Research Article: Journal of Computational Social Dynamics Volume: 08

Ethical Considerations and Potential Risks in the Deployment of Large Language Models in Diverse Societal Contexts

Udara Piyasena Liyanage
Department of Mathematics, University of Kelaniya, Kelaniya 11600, Sri Lanka

Nimnaka Dilshan Ranaweera
Department of Computer Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka

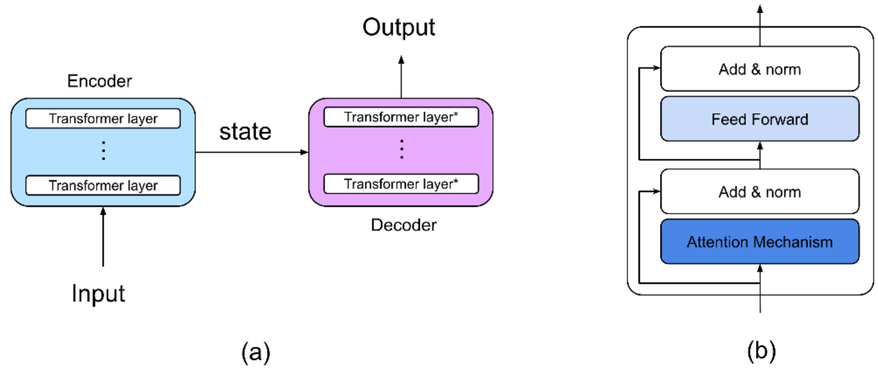
Abstract

The integration of large language models (LLMs) into various societal applications brings forth a plethora of ethical challenges and potential risks. One primary concern is the perpetuation and amplification of biases present in the training data of LLMs. This risk is acutely pronounced in diverse societal contexts, potentially leading to unfair or harmful outcomes, especially for marginalized groups through stereotypical or discriminatory content generation. Additionally, privacy concerns emerge due to LLMs' potential to inadvertently memorize and disclose sensitive personal information, a significant issue in environments with varying data protection norms. Another critical risk is the use of LLMs in spreading misinformation and manipulation, with profound implications in political, social, and personal spheres, such as influencing elections or facilitating scams. The increasing dependence on LLMs for various tasks might result in human skill degradation, particularly concerning in educational and professional contexts. Accountability and transparency issues also arise, given the difficulty in pinpointing responsibility for LLM outputs and the "black box" nature of these models. Economically, LLMs pose a threat of job displacement in certain sectors due to their ability to automate tasks traditionally performed by humans, necessitating societal adjustments and new workforce training approaches. Cultural homogenization is another concern, as the dominance of specific languages and cultures in LLM training data might lead to the underrepresentation or misrepresentation of minority cultures. Furthermore, the uneven distribution of LLM benefits exacerbates the digital divide, potentially leaving individuals without advanced technology access behind. The application of LLMs in sensitive fields like healthcare, legal, or law enforcement raises unique ethical considerations, as incorrect or biased advice can have dire consequences.

Introduction

Socialization, Large Language Models (LLMs) represent a transformative breakthrough in the field of artificial intelligence and natural language processing. These models are essentially neural networks that have been trained on massive amounts of text data, enabling them to understand and generate human-like text with remarkable accuracy [1], [2]. The history of LLMs can be traced back to the early development of neural networks and deep learning techniques, but their recent explosion in popularity can be attributed to several key milestones.

