MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 02 (2026) | Pages: 386-392

Challenges in AI Development: A Multi-Dimensional Study of Risks and Solutions

Swapnaja Patwardhan[1], Omkar Babar[2], Manish Kenjale[3], Atharva Mazire[4], Ajinkya Thakur[5], Onkar Savaratkar[6]

[1,2,3,4,5,6]Department of MCA, MES' IMCC, Pune

[1]sp.imcc@mespune.in, [2]ombabar04@gmail.com, [3]manishkenjale07@gmail.com, [4]atharvamazire7245@gmail.com, [5]ajinkyaathakur@gmail.com, [6]onkars9818@gmail.com

Peer Review Information

***Type**: Article*
***Received**: 20 March 2026*
***Revised**: 03 April 2026*
***Accepted**: 21 May 2026*
***Published**: 03 June 2026*

Abstract

Artificial Intelligence (AI) is playing an increasingly important role in transforming various industries such as healthcare, finance, education, and business operations. Although its adoption has grown significantly and offers numerous advantages, the development of AI systems still faces several major challenges that affect their reliability and ethical use. This study focuses on important issue including algorithmic bias, issues related to data privacy, lack of transparency in decision-making, unclear legal accountability, high energy consumption, and the impact of automation on employment. By combining real-world case studies with existing theoretical perspectives, this research aims to provide a deeper understanding of the complex and interconnected nature of these challenges.

The findings reveal that biased datasets can lead to unfair and discriminatory outcomes, while the lack of Understandability in AI systems reduces user trust and accountability. Additionally, increasing dependence on large-scale AI models raises critical issue regarding data security and environmental impact due to high computational requirements. The study also focuses on the growing impact of automation on employment, emphasizing the need for reskilling and workforce adaptation.

Based on the analysis, the paper suggests practical solutions such as implementing fairness- aware algorithms, adopting privacy-preserving techniques, promoting explainable AI models, and encouraging sustainable AI practices. The results shows that a balanced approach combining technical improvements, ethical considerations, and regulatory frameworks is essential for developing responsible and trustworthy AI systems. Overall, this research provides a comprehensive explainability of AI challenges and offers strategies to ensure its sustainable and ethical growth.

Keywords: Artificial Intelligence; Algorithmic Bias; Data Privacy; Explainable Artificial Intelligence; Sustainable Artificial Intelligence; Workforce Adaptation.

How to Cite This Article

Patwardhan, S., Babar, O., Kenjale, M., Mazire, A., Thakur, A., & Savaratkar, O. (2026). Challenges in AI development: A multi-dimensional study of risks and solutions. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2), 386–392.

Introduction

Artificial Intelligence (AI) has become one of the most influential technologies in recent years, transforming the way organizations operate in sectors such as healthcare, banking, education, law, and manufacturing. By enabling machines to process data, recognize patterns, and make predictions, AI has improved efficiency and automation in many complex tasks. From simple rule-based systems to advanced deep learning architectures, AI has evolved into a powerful tool capable of solving real-world problems at scale [1][13].

Recent progress in generative AI has further accelerated this transformation. Technologies such as large language models and image-generation systems are now capable of producing human-like text, code, and visual content. These innovations are creating new opportunities in business operations, research, communication, and creative industries. However, despite their impressive capabilities, such systems may still generate misleading, inaccurate, or fabricated outputs, which creates risks when used in sensitive or critical applications [3][10][11].

Alongside technical advancements, AI development also introduces major ethical and societal concerns. One of the most significant issues is algorithmic bias, where systems trained on unbalanced or historical datasets may produce unfair or discriminatory outcomes. This becomes especially problematic in domains like hiring, criminal justice, and financial decision-making, where biased predictions can directly affect human lives. Ensuring fairness in AI systems has therefore become a major research priority [4][5][16].

Another growing challenge is the issue of privacy and data security. AI models often depend on massive datasets that may include personal or sensitive information. Improper data collection, weak protection mechanisms, or unauthorized access can expose individuals to privacy violations and cyber risks. At the same time, many AI systems operate with limited explainability, making it difficult for users to understand how decisions are reached. This lack of transparency reduces trust and complicates accountability when errors occur [14][17][18].

Environmental sustainability is also emerging as an important concern in AI development. Training and deploying large-scale AI models requires extensive computational resources, which consume significant electricity and contribute to carbon emissions. As AI adoption expands globally, balancing innovation with sustainable computing practices has become increasingly necessary [2][6][7]. In addition, AI-driven automation is changing workforce structures, replacing some traditional jobs while simultaneously creating demand for new digital skills [9][20].

