Challenges in AI–Cloud Integration: A Comprehensive Review

Vividha Bahety
School of Information Technology Indira University, Pune

Dr. Dhanashri Kulkarni
School of Information Technology Indira University, Pune

DOI: 10.29322/IJSRP.16.02.2026.p17028
https://dx.doi.org/10.29322/IJSRP.16.02.2026.p17028

Paper Received Date: 5th January 2026
Paper Acceptance Date: 10th February 2026
Paper Publication Date: 12th February 2026

Abstract

Artificial Intelligence (AI) has been broadly integrated with cloud computing. The integration of these technologies has involved a lot of attention as it can deliver scalable, smart, and economically feasible solutions in numerous fields. Despite this opportunity for scalability, intelligence and economic viability for various applications using AI and cloud computing combined, there are many challenges as well. This paper provides a systematic insight and discussion regarding the key challenges in AI and cloud computing integration. A systematic literature review has been done in compliance with PRISMA. In this process, scientific publications in journals from 2019 to 2025 were considered for this research. Based on this systematic literature survey, it has been found that even though AI and cloud computing are helping in scalability, automation and optimizing processes in terms of time and economic viability, risks and concerns regarding security, privacy, latency, regulatory compliance and ethical considerations are still to be properly addressed.

Keywords: Artificial Intelligence, Cloud Computing, Data Security, Data Privacy, Ethical concerns

Introduction

The coupling of Artificial Intelligence (AI) with cloud computing has been noticed as a major technological paradigm for developing intelligent, elastic and economical digital solutions in various fields such as healthcare, finance, smart systems and enterprise solutions. Cloud computing infrastructure extends elastic computing capabilities and storage capacities needed for handling data-intensive tasks in AI whereas AI optimizes cloud services by introducing automation, smart analytics and intelligence in decision-making. This has significantly contributed in hastening the adoption of AI-based applications in various fields.

However, in spite of the above benefits there are number of challenges associated with the integration of cloud and AI technology combined. Normally the systems require tremendous amounts of data which are all processed in the cloud environment. This makes the systems prone to various security issues and potential breaches of various regulations. Also issues like network latency and inefficient resource allocation result in the unfitness of the cloud AI systems in contexts requiring mission-critical tasks to be performed in real-time. Other issues associated with these systems regard ethical considerations which might not be suitable in the context of the cloud environments.

Many of such challenges have already been investigated by existing literature. However existing review studies on this subject have mostly focused on singular dimensions of such challenges such as security measures, performance issues or regulatory issues. Consequently, existing research has still not explored such issues related to AI-cloud integration from an adequate number of dimensions. Differing from existing surveys on this subject that have mostly focused on isolated issues. This research paper aims to conduct an integrated review on such issues related to AI-cloud integration using a systematic literature review following the guidance of PRISMA.

As such the focus of this manuscript is on the systematic analysis and synthesis of existing work carried out on the integration of AI & Cloud. This is in a bid to determine the major challenges, trends and research issues surrounding the integration. This is expected to provide a comprehensive outlook that will enable people to design cloud systems that are safe, ethical, and sustainable using AI.

Objectives:

To identify major security and privacy concerns and ethical issues arising from the integration of AI with Cloud Computing to explore rules and policies that come with combining AI and Cloud.

Background of AI and Cloud Integration:

Artificial Intelligence comprises processes like machine learning, deep learning, natural language processing etc. Where a large number of computations and constant accessibility to a large number of data points are required. These needs are fulfilled by cloud computing on a service model defined by its Infrastructure as a Service model, Platform service model, Software service model, Software service model, Software service model respectively. Currently the service model for Artificial Intelligence is gaining motion. The Artificial Intelligence Service Model by which clients can access pre-built models using cloud infrastructure. Artificial Intelligence cloud integration offers innovation and flexibility but network connectivity issues, issues with sharing infrastructure, third-party cloud service providers are involved with these systems.

Research Methodology:

This study will follow the SLR methodology informed by the PRISMA framework to ensure transparency, rigor and duplicability of the review process. The PRISMA approach has been chosen for its clear systematic stages of identification, screening, assessment and synthesizing the literature on integrating AI with cloud computing. The review process in general is visually summarized in a PRISMA flow diagram showing visually how studies were selected and filtered through the different stages of the review process.

