

Risks Associated with the Use of Artificial Intelligence: A Comprehensive Review

Sujal Mutgekar¹, Sanika More², Vipul Zope³, Vedant Bhangale⁴

¹⁻⁴Research Scholars, Vishwakarma Institute of Information Technology, Pune, India
Correspondence: mutgekarSR@gmail.com

Abstract

Artificial Intelligence (AI) is increasingly being adopted across healthcare, finance, manufacturing, cybersecurity, education, and other sectors because of its ability to support prediction, automation, pattern recognition, and decision-making. However, the growing use of AI also introduces technical, security, economic, ethical, and social risks. This study examines the major risks associated with AI and classifies them into data privacy and security, algorithmic bias, lack of transparency and explainability, cybersecurity threats, economic and employment effects, and ethical and social concerns. A qualitative review-based methodology is used to examine relevant research and established AI risk-management frameworks. The study further discusses how these risks can affect different industries and reviews mitigation approaches including data governance, explainable AI, cybersecurity controls, human oversight, continuous monitoring, and governance mechanisms. The findings indicate that AI risk management should be addressed throughout the AI lifecycle rather than only after system deployment. A coordinated approach involving technical safeguards, organizational practices, human oversight, and appropriate governance can support more responsible and risk-aware deployment of AI systems.

Keywords: Artificial Intelligence; AI Risks; Data Privacy; Algorithmic Bias; Explainable AI; Cybersecurity; AI Governance; Responsible AI; Risk Management.

1. Introduction

Artificial Intelligence (AI) has evolved into a widely used technology for tasks such as prediction, pattern recognition, decision-making, and automation. Its applications extend across healthcare, finance, manufacturing, education, transportation, and cybersecurity. The growing adoption of AI has improved efficiency, scalability, and decision support in many applications. However, its increasing integration into real-world systems also introduces technical, security, ethical, and societal risks [1], [2].

AI systems often depend on large datasets and complex models whose behaviour can be difficult to interpret or predict. Changes in data, model limitations, and system context can affect the reliability and safety of AI-based decisions. These concerns become more significant when AI is used in applications involving healthcare, financial services, employment, or other decisions that may directly affect individuals [1].

Data privacy and security are also important concerns because AI systems may process large quantities of personal or sensitive information. In addition, risks related to bias, lack of transparency, misinformation, malicious use, and socioeconomic effects have been identified in research on AI and advanced AI systems [2], [3].

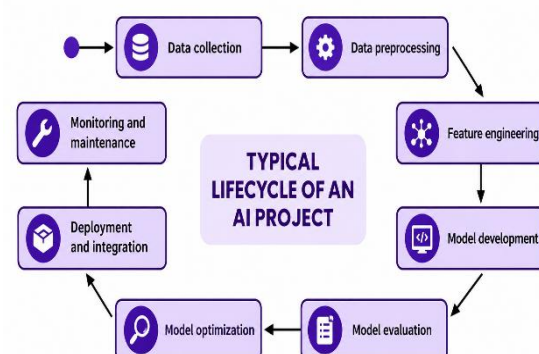


Figure 1. General workflow of an Artificial Intelligence system

Therefore, AI risk assessment requires consideration of both technical and socio-technical factors. Effective risk management should address relevant risks throughout the AI lifecycle, including system design, development, deployment, evaluation, and monitoring [1]. This study examines major risks associated with the use of AI, their effects across different sectors, and approaches for reducing these risks.

2. Objectives of the Study

The objectives of this study are:

- To identify the major risks associated with the use of Artificial Intelligence (AI).
- To classify AI risks into technical, security, ethical, economic, and social categories.
- To examine the potential impact of AI-related risks across different industries.
- To review approaches for mitigating identified AI risks through appropriate technical, organizational, and regulatory measures.
- To highlight the importance of responsible and risk-aware AI deployment.

3. Literature Review

Artificial Intelligence (AI) risks have been studied across technical, ethical, security, and societal dimensions. Existing research identifies concerns related to data quality, privacy, fairness, transparency, cybersecurity, misuse, and the broader effects of AI deployment [1], [2].

Data-related risks are important because AI systems depend on datasets for training, testing, and evaluation. Inadequate documentation, incomplete datasets, and inappropriate data collection can affect model reliability, transparency, and accountability [4]. Research on algorithmic fairness has also demonstrated that differences in training data and model performance can result in unequal outcomes across demographic groups [5].