Because of these interconnected challenges, AI development is no longer only a technical matter but also a multidisciplinary issue involving ethics, law, governance, sustainability, and social adaptation. Addressing these concerns requires responsible design practices, clear regulations, and collaboration among researchers, policymakers, and industry leaders. This paper examines the major challenges in AI development and explores practical solutions for building transparent, fair, secure, and sustainable AI systems.

Adoption of Generative AI Across Industry Sectors

Generative AI has rapidly become an important technology across many sectors, including business, education, healthcare, and digital media. Applications such as ChatGPT and similar intelligent systems are designed to create human-like text and assist users in completing a wide variety of tasks. Their growing adoption is changing how organizations manage communication, automate services, and improve productivity in day-to-day operations.

In business environments, generative AI supports several functional areas such as customer interaction, marketing, finance, technical assistance, and administrative management. These systems can respond to customer inquiries through virtual chat interfaces, assist employees with repetitive office tasks, and help generate promotional material such as campaign content and advertising drafts. They are also useful in collaborative settings where teams use AI tools for brainstorming, drafting reports, and improving workflow efficiency.

Although generative AI offers many advantages, it also introduces several important risks. These systems can sometimes generate incorrect, misleading, or fabricated responses, which may affect decision-making if outputs are accepted without verification. While they can save time, lower operational expenses, and enhance creativity, human review remains essential before applying AI-generated results in real-world scenarios.

Another major concern involves privacy and confidential data protection. When organizations input sensitive business information into AI platforms, there is a possibility of unintended exposure if proper safeguards are not followed. To reduce such risks, companies often limit confidential data sharing, disable storage of conversation histories, and restrict AI access in highly sensitive departments. In response to growing privacy concerns, some institutions and governments have introduced regulations controlling the use of generative AI technologies [11][12].

Case Studies and Practical Solutions

AI challenges can be broadly categorized in ethical, technical, legal, and socio- economic dimensions. These dimensions are interconnected

and often influence each other. For instance, biased data can lead to discrimination (ethical issue), which may further result in legal consequences.

The following case studies illustrate real- world failures of AI systems and provide insights into practical solutions that can be implemented to mitigate these risks.

Algorithmic Bias

Algorithmic bias is basically unintentional but deeply embedded in data collection and model design processes. Machine learning models learn patterns from historical data, and if this data reflects societal inequalities, the model will replicate and amplify them. Bias can be categorized into several types:

- Selection Bias – when training data is not representative
- Measurement Bias – incorrect data labeling
- Algorithmic Bias – flaws in model design

In recent research, synthetic data generation and fairness-aware algorithms have been proposed to address bias. However, completely eliminating bias remains a challenge due to the complexity of real-world data.

Case 1 (Recruitment): Amazon developed an AI-based hiring tool to automatically screen resumes and identify suitable candidates efficiently. The system was trained using previous hiring data collected over a period of ten years. Thus, this data mainly reflected a male-dominated workforce, which introduced an unintended bias into the model. As a result, the AI began favouring resumes that resembled past male applicants. It penalized resumes containing the word "women's," even in positive contexts such as leadership roles or achievements. Additionally, candidates from women's colleges were ranked lower, regardless of their qualifications. This issue highlighted how AI systems can inherit and amplify biases present in training data. The problem was not due to intentional discrimination but rather a lack of balanced data and fairness checks during development. Eventually, Amazon discontinued the system after identifying its biased behaviour.

Case 2 (Justice System): The COMPAS algorithm is widely used in the United States judicial system to assess the likelihood of a defendant committing another crime. It is often used to support decisions related to bail, sentencing, and parole. However, investigations revealed that the system exhibited racial bias in its predictions. It was found to incorrectly classify Black defendants as high-risk at nearly twice the rate compared to white defendants. At the same time, some white defendants were inaccurately labelled as low-risk. This raised serious issue about fairness and equality in legal decision- making. The root of the problem lay in the previous data used to train the model, which reflected existing societal and systemic biases. Furthermore, the lack of transparency in how the algorithm made decisions made it difficult to question or correct its outcomes. This case sparked widespread debate on the ethical use of AI in critical domains.

Solution:

To reduce such biases, it is important to conduct regular bias audits using tools like the What-If Tool and other fairness evaluation techniques. Training data should be carefully pre-processed to ensure it represents diverse groups and does not favour any particular demographic. Post-processing methods can also be applied to adjust model outputs and improve fairness. In addition, explainable AI techniques should be used so that decisions made by the system can be clearly understood and evaluated. Organizations must also ensure continuous monitoring of AI systems after deployment. Most importantly, human oversight should be maintained in critical decision-making processes to prevent unfair outcomes and ensure ethical use of AI.

