



Artificial Intelligence and Big Data: A Comprehensive Review

¹G. P. Mulla, ²D. T. Kumbhar, ³N. R. Jagatap

¹Department of Computer Science, Yashwantrao Chavan Mahavidyala, Islampur

²Department of Computer Application, Krishan Institute of Computer Application and Management, Wathar

³Department of Computer Science, Yashwantrao Chavan Mahavidyala, Islampur

Email: ¹gulzar.p.mulla@gmail.com, ²dattatrayakumbhar4@gmail.com ³nilesh.jagatap@gmail.com

Peer Review Information	Abstract
<p><i>Submission: 15 April 2026</i></p> <p><i>Revision: 30 April 2026</i></p> <p><i>Acceptance: 09 May 2026</i></p> <p>Keywords</p> <p><i>Artificial Intelligence, Big Data, Machine Learning, Industry 4.0, Security, Education Analytics, AI 2.0</i></p>	<p>The proliferation of large-scale, heterogeneous, and high-velocity data necessitates the integration of Artificial Intelligence (AI) techniques with big data analytics for efficient knowledge discovery and intelligent decision-making. This paper presents a systematic technical review of AI-based big data analytics, synthesizing recently published peer-reviewed surveys and studies. The study analyzes how machine learning and deep learning models enhance big data processing through automated feature extraction, predictive modeling, and anomaly detection, while also identifying key limitations related to scalability, interpretability, data imbalance, and computational overhead. Application-centric insights are discussed across critical domains such as Industry 4.0, cyber security, decision support systems, and learning analytics, emphasizing real-time processing, distributed architectures, and privacy-preserving mechanisms. Furthermore, this paper highlights fundamental and algorithmic challenges, including streaming data processing, multimodal learning, adversarial robustness, and energy-efficient deployment in distributed and edge computing environments. Emerging trends towards AI 2.0 are examined, focusing on knowledge-based learning, neuro-symbolic integration, and explainable AI frameworks. In this review, the open research challenges and future directions for developing scalable, transparent, and secure AI-enabled big data systems are identified, while providing a unified technical reference for researchers and practitioners.</p>

Introduction

The importance of analytical systems that can learn from large-scale, high-dimensional data has increased due to the rapid spread of vast, heterogeneous datasets, or "Big Data." To unlock the value contained in Big Data, artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), is being used more and more. On the other hand, the availability of large, varied datasets is essential to the development of AI capabilities.

There is increasing research being conducted on how artificial intelligence and big data interact with each other. Leading surveys on ML-driven

analytics (Nti et al., 2022), AI applications in Industry 4.0 (Jagatheesaperumal & Rahouti, 2021), algorithmic challenges in Big Data (Kersting & Meyer, 2018), AI-enhanced security mechanisms (Dai & Boroomand, 2022), and education-focused Big Data + AI systems (Luan et al., 2020) are summarized in this review.

AI-Driven Big Data Analytics -Large-Scale Data Machine Learning

Through automated feature extraction, pattern recognition, and predictive modelling, AI improves Big Data analytics. A review of machine learning in big data analytics claims that while

ML helps with forecasting, classification, and clustering scenarios, it has problems with data imbalance, computational cost, and interpretability (Nti et al., 2022).

A taxonomy of AI strategies - supervised, unsupervised, hybrid, and reinforcement approaches - is provided by a thorough analysis by Rahmani et al. (Rahmani et al., 2021), illustrating their appropriateness for different Big Data contexts.

Industry 4.0: AI and Big Data

Smart factories, distributed industrial networks, and cyber-physical systems are all powered by AI and Big Data. Researchers describe the use of AI in robotics, intelligent monitoring, predictive maintenance, and optimization in The Duo of Artificial Intelligence and Big Data for Industry 4.0 (Jagatheesaperumal & Rahouti, 2021).

Despite rapid expansion, obstacles still exist: Data sources that are fragmented and interoperability issues. Limitations in latency and edge computing; limitations in data quality and annotation

These limitations highlight the need of efficient, distributed AI models and consistent data pipelines.

Protecting Systems for Big Data Applying AI

The huge data environment has created vulnerabilities such as insider threats, unusual patterns, distributed denial of service attacks, and data poisoning. Neural networks, anomaly detectors, and intelligent filters reduce cyber security risks through fast pattern identification and threat classification, according to a review of artificial intelligence to improve the security of big data systems (Dai & Boroomand, 2022).

Among the crucial difficulties are:

- AI model manipulation by adversaries.
- Maintaining resilience in streaming scenarios.
- Safeguarding privacy in distributed systems

To keep security effective, AI must change along with these attack surfaces.

