Artificial Intelligence techniques for complex big data environments: methods and perspectives

Zice Gao

University of Rochester, Rochester, USA

395001740@qq.com

Abstract. In today's data-driven world, big data environments are becoming increasingly complex, characterized by high volume, variety, and velocity. Traditional data processing methods are no longer sufficient to handle such challenges. Artificial Intelligence (AI) provides powerful solutions for extracting value from diverse and dynamic data sources. This paper reviews key AI techniques—including machine learning, deep learning, natural language processing, graph-based models, and federated learning—and discusses their applications in complex scenarios such as healthcare, finance, smart cities, and Industry 4.0. It also highlights major challenges, including data quality, model interpretability, computational cost, and privacy concerns. Finally, the paper explores future directions in AI development, such as multimodal learning and real-time decision-making. These advancements will play a vital role in enabling intelligent, efficient, and ethical data analytics in the years to come.

Keywords: Artificial Intelligence, big data, machine learning, deep learning, natural language processing

1. Introduction

With the rapid expansion of digital technologies, data is being generated at an unprecedented rate. Today's big data environments are not only massive in volume, but also complex in structure. They often involve multiple data sources, including text, images, audio, and real-time sensor data, making them highly heterogeneous and dynamic. These complex environments pose serious challenges for traditional data processing methods, which struggle with scalability, speed, and adaptability.

Traditional techniques rely heavily on manual feature design and are typically limited to structured data and offline processing. As a result, they fall short in handling unstructured content, real-time analysis, and high-dimensional data, especially in fast-changing and uncertain conditions [1].

Artificial Intelligence (AI) provides new opportunities to address these issues. Techniques like machine learning, deep learning, and natural language processing can automatically learn from data, adapt to changing patterns, and process information at scale. AI has become an essential tool for extracting value from complex big data, enabling smarter decision-making in fields such as healthcare, finance, and smart cities.

2. Key AI techniques for big data

Artificial Intelligence (AI) provides powerful tools to manage and analyze complex big data. Four key techniques stand out for their effectiveness: machine learning, natural language processing, graph-based models, and privacy-aware approaches like federated learning [2].

2.1. Machine learning and deep learning

Machine Learning (ML) helps find patterns in large datasets and supports tasks like classification, clustering, and prediction. Traditional ML models—such as decision trees or support vector machines—are useful for structured data.

Deep Learning (DL), especially neural networks, is more suitable for complex and unstructured data such as images, sound, and time series. Models like CNNs (for images) and Transformers (for sequences and text) can automatically learn deep features without manual effort. These techniques are widely used in applications such as recommendation systems, anomaly detection, and predictive maintenance.

2.2. Natural Language Processing (NLP) and large language models

Much of today's data is in text form. NLP enables machines to understand and process human language. It is widely used for sentiment analysis, topic detection, and document classification [3].

Recent Large Language Models (LLMs) like GPT and BERT have brought major improvements. These models can handle tasks like summarizing documents, answering questions, and generating text. In big data systems, LLMs help extract insights from reports, social media, emails, and customer feedback.

2.3. Graph-based methods

Many real-world problems involve relationships between entities—such as people, products, or locations. Graph-based methods can model these connections. Graph Neural Networks (GNNs) learn patterns from linked data, making them useful for fraud detection, social networks, and supply chain analysis. For example, GNNs are used in fraud detection by identifying unusual patterns of transactions between accounts, which could indicate fraudulent activity. In social networks, GNNs help to predict user behavior by analyzing interactions and connections, enabling personalized recommendations. In supply chain analysis, GNNs can predict delays or inefficiencies by analyzing the relationships between suppliers, manufacturers, and distributors.

Knowledge Graphs help organize information across sources and improve search and reasoning. For example, Google's Knowledge Graph enhances search results by understanding the relationships between entities like people, places, and events, enabling more relevant answers. Another case is IBM Watson, which uses knowledge graphs to enhance its AI's ability to reason over vast amounts of structured and unstructured data, improving decision-making in industries like healthcare and finance.

