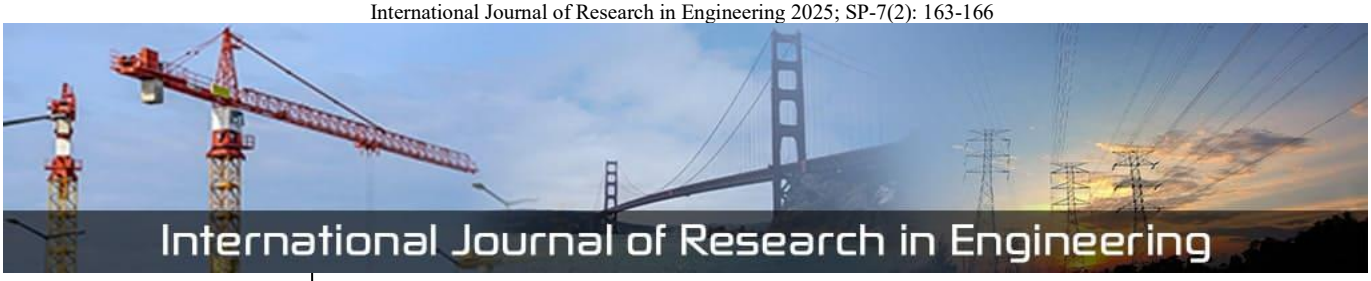

E-ISSN: 2664-8784
P-ISSN: 2664-8776
IJRE 2025; SP-7(2): 163-166

www.engineeringpaper.net
Received: 16-05-2025
Accepted: 18-06-2025

Sucheta Mor
Assistant Professor, DPG College, Gurugram, Haryana, India

Correspondence
Sucheta Mor
Assistant Professor, DPG College, Gurugram, Haryana, India

Two-Days National Conference on Multidisciplinary Approaches for Innovation and Sustainability: Global solution for contemporary Challenges-NCMIS (DPG Degree College: 17 th-18th 2025)

Balancing innovation and risk: Navigating the challenges of generative AI

Sucheta Mor

DOI: https://www.doi.org/10.33545/26648776.2025.v7.i2c.105

Abstract
Generative AI refers to a class of artificial intelligence models that can create new content such as text, images, music and more based on the data they have been trained on. This differs from traditional predictive ML that used to have narrow outcomes example spam classification. Though the generative AI can completely transform industries, improve productivity, creativity and enhance overall business operations, there should be rigorous research to evaluate Gen AI's effect on human cognition, metacognition and creativity. Education community must ensure that Gen AI must become powerful tool for enhancing knowledge and innovation rather than weakening our intellectual abilities and impacting academic integrity. The launch of open AI ChatGPT language generation model has raised alarms in academic sector also as the business world integrate large language models and artificial intelligence into workflows serious ethical concerns demand immediate action. Advancements in unchecked AI bring risk like bias in decision making, the spread of misinformation and growing copyright disputes.
In this author examines the use of Gen AI as a tool for human learning addressing its implications on human abilities.

Keywords: Generative AI, large language models, Chat GPT, innovation, challenges, education sector

1. Introduction
Now days Gen AI is transforming industries and enterprises and is no longer a futuristic concept. In every sector like healthcare, education, marketing, cybersecurity Ai models are revolutionizing how business think, interpret and operate. The global generative AI (Gen AI) market is projected to surge from $11.3 billion in 2023 to $51.8 billion by 2028, [1] signalling the speed at which this technology is reshaping the world. Although industries are racing to incorporate LLM and AI models into their workflows, serious ethical concerns demand immediate action.
Unchecked AI advancements bring tangible risks, including bias in decision-making, the spread of misinformation, and growing copyright disputes. Organizations that fail to implement responsible AI practices risk regulatory penalties, reputational damage, and eroded public trust.
The swift incorporation of Artificial Intelligence (AI) into education has brought significant advantages, but it also raises important risks and ethical issues. Traditional AI, which relies on rule-based systems and data-oriented decision-making, has been widely embraced in educational environments to improve administrative functions, tailor learning experiences, and offer predictive insights. While these systems are effective at automating routine tasks and delivering analyses based on past data, they have clear limitations. They frequently fall short in generating new content, engaging in complex interactions, or dynamically adapting to individual student needs. Consequently, their influence on the deeper elements of teaching and learning is limited.
In contrast, generative AI-a branch of AI capable of producing new content like text, images, and music-offers a significant advance over traditional forms of AI. This technology has the potential to transform learning experiences and teaching strategies by creating more engaging and personalized educational tools.