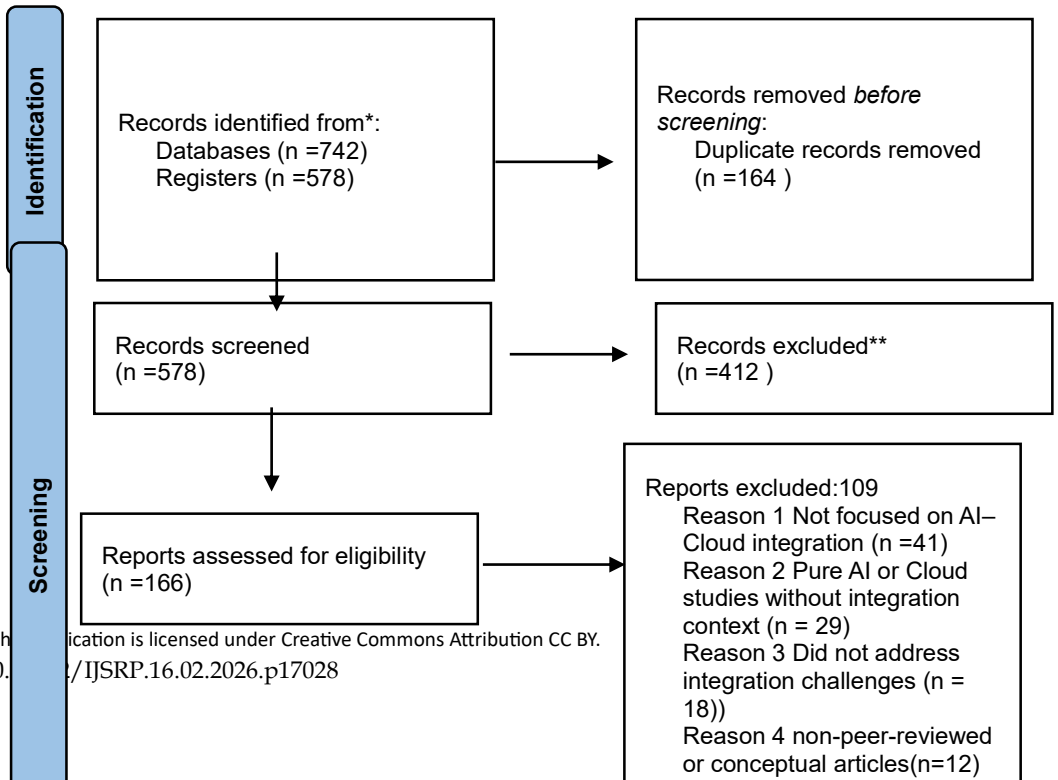
Data Sources and Search Strategy

To ensure a fair coverage of quality research in this domain several research databases were searched collectively: Scopus, IEEE Xplore Digital Library, ACM Digital Library, ScienceDirect (hosted by Elsevier) and Google Scholar. These databases were selected for their vast coverage of research literature in the form of journals and proceedings published in the domains of artificial intelligence, cloud computing and distributed computing. The research was constrained to papers published between 2019 and 2025 to include current advancements in cloud-AI integration.

A structured searching approach was adopted by utilizing key terms and logical operators to focus the searches and increase their relevancy. Some of the important search terms utilized were Artificial Intelligence, Cloud Computing, AI-Cloud Integration, AI-as-a-Service, Cloud Security and Privacy, Performance and Latency in Cloud AI, Ethical AI, Regulatory Compliance and Edge-Cloud Computing. Additionally backwards searching was also adopted as an approach by searching the reference lists of the articles chosen to find more studies that met the selection criteria.

Study Selection Process

Study identification was conducted according to the regular PRISMA protocol involving four phases of identification, screening, eligibility and inclusion. During the identification phase the databases generated an initial set of 742 records. However, the removal of 164 records representing duplicate literature resulted in the consideration of the remaining 578 articles. During the screening phase of the literature search the review of the study tiles and summaries led to the removal of 412 studies as they did not relate to the integration of AI and clouds. During the subsequent eligibility phase the assessment of the full text of 166 articles led to the removal of 109 studies as they did not relate appropriately to the reviewed objective. Only 57 studies satisfied the inclusion criteria to proceed to the analysis phase. This literature search flow is depicted in the accompanying PRISMA flow diagram in this paper.



