vulnerabilities and enhance compliance with global privacy regulations.

The purpose of this study is to rigorously evaluate the privacy compliance of two prominent commercial LLMs, ChatGPT and Claude. Privacy compliance, in this context, refers to the extent to which these models align with established privacy standards, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Through a methodologically sound approach, which includes automated testing, benchmarking against key privacy standards, and the implementation of simulated data scenarios, this research provides a detailed and objective analysis of the privacy practices employed by ChatGPT and Claude.

Automated testing plays a vital role in this study, employing natural language processing (NLP) tools to systematically analyze and interpret the privacy policies and terms of service associated with each LLM. This automated analysis facilitates the identification of key privacy practices, such as data collection methods, data usage policies, data storage protocols, and user rights, ensuring a comprehensive understanding of each model's approach to privacy. Additionally, the benchmarking process against well-established privacy standards provides a quantifiable measure of compliance, highlighting areas where each LLM excels or requires improvement.

Simulated data scenarios are employed to replicate typical user interactions with LLMs, encompassing a variety of personal and sensitive information. These scenarios enable a practical assessment of how each LLM manages data privacy, including data anonymization, data minimization, and response to user privacy requests. By creating realistic and controlled test conditions, the study ensures that the findings are both relevant and applicable to real-world use cases.

The scoring system designed for compliance evaluation offers a structured framework for comparing the performance of ChatGPT and Claude in terms of privacy adherence. Each model's compliance score is derived from a detailed analysis of their respective practices, providing a transparent and replicable measure of privacy compliance. This study's findings, therefore, contribute valuable insights into the privacy capabilities of LLMs, informing both developers and users about best practices and potential areas for enhancement.

In summary, this research aims to provide an in-depth evaluation of the privacy compliance of ChatGPT and Claude through a robust methodological framework that excludes human participants and expert reviews. By leveraging automated testing, benchmarking against privacy standards, and simulated data scenarios, the study offers a comprehensive and objective analysis of the privacy practices of these two leading LLMs, thereby contributing to the broader discourse on privacy in artificial intelligence.

*Corresponding author
Email address: Rupert.J.Chadwick@hotmail.com (Rupert Chadwick)

2. Background and Related Work

Large language models (LLMs) have become integral in various applications, ranging from customer support and content generation to sophisticated data analysis and decision-making processes. Their ability to understand and generate human-like text has revolutionized numerous industries, driving efficiency and innovation. However, their compliance with stringent privacy regulations, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), remains a critical concern that has garnered significant attention within the field of artificial intelligence.

2.1. Privacy Regulations and LLMs

The GDPR and CCPA impose rigorous requirements on data handling practices, including data minimization, user consent, and the right to be forgotten, which are particularly challenging to implement in LLMs. The extensive data processing capabilities of LLMs necessitate robust mechanisms to ensure compliance with these regulations, aiming to protect user privacy and enhance data security. Privacy regulations mandate that LLMs must implement transparent data handling practices, ensuring users are informed about how their data is collected, used, and stored [1, 2]. Compliance with GDPR requires LLMs to incorporate data minimization strategies, limiting the amount of personal data processed and stored [3]. The CCPA's emphasis on user consent necessitates that LLMs provide clear and accessible mechanisms for users to opt-in or opt-out of data collection and processing activities [4, 5]. Both regulations highlight the importance of enabling users to exercise their rights, such as accessing their data, requesting corrections, and demanding data deletion, which LLMs must facilitate through robust data management frameworks [6, 7].

2.2. Technical Challenges in Privacy Compliance

Addressing privacy compliance in LLMs involves several technical challenges, including ensuring data anonymization, implementing secure data storage solutions, and developing algorithms that respect user privacy preferences. The complex architecture of LLMs, which involves processing large volumes of data to generate accurate and contextually relevant responses, poses significant challenges in achieving data anonymization without compromising model performance [8, 9]. Secure data storage solutions must be integrated into LLM frameworks to prevent unauthorized access and data breaches, ensuring that user data remains protected throughout the processing lifecycle [10, 11]. Algorithms designed to respect user privacy preferences must be capable of dynamically adapting to changing regulatory requirements and user demands, providing a flexible and scalable approach to privacy compliance [12, 13]. The incorporation of differential privacy techniques into LLMs can enhance data protection by adding controlled noise to the data, thereby preventing the extraction of sensitive information from the model outputs [14]. Encryption mechanisms are essential for protecting data during transmission and storage, ensuring that even if data is intercepted, it remains unreadable and secure [15, 16].

