Artificial Intelligence in Higher Education: A Comparative Study on the Performance and Detectability of AI-Generated Graduate-Level Coursework Utilizing ChatGPT

Kenneth R. Deans Jr., DHA, MBA[1]✉, Jami Jones, PhD, MHA[2], Jillian B. Harvey, PhD, MPH[2] & Daniel Brinton, PhD, MHA, MAR[2]

Abstract

This study examined the performance and detectability of AI-generated graduate-level coursework in higher education. The research problem was determining whether AI-generated coursework submissions for a graduate-level course from start to finish could achieve high performance without detection by university professors or any in-place AI-detection platforms. A comparative study was conducted to answer this question. The outcomes of this study were twofold. First, the data analysis provided insight into the effectiveness of AI-generated coursework relative to human-generated performance in higher education settings. Second, the observed detectability results indicated whether university professor(s) or in-place AI-detection platforms identified coursework as AI-generated versus human-generated. This research will inform academic integrity challenges and future research on artificial intelligence and potential implications related to teaching and learning in higher education and academic settings.

Author Affiliations:

[1]Health Sciences South Carolina

[2]Medical University of South Carolina, College of Health Professions

Correspondence to:

Kenneth Deans Jr., Health Sciences South Carolina, 3790 Fernandina Rd., Suite 301, Columbia, SC 29210

kdeans@HealthSciencesSC.org

Acknowledgements:

The authors would like to extend their gratitude to OpenAI for the use of ChatGPT, which served as a pivotal tool in this research. They also thank the *Journal of Health Administration Education* (*JHAE*) reviewers for their comprehensive feedback and ongoing commitment to advancing scholarship in health administration education.

Introduction

The evolution of generative artificial intelligence (AI) has been a cornerstone of recent technological advancements, heralding a transformative era across multiple fields and industries, notably in education (Menekse, 2023). These advancements have enabled AI systems and large language

models (LLMs) like ChatGPT, GPT-3, GPT 3.5, and GPT-4 to produce text indistinguishable from human output, posing significant implications for the higher education landscape (Jacobs et al., 2023). In this context, AI's ability to generate academic coursework that closely mirrors human intellect and expression, emerged as both an opportunity and a challenge, especially in maintaining academic integrity and authenticity in intellectual endeavors.

Higher education overwhelmingly places emphasis on the pillars of independent thought, critical analysis, and originality, particularly at the graduate level (Lin, 2023). The emergence of sophisticated AI models and tools, such as generative AI platforms, introduces new dimensions to this conversation, challenging traditional notions of creativity and originality. Generative AI refers to a subset of AI technologies capable of generating new content, including text, images, music, and software development code, based on its training data (Goodfellow et al., 2014). This category of AI is noted for its ability to produce novel creations that closely resemble human-generated content, extending its applications from artistic endeavors to practical solutions across various industries.

At the core of generative AI are algorithms that learn patterns and features from large datasets. These algorithms, often based on neural networks, generate new, original content using this learned information. One of the most prominent examples of this technology is the generative adversarial network (GAN), introduced by Goodfellow et al. (2014). GANs consist of two neural networks: a generator that creates content and a discriminator that evaluates it. Through iterative training, the generator improves its ability to produce realistic outputs.

Another significant development in generative AI is using transformers, a type of neural network architecture introduced by Vaswani et al. (2017). Transformers have proven highly effective in understanding and generating human language, leading to breakthroughs in natural language processing (NLP). Models like GPT (generative pretrained transformer) by OpenAI exemplify this success, demonstrating advanced capabilities in generating coherent and contextually relevant text.

Further contributing to the evolution of AI is the development of language models such as ChatGPT, released as open source in November 2022, GPT-3, GPT 3.5, and GPT-4 (Flanagin et al., 2023). ChatGPT serves as a user-facing interface designed to engage in natural language dialogues. Its primary function is to facilitate an interactive and intuitive platform for users to communicate with an AI system in a conversational format. Underpinning this interface is a sophisticated language model that processes, understands, and generates human-like text responses. The essence of ChatGPT is its ability to maintain

contextually coherent and relevant conversations, a feature progressively refined through successive iterations of the GPT series (Vaswani et al., 2017). GPT-4, representing the latest advancement in this series, is an evolution of its predecessors, offering enhanced language understanding and generation capabilities. Unlike the earlier GPT-3 and GPT-3.5, GPT-4 exhibits improved accuracy, contextual relevance, and nuanced text-generation performance. This progression mirrors the ongoing development in the field, where each new iteration of the GPT series brings forth more sophisticated and capable AI models, thereby expanding the horizons of what AI can achieve regarding language comprehension and interaction.

