

Original Article

The Importance of Governance and Guardrails in Generative Artificial Intelligence

Christopher Maduabuchi Okonkwo

¹Christopher is an Alumni of the Texas A&M– Commerce School of Business, Texas, USA.

¹Corresponding Author : mokonkwo@leomail.tamuc.edu

Received: 24 May 2024

Revised: 02 July 2024

Accepted: 21 July 2024

Published: 31 July 2024

Abstract - Generative AI is revolutionizing multiple industries through the creation of content, as well as automating selective and accelerated processes to enhance productivity and growth. This paper summarizes the past and presents various uses of generative AI and its potential future application. This work looks at the various aspects of governance in ensuring that the technology is optimally utilized and draws upon the various duties of key stakeholders in the governance structures and the benefits of sound governance frameworks for society.

Keywords - Generative AI, Governance, Machine Learning, Ethical considerations, Productivity, Innovation.

1. Introduction

Generative AI – a subset of machine learning technology that enables models to self-create new and original works – has actually been around for decades. Early work in the field was developed by computer science pioneers such as Alan Turing, while more recent approaches to generative technology emerged in the 1980s-1990s with some of the early scalable instantiation of coherent generation. These advancements developed methods like Hidden Markov Models and Gaussian Mixture Models. However, it was not until 2014, when Ian Goodfellow introduced Generative Adversarial Networks, that modern approaches to generative technology started to take form. Since then, the quality of generated images, sounds and text has become significantly more realistic thanks to generative AI—and it is this applicability across different types of content that is making so many take notice.

The contemporary use of generative artificial intelligence (AI) is revolutionizing the productivity and the creative potential of different sectors. In healthcare, generative AI uses patterns in large-scale datasets to enhance diagnostics for better patient outcomes. Generative AI is also being used to create content in post-production and editing, as well as new frontiers in music and literature in media and entertainment. However, this powerful technology comes with ethical, legal, and social implications that are practical barriers to its rapid implementation: job disruption due to automation, privacy and security concerns, unintended bias in training datasets, and regulatory challenges, among others. We highlight the governance and regulation needed for the ethical and responsible development of generative AI systems. Holistic governance is required for maximization of benefits from generative AI while minimizing risks, and imbuing trust and transparency among stakeholders.

2. Literature Survey

2.1. Governance and Guardrails in Generative Artificial Intelligence

Generative AI is the future generator of concepts, efficiency, and progress in different fields via creating new material and performing certain tasks. In their review, which Gupta et al. (2024) conducted, they outline the historical background of generative AI and its evolution, including its modern use compared to the possible future application of generative AI, as well as the importance of the technology's regulation. This paper discusses the history, from the past until now and furthermore in the future of generative AI, the liberal and adverse impacts for most industries and the necessity to continue the enhancement and correct utilization of AI through governance. Many of these also outline the roles of all the stakeholders in the governance structure and the benefits of having governance systems in society.

2.2. Generative AI: Its History, Present, and Future

Generative AI has been defined in a variety of ways, starting with the earliest AI concepts. Pioneers like Alan initiated the foundational work in the mid-20th century, but the development of machine learning methods in the 1980s and 1990s led to the emergence of accurate generative models (Liu et al., 2023). Early models, such as Hidden Markov and Gaussian Mixture Models, shaped subsequent development. According to Bandi et al. (2023), the 2010s saw significant development due to profound learning system innovation. Ian Goodfellow and others launched VAEs and GANs in 2014, revolutionizing AI and ML. With a two-tier structure, a generator creates fake data, and a discriminator checks and rates it to make it seem like competition (Bandi et al., 2023). This is how GANs changed generative AI. Generational AI has arisen across several domains and has excellent potential to generate vivid images, sounds, and realistic speech. OpenAI's lower GPTs



and GPT-4 release show how generative models can create a context-driven language that revolutionizes content creation, customer service, and other sectors.

There are exciting developments in DeepFakes and AI artistry, prompting critical ethical considerations about this groundbreaking technology. Further studies, including those by Benbya et al. (2020), indicate that generative AI will become increasingly prevalent. Its processing power and use of advanced algorithms will improve, allowing it to generate more relevant data in context. However, the drawbacks include ethical issues related to the development and application of AI, necessitating the adoption of more moral and transparent design principles (Konda, 2022). Therefore, people should not allow technological complexity to overshadow the issue of blame, as this would undermine human contribution in creating innovative and simple solutions to prevailing authoritative notions of superiority and contempt for humanity, especially when humanity has demonstrated its ability to reciprocate in the embrace of AI.