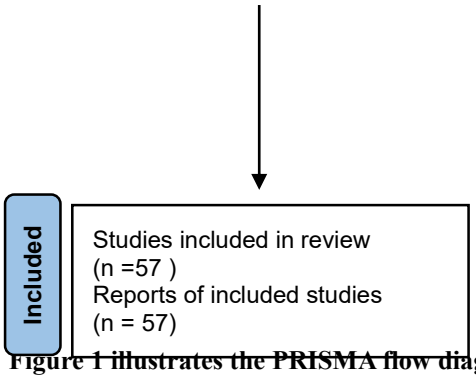


Figure 1 illustrates the PRISMA flow diagram of the study identification, screening, eligibility, and inclusion process.

Inclusion and Exclusion Criteria

To ensure consistency and relevance, specific criteria were formulated for both inclusion and exclusion in the course of reviewing literature. These criteria include studies that were published in the form of refereed articles or papers in English-language conferences/journals between 2019 and 2025 focusing on the integration of AI in either cloud or hybrid clouds/edge clouds. Studies that covered matters concerning security, privacy, performance, ethics, regulatory, organizational or cost viewpoints were considered for eligibility. Those that centered specifically on AI or clouds, but unrelated to the combination. In addition to articles that were not refereed like editorials, tutorials, blog posts or books were considered for exclusion.

Data Extraction and Synthesis

From each of these selected studies relevant data were extracted based on a predefined data extraction framework. The data extracted were on publication details the context of application, challenges identified, proposed solutions and limitations reported. Thematic synthesis was applied for analyzing the extracted data which facilitated the identification of patterns, converging findings and unresolved research gaps. Accordingly, this method allowed performing a comparative analysis throughout the selected studies and facilitated identifying integrated thematic insights.

Though 57 studies included in the thematic synthesis meet the criteria for inclusion, only a few selected and representative studies have been discussed above to counter repetition to focus only on evidence that has played a very prominent role in defining these emerging themes.

Methodological Rigor and Reliability

This review is based on the PRISMA framework and clear selection criteria enhance its methodological rigor and make its findings more reliable. Systematic documentation of each stage in the processes leading to the literature selection and presentation of the workflow through a PRISMA flow diagram minimizes selection bias. Hence enhancing reproducibility. This structured approach ensures that the findings are informed by careful selection and critical evaluation of the literature. Thus, offering a very strong foundation for future research on AI-cloud integration.

Literature Review

Contrary to offering individual study summaries for isolated findings this research paper uses a comparative theme analysis of past studies to clarify overarching challenges and unresolved study questions surrounding AI and cloud integration. This analysis shows that technical, governance-oriented and organizational aspects of AI/cloud integration are interlinked but often viewed singularly.

Security and Privacy Challenges

Security and privacy issues have always been highlighted as the main areas of concern in AI-cloud based systems. Many research studies [1, 2, 12, 13] have indicated that the use of AI applications is associated with immense data processing tasks being performed on AI-cloud systems thus making them more vulnerable to unauthorized access, data leakage and inference attacks. However, to address such issues associated with AI-cloud systems, AI-based security solutions like machine learning-powered intrusion detection and anomaly-based systems detection have emerged as prominent solutions proposed by research studies [1, 9,10].

Privacy preservation faces some other issues especially when it comes to regulated fields. Various researchers [2, 12, 13] point out that traditional privacy preservation techniques like encryption and access control can fail when large-scale machine learning tasks come into focus. Some privacy-preserving techniques like federated learning and decentralized computations have been proposed in literature to handle expose of sensitive data. However extensive experimentation of such techniques has yet to be conducted in a production setup of a cloud platform.

Performance, Latency, and Resource Management

Centralized cloud-based deployments of AI have limitations due to latency, bandwidth constraints and ineffective use of resources. Various research studies [3,6,10] have experimented and shown that data transfer delays and resource contention badly

affect inference speed and system responsiveness, especially for real-time applications. Hybrid cloud-edge architectures have been gaining increasing attention in recent times. Empirical results confirm that distributing computation closer to the sources of data reduces latency and network congestion [4,16].

However, comparisons across these studies show that this performance gain comes with significant added complexity in the architecture. All studies [4, 8], point out difficulties arising from workload coordination, system interoperability and dynamic resource management in distributed environments. Furthermore, without standardized performance benchmarks and deployment models, comparability of the results across different platforms is heavily restricted. This reflects a need for common investigation methodology.