Data Privacy Issue

The rapid growth of AI has led to the accumulation of vast amounts of personal data. This raises serious issue regarding user consent, data ownership, and surveillance.

Modern AI systems often rely on centralized data storage, which increases vulnerability to cyberattacks. Additionally, the use of AI in surveillance systems has sparked debates about civil liberties and human rights.

Advanced privacy-preserving techniques such as:

- Differential Privacy
- Secure Multi-party Computation
- Homomorphic Encryption

are being explored to enhance data security while maintaining model performance.

Case 1 (Surveillance): Facial recognition technology has been widely adopted for surveillance purposes in many regions, often without the

clear knowledge or consent of the public. These systems collect and store large amounts of biometric data, including facial features of millions of individuals. While such technology is promoted for security and law enforcement, it raises serious issue about privacy and misuse. In many cases, people are monitored continuously in public spaces without being aware of it. The collected data can be vulnerable to cyberattacks, leading to potential identity theft or unauthorized access. There is also a risk that governments or organizations may misuse this data for mass tracking or control. The lack of proper regulations and transparency further increases these issue. This case highlights the growing tension between technological advancement and individual privacy rights.

Case 2 (Data Exposure): In early 2023, a technical bug in an open-source library used by OpenAI led to an unexpected data exposure issue. Due to this flaw, some users were able to view titles of other users' chat histories. Although full conversations were not exposed, even partial access raised serious issue about data privacy. Chat history titles can still reveal sensitive or personal information about users. This incident showed how even small technical errors can lead to significant privacy risks. It also highlighted the challenges of relying on third-party libraries in large-scale systems. The issue was quickly identified and resolved, but it emphasized the importance of strong security practices. This case demonstrates the need for continuous testing and monitoring in AI systems handling user data.

Solution:

Privacy risks can be reduced by adopting federated learning, where data remains on user devices instead of being centrally stored. Strong end-to-end encryption should be used to protect data both in storage and during transmission. Regular security audits and vulnerability testing must be conducted to identify potential weaknesses. Developers should carefully evaluate and update third-party libraries to avoid hidden risks. Access control mechanisms should be implemented to limit unauthorized data exposure. Transparency in data collection and user consent is also essential. Together, these steps can help build more secure and privacy-focused AI systems.

Legal Responsibility

The legal landscape of AI is still evolving, and there is no universally accepted framework for assigning responsibility in case of AI failures. One major challenge is the **"**black box problem**"**, where it is difficult to trace how an AI system arrived at a particular decision. This makes accountability unclear and complicates legal proceedings.

Emerging approaches suggest:

- Defining AI liability laws
- Creating audit trails for AI decisions
- Establishing regulatory bodies for AI governance

Case 1 (Autonomous Safety): In 2018, an incident involving an Uber self- driving car resulted in the tragic death of a pedestrian during a road test. The vehicle was operating in autonomous mode, with a human safety driver present inside the car. However, the system failed to correctly detect and respond to the pedestrian in time. Investigations revealed that both the software and human monitoring had limitations that contributed to the accident. This case raised serious issue about the safety and reliability of autonomous vehicles. It also created confusion about accountability, as it was unclear whether the responsibility lay with the driver, the developers, or the manufacturer. The incident highlighted the risks of deploying AI systems in real- world environments without full reliability. It also emphasized the need for clear legal and ethical frameworks in autonomous technologies.

Solution:

To improve safety, human-in-the-loop systems should be implemented so that crucial decisions are not left entirely to AI. Clear guidelines must define the roles and responsibilities of developers, operators, and organizations. Detailed documentation, such as Model Cards, should be maintained to explain system limitations and expected behaviour. Regular safety testing and real-world simulations should be conducted before deployment. Transparency in system design can help identify risks early. Legal frameworks must also be developed to assign accountability in case of failures. Together, these steps can ensure safer and more responsible use of autonomous systems.

Energy and Computational Demand

The environmental impact of AI is becoming increasingly significant due to the large-scale infrastructure required for training and deployment.

Data centers consume enormous amounts of electricity and contribute to global carbon emissions. Additionally, water consumption for cooling systems adds to environmental issue.

Recent advancements aim to reduce this impact through:

- Efficient hardware (TPUs, GPUs optimization)

- Edge computing
- Energy-efficient algorithms

This concept, often referred to as Sustainable AI or Green AI, is gaining importance in research communities.

