Babylonian Journal of Artificial Intelligence
Vol. **(2023)**, 2023, **pp**. 5–9
DOI: https://doi.org/10.58496/BJAI/2023/002 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Exploring GPT-4's Characteristics Through the 5Vs of Big Data: A Brief Perspective

Osamah Mohammed Alyasiri [1,2,*], Dua'a Akhtom [2], Ahmed Hussein Ali [3]

[1] *Karbala Technical Institute, Al-Furat Al-Awsat Technical University, Karbala 56001, Iraq*
[2] *School of Computer Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia*
[3] *Department of Computer, College of Education, AL-Iraqia University, Baghdad, Iraq*

ARTICLE INFO

Article History
Received 21 Aug 2023
Accepted 02 Oct 2023
Published 27 Oct 2023

Keywords
Artificial Intelligence
Big Data
Big data analytics
GPT-4's
5V's

ABSTRACT

In the dynamic field of modern artificial intelligence, GPT-4 emerges as a key participant, addressing challenges similar to Big Data's 5Vs—Volume, Velocity, Variety, Veracity, and Value. This study explores the convergence of GPT-4's operational framework with the core aspects of Big Data, highlighting the model's flexibility and efficacy in handling intricate datasets. GPT-4 excels in managing extensive textual data, aligning with Big Data's voluminous nature, and demonstrates real-time processing capabilities to match the rapid evolution of Big Data. While initially text-oriented, GPT-4 expands into image recognition, enhancing versatility and aligning with Big Data's Variety aspect. The model's evolving proficiency in non-textual domains broadens its utility. Addressing Veracity, GPT-4 critically evaluates diverse training data, mirroring Big Data's challenges in ensuring accuracy. Its outputs, offering context and insights, contribute to actionable knowledge and align with Big Data's objectives. Despite differences, GPT-4 serves as a microcosm, providing scalable and accessible data processing capabilities, establishing itself as a crucial tool in the AI domain. This paper emphasizes the parallels and underscores GPT-4's adaptability in handling complex datasets.

1. INTRODUCTION

In the era of digital transformation, data has become the lifeblood of industries, shaping the way businesses operate and make decisions. The advent of Big Data technology has revolutionized the landscape, aiming to predict future trends by uncovering patterns within extensive datasets. However, the effective utilization of Big Data requires advanced monitoring and analysis tools, as well as overcoming challenges related to storage, management, and analysis. Traditional data processing systems often struggle to cope with the demands of Big Data. To address these challenges, the concept of Big Data has evolved, characterized by its 5Vs—volume, velocity, variety, veracity, and value [1]. These 5Vs highlight the intricate nature of navigating modern information landscapes through data analytics.

In parallel, the emergence of GPT-4, the latest iteration of OpenAI's language generation technology, adds a new dimension to the landscape of data processing [2]. GPT-4, built upon the advancements of its predecessors, is a state-of-the-art language model capable of understanding and generating human-like text at an unprecedented scale. GPT-4 has shown remarkable results in various domains, such as healthcare [3]-[5], education [6]-[8], language editing services, assistant writing tools [9]-[11], and engineering [12], [13], among others [14]-[16].

This breakthrough in natural language processing presents an innovative approach to interpreting and communicating vast amounts of textual information.

Just as Big Data addresses the challenges of handling massive datasets, GPT-4 tackles the complexities of language understanding and generation. Its capabilities extend beyond conventional data processing, offering a unique synergy with Big Data analytics. As organizations continue to grapple with the intricacies of data in the digital age, the integration of Big

*Corresponding author. Email: osama.alyasiri@atu.edu.iq

Data and GPT-4 represents a powerful alliance, promising enhanced insights, communication, and decision-making in the dynamic landscape of information technology.

2. 5Vs CHARACTERISTICS OF BIG DATA

The 5Vs characteristics of Big Data are illustrated in Figure 1. These 5Vs collectively form a framework for comprehending and tackling the intricacies and nuances associated with handling large and diverse datasets in the field of Big Data.