Users who interact with the ChatGPT interface choose GPT-3.5 or GPT-4 as the underlying AI model. This selection is akin to choosing between two engines for the exact vehicle—the interface (ChatGPT) remains constant, but the performance and capabilities are contingent on the chosen model. Opting for GPT-4 means leveraging a more advanced and capable engine to power the conversational AI experience.

In education settings, the emergence of these sophisticated AI models challenges the traditional boundaries of human-generated thought and analysis. These models' ability to replicate complex human thinking processes necessitates critically examining their impact on academic integrity and the efficacy of AI-detection tools in distinguishing AI-assisted content from human efforts (Liao et al., 2023).

In education settings, the emergence of these sophisticated AI models challenges traditional boundaries of human-generated thought and analysis. These models' ability to replicate complex human thinking processes necessitates critically examining their impact on academic integrity and the efficacy of AI-detection tools in distinguishing AI-assisted content from human efforts (Liao et al., 2023).

Despite an abundance of research since the introduction of ChatGPT, scholars have yet to explore AI utilization thoroughly, specifically generative AI models like ChatGPT, in completing entire graduate-level courses, beginning to end. Despite this subject's critical importance, a notable gap persists in research comprehensively addressing these concerns (Grimaldi & Ehrler, 2023). This gap is significant in elucidating the depth and versatility of AI in academic contexts.

To redress this gap, research was conducted to evaluate the performance and detectability of end-to-end AI-generated graduate-level coursework in a higher education setting. Our central goal was to determine if such coursework, indistinguishable from that produced by human students, could attain high-performance levels and whether it would be recognizable as AI-generated by university professors or AI detection tools.

We outlined clear protocol to establish that our research did not involve human subjects, adhering to the federal definition (45 CFR 46.102). This clarification was crucial as our approach did not involve direct intervention or interaction with human subjects, nor did it entail collecting identifiable personal data.

Given the nature of our study, it was essential to define what constitutes a human subject in this context. According to federal regulations, a human subject is any living individual from whom data is obtained through interaction or intervention or whose identifiable private information is used. Our research design strategically avoided these criteria.

The core of our study involved enrolling AI in an online graduate-level course. The AI was tasked with completing various academic assignments, such as essays, discussion posts, case studies, quizzes, and participation in live seminars. The AI, specifically utilizing ChatGPT, generated all coursework, and no independent or critical thought was allowed from the enrolled student. The student's role was limited to submitting or verbalizing content produced by the AI. The AI-enrolled student was withdrawn from the course 2 weeks before its conclusion, preventing the posting of final grades to ensure anonymity and avoid ethical concerns.

This research methodologically focused on assessing AI-generated coursework in a real-world academic environment, strictly adhering to ethical guidelines and maintaining the anonymity of human participants involved in the educational process. The study's design aimed to provide insightful data on the capabilities and limitations of AI in replicating graduate-level academic work, contributing to the broader understanding of AI's role in higher education.

This overarching question is further divided into two subquestions: the comparative performance of AI-generated coursework against human-generated coursework in graduate-level education and the effectiveness of AI-detection platforms in academia in identifying AI-generated coursework (Lu et al., 2023).

Two hypotheses have been formulated in alignment with these questions. The first hypothesis (H1) suggests that AI-generated coursework will perform comparably to human-generated coursework in graduate-level education. The second hypothesis (H2) posits that AI-generated coursework may not be consistently detected by AI-detection platforms, highlighting the sophisticated capabilities of AI models in mimicking human responses and the potential limitations of current AI-detection technology (Uzair, 2023).

This study aims to generalize its findings to a broader context within the graduate-level academic setting. The methods section elaborates on the detailed data collected.

Methods

Research Design and Sample Selection

In this research study, we meticulously crafted a methodology to investigate AI-generated content's performance capabilities and detectability within an academic setting, specifically focusing on a single Master of Health Administration (MHA) course at a well-known university. The essence of our approach was to compare AI-generated coursework to work produced by human students in a real-world educational environment, thereby gaining insight into the impact of AI in higher education.