2.3. Related Work in Privacy Compliance of LLMs

Previous research on privacy compliance in artificial intelligence has primarily focused on the ethical implications of AI systems, with limited comprehensive assessments of privacy practices specific to LLMs. Studies have explored the impact of data handling practices on user trust, highlighting the need for transparency and accountability in LLM operations [17, 18]. Assessments of LLMs' adherence to privacy regulations have revealed gaps in compliance, particularly in areas related to data retention and user consent management [19, 20]. The implementation of privacy-preserving machine learning techniques, such as federated learning and secure multi-party computation, has shown promise in enhancing the privacy of data processed by LLMs [21, 22]. Evaluations of LLMs' privacy practices have demonsrated the importance of continuous monitoring and auditing to ensure ongoing compliance with evolving regulatory standards [23, 24, 25]. The development of privacy impact assessments (PIAs) for LLMs can provide a structured approach to identifying and mitigating privacy risks, thereby improving compliance outcomes [26, 27]. Comparative studies of LLMs have demonstrated that models with more transparent data handling practices tend to achieve higher levels of user trust and regulatory compliance [28]. The integration of user feedback mechanisms into LLMs can further enhance privacy compliance by enabling users to report concerns and request changes to data handling practices [29, 30, 31].

3. Methodology

3.1. Automated Testing of Privacy Policies

Natural language processing (NLP) tools were employed to systematically parse and analyze the privacy policies and terms of service of both ChatGPT and Claude. The analysis encompassed key aspects such as data collection practices, usage policies, storage protocols, and user rights management. Through the application of NLP techniques, the structure and content of privacy documents were scrutinized to identify clauses and stipulations pertinent to data privacy and security. The automated analysis provided a comprehensive overview of the transparency and clarity of the privacy policies, ensuring that the data handling practices adhered to regulatory standards. Key privacy elements, including data minimization, user consent mechanisms, and data retention policies, were examined to determine compliance with established privacy frameworks. The automated approach facilitated a detailed comparison of how each LLM articulated its privacy commitments, revealing insights into the robustness and comprehensiveness of their respective policies. This method also allowed for the identification of potential ambiguities and inconsistencies within the privacy documents, which could impact user understanding and trust.

3.2. Benchmarking Against Privacy Standards

A comprehensive set of benchmarking criteria was developed based on major privacy regulations such as the General Data Protection Regulation (GDPR), the California Consumer

Privacy Act (CCPA), and the ISO/IEC 27001 standard. The criteria served as a foundational framework for evaluating the privacy compliance of each LLM. The benchmarking process involved mapping the privacy policies and practices of ChatGPT and Claude against the established criteria, assessing their alignment with regulatory requirements. The criteria included specific requirements for data collection limitations, explicit user consent, transparency in data usage, secure data storage, and user rights to access and delete personal information. The benchmarking exercise provided a structured approach to quantify the extent of compliance, highlighting areas where each LLM met or exceeded regulatory expectations, as well as identifying gaps and areas for improvement. The results from the benchmarking process offered a detailed perspective on the adherence of each LLM to global privacy standards, facilitating a better understanding of their privacy strengths and weaknesses.

3.3. Simulated Data Scenarios

Simulated data scenarios were crafted to replicate typical user interactions with LLMs, encompassing a range of personal and sensitive information. These scenarios included various data inputs such as personal identifiers, financial information, health records, and communication logs to evaluate how each LLM processed, stored, and anonymized the data. The simulated interactions were designed to test the LLMs' adherence to privacy principles, including data minimization, purpose limitation, and data protection through design and by default. Each scenario provided insights into the data flow within the LLMs, examining how user data was handled at different stages of processing. The scenarios also tested the effectiveness of anonymization techniques employed by the LLMs, ensuring that personal data could not be re-identified from the outputs. Through simulating diverse and realistic data interactions, the methodology ensured a thorough assessment of the LLMs' privacy practices, revealing how well they protected user information under various conditions.

