Enhancing Data Science Education with AI: Case Studies on the Integration of ChatGPT in Machine Learning

Jin Soung Yoo
yooj@pfw.edu
Department of Computer Science
Purdue University Fort Wayne
Fort Wayne, IN 46805, U.S.A

Ae-Sook Kim
ae-sook.kim@quinnipiac.edu
Department of Business Analytics and Information Systems
Quinnipiac University
Hamden, CT 06518, USA

Abstract

The growing influence of data science, fueled by technological advancements and an increasing reliance on data-driven decision-making, emphasizes the need for innovative educational strategies in this evolving field. This paper explores the integration of ChatGPT, a large language model developed by OpenAI, into data science education, with a focus on a graduate-level machine learning course. Through two case studies, we evaluate ChatGPT's potential to enhance test question generation and programming code solutions. The first case study, based on student feedback, compares perceptions of ChatGPT-generated test questions with those created by the instructor, assessing various aspects such as relevance to the course material, contribution to learning, difficulty level, challenge, engagement, and introduction of new perspectives. The second case study examines ChatGPT's ability to generate functional programming code for a machine learning and visualization problem, highlighting the importance of human review and debugging. These findings offer valuable insights into both the benefits and limitations of integrating AI tools like ChatGPT into data science education, guiding educators in their efforts to effectively enhance the learning experience.

Keywords: Data Science Education, ChatGPT, Large Language Models, Machine Learning, AI-assisted Student Assessment, Pedagogical Innovation

1. INTRODUCTION

The rapid growth of data has elevated the importance of data science. This interdisciplinary field combines statistical analysis, computational techniques, and domain-specific knowledge to extract insights from data (Cao, 2017). Despite advancements in technology, the field still encounters notable challenges. A shortage of skilled professionals limits the ability of organizations to fully leverage data science techniques. Furthermore, the dynamic nature of data science, particularly in areas such as machine learning, requires continuous evolution in educational curricula. Instructors face the ongoing challenge of keeping their courses up to date with the latest research, methods, and industry practices, all while navigating the fast-paced development of machine learning tools and algorithms. Assessing students' proficiency in this

ever-evolving field presents additional complexities.

Artificial intelligence (AI), defined as the simulation of human intelligence by machines, plays an increasingly critical role in fields like data science and machine learning. A notable development in AI is the Chat Generative Pre-Trained Transformer (Chat-GPT), a state-of-the-art large language model (LLM) developed by OpenAI (OpenAI, n.d.-a). ChatGPT utilizes deep neural network architecture for natural language processing, enabling it to generate coherent, contextually relevant text. While initially applied in areas such as chatbots and language translation, ChatGPT is now recognized for its potential to transform educational practices.

This paper explores the integration of ChatGPT into data science education, specifically focusing on its role in a graduate-level machine learning course. Through two case studies, we evaluate ChatGPT's ability to enhance both test question generation and programming code solutions. The first case study examines student feedback, comparing perceptions of ChatGPT-generated test questions with those crafted by the instructor, evaluating aspects such as relevance, contribution to learning, difficulty, challenge, engagement, and introduction to new perspectives. The second case study investigates ChatGPT's ability to generate functional programming code for machine learning tasks, underscoring the importance of human oversight in reviewing and debugging AI-generated solutions. These findings provide insights into both the opportunities and challenges of integrating AI tools like ChatGPT into machine learning education.

This paper is structured as follows: Section 2 reviews AI and ChatGPT's role in data science education, with applications in statistics, exploratory data analysis, programming, and machine learning. Section 3 provides a literature review. Section 4 presents two case studies on the use of ChatGPT in machine learning education and their outcomes. Finally, Section 5 concludes with key findings, contributions, limitations, and directions for future research.

2. AI AND CHATGPT IN DATA SCIENCE EDUCATION

AI and ChatGPT

Artificial Intelligence (AI) involves the replication of human intelligence processes through machines, especially computer systems (Russell & Norvig, 2020). These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions), and self-correction. AI enables machines to perform tasks that typically require human intelligence, such as decision-making, problem-solving, and natural language understanding.

A large language model (LLM) is an advanced AI system designed to understand and generate human language based on extensive datasets and sophisticated algorithms (Vaswani et al., 2017). LLMs, such as OpenAI's GPT, utilize deep learning techniques based on the Transformer model. As a generative model, an LLM can produce new text based on the input it receives, making it highly versatile in various applications. The model is pre-trained on vast amounts of text data, allowing it to generate coherent and contextually appropriate responses. ChatGPT, a prominent example of an LLM, leverages these capabilities in numerous applications, ranging from customer service chatbots to language translation, demonstrating significant potential in enhancing human-computer interaction. OpenAI has released several versions of ChatGPT, such as GPT-2, GPT-3.5, GPT-4, and GPT-4o, each differing in size, number of parameters, and the breadth of languages included in their pre-training data.