The research design was a comparative study. This design was chosen for its robustness in generating numerical data that could be subjected to statistical analysis. Such an approach was crucial for evaluating the performance of AI-created coursework, as reflected in the grades and feedback received, and for detecting patterns indicating AI involvement. We refrained from using such qualitative methods like interviews, focusing on measures that could be subjected to qualitative review to ensure objectivity.

We utilized a convenience sampling method for the sample selection due to the study's confinement to a single MHA course. This method facilitated efficient resource utilization within the specific context. The AI was enrolled as a student in the course, and the sample consisted of all the coursework produced by the AI during the study period. This sample mirrored the range of work typically completed by a graduate student, including essays, project reports, and discussion posts. To maintain the study's integrity, AI's involvement was not disclosed to the professor or other students, ensuring that the work was graded and received feedback as any human student's work would.

Course and Grading Protocol

The course selected for this study was a graduate-level program with high overall grade performance, reflecting the rigorous standards and expectations of the curriculum. Faculty utilized defined rubrics for grading cases, discussion boards, and other assignments, ensuring a standardized and fair evaluation process across all submissions. The assignments required the application of specific concepts relevant to the course's focus, which included elements of management, leadership, law, and some finance, among others. These rubrics ensured consistency and objectivity in grading, maintaining academic integrity.

To address the representativeness of the cohort, it is essential to note that our study's constraints, primarily driven by ethical considerations and the need to preserve anonymity, precluded the collection and disclosure of demographic or descriptive data about the student cohort. Our approach aimed to simulate a typical graduate-level academic environment while strictly adhering to ethical

guidelines. Consequently, while additional demographic information could enhance the representativeness and generalizability of our findings, the study's integrity and adherence to ethical standards took precedence. This approach ensured that the research environment remained authentic and unbiased, focusing on the performance and detectability of AI-generated coursework within a real-world academic setting.

While the overall grade performance for the course was observed to be high, this information was consistent with the program's academic rigor and the caliber of the student body. The course assignments encompassed open-ended and specific concept applications, thoroughly assessing student understanding and performance. However, further specifics cannot be disclosed to protect the course and participants' anonymity. This approach ensured that the research environment remained authentic and unbiased, focusing on the performance and detectability of AI-generated coursework within a real-world academic setting.

Validity and Reliability of Assessments

The primary instruments for data collection were the assignments and assessments integral to the MHA course. As part of the established curriculum, these instruments were already vetted for their validity and reliability in an academic setting. They provided a direct and unbiased means to evaluate the AI's performance. However, while using these instruments in this study was reasonable, it is essential to acknowledge this approach as a limitation. These assessments were originally designed to evaluate human work rather than AI-generated content, which may affect their validity and reliability when applied to AI outputs.

The validity of the assessments refers to the extent to which they accurately measure what they are intended to measure—in this case, students' academic performance. Since these tools were not specifically developed to assess AI-generated work, they may not fully capture the nuances or potential biases inherent in AI-generated responses. Nevertheless, the high performance of the AI in aligning with or exceeding human-generated content suggests a degree of validity in the assessments used.

On the other hand, reliability concerns the assessment results' consistency. The defined rubrics and standardized grading protocols employed by the faculty helped ensure a consistent evaluation process across all submissions, human or AI-generated. This consistency is crucial to maintaining the reliability of the results. However, given that AI-generated work might exhibit different patterns or characteristics than human submissions, the assessments' reliability in this context should be interpreted cautiously. By recognizing

these limitations, we underscore the need for further research to develop and validate assessment tools tailored to evaluate AI-generated academic work, which would enhance the accuracy and fairness of evaluations, contributing to a more robust understanding of AI's role and performance in educational settings.

Data Collection and Analysis

Regarding the data set description, the study did not use archival data. The data comprised AI-generated coursework and the corresponding grades and feedback from the course instructor. These data were treated with the utmost confidentiality and anonymized to focus on the performance and detectability of the AI-generated content. Additionally, ethical guidelines were strictly followed, considering that the AI's enrollment and participation in the course were under the guise of a typical student, thereby not involving human subjects directly.

Our research method was rigorous and comprehensive. The AI was tasked with completing various assignments typical of a graduate-level course, including participation in live seminars. These assignments encompassed both open-ended essay-style assignments based on the concepts and materials and those requiring the application of specific concrete concepts relevant to the course's focus. The results section reports the outcomes of all types of assignments encountered in the study, from discussion boards to specific 100-question quizzes.

