The Accuracy of AI Graph Interpretations

Danika Wagner

PSYC 6135: The Psychology of Data Visualization

Dr. Michael Friendly

April 21, 2025

The Accuracy of AI Graph Interpretations

Background and Relevant Research

Recent advances in computer systems suggest that Artificial Intelligence (AI) is capable of simulating human cognition by performing various tasks that traditionally require human intelligence. These tasks include complex functions such as visual perception, language understanding, and decision-making processes (Rashidi et al., 2024). AI systems operate by meticulously observing patterns within vast amounts of data. They employ sophisticated mathematical models to analyze this data and make predictions based on the identified patterns (University of Illinois, 2024). Moreover, these systems continuously test and measure their performance through various metrics to refine their predictions. As more data is collected and analyzed, the AI adjusts its models to enhance accuracy and relevance (University of Illinois, 2024). Consequently, AI systems develop a form of "expertise" through ongoing self-assessments every time they process new data. This includes information shared by individual users who interact with the system, which is particularly relevant in the case of conversational AI like ChatGPT (OpenAI, 2025). By learning from these interactions, AI can improve its understanding and responses, creating a more personalized and practical user experience.

Within the expansive and evolving landscape of AI, it is crucial to recognize the diverse models that contribute to its framework. Prominent among these are machine learning, deep learning, natural language processing, generative AI, and large language models, each offering unique capabilities and applications (Iqbal, 2024). Machine learning is the foundation of AI technologies, using algorithms that identify patterns in vast datasets to make informed predictions about future outcomes. This approach has revolutionized industries by enabling systems to learn from data without explicit programming, significantly enhancing decision-making processes (IBM, 2023). Deep learning is a sophisticated subset of machine learning, characterized by its ability to process and analyze information across multiple layers of data. This multi-layered approach facilitates the prediction of clustered outcomes, providing deeper insights into complex datasets and further advancing the capabilities of technology in automating intricate tasks (IBM, 2023).

Beneath the broader umbrella of deep learning lie natural language processing and generative AI, two interconnected fields that enhance human-computer interaction. Natural language processing focuses on understanding and interpreting human language by parsing text and speech, analyzing semantics, and translating recognized patterns into a format that machines can understand (Vaniukov, 2024). This technology underpins many applications, from voice-activated assistants to real-time translation services. Generative AI takes this a step further by producing meaningful content in response to user prompts. It can generate a wide range of outputs, such as written text, images, or other graphical forms, demonstrating creativity and adaptability in responding to inputs. This ability to create new content marks a significant leap in AI capabilities, offering exciting possibilities in various fields, including art, design, and content creation.

Large language models represent the convergence of natural language processing and generative AI, merging the power of pattern recognition with the ability to create coherent, human-readable content. These models leverage extensive datasets to train on language structures, enabling them to generate responses that mimic human abilities (Callabar, 2024; Ghosh, 2024). As AI technology continues to advance, the significance of these models will only grow, reshaping how we interact with machines and rely on their capabilities to augment human endeavours, such as graph reading, interpretation, and generation.