3. The Benefits of Generative AI Across Various Industries are Significant

Generative AI enables the transference of the concept to numerous industries and the enhancement of many sides connected with productivity, creativity, and trend growth. Generational AI has the potential to revolutionize the development of interim solutions and assessments of healthcare services (Ahmad et al., 2021). For example, it can sift through large data sets to identify relationships beyond the reach of human diagnostic officers and improve disease prognosis and diagnosis. This could go a long way toward improving the quality of patient care and overcoming most of the daily difficulties that healthcare workers face.

The creative aspects of AI could store innovative content, create tools, and enhance business operations in the media and entertainment sectors, necessitating an improvement in entertainment analysis (Ahmad et al., 2021). It enhances the development of various ideas in literature, music, and the arts that are unimaginable to ordinary people. Because these phases take minimal time and money, this technology is still useful in the post-production and editing of movies and television shows. According to Bail (2024), generative AI can create social media content that targets various age categories and genders, thereby increasing awareness and boosting sales. It helps construct better, more powerful ads that will create an irresistible appeal, increasing the number of responses. The industrial sector, which also benefits from generative AI, encompasses manufacturing, where generative AI enables the development and enhancement of designs and efficient production cycles (Elahi et al., 2023). Furthermore, the application of AI in logistics and supply chain management can help forecast demand, determine the correct stock levels, and choose the best delivery pathways, resulting in notably lower expenditures and increased productivity. The application of generative AI in algorithmic trading, fake detection, and other essential financial planning components

in finance can yield superior and more efficient results for clients (Petratos, 2021).

4. Negatives of AI in Different Industries and Sectors

As illustrated below, generative AI has challenges in various fields. This is because job outsourcing is one of the most formidable threats facing a company. Further development in AI systems has raised concerns about its potential impact on employment in manufacturing, retail, customer service, and other clerical functions (Elahi et al., 2023). Individuals should always avoid this because it has a strong probability of increasing economic risks and disparities. Strengthening security for generative AI is challenging due to fierce adversarial models.

Experts have identified manufactured content, particularly fake news, as a key component of fraud and other related crimes, such as hacking and burglary (Petratos, 2021). It erodes the credibility of digital content and presents specific opportunities, threats, and challenges to law enforcers and regulators. According to Gupta et al. (2024), the use of generative AI in writing papers or conducting legal research within the legal profession may limit the job availability of new lawyers and produce low-quality work or papers.

Furthermore, generative AI increases the following ethical concerns: reliability, justice, and prejudice. An AI obtains information from a data set and uses the training information set to reinforce already-established prejudices. Lam et al. (2024) suggest that ML antennas may impact different fields of human life, such as criminal justice, credit, and employment.

Having constant and consistent control is critical because AI decision-making systems often need to be more transparent and subjective (Bandi et al., 2023). This has caused the development of artificial intelligence to have dire environmental consequences.

Training an advanced AI model requires heavy computational infrastructure and human labor, both of which lead to increased energy consumption and, consequently, a larger carbon footprint. Ahmad et al. (2022) suggest that the advanced application of AI technologies increases energy consumption due to the energy-intensive nature of computers, prompting the development of new AI techniques and the integration of energy-efficient methodologies.

5. Importance of Implementing Governance and Guardrails in AI

Gupta et al. (2024) contribute to the debate by postulating that advancing good governance and more reflective utilization of generative AI is vital. With the increasing deployment and utilization of generative AI systems across various industries, it is crucial to establish appropriate security measures and guidelines for their safe and ethical deployment.

5.1. Important Moral Factors

Naturally generated text by an AI may contain bias, errors, or downright false information. Determining the ethical standards for artificial intelligence's output based on societal requirements and norms, achievable through the adoption of governance frameworks, is also crucial. The government needs to intervene in the last stage and prevent generating such undesirable content as deepfakes, hatred, and misusing information (Bontridder & Pouillet, 2021). As a result, the problem could potentially resurface in the future, leading to societal issues and a decline in stakeholder confidence.

5.2. Privacy and Security

Individuals must take all the essential risks that AI introduces to the system seriously, including data hacking (Bail, 2024). At each step described above, we prioritize protecting users' data and information when creating AI applications.

5.3. Regulatory Compliance

The use of data, its protection, and the application of artificial intelligence differ considerably across nations or countries, depending on the laws that govern them. According to Crepax (n.d.), legal structures enable organizations to operate within these paradigms of law, such as GDPR in Europe and CCPA in California, among others. They help eliminate legal issues where appropriate and contribute to the best international practices in AI applications.

