AI Business Review

Journal homepage: https://theaibr.com/index.php/aibr

Volume 1| No. 1 |Spring 2025

Bias in AI: A Comparative Analysis of DeepSeek and ChatGPT

Hesham Allam*[1], David Hua[1], Luis Orozco[1], Benjamin Kwasi Gyamfi[1], Daniel Ahengua[2]

1 Center of Information and Communication Science (CICS), College of Communication, Information & Media, Ball State University, Muncie, IN, 47304; hesham.allam@bsu.edu, dhua@bsu.edu, leo-rozco@bsu.edu,benjamine.gyamfi@bsu.edu

2 Instruction Technology, Northern Illinois University, DeKalb, Illinois, USA; dahengua1@niu.edu

*Correspondence: hesham.allam@bsu.edu

Abstract: Language models trained using artificial intelligence (AI) have now become ubiquitous in various fields, such as education, business, healthcare, and entertainment. However, these systems invite ethical questions, not the least of which is how to manage biases and maintain fairness. In this paper, two state-of-the-art AI language models are analyzed and compared: DeepSeek and ChatGPT. It examines how the ethical beliefs and practices of model developers seeking to mitigate bias influence the models' outputs and their real-world implications, using a literature review process. Through examining the strengths and limitations of each model in the context of ethical considerations, this study demonstrates key differences in how responses are generated, informative, and fair. Insights are presented in the context of responsible AI, including recommendations to improve governance and move toward a more equitable AI systems.

Academic Editor(s): Juan Dempere
Received: 10 June 2025
Revised: 15 June 2025
Accepted: 26 June 2025
Published: date: 4 July, 2025
Citation: Allam et al. (2025). Bias in AI: A Comparison of DeepSeek and ChatGPT

AI Business Review, 1(1), x.
https://doi.org/10.64044xxxx

Keywords: Bias in AI, DeepSeek, ChatGPT, Ethical AI, Comparative Analysis.

1. Introduction

A. Background and Rationale

In recent years, artificial intelligence (AI) has made great progress, and NLP has been widely used in AI (Hesham Allam et al., 2024; H. M. Allam, Gyamfi, & AlOmar, 2025). DeepSeek and ChatGPT are two examples of model-building, both at scale and in different ways. These models, which are fine-tuned on large text corpora, can also be used to perform state-of-the-art intelligent reasoning, writing, and other tasks such as chatbot or content generation (H. Allam, Dempere, Akre, & Flores, 2023; H. Allam et al., 2025a; Mourtzis, Angelopoulos, & Panopoulos, 2023).

The popularity of AI models is essentially a result of advancements in architecture, training methods, and computational capabilities. These developments have enabled AI tools to be very efficient in ways that can exceed human performance in terms of speed and accuracy (Ray, 2023). Consequently, AI is now being applied to many industries, from drafting documents to programming and from data analysis to health services (H. Allam, 2025; H. Allam et al., 2024).

However, there are still some issues with AI models. If biased information is learned, then discriminatory or misleading results will be obtained. Furthermore, it is hard to understand how certain outputs are produced since the AI process is a "black box" when it comes to decision-making(Ferrara, 2023) .This can lead to concerns over-reinforcing stereotypes, spreading misinformation, and diminishing user trust. DeepSeek and ChatGPT are two distinct approaches to addressing the same problem of bias and increasing transparency. DeepSeek specializes in open-source AI development enhanced with theorem proving and code intelligence (Wu, Duan, & Ni, 2024). ChatGPT, developed by OpenAI, has made headlines for its accessible interface, rapid updates, and the importance of content moderation and reinforcement learning (OpenAi, 2024). In this study, we conduct a comprehensive comparison of these two models regarding their ethical implications and performance in bias mitigation. It reviews existing research, presents the evaluation process, discusses patterns of bias, and makes recommendations for AI development & deployment.