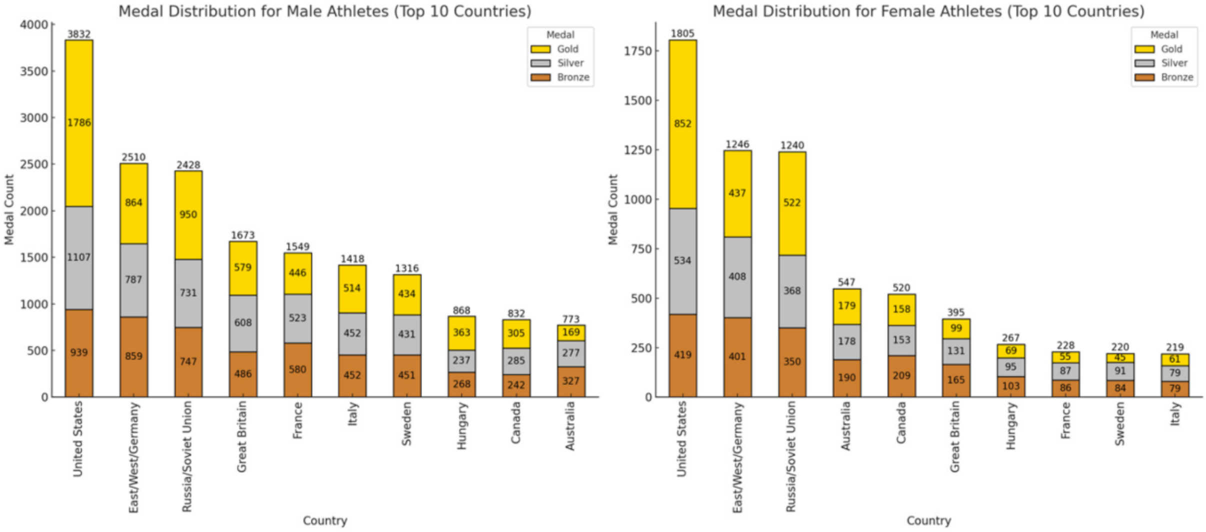
The field of artificial intelligence has witnessed remarkable advancements in recent years, leading to a substantial rise in public and industry interest in AI technology. Notably, Google Trends data from April 19, 2025, reveals that the number of searches containing the term "AI" has surged by 5,000% since March 2015. This statistic highlights the increasing curiosity and engagement with artificial intelligence among internet users. Indeed, recent large-scale surveys have found that up to 90% of American citizens have heard a little about AI, though few reported a high level of familiarity (Faverio & Tyson, 2023). This search pattern and high level of public awareness show a clear and pronounced increase in the public's fascination with AI, including its potential applications and implications for visual perception and understanding, particularly in relation to graphs.

Recent data suggests that AI has become increasingly prevalent in both business and academia. Notably, in the second quarter of 2024, 6.1% of Canadian businesses reported using AI in the production of goods and the delivery of services, as indicated by a report from Statistics Canada (Bryan, Sood, & Johnston, 2024). This is a significant milestone, reflecting the growing integration of AI into various industries. Furthermore, the AI Index Report (Stanford University, 2024) illustrates broader trends within the education and research sector, revealing that research focused on AI has tripled between 2010 and 2022. This rapid increase underscores the importance of AI in shaping future innovations and educational methodologies, as institutions seek to enhance their capabilities and outcomes through these advanced technologies. As capabilities and outcomes increase through the use of AI, quantifying their accuracy will become increasingly important.

Generative AI technologies, such as ChatGPT and Claude 3, show significant promise in both creating and interpreting graphs (Brownell, 2025; Friendly & Claude, 2025). These AI systems rely on iterative processes, which have proven effective in producing graphs that adhere to established best practices in data visualization (DeJeu, 2024). However, it is essential to note that, according to DeJeu (2024), the generation of these specific types of graphs, illustrated in Figure 1, necessitates multiple prompts to achieve satisfactory results. Furthermore, while generative AI can excel in many aspects, there have been instances where ChatGPT produced ineffective data visualizations and made incorrect calculations in its attempts to create these graphs (DeJeu, 2024). This indicates that, despite their advancements, limitations and challenges remain in applying generative AI for data visualization tasks.

Figure 1

Example of ChatGPT-Generated Graph



Note. Taken from DeJeu (2024), who used ChatGPT for data analysis and visualizations.

Large language models and generative AI have gained significant attention for their capability to generate various forms of content, including graphs. However, it remains an open question how effectively these models interpret graphical information and convey its meaning. Brownell (2024) explored this issue and found ChatGPT demonstrated a foundational understanding of graphs, performing particularly well in interpreting simple graphical representations. In contrast, it encountered difficulties when tasked with analyzing more complex figures, especially those that presented multiple metrics and indices, which can be inherently challenging to synthesize into clear interpretations. In a similar vein, Friendly and Claude (2025) found that Claude 3 was capable of articulating key points derived from the textual descriptions accompanying graphs. However, they noted that Claude 3 tended to focus more on the textual information rather than the graphical elements themselves. When given a prompt to elaborate on relationships between different variables and to represent those relationships through graphs, Claude 3 encountered significant challenges (Friendly & Claude, 2025). The graphs it produced often fell short of adhering to established best practices in graph creation, indicating a gap in its capacity to generate graphical representations that are both informative and visually effective (Friendly & Claude, 2025; Tufte, 1983).