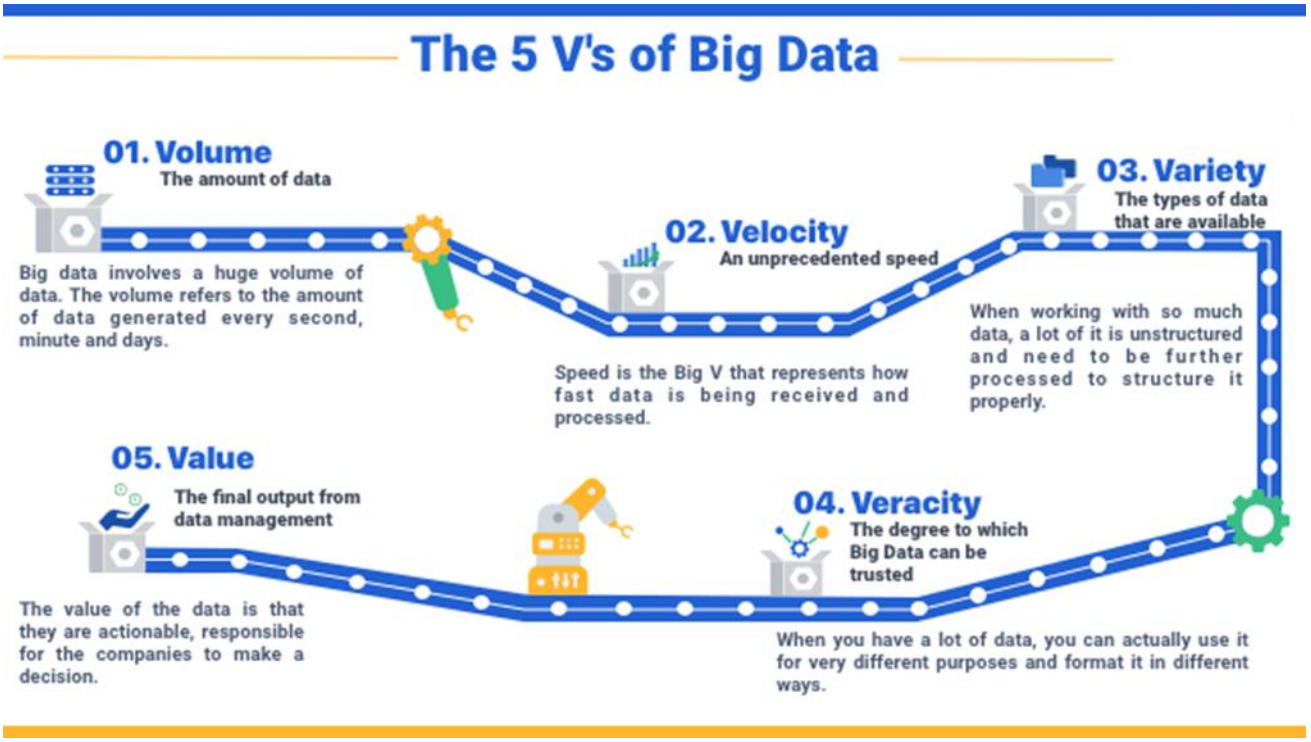

Fig. 1. 5Vs Characteristics of big data [17]

3. GPT-4's CHARACTERISTICS THROUGH THE 5VS OF BIG DATA

In this section, a comparative analysis is undertaken, wherein the 5Vs of Big Data provide a framework to understand and utilize vast datasets effectively [1], [17], [18]. By comparing these characteristics to GPT-4, we are essentially evaluating how GPT-4 handles each aspect:

3.1 **Volume:** The amount of data

- **Big Data:** Refers to the immense volume of data generated every second from multiple sources [19].
- **GPT-4:** As an AI, GPT-4 is designed to process and generate responses from large volumes of data. Its training involves massive datasets to comprehend language patterns and nuances.

3.2 **Velocity:** An unprecedented speed

- **Big Data:** Indicates the high speed at which data is created, processed, and made available for use [19].
- **GPT-4:** Showcases the ability to quickly process inputs and generate outputs, reflecting real-time interaction capabilities with users.