Integration Complexity and Vendor Lock-in

Integration complexity and vendor lock-in are the issues under constant focus in the research involving AI and cloud computing. The dependence on cloud computing platforms and services along with their APIs creates restrictions on the portability of the data and makes it difficult to migrate as confirmed in various research studies [5, 7, 8]. Despite the potential of containers and microservices in improving the portability of the data between the cloud computing platforms its application in production is an un-explored area in the existing research literature [5, 6].

Regulatory Compliance Challenges

Regulatory compliance is a complementary but related challenge in the integration of AI and cloud. Works that have focused on the regulatory aspects [2, 12, 18] highlighted the challenges of complying with data protection laws, accountability requirements and cross-border data governance with AI workloads distributed across several cloud regions. These works highlight that mechanisms for regulatory compliance are normally applied as an external constraint but not embedded into the design of the system. hence limiting their applicability to dynamic cloud environments.

Ethical Considerations in AI–Cloud Systems

Ethical issues range beyond regulatory compliance into areas such as fairness, transparency, explainability, and accountability of AI-driven decision-making. Various works [9, 10, 15] show that cloud-based AI systems heighten these ethical risks due to their limited ability to Peer into model behaviors and due to the general opacity of proprietary platforms. Explainable AI and ethical governance frameworks are widely discussed. Still there is an indication from the literature that most of these approaches are conceptual with limited real-world implementation or evaluation. This gulf between ethical intent and operational practice points out the need for lifecycle level ethical embedding.

Organizational and Cost-Related Constraints

Additionally, organizational and economic considerations play an important part in the adoption of solutions based on AI cloud technology. Various research articles [8, 10] reveal that a lack of professionals with combined expertise in cloud computing and AI causes a delay in its efficient implementation. Another major hindrance is the increased cost for computations for the implementation of AI models in the cloud on a small to medium-scale level. Although methods for cost-effective resource management have already been suggested in previous research articles, a comparative study on its influence on cost-effective security, performance and ethical management is relatively rare.

Synthesis and Research Gaps

Overall, the literature reviewed above has confirmed that AI-cloud integration leads to greater scalability and efficiency. There are number of studies [1, 3, 10, 12] that clearly state that the literature has mostly treated individual challenges and has in consequence provided fragmentary solutions. The lack of comprehensive frameworks that take into consideration technical, security, regulatory, accountability, as well as organizational requirements, has, in fact, become a very pressing issue that has yet to be resolved by further research work as will be addressed in this paper by way of thematic synthesis.

Discussion

In a critical analysis of existing studies. It has been revealed that despite its immense benefits a majority of studies on the integration of AI and clouds focus on addressing the challenges independently instead of considering their relationships. The majority of studies paid attention on security focus on threat management while studies focusing on performance primarily ignore its relationships with management considerations. Cloud and edge infrastructures for hybrids increase performance but result in complexities related to governance, interoperability and compliance that remain unaddressed in a majority of studies.

Frequently regulatory and ethical aspects are considered on the conceptual plane with minimal practicality and empirical support. Organizational and cost-related issues make it even more difficult to apply them especially with respect to organizations with limited resources. It is important that the research gap of missing unifying and adaptive strategies addressing technical proficiency, security, ethics, regulatory aspects, and economic viability be recognized and filled to make AI cloud ecosystems responsibility and scalable.

Thematic Summary Table: Key Findings and Research Gaps

Theme	Key Challenges Identified	Synthesized Findings	Research Gaps
Security and Privacy	Data breaches, model inference attacks, multi tenancy risks	There are numerous pieces of research claiming to improve threat detection using AI based security systems but with greater system complexity overheads [1, 2, 12, 16]. Privacy-preserving approaches like federated learning result in limited data exposure but are not empirically tested at large scales for their effectiveness and privacy impact in IoT systems [2, 12].	The requirement for adaptive security-privacy frameworks covering protection, scalability, and performance requirements
Performance and Resource Management	Latency, bandwidth constraints, inefficient resource allocation	Some studies have indicated that latency and resource competition are problems in centralized cloud-based AI systems, and that cloud-edge systems enhance latency but add complications in orchestration [3, 17, 6, 20].	A lack of standard benchmarks and performance evaluation criteria
Integration Complexity and Vendor Lock-in	Platform dependency, limited portability, migration costs	There are indications that if Open Clouds are relied upon, it may result in limited interoperability, and this is recommended to be addressed using either containers or the 'micro-service' paradigm, but its efficacy in dealing with large-scale machine learning tasks is not adequately validated at this stage [5]	Empirical validation of interoperability solutions in production environments.
Regulatory Compliance	Data protection laws, cross-border data governance, accountability	That is, most of the research points to a limitation in the effectiveness within dynamic deployments due to compliance mechanisms being applied as an external constraint rather than an integral part of AI-cloud system design. [2, 12].	Embedded compliance by-design models in AI cloud systems
Ethical Considerations	Bias, lack of transparency, limited explainability	Various studies argue that cloud-based AI systems magnify the ethical risks since model behaviours are not transparent, and visibility of decision processes is limited. The already proposed ethical frameworks have remained	Operationalization of Explainable and Accountable AI on Cloud Infrastructure