B. Need for Ethical Oversight

Previous research(Hesham Allam et al., 2023; Hesham, Dempere, Akre, & Flores, 2023; Raji et al., 2020; Samala & Rawas, 2024) indicated that biased AI can negatively affect processes such as employee hiring, war perspectives, political debates, and prolonged systemic biases in education and healthcare. Accordingly, policymakers, academics, and advocacy organizations tend to emphasize the urgency of a holistic AI governance to mitigate such an effect. Although AI practitioners strive to create guidelines that emphasize fairness, accountability, transparency, and safety, addressing AI bias in AI models remains a challenge(Ashwini & Padhy, 2024).

OpenAI's ChatGPT learns over time through reinforcement and filtering mechanisms to adapt its responses; however, reported cases of subtle biases remain (Heaton, Nichele, Clos, & Fischer, 2024; Li, Fan, Atreja, & Hemphill, 2024). DeepSeek, which is more explicitly oriented towards transparency, has been subject to fewer public audits, and its attempts to mitigate bias are less evaluated (Sapkota, Raza, & Karkee, 2025).

C. Importance of Comparing DeepSeek and ChatGPT

The comparison of AI models is vital concerning the training data used, response generation, and the ethical approaches employed. With such an understanding, policymakers, developers, and end users can emphasize fairness when implementing AI models. The following are some of the advantages of comparing AI models:

- **Building better AI:** Understanding the strengths and weaknesses of DeepSeek and ChatGPT in addressing bias could lead to more effective pre-training and fine-tuning techniques, resulting in greater transparency and fairness (Kheya, Bouadjenek, & Aryal, 2024; Sreerama & Krishnamoorthy, 2022).

- **Regulatory Perspectives:** To develop regulations and robust accountability that minimize ethical concerns associated with AI, provided that regulators must have empirical evidence (Nathim et al., 2024; Samala & Rawas, 2024).

- **User Trust and Engagement**: Engaging users in the development process can enhance trust and ensure that AI systems meet ethical standards (Sreerama & Krishnamoorthy, 2022).

- **Contributions**: Comparative Studies provide valuable insights into algorithmic fairness, data governance, and ethical innovation (Sreerama & Krishnamoorthy, 2022).

D. Objectives of the Study

1. **Examine Ethical Frameworks in AI Language Models**: Analyze how ChatGPT and DeepSeek address key ethical principles, such as fairness, transparency, and inclusivity, based on existing scholarly literature and publicly available technical documentation.

2. **Interpret Literature on Bias and Sociopolitical Contexts**: Investigate how prior research depicts each model's handling of sensitive topics, including political ideologies, racial equity, and gender representation, across different cultural and regulatory environments.

3. **Develop Evidence-Based Recommendations:** Offer literature-informed recommendations for developers, policymakers, and AI practitioners to enhance ethical accountability and minimize bias in future model development and deployment.

E. Research Significance

It is crucial to guarantee that the AI model makes ethical decisions in order to ensure fairness and accuracy in different cases. This paper analyzes and evaluates DeepSeek and ChatGPT in the context of their handling of bias and ethical AI governance. Given the negative effects of biased AI, including misinformation, discrimination, and loss of trust, bias debiasing in AI has the potential to provide crucial lessons for the improvement of AI systems and their regulations. This paper justifies why, when considering predictions under risk, we may want to improve model monitoring and fairness analysis by discussing how these models succeed and fail to contribute to the development of AI solutions that are fairer and more responsible.

2. Literature Review

Prior work has studied the ethical implications of AI language models, but it has primarily been descriptive. To advance the conversation, we examine existing frameworks for probing AI bias, particularly those centered on fairness-aware learning and ethical alignment strategies. For example, Chopra and Singh (Chopra & Singh, 2018) emphasize the sociotechnical complexity of AI systems, noting that ethical issues arise not only from biases in the training data but also from the broader socio-technical contexts of system deployment. Building upon these observations, we evaluate whether these dependencies are mitigated or instead exacerbated in models such as DeepSeek and ChatGPT. This extended lens allows us to move beyond simply identifying problems, as in prior work, and instead test the effectiveness of empirical mitigation strategies that developers used to address ethical concerns.

A. DeepSeek Research Literature

1) The Multifaceted Approach of DeepSeek in Brief

DeepSeek has demonstrated the universality of its AI model that is suitable for code intelligence, theorem proving, vision-language understanding, and image retrieval. DeepSeek is studied for its ethical aspects and model efficiency with a focus on open-source availability. The following paragraphs present the background that underpins DeepSeek's approach to AI performance and responsible AI design (Gan, Ning, Qi, & Yu, 2025).