This suggests that while advances in generative AI hold great potential for enhancing graph creation, there is still much to be done to improve their ability to interpret and represent complex data accurately and effectively. The discrepancies highlighted in the studies emphasize the need for further research to bridge the current gaps in understanding and improving the graphical competencies of AI models, particularly as their applications continue to expand in various fields of research and data analysis.

Moreover, given that the stated goal of AI is to simulate human abilities, it is important to understand whether AI is meeting this significant challenge. Cleveland and McGill (1984) define human graphical perception as the process by which the visual system decodes the information in a graph into a numerical or categorical understanding. They found that humans interpret visual information in graphs most accurately when it is encoded as position along a common scale, followed by position along non-aligned scales, then length, angle, area, volume, and finally colour. The researchers discovered that as encoding shifts from position to colour, human perceptual accuracy decreases, indicating that graphs relying on position and length are more effective for conveying precise information than those depending on angle, area, or colour. Yet, it remains unclear if AI systems adhere to the same graphical perception accuracy hierarchy.

Padilla et al. (2018) propose a cognitive framework for human decision-making that employs visualizations integrating dual-process theories of cognition. They identify four aspects of visualization design and human visual reasoning influencing interpretation and decision-making: bottom-up attention, visual-spatial biases, cognitive fit, and knowledge-driven processes. Visualizations can aid or hinder visual reasoning through bottom-up processes by guiding viewer attention (Padilla et al., 2018). The encoding techniques used in visualizations, such as those highlighted by Cleveland and McGill (1994), also affect visual reasoning and can introduce biases (Padilla et al., 2018). Additionally, Padilla et al. (2018) emphasize that visualizations aligning well with the viewer's cognitive processes can lead to more effective decisions and greater acuity in visual reasoning. Finally, the researchers underscore that a viewer's prior knowledge can interact with visualization design, thereby influencing decision-making and interpretation. Taken together, Padilla et al. (2018) highlight the importance of considering both automatic (i.e., bottom-up attention, visual-spatial biases) and deliberate (i.e., cognitive fit and knowledge-driven processes) factors in the design and interpretation of graphs. Yet, whether AI systems employ automatic and deliberative processes when interpreting graphs remains unclear.

Research Questions

AI primarily learns and hones its skills by engaging in a process of self-testing alongside training datasets. This self-evaluation phase is crucial, as it enables the AI to refine its algorithms and deepen its understanding of various types of data. However, equally important is the AI's proficiency in accurately interpreting new, unseen data, especially when that data is presented in a graphical format provided by users. Therefore, it is essential to assess how well large language models can interpret not only straightforward graphs but also more intricate and complex graphical representations. This raises a significant question in the field of AI: how does the accuracy of large language models compare to that of a human when dealing with simple graphs versus complex ones? Do AI models make similar mistakes to humans (e.g., struggling more with pie charts than bar charts)?

Moreover, given the significant claims that AI simulates human cognition by performing tasks that typically require human visual perception and reasoning abilities, it is crucial to evaluate the validity of these assertions thoroughly. AI has made great strides in processing information and interpreting graphical data, but it is essential to critically assess the extent to which AI's capabilities can truly mirror those of human cognition. Specifically, how does AI's performance in the realm of graphical interpretation compare to that of humans?

Methods

The proposed study will employ a comparative experimental design to assess and compare how accurately humans and an AI model interpret common types of data visualizations. The goal is to evaluate how well AI "understands" graphs in ways similar to humans, using a shared set of visual tasks. The proposed study will recruit 60 young adult participants from a diverse university in a metropolitan area. Participants will report their sex, gender, socioeconomic status, and racial background before completing a measure that assesses their prior experience interpreting and producing data visualizations. Following this, participants will complete a brief online task involving visual reasoning with charts. The AI model will be a multimodal large language model capable of interpreting visual inputs (e.g., ChatGPT-4). The model will be shown the same graphs and asked the same questions as human participants, with prompts crafted to closely mirror the human task instructions.