Figure 1. Architecture Encoder-Decoder, transformer layer

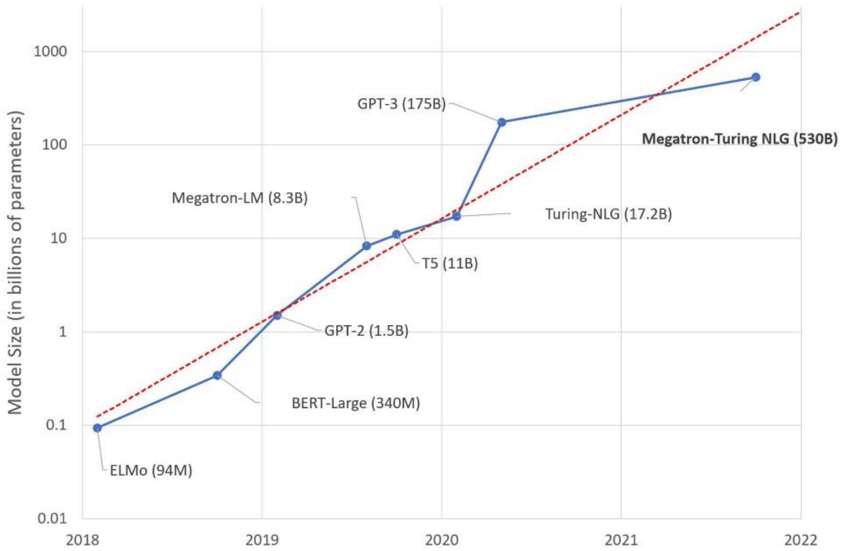


The concept of LLMs can be traced back to the early days of artificial neural networks, which were inspired by the structure and function of the human brain. Researchers in the 1950s and 1960s began developing the mathematical foundations for these networks, but their computational limitations at the time hindered their practical applications. It wasn't until the 2010s that significant progress was made, thanks to advancements in hardware and the availability of vast amounts of text data on the internet. Researchers began to build increasingly deep and complex neural network architectures, such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), which showed promise in processing natural language [3].

The breakthrough in LLMs came with the introduction of models like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers) in the mid-2010s. These models leveraged the power of transformers, a novel neural network architecture, to achieve state-of-the-art results in a wide range of natural language understanding and generation tasks. GPT-3, for instance, released by OpenAI in 2020, demonstrated an unprecedented ability to generate coherent and contextually relevant text by training on an enormous corpus of text data [4]. This marked the beginning of a new era in AI, with LLMs becoming increasingly prevalent in applications ranging from chatbots and virtual assistants to machine translation and content generation [5], [6]. Today, large language models continue to evolve, with researchers exploring ways to make them even more efficient, ethical, and capable of understanding and generating text in multiple languages and domains.

Large Language Models (LLMs) consist of several integral components that collectively empower them to comprehend and generate human-like text with remarkable proficiency. The first key component is the embedding layer, which converts words or tokens into highdimensional vectors, capturing semantic information and facilitating context understanding [7]. The transformer architecture forms the backbone of most LLMs, comprising multiple layers of attention mechanisms and feedforward neural networks. These transformers excel at capturing long-range dependencies in text and weigh the significance of different words through self-attention mechanisms [8].

Figure 2. Number of parameters of Language Models (2018-2022)



LLMs undergo a two-stage training process: pre-training and fine-tuning. In pre-training, the model learns language patterns and general knowledge from a massive text corpus. Fine-tuning follows, where the model specializes in specific tasks, making it versatile and adaptable across diverse applications. Attention heads within the transformer architecture play a crucial role, as they focus on different parts of the input sequence, enabling the model to capture various linguistic features and dependencies. Positional encoding, another vital component, imparts information about word order and position in a sequence, addressing a limitation of transformers [9].

The output layer of an LLM is responsible for generating text or predictions, depending on the task. It may take various forms, such as a softmax layer for classification or a decoder for text generation. These components, in concert, empower LLMs to revolutionize natural language processing, leading to significant advancements in various AI applications, including chatbots, virtual assistants, language translation, and content generation. The synergy between these components underpins the transformative impact of LLMs on the field of artificial intelligence and natural language understanding [10], [11].

Ethical Considerations and Potential Risks

Bias and fairness are critical concerns in the deployment of Large Language Models (LLMs). These models, while powerful and versatile, can inherit and even exacerbate biases present in their training data. This issue is particularly pronounced in diverse societal contexts where biases related to culture, gender, race, and other factors can result in unfair or harmful outcomes. The potential for LLMs to perpetuate stereotypes or generate discriminatory content is a significant challenge that must be addressed to ensure ethical and equitable use of these technologies.