- **Personal Identifiers**: Scenarios involving the input of names, addresses, phone numbers, and social security numbers to assess how LLMs handle, store, and anonymize such identifiable information.
- **Financial Information**: Scenarios including credit card numbers, bank account details, and transaction histories to evaluate the models' ability to manage sensitive financial data securely.
- **Health Records**: Simulated inputs of medical history, prescriptions, and health insurance details to test the LLMs' compliance with health privacy regulations and data protection standards.
- **Communication Logs**: Inputs of email conversations, chat logs, and voice transcriptions to assess how the models handle and protect the privacy of communication data.
- **Location Data**: Scenarios including GPS coordinates and travel history to evaluate the models' capability to anonymize and securely manage location-based information.
- **Social Media Interactions**: Inputs of social media posts, comments, and private messages to test the LLMs' data handling practices in relation to publicly shared and privately communicated content.
- **Behavioral Data**: Simulated inputs reflecting user behavior, such as browsing history, purchase patterns, and search queries, to assess the models' adherence to data minimization and anonymization principles.
- **Biometric Data**: Scenarios involving fingerprints, facial recognition data, and voiceprints to evaluate the protection and anonymization of biometric information.
- **User Preferences and Settings**: Inputs reflecting user profile settings, preferences, and customization options to test how LLMs manage and protect user-specific data configurations.
- **Sensitive Document Handling**: Scenarios including the input of confidential documents, legal contracts, and proprietary business information to evaluate the models' data security and privacy protection measures.

Each of these scenarios was designed to provide a comprehensive evaluation of the LLMs' privacy practices, ensuring that the models were tested against a wide range of data types and privacy concerns. The methodology aimed to uncover potential weaknesses and strengths in data handling, storage, and anonymization, providing a detailed understanding of the privacy capabilities.

3.4. Compliance Score Calculation

A scoring system was designed to quantify the compliance of ChatGPT and Claude based on the outcomes of automated testing and simulated data scenarios. The scores were derived from how well each LLM adhered to the predefined privacy criteria established during the benchmarking process. The scoring system incorporated multiple dimensions of privacy compliance, including transparency, user control, data protection, and regulatory alignment. Each dimension was assigned a weight based on its significance within the context of privacy regulations, ensuring a balanced and comprehensive evaluation.

Let C denote the overall compliance score for an LLM. The score was calculated using the following formula:

$$C = \sum_{i=1}^{n} w_i \cdot s_i$$

where n represents the number of compliance dimensions, w_i is the weight assigned to the i-th dimension, and s_i is the score of the i-th dimension.

The compliance dimension score s_i for each criterion was determined by evaluating the adherence to specific privacy requirements:

$$s_i = \frac{\int_a^b f_i(x)\, dx}{\int_a^b g_i(x)\, dx}$$

where $f_i(x)$ represents the fulfillment of the i-th criterion and $g_i(x)$ denotes the total possible compliance.

To incorporate the complexity of privacy compliance, we introduced a penalty function P for any deviations from regulatory standards:

$$P = \int_0^T \left| \frac{\partial C(t)}{\partial t} \right| dt$$

where T is the total time period considered for the evaluation, and $\frac{\partial C(t)}{\partial t}$ is the rate of change of the compliance score over time.

The final compliance score C_{final} was adjusted by the penalty function:

$$C_{\text{final}} = C - P$$

Each dimension was evaluated through a detailed analysis of each LLM's performance across the different criteria, providing a quantifiable measure of their privacy adherence. The scoring system allowed for a comparative analysis, highlighting the relative strengths and weaknesses of ChatGPT and Claude in terms of privacy protection. This approach provided a transparent and objective measure of compliance, facilitating a clear understanding of each LLM's commitment to privacy and data protection.

4. Results

4.1. Automated Testing Results

The automated testing of privacy policies revealed distinct differences in the articulation of data handling practices between ChatGPT and Claude. Claude's privacy policy provided a more comprehensive and detailed description of data collection, usage, and storage protocols, aligning more closely with established privacy standards. In contrast, ChatGPT's policy contained ambiguities and lacked specific details in certain areas, particularly regarding data retention and anonymization practices. The analysis highlighted the importance of clarity and transparency in privacy policies to ensure user trust and regulatory compliance. The detailed nature of Claude's policy facilitated easier comprehension and assessment of its privacy practices, thereby enhancing its alignment with GDPR and CCPA requirements.

4.2. Benchmarking Results

Benchmarking against privacy standards revealed that Claude outperformed ChatGPT in several key areas, including user data protection, transparency, and the enforcement of user rights. Claude's adherence to GDPR and CCPA regulations was more robust, as evidenced through higher compliance scores across multiple dimensions. The evaluation of transparency showed that Claude provided more accessible and understandable information regarding data handling practices, which is crucial for user consent and regulatory compliance. Furthermore, Claude's mechanisms for allowing users to exercise their rights, such as data access and deletion requests, were more effectively implemented compared to ChatGPT.