Data Science Education

Data science is an interdisciplinary field focused on extracting meaningful insights and knowledge from both structured and unstructured data (Cao, 2017). It involves various processes, including data collection, cleaning, transformation, analysis, visualization, and interpretation. Data scientists employ advanced tools and methodologies from machine learning, artificial intelligence, and big data technologies to identify patterns, make predictions, and support data-informed decision-making across diverse industries and scientific fields, including healthcare, finance, marketing, and environmental science. The ultimate goal of data science is to leverage data to solve complex problems and create value.

A data science curriculum typically begins with foundational courses in statistics, probability, and linear algebra, which provide essential mathematical groundwork for analyzing data. Students are also introduced to key programming languages, such as Python and R, which are integral to data manipulation and analysis. Core courses includes machine learning, data mining, data visualization, and database management, equipping students with the skills necessary to

manage, analyze, and effectively present data. Advanced topics often cover big data technologies and natural language processing (NLP), further broadening students' skill sets for handling large datasets and understanding human language data.

Within the data science curriculum, a machine learning course is pivotal. It provides students with a deep understanding of both the theoretical foundations and practical applications of machine learning algorithms. Key topics include supervised learning (e.g., decision trees, logistic regression, random forest, neural networks), unsupervised learning (e.g., k-means clustering, principal component analysis), and reinforcement learning (e.g., Q-learning). Students may also be introduced to more advanced concepts such as deep neural networks and transformers. The course emphasizes hands-on projects, allowing students to apply these techniques to real-world datasets, develop models, and evaluate their performance.

ChatGPT In Data Science Education

In educational settings, ChatGPT offers significant potential to enhance both teaching and learning by providing personalized, interactive experiences and creating tailored questions for continuous student assessment (David & Ansah, 2023; Lo, 2023). As an intelligent tutor, ChatGPT delivers instant feedback, detailed explanations, and clarifications in response to student queries, thereby helping students grasp complex concepts more effectively. While many students may be familiar with ChatGPT, offering specific guidance on how to leverage its full potential can further enhance its role as a learning tool. Additionally, ChatGPT supports educators by generating customized lesson plans and creating a variety of educational materials, including quizzes, summaries, and study guides (Modal et al., 2023).

In data science applications, ChatGPT automates key aspects of the workflow, such as data cleaning, preprocessing, model training, and result interpretation (Hassani & Silva, 2023). Within data science education, ChatGPT's potential can be observed across several key areas, which are discussed in the following subsections.

Statistics and Exploratory Data Analysis

ChatGPT acts as a valuable resource for explaining complex statistical concepts and methodologies by offering clear explanations and illustrative examples. Particularly in exploratory data analysis (EDA) (Komorowski et al., 2016; Tukey, 1977), ChatGPT aids students in preparing datasets, performing preprocessing tasks, and generating visualizations that help uncover patterns, trends, and correlations within the data. It can assist with generating descriptive statistics and producing visualizations such as histograms, scatter plots, and correlation heatmaps (OpenAI, n.d.-b), further helping students interpret these outputs and gain deeper insights.

Despite its usefulness, ChatGPT has limitations in handling statistical analysis. As noted by Ordak (2024), it can make calculation errors and may struggle with advanced statistical techniques, underscoring the need to verify the assumptions of statistical tests and confirm accuracy. Liu et al. (2023) also highlight that ChatGPT has difficulty processing large datasets and often requires external tools and precise instructions to manage complex data effectively.

Programming for Data Analytics

ChatGPT can be a valuable resource for both students and instructors in learning and teaching programming languages like Python and R, essential for data manipulation and analysis in data science. For students, ChatGPT provides instant, detailed explanations and code snippets to help them understand key programming concepts, syntax, and best practices (Biswas, 2023). It also assists in efficiently identifying and correcting errors in code (Surameery & Shakor, 2023). For instructors, ChatGPT can streamline the development of teaching materials, including coding examples and homework assignments, by generating content aligned with course objectives.

However, despite its advantages, ChatGPT has several limitations in generating program code. One major limitation is that ChatGPT does not compile or execute the generated code, meaning it cannot verify the correctness or functionality through actual execution (Liu et al., 2024). Instead, it provides outputs based on its training data and learned patterns, which can sometimes result in incorrect or non-functional code. Additionally, ChatGPT may not account for specific nuances and dependencies of a programming environment or project, potentially causing integration issues with existing systems.