3.3 **Variety:** The types of data that are available

- **Big Data:** Encompasses the different types of data available, including structured, unstructured, and semi-structured data from diverse sources [19].
- **GPT-4:** Can handle a wide range of input types, from plain text to more complex queries, and is also capable of generating diverse forms of content, demonstrating adaptability to various data types.

3.4 **Veracity:** The degree to which big data can be trusted

- **Big Data:** Concerns the trustworthiness and quality of the data, addressing issues of accuracy and consistency [20].
- **GPT-4:** Its effectiveness depends on the veracity of the data it was trained on. While it can produce information that seems accurate, its outputs must be evaluated for reliability, as it can replicate biases or inaccuracies present in the training data

3.5 **Value:** The final output from data management

- **Big Data:** Refers to the actionable insights and benefits that can be derived from processing big data [1].
- **GPT-4:** Decision-making and rough its interactions by providing accurate information, automating tasks, aiding decision-making, and enhancing user experiences.

4. DISCUSSION

GPT-4, as a product of the contemporary AI landscape, mirrors the challenges presented by Big Data's 5Vs: Volume, Velocity, Variety, Veracity, and Value. Regarding Volume, GPT-4 is designed to handle extensive textual data, showcasing an ability to process and analyze large datasets with which Big Data is synonymous. When it comes to Velocity, the AI model excels in its rapid processing capabilities, offering real-time analysis and response generation that align with the swift pace at which Big Data grows and evolves.

Although Big Data encompasses a broad spectrum of data types—from social media interactions to visual content—GPT-4 has traditionally focused on textual information. However, its recent forays into image analysis [21], [22] signal an expanding versatility, aligning it more closely with the Variety aspect of Big Data. This expansion enhances GPT-4's applicability across a wider array of data formats, though it is important to note that its capabilities in non-textual domains are still developing compared to its textual proficiency.

In terms of Veracity, GPT-4 must critically assess and filter through its training data, which is as varied and extensive as Big Data itself, to generate reliable outputs. This parallels the challenges of ensuring accuracy and trustworthiness within Big Data, where the model must discern between accurate and misleading information to maintain the integrity of its responses.

Finally, the Value that GPT-4 provides is in its ability to generate contextually relevant and insightful content, paralleling the goal of Big Data to yield actionable insights. GPT-4's outputs aim to be more than information retrieval; they seek to add context and understanding, turning raw data into useful knowledge.

While it's not entirely fair to compare GPT-4's data processing directly with all forms of Big Data due to differences in data types and maturity in handling diverse content, GPT-4 does embody a microcosm of Big Data's challenges and objectives. With its cloud-based framework, GPT-4 offers scalable and accessible data processing capabilities, transcending geographic limitations and propelling it as a vital tool that reflects and addresses the 5Vs of Big Data. As such, GPT-4 stands as a robust counterpart within the AI sphere, capable of meeting the demands and leveraging the potential of the Big Data landscape.

5. FINDING

From the analysis and ensuing discussion, we can outline the salient parallels between the 5Vs of Big Data and GPT-4's operational framework, highlighting the model's adaptability and efficacy in managing complex datasets.

Volume: Mirroring the expansive nature of Big Data, GPT-4 exhibits an impressive capacity for managing and interpreting substantial volumes of textual data, affirming its competency in engaging with large-scale information landscapes.

Velocity: GPT-4 aligns with Big Data's demands for swift data processing by exhibiting real-time comprehension and generation capabilities. This rapid processing power mirrors the essential quickness required to navigate and utilize the ever-expanding data deluge characteristic of Big Data environments.

Variety: Despite Big Data's multifaceted nature, encompassing various data forms like social media posts, images, and blog content, GPT-4 has demonstrated proficiency in processing a wide array of textual inputs. Its ability to understand and interact with different forms and nuances of language showcases its versatility within the realm of textual data variety.

Veracity: GPT-4’s approach to data accuracy and reliability showcases its alignment with Big Data's veracity challenge. By sifting through large volumes of information, GPT-4 is designed to determine the credibility of sources and the likelihood

of truth, which is crucial in an age where information accuracy is paramount. Its performance in this domain is dependent on the integrity of its training data and its ability to navigate and interpret nuanced and potentially conflicting information.