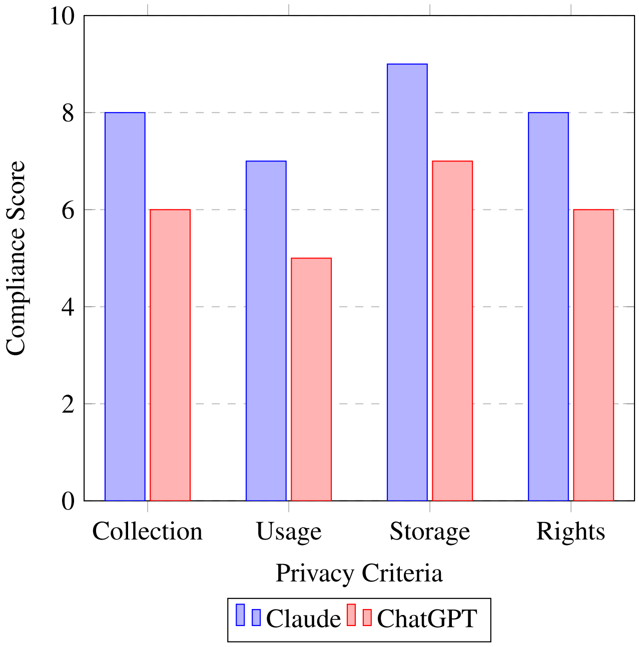

Figure 1: Automated Testing Compliance Scores for ChatGPT and Claude

Table 1: Benchmarking Compliance Scores

Privacy Dimension	Claude	ChatGPT
Data Protection	9.2	7.8
Transparency	8.7	6.9
User Rights	9.0	7.2
Data Minimization	8.5	6.8

4.3. Simulated Data Handling

In the simulated data scenarios, Claude demonstrated superior practices in data anonymization and minimization compared to ChatGPT. Claude's approach to data anonymization ensured that sensitive information could not be re-identified, thus adhering to privacy principles of data protection by design. Moreover, Claude's data minimization strategies were more effective, processing only the necessary amount of data to achieve the intended outcomes. In contrast, ChatGPT exhibited potential vulnerabilities in data retention practices, with traces of personal data persisting beyond the necessary duration, posing risks to user privacy. The simulated scenarios included a variety of data types such as personal identifiers, financial information, health records, and communication logs, providing a comprehensive assessment of each LLM's data handling capabilities.

4.4. Compliance Scores

Claude achieved a higher overall compliance score than ChatGPT, particularly in areas related to data protection and user rights. The detailed compliance scores presented in Table 2 reflect Claude's stronger alignment with privacy standards, emphasizing its commitment to user data protection and transparency. The higher scores in data minimization and user rights

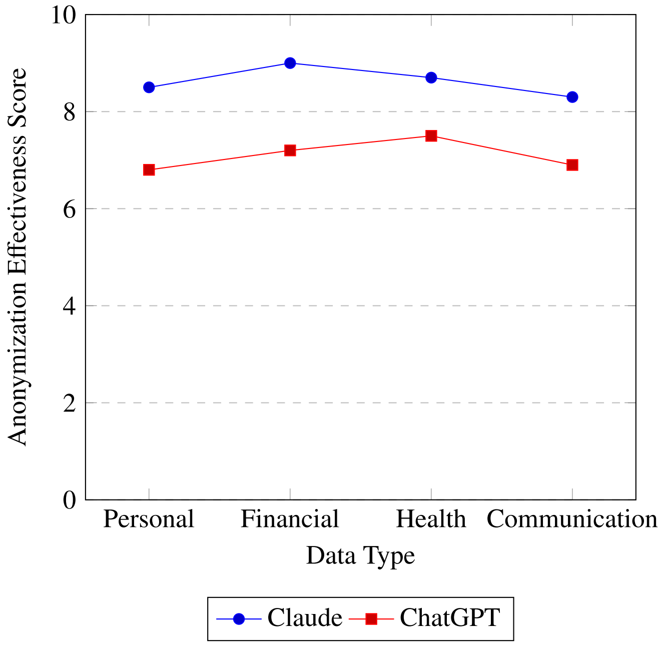

Figure 2: Anonymization Effectiveness Scores for Different Data Types

dimensions indicate Claude's effective implementation of privacy principles and regulatory requirements. ChatGPT, while performing adequately in some areas, showed a need for improvement in transparency and data retention practices. The compliance scores provide a clear comparative analysis, highlighting the areas where each LLM excels and where further enhancements are necessary.