Enhancing Machine Learning Education

ChatGPT holds significant potential for enhancing machine learning education for both students and instructors. For students, it serves as a responsive educational assistant, simplifying complex machine learning concepts such as gradient descent and regularization techniques.

ChatGPT also aids in implementing machine learning models by suggesting relevant libraries, algorithms, and approaches. Additionally, it helps students brainstorm project ideas and methodologies, enabling them to experiment more easily with machine learning concepts.

For instructors, ChatGPT can assist with various course-related tasks, enabling them to focus more on higher-level instruction. It can generate machine learning-focused assignments and quizzes tailored to the course content, covering topics like classification, clustering, and deep learning. By creating problem sets, coding challenges, and case studies, ChatGPT supports instructors in delivering a diverse range of assessments.

Despite its advantages, ChatGPT has certain limitations in machine learning education. While it is adept at automating routine coding tasks and providing theoretical explanations, it may face difficulties with advanced topics such as hyperparameter optimization or navigating trade-offs in model performance metrics (e.g., balancing precision and recall or addressing overfitting issues). Instructors need to carefully review ChatGPT-generated content to ensure it maintains the required depth and rigor for machine learning coursework.

3. LITERATURE REVIEW

Recent advances in AI have significantly influenced educational practices (Chen, 2020). Researchers have explored ChatGPT's applications in various fields, such as mathematics, economics and finance, medical education, computer programming, and the generation of questions and exercises.

Wardat et al. (2023) examined ChatGPT's role in teaching mathematics, particularly geometry, noting that while it improves instruction in basic concepts, it lacks depth in more advanced topics. Geerling et al. (2023) emphasized ChatGPT's strong performance on the Test of Understanding in College Economics (TUCE), suggesting that AI tools like ChatGPT could prompt educators to reconsider traditional assessment methods by incorporating more experiential and AI-enhanced instruction. Lee (2024) explored ChatGPT's potential in medical education, suggesting its use as a virtual teaching assistant to enhance student engagement, though further research is necessary to address ethical concerns and ensure its integration into medical curricula. Cheung et al. (2023) evaluated the quality of multiple-choice questions generated by ChatGPT for medical graduate examinations and compared them to those crafted by university professors.

In computer programming, Biswas (2023) found that ChatGPT performs tasks such as code completion, error correction, and code optimization, while also providing explanations to help developers diagnose and resolve issues. Surameery & Shakor (2023) highlighted ChatGPT's debugging capabilities, noting its ability to predict and explain bugs, but emphasized the need for additional tools to validate these predictions. Sarsa et al. (2022) explored the potential of using large language models to generate programming exercises for educational purposes.

In data science, Hassani and Silva (2023) highlighted ChatGPT's ability to automate workflows such as data preprocessing, model training, and result interpretation, while addressing concerns around bias and output interpretability. Zheng (2023) examined the application of ChatGPT in data science education, highlighting its unique contributions and challenges compared to other academic fields. Ellis and Slade (2023) explored ChatGPT's potential in statistics and data science education, advocating for responsible use in developing course materials while acknowledging concerns about misuse.

While these studies underscore ChatGPT's broad applicability in education, none have explicitly examined its integration in machine learning education through practical case studies. Our research addresses this gap by evaluating ChatGPT's role on both test question generation and programming solution validation within a graduate-level machine learning course, offering practical insights into its potential to enhance machine learning education.

4. CASE STUDIES IN MACHINE LEARNING EDUCATION

This study examines the integration of ChatGPT in a graduate-level machine learning course through two case studies: (1) a comparative analysis of ChatGPT-generated and instructor-created questions in assessments, and (2) an evaluation of ChatGPT-generated programming solutions for usability.

The course, offered every other semester, covers topics such as supervised, unsupervised, and reinforcement learning. It balances theory with practical applications through quizzes, exams, assignments, hands-on practice, and a term

ISSN: 2473-4901
v10 n6234

project. Students are expected to have a strong background in statistics, probability, linear algebra, and proficiency in programming languages like Python and R.