One of the primary reasons LLMs can perpetuate bias is their reliance on vast corpora of text data from the internet, which often contains inherent biases present in human language. These biases can manifest in various ways, from gender stereotypes to racial prejudices, and the models can learn and reproduce these biases when generating text [12]. For example, an LLM might produce gendered language when discussing certain professions or racial stereotypes when generating content related to specific communities. These biases not only reinforce harmful stereotypes but can also lead to real-world consequences, such as discrimination in automated decision-making systems [13].

To address bias and fairness concerns, researchers and practitioners are actively working on developing techniques and guidelines for mitigating bias in LLMs. This includes pre-processing training data to reduce biases, fine-tuning models on specific fairness criteria, and actively seeking diverse perspectives and input in the development process. Additionally, promoting transparency and accountability in LLM development and deployment is crucial. Users and developers should be aware of the potential biases in these models and take steps to evaluate and mitigate them in real-world applications [14].

Privacy concerns surrounding Large Language Models (LLMs) are indeed substantial, and they stem from the vast amounts of training data these models rely on. One major issue is the potential for LLMs to inadvertently memorize and subsequently regurgitate sensitive personal information. This poses significant challenges in the realm of data protection, especially in contexts where privacy norms and regulations can vary widely [15].

The primary source of this privacy concern lies in the nature of LLM training data. These models are typically trained on large and diverse datasets that may contain personal information, such as names, addresses, phone numbers, and more. During the training process, LLMs can learn to associate specific tokens with individuals, even when the data itself is anonymized or pseudonymized. This can lead to the unintended disclosure of private information when LLMs generate text or responses. Furthermore, the global nature of the internet and the varying levels of data protection regulations across different jurisdictions complicate matters. What may be considered a privacy violation in one region may not be viewed the same way in another. LLMs, if not properly controlled, can inadvertently violate privacy norms, which can have legal and ethical consequences.

To address these privacy concerns, researchers and developers working on LLMs are actively exploring techniques to enhance privacy protection. These approaches include refining data anonymization methods, implementing differential privacy mechanisms, and developing guidelines for responsible data handling. Additionally, LLM developers are working on ways to allow users more control over the information their models have access to and generate.

The risk of misinformation and manipulation is a pressing concern when it comes to Large Language Models (LLMs). These models, with their remarkable text generation capabilities, have the potential to create convincing but false or misleading information, and this poses serious implications across various domains, including politics, social discourse, and personal interactions [16], [17].

In the realm of politics, LLMs can be weaponized to generate persuasive political propaganda, fake news articles, or social media posts. These manipulative texts can spread rapidly, sway public opinion, and influence elections. The ability to create highly engaging and emotionally charged content can make it challenging for individuals to discern fact from fiction, leading to the erosion of trust in reliable sources of information [18].

In the social sphere, LLMs can be used to fabricate social media posts, comments, and reviews. This can undermine the credibility of online discussions and reviews, making it difficult for users to make informed decisions or engage in meaningful conversations. The proliferation of fake accounts and content can also contribute to the creation of echo chambers, where individuals are exposed only to information that reinforces their existing beliefs [19].

On a personal level, LLMs can facilitate various scams and online fraud. Cybercriminals can use these models to craft convincing phishing emails, fraudulent product listings, or even impersonate individuals for identity theft. Victims may be lured into sharing sensitive personal information or falling for financial scams [20].

To mitigate the risks associated with misinformation and manipulation, it is essential to develop and implement strategies that enhance the accountability and transparency of LLMs. This includes developing techniques to detect generated content, promoting media literacy and critical thinking skills, and encouraging responsible use of these technologies. Moreover, ethical guidelines and regulations may be necessary to address the responsible deployment of LLMs and hold those who misuse them accountable.

The growing dependence on Large Language Models (LLMs) for various tasks, such as writing, decision-making, and problem-solving, raises concerns about potential skill degradation in humans. This over-reliance on LLMs can have significant implications in both educational and professional settings, impacting the development of critical skills and undermining the traditional roles of humans in these domains.