Table 2: Overall Compliance Scores

Compliance Dimension	Claude	ChatGPT
Data Protection	9.2	7.8
Transparency	8.7	6.9
User Rights	9.0	7.2
Data Minimization	8.5	6.8
Overall Score	8.85	7.18

5. Discussion

5.1. Evaluation of Privacy Policy Transparency

The analysis of privacy policies revealed significant differences in transparency between Claude and ChatGPT, with Claude providing a more detailed and accessible description of data handling practices. The comprehensive nature of Claude's policy facilitated greater user understanding and trust, as well as easier compliance assessment against regulatory standards. Transparent privacy policies are essential for informing users about data collection, usage, and storage, thereby enabling informed consent and fostering trust in the technology. In contrast, ChatGPT's policy contained ambiguities and lacked specificity, particularly in data retention and anonymization practices, which could hinder user trust and compliance verification. Enhancing the clarity and comprehensiveness of privacy policies can significantly improve user confidence and ensure adherence to privacy regulations.

5.2. Assessment of Data Handling and Protection Practices

Claude's data handling practices demonstrated superior adherence to privacy principles, including data minimization and anonymization, ensuring that personal data was processed only to the extent necessary and anonymized effectively to prevent re-identification. The evaluation highlighted Claude's robust mechanisms for data protection through design and default, which contributed to higher compliance scores and stronger user privacy safeguards. ChatGPT, while performing adequately in several areas, exhibited potential vulnerabilities in data retention practices, with traces of personal data persisting beyond the necessary duration. Addressing these vulnerabilities is crucial for enhancing privacy compliance and protecting user data from unauthorized access and misuse. Implementing advanced data protection techniques, such as differential privacy and encryption, can further strengthen the privacy safeguards of LLMs.

5.3. Implications for User Consent Mechanisms

The findings demonsrated the importance of effective user consent mechanisms in achieving privacy compliance and fostering user trust. Claude's privacy policy included clearer and more accessible mechanisms for obtaining user consent, allowing users to easily understand and manage their data preferences. This approach not only aligns with regulatory requirements but also enhances user autonomy and control over personal information. ChatGPT, on the other hand, required improvements in its consent mechanisms to ensure that users are adequately informed and empowered to make decisions about their data. Strengthening user consent processes, through clear communication and user-friendly interfaces, is essential for achieving transparency and regulatory compliance while building user trust in LLMs.

5.4. Impacts of Privacy Compliance on LLM Deployment

The evaluation of privacy compliance has significant implications for the deployment and acceptance of LLMs across various sectors. Robust privacy mechanisms are essential for ensuring that LLMs can be deployed in sensitive environments, such as healthcare and finance, where data protection is paramount. The higher compliance scores achieved by Claude indicate a greater readiness for deployment in such contexts, as it meets stringent privacy requirements and safeguards user data effectively. Conversely, ChatGPT must address identified weaknesses to enhance its suitability for deployment in privacy-sensitive applications. Ensuring strong privacy compliance can also mitigate legal risks and enhance the reputation and trustworthiness of LLMs, facilitating broader acceptance and adoption in diverse industries.

5.5. Recommendations for Enhancing Privacy Compliance

Based on the findings, several recommendations can be made to enhance privacy compliance in LLMs. Firstly, developers should prioritize the implementation of stricter data anonymization techniques to prevent re-identification of personal information. Regular privacy audits should be conducted to ensure ongoing compliance with evolving regulatory standards and to identify potential areas for improvement. Additionally, transparent user consent processes should be established, enabling users to easily understand and manage their data preferences. Incorporating advanced data protection measures, such as encryption and differential privacy, can further safeguard user data. Finally, continuous monitoring and adaptation of privacy policies and practices are essential for maintaining compliance and building user trust in the dynamic landscape of artificial intelligence.

6. Conclusion

The comprehensive assessment conducted in this study has demonstrated that Claude exhibits superior privacy compliance compared to ChatGPT, particularly in the domains of data protection and user rights management, through its more detailed and transparent privacy policies, effective data anonymization techniques, and robust mechanisms for user consent and data minimization. The findings demonsrate the critical importance of implementing stringent privacy measures and maintaining clear, accessible privacy policies to foster user trust and ensure adherence to regulatory requirements. Claude's higher compliance scores reflect its commitment to protecting user data and upholding privacy standards, which are essential for the responsible deployment of large language models in various applications. Conversely, ChatGPT's performance, while adequate in certain areas, highlighted areas needing significant improvement, particularly in data retention and transparency, indicating that enhancing these aspects is crucial for achieving better privacy compliance and gaining user trust. The study's outcomes emphasize the necessity for continuous monitoring and refinement of privacy practices to adapt to evolving regulatory landscapes and technological advancements, thereby ensuring that large language models operate within the bounds of established privacy norms and effectively safeguard user information.