2.4. Federated learning and privacy-preserving AI

Data privacy is a major concern in big data environments. Federated Learning (FL) allows AI models to be trained on local devices without sending raw data to a central server. This protects user privacy while enabling large-scale learning [4].

Other privacy-aware methods—like differential privacy and encryption—further reduce risks. These approaches are useful in fields like healthcare, finance, and mobile computing.

3. Applications in complex scenarios

AI techniques have been widely adopted to address real-world challenges in complex big data environments. Their ability to analyze diverse, high-volume, and fast-changing data makes them highly suitable across sectors. This section outlines four key application areas: healthcare, finance, smart cities, and Industry 4.0.

3.1. Healthcare

Healthcare is one of the most data-intensive fields, with large volumes of patient records, medical images, and real-time monitoring data. AI helps make sense of this complexity [5].

In medical imaging, deep learning models such as Convolutional Neural Networks (CNNs) are used to detect abnormalities in X-rays, MRIs, and CT scans with high accuracy. These systems assist doctors in early diagnosis of diseases like cancer, pneumonia, and stroke.

AI also plays a key role in analyzing Electronic Health Records (EHRs). Natural Language Processing (NLP) techniques help extract critical information from unstructured clinical notes, improving patient profiling and treatment recommendations.

3.2. Finance

In the financial sector, data arrives at high speed and from many sources, including transactions, customer profiles, social media, and market news. AI enhances risk management, fraud prevention, and customer service [6].

Machine learning models are commonly used for credit scoring, predicting a customer's likelihood to repay loans. These models can analyze thousands of variables and adapt as more data becomes available.

Fraud detection systems rely on anomaly detection and real-time analysis. For example, graph-based AI methods can identify suspicious transaction patterns by analyzing relationships among accounts [7].

AI-powered chatbots and NLP tools streamline customer interactions, improving satisfaction while reducing operational costs. For example, Zendesk's AI-powered chatbots are used to handle customer service queries, reducing wait times and freeing up human agents to tackle more complex issues. Similarly, Bank of America's Erica helps customers manage their finances by offering personalized advice and performing banking transactions, significantly enhancing customer experience.

In investment, AI helps in market trend prediction and portfolio optimization. For instance, Robo-advisors like Betterment and Wealthfront use AI algorithms to analyze market trends and optimize investment portfolios based on individual risk

preferences. Additionally, AI-powered platforms like Sentiment Investor analyze social media and news sentiment to predict market movements, helping investors make more informed decisions.

3.3. Smart cities

As urban areas grow, cities are becoming increasingly complex systems that rely on data for decision-making. AI supports smart cities by helping manage transportation, public safety, and environmental monitoring.

In traffic management, AI models analyze real-time data from cameras and sensors to optimize traffic light control, reduce congestion, and predict traffic flow. Combined with GPS and weather data, these systems improve route planning and emergency response [8].

4. Challenges and future directions

While AI has demonstrated significant potential in complex big data environments, several challenges remain that limit its effectiveness and broader adoption. Addressing these issues is essential for building robust, trustworthy, and scalable AI systems. This section outlines key challenges and future research directions.

4.1. Data quality and labeling costs

AI models rely heavily on large volumes of high-quality data. However, in real-world settings, data is often noisy, incomplete, or inconsistent. Poor data quality can lead to unreliable outputs, especially in safety-critical fields like healthcare or finance.

Additionally, supervised learning methods require large amounts of labeled data, which can be expensive and time-consuming to obtain. In domains where expert knowledge is needed—such as medical diagnostics or legal analysis—the labeling cost is particularly high. Future research should explore data-efficient approaches like self-supervised learning, active learning, and synthetic data generation to reduce dependence on labeled data.

4.2. Model interpretability, transparency, and robustness

Many state-of-the-art AI models, especially deep learning architectures, are often seen as "black boxes." This lack of interpretability hinders trust and limits adoption in regulated industries.