In educational contexts, the availability of LLMs for tasks like essay writing, research, and homework assistance can lead to a reduction in students' motivation to develop essential writing and research skills. If students rely too heavily on LLMs to generate their academic work, they may miss out on the opportunity to cultivate their creativity, critical thinking, and information synthesis abilities [21], [22]. Furthermore, educators may struggle to assess students' genuine knowledge and capabilities when faced with the possibility of plagiarized or machine-generated content [23].

In the professional realm, an over-reliance on LLMs for decision-making and problem-solving could lead to a degradation of cognitive skills in individuals and teams. Professionals may become complacent, relying on LLMs to provide solutions without fully understanding or critically evaluating the underlying reasoning. This can result in a diminished capacity for independent thinking and problem-solving, potentially compromising the quality and innovation of work.

Additionally, issues of dependence on LLMs can extend to issues of bias and accountability. If LLMs are used as decision-making tools in fields like law, finance [24], or healthcare, there is a risk that biased or incorrect recommendations from these models may be accepted without question, leading to unjust or harmful outcomes. This underscores the importance of maintaining human oversight and critical evaluation when integrating LLMs into decision-making processes.

To address these concerns, it is crucial to strike a balance between leveraging LLMs' capabilities and preserving human skills and autonomy. In educational settings, educators can emphasize the importance of developing foundational skills even when using technology as a tool. In professional contexts, organizations should establish guidelines for responsible and ethical LLM use, ensuring that human expertise and judgment remain central to decision-making processes.

Accountability and transparency present intricate challenges in the realm of Large Language Models (LLMs). Determining responsibility for the outputs of these models is often unclear, as it involves a complex interplay between developers, users, and the model itself. Developers have a role in training and fine-tuning LLMs, but they cannot foresee all potential applications or misuses. Users influence the outcomes through their inputs and context, but holding them solely responsible for unexpected or harmful content may not be equitable. Balancing accountability among these stakeholders requires careful consideration and the establishment of clear guidelines. Moreover, the "black box" nature of LLMs compounds these challenges. LLMs consist of millions or even billions of parameters, making it exceedingly difficult to decipher the intricate decision-making processes underlying their outputs. This lack of interpretability hinders efforts to understand, scrutinize, and rectify potential biases or errors in the generated content.

Cultural homogenization is a critical concern when it comes to the development and deployment of large language models (LLMs). These models rely heavily on vast amounts of text data to learn and generate human-like text. However, the data used to train LLMs often comes from dominant languages and cultures, leading to a lack of representation for minority cultures and languages. As a result, LLMs may struggle to understand or generate content that accurately reflects the nuances, idioms, and cultural diversity of these underrepresented groups. This can reinforce the dominance of certain cultures and languages on the internet and in digital communication, stifling the preservation and celebration of cultural diversity [25], [26].

Furthermore, the issue of accessibility and the digital divide is closely intertwined with the deployment of LLMs. While these models hold tremendous potential for various applications, they require access to advanced technology and the internet. This creates a digital divide where individuals and communities without such access are at a significant disadvantage. Those who lack access to the internet or advanced computing resources may miss out on the benefits of LLMs, including improved communication, education, and information retrieval. This digital divide can exacerbate existing inequalities and hinder the potential for equitable societal progress [27].

To address these challenges, it is crucial to consider strategies that promote cultural diversity and inclusivity in the development and training of LLMs. This may involve actively seeking out and incorporating data from underrepresented cultures and languages, as well as developing models that are more sensitive to cultural differences. Additionally, efforts should be made to ensure that the benefits of LLMs are accessible to a wider range of people, including those in underserved communities. This could involve initiatives to provide affordable internet access and computing resources to marginalized populations and to develop LLM applications that work effectively on low-end devices.

Moreover, it's essential to foster a global dialogue on the ethical and societal implications of LLMs. This discussion should involve various stakeholders, including technology companies, policymakers, academics, and representatives from different cultural and linguistic backgrounds. By engaging in a collaborative effort to address these challenges, we can work towards a more inclusive and equitable future where the potential of LLMs is harnessed for the benefit of all.