Value: The pursuit of extracting meaningful insights is a shared objective of Big Data analytics and GPT-4's functionality. GPT-4 enhances this endeavor by providing responses that are not only relevant to the context but also deepen the understanding of the information, thereby adding considerable value to the data it processes.

6. CONCLUSION AND FUTURE WORK

In the ever-evolving landscape of information processing, GPT-4 emerges as a formidable force, seamlessly addressing the challenges posed by the 5Vs of Big Data. Its advanced language understanding, adaptability, and real-time processing capabilities position it as a powerful ally in the quest for meaningful insights in the era of massive datasets. As we witness the convergence of language models and Big Data technologies, GPT-4 stands as a beacon, illuminating the path towards a more sophisticated and nuanced understanding of the vast sea of information. The synergy between Big Data and GPT-4 exemplifies the convergence of advanced data analytics and natural language processing. While Big Data empowers businesses to harness the potential of vast and varied datasets, GPT-4 showcases the remarkable progress in language generation technology, offering new avenues for creative expression, communication, and human-computer interaction. As these technologies continue to evolve, their combined potential holds the promise of transforming industries and shaping the future of intelligent data-driven decision-making.

In future research, an intriguing avenue lies in exploring GPT-4 as a pivotal tool within the realm of big data models, particularly within the context of the ten big characteristics. The investigation should delve into the non-linear interrelationships between these characteristics, examining how GPT-4 aligns with and contributes to the three levels of the unified framework—fundamental, technological, and socio-economic. Assessing GPT-4's capabilities at the fundamental level involves scrutinizing its proficiency in handling voluminous textual data, real-time processing, diverse data formats, and critical veracity assessment, while also evaluating its outputs' contribution to actionable knowledge. At the technological level, the focus should be on understanding the synergy between GPT-4 and big data technologies, considering its cloud-based framework, scalable data processing, and adaptability for complex datasets. Additionally, an exploration of GPT-4's impact on socio-economic aspects, such as accessibility, scalability, and its role within the AI sphere, would provide valuable insights. Adopting a service-oriented perspective, this future work aims to elucidate GPT-4's role in addressing challenges and objectives across the different levels of the proposed framework, thereby advancing understanding in the fields of big data, big data analytics, business intelligence, and business analytics.

Conflicts Of Interest

The authors declare no conflicts of interest.

Funding

None.

Acknowledgment

The corresponding author would like to express gratitude to Al-Furat Al-Awsat Technical University for its invaluable support throughout this research project.

References

[1] Anuradha, J., "A brief introduction on Big Data 5Vs characteristics and Hadoop technology," Procedia Computer Science, vol. 48, pp. 319-324, 2015.

[2] Hassani, H., and Silva, E. S., "The role of ChatGPT in data science: how ai-assisted conversational interfaces are revolutionizing the field," Big Data and Cognitive Computing, vol. 7, no. 2, art. no. 62, 2023.

[3] Vaishya, R., Misra, A., and Vaish, A., "ChatGPT: Is this version good for healthcare and research?," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 17, no. 4, art. no. 102744, 2023.

[4] Scheschenja, M., Viniol, S., Bastian, M. B., Wessendorf, J., König, A. M., and Mahnken, A. H., "Feasibility of GPT-3 and GPT-4 for in-Depth Patient Education Prior to Interventional Radiological Procedures: A Comparative Analysis," CardioVascular and Interventional Radiology, 2023.

[5] Currie, G., Robbie, S., and Tually, P., "ChatGPT and Patient Information in Nuclear Medicine: GPT-3.5 Versus GPT-4," Journal of Nuclear Medicine Technology, 2023.

[6] Firdaus, M. F., Wibawa, J. N., and Rahman, F. F., "Utilization of GPT-4 to Improve Education Quality Through Personalized Learning for Generation Z in Indonesia," IT for Society, vol. 8, no. 1, 2023.