5.4. Transparency and Accountability

Getting to the guardrails in AI governance increases the explanation of how the organization arrived at specific AI decisions. It fosters accountability by enabling a legal interpretation of the AI system's actions and decisions (Liu et al., 2023). Transparency similarly ensures that users, stakeholders, and regulators trust the system.

5.5. Mitigating Unintended Consequences

AI systems may have several adverse side effects, primarily due to their non-linear behavior. Governance structures always put in place checks and balances to review and optimize AI systems' needs for the delivery of their intended function, as well as to contain adverse effects should they occur.

6. Roles of Stakeholders in Developing and Implementing Governance and Guardrails in AI

Corporate governance and regulation of AI include all institutions, norms, and procedures that may prevent undesirable outcomes from AI deployment and encourage safe and successful use. Gupta et al. (2024) identified stakeholders as policymakers, business leaders, academics, civic society, and users who help oversee AI. Today, policymakers must enforce norms for AI technology use and innovation. They provide privacy, security, and ethical guidelines to prevent AI system misuse and ensure intended

conduct (Gupta et al., 2024). Industry elites, AI innovators, and technology employers still self-regulate using the highest standards and ethical norms (Ahmad et al., 2022). These pioneers established formal frameworks for AI products, including bias detection, explainability, and radical openness.

Academic thinking may benefit research on policy and business standards. Researchers examine AI's social impacts, abuse, and technological solutions. They also prepare the next generation of AI experts and policymakers. Civil society groups monitor AI usage and its effects. They want increased transparency in AI development and implementation, as well as the inclusion of people of color, women, and LGBTQIA+ individuals. The university teaches customers or corporations who want to incorporate artificial intelligence into their businesses that they should be knowledgeable about AI's benefits and shortcomings (Benbya et al., 2020). It is critical for overseers to understand the workings of AI, the permissions granted to it, and the ethical and correct operation of its tasks. This diversification of governance could involve several stakeholders addressing the development and implementation issues associated with the intricate issues of AI. Gupta et al. (2024) suggest that people must educate the public about the potential benefits of AI and demonstrate how we can utilize it to achieve positive outcomes without the current negative consequences.

7. Benefits of Implementing Governance and Guardrails in AI

According to Gupta et al. (2024), AI has the potential to assist organizations in enhancing their ethical actions, tasks, and protocols within the context of information technology, particularly given the need for increased tones and checks among organizations. The governance theoretical framework recommends four strategies for using AI in a morally responsible manner. On the other hand, responsiveness refers to a consumer's readiness to interact with a good before a seller or any other form of stimulus compels them to do so. Recoiling happens when buyers give a wide berth to things that capture their curiosity. Its coverage is defined by its extent; on the other hand, its evaluation reveals how consumers feel about the product (Bail, 2024). These enable AI systems to analyze end users without the contamination of biased feelings critically. Accounting controls allow responsible governance to evaluate such technologies, and regular audits can help eliminate biases (Barrett, 2002).

Gasser (2024) outlined that guardrail-equipped procedures are more formal and safer than those lacking well-defined standards. This framework applies guardrails to data security in model fitting, deployment, and management, thereby maintaining organizational integrity. In terms of the three dials, if they conduct and monitor exercises and quickly set up alert systems, anti-safe measures may prevent issues. Therefore, trustworthy AI systems in industries such as healthcare, finance, and AVs will always incorporate algorithms (Taeihagh, 2021).

Additionally, these two concepts govern decision making processes and ensure the legal admissibility of an entity's activities. The laws that governments enact across the world pertaining to artificial intelligence provide good governance for companies to ensure they operate within the provisions of the law and to prevent fines and other detriments that would potentially affect their international operations.

Gaining approval from customers, lenders, and government agencies can enhance the likelihood of receiving a favourable compliance audit (Barrett, 2002).

8. Summary

The innovative applications of AI, optimization, and other technologies constitute severe threats. Optimization is a toolkit of artificial intelligence, and it has its risks inbuilt into it. Based on such reasoning, people can deduce that compliance with rules and practices compatible with the management and governance systems under implementation could improve the job and reduce the risk associated with the same.

This means that all the various players, whether in the public or the private domain, need to unite with universities and Non-Governmental Organizations (NGOs) to forge governance. Thus, it may be possible to entice AI to produce innovations, act ethically, and be responsible for the outcomes. That is why when showing better governance, we eliminate the risks of harming trust and ethical standards for society to experience the positive effects of generative AI.