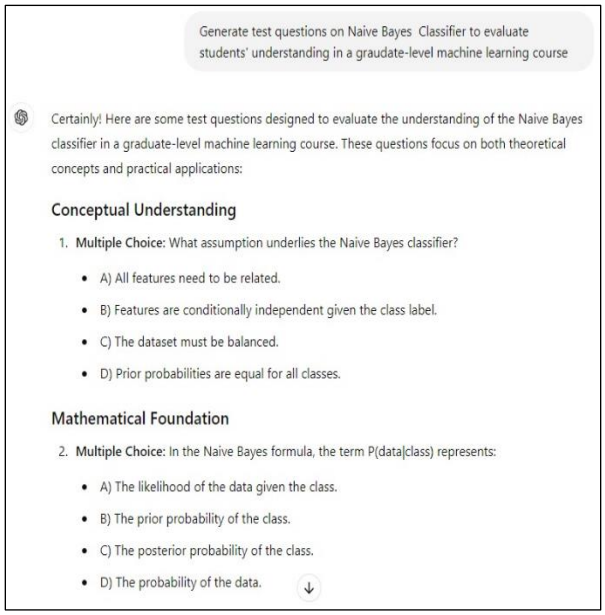
We utilized GPT-4, the latest version of ChatGPT available during the study, to generate test questions and programming code, which were assessed through both student feedback and expert validation.

Case 1: Comparative Analysis of Test Assessment

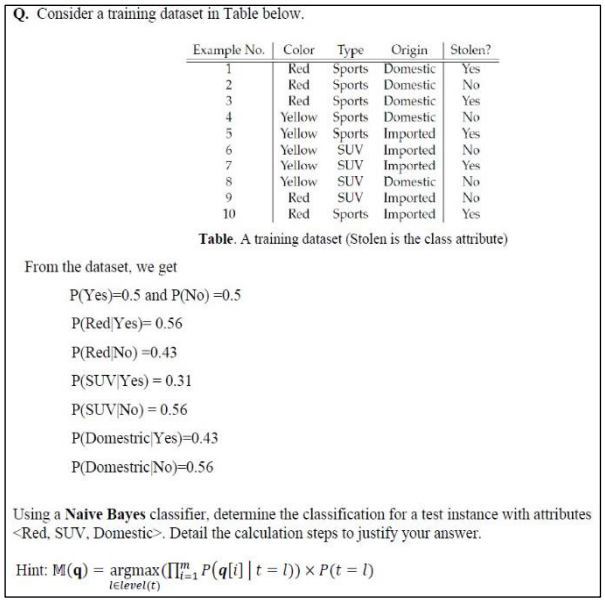
To compare ChatGPT-generated questions with instructor-created ones used in quizzes and exams, we conducted an assessment based on student feedback.

Methodology and Procedure

In this analysis, the first set of questions was generated by ChatGPT, utilizing its advanced language capabilities. These questions covered selected course topics such as decision-tree induction, ensemble learning, instance-based learning, Bayesian learning, model evaluation, and neural networks. We only provided the specific topic name, without additional context or course materials. No ablation studies were conducted to determine the best input or approach for generating test questions. Instead, we employed a zero-shot learning approach (Xian 2019), where the model performs the task without prior examples, relying solely on its pre-existing knowledge. Figure 1 (a) shows examples of test questions generated by ChatGPT.



(a) By GPT



(b) By Instructor

Figure 1 Example of Test Questions in Naïve Bayes Classifier

The second set consisted of questions carefully crafted by the course instructor, leveraging years of expertise in the field of machine learning. Figure 1 (b) illustrates an example of test questions developed by the instructor on the same topics. Both sets of questions were used in pop-up quizzes and examinations throughout the semester.

To evaluate the effectiveness of the questions, a detailed student feedback survey was designed. The survey included questions on (i) relevance to the course material, (ii) contribution to student learning, (iii) difficulty levels, (iv) challenge, (v) creativity and engagement, and (vi) introduction of new perspectives, along with additional questions on (vii) effectiveness in assessment and (viii) preference for future tests.

The complete survey, along with response options, is available in Appendix A. It includes a combination of Likert-scale and categorical questions. For example, "How well do you think ChatGPT-generated test questions covered the course material?" used a Likert scale (e.g., "extremely relevant" to "not relevant at all"), while "Did the ChatGPT-generated test questions introduce new perspectives?" used categorical options ("yes," "no," "somewhat," etc.). Responses were carefully counted, and percentages were calculated for accurate data.
The survey was administered at the end of the semester. Participation was voluntary, with anonymity and confidentiality assured. Students

ISSN: 2473-4901
v10 n6234

were informed that the survey results would be used for research purposes to improve future teaching practices. The survey was distributed online via the Qualtrics survey tool, with 32 out of 54 enrolled students responding, yielding a 59% response rate.

Results: Feedback From Students

Relevance to Course Material: As shown in Figure 2(a), 79% of students rated ChatGPT-generated questions as "Very relevant" or "Extremely relevant" to the course material, indicating strong alignment with the machine learning content.