Ethical considerations play a pivotal role in the deployment of large language models (LLMs), especially when they are applied in sensitive domains like healthcare, legal, or law enforcement. The use of LLMs in these critical areas gives rise to specific ethical concerns that must be carefully addressed. One of the primary concerns is the potential for LLMs to provide incorrect or biased advice or information, which could lead to serious consequences for individuals and society as a whole [28], [29].

In the realm of healthcare, for instance, LLMs are being employed to assist with diagnosis, treatment recommendations, and medical research. However, if these models are not trained on diverse and representative healthcare data, they may provide recommendations that are inaccurate, incomplete, or biased. This could result in misdiagnoses, inappropriate treatments, or disparities in healthcare outcomes, especially for marginalized or underrepresented patient populations.

Similarly, in the legal field, the use of LLMs for legal research and document review has the potential to streamline processes and increase efficiency. However, if these models are not trained to recognize and mitigate biases present in legal texts or to uphold legal principles and standards, they may inadvertently perpetuate injustice and inequality. Biased legal advice or decisions can have severe consequences for individuals' rights and the integrity of the legal system itself.

In the context of law enforcement, LLMs are sometimes used to assist in decision-making, such as predicting crime trends or identifying suspects. The use of biased or incomplete data to train these models can result in discriminatory practices and unjust profiling, exacerbating social inequalities and eroding trust in law enforcement agencies [30].

To address these ethical concerns, it is imperative to establish rigorous guidelines and oversight mechanisms for the development and deployment of LLMs in sensitive applications. This includes ensuring that training data is diverse and representative, regularly auditing the models for biases, and involving domain experts to validate and interpret the model's outputs. Additionally, transparency and accountability should be key principles in the deployment of LLMs in these domains, with clear mechanisms for recourse in case of errors or biases.

Conclusion

The deployment of large language models (LLMs) in diverse societal contexts raises several ethical considerations and potential risks that demand careful scrutiny. One of the most pressing concerns is the issue of bias and fairness. LLMs have the capacity to perpetuate or even amplify biases present in their training data, which poses a significant risk when applied in diverse settings. In societies characterized by cultural, gender, racial, or other forms of bias, these models can inadvertently generate content that is stereotypical or discriminatory, disproportionately affecting marginalized groups. This not only reinforces existing inequalities but also raises questions about accountability and the responsibility of those developing and deploying these technologies to mitigate bias.

Privacy concerns are another paramount issue tied to the use of LLMs in diverse contexts. These models, trained on vast datasets, possess the capability to inadvertently memorize and regurgitate sensitive personal information, raising significant privacy concerns. In societies where data protection and privacy norms vary widely, the deployment of LLMs could result in violations of individuals' privacy rights. Striking the right balance between harnessing the power of LLMs for innovation and respecting individuals' right to privacy is a challenge that must be addressed proactively.

Misinformation and manipulation represent a considerable risk associated with LLMs in diverse societal contexts. These models can be exploited to generate convincingly false or misleading information, with far-reaching consequences. In political arenas, they could be used to influence elections by disseminating false narratives, while in social and personal contexts, they might facilitate scams and disinformation campaigns. The ability of LLMs to generate content that appears credible and authentic underscores the urgency of developing robust mechanisms to verify information generated by these models [31].

In addition to these concerns, the deployment of LLMs in diverse contexts also raises issues related to accountability and transparency. It becomes challenging to trace the source of generated content back to the model, which can complicate matters when addressing issues of bias, misinformation, or privacy breaches. Clear guidelines and mechanisms for accountability and transparency are essential to ensure responsible usage of LLMs in diverse societal contexts.

Furthermore, there is a risk of LLMs exacerbating the digital divide in diverse societies. If these advanced language models are predominantly accessible to a privileged few, it can further marginalize underserved communities [32]. This access gap can result in disparities in information dissemination, exacerbating existing inequalities. Thus, efforts must be made to ensure equitable access and usage of LLMs to avoid widening societal divides.

Dependence on LLMs presents a unique concern as their widespread use for tasks like writing, decision-making, or problem-solving could lead to the degradation of human skills. Over-reliance on these models might erode individuals' ability to think critically, write coherently, or make informed decisions independently. This dependence could manifest as a challenge in educational and professional settings, where the development of essential cognitive and analytical skills could be hindered by the excessive use of LLMs, raising questions about how to strike a balance between leveraging the technology and maintaining human capabilities.