The proposed study will include several types of data visualizations, categorized into two types: simple and complex. Simple graphs consist of visualizations that are ubiquitous, such as bar charts, pie charts, line graphs, and histograms. Complex graphs are visualizations that are less commonly found in everyday life, such as scatterplots, area charts, heat maps, 3-dimensional plots, boxplots, and violin plots. Each type of graph will be presented in several forms, according to the hierarchy of graphical perception described by Cleveland and McGill (1984). Moreover, each will be designed to test comprehension of a basic quantitative relationship. Participants will be asked a series of questions that assess their ability to accurately interpret and draw inferences about the variables used within the graphs. Specifically, each graph will be accompanied by three open-ended questions assessing literal comprehension(e.g., "What value does X reach at point Y?"), trend or pattern detection (e.g., "Which category shows the greatest increase over time?"), and inference (e.g., "What can you predict about Z based on this trend?"). The AI model (ChatGPT-4) will then be asked the same series of questions after being presented with the same graphs.

Three main metrics will be assessed to compare the AI model's performance with human performance. First, accuracy metrics will be calculated based on the percentage of correct answers for each question, graph type, and comprehension category (simple versus complex). Second, a consistency metric will be calculated based on the internal coherence of responses across similar questions or repeated trials (e.g., whether an AI or human gives logically consistent answers across related prompts). Finally, an error analysis will be conducted to determine whether certain graph types and comprehension categories (simple versus complex) are more prone to misunderstanding by AI or humans. Errors will be classified according to type: perceptual errors (e.g., misreading a value or axis), inferential errors (e.g., incorrect trend-based predictions), and graph-specific errors (e.g., misinterpreting scatterplots more than bar charts).

This study will utilize descriptive statistics to provide a comprehensive summary of performance metrics across different types of graphs and comprehension categories. This approach will enable us to capture essential trends and variations in the data, facilitating a clearer understanding of how well individuals perform when interpreting various graphical representations. Additionally, to evaluate the differences in graphical acuity between artificial intelligence systems and human participants, inferential statistical analysis will be conducted. Specifically, t-tests will be employed to compare the performance outcomes of AI and human subjects. This method will allow us to determine whether there are statistically significant differences between the two groups, contributing valuable insights into the comparative effectiveness of AI versus human interpretation of graphical data. Finally, given that individuals may differ in their prior experience with data visualization, including previous coursework in such topics, additional analyses exploring the differences in graphical perception acuity between AI and those with high and low experience may provide more estimates of the accuracy of AI graphical interpretations.

While the core focus of the proposed research does not centre on this aspect, delving into qualitative analyses of open-ended AI responses could yield significant insights. By examining how AI interprets and responds to graph-related inquiries, we may uncover a deeper understanding of its operational mechanisms. As demonstrated by Brownell (2024), the explanations that AI provides for its responses can be particularly illuminating. They reveal much about the often opaque, or black-box, nature of the functionalities inherent in AI systems. This exploration could open avenues for further research into the cognitive processes of AI, whether deliberative or automatic, enhancing our ability to comprehend the reasoning behind its outputs and improving overall transparency in AI technologies.

Conclusion

Given that AI purports to simulate human cognition, it is imperative to formally assess its abilities in comparison to those of humans. In emerging areas such as AI graphical perception and visual reasoning, it remains an open question how accurately AI systems can interpret and understand both simple and complex data visualizations compared to human capabilities. We should examine various benchmarks, tasks, and contexts in which both AI systems and humans operate, shedding light on whether AI can reach, meet, or even exceed human performance in these areas. The proposed research not only has implications for understanding the current abilities of AI but also for its future development and integration into tasks that traditionally rely on human insight, perception, and reasoning. By recognizing the differences in AI capabilities, developers can create more sophisticated and reliable systems. These improved systems will not only enhance user experiences but also be equipped to assist with a broader spectrum of analytical tasks, ranging from simple data interpretation to complex problem-solving. As a result, users will have access to AI tools that are not only efficient but also versatile, enabling them to tackle challenges across various domains effectively.