Contribution to Learning: In Figure 2(b), 50% of students found instructor-created questions more helpful in understanding key concepts, while 22% felt both types were equally beneficial. Only 9% favored ChatGPT-generated questions, suggesting a preference for instructor-designed questions in enhancing comprehension.

Difficulty Level: According to Figure 2(c), 28% of students found instructor-created questions "Too difficult," while 66% rated them "Challenging but manageable." For ChatGPT-generated questions, 51% rated them "Just right," indicating they were seen as more balanced in difficulty compared to instructor-created ones.

Challenge: Figure 2(d) shows 69% of students found instructor-created questions more challenging, while only 6% preferred the ChatGPT-generated ones, further reinforcing the perception that instructor-created questions are more rigorous.

Creativity and Engagement: As indicated in Figure 2(e), 63% found instructor-created questions more engaging and thought-provoking, while 19% found both equally engaging. A minority, 13%, favored ChatGPT-generated questions, showing some value in the AI-generated content.

Introduction of New Perspectives: In Figure 2(f), 38% of students noted that ChatGPT-generated questions introduced new perspectives, though 31% felt they did not offer anything new, highlighting mixed views on the novelty of AI-generated content.

Effectiveness in Assessment: Figure 2(g) shows 35% of students rated ChatGPT-generated questions as effective as instructor-created ones. However, 26% found them somewhat less effective, while another 26% found them somewhat more effective, reflecting varying perceptions of their utility in assessments.

Preferences for Future Tests: According to Figure 2(h), 56% of students preferred instructor-created questions for future assessments, while 34% favored a mix of both ChatGPT and instructor-created questions, indicating interest in combining AI and human-generated content for future exams.

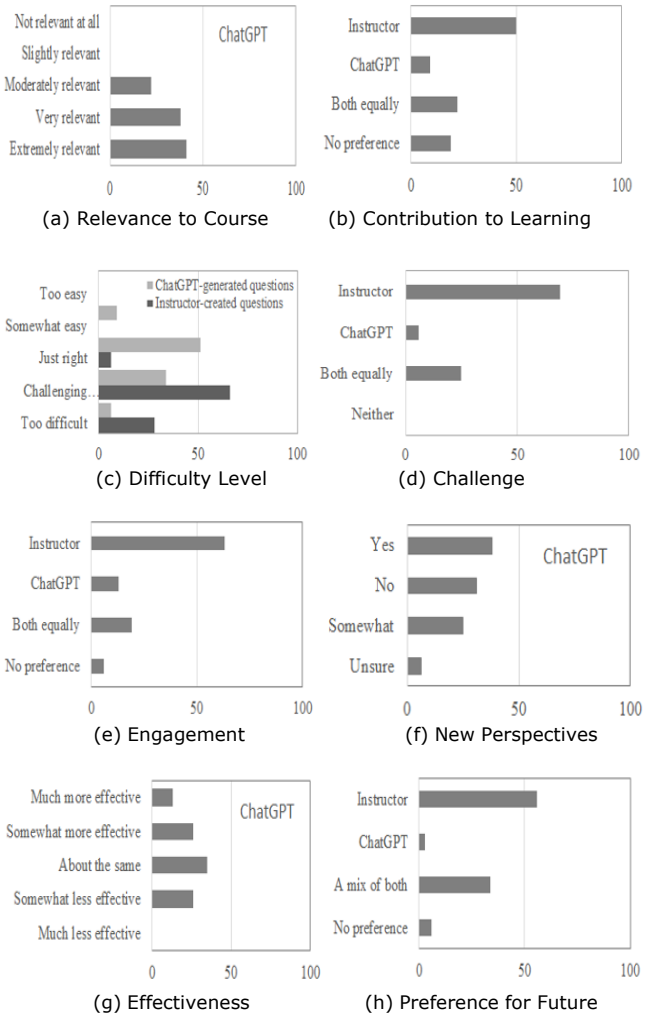


Figure 2: Student Survey Results

Discussion

The feedback highlights key findings regarding the comparison between ChatGPT-generated and instructor-created questions. While 79% of students found ChatGPT-generated questions highly relevant, 50% felt instructor-created questions better supported their learning, indicating a preference for the depth of human-generated questions. Instructor-created questions were also seen as more challenging and

ISSN: 2473-4901
v10 n6234

engaging, with 69% of students rating them as more difficult. However, 38% of students appreciated ChatGPT-generated questions for introducing new perspectives. Notably, 34% of students preferred a hybrid approach, suggesting potential in blending both AI and instructor-created questions for future assessments.