2) DeepSeek-Coder-V2

DeepSeek-Coder-V2 has been developed to utilize Mixture-of-Experts (MoE) for complex coding tasks, enhancing code intelligence and mathematical reasoning (DeepSeek-AI, 2024). It demonstrates versatility, supporting 86 to 338 programming languages and an extended context length of 128K tokens. However, questions remain as to how well it can control biases, specifically regarding ensuring that groups of underrepresented coders are safe and can contribute to open-source projects. (DeepSeek-AI, 2024).

4) DeepSeek-V2

DeepSeek-V2 is a large-scale MoE language model with 236B parameters (DeepSeek-AI, 2024). This version reduces the training cost by nearly 42.5% and simultaneously enhances throughput efficiency. Accordingly, the question remains whether such resource-intensive optimization handles fairness in conversational responses. Such an ongoing critique is paramount to ensure that efficiency is not compromised for the sake of ethical AI policy (DeepSeek-AI et al., 2024).

5) DeepSeek-VL

DeepSeek-VL has been created for vision-language applications to handle complex multimodal datasets such as web screenshots, PDFs, and graphs. (Lu et al., 2024). This ability to break down high-resolution photos is a considerably powerful tool, but it still pertains to ethical concerns about how it processes sensitive images and graphs. Evaluating its effectiveness under various cultural perspectives or dataset biases is essential for fair AI deployment(Deng et al., 2025).

6) DeepSeek-Coder and DeepSeek LLM

While achieving strong performance with relatively smaller parameter sizes in the scale range from 1.3B to 33B (DeepSeek-Coder) and even up to 67B (DeepSeek LLM), DeepSeek's paradigm scaling unfolds great success in pushing the frontier of code intelligence (Guo et al., 2024; Wu et al., 2024). These models all apply fill-in-the-blank learning strategies that outperform all ad-hoc proprietary approaches. But there are concerns about the open-source nature of DeepSeek when it comes to reducing biases and misuse when generating code (Wu et al., 2024).

7) **Content-Based Image Retrieval and Search**

DeepSeek features AI-based image search through natural language processing to improve semantic search precision (Deng et al., 2025; Piplani & Bamman, 2020). Bias issues are also relevant to the task of describing images for retrieval, as the unevenness in image annotation quality can bias evaluations, and the misrepresentation of cultural content can affect the fairness of descriptions (Piplani & Bamman, 2020). These challenges highlight the need for unbiased training data and responsible AI governance.

B. Literature on ChatGPT Research

OpenAI created ChatGPT and has been extensively evaluated for enhancing language capabilities, while focusing on ethics and reducing biases. Using transformer models and reinforcement learning through feedback (RLHF), ChatGPT is widely utilized in settings as well, as in business and customer service contexts. (Raj, Singh, Kumar, & Verma, 2023; Sajja, Sermet, Cikmaz, Cwiertny, & Demir, 2024). Despite being used by many users and experts, researchers still find flaws such as misinformation, political bias, and ethical considerations, according to a study conducted by previous studies (H. Allam et al., 2025b; Nathim et al., 2024; Samala & Rawas, 2024)

1) Ongoing Improvements

OpenAI recently implemented updates to improve the content moderation and fact checking abilities of ChatGPT (OpenAi, 2024). The updates include restrictions on types of prompts and enhancements in ensuring accuracy using refined RLHF methods to minimize biases(Heaton et al., 2024). Despite these efforts to reduce biases and enhance accuracy in AI generated content moderation and fact checking processes by OpenAI, scholars express concerns about censorship implications and the impact on user autonomy in politically sensitive conversations. (Heaton et al., 2024; Li et al., 2024).

2) Bias, Persistence, and Ethical Limitations

Research has shown that Chatbot GPT sometimes perpetuates stereotypes as a result of biases, in its training data(Zack et al., 2023). Despite efforts by OpenAI to use methods, like content moderation to address this issue some biases still exist in situations, making it difficult to completely eradicate the bias effect(Gichoya et al., 2023). This underscores the importance of having a range of data sets and clearer monitoring systems in place.