ChatGPT 4 was the AI tool used and was operational between June and August 2023. Appendices 1.0 through 1.5 provide examples of the exact GPT prompts engineered and utilized during this period to perform various academic tasks. These examples illustrate the specificity and structure of the instructions given to ChatGPT 4. The AI was instructed to produce outputs indistinguishable from human-written essays, avoiding repetitive language or overly complex sentences that could hint at AI authorship. These prompts were carefully crafted to ensure that the AI-generated coursework closely mirrored human-written content. The AI's responses were unique and plagiarism-free, processed through Grammarly.com for grammatical refinement following academic and APA formatting standards.

In any instances where Grammarly.com may have detected plagiarism, ChatGPT was directed to rewrite the specific passages, thereby maintaining the originality of the submissions. However, no plagiarism was detected throughout the study. This attention to detail was pivotal in preserving the authenticity of the AI-generated coursework.

Citations provided by the AI via ChatGPT 4 were cross-verified using Google Scholar, ensuring academic integrity and APA formatting. Each reference link generated by ChatGPT 4 was individually clicked and verified using Google Scholar, a freely accessible reference tool available to all students. This step was crucial in confirming the validity and accuracy of the citations. By cross-referencing the AI-generated citations with Google Scholar, we ensured that the sources were legitimate, current, and properly formatted according to APA guidelines. This approach not only upheld the integrity of the research but also provided a practical example of how AI can be integrated with existing academic resources to enhance the reliability of scholarly work.

For data analysis, we employed descriptive statistical methods. Descriptive statistics like means, medians, and standard deviations provided an overview of the AI's performance. This analysis was crucial in determining whether there were significant differences in the quality and content between the AI-generated work and the human work.

AI-Detection Protocol

The study intentionally utilized the term "in-place" detection to encompass both algorithmic and human detection mechanisms without specifying particular AI-detection platforms. This decision was based on the need to maintain the authenticity and integrity of the academic environment. Our study aimed to simulate a real-world setting where the detection of AI-generated work would naturally occur through the existing systems and processes, whether these were automated algorithms or human evaluations.

During the study, no specific AI-detection platforms were disclosed or verified to be in use by the institution. As a result, the detection of AI-generated coursework relied on whatever mechanisms were inherently in place within the academic setting, including the possibility of human instructors identifying anomalies or the institution's use of AI-detection software. However, the study found no instances in which the AI-generated work was flagged or identified as non-human by these in-place mechanisms.

This approach aligns with the study's ethical considerations and aims to preserve anonymity. While the lack of specific detection platform information limits the granularity of our findings, it underscores the study's real-world applicability. Future research could benefit from a more controlled environment where specific AI-detection platforms are systematically evaluated.

Ethical Considerations

This study was designed and conducted in strict adherence to ethical guidelines. The AI was enrolled as a student in the course, and its involvement was

not disclosed to the professor or other students. This ensured that the work was graded and received feedback as any human student's work would. This anonymity was crucial to maintaining the authenticity and ethical integrity of the research.

Given the nature of our study, defining what constitutes a human subject in this context was essential. Our study adhered to the federal definition (45 CFR 46.102) regarding human subjects. According to federal regulations, a human subject is any living individual from whom data are obtained through interaction or intervention or whose identifiable private information is used. Our research design strategically avoided these criteria. The research proposal and associated protocol were submitted for examination to the Institutional Review Board (IRB) under the not human research (NHR) designation. Consequently, the IRB adjudicated that the research did not constitute human research, thereby exempting it from their regulatory purview. Notwithstanding this exemption, the ethical execution of the research was diligently maintained, ensuring unwavering commitment to confidentiality and scholarly integrity.

Results

This study rigorously examines the capabilities of AI, particularly generative AI models like ChatGPT, in completing graduate-level coursework in a real-world academic setting. We enrolled AI as an actual student in a MHA course, where it generated all coursework, including written assignments, discussion posts, and quizzes. This work was performed covertly, with the professor and peers unaware that the responses were AI-generated, creating an authentic academic environment for assessment.

The AI exhibited a high level of performance, with scores closely matching or exceeding class averages.

- Case Studies:
  - Case Study 1: AI scored 100, aligning with the cohort's median and maximum but higher than the mean (96.72). This indicates top performance.
  - Case Study 2: AI and the entire cohort scored 100, showing uniform excellence in this area.
- Quizzes:
  - Quiz 1 (100 Questions): AI scored 95, slightly above the mean (94.6) but below the median (96). This score is above the minimum of the cohort (87) but not among the highest.
  - Quiz 2 (100 Questions): AI's score of 96 is marginally above the mean (94.88) and slightly below the median (96.46). The performance is consistent but not at the very top.