Case 2: Validation of Programming Code Solutions

To evaluate ChatGPT's effectiveness in generating accurate programming solutions for machine learning, we conducted a validation study focused on the well-known k-Nearest Neighbor classifier.

Methodology and Procedure

k-Nearest Neighbors (kNN) is an instance-based, lazy-learner algorithm that classifies a query instance by identifying its nearest neighbors in the dataset (Kelleher et al., 2020). Unlike model-based methods, kNN does not create a model during the training phase, instead making predictions based on the stored dataset.

A simple kNN classifier can be implemented using open-source libraries like scikit-learn (Scikit-learn developers, n.d.) to assess model performance through metrics such as accuracy, precision, and recall. However, this approach can obscure the algorithm's internal mechanics, making it challenging for students to understand how the classifier works.

To bridge this gap, we employed Voronoi diagrams as a visual tool to help students grasp the classification boundaries of the kNN algorithm. By visualizing each region associated with a data point and its nearest neighbors, students can better understand how kNN determines the class of a query point. This visualization clarifies the link between theory and practice in instance-based learning.

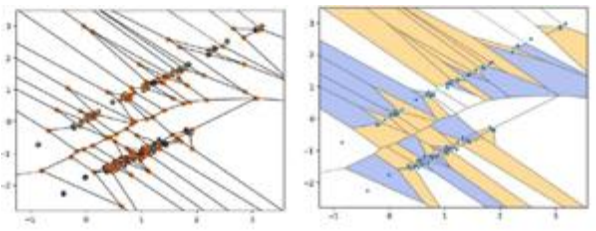
For this case study, we used the VoronoiKNN method, which combines Voronoi tessellation with kNN. Voronoi diagrams divide the data space into cells, with each cell representing a region where any query point is closer to the central point of that cell than to any other point (Okabe, 2000). When applied to kNN, Voronoi diagrams can illustrate how each region, defined by a training point, is associated with its class label.

We tasked ChatGPT with generating Python code for the VoronoiKNN classifier, using a toy dataset designed for binary classification with two classes, "blue" and "orange." We applied a zero-shot learning approach, meaning that ChatGPT was asked to generate the solution without having been specifically trained on examples for this particular task. Instead, it relied on its pre-existing knowledge, learned from a wide variety of data, to infer how to implement the algorithm. The generated code was then run on Google Colab, an online platform for executing Python code in Jupyter notebooks (Google Colab developers, n.d.).

Results

Figure 3 illustrates the Voronoi diagrams generated the toy dataset using ChatGPT's Python code. Initially, the results were not fully correct. In Figure 3(a), the Voronoi diagram correctly represents regions for the blue data points, but not for the orange ones, which were incorrectly placed along the borders.

To address this issue, we refined our prompt and provided ChatGPT with additional instructions. Specifically, we asked it to perform Voronoi partitioning without initially considering class labels and to later color each region based on the class label of the central point. This refinement was based on the rationale that performing the partitioning first would allow for a more accurate representation of the class boundaries. The revised output, shown in Figure 3(b), reflects this improvement, although some regions still remained uncolored.



(a) Initial (b) Revised

Figure 3: Initial and Revised Program Results Generated by ChatGPT

Discussion

While ChatGPT demonstrated the ability to generate functional code, its outputs were not flawless. Users must review and debug the code to ensure it follows best practices and meets the desired outcome. This study highlights ChatGPT's potential as a useful tool in programming tasks, but underscores the needs of human oversight, especially when handling complex algorithms or intricate machine learning tasks.

5. CONCLUSION

This paper examined the integration of ChatGPT, a large language model, into data science

education, specifically within a graduate-level machine learning course. Through two case studies, we evaluated ChatGPT's usability and effectiveness in generating test questions and programming solutions based on student perceptions and manual validation.

In the first case study, student feedback was collected on various aspects of the test questions. While instructor-created questions were generally preferred for their challenge and perceived effectiveness in enhancing learning, ChatGPT-generated questions provided valuable new insights. Students expressed significant interest in blending both types of questions, suggesting that a hybrid approach could leverage human expertise and AI innovation to create a more dynamic learning environment.

The second case study explored ChatGPT's ability to generate programming solutions for an instance-based classification problem. While ChatGPT produced functional code, human intervention was essential for reviewing and debugging to ensure accuracy and relevance, especially for more complex tasks. This highlights the limitations of relying solely on AI tools in educational setting and emphasizes the need for critical human oversight to ensure AI-generated solutions meet educational goal and quality standards.