The limited interpretability of complex AI models has increased the importance of Explainable Artificial Intelligence (XAI). Methods such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) provide

approaches for examining the factors contributing to model predictions [6], [7]. Such techniques can support transparency and assist users in understanding model behaviour.

AI systems are also exposed to cybersecurity threats. Research on adversarial examples has shown that carefully designed inputs can cause machine-learning models to produce incorrect predictions [8]. Similarly, poisoning attacks can manipulate training data and influence the behaviour of the resulting models [9]. These findings demonstrate that AI security requires consideration of both the model and its underlying data.

Recent research has further examined risks associated with advanced AI systems, including misinformation, harmful content, privacy concerns, malicious use, and other societal effects [2], [3]. These studies indicate that AI risk is multidimensional and requires consideration beyond model accuracy alone.

Recent work published in IRJIST has also examined AI-based cybersecurity applications. Bisht [10] proposed an explainable machine-learning framework for detecting phishing and deepfake-related threats in digital payment systems. The framework combines behavioural analysis with SHAP and LIME and emphasizes interpretability, continuous monitoring, and human review for security-related decisions [10].

Overall, the reviewed literature shows that AI risks can arise from data, model behaviour, security vulnerabilities, limited transparency, and wider societal impacts. These findings support the need for a comprehensive approach to identifying and mitigating AI-related risks throughout the AI lifecycle [1]–[10].

4. Methodology

This study adopts a qualitative review-based methodology to examine the major risks associated with the use of Artificial Intelligence (AI). Relevant information was considered from peer-reviewed research papers, technical publications, and established AI risk-management frameworks [1]–[10].

The identified literature was analyzed and organized into major risk categories, including data privacy and security, algorithmic bias, transparency and explainability, cybersecurity, economic and employment effects, and ethical and social risks.

The study further examines the potential effects of these risks across different industries and reviews commonly proposed mitigation approaches, including data governance, explainable AI, cybersecurity measures, human oversight, continuous monitoring, and regulatory mechanisms.

The findings are synthesized to provide a concise overview of AI-related risks and the approaches that can support responsible and risk-aware AI deployment.

5. Classification of AI Risks

The risks associated with Artificial Intelligence can be grouped into technical, security, economic, ethical, and social dimensions. These risks may originate from the data used by AI systems, model behaviour, system vulnerabilities, deployment practices, and the broader context in which AI is applied [1], [2].

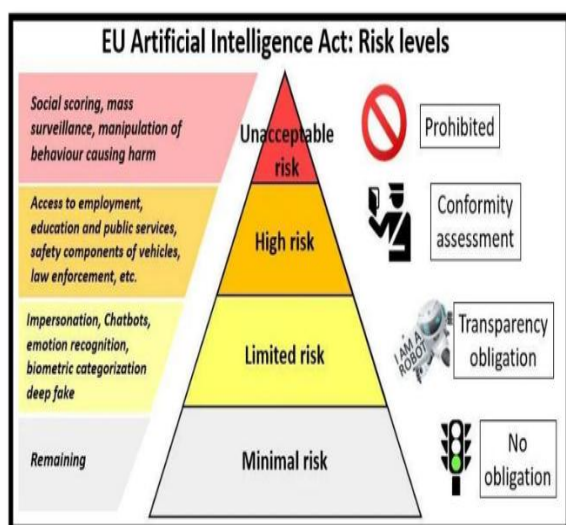


Figure 2. Risk levels under the EU Artificial Intelligence Act. Source: Telefónica [11].

5.1. Data privacy and security risks

AI systems often require large volumes of data for training and operation. When these datasets contain personal or sensitive information, improper collection, storage, processing, or access can create privacy and security risks [1], [4].

Inadequate data protection may result in unauthorized access, data exposure, or inappropriate use of information. Effective data governance, access controls, and appropriate data-handling practices are therefore important for reducing these risks [1].