[7] Moore, S., Nguyen, H. A., Chen, T., and Stamper, J., "Assessing the quality of multiple-choice questions using gpt-4 and rule-based methods," in European Conference on Technology Enhanced Learning, Cham: Springer Nature Switzerland, pp. 229-245, 2023.

[8] Yu, H., "Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching," Frontiers in Psychology, vol. 14, art. no. 1181712, 2023.

[9] Alyasiri, O. M., Salisu, S., Younis, H. A., Salman, A. M., Sahib, T. M., Akhtom, D., and Hayder, I. M., "ChatGPT Revisited: Using ChatGPT-4 for Finding References and Editing Language in Medical Scientific Articles," SSRN Electronic Journal, 2023. [Online]. Available: https://ssrn.com/abstract=4621581 or http://dx.doi.org/10.2139/ssrn.4621581

[10] Mohammed, O., Sahib, T. M., Hayder, I. M., Salisu, S., and Shahid, M., "ChatGPT Evaluation: Can It Replace Grammarly and Quillbot Tools?," British Journal of Applied Linguistics, vol. 3, no. 2, pp. 34-46, 2023.

[11] Sahib, T. M., Alyasiri, O. M., Younis, H. A., Akhtom, D., Hayder, I. M., Salisu, S., and Muthmainnah, "A comparison between ChatGPT-3.5 and ChatGPT-4.0 as a tool for paraphrasing English Paragraphs," International Applied Social Sciences (C-IASOS-2023) Congress, 2024. In press.

[12] Cheng, K., Guo, Q., He, Y., Lu, Y., Gu, S., and Wu, H., "Exploring the potential of GPT-4 in biomedical engineering: the dawn of a new era," Annals of Biomedical Engineering, 2023.

[13] Sahib, T. M., Younis, H. A., Alyasiri, O. M., Ali, A. H., Salisu, S., Noore, A. A., Hayder, I. M., and Shahid, M., "ChatGPT in Waste Management: Is it a Profitable," Mesopotamian Journal of Big Data, 2023.

[14] H. A. Dida, D. S. K. Chakravarthy, and F. Rabbi, "ChatGPT and Big Data: Enhancing Text-to-Speech Conversion," Mesopotamian Journal of Big Data, vol. 2023, pp. 33-37, 2023.

[15] P. Sharma and B. Dash, "Impact of big data analytics and ChatGPT on cybersecurity," in Proc. 4th Int. Conf. Computing and Communication Systems (I3CS), 2023, pp. 1-6.

[16] D. Akhtom, O. M. Alyasiri, E. Allogmani, A. M. Salman, and T. M. Sahib, "Unlocking Chatgpt's Title Generation Potential: An Investigation Of Synonyms, Readability, And Introduction-Based Titles," Journal Of Theoretical And Applied Information Technology, vol. 101, 2024, in press.

[17] C. Gaur, "What is Big Data: Characteristics, Challenges, Tools & Use Cases," Xenonstack, Oct. 24, 2020. [Online]. Available: https://www.xenonstack.com/blog/what-is-big-data

[18] J. Arcondara, K. Himmi, P. Guan, and W. Zhou, "Value oriented big data strategy: analysis & case study," in Proc. of the 50th Hawaii Int. Conf. on System Sciences, 2017, pp. 1053-1062.

[19] R. Dautov and S. Distefano, "Quantifying volume, velocity, and variety to support (Big) data-intensive application development," in Proc. IEEE Int. Conf. on Big Data (Big Data), 2017, pp. 2843-2852.

[20] A. P. Reimer and E. A. Madigan, "Veracity in big data: How good is good enough," Health Informatics Journal, vol. 25, no. 4, pp. 1290-1298, 2019.

[21] O. V. Johnson, O. M. Alyasiri, D. Akhtom, and O. E. Johnson, "Image Analysis through the lens of ChatGPT-4," Journal of Applied Artificial Intelligence, 2023, in press.

[22] O. M. AL-Janabi, O. M. Alyasiri, and E. A. Jebur, "GPT-4 Versus Bard and Bing: LLMs for Fake Image Detection," in Proc. 3rd Int. Conf. on Intelligent Cybernetics Technology & Applications (ICICyTA), 2024, in press.