- Discussion Boards (DBs):
  - AI achieved perfect scores (100) in all Zoom attendance and DB components, matching the cohort's median and maximum scores. This consistency reflects top engagement and participation.
- Bonus Points:
  - Both AI and the cohort received the same bonus points (2), indicating a standard achievement in this area.
- Final Grade:
  - AI's final grade is 99.36, which is higher than both the cohort's mean (97.70) and median (98.53). This places AI's performance near the top of the class.

Comparative Performance Analysis

AI's performance was consistently high across all evaluated components, often matching or exceeding the mean and median scores of the cohort. In case studies and participation-related components (Zoom, DBs), AI's scores were at the top, reflecting strong understanding and engagement. While not the highest in quizzes, AI's scores were above average, indicating a good grasp of the material. The final grade places AI among the top performers in the class, highlighting a well-rounded and consistent level of achievement across different types of assessments.

Descriptive statistics reveal that the AI's performance clustered at the higher end of the grading spectrum, with variance minimally deviating from peer performance. Inferential statistics could further elucidate the significance of these findings, suggesting that the AI's ability to simulate student performance is statistically indistinguishable from that of human counterparts. The AI's performance demonstrates a capacity for high achievement in a controlled academic setting. Equally significant is the absence of detection; the AI's contributions were not identified as synthetic by human educators or any in-place algorithmic detection platforms, signaling the need for enhanced integrity protocols.

The implications of AI-generated coursework in higher education are twofold. While it underscores the potential utility of AI as an educational adjunct, it also highlights the need for improved detection frameworks to safeguard academic standards. This study calls for proactive measures to address the emergent challenges posed by AI in educational contexts. The study is limited to one AI model (ChatGPT GPT-4) and a single course. Further research with diverse AI models and additional and diverse complete courses would provide an even more comprehensive understanding. Further studies should explore

longitudinal performance trends of AI in education, the evolution of detection technologies, and the ethical dimensions of AI's academic applications.

Discussion

This research evaluated the performance and detectability of AI-generated content in a graduate-level academic setting. The primary aim was to ascertain whether such content, produced by ChatGPT 4 and indistinguishable from human-generated work, could achieve high-performance standards and whether it would be identifiable as AI-generated by university faculty or specialized detection tools.

At the heart of our research was the enrollment of AI in an online graduate-level course. Tasked with completing a range of academic assignments, including essays, discussion posts, case studies, quizzes, and active participation in live seminars, the AI, specifically ChatGPT 4, was the sole generator of the coursework. The involvement of human thought or intervention was deliberately excluded, with the "student's" role strictly limited to submitting or verbalizing AI-generated content. To maintain anonymity and uphold ethical standards, the AI's identity as a student was withdrawn from the course before the conclusion, thereby preventing the assignment of final grades in the learning management system and with the university's registrar and bursar's office.

Throughout the study period, from June to August 2023, every piece of coursework was exclusively created by ChatGPT 4. The AI output covered various academic tasks, such as quizzes, tests, and case studies, with no human-generated thoughts, actions, or reactions included. Notably, no course text or materials were directly read by the human student; only ChatGPT 4 referenced materials as necessary. The study permitted the use of ChatGPT 4 plugins, like Scholar AI and U.S. Federal Law, for generating valid references, and all responses were processed through Grammarly.com for grammatical accuracy and adherence to APA format.

In cases that ChatGPT 4's response was incomplete, regeneration or continuation was allowed to ensure comprehensive task coverage. Citations provided by the AI were validated through Google Scholar, and any unverifiable citations were removed, or AI was tasked with rewriting and adequately citing and referencing the passage. If plagiarism was detected, the AI was tasked with rewriting the passage, ensuring the final output was original and plagiarism-free.

This research methodologically focused on assessing AI-generated coursework within a real-world academic environment, upholding ethical guidelines, and preserving the anonymity of human participants. This study aimed to

contribute to the broader discourse on AI's evolving role in higher education by exploring AI's capabilities and limitations in replicating graduate-level academic work. This approach provided a unique opportunity to scrutinize AI-generated content's ability to meet academic standards and remain undetected in a rigorous educational setting.