However, it also presents challenges regarding academic integrity, data privacy, and algorithmic biases. It is essential to recognize these risks and establish ethical guidelines for the application of generative AI in education to ensure that its advantages are maximized while its potential drawbacks are minimized.

Although conventional AI will still have a place in education, its shortcomings indicate that it might be gradually supplemented or replaced by more advanced generative AI technologies. This transition to generative AI is not merely an improvement; it represents a fundamental change that necessitates thoughtful evaluation of its transformative abilities as well as the ethical dilemmas it introduces.

As generative artificial intelligence (GenAI) advances, its incorporation into business operations will become more common. We observe its usage in areas like customer service automation with chat bots, creation of marketing content, and analysis of extensive datasets to generate valuable insights that inform business decisions. While these applications significantly improve process efficiency and streamline operations, they also pose risks if used without adequate human supervision.

One of the major challenges confronting generative AI (GenAI) is the propensity for large language models (LLMs) to produce hallucinations. These models are trained on extensive datasets sourced from the internet, which allows them to comprehend and generate human language. However, the quality of this data can vary significantly, which can lead to the generation of misleading or nonsensical information. When these inaccuracies are presented with confidence, they can be hard to differentiate from true statements. This has already led to instances of misplaced trust and, in certain cases, potentially harmful outcomes.

While hallucinations represent mistakes made by AI systems, there is a similarly alarming issue regarding the intentional use of AI to manipulate information, commonly referred to as deepfakes. Technologies such as deepfakes and voice cloning have already been exploited to imitate political figures, sway public opinion, and create discord. For instance, an AI-generated robocall once impersonated a U.S. presidential candidate, urging voters to abstain from participating in the New Hampshire primary. Although these tactics can be detected at the national level, they are often more challenging to identify in state or local elections, where resources for cybersecurity tend to be more limited.

Deepfakes exploit a broader societal dilemma regarding trust. Unlike hallucinations, where AI inadvertently produces false information, deepfakes are deliberately crafted to mislead. The core issue lies in AI's capacity to undermine confidence in our sensory perceptions, as individuals start to question what they see and hear. This type of manipulation threatens to destabilize society's trust in reality, making it increasingly difficult to distinguish between truth and falsehood, and posing a significant risk to democracy, governance, and public discourse. Whether we are dealing with the unintentional errors of large language models or the intentional deceit of deepfakes, the overarching concern is evident: as AI technology progresses, our methods for establishing trust and protecting against both inadvertent and deliberate misuse must also evolve. Without these protective measures, the increasing reliance on AI could jeopardize critical decision-making processes in various sectors, from business to public policy.

Operational efficiency and product innovation are crucial for any company, especially within the financial services sector, where consumer obligations take precedence. While generative AI holds the potential to elevate the financial services industry, its integration carries certain risks, such as data privacy issues, security concerns, and the appropriate handling of data. Utilizing generative AI and AI, in general, responsibly and ethically is not only a regulatory necessity but also a critical business requirement. Consumers depend on the financial services sector for various needs, including banking, loans, credit access, and wealth management. This dependence means that organizations have access to sensitive consumer data, and clients must trust that their information is safeguarded and used in a responsible manner. Additionally, any data-driven decisions should be made without bias. A financial institution's failure to uphold this trust can damage its relationship with customers. Therefore, we consider the responsible and ethical use of AI fundamental to our strategy. Our governance frameworks are built on intentional and thoughtful decisions that allow us to foster innovation while simultaneously building trust within the wider community.

2. The Issue of Bias in Gen AI

One of the most pressing concerns in LLM AI and generative models is bias. AI systems learn from vast datasets that may reflect historical prejudices or systemic inequalities. If these biases go unchecked, AI-generated content can reinforce discrimination, leading to ethical and reputational risks. For example, Amazon scrapped its AI recruitment tool in 2018 after it was discovered that the system systematically discriminated against female candidates, favouring male applicants due to historical hiring data. Similarly, facial recognition technology has demonstrated significant racial bias, with studies from the National Institute of Standards and Technology in 2019 showing that Black and Asian individuals were misidentified at rates up to 100 times higher than white individuals. Such algorithmic biases not only deepen societal disparities but also pose legal and ethical challenges for organizations deploying AI-driven solutions [2].

