Intl. J. Engg. Sci.Adv. Research 2025 Sept; 11(3): 1-6 **ISSN NO: 2395-0730**

Transforming Big Data Analysis Through Modern AI Techniques

Gyanendra K. Gautam Devendra K. Mishra

[1]Research Scholar, Computer Science & Engineering, Amity School of Engineering and Technology, Amity University Madhya Pradesh Gwalior, MP,India

E-mail: gyanendragautam26@gmail.com

[2]Associate Professor, Computer Science & Engineering, Amity School of Engineering and Technology, Amity University Madhya Pradesh Gwalior, MP,India

dkmishra@gwa.amity.edu

ABSTRACT

Modern techniques of AI have revolutionized the entire management, analysis, and utilization processes of big data in any company. Indeed, the problem is that Big Data puts the traditional model in danger owing to its volume, velocity, and variety. yet Artificial Intelligence, and notably machine learning and deep learning, provides a strong set of solutions for the same issues. This paper will discuss the interaction between AI and big data, where AI applications like neural networks, reinforcement learning, and NLP are recognized as advanced techniques to produce useful concepts, such as relations, trends, forecasts or actionable intelligence, from large volumes of information. Such information is processed fully with the help of AI integrated systems in such a way as to remove the need of integration, cleansing, and transforming data focusing on saving effort and time. These technologies lead to better precision and because of this, inform choices made by using predictive analytics, clustering, and classification that make them essential in industries such as marketing, banking, and healthcare. In any case, the class of Big Data is not just structured datasets since AI models may be trained on various unstructured data including text, images, video. This paper also examines the issues connected with AI in Big Data analysis, like computational resource demands, algorithm transparency, and bias in data models. It focuses how important scalable, distributed AI architectures are for managing the enormous computational demands of big data, such as edge and cloud computing.

***Keywords*- Big Data, Natural Language Processing, Reinforcement Learning, Predictive Analytics, Artificial Intelligence, Machine Learning, Deep Learning**

I. INTRODUCTION

The vast amount of data fabricated in a second is termed big data in our digital stage. It can be analysed to provide useful information to governments, corporations, and others. However, with this volume and speed, the conventional data analysis techniques are not always efficient. To deal with these problems more effectively, contemporary AI techniques have been developed. It is highly important to understand this change, as it shows how AI can develop many industries. Artificial intelligence methods, especially ML and DL, have emerged to be mighty solutions to those problems as they provide seamless, automatic data processing, thereby speeding up the workflow and adding accuracy.

I.1 Conceptual basis of Big Data Analytics

At the core of Big Data Analytics is the ability to deal with the three Vs – volume, variety and velocity[1]. The sheer volume of data generated daily requires scalable solutions for storage and processing. The velocity at which data is produced demands real-time analytics capabilities. Moreover, the variety of data, including structured and unstructured formats, necessitates flexible processing methods. Big Data platforms, exemplified by Apache Hadoop and Apache Spark, provide the infrastructure to navigate and analyse these vast datasets.

Fig. 1 Basic Principal of Big Data Analysis

Where traditional analytics fails in the presence of Big Data, advanced analytics takes a front seat of glory[2]. The presence of Machine Learning and Artificial Intelligence in the Big Data framework enhances the depth and precision of the insights. Cluster and regression-based algorithms can learn large-scale data well. Artificial intelligence-based algorithms with their cognitive abilities improve understanding, reasoning, and decision-making capabilities.

I.2 Effect on Big Data Analytics Decision-Making Techniques in Trendy Intelligence

Big Data Analytics transforms decision-making processes by allowing data-driven insight in strategic planning and risk management across different industries. It lets businesses foresee market trends and mitigate the risks involved by aligning decisions with the organizational objectives. In finance, fraud detection and risk assessment get improved, whereas in health care, predictive models for personalized medicine get advanced with better patient outcomes [3]. Integration of modern AI techniques further transforms big data analysis through automation of tasks, improved accuracy, and the unveiling of patterns in unstructured data. Even though the issues that affect the development of these emerging solutions include data quality, algorithm transparency, ethical issues, and computational requirements, edge computing, federated learning, and Explainable AI (XAI) are among the emerging solutions that help solve the problem. XAI especially fosters accountability and transparency in critical applications. By handling these challenges, the synergy between AI and big data continues to unlock transformative potential, driving innovation and efficiency across various sectors.