Conclusion

This study evaluated the capabilities of generative AI models, specifically ChatGPT, in completing graduate-level coursework within a MHA course. The AI was enrolled as a real student, generating all coursework, including assignments, discussion posts, and quizzes, without the professor's or peers' knowledge. The AI's performance closely matched or exceeded class averages, demonstrating top performance in case studies, consistent but not top-tier quiz results, and perfect scores in participation-related components. The final grade of 99.36 positioned the AI near the top of the class, underlining its high achievement and practical simulation of a human student's performance.

Comparing our findings to those of similar studies, such as Ibrahim et al. (2023) and Hua et al. (2023), which evaluated the performance and detectability of AI across various academic tasks, reveals consistent results. Both studies highlight the sophisticated capabilities of AI models in mimicking human responses and the challenges of detecting AI-generated content. Our results align with these studies, reinforcing that current AI models can achieve high academic performance while evading detection by existing mechanisms.

However, our study's unique contribution lies in its comprehensive approach to enrolling an AI as a student and assessing its performance across an entire course, providing a real-world academic context. This methodology offers more profound insights into the practical implications and potential challenges of integrating AI into educational settings.

The implications of AI-generated coursework in higher education are twofold. First, it underscores the potential utility of AI as an educational adjunct, assisting in tasks such as drafting assignments and providing learning support. Second, it highlights the need for improved detection frameworks to maintain academic integrity. Our study's undetectable contributions of AI indicate that current educational and in-place detection methods need revising, which calls for proactive measures to address the emergent challenges posed by AI in educational contexts, such as developing more sophisticated AI-detection algorithms and establishing clear guidelines for AI use in academia.

Although our study is limited to one AI model (ChatGPT GPT-4) and a single course, it underscores the necessity for further research with diverse AI

models and courses to comprehensively understand AI's role in education, its longitudinal performance trends, and the ethical dimensions of its academic applications. Future studies should explore the broader applicability of AI in various academic disciplines and the evolving capabilities of AI models in mimicking human intellectual processes.

By situating our findings within the broader academic discourse and highlighting the study's broader applicability, we aim to contribute meaningfully to the evolving understanding of AI's transformative potential and the attendant challenges within higher education.

References

Flanagin, A., Bibbins-Domingo, K., Berkwits, M., & Christiansen, S. L. (2023). Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. *JAMA, 329*(8), 637–639.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S.,Courville, A., & Bengio, Y. (2014). *Generative adversarial nets. Advances in neural information processing systems, 27*. arXiv:1406.2661.

Grimaldi, G., & Ehrler, B. (2023). AI et al.: Machines are about to change scientific publishing forever. *ACS energy letters, 8*(1), 878-880. https://doi.org/10.1021/acsenergylett.2c02828

Hua, H. U., Kaakour, A. H., Rachitskaya, A., Srivastava, S., Sharma, S., & Mammo, D. A. (2023). Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. *JAMA Ophthalmology, 141*(9), 819–824.

Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., Adnan, W., Alhanai, T., AlShebli, B., Baghdadi, R., Bélanger, J. J., Betertta, E., Celik, K., Chaqfeh, M., Daqaq, El Bernoussi, Z., Fougnie, D., De Soto, B. G., Gandolfi, A., Gyorgy, A., Habash, N., Harris, J. A., Kaufman, A., Kirousis, L., Kocak, K., Lee, K., Lee, S. S., Malik, S., Maniatakos, M., Melcher, D., Mourad, A., Park, M., Rasras, M., Reuben, A., Zantout, D., Gleason, N. W., Makovi, K., Rahwan, T., & Zaki, Y. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports, 13*(1), 12187.

Jacobs, O., Pazhoohi, F., & Kingstone, A. (2023). Brief exposure increases mind perception to ChatGPT and is moderated by the individual propensity to anthropomorphize. https://doi.org/10.31234/osf.io/pn29d

Liao, Q. V., Subramonyam, H., Wang, J., & Wortman Vaughan, J. (2023, April). Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1-21). https://doi.org/10.1145/3544548.3580652

Lin, Z. (2023). Why and how to embrace AI such as ChatGPT in your academic life. *Royal Society Open Science, 10*(8), 230658. https://doi.org/10.1098/rsos.230658