3) Cultural and Multilingual Usage

ChatGPT is used in cultural settings because it is highly popular, among users worldwide. For some minority languages, it might not provide the accuracy and fluency, hence possibly excluding users from these communities(Tuna, Schaaff, & Schlippe, 2024). Research suggests that particular dialects and underrepresented languages may not receive the same level of accuracy and fluency, potentially marginalizing users from these communities. (Song & Song, 2023; Tuna et al., 2024). Addressing these disparities is critical for ensuring inclusive AI and fairness.

C. Bias in AI Language Models

AI bias can stem from various sources, such as training data skew, cultural bias, and algorithm design(Gallegos et al., 2024; Kotek, Dockum, & Sun, 2023). Although fine-tuning methods help nudge AI models toward ethical decisions, they can also introduce new biases among models due to the incoherence of available annotations contributed by humans. On the one hand, ChatGPT is criticized for its sociopolitical biases(Rozado, 2023)while DeepSeek is open source, without enough public audit for fairness claims (Gallegos et al., 2024).

D. Ethical AI Development

ChatGPT and DeepSeek have attained maturity in their development phases and are now integral to the historical progression of the AI landscape. This emphasizes the frequency of model checkpointing and the suitable duration of training, rather than the loss of functions. Despite their apparent similarities, ChatGPT and DeepSeek constitute two separate disciplines of artificial intelligence(IS & DR, 2025). ChatGPT stands as a model created using Reinforcement Learning from Human Feedback (RLHF), constantly undergoing improvements and advancements(Kotsis, 2025). DeepSeek is a source that is transparent in nature. It encourages researchers to examine and contribute to the refinement of its model. (DeepSeek, 2024). Each model possesses distinct advantages, as ChatGPT leverages real-world feedback mechanisms, whereas DeepSeek facilitates independent evaluations of its bias mitigation approaches(IS & DR, 2025).

E. Case Studies of AI Bias

Previous research highlights how AI bias can impact real-world uses such as recruitment algorithms and facial recognition technology(Gallegos et al., 2024; Kotek et al., 2023). These concerns have led to calls for regulations and fairness evaluations from government bodies, as well as advocacy groups and institutions. Another example is related to databases, which often arises from unrepresentative datasets, leading to skewed outcomes that can perpetuate discrimination(Oyeniran, Adewusi, Adeleke, Akwawa, & Azubuko, 2022). Human-induced bias can inadvertently affect AI training, impacting decision-making processes in recruitment and other applications(Mujtaba & Mahapatra, 2019). Furthermore, algorithm-based systems are another concern resulting from assumptions about floating-point algorithms, which can exacerbate existing social inequalities. This study investigates these instances to identify methods for enhancing model accountability and promoting ethical AI governance.

Table 1: Summary Of Recent Studies On Deepseek

Title	Objectives	Findings	Authors
DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence	Introduce an open-source Mixture-of-Experts (MoE) model for coding tasks. Expand language support and context length to 128K tokens.	Matches GPT-4 Turbo performance; broadens language coverage.	(DeepSeek, 2024)
DeepSeek-Prover: Advancing Theorem Proving in LLMs Through Large-Scale Synthetic Data	Develop theorem-proving AI using synthetic data. Translate informal math problems into formal proofs.	Outperforms GPT-4 in proof generation, achieving higher accuracy.	(Xin et al., 2024)
DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts	Enhance MoE efficiency and reduce training costs. Provide bilingual text generation.	Reduces training costs by 42.5%; shows better human alignment.	(DeepSeek et al., 2024)
DeepSeek-VL: Towards Real-World Vision-Language Understanding	Create a vision-language model for real-world data such as PDFs, OCR, and charts.	Processes high-resolution images effectively, improving real-world vision-language tasks.	(Lu et al., 2024)