Accountability and transparency issues emerge as central dilemmas in the deployment of LLMs. Determining responsibility for the outcomes generated by these models is often unclear. Should the developers, users, or the model itself be held accountable for unintended consequences or biases in its output? The inherent "black box" nature of these models further compounds the problem, making it challenging to understand and explain how the models arrive at certain outputs. This opacity complicates the assignment of accountability and underscores the need for greater transparency and oversight.

The economic impact and job displacement potential of LLMs are significant considerations in diverse societal contexts. These models have the capacity to automate tasks traditionally performed by humans, potentially leading to job displacement in specific sectors. The resulting shift in the job market would necessitate societal adjustments and a reevaluation of workforce training and education programs to ensure that individuals are equipped with skills relevant to an increasingly automated world.

Cultural homogenization is another concern arising from the dominance of certain languages and cultures in the data used to train LLMs. This dominance might lead to the underrepresentation or misrepresentation of minority cultures and languages. In diverse societies, this can perpetuate cultural inequalities and hinder efforts to preserve linguistic and cultural diversity. Addressing this issue requires careful consideration of the data sources used in LLM training and efforts to ensure that a broader range of cultures and languages are adequately represented.

The accessibility and digital divide concern associated with LLMs highlight the risk that the benefits of these technologies may not be evenly distributed across society. Those without access to advanced technology or the internet may be left behind, further widening existing disparities. To harness the full potential of LLMs, efforts must be made to bridge this digital divide, ensuring equitable access and usage across different socio-economic and geographical contexts.

The ethical use of LLMs in sensitive applications such as healthcare, legal, or law enforcement introduces specific ethical considerations. The potential for incorrect or biased advice generated by these models in such critical domains can have profound consequences. Striking a balance between leveraging the capabilities of LLMs in sensitive contexts while mitigating ethical risks becomes imperative, necessitating clear guidelines and rigorous scrutiny of their deployment.

References

[1] G. Xiao, J. Lin, M. Seznec, and H. Wu, “Smoothquant: Accurate and efficient post-training quantization for large language models,” *International*, 2023.

[2] Z. Fan, X. Gao, M. Mirchev, A. Roychoudhury, and S. H. Tan, “Automated Repair of Programs from Large Language Models,” in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 2023, pp. 1469–1481.

[3] Y. Huang *et al.*, “Behavior-driven query similarity prediction based on pre-trained language models for e-commerce search,” 2023.

[4] S. Mandvikar, “Factors to Consider When Selecting a Large Language Model: A Comparative Analysis,” *International Journal of Intelligent Automation and Computing*, vol. 6, no. 3, pp. 37–40, Aug. 2023.