References

- [1] Tanveer Ahmad et al., "Energetics Systems and Artificial Intelligence: Applications of Industry 4.0," *Energy Reports*, vol. 8, pp. 334-361, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Zubair Ahmad et al., "Artificial Intelligence (AI) in Medicine, Current Applications and Future Role with Special Emphasis on its Potential and Promise in Pathology: Present and Future Impact, Obstacles Including Costs and Acceptance Among Pathologists, Practical and Philosophical Considerations. A Comprehensive Review," *Diagnostic Pathology*, vol. 16, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Christopher A. Bail, "Can Generative AI improve Social Science?," *Proceedings of the National Academy of Sciences*, vol. 121, no. 21, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Ajay Bandi, Pydi Venkata Satya Ramesh Adapa, and Yudu Eswar Vinay Pratap Kumar Kuchi, "The Power of Generative AI: A Review of Requirements, Models, Input-Output Formats, Evaluation Metrics, and Challenges," *Future Internet*, vol. 15, no. 8, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Pat Barrett, "Achieving Better Practice Corporate Governance in the Public Sector," *Australian National Audit Office*, 2002. [[Google Scholar](#)]
- [6] Hind Benbya, Thomas H. Davenport, and Stella Pachidi, "Artificial Intelligence in Organizations: Current State and Future Opportunities," *MIS Quarterly Executive*, vol. 19, no. 4, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Noémi Bontridder, and Yves Pouillet, "The Role of Artificial Intelligence in Disinformation," *Data & Policy*, vol. 3, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Tommaso Crepax, "Global Privacy Control and Portability of Privacy Preferences through Browser Settings: A Comparative Study of Techno-Legal Challenges Under the Ccpa/Cpra and the Gdpr," *SSRN*, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Mahboob Elahi et al., "A Comprehensive Literature Review of the Applications of AI Techniques Through the Lifecycle of Industrial Equipment," *Discover Artificial Intelligence*, vol. 3, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Urs Gasser, *Guardrails: Guiding Human Decisions in the Age of AI*, Princeton University Press, pp. 1-232, 2024. [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Ruchi Gupta et al., "Adoption and Impacts of Generative Artificial Intelligence: Theoretical Underpinnings and Research Agenda," *International Journal of Information Management Data Insights*, vol. 4, no. 1, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Sreedhar Reddy Konda, "Ethical Considerations in the Development and Deployment of AI-Driven Software Systems," *International Journal of Computer Science and Technology*, vol. 6, no. 3, pp. 86-101, 2022. [[Google Scholar](#)] [[Publisher Link](#)]

9. Conclusion

Generative AI will be at the top of new technologies that can definitely change industries and make productivity and creativity significantly higher. However, its advancement and implementation are relatively fast and present serious ethical and practical issues like job losses, sensitive data vulnerability, and existing prejudice. For such difficulties, the development of integral overall management systems is inevitable. These frameworks must make certain that the AI systems are functioning ethically and legally and are effectively transparent and accountable systems. Through engaging with these questions, it is possible for governance structures to build trust between stakeholders and enhance the proper implementation of generative AI in the various sectors.

Also note that Generative AI governance only works if implemented with the cooperation of policymakers, industry players, academics, civil sectors, and users. All the people involved are crucial in dictating the ethical and practical aspects of AI. The government should ensure that the laws safeguarding privacies and security are properly implemented, leaders in industries should act professionally and ethically, academics are beneficial in offering research data and training, and the public can support the causes that are sensitive to the minority. Consumers need to be enlightened on the pros and cons of AI in their use of it. These players can work in harmony to make certain that generative AI offers benefits that are not only economic but also social, consequently enhancing society.

- [13] Chiranji Lal Chowdhary et al., *Machine Learning for Mobile Communications*, CRC Press, pp. 1-214, 2024. [[Publisher Link](#)]
- [14] Yue Liu et al., "Generative Artificial Intelligence and its Applications in Materials Science: Current Situation and Future Perspectives," *Journal of Materiomics*, vol. 9, no. 4, pp. 798-816, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Pythagoras N. Petratos, "Misinformation, Disinformation, and Fake News: Cyber Risks to Business," *Business Horizons*, vol. 64, no. 6, pp. 763-774, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Araz Taeihagh, "Governance of Artificial Intelligence," *Policy and Society*, vol. 40, no. 2, pp. 137-157, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]