		largely conceptual till now [9, 10, 15].	
Organizational and Cost Constraints	Skill shortages, high operational and training costs	Lack of knowledge about AI cloud technology and high costs of computation have been identified as factors that deter adoption, especially in SMEs [8, 10]. Cost-effective strategies have been proposed but never assessed comprehensively.	Integrated assessment of cost, security, performance, and governance impacts.

From the systematic review conducted in this research it is evident that the application of Artificial Intelligence (AI) technologies to the field of cloud computing holds immense potential in reshaping the digital landscape in the present and the coming ages. However from the data analysis, it is apparent that the present research is full of gaps in identifying the associated challenges of the application of Artificial Intelligence to cloud computing. This is because the various challenges are interlinked in the application of Artificial Intelligence in the cloud.

A large number of studies emphasize the beneficial impact of AI-powered security systems on cloud infrastructure for opposing advanced cyber-attacks. On the other hand studies focusing on performance issues indicate that these security improvements bring increased complexity to computations thus increasing latency values. These two illustrations show that a limitation in a large number of research papers is the lack of consideration for the combined impact on different aspects while enhancing one area against the other.

The analysis further confirms the importance of hybrid cloud and edge architectures as a way of overcoming the latency and performance challenges of the cloud computing paradigm. Even though it is factual evidence that edge computing solves the performance and latency challenges of the cloud very few research pieces discuss the implications of such solutions. System coordination, platform interoperability, and government regulations become even more complex as one move from the cloud to the edge of the network.

However, the regulatory/ethical aspects that include the integration of AI and cloud computing have been widely reported in literature although they mostly contain conceptual treatments. Regulatory research mainly focuses on data protection regulations concerning accountability whereas ethics research mainly focuses on aspects such as fairness, interpretability/transparent aspects of AI. This appears to suggest that system designs take governance/ethical aspects as secondary aspects as they appear to treat efficiency concepts separately.

Organizational and economic constraints also play a part in the adoption of cloud solutions fueled by AI. The lack of qualified professionals knowledgeable in both AI technology and cloud computing solutions has been inhibiting the adoption process in this field. Despite strategies for cost optimization being offered it has been found that current research works to a very limited extent in understanding impacts on other aspects such as system security and ethical accountability.

From the discussion above it is apparent that there is an emerging research gap in the existing literature regarding the application of frameworks to cope with the various associated challenges associated with the integration of AI in the cloud. Moving forward it is recommended that research needs to focus on holistic solutions to strike the right balance between associated aspects of efficiency, security, regulatory aspects, accountability and practicality. Such research would help in ensuring the sustainable application of AI in the cloud.

Conclusion

The above systematic review brings forth the fact that the amalgamation of Artificial Intelligence with cloud computing has immense potential for achieving intelligent, scalable and efficient computing systems. However, it is very clear from the above results that for an efficient amalgamation of AI with cloud computing it is not feasible to focus on isolated developments in this area. The main theme of this research work is that scalability, intelligence, security, ethics and readiness are interrelated.

Beginning with research aspect view this book review highlights a fragmented body of existing research attempting to address these issues one at a time either working on improving security, reducing latency and enforcing compliance. The research underlines a need for more integrated and adaptive solutions estimating these AI-cloud systems holistically and taking into consideration interdependence issues regarding performance and cost parameters apart from governance issues as well.

For application developers and other stakeholder groups it is apparent that research findings indicate that there are performance-based goals that should be met in terms of the deployment of AI and cloud technologies and there are ethical, compliance and cost-based concerns that must be addressed as well.