DeepSeek-Coder: When the Large Language Model Meets Programming – The Rise of Code Intelligence	Provide open-source code intelligence with model sizes from 1.3B to 33B parameters and training on 2 trillion tokens.	Surpasses Codex and GPT-3.5 in code generation and infilling tasks.	(Guo et al., 2024)
DeepSeek LLM: Scaling Open-Source Language Models with Longtermism	Explore scaling strategies in open-source LLMs with up to 67B parameters.	Outperforms LLaMA-2 70B in reasoning, coding, and math.	(Wu et al., 2024)
DeepSeek: Content-Based Image Search and Retrieval	Implement NLP-based image retrieval systems.	Enhances semantic and contextual image retrieval.	(Piplani & Bamman, 2020)

3. Methodology

This paper employs a comparative literature-based approach to analyze and interpret the ethical dimensions and bias mitigation strategies of two prominent large language models: ChatGPT and DeepSeek. Instead of conducting a new experimental evaluation, we relied on a diverse body of existing scholarly papers, technical documentation, and external audits to assess how these models address fairness, transparency, and inclusivity.

A. Theoretical Framework

Established frameworks in AI ethics and bias research guide our analysis. In particular, we adopt the dimensions outlined by Blodgett et al., Bender et al., and Sap et al.) Bender, Gebru, McMillan-Major, & Shmitchell, 2021; Blodgett, Barocas, Daumé Iii, & Wallach, 2020; Sap, Card, Gabriel, Choi, & Smith, 2019(which assesses AI systems according to:

- **Equity and Representational Fairness**: Evaluating whether models reflect biased assumptions or reinforce societal stereotypes.
- **Misinformation Management**: Analyzing the model's strategies for recognizing and mitigating misleading or false information.
- **Ethical Alignment**: Assessing conformity with ethical principles such as justice, transparency, and respect for human dignity.

These lenses enable a structured critique of both ChatGPT and DeepSeek in ethically complex scenarios.

B. Comparative Literature Review Method

1. **Source Selection**: We synthesized more than 50 sources, including peer-reviewed articles, AI audits, official documentation, and white papers that explicitly address bias and ethics in ChatGPT and DeepSeek.
2. **Qualitative Thematic Analysis**: Sources were coded according to recurring ethical themes, including censorship tendencies, cultural bias, transparency in moderation, and openness of data and architecture. Contrasting perspectives from Western and Eastern AI ethics discourses were included to ensure cultural sensitivity.

3. **Case Integration**: Where appropriate, we incorporate previously published examples that compare the models on controversial topics such as political discourse, gender roles, and historical representation. These published instances help contextualize abstract ethical principles into model-specific behavior.

4. Findings and Analysis

The literature uncovers paths in which Chatbot GPT and DeepSeek tackle the dilemmas that come with extensive language modeling on a large scale.

A. Bias and Cultural Framing

ChtGPT, from OpenAI, has undergone improvements using reinforcement learning and moderation techniques to enhance its performance in engaging with users over time. Various research studies (Gichoya et al., 2023; Heaton et al., 2024) suggest that the GPT chatbot still exhibits remnants of its principles and may not adequately represent conservative or non-Western viewpoints.

DeepSeek stands out as a source, prioritizing transparency (Wu et al., 2024), however, it shows a sociopolitical filtering approach that adheres to Chinese regulatory standards, as highlighted by Sapkota et al(Sapkota et al., 2025) . The biases observed within DeepSeek are usually not due to misrepresentation but stem from the exclusion or avoidance of politically sensitive material.

B. Moderation and Accountability

Chatbot GPT incorporates a method called Reinforcement Learning from Human Feedback (RLHF), which offers a way to refine and adjust responses based on input from users or humans giving feedback. However, its exclusive design restricts assessment by sources, resulting in criticisms concerning the lack of transparency in how content moderation is carried out (Bender et al., 2021).

DeepSeek encourages community scrutiny; however, the efficacy of this decentralization is a topic of debate within the literature realm. Studies indicate that while open source contributions enhance transparency levels, the absence of audit mechanisms may overlook ethical dilemmas (Lu et al., 2024).

C. Model Governance and Documentation

ChatGPT is periodically updated with structured release notes and alignment papers(OpenAi, 2024). DeepSeek's open documentation offers greater customization, but lacks the robust communication channels seen in OpenAI's infrastructure. Literature points out that this gap may hinder non-technical users from fully understanding the implications of DeepSeek's model outputs.