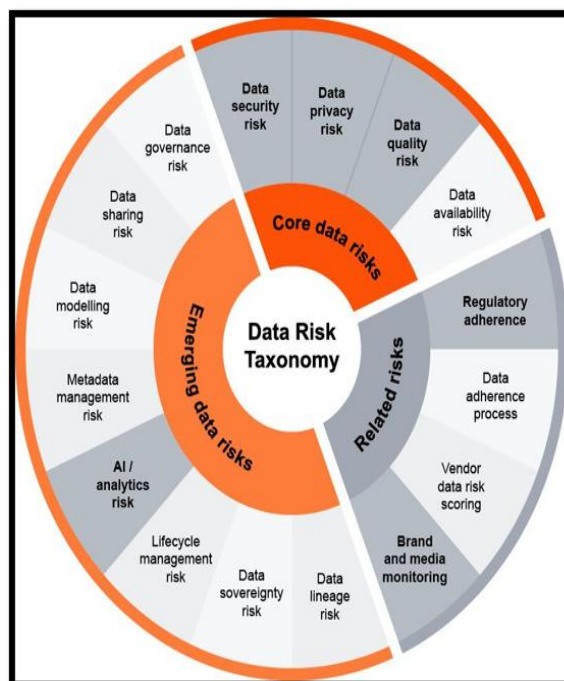


Figure 3. Data risk taxonomy. Source: PwC Singapore [13].

5.2. Algorithmic bias and fairness issues

AI models can reflect biases present in their training data or development processes. Such biases may produce different outcomes for different groups, particularly when AI is applied to areas such as employment, financial services, or other high-impact decisions [5].

Bias reduction requires appropriate dataset selection, evaluation across relevant groups, and continuous monitoring of model performance. Fairness should therefore be considered throughout the AI development and deployment process [1], [5].

5.3. Lack of transparency and explainability

Complex AI models may provide predictions without clearly revealing the factors responsible for individual outputs. Limited interpretability can make it difficult for users and organizations to understand, evaluate, and challenge AI-supported decisions [1], [6].

Explainable AI methods such as LIME and SHAP can provide information about the contribution of features to model predictions and support greater transparency [6], [7]. However, explanations should be interpreted according to the model and application context.

5.4. Cybersecurity threats and vulnerabilities

AI systems can be affected by attacks targeting either the model or its training data. Adversarial inputs may cause incorrect model predictions, while data-poisoning attacks can influence model behaviour by manipulating training information [8], [9].

AI can also be incorporated into malicious activities, increasing the importance of security testing, monitoring, robustness evaluation, and appropriate access controls [1], [2].

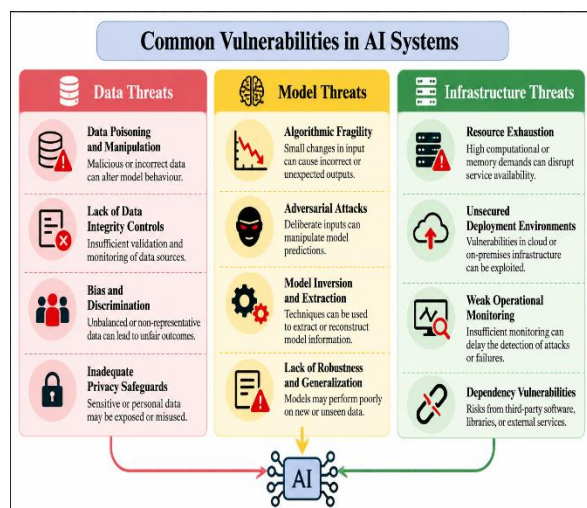


Figure 4. Common vulnerabilities in AI systems. Adapted from Markovate [12].

5.5. Economic and employment risks

The adoption of AI can change the nature and distribution of work by automating selected tasks and altering workforce requirements. These changes may create demand for new skills while affecting occupations involving tasks that can be automated.

The economic effects of AI therefore depend on factors such as the type of work being automated, the rate of technological adoption, workforce adaptation, and the availability of reskilling opportunities.

5.6. Ethical and social risks

AI deployment can create ethical and social concerns related to accountability, fairness, human control, privacy, misinformation, and misuse. These concerns become particularly important when AI systems influence decisions affecting individuals or society [2], [3].

Responsible AI deployment therefore requires appropriate human oversight, transparency, accountability, and consideration of the potential social consequences of AI-based systems [1], [2].

6. Impact on Industries

AI-related risks can affect industries differently depending on the nature of their applications, data requirements, and level of dependence on automated decision-making [1], [2].

In healthcare, errors in AI-based diagnostic or decision-support systems may affect clinical decisions and patient safety. Data privacy is also important because healthcare applications may process sensitive personal information [1].

In the financial sector, AI systems are used for activities such as risk assessment, fraud detection, and decision support. Bias, inaccurate predictions, security vulnerabilities, and limited explainability can affect financial decisions and user trust [1], [5], [10].

In manufacturing, AI is increasingly associated with automation, monitoring, prediction, and process optimization. Model errors, inadequate monitoring, or system failures may affect production processes and operational continuity.