[5] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy, “Challenges and Applications of Large Language Models,” *arXiv [cs.CL]*, 19-Jul-2023.
[6] Z. Wang, J. Wohlwend, and T. Lei, “Structured Pruning of Large Language Models,” *arXiv [cs.CL]*, 10-Oct-2019.
[7] I. Skelac and A. Jandrić, “Meaning as Use: From Wittgenstein to Google’s Word2vec,” in *Guide to Deep Learning Basics: Logical, Historical and Philosophical Perspectives*, S. Skansi, Ed. Cham: Springer International Publishing, 2020, pp. 41–53.
[8] S. Mandvikar, “Augmenting Intelligent Document Processing (IDP) Workflows with Contemporary Large Language Models (LLMs),” *International Journal of Computer Trends and Technology*, vol. 71, no. 10, pp. 80–91, 2023.
[9] J. Gesi, H. Wang, B. Wang, A. Truelove, J. Park, and I. Ahmed, “Out of Time: A Case Study of Using Team and Modification Representation Learning for Improving Bug Report Resolution Time Prediction in Ebay,” *Available at SSRN 4571372*.
[10] J. Wei, X. Wang, and D. Schuurmans, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in*, 2022.
[11] J. Austin, A. Odena, M. Nye, and M. Bosma, “Program synthesis with large language models,” *arXiv preprint arXiv*, 2021.
[12] N. Kandpal, H. Deng, A. Roberts, E. Wallace, and C. Raffel, “Large Language Models Struggle to Learn Long-Tail Knowledge,” in *Proceedings of the 40th International Conference on Machine Learning*, 23--29 Jul 2023, vol. 202, pp. 15696–15707.
[13] J. Gesi, X. Shen, Y. Geng, Q. Chen, and I. Ahmed, “Leveraging Feature Bias for Scalable Misprediction Explanation of Machine Learning Models,” in *Proceedings of the 45th International Conference on Software Engineering (ICSE)*, 2023.
[14] D. Zan, B. Chen, F. Zhang, D. Lu, and B. Wu, “Large language models meet NL2Code: A survey,” *Proceedings of the*, 2023.
[15] F. N. U. Jirigesi, “Personalized Web Services Interface Design Using Interactive Computational Search.” 2017.
[16] Y. Liu *et al.*, “Summary of ChatGPT-Related research and perspective towards the future of large language models,” *Meta-Radiology*, vol. 1, no. 2, p. 100017, Sep. 2023.
[17] Y. Shen *et al.*, “ChatGPT and Other Large Language Models Are Double-edged Swords,” *Radiology*, vol. 307, no. 2. pubs.rsna.org, p. e230163, Apr-2023.
[18] F. Jirigesi, A. Truelove, and F. Yazdani, “Code Clone Detection Using Representation Learning,” 2019.
[19] K. Singhal *et al.*, “Large language models encode clinical knowledge,” *Nature*, vol. 620, no. 7972, pp. 172–180, Aug. 2023.
[20] H. P. Kothandapani, “Applications of Robotic Process Automation in Quantitative Risk Assessment in Financial Institutions,” *International Journal of Business Intelligence and Big Data Analytics*, vol. 6, no. 1, pp. 40–52, 2023.
[21] T. Brants, A. C. Popat, P. Xu, F. J. Och, and J. Dean, “Large Language Models in Machine Translation,” 2007. [Online]. Available: http://research.google/pubs/pub33278.pdf.
[22] S. Biderman *et al.*, “Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling,” in *Proceedings of the 40th International Conference on Machine Learning*, 23--29 Jul 2023, vol. 202, pp. 2397–2430.
[23] J. Gesi *et al.*, “Code smells in machine learning systems,” *arXiv preprint arXiv:2203.00803*, 2022.
[24] H. P. Kothandapani, “Drivers and Barriers of Adopting Interactive Dashboard Reporting in the Finance Sector: An Empirical Investigation,” *Reviews of Contemporary Business Analytics*, vol. 2, no. 1, pp. 45–70, 2019.

[25] M. Anderson and S. L. Anderson, “Machine Ethics: Creating an Ethical Intelligent Agent,” *AIMag*, vol. 28, no. 4, pp. 15–15, Dec. 2007.

[26] J. Torresen, “A review of future and ethical perspectives of robotics and AI,” *Frontiers in Robotics and AI*, 2018.

[27] A. Groce *et al.*, “Evaluating and improving static analysis tools via differential mutation analysis,” in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS)*, 2021, pp. 207–218.

[28] I. J. Martínez-Martínez, “Ethical implications of digital advertising automation: The case of programmatic advertising in Spain,” *Profesional de*, 2017.

[29] M. Ryan, “The social and ethical impacts of artificial intelligence in agriculture: mapping the agricultural AI literature,” *AI Soc.*, Jan. 2022.

[30] J. Gesi, J. Li, and I. Ahmed, “An empirical examination of the impact of bias on just-in-time defect prediction,” in *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2021, pp. 1–12.

[31] C. B. Head, P. Jasper, M. McConnachie, L. Raftree, and G. Higdon, “Large language model applications for evaluation: Opportunities and ethical implications,” *New Dir. Eval.*, vol. 2023, no. 178–179, pp. 33–46, Jun. 2023.

[32] L. Weidinger *et al.*, “Ethical and social risks of harm from Language Models,” *arXiv [cs.CL]*, 08-Dec-2021.