Contributions

This study contributes to the growing body of research on AI in education. It provides an overview of the potential benefits and limitations of using ChatGPT in data science education. Our comparative assessment of human- and AI-generated test questions, along with student feedback and the validation of AI-generated code, offers practical insights for educators looking to incorporate AI tools like ChatGPT into their curricula.

Limitations

This study has several limitations. First, our experiments and student feedback were collected from a single graduate-level course, limiting the generalizability of the results to other courses or institutions. Additionally, the voluntary nature of the survey may have introduced response biases, and individual differences—such as prior knowledge, learning styles, and preferences—may have influenced students' perceptions of the questions' relevance and difficulty. This variability could have impacted the survey results.

Although we compared ChatGPT- and instructor-generated question sets, the feedback primarily reflected students' perceptions rather than actual learning outcomes. The lack of pre- and post-test evaluations limited our ability to directly compare learning gains between the two question types. Furthermore, the survey was not designed to capture detailed qualitative feedback, restricting our ability to explore nuances behind students' multiple-choice responses.

In Case Study 1, while ChatGPT could generate questions comparable to those created by the instructor, it often required multiple refined prompts. The quality of the output was highly dependent on the specificity of the input prompts, with variations in phrasing leading to different results. Moreover, ChatGPT is limited to text generation and cannot autonomously create charts or graphs, which are often necessary for assessments. Although GPT-4 can interpret some figure images, its performance in this area remains limited.

In Case Study 2, while ChatGPT successfully generated functional programming code for instance-based classification, the outcomes were highly dependent on the quality of the initial prompts and the user's level of coding expertise. The AI often required specific, detailed prompts to produce accurate and contextually relevant code, which suggests that without adequate guidance, the solutions might not meet expectations. Additionally, the effectiveness of the generated code varied based on the complexity of the problem, highlighting the need for human oversight to refine, debug, and optimize the AI-generated solutions for practical use.

Future Research

Future research should explore the broader potential of ChatGPT and other AI tools in data science and machine learning education. Further studies could focus on improving the integration of AI tools with human input, particularly in generating complex assessments, programming solutions, and adaptive learning materials. Longitudinal research on the effects of combining AI-generated and instructor-created materials on student performance, as well as comparative analyses across different academic levels and institutions, could provide deeper insights into the optimal use of AI in modern education.

6. REFERENCES

Baidoo-Anu, D. & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching

and learning. *Journal of AI*, 7(1): 52-62. https://doi.org/10.61969/jai.1337500

Biswas, S. (2023). Role of ChatGPT in computer programming. *Mesopotamian Journal of Computer Science*, 9–15. https://doi.org/10.58496/MJCSC/2023/002

Cao, L. (2017). Data science: A comprehensive overview. *ACM Computing Surveys*, 50(3):1–42. https://doi.org/10.1145/3076253

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., Wong, R., & Co, M. T.-H. (2023). ChatGPT versus human in generating medical graduate exam multiple-choice questions—A multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). *PLoS One*. https://doi.org/10.1371/journal.pone.0290691

David, B. & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning, *Journal of AI*, 7(1), 52–62. https://doi.org/10.61969/jai.1337500

Ellis, A. R. & Slade , E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. https://doi.org/10.1080/26939169.2023.2223609

Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has aced the test of understanding in college economics: Now what? *The American Economist, 68(2), 233-45*. https://doi.org/10.1177/05694345231169654

Google Colab Developers. (n.d.). *Google Colab: Free Jupyter notebook environment*. Retrieved August 10, 2024 from https://colab.research.google.com/

Hassani, H. & Silva. E. S. (2023). The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the field. *Big Data and Cognitive Computing*, 7(2), 62. https://doi.org/10.3390/bdcc7020062

Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020*). Fundamentals of machine learning for predictive data analytics* (2nd ed.). The MIT Press.

Komorowsk, M., Marshall, D. C., Salciccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. In *Secondary Analysis of Electronic Health Records* (pp. 185-203).

Lee, H. (2024). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*. 17: 926–931. https://doi.org/10.1002/ase.2270

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Chen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2), 100017. doi:https://doi.org/10.1016/j.metrad.2023.100017

Liu, Y., Le-Cong, T., Widyasari, R., Tantithamthavorn, C., Li, L., Le, X.-B. D., & Lo, D. (2024). Refining ChatGPT-generated code: Characterizing and mitigating code quality issues. *ACM Transactions on Software Engineering and Methodology*, 33(5), 1-26. https://doi.org/10.1145/3643674

Lo, C. K. (2023), What is the impact of ChatGPT on education? A rapid review of the literature, *Education Sciences*, 13(4), 410. https://doi.org/10.3390/educsci13040410

Mondal, H., Marndi, G., Behera, J. K., & Mondal, S. (2023). ChatGPT for teachers: Practical examples for utilizing artificial intelligence for educational purposes. *Indian Journal of Vascular and Endovascular Surgery*, 10(3), 200-205. https://doi.org/10.4103/ijves.ijves_37_23

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and Applications of Voronoi Diagrams* (2nd ed.). Wiley.