Mitigating AI Bias

Addressing bias in AI systems require proactive, multi-layered approach to ensure fairness and accuracy. Organizations can implement several key strategies to minimize unintended discrimination in AI-generated outputs.

- **Dataset should be of diverse demographics and perspectives:** AI models should be trained on larger datasets that are balanced and are of high-quality that represent diverse demographics and perspectives. This helps reduce biases that stem from historically skewed data.
- **Regular assessment:** Regular assessments of AI-generated outputs can help detect disparities in decision-making and ensure the system is functioning equitably. Independent audits and fairness metrics can be used to continuously refine AI models.
- **Bias Correction Algorithms:** Techniques such as reweighting datasets, adversarial debiasing, and fairness constraints can be applied to mitigate biases and

improve model transparency.

3. Misinformation and Deepfake Risks

Gen AI's ability to create human-like text, images, and videos introduces risks of misinformation and deepfakes. Fake news articles, AI-generated social media posts, and manipulated images can deceive the public, spread false narratives, and harm trust in digital information. The content that has been created using Gen AI is often indistinguishable from human-created content, making it difficult to verify authenticity. Additionally, bad actors can leverage Generative AI to craft deceptive political propaganda, falsified financial reports, or harmful deepfakes videos that erode public trust in the media.

1. Strategies to Mitigate Misinformation Risks

As Gen AI becomes more advanced, the risk of AI-generated misinformation and deepfakes grows. Organizations can implement proactive strategies to address these challenges that improve transparency, accountability, and content integrity.

- **AI-generated Content Verification Tools:** Automated verification systems can cross-check AI-generated outputs against reliable sources to detect and flag misleading or fabricated content.
- **Regulatory Compliance Measures:** Aligning AI deployments with emerging regulations [3] and ethical guidelines ensures transparency and responsible use of AI in content creation.
- **Digital Literacy Initiatives:** Employees and consumers can be educated on how to identify AI-generated misinformation reduces susceptibility to fake content and increase public trust in digital media.

4. Copyright and Intellectual Property Concerns

The rise of Gen AI has blurred the lines of intellectual property (IP) ownership. AI-generated works, including written content, music, and artwork, raise questions about authorship and legal rights.

How AI Challenges Copyright Laws

As the existing copyright laws are not designed to accommodate AI generated content so the courts worldwide are now debating how to regulate AI driven creativity, who holds the right whether it is AI developer, the user or the original dataset owner

Best Practices for Managing AI and Copyright

As Gen AI continues to develop, organizations must navigate the complexities of intellectual property (IP) and copyright laws to ensure ethical and legal compliance. Several best practices can help mitigate risks associated with AI-generated content.

- **Watermarking:** To track ownership, improve transparency, and prevent unauthorized replication AI generated content can be embedded with watermark.
- **Copyright-Compliant Training Data:** AI models should be trained on properly licensed, open-source, or public domain datasets to reduce the risk of unintentional copyright infringement.
- **Legal Compliance and Governance:** Organizations must stay informed about evolving copyright laws and industry regulations to ensure AI-generated works align with intellectual property rights.

5. Threat to academic integrity

With ChatGPT professors and lecturers should be worried about the future of essays as a form of assessment. According to the International Centre for Academic Integrity (2021) [4], academic integrity is defined as a commitment to six fundamental values: honesty, trust, fairness, respect, responsibility and courage. As such, when a person uses ChatGPT to generate essays or other forms of written texts that are then passed off as original work, it violates the core principles of academic integrity. ChatGPT raises similar concerns as the well documented commercial 'contract cheating' in higher education (Newton, 2018). The only difference is that ChatGPT is free and easily accessible to all users. It also offers users the opportunity of interaction. Users can tweak their queries to know how different the responses can be. This means that there are possibilities of generating different texts/essays and the user can pick the best out of the lot. One academic was quoted in a Nature commentary recently (Stokel-Walker, 2022) [5] saying that "at the moment, it's looking a lot like the end of essays as an assignment for education".