This study has the potential to impact the field of AI in several ways significantly. One of the most exciting aspects is its influence on AI education tools, which are critical for equipping future generations of technologists with the skills they need to thrive in an increasingly digital world. Furthermore, the research addresses crucial aspects of safety in data-driven decision-making, ensuring that AI technologies are utilized responsibly and ethically. In addition to these educational and safety implications, the results of this research can extend into multiple domains, including healthcare, finance, and innovative technologies. By enhancing our understanding of AI systems and their applications, this study contributes to theoretical frameworks and practical implementations that potentially transform industries. The far-reaching implications of this research may lead to innovative solutions and advancements in the AI landscape, promoting a deeper integration of AI technologies into everyday life and enhancing how we interact with intelligent systems.

References

Brownell, B. (2025). ChatGPT data analysis in 2025: Interpret charts & diagrams. *Descript* [Online]. https://www.descript.com/blog/article/using-chatgpt-data-analysis-to-interpret-charts--diagrams

Bryan, V., Sood, S. & Johnston, C. (2024). Analysis on artificial intelligence use by businesses in Canada, second quarter of 2024, *Analysis in Brief - Statistics Canada 2024* [online]. https://www150.statcan.gc.ca/n1/pub/11-621-m/11-621-m2024008-eng.htm

Caballar, R. D. (2024). Generative AI vs. Predictive AI: What's the difference? IBM. https://www.ibm.com/think/topics/generative-ai-vs-predictive-ai-whats-the-difference

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

DeJeu, E. B. (2024). Using Generative AI to Facilitate Data Analysis and Visualization: A Case Study of Olympic Athletes. *Journal of Business and Technical Communication, 38*(3), 225-241. https://doi.org/10.1177/10506519241239923

Faverio, M. & Tyson, A. (2023). What the data says about Americans' views of artificial intelligence, *Pew Research Center* [Online]. https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/

Friendly, M., & Claude, H. (2025). What does AI understand about a graph?, *Nightingale* [Online]. https://nightingaledvs.com/what-does-ai-understand-about-a-graph/

Ghosh, P. (2024). Generative AI vs. traditional AI. Dataversity. https://www.dataversity.net/generative-ai-vs-traditional-ai/

Gierhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). International Conference on Learning Representations. New Orleans. Retrieved 2025, from https://arxiv.org/abs/1811.12231

Google. (n.d.). *Search term: AI.* Google Trends [Online]. https://trends.google.com/trends/explore?date=2015-03-19%202025-04-19&geo=CA&q=ai&hl=en-GB

IBM Data and AI Team. (2023). AI vs. Machine Learning vs. Deep Learning vs. neural networks. https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Iqbal, M. U. (2024). AI, ML, DL, NLP, and Generative AI: A beginner's guide. *LinkedIn* [Online]. https://www.linkedin.com/pulse/ai-ml-dl-nlp-generative-beginners-guide-muhammad-usman-iqbal-8cgcf

OpenAI. (2025). Image inputs for ChatGPT - FAQ, *OpenAI Help Center* [Online]. https://help.openai.com/en/articles/8400551-image-inputs-for-chatgpt-faq

Rashidi, H. H., Pantanowitz, J., Hanna, M. G., Tafti, A. P., Sanghani, P., Buchinsky, A., Fennell, B., Deebajah, M., Wheeler, S., Pearce, T., Abukhiran, I., Robertson, S., Palmer, O., Gur, M., Tran, N. K., & Pantanowitz, L. (2025). Introduction to artificial intelligence and machine learning in pathology and medicine: Generative and Nongenerative Artificial Intelligence Basics. Modern Pathology, 38(4), 100688. https://doi.org/10.1016/j.modpat.2024.100688

Stanford University Human-Centered Artificial Intelligence. (2024). (rep.). *Artificial Intelligence Index Report 2024* [Online]. https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

University of Illinois. (2024). Traditional AI vs. Generative AI: What's the difference? University of Illinois Urbana-Champaign - College of Education. https://education.illinois.edu/about/news-events/news/article/2024/11/11/what-is-generative-ai-vs-ai