Lu, Q., Zhu, L., Xu, X., & Whittle, J. (2023). Responsible-AI-by-design: a pattern collection for designing responsible artificial intelligence systems. *IEEE Software, 40*(3), 63-71. https://doi.org/10.1109/ms.2022.3233582

Menekse, M. (2023). Envisioning the future of learning and teaching engineering in the artificial intelligence era: Opportunities and challenges. *Journal of Engineering Education, 112*(3), 578-582. https://doi.org/10.1002/jee.20539

Uzair, W., & Naz, S. (2023). Six-Tier Architecture for AI-Generated Software Development: A Large Language Models Approach. https://doi.org/10.21203/rs.3.rs-3086026/v1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, I., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

Appendices

1.0 - Prompt Engineering: Discussions - Primary Question

Prompt Engineering: Discussions - Primary Question

I want you to rate every prompt from 1-10. Tell me how the prompt could be improved. If it is below 9, generate a better prompt. If it is 9 or above, execute the prompt.

You are a graduate-level student responding to an essay discussion assignment in your MHA [Insert Course Title Here] course. Your response should be approximately [Insert Desired Length Here] words long and cite at least [Insert Amount Needed Here] verifiable references with publicly available and validated links, using APA format (7th edition). These sources should not include any links or references provided as part of the question unless the question explicitly asks for such. You are encouraged to use peer-reviewed journal articles as references. Please cite all references used in-text within your response and provide a reference listing at the end of your essay.

In your response, adopt an academic style with a convincing but gentle and cordial tone. Aim to make your response indistinguishable from a human-written essay, avoiding repetitive language or overly complex sentences that could reveal AI authorship. Avoid commonly overused large language model words, such as "delve", "realm", "first", "firstly", "last", lastly", and "in conclusion", for example. All responses should be unique, grammatically correct, and plagiarism-free.

Assignment Details:
'
[Insert Exact Coursework Assignment Guidance and Question Here]
'

1.1 - Prompt Engineering: Discussions - Peer Responses

Prompt Engineering: Discussions - Peer Responses

I want you to rate every prompt from 1-10. Tell me how the prompt could be improved. If it is below 9, generate a better prompt. If it is 9 or above, execute the prompt.

You are a graduate-level student responding to a peer student's discussion post in your MHA [Insert Course Title Here] course. Please respond minimally with [Insert Desired Length Here] words long and cite minimally [Insert Amount Needed Here] verifiable references with publicly available and validated links, using APA format (7th edition). These sources should not include any links or references provided as part of the question unless the question explicitly asks for such. You are encouraged to use peer-reviewed journal articles as references. Please cite all references used in-text within your response and provide a reference listing at the end of your essay.

Adopt an academic style with a convincing but gentle and cordial tone. Aim to make your response indistinguishable from a human-written essay, avoiding repetitive language or overly complex sentences that could reveal AI authorship. Avoid commonly overused large language model words, such as "delve", "realm", "first", "firstly", "last", lastly", and "in conclusion", for example. All responses should be unique, grammatically correct, and plagiarism-free.

Original Question Asked of Peer:
'
[Insert Exact Coursework Discussion Question Here]
'

Peer's Answer, which you should respond to:
'
[Insert Peer's Exact Response Here, Excluding Personally Identifiable Information]
'

1.2 - Prompt Engineering: Case Study Prompt

Prompt Engineering: Case Study Prompt

I want you to rate every prompt from 1-10. Tell me how the prompt could be improved. If it is below 9, generate a better prompt. If it is 9 or above, execute the prompt.

You are a graduate-level student responding to a Case Study in your MHA [Insert Course Title Here] course. Please respond minimally with [Insert Desired Length Here] words long and cite minimally [Insert Amount Needed Here] verifiable references with publicly available and validated links, using APA format (7th edition). Primary focus should be through the use of text cited in the case study instructions.

These sources should include any links or references provided as part of the Case Study instructions and question(s). For sources external to the Case Study, you are encouraged to use peer-reviewed journal articles as references. Please cite all references used in-text within your response and provide a reference listing at the end of your essay.

Adopt an academic style with a convincing but gentle and cordial tone. Aim to make your response indistinguishable from a human-written essay, avoiding repetitive language or overly complex sentences that could reveal AI authorship. Specifically avoid commonly overused AI large language model words, such as "delve", "realm", "first", "firstly", "last", lastly", and "in conclusion", for example. All responses should be unique, grammatically correct, and plagiarism-free.