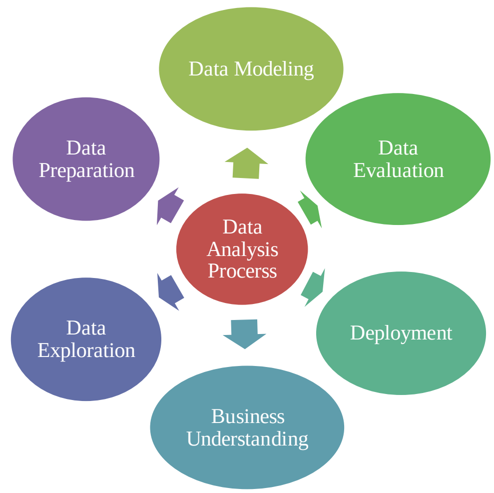

Fig. 2 Data Analysis Process Diagram

II. Modern AI Techniques for Big Data Analysis

AI has grown over the years, it has expanded to include a number of massive data handling strategies. Some technique to handle data are as follows-

II.1 Machine Learning

A subfield of artificial intelligence called machine learning is primarily concerned with developing various techniques and algorithms that enable a computer to learn on its own by utilizing prior knowledge and experience. It allows machine to learn automatically from the data, better performance from experiences, and predict the things. It is capable to creates a mathematical model to make predictions or decisions without being distinctly programmed, rather than the help of specimen historical data, or training data. Machine learning combines statistics and computer science in developing predictive models. Algorithms that learn from historical data are either fabricate or used in machine learning. Various algorithms are used in machine learning to address data issues. Data scientists prefer to emphasize that there is not a single, universally applicable method that works well for every problem[4]. The more details we give, the better the results will be.

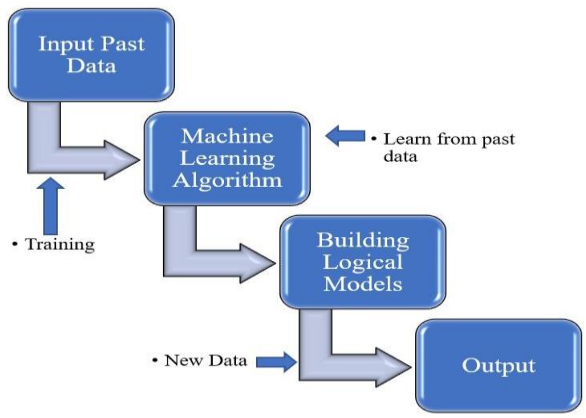

Fig. 3 Process or Working of Machine Learning Technique

II.2 Neural Network

Neural networks is a highly evocative term. It alludes to devices that resemble brains. A collection of interconnected units known as neurons that communicate with one another is called a neural network. Neurons can be mathematical models or real cells. Many neurons working together in a network can accomplish complicated tasks, even though individual neurons are basic. An interconnected grouping of basic processing components, also known as nodes or units, that functions much like an animal neuron is called a neural network. The network's ability to process is stored in the interunit connection weights or strengths that are learned from a series of training patterns.