Future Scope

Future studies need to concentrate on creating and verifying complete frameworks to cope with technological efficiency, security, regulatory requirements, ethical accountability and economic viability in both AI and cloud environments. Empirical research in actual cloud settings is required to understand the effectiveness of the proposed solutions in the presence of system and work-load dynamics.

New paradigms like Hybrid Cloud-EDGE architectures, Federated Learning and Privacy Preserving AI hold great promise in addressing latency issues in a significant manner, improving data privacy and building trust in AI systems. More work is needed to understand their scalability and governance implications in multi-cloud and trans-border environments.

Moreover, there is a need for future research to investigate cost-effective and skill-adaptive strategies on how to deploy these services to small and medium businesses which might be facing limitations when it comes to resources and expertise. Moreover, there is a growing need for dynamic approaches to ethics and regulations which would be able to develop alongside technological advancements at a faster rate than ever before. There is also a need for improved cooperation between researches and policymakers when it comes to development and implementation.

References

1. Gill, A. Tuli, M. Xu, I. Singh, and R. Buyya, “AI-enabled cloud computing: Security and privacy challenges,” Journal of Cloud Computing, vol. 8, no. 1, pp. 1–15, 2019.
2. S. Belgaum, R. Alansari, S. Musa, and M. Alam, “Privacy and trust issues in cloud-based AI systems,” Future Generation Computer Systems, vol. 98, pp. 1–12, 2019.
3. N. Khatoon, A. Rizvi, and S. Qamar, “Performance evaluation of AI workloads on cloud platforms,” Future Generation Computer Systems, vol. 115, pp. 1–12, 2021.
4. A. Botta, W. de Donato, V. Persico, and A. Pescapé, “Integration of cloud and edge computing for real-time AI applications,” Future Generation Computer Systems, vol. 87, pp. 350–362, 2018.
5. S. Pahl, A. Brogi, J. Soldani, and P. Jamshidi, “Cloud container technologies: A state-of-the-art review,” IEEE Transactions on Cloud Computing, vol. 7, no. 3, pp. 677–692, 2019.
6. Y. Zhang, M. Chen, L. Hu, and R. Boutaba, “Challenges of infrastructure control in cloud environments,” IEEE Transactions on Cloud Computing, vol. 8, no. 3, pp. 850–861, 2020.
7. M. Armbrust et al., “Multi-cloud challenges for enterprise AI applications,” Communications of the ACM, vol. 64, no. 6, pp. 44–52, 2021.
8. H. Li, Y. Li, and X. Zhou, “Cost-efficient resource management for AI workloads in cloud computing,” IEEE Access, vol. 10, pp. 33421–33435, 2022.
9. S. Gill and R. Buyya, “Transformative effects of AI-driven security in cloud computing,” IEEE Cloud Computing, vol. 9, no. 3, pp. 34–43, 2022.
10. S. Gill, A. Tuli, M. Xu, and R. Buyya, “AI-driven cloud security: Emerging challenges and future directions,” Journal of Cloud Computing, vol. 12, no. 1, pp. 1–18, 2023.
11. M. Hakimi, G. A. Amiri, S. Jalalzai, and F. A. Darmel, “Exploring the integration of artificial intelligence and cloud computing: Opportunities and challenges,” TIERS Information Technology Journal, vol. 5, no. 1, pp. 57–69, 2024.
12. L. Albshaier, S. Almarri, and A. Albuali, “Federated learning for cloud and edge security: A systematic review of challenges and AI opportunities,” Electronics, vol. 14, no. 5, p. 1019, 2025.
13. M. Almutairi and F. T. Sheldon, “Security challenges in IoT–cloud and AI-enabled environments,” Electronics, vol. 14, no. 7, p. 1394, 2025.
14. A. Yanamala, “Emerging challenges in cloud computing security: A comprehensive review,” International Journal of Advanced Engineering Technology and Innovation, vol. 1, pp. 448–479, 2024.
15. A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” Nature Machine Intelligence, vol. 1, pp. 389–399, 2019.
16. Q. Zhang, M. Chen, L. Li, and R. Boutaba, “Dynamic security management for cloud-based AI services,” IEEE Transactions on Network and Service Management, vol. 20, no. 1, pp. 112–126, 2023.
17. W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,”

IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2456–2472, 2022.