OpenAI. (n.d.-a). ChatGPT: Optimizing Language Models for Dialogue. Retrieved August 10, 2024 from https://openai.com/

OpenAI. (n.d.-b). Data analysis with ChatGPT. Retrieved August 10, 2024 from https://help.openai.com/en/articles/8437071-data-analysis-with-chatgpt

Ordak, M. (2024). Using ChatGPT in Statistical Analysis: Recommendations for JACC

ISSN: 2473-4901
v10 n6234

Journals. *JACC: Advances*, 3(2). https://doi.org/10.1016/j.jacadv.2023.100776

Russell, S. & Norvig, P. (2020). *Intelligence: A Modern Approach* (4th). Pearson.

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. *Association for Computing Machinery*. https://doi.org/10.1145/3501385.3543957

Scikit-learn developers. (n.d.). Scikit-learn: Machine learning in Python. Retrieved August 10, 2024 from https://scikit-learn.org/stable/

Surameery, N. M. S. & Shakor, M. Y. (2023). Use ChatGpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering*, 3(1), 17–22. https://doi.org/10.55529/ijitc.31.17.22

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Proceedings of the International Conference on Neural Information Processing Systems*, 6000-6010. https://doi.org/10.48550/arXiv.1706.03762

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah , A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7): em2286. https://doi.org/10.29333/ejmste/13272

Xian, Y., Lampert, C. H., Schiele, B., & Akata, Z. (2019). Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9), 2251-2265. https://doi.org/10.1109/TPAMI.2018.2857768

Zheng, Y. (2023). ChatGPT for teaching and learning: An experience from data science education. *Proceedings of the Conference on Information Technology Education.* https://doi.org/10.1145/3585059.3611431

APPENDIX A
STUDENT FEEDBACK SURVEY

Relevance to course material:
Q1: Compared to the instructor-created test questions, how well do you think ChatGPT-generated test questions covered the course material?
- □ *Not relevant at all*
- □ *Slightly relevant*
- □ *Moderately relevant*
- □ *Very relevant*
- □ *Extremely relevant*

Contribution to learning:
Q2: Among the ChatGPT-generated test questions and the instructor-created test questions, which type of questions helped you better understand the key concepts of the course?
- □ *Instructor*
- □ *ChatGPT*
- □ *Both equally*
- □ *No preference*

Difficulty level:
Q3: How would you rate the difficulty level of the ChatGPT-generated test questions?
- □ *Too easy*
- □ *Somewhat easy*
- □ *Just right*
- □ *Challenging but manageable*
- □ *Too difficult*

Q4: How would you rate the overall difficulty of the instructor-created test questions?
- □ *Too easy*
- □ *Somewhat easy*
- □ *Just right*
- □ *Challenging but manageable*
- □ *Too difficult*

Challenge:
Q5: Among the ChatGPT-generated test questions and the instructor-created test questions, which set of questions did you find more challenging?
- □ *Instructor*
- □ *ChatGPT*
- □ *Both equally*
- □ *Neither*

Creative and engagement:
Q6: Among the ChatGPT-generated test questions and the instructor-created test questions, which set of questions did you find more creative and engaging?
- □ *Instructor*
- □ *ChatGPT*
- □ *Both equally*
- □ *Neither*

Introduction of new perspectives:
Q7: Did the ChatGPT-generated test questions introduce new perspectives or ways of thinking about the subject matter compared to the instructor-created questions?
- □ *Yes*
- □ *No*
- □ *Somewhat*
- □ *Unsure*

Effectiveness in assessment:
Q8: In your opinion, how effective are the ChatGPT-generated test questions in assessing your knowledge and skills compared to instructor-created or verified questions?

- □ *Much more effective*
- □ *Somewhat more effective*
- □ *About the same*
- □ *Somewhat less effective*
- □ *Much less effective*

Preferences for future tests:
Q9: Among the ChatGPT-generated test questions and the instructor-created test questions, which type of questions would you prefer for future tests and learning?

- □ *Instructor*
- □ *ChatGPT*
- □ *A mix of both*
- □ *No preference*