In cybersecurity, AI can support threat detection and security monitoring while also creating new attack surfaces. Adversarial manipulation, data poisoning, and malicious use of AI can therefore introduce additional security challenges [8], [9].

The impact of AI risks is consequently dependent on the application context. Effective risk management should consider the potential consequences of system failure, the sensitivity of processed data, and the level of human oversight required in each industry [1].

7. Mitigation Strategies

AI-related risks require coordinated technical, organizational, and regulatory measures. Risk mitigation should be incorporated throughout the AI lifecycle rather than addressed only after deployment [1]. Improving data quality, documentation, and diversity can help reduce errors and potential bias in AI systems [1], [4], [5]. Appropriate data governance should also address access, storage, privacy, and responsible data use.

Explainable AI methods can improve transparency by providing information about the factors contributing to model predictions [6], [7]. Human oversight remains important, particularly for high-impact applications where automated decisions may have significant consequences.

Cybersecurity controls should include secure data handling, access control, robustness testing, continuous monitoring, and evaluation against adversarial threats [1], [8], [9]. Organizations should also monitor deployed models for changes in performance and emerging risks.

Regulatory and organizational frameworks can further support responsible AI deployment by defining accountability, risk-management procedures, documentation requirements, and appropriate oversight [1], [2]. The combination of technical safeguards, human supervision, organizational controls, and governance mechanisms can help reduce AI-related risks.

8. Future scope

As AI technologies continue to develop, new forms of risk may emerge with their integration into connected and autonomous systems. Future research should focus on improving the robustness, transparency, security, and reliability of AI systems [1], [2].

Further work can examine AI risk in emerging applications involving interconnected systems, advanced generative models, and large-scale automated decision-making. Greater attention is also required for continuous risk assessment, explainability, privacy protection, and responsible deployment.

The development of effective governance frameworks and continued collaboration between researchers, organizations, and regulators can support safer and more accountable use of AI technologies [1], [2].

9. Conclusion

Artificial Intelligence offers significant opportunities for automation, prediction, and decision support across different sectors. However, its use also introduces risks related to data privacy, algorithmic bias, transparency, cybersecurity, economic effects, and ethical and social concerns [1], [2].

These risks can affect both AI systems and the individuals or organizations that depend on them. Effective risk management therefore requires appropriate data governance, explainability, cybersecurity measures, human oversight, continuous monitoring, and suitable governance frameworks [1], [4], [6].

A lifecycle-based approach to AI risk management can help identify potential risks during system development, deployment, and operation. Addressing these risks through coordinated technical, organizational, and regulatory measures can support more responsible and reliable use of AI technologies.

References

- [1] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology, NIST AI 100-1, 2023, DOI: 10.6028/NIST.AI.100-1.
- [2] L. Weidinger et al., "Taxonomy of Risks Posed by Language Models," *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 214–229, 2022, DOI: 10.1145/3531146.3533088.
- [3] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623, 2021, DOI: 10.1145/3442188.3445922.
- [4] T. Gebru et al., "Datasheets for Datasets," *Communications of the ACM*, vol. 64, no. 12, pp. 86–92, 2021, DOI: 10.1145/3458723.
- [5] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research*, vol. 81, pp. 77–91, 2018.
- [6] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016, DOI: 10.1145/2939672.2939778.
- [7] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

- [8] A. Ilyas et al., “Adversarial Examples Are Not Bugs, They Are Features,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [9] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Šrđić, P. Laskov, G. Giacinto, and F. Roli, “Poisoning Attacks against Support Vector Machines,” *Proceedings of the 29th International Conference on Machine Learning*, pp. 1467–1474, 2012.
- [10] P. Bisht, “AI-Based Detection of Phishing Attacks in Digital Payment Systems Using Explainable Machine Learning,” *International Research Journal of Innovation in Science and Technology (IRJIST)*, vol. 1, no. 3, 2026, Article ID IRJIST-2026-083, DOI: 10.67308/irjist.083.
- [11] Telefónica, “EU Artificial Intelligence Act: Risk levels,” Telefónica, 2022.
- [12] R. Sharma, “AI Security: Threats, Frameworks, Best Practices & More,” Markovate, July 15, 2025.
- [13] PwC Singapore, “Data Risk Framework,” PwC Singapore, 2026.

Publisher’s Note & Copyright

IRJIST Journals remains neutral regarding jurisdictional claims in published maps and institutional affiliations; the views expressed are solely those of the authors.

© 2026 by the authors. Open access under the CC BY 4.0 license.