Case 1 (Electricity): Training large-scale AI models such as GPT-3 requires massive computational power and energy consumption. These models are trained on powerful hardware like GPUs and TPUs that run continuously for days or even weeks. As a result, the electricity used during training leads to a significant carbon footprint. Studies have shown that training a single large model can produce emissions comparable to multiple cars over their entire lifetime. This raises serious environmental issue, especially as AI adoption continues to grow. The demand for more complex models further increases energy usage. This case highlights the hidden environmental cost of advanced AI technologies. It emphasizes the need to balance innovation with sustainability.

Case 2 (Water Usage): Data centres that support AI systems require efficient cooling systems to maintain high-performance hardware. These cooling systems often rely on large amounts of water to prevent overheating. In some cases, millions of gallons of water are used daily, especially in large-scale facilities. This becomes a serious issue in regions that already face water scarcity or drought conditions. The continuous demand for water can strain local resources and affect surrounding communities. Many people are unaware of this indirect environmental impact of AI infrastructure. This case highlights how AI development can also affect natural resources beyond electricity. It calls attention to the importance of sustainable infrastructure planning.

Solution:

To reduce environmental impact, organizations should adopt Green AI practices that focus on efficiency and sustainability. Techniques like model quantization and pruning can be used to reduce model size and computational requirements. Energy-efficient hardware and optimized algorithms should be prioritized during development. Whenever possible, AI systems should be hosted on cloud platforms powered by renewable energy sources. Data centres can also adopt advanced cooling methods that use less water. Monitoring energy and resource usage can help identify areas for improvement. These steps can make AI development more environmentally responsible.

Transparency and Explainability

Explainability is crucial for building trust in AI systems, especially in sensitive domains such as healthcare and finance.

- Lack of transparency can lead to:
- Reduced user trust
- Difficulty in debugging models
- Legal and ethical issues
- Explainable AI techniques can be divided into:
- Model-based (interpretable models)
- Post-hoc explanations (SHAP, LIME)

These techniques help stakeholders understand and validate AI decisions.

Case 1 (Medical Trust): IBM Watson Health was developed to assist doctors by providing AI-based recommendations for cancer treatment. The system aimed to analyse large amounts of medical data and suggest effective treatment options. However, it later faced criticism when some of its recommendations were found to be unsafe or not based on real patient data. In certain cases, the system relied on hypothetical or incomplete information, leading to unreliable suggestions. Doctors began to lose trust in the tool because it could not clearly explain how it arrived at its decisions. This lack of transparency made it difficult for medical professionals to verify or rely on its outputs. As a result, the adoption of the system declined significantly. This case highlights the importance of trust and reliability in healthcare AI systems.

Case 2 (Financial Rejection): AI-driven credit scoring systems are widely used by financial institutions to evaluate loan applications. These systems analyse various factors such as income, credit history, and spending patterns to make decisions. However, many applicants face rejection without receiving a clear explanation for the outcome. This lack of transparency makes it difficult for individuals to understand what went wrong or how they can improve their eligibility. In some cases, the models may also reflect hidden biases present in the data. Customers are left with no clear way to challenge incorrect or unfair decisions. This can lead to frustration and loss of trust in financial systems. The issue highlights the need for fairness and clarity in automated decision-making.

Solution:

To improve transparency, Explainable AI (XAI) techniques such as SHAP and LIME should be used to clearly show how decisions are

made. These methods help identify which features influenced the final outcome. Providing users with understandable explanations can increase trust in AI systems. It also allows errors or biases to be detected and corrected more easily. Organizations should ensure that explanations are simple and user-friendly. Regular validation of models can further improve reliability. Combining transparency with fairness can make AI systems more accountable and trustworthy.

Social Impact: Job Displacement

AI-driven automation is reshaping the global workforce. While some jobs are being replaced, new roles are also emerging in AI development, data analysis, and system management.

The key challenge is the skill gap between traditional workers and AI-driven job requirements.

To address this:

- Governments must invest in education and training
- Companies should promote reskilling programs
- Educational institutions must integrate AI literacy.

Case 1 (Service Sector): The rapid growth of generative AI chatbots has significantly impacted the customer service industry. Many companies are increasingly adopting automated systems to handle customer queries, as they are faster and more cost- effective than human employees. As a result, thousands of workers in call centres and support roles have faced job losses. While these AI systems can efficiently manage routine and repetitive tasks, they often lack emotional understanding and human interaction skills. This shift has created issue about job displacement and the future of employment in service-based sectors. Employees with traditional skill sets may struggle to adapt to changing job requirements. At the same time, new roles are emerging in AI development and management. This situation highlights the growing gap between technological advancement and workforce readiness.