4. Discussion

A. Relative Strengths and Weaknesses

The centralized control system of ChatGPT, combined with its rigorous moderation process and continuous tuning, helps minimize overt bias. The system occasionally demonstrates excessive caution through self-restraint when discussing political or cultural matters

(Sapkota et al.,2025). DeepSeek promotes transparency through its open-source development model and community contribution process. The decentralized nature of DeepSeek creates challenges with update consistency and ethical oversight, according to Lu et al. and Sapkota et al. (Lu et al., 2024; Sapkota et al., 2025).

B. Implications for AI Ethics

The literature shows that ethical alignment in AI goes beyond technical aspects, because it involves sociopolitical considerations(Bender et al., 2021; Sap et al., 2019). ChatGPT's behavior follows Western liberal-democratic norms, but DeepSeek's alignment follows Chinese cultural and regulatory expectations(Sapkota et al., 2025; Wu et al., 2024). The tendencies demonstrate the challenges of creating AI systems that function ethically in different worldwide settings. The comparative analysis demonstrates that AI ethical design needs to address regional regulatory systems and historical backgrounds, together with changing social standards(Al-Zahrani & Alasmari, 2025).

C. Path Forward for Practitioners

Future development of LLMs should prioritize region-sensitive data source, transparent governance mechanisms, and third-party auditability. Several scholars suggest integrating culturally diverse training data and expanding AI literacy to empower users in critically engaging with AI-generated content(Davoodi, 2024; Nguyen & Pham, 2024; Orosoo et al., 2024; Raji et al., 2020). The implementation of formal ethical auditing frameworks and feedback systems that include underrepresented voices should be applied to both proprietary and open-source platforms(Raji et al., 2020). Through this approach, developers and policymakers can work together to develop AI models, which are fair and transparent and suitable for specific contexts.

5. Conclusion

This paper highlights how two leading LLMs—ChatGPT and DeepSeek—have approached ethical alignment and bias mitigation from fundamentally different standpoints. The research uses comparative literature analysis to demonstrate how these differences represent fundamental design principles, together with geopolitical factors and regulatory frameworks (Bender et al., 2021; Gichoya et al., 2023). The centralized moderation system of ChatGPT maintains safety standards and consistency; however, it creates transparency and bias concerns in political situations. On the other hand, the open-source nature of DeepSeek promotes accountability and community involvement; however, it requires standardized audit systems to ensure consistent fairness.

Future research should track the development of these systems, as they will operate in complex, multimodal, and cross-lingual environments. The development of AI models that meet ethical standards in various societal contexts requires ongoing critical evaluation, which must be supported by interdisciplinary expertise in the field (H. Allam et al., 2025a; Williamson & Prybutok, 2024).

Acknowledgement

The authors gratefully acknowledge the use of AI-assisted tools, such as Grammarly, to improve the clarity, grammar, and readability of the manuscript. All conceptual development, research design, data analysis, and interpretation presented in this study were conducted solely by the authors without reliance on AI for content generation or intellectual contributions.