There is growing demand for Explainable AI (XAI) methods that provide transparent, human-understandable reasoning for decisions. At the same time, models must be robust to adversarial attacks, input noise, and data distribution shifts. Future work should focus on building interpretable and resilient models without compromising performance.

4.3. Computational bottlenecks and green AI

Training large AI models requires significant computational resources, often involving high energy consumption and specialized hardware. This limits access to advanced AI tools for small organizations and raises environmental concerns.

"Green AI" is an emerging direction that emphasizes energy-efficient algorithms and sustainable model development. Techniques like model pruning, quantization, and efficient architectures (e.g., TinyML) aim to reduce the carbon footprint of AI systems while maintaining accuracy.

4.4. Ethical, legal, and privacy concerns

AI systems that process personal, financial, or sensitive data raise serious ethical and legal concerns. Issues include algorithmic bias, discrimination, data misuse, and lack of accountability.

Regulations such as the GDPR and AI Act highlight the importance of data protection and ethical standards. Future systems must be designed with built-in privacy safeguards (e.g., differential privacy, federated learning) and mechanisms for human oversight and legal compliance.

5. Conclusion

Artificial Intelligence has become a key enabler in extracting value from complex big data environments. Through techniques such as machine learning, deep learning and federated learning, AI allows for more accurate, scalable, and intelligent data analysis across various industries.

These tools help address challenges posed by the volume, variety, and velocity of modern data, supporting smarter decision-making in fields like healthcare, finance, smart cities, and manufacturing. However, significant obstacles remain,

including data quality, interpretability, computational demands, and privacy concerns. Continued research is needed to develop more robust, transparent, and energy-efficient AI systems. Future directions include multimodal learning, real-time AI, and more ethical and privacy-aware frameworks. As data continues to grow in complexity, the integration of AI will be critical in transforming how we understand and act upon information. A responsible and innovative approach to AI development will shape the future of big data analytics and its societal impact.

References

[1]   Ghodmare, S. D., Khode, B. V., & Ladekar, S. M. (2021). The role of artificial intelligence in Industry 4.0 and smart city development. In *Advances in Civil Engineering Infrastructures Development: Selected Proceedings of ICRACEID 2019* (pp. 1–12). Springer Singapore. https://doi.org/10.1007/978-981-15-6463-5_58

[2]   Ahmed, I., Jeon, G., & Piccialli, F. (2022). From artificial intelligence to explainable artificial intelligence in Industry 4.0: A survey on what, how, and where. *IEEE Transactions on Industrial Informatics, 18*(8), 5031–5042. https://doi.org/10.1109/TII.2022.3146552

[3]   Chataut, R., Phoummalayvane, A., & Akl, R. (2023). Unleashing the power of IoT: A comprehensive review of IoT applications and future prospects in healthcare, agriculture, smart homes, smart cities, and Industry 4.0. *Sensors, 23*(16), 7194. https://doi.org/10.3390/s23167194

[4]   Luckey, D., Cohn, G., & Hou, L. (2021). Artificial intelligence techniques for smart city applications. In *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering (ICCCBE 2020)* (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-51295-8_1

[5]   Ahmed, I., Jeon, G., & Piccialli, F. (2022). A blockchain- and artificial intelligence-enabled smart IoT framework for sustainable city. *International Journal of Intelligent Systems, 37*(9), 6493–6507. https://doi.org/10.1002/int.22852

[6]   Javed, A. R., Farooq, M. I., & Yaseen, Z. (2023). A survey of explainable artificial intelligence for smart cities. *Electronics, 12*(4), 1020. https://doi.org/10.3390/electronics12041020

[7]   Wolniak, R., & Stecuła, K. (2024). Artificial intelligence in smart cities—Applications, barriers, and future directions: A review. *Smart Cities, 7*(3), 1346–1389. https://doi.org/10.3390/smartcities7030057

[8]   Sood, S. K., Rawat, K. S., & Kumar, D. (2022). A visual review of artificial intelligence and Industry 4.0 in healthcare. *Computers and Electrical Engineering, 101*, 107948. https://doi.org/10.1016/j.compeleceng.2022.107948