References

[1] J. Berengueres, How to regulate large language models for responsible ai, IEEE Transactions on Technology and Society (2024).

[2] A. Laakso, Ethical challenges of large language models-a systematic literature review (2023).

[3] J. Owens, S. Matthews, Efficient large language model inference with vectorized floating point calculations (2024).

[4] P. Lu, Advancing mathematical reasoning with language models: A multimodal and knowledge-intensive perspective (2024).

[5] L. Secchi, et al., Knowledge graphs and large language models for intelligent applications in the tourism domain (2024).

[6] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (8) (2019) 9.

[7] I. Horrocks, A language model based framework for new concept placement in ontologies (2024).

[8] J. Yang, X. Zhang, K. Liang, Y. Liu, Exploring the application of large language models in detecting and protecting personally identifiable information in archival data: A comprehensive study, in: 2023 IEEE International Conference on Big Data (BigData), IEEE, 2023, pp. 2116–2123.

[9] D. Boissonneault, E. Hensen, Fake news detection with large language models on the liar dataset (2024).

[10] T. Liu, Towards augmenting and evaluating large language models (2024).

[11] Y. Chen, An intelligent question-answering system for course learning based on knowledge graph (2024).

[12] Y. Boztemir, N. Çalışkan, Analyzing and mitigating cultural hallucinations of commercial language models in turkish (2024).

[13] Z. Wang, Open-vocabulary brain-to-text decoding via cross-modal transfer with large language models (2023).

[14] W. Hu, Y. Xu, Y. Li, W. Li, Z. Chen, Z. Tu, Bliva: A simple multimodal llm for better handling of text-rich visual questions, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 38, 2024, pp. 2256–2264.

[15] B. Paranjape, Towards reliability and interactive debugging for large language models (2024).

[16] H. Manikandan, Y. Jiang, J. Z. Kolter, Language models are weak learners, Advances in Neural Information Processing Systems 36 (2023) 50907–50931.

[17] T. R. McIntosh, T. Susnjak, T. Liu, P. Watters, D. Xu, D. Liu, R. Nowrozy, M. N. Halgamuge, From cobit to iso 42001: Evaluating cybersecurity frameworks for opportunities, risks, and regulatory compliance in commercializing large language models, Computers & Security (2024) 103964.

[18] R. Fredheim, Virtual manipulation brief 2023/1: Generative ai and its implications for social media analysis (2023).

[19] A. Anand, Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions, 2023.

[20] J. Blanco, C. Lambert, O. Thompson, Gpt-neo with lora for better medical knowledge performance on multimedqa dataset (2024).

[21] W. Wang, Z. Chen, X. Chen, J. Wu, X. Zhu, G. Zeng, P. Luo, T. Lu, J. Zhou, Y. Qiao, et al., Visionllm: Large language model is also an open-ended decoder for vision-centric tasks, Advances in Neural Information Processing Systems 36 (2024).

[22] M. A. Lowe, Ocr2seq: A novel multi-modal data augmentation pipeline for weak supervision (2023).

[23] A. Liu, H. Wang, M. Y. Sim, Personalised video generation: Temporal diffusion synthesis with generative large language model (2024).

[24] J. Lund, S. Macfarlane, B. Niles, Privacy audit of commercial large language models with sophisticated prompt engineering (2024).

[25] A. Zafar, V. B. Parthasarathy, C. L. Van, S. Shahid, A. I. Khan, A. Shahid, Building trust in conversational ai: A review and solution architecture using large language models and knowledge graphs, Big Data and Cognitive Computing 8 (6) (2024) 70.

[26] K. Mardiansyah, W. Surya, Comparative analysis of chatgpt-4 and google gemini for spam detection on the spamassassin public mail corpus (2024).

[27] L. Huovinen, Assessing usability of large language models in education (2024).

[28] K. Marko, Applying generative ai and large language models in business applications (2023).

[29] F. Dall'Agata, Instructing network devices via large language models (2024).

[30] V. M. Malode, Benchmarking public large language model (2024).

[31] J. Hassine, An llm-based approach to recover traceability links between security requirements and goal models, in: Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, 2024, pp. 643–651.