6. References

1. Al-Zahrani, A. M., & Alasmari, T. M. (2025). A comprehensive analysis of AI adoption, implementation strategies, and challenges in higher education across the Middle East and North Africa (MENA) region. *Education and Information Technologies*, 1-51.
2. Allam, H. (2025). Prescribing the future: The role of artificial intelligence in pharmacy. *Information, 16*(2), 131.
3. Allam, H., Allen, C., Allred, A., Boersma, M., Davison, C. B., Lazaros, E. J., & Truell, A. D. (2024). Practical wisdom and intelligent machines: toward AI-human socio-technical decision making and resource allocation. *Issues in Information Systems, 25*(2).
4. Allam, H., Dempere, J., Akre, V., & Flores, P. (2023). *Artificial intelligence in education (AIED): Implications and challenges.* Paper presented at the Proceedings of the HCT International General Education Conference (HCT-IGEC 2023).
5. Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., & Ahamed, J. (2023). *Artificial intelligence in education: an argument of Chat-GPT use in education.* Paper presented at the 2023 9th International Conference on Information Technology Trends (ITT).
6. Allam, H., Dempere, J., Lazaros, E., Davison, C., Kalota, F., & Hua, D. (2025a). *Unleashing educational potential: Integrating chatgpt in the classroom*. Paper presented at the HCT International General Education Conference (HCTIGEC 2024). .
7. Allam, H. M., Gyamfi, B., & AlOmar, B. (2025). Sustainable innovation: Harnessing AI and living intelligence to transform higher education. *Education Sciences, 15*(4), 398.
8. Ashwini, P., & Padhy, P. C. (2024). Cognitive Bias and Fairness Challenges in AI Consciousness. In (pp. 89-109): IGI Global.
9. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Paper presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.
10. Blodgett, S. L., Barocas, S., Daumé Iii, H., & Wallach, H. (2020). *Language (technology) is power: A critical survey of "bias" in NLP*. arXiv preprint arXiv:2005.14050. Retrieved from https://arxiv.org/abs/2005.14050
11. Chopra, A. K., & Singh, M. P. (2018). *Sociotechnical systems and ethics in the large.* Paper presented at the Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society.
12. Davoodi, A. (2024). EQUAL AI: A Framework for Enhancing Equity, Quality, Understanding and Accessibility in Liberal Arts through AI for Multilingual Learners. *Language, Technology, and Social Media, 2*(2), 178-203.
13. DeepSeek-AI. (2024). Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model Retrieved from https://arxiv.org/abs/2405.04434
14. DeepSeek, A. I. (2024). *DeepSeek: Advancing open-source AI for reasoning and coding*. Retrieved from https://deepseek.com/research
15. DeepSeek, A. I., Zhu, Q., Guo, D., Shao, Z., Yang, D., Wang, P., & Liang, W. (2024). DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence. *Artificial Intelligence Review, 12*(2), 45-60.
16. Deng, Z., Ma, W., Han, Q. L., Zhou, W., Zhu, X., Wen, S., & Xiang, Y. (2025). Exploring DeepSeek: A Survey on Advances, Applications, Challenges and Future Directions. *IEEE/CAA Journal of Automatica Sinica, 12*(5), 872-893.
17. Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), 3.
18. Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., . . . Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 1-79.
19. Gan, W., Ning, Z., Qi, Z., & Yu, P. S. (2025). Mixture of Experts (MoE): A Big Data Perspective. *arXiv Preprint*.
20. Gichoya, J. W., Thomas, K., Celi, L. A., Safdar, N., Banerjee, I., Banja, J. D., & Purkayastha, S. (2023). AI pitfalls and what not to do: mitigating bias in AI. *The British Journal of Radiology, 96*(1150), 20230023. doi:10.1259/bjr.20230023
21. Guo, D., Zhu, Q., Yang, D., Xie, Z., Dong, K., Zhang, W., & Liang, W. (2024). DeepSeek-Coder: The Rise of Code Intelligence. *AI & Software Engineering Journal, 8*(1), 32-49.
22. Heaton, D., Nichele, E., Clos, J., & Fischer, J. E. (2024). “ChatGPT says no”: Agency, trust, and blame in Twitter discourses after the launch of ChatGPT. *AI and Ethics*, 1-23.