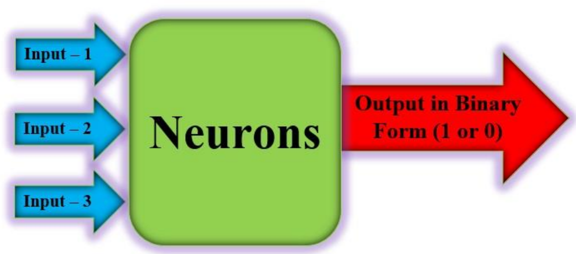

Fig. 4 Block Diagram of Neural Network

II.3 Natural Language Processing

Natural language processing is the study of how computers could comprehend and alter natural language speech or text to carry out advantageous activities[5]. It is designs computers to recognise like human language. NLP applies the field of computational and models that depend on statistics, deep learning and machine learning. These innovations made it possible for computers to process and analyse text or audio data and fully understand their meaning, including the motive and feeling of the announcer or editor. NLP is used in many applications, such as chatbots, speech recognition, text summarization, and text translation. we can use several other things, it includes GPS enabled voice-activated system, digital voice assistants, speech-to-text converter software, and customer service chatbots. NLP also assists companies in increasing their performance, productivity, and efficiency by streamlining common language-related tanks.

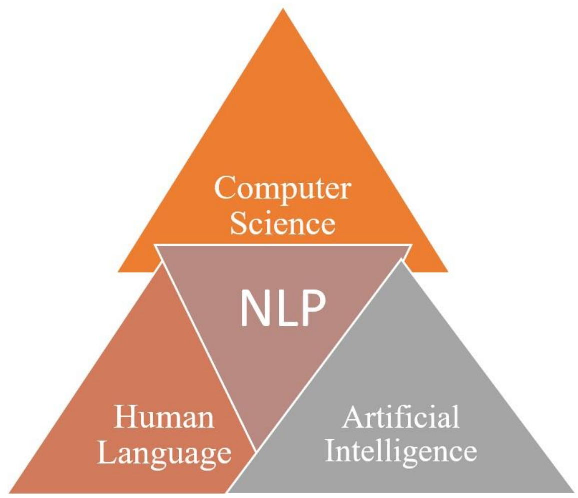

Fig. 5 Natural Language Processing

II.4 Data Visualization and Interpretation

These components are crucial for investigation and thinking with big, high-dimensional data. Latest techniques of AI have led to the development of two approaches Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (t-SNE) widely being applied to decrease dimensions with which meaningful visualizations may be possible. This not only provides intuitive insight but makes evident patterns, clusters, and anomalies that were buried deep within raw numerical presentations. Data visualization shows information about recently identified connections, like an unidentified client segment or a series of occurrences that could be predicted attrition among customers. therefore, big data discovery or exploratory analytics are more frequently linked to data visualization[6].

III. Problem and Challenges when Dealing with AI and Big Data Analysis

Although AI and big data analysis has proved to be one of the most effective technologies, numerous drawbacks are also associated with it. Some major problems faced are outlined below in the table.

Table. 1

S. No.	Category	Issue	Description
1.	Data quality and availability	Incomplete or Noisy data	AI based models require huge amounts of high potential data, but real-world data often contains missing. Inconsistent and irrelevant data or noisy entries
		Data repository	Organization often store data in isolated systems, making it difficult to integrate and analyse comprehensively.
2.	Resource Requirement	Energy Consumption	Training large AI models, especially big data, requires great energy, This has been an issue for environment.
		Specialized Talent	The developing and deployment of AI solutions for big data require expertise that might be costly and difficult to find.
3.	Integration with Existing Systems	Legacy Systems	The implementation of AI-driven big data analysis in traditional IT systems is technically difficult.
		Interoperability	Ensuring AI solutions work seamlessly across different platforms is complex.
4.	Security and Privacy related issue	Data Breaches	AI systems are vulnerable to cyber-attacks, this type of data poisoning or model theft.
		Sensitive Data	Big data typically involves personal or sensitive information, hence privacy-related worries
5.	Ethical Issue	Accountability	AI makes crucial decisions, establishing liability in cases of a mistake even bias turns problematic.
		Fairness	Ensuring that AI does not disproportionately disadvantage certain group requires deliberate fairness interventions.