Solution:

To address this challenge, the focus should be on using AI for augmentation rather than complete replacement of human workers. AI systems should handle repetitive tasks while humans manage complex and sensitive interactions. Organizations must invest in reskilling and upskilling programs to help employees adapt to new roles. Governments should support initiatives that promote AI literacy and digital skills. Educational institutions also need to update their curriculum to match industry demands. Collaboration between companies and policymakers can create better employment opportunities. This balanced approach can ensure both innovation and job sustainability.

Comparative Analysis of AI Frameworks

In addition to the above comparison, it is important to note that different industries adopt different AI frameworks based on their requirements.

For example:

- Healthcare prioritizes accuracy and safety
- Finance emphasizes risk management and transparency
- Education focuses on personalized learning and accessibility
- Manufacturing emphasizes automation and operational efficiency
- Retail prioritizes customer experience and demand prediction

Table 1. AI Challenges and Solutions

Challenge Category	Industry Impact	Primary Mitigation Strategy	Technical Goal
Algorithmic Bias	Systematic discrimination	Fairness-aware preprocessing	Eliminate historical bias
Data Privacy	Sensitive information leakage	Federated learning and encryption	Zero-trust data architecture
Environmental Impact	High carbon footprint	Model pruning and Green AI	Energy-efficient inference
Explainability	Lack of user trust	SHAP/LIME interpretability methods	Transparent decision-making
Social Impact	Workforce displacement	Collaborative Human–AI roles	Economic resilience

Conclusion

The development of Artificial Intelligence has reached a stage where its impact is not only technological but also social, ethical, and environmental. This study highlights that challenges such as algorithmic bias, data privacy risks, lack of transparency, high energy consumption, and job displacement are interconnected and require careful consideration. These issues show that AI systems cannot be treated as purely technical tools, as they directly influence human lives and decision-making processes.

The findings suggest that addressing the challenges requires approach that integrates technological advancement with ethical considerations and appropriate regulatory measures support. Implementing fairness-aware algorithms, improving data protection methods, promoting explainable AI models, and adopting sustainable computing practices can significantly improve the reliability and trustworthiness of AI systems. Additionally, preparing the workforce through reskilling and education is essential to reduce the negative impact of automation.

In conclusion, the future of AI depends on developing systems that are not only intelligent but also transparent, fair, and accountable. A collaborative effort between researchers, policymakers, and industry professionals is necessary to ensure that AI continues to benefit society while minimizing its risks.

References

1. Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson, 2021.
2. Zhou, Zhi-Hua, et al. “Green AI: Energy-Efficient Algorithms.” *Nature Machine Intelligence*, 2022.
3. OpenAI. “March 20 ChatGPT Outage and Data Leak Post-Mortem.” 2023.
4. Angwin, Julia, et al. “Machine Bias: Risk Assessment Algorithms in Criminal Sentencing.” *ProPublica*, 2016.
5. American Civil Liberties Union. “Amazon’s Automated Hiring Tool and Gender Bias.” 2018.
6. “AI Models and Their Carbon Footprint Impact.” *CarbonCredits.com*, 2026.
7. “Data Centers and Environmental Impact of AI.” *Lincoln Institute of Land Policy*, 2025.
8. “The Rise and Fall of IBM Watson Health.” *Healthcare Digital*, 2026.
9. “Statistics on AI Replacing Jobs.” *AIPRM*, 2024.
10. Chui, Michael, et al. “The Economic Potential of Generative AI.” *McKinsey & Company*, 2022.
11. Ray, Siladitya. “ChatGPT Bans and Restrictions Across Countries.” *Forbes*, 2023.
12. Stancati, Margherita, and Sam Schechner. “Global Issue Over AI Privacy Risks.” *The Wall Street Journal*, 2023.
13. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
14. Floridi, Luciano, et al. “AI Ethics: Challenges and Solutions.” *Philosophy & Technology*, 2018.
15. Jobin, Anna, Marcello Ienca, and Effy Vayena. “The Global Landscape of AI Ethics Guidelines.” *Nature Machine Intelligence*, 2019.
16. Mittelstadt, Brent D., et al. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society*, 2016.
17. Kairouz, Peter, et al. “Advances and Open Problems in Federated Learning.” *Foundations and Trends in Machine Learning*, 2021.
18. Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. “Why Should I Trust You? Explaining the Predictions of Any Classifier.” *Proceedings of the ACM SIGKDD Conference*, 2016.
19. Lundberg, Scott M., and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems*, 2017.
20. Brynjolfsson, Erik, and Andrew McAfee. *The Second Machine Age*. W.W. Norton & Company, 2014.