23. IS, S., & DR, A. (2025). Advancements in AI-Powered NLP Models: A Critical Analysis of ChatGPT and DeepSeek. *Available at SSRN 5125445*.
24. Kheya, T. A., Bouadjenek, M. R., & Aryal, S. (2024). The Pursuit of Fairness in Artificial Intelligence Models: A Survey. *arXiv.Org*. Retrieved from https://doi.org/10.48550/arxiv.2403.17333
25. Kotek, H., Dockum, R., & Sun, D. (2023). *Gender bias and stereotypes in large language models*.
26. Kotsis, K. T. (2025). ChatGPT and DeepSeek Evaluate One Another for Science Education. *EIKI Journal of Effective Teaching Methods, 3*(1).
27. Li, L., Fan, L., Atreja, S., & Hemphill, L. (2024). "HOT" ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media. *ACM Transactions on the Web, 18*(2), 1-36.
28. Lu, H., Liu, W., Zhang, B., Wang, B., Dong, K., Sun, J., & Ruan, C. (2024). DeepSeek-VL: Towards Real-World Vision-Language Understanding. *Machine Learning in Vision Research, 10*(4), 89-105.
29. Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2023). The future of the human–machine interface (HMI) in society 5.0. *Future Internet, 15*(5), 162.
30. Mujtaba, D. F., & Mahapatra, N. R. (2019). *Ethical Considerations in AI-Based Recruitment*. Paper presented at the International Symposium on Technology and Society (ISTAS). https://doi.org/10.1109/ISTAS48451.2019.8937920
31. Nathim, K. W., Hameed, N. A., Salih, S. A., Taher, N. A., Salman, H. M., & Chornomordenko, D. (2024). Ethical AI with Balancing Bias Mitigation and Fairness in Machine Learning Models. Retrieved from https://doi.org/10.23919/fruct64283.2024.10749873
32. Nguyen, V., & Pham, C. (2024). Leveraging Large Language Models for Suicide Detection on Social Media with Limited Labels. doi:https://doi.org/10.48550/arxiv.2410.04501
33. OpenAi. (2024). *ChatGPT and AI safety: An evolving approach*. Retrieved from https://help.openai.com/en/articles/
34. Orosoo, M., Goswami, I., Alphonse, F. R., Fatma, G., Rengarajan, M., & Bala, B. K. (2024). *Enhancing Natural Language Processing in Multilingual Chatbots for Cross-Cultural Communication.* Paper presented at the 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).
35. Oyeniran, O. C., Adewusi, A. O., Adeleke, A. G., Akwawa, L. A., & Azubuko, C. F. (2022). Ethical AI: Addressing bias in machine learning models and software applications. *Computer Science & IT Research Journal, 3*(3), 115-126. doi:10.51594/csitrj.v3i3.1559
36. Piplani, T., & Bamman, D. (2020). DeepSeek: Content-Based Image Search & Retrieval. *Journal of Digital Media & AI, 9*(2), 56-72.
37. Raj, R., Singh, A., Kumar, V., & Verma, P. (2023). Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3*(3), 100140.
38. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., . . . Barnes, P. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. Paper presented at the Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20).
39. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121-154.
40. Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences, 12*(3), 148. doi:10.3390/socsci12030148
41. Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information, 15*(10), 596.
42. Samala, A. D., & Rawas, S. (2024). Bias in artificial intelligence: smart solutions for detection, mitigation, and ethical strategies in real-world applications. *IAES International Journal of Artificial Intelligence*. Retrieved from https://doi.org/10.11591/ijai.v14.i1.pp32-43
43. Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). *The risk of racial bias in hate speech detection*. Paper presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.

44. Sapkota, R., Raza, S., & Karkee, M. (2025). *Comprehensive analysis of transparency and accessibility of ChatGPT, DeepSeek, and other SOTA large language models*. arXiv preprint. Retrieved from https://arxiv.org/abs/2502.18505
45. Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in psychology*, *14*, 1260843.
46. Sreerama, J., & Krishnamoorthy, G. (2022). Ethical Considerations in AI Addressing Bias and Fairness in Machine Learning Models. *Online*. Retrieved from https://doi.org/10.60087/jklst.vol1.n1.p138
47. Tuna, M., Schaaff, K., & Schlippe, T. (2024). Effects of Language- and Culture-Specific Prompting on ChatGPT. In.
48. Williamson, S. M., & Prybutok, V. (2024). The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation. *Information*, *15*(6), 299.
49. Wu, X., Duan, R., & Ni, J. (2024). Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence*, *2*(2), 102-115.
50. Xin, H., Guo, D., Shao, Z., Ren, Z., Zhu, Q., Liu, B., & Liang, X. (2024). DeepSeek-Prover: Advancing Theorem Proving in LLMs through Large-Scale Synthetic Data. *Computational Mathematics Journal*, *14*(1), 78-94.
51. Zack, T., Lehman, E., Suzgun, M., Rodriguez, J. A., Celi, L. A., Gichoya, J., . . . Abdulnour, R. E. E. (2023). Coding Inequity: Assessing GPT-4's Potential for Perpetuating Racial and Gender Biases in Healthcare. *medRxiv*.