IV. Literature Review

This is due to the combination of AI and big data. It has changed the way business handle and evaluate enormous volumes of data. Traditional data management methods have been put to the test by the exponential development of data in terms of volume, velocity, and variety. However, artificial intelligence, especially in the form of methods like deep learning, provides practical answers to these problems. This literary review explores the relationship of artificial intelligence and big data emphasizing its multiple applications advantages related challenges[7]. The application of artificial intelligence (AI) and big data to smart manufacturing has been made possible by the growing demand for smart manufacturing that is safe, affordable, and sustainable as well as by new technical enablers. This suggests a significant integration of artificial intelligence (AI), robotics, big data, blockchain, 5G connectivity, and the Industrial Internet of Things (IIoT) to support smart production and the dynamic processes in contemporary industries. With an emphasis on important applications, methodologies, the ideas involved, important enabling technologies, difficulties, and research perspectives toward the implementation of Industry 5.0, we present a thorough review of the various facets of AI and Big Data in Industry 4.0 in this article. We specifically highlight and examine the ways in which AI and Big Data work together to support various Industry 4.0 applications. According to the literature, AI will power future industries through the use of robotics, high-speed communication systems, blockchain, smart machines, Big Data, IIoT, and the overall economic transformation[8]. Although the application of cognitive technology to address business issues is growing, many of the most ambitious AI initiatives experience failures or setbacks[9]. Most information in companies is held in unstructured models. Data extraction and retrieval are a couple of the important operations and major in semantic web applications. The success of storage and processing of the unstructured data will govern most of these metrics. To the analyst, the amount and richness of unstructured data bring into existence an enormous amount of new potential. We make both single and multiple group analyses of structured and unstructured data. Two methods of extracting knowledge from textual context in documents are text mining and natural language processing. This paper will show examples of both text mining and natural language approaches. NLP utilizes specific algorithms to attempt to understand textual concepts[10]. The textual content found in emails, blogs, tweets, forums, and similar platforms is what we call text analytics. Often termed text mining, Text analytics is one of the earliest subfields of Artificial intelligence developed during the 1950s when a desire to comprehend text first emerged. In this present era, Text analytics is often regarded as the next phase of big data analysis. Information extraction, Named Entity Recognition, Semantic Web annotated domain representation, and many other subcategories form the domain of Text Analytics. There are many approaches being used, and some, like machine learning, have brought significant attention as they exhibited a semi supervised improvement of systems. At the same time, these methods have quite a few limitations, which sometimes make them not the best or exclusive option[11].

V. Opportunities for Contemporary AI in Big Data Analytics in the Future

Future prospects of AI in big data analytics are bright and ongoing efforts are being made for making it more efficient, interpretable, and ethics-compliant. Another living direction is scalable distributed AI architectures essential for handling huge data and computational demands inherent to big data. Edge computing and cloud-based platforms become pivotal, enabling AI algorithms processing data closer to its source to reduce latency while distributing the computational load[12]. These structures will enable real-time analysis, important for applications like autonomous vehicles, smart cities, and industrial IoT by improving the efficiency of data handling.

Explainable AI is another future direction with significant importance to those industries requiring transparent decision-making processes. Techniques such as attention mechanisms, surrogate models, and feature importance analysis are being researched to enhance the interpretability of traditionally opaque models, thus allowing stakeholders to understand the rationale behind predictions[13]. This development will most likely support compliance with regulations and the responsible use of AI, particularly in fields like finance and healthcare, where decisions must be accountable and trustworthy.

To reduce biases, the future will also include more powerful approaches to fair and unbiased AI, which would include methods for detecting and correcting biases in training data and algorithms. This would help in addressing social and ethical challenges because fair AI models are important for equitable decision-making, especially when algorithms impact diverse demographic groups[14]. Moreover, the rise of Federated Learning (FL) is a move toward de-centralized data processing such that models can be updated across multiple devices without a need to share raw data, thus improving privacy. FL is promising in the health sector because data privacy is a sensitive issue such that collaborative model training with institutions can be done by maintaining patient confidentiality[15]. In addition to this, quantum computing could have an impact on big data analysis and could potentially reduce by orders of magnitude the computing time of complex computations thereby making feasible work with earlier intractable datasets. Future study will probably focus on striking a compromise between efficiency and accuracy. Lightweight yet powerful models should emerge to be able to function in resource-constrained environments. Algorithmic efficiency, combined with ethical AI practice, will continue to take AI integration in big data analysis to new heights and establish it as a core technology in the data-driven world.