Also a tested, validated and accepted tool to identify dishonest use of AI text generators in academia is not yet available. That means that it is still easy to pass off an output from ChatGPT as an original work without detection. To address this challenge, OpenAI has developed a free tool (AI text Classifier13) trained to distinguish between AI-written and human-written texts. Unfortunately, this has been described as an 'imperfect tool' by OpenAI who warned that it should not be used as a primary decision-making tool. How academic institutions and publishers will implement OpenAI's imperfect tool or develop better tools remains unclear. It is important to note that there is a greater concern for institutions in Low-and-middle-income countries where Turnitin and other plagiarism tools are yet to be integrated as measures for academic integrity. Technical integration of these tools costs money which many of the institutions in these countries do not have (Damian Okaibedi Eke) [6].

Research indicates that while AI systems can enhance educational experiences by providing personalized learning and real-time feedback, they also raise significant privacy and ethical issues. For instance, AI tools like Proctorio, which monitors students during exams to prevent cheating, have been criticized for increasing test-taking anxiety and creating an uncomfortable atmosphere for students (McArthur, 2020) [7]. Moreover, AI systems like Squirrel AI, designed to provide adaptive learning experiences, can sometimes restrict students' creative learning by overly tailoring educational content to individual needs, potentially limiting broader educational development (Beard, 2020) [8]. These concerns highlight the need for careful monitoring and regulation of AI technologies in educational settings to ensure they support rather than hinder the learning process [9].

As the use of AI by students in assessments has become more prevalent and difficult to detect, there are increased concerns in the academic community regarding academic integrity. Inviting open dialogue and involvement from students in assessment and feedback mechanisms can bring promising results in student success and academic quality. The key consideration for educators is to ensure students are using AI to develop creative thinking, not replace it.

The widespread use of AI among students, utilizing it

weekly or daily, indicates that generative AI has already integrated significantly into education. This swift adoption calls for a proactive response from educators and administrators to maximize the benefits of AI while minimizing potential risks. The favourable views on AI's role in enhancing understanding and efficiency create opportunities for innovative teaching methods. However, the ethical issues and risks of academic dishonesty underscore the necessity for clear guidelines and instruction on responsible AI use. As proposed by Chiu *et al*. (2023) [10], creating comprehensive AI literacy programs could help tackle these challenges by promoting critical thinking and ethical considerations regarding AI usage.

6. Conclusion

I contend that the use of ChatGPT and other AI text generators could potentially jeopardize academic integrity, while also having the power to transform academia. It's our collective duty to address the risks associated with maintaining academic integrity for the sake of progress. Achieving this requires a collaborative effort from various stakeholders, including technical developers, academic policymakers, publishers, professors, lecturers, and students. While academic writing, essay assignments, and technical coding assessments are not obsolete, it's essential to rethink and implement significant changes to uphold sustainable integrity within academia.

7. References

1. MarketsandMarkets. Generative AI market https://www.marketsandmarkets.com/Market-Reports/generative-ai-market-142870584.html
2. Kloud9. Navigating ethical dilemmas in generative AI: Balancing innovation and responsibility https://www.kloud9.nyc/blogs/navigating-ethical-dilemmas-in-generative-ai-balancing-innovation-and-responsibility
3. AIDI. Cloud infrastructure providers | AI infrastructure | AI service providers | AI driven infrastructure https://aidi.ai
4. International Center for Academic Integrity. The fundamental values of academic integrity. 3rd ed. https://academicintegrity.org/aws/ICAI/pt/sp/home_page
5. Stokel-Walker C. AI bot ChatGPT writes smart essays - should professors worry? Nature https://www.nature.com/articles/d41586-022-04397-7
6. Eke DO. ChatGPT and the rise of generative AI: Threat to academic integrity? https://www.researchgate.net/publication/368713523
7. McArthur S. The effects of AI monitoring on student anxiety. Int J Educ Technol. 2020.
8. Beard L. Adaptive learning systems and their limitations. Educ Technol J. 2020.
9. Nistala S, Liu R, Yi Z, Lu M, Huang H. Balancing innovation and ethics: Transformative potential and ethical challenges of generative AI in education. 2024.
10. Chiu MM, *et al*. AI literacy in K-12: A systematic literature review. Int J STEM Educ. 2023.