Using AI to Make Decisions in the Big Data Era

Artificial intelligence is being increasingly used by decision support systems to transform unprocessed data into useful information. According to Duan et al. (2019), artificial intelligence can improve strategic planning, risk prediction, and automated reasoning in organizational contexts when used for decision making in the Big Data era.

But the researchers stress the necessity of:

- Transparent and comprehensible models

- Integration of qualitative knowledge with data-driven systems;
- ethical governance

Infrastructure and Algorithmic Difficulties in Big AI

Algorithms that can handle large-scale, high-dimensional, high-velocity and frequently noisy data are essential for Big Data. Kersting and Meyer describe the algorithmic bottlenecks limiting contemporary AI systems in from large data to huge artificial intelligence? (Kersting & Meyer, 2018).

Important algorithmic difficulties:

- Managing multimodal, partial, and imbalanced datasets.
- Memory and streaming computing.
- Robustness in dispersed contexts more scalable, adaptable, and energy-efficient AI architectures are required as a result of these limitations.

Domain-Specific Integration: Big Data + AI in Education

Learning management systems, assessment tools, and student interactions all generate multimodal, high-frequency data for education. Advances in intelligent tutoring, early warning systems, and individualized learning are highlighted in the challenges and future directions of artificial intelligence and big data in education (Luan et al., 2020).

Among the difficulties noted are:

- Predictive analytics' ethical application.
- Algorithmic bias in educational settings.
- Requirement for scalable, privacy-preserving infrastructures.

Bibliometric insights and cross-domain trends

The fast growth of AI and Big Data in healthcare, business analytics, agriculture, cyber security, and education is mapped in a bibliometric review, State-of-the-art of artificial intelligence and big data analytics reviews in five distinct fields (Thayyib et al., 2023).

Emergent research clusters are identified by the researcher in:

- Automation made possible by deep learning;
- Cyber-physical systems;
- Sustainable smart habitats; and
- Human-AI collaboration models

Big Data as a Catalyst for AI 2.0

Big Data is hastening the shift toward AI 2.0— systems that are context-aware, knowledge-driven, and capable of high-level reasoning

beyond pattern recognition—according to Challenges and Opportunities: from Big Data to Knowledge in AI 2.0 (Zhuang et al., 2017).

The researchers point up a number of potential future paths:

- Semantic-level learning;
- Cross-domain knowledge transfer;
- Cognitive and neuro-symbolic architectures; and
- Unified knowledge bases for AI with multiple tasks

Conclusion

Big data and AI are closely related foundational pillars of contemporary intelligent systems. While big data itself enhances the scope, accuracy, and autonomy of artificial intelligence, artificial intelligence provides the analytical skills necessary to interpret large datasets. AI-driven big data applications are transforming processes, infrastructures, and societal capabilities across a wide range of areas, from industry and cyber security to education and decision sciences. However, there are still significant obstacles to overcome, such as algorithmic scalability, ethical issues, security flaws, and the requirement for transparent, interoperable, and energy-efficient AI ecosystems.

Interdisciplinary techniques, privacy-preserving designs, explainable AI, and the creation of AI 2.0 systems that can truly integrate knowledge must be given top priority in future research.

References

I. K. Nti, J. A. Quarcoo, and J. Aning, "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects," *IEEE Access*, 2022.

S. K. Jagatheesaperumal and M. Rahouti, "The duo of artificial intelligence and big data for Industry 4.0: Applications, techniques, challenges, and future research directions," *IEEE Access*, 2021.

D. Dai and S. Boroomand, "A review of artificial intelligence to enhance the security of big data systems: State-of-art, methodologies, applications, and challenges," *Evolutionary Intelligence*, 2022.

Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data—Evolution, challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, pp. 63–71, 2019.

K. Kersting and U. Meyer, "From big data to big artificial intelligence? Algorithmic challenges and opportunities of big data," *KI - Künstliche Intelligenz*, vol. 32, pp. 143–148, 2018.

H. Luan et al., "Challenges and future directions of big data and artificial intelligence in education," *Frontiers in Psychology*, vol. 11, Art. no. 580820, 2020.

A. M. Rahmani et al., "Artificial intelligence approaches and mechanisms for big data analytics: A systematic study," *PeerJ Computer Science*, vol. 7, p. e488, 2021.

P. V. Thayyib, R. Mamilla, M. Khan, H. Fatima, and M. Asim, "State-of-the-art of artificial intelligence and big data analytics reviews in five different domains: A bibliometric summary," *Sustainability*, vol. 15, no. 5, Art. no. 4026, 2023.

Y. Zhuang, F. Wu, C. Chen, and Y. Pan, "Challenges and opportunities: From big data to knowledge in AI 2.0," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, pp. 3–15, 2017.