VI. Conclusion

Modern AI techniques transformed the way big data analytics take place. Organizations may finally derive actionable insights out of vast datasets. It further automates data analysis via techniques like machine learning and deep learning, raises accuracy, and reveals invisible trends in unstructured data for fields such as finance, marketing, and even in healthcare. However, challenges remain Data quality issues, algorithm transparency, ethical concerns, and high computational demands persist. Most of the AI systems operate as black boxes, and thus, there is a need for better interpretability in order to be trusted. The need for fairness and inclusivity is best achieved through addressing bias in AI models when the decision-making process affects different populations. Future advancements will be on scalable structures for AI, such as edge computing and federated learning, to manage the computational requirements of big data while protecting privacy. Explainable AI (XAI) techniques will also become more important in developing accountability and transparency, especially in sensitive industries. To reduce social trouble and encourage fair outcomes, moral principles in AI evolution will remain crucial. Although AI has, in a sense, remould big data analysis, its fully capable and the growth of various industries depends on resolving present issues and encouraging morally sound, open technologies.

References

[1] A. Rehman, S. Naz, and I. Razzak, “Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities,” *Multimed Syst*, vol. 28, no. 4, pp. 1339–1371, 2022, doi: 10.1007/s00530-020-00736-8.

[2] M. G. Dekimpe, “Retailing and retailing research in the age of big data analytics,” *International Journal of Research in Marketing*, vol. 37, no. 1, pp. 3–14, 2020, doi: 10.1016/j.ijresmar.2019.09.001.

[3] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” *Decis Support Syst*, vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/j.dss.2010.08.006.

[4] B. Mahesh, “Machine Learning Algorithms - A Review,” *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/art20203995.

[5] G. Chowdhury, “Natural language processing . Annual Review of This is an author-produced version of a paper published in The Annual Review of Information Science and Technology ISSN 0066-4200 . This version has been peer-reviewed , but does not,” *The annual review of Information science and technology*, vol. 37, pp. 51–89, 2003.

[6] J. Berengueres, M. Sandell, and A. Fenwick, “Data Visualization and Storytelling,” no. May, pp. 4–10, 2546.

[7] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014, doi: 10.1007/s11036-013-0489-0.

[8] S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, “The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions,” *IEEE Internet Things J*, vol. 9, no. 15, pp. 12861–12885, 2022, doi: 10.1109/JIOT.2021.3139827.

[9] T. H. Davenport and R. Ronanki, “Intelligence for the real world,” *Harv Bus Rev*, vol. 96, no. 1, pp. 108–116, 2018.

[10] F. S. Gharehchopogh and Z. A. Khalifelu, “Analysis and evaluation of unstructured data: Text mining versus natural language processing,” *2011 5th International Conference on Application of Information and Communication Technologies, AICT 2011*, no. April, 2011, doi: 10.1109/ICAICT.2011.6111017.

[11] A. Moreno and T. Redondo, “Text Analytics: the convergence of Big Data and Artificial Intelligence,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 6, p. 57, 2016, doi: 10.9781/ijimai.2016.369.

[12] Q. Wang, G. Jin, Q. Li, K. Wang, Z. Yang, and H. Wang, “Industrial Edge Computing: Vision and Challenges,” *Information and Control*, vol. 50, no. 3, pp. 257–274, 2021, doi: 10.13976/j.cnki.xk.2021.1030.

[13] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models,” *ACM Comput Surv*, vol. 51, no. 5, 2018, doi: 10.1145/3236009.

[14] S. Passi and S. Barocas, “Problem formulation and fairness,” *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, pp. 39–48, 2019, doi: 10.1145/3287560.3287567.

[15] P. Kairouz *et al.*, *Advances and open problems in federated learning*, vol. 14, no. 1–2. 2021. doi: 10.1561/2200000083.