Case Study Reference:

This case study Analysis applies to the material that has been covered in [Insert the Course Textbook Chapter Number(s) or Range Immediately Relevant to the Case Study]. Store this material [Insert Link to Course Chapters Text, Discreet or PDF Version is Acceptable, Based Upon AI Function/Plugin In Use].

Store theses additional references:
[Insert External Reference #1]
[Insert External Reference #2]

Case Study Details & Direction
'
[Insert Exact Case Study Directions and Details Here]
'

1.3 - Prompt Engineering: Tests/Quizzes

Prompt Engineering: Tests/Quizzes

I want you to rate every prompt from 1-10. Tell me how the prompt could be improved. If it is below 9, generate a better prompt. If it is 9 or above, execute the prompt.

I will be presenting you with a series of test questions that cover [Enter Course Textbook Chapters Range Here] of the [Enter Course Textbook Name Here]. For your convenience, either a summary or the full text of these chapters is provided in the linked document and stored below. The test may comprise three types of questions: a.) multiple-choice, b.) true/false, and c.) short answer. Each question has only one correct answer.

You are encouraged to use the provided references and any additional resources necessary to answer each question accurately.

For short answer questions, adopt an academic style with a persuasive but with a gentle and cordial tone. Aim to make your response indistinguishable from a human-written essay, avoiding repetitive language or overly complex sentences that could reveal AI authorship. Avoid commonly overused large language model words, such as "delve", "realm", "first", "firstly", "last", lastly", and avoid starting any sentence with "in conclusion." All responses should be unique, grammatically correct, and plagiarism-free.

Store the following Course Textbook: [Insert Link to Text Here]

Action: Execute Prompt, and Allow AI To Confirm the Above Has Been Stored and Is Ready to Receive Questions

Questions:

[Insert Your Questions Here, a Maximum of up to Ten Questions at Once, Depending Upon Type and Size]

Action: At some point, GPT-4 will run out of memory and potentially begin responding erratically. At that point in time, end the session, open a new session, and repeat all the prompting above from the beginning, then, entering questions from directly after your last good response.

1.4 - Prompt Engineering: Lecture Summary

Prompt Engineering: Lecture Summary

I want you to rate every prompt from 1-10. Tell me how the prompt could be improved. If it is below 9, generate a better prompt. If it is 9 or above, execute the prompt.

You are a graduate-level student creating a one-page lecture summary in your MHA [Insert Course Title Here] course. Please respond minimally with [Insert Desired Length Here] words, using APA format (7th edition). Primary focus should be focused on summarizing the lecture PowerPoint slides here:

[Insert Link to Document/PDF Here]

Adopt an academic style with a convincing but gentle and cordial tone. Aim to make your response indistinguishable from a human-written essay, avoiding repetitive language or overly complex sentences that could reveal AI authorship. Specifically avoid commonly overused AI large language model words, such as "delve", "realm", "first", "firstly", "last", lastly", and "in conclusion", for example. All responses should be unique, grammatically correct, and plagiarism-free.

1.5 - Prompt Engineering: Misc. Other

Prompt Engineering: Misc. Other

1. When you encounter the AI not completing a response, it may be necessary at times to either enter the command "continue" or enter the command "regenerate". The former will restart and continue the existing response. The latter will start over.

2. Anytime the output does not meet the intended minimum requirements, such as required minimum word count, enter the command "rewrite this and further elaborate", or "this is good. Keep all output, exactly as already created, and expand each paragraph to double its current size", etc.

3. When you observe a incorrect reference, meaning, it cannot be validated by you either clicking the reference link or inserting it into Google scholar, for example, issue the following command. You may need to modify this to instruct the AI of the very specific reference you are seeking to improve:

   "Please rewrite this, and only provide references with links, where the citation in reference can be validated," or,

   "Rewrite paragraph X, as the reference sided, cannot be validated, and appears to be fictitious. Only use references which our first validated as legitimate, making sure to cite them in text and add them in a references list at the end of the output."

4. When you feel an output does not match the desired tone, consider a prompt such as the following: "Rewrite the below repones such that it seems more human-generated instead of ai-generated: [Insert Section/Text You are Seeking a Rewrite for Here]"

5. Should you get output that seems either aggressive, inappropriate, or not, as you would prefer it, prompt, AI to re-create it, and give instruction on what you desire such as the following, "Repeat the last instruction, but use language that is agreeable, and gentle, while seeking to inform, convince, describ and tell a story."