

Work-in-Progress: Course-based Undergraduate Research Experience (CURE) with Generative AI in a Computer Science Course

Paula Lauren

*Department of Mathematics and Computer Science
Lawrence Technological University
Southfield, Michigan, USA
plauren@ltu.edu*

Abstract—This work-in-progress innovative practice paper describes a novel integration of Generative AI with Course-based Undergraduate Research Experiences (CUREs). CUREs integrate research activities into the curriculum, allowing all students in a course to participate in inquiry-based research projects. Generative Artificial Intelligence (AI) applications are advanced AI designed to generate human-like responses by processing natural language inputs. These applications leverage machine learning models to produce outputs that can assist users in a variety of tasks from writing to coding. The integration of Generative AI with CURE had been adopted in a text-based machine learning course during the Fall 2023 semester. A comparative analysis had been conducted on student survey responses from Fall 2022 and Fall 2023 to evaluate the effectiveness of Generative AI in a CURE integrated course. Descriptive statistics and statistical tests were conducted to assess differences in student perceptions between the two semesters. Although the differences were not statistically significant, the results indicate a promising trend towards improved student perceptions of both the overall course effectiveness and the benefits of Generative AI in enhancing various aspects of the research process, especially the literature review.

Index Terms—Course-based Undergraduate Research Experiences (CUREs), Generative AI, Artificial Intelligence in Education (AIED)

I. INTRODUCTION

Course-based Undergraduate Research Experiences (CUREs) are considered an inquiry-based pedagogical approach that integrates research into the undergraduate curriculum. Unlike traditional undergraduate research experiences that typically involve working on a specific research project under the guidance of a faculty mentor, CUREs are designed to engage all students in authentic research as part of their regular coursework [1]. The benefits of CUREs have been shown to enhance learning, develop skills, increase engagement, prepare for graduate studies, and uphold equity in education [2]–[4].

Generative Artificial Intelligence (AI) applications are designed to provide customized responses to user prompts. These applications are based on the Generative Pre-trained Transformer (GPT), a family of neural network models that utilize the transformer architecture [5]. GPT powers Generative AI

applications like ChatGPT¹, a widely used tool for generating natural language responses to user inputs [6]. The use of Generative AI tools presents both challenges and opportunities for students and educators, driving necessary changes in teaching and assessment, particularly in higher education [7]–[13]. As Generative AI applications like ChatGPT become more pervasive, they offer educators an opportunity to redefine the content of individual courses and entire curricula. A shift toward personalized adaptive learning could enable teaching strategies based on real-time monitoring [14].

This work-in-progress paper builds on ideas from previous work using Course-based Undergraduate Research Experiences (CUREs) in a text-based machine learning course and a preliminary study on the use of Generative AI in a computer science course [15], [16]. This study presents a novel approach by incorporating Generative AI into an existing CURE course in computer science, specifically focusing on a text-based machine learning course.

II. COURSE OVERVIEW

A. Course Description

The CURE-based course selected for this study is an upper undergraduate computer science course entitled *Text Mining and Analytics*, a required course in the Bachelor of Science in Computer Science program for the AI concentration at Lawrence Technological University. The course description entails gaining a theoretical and practical application of text-based machine learning algorithms. Embedded in the course is a course-based research experience over four weeks, which involves a research overview, research problem and literature review, research design and methods, and research paper construction.

B. Course Content and Structure

Core learning objectives for the course are as follows:

- 1) Become conversant in the terminology of text mining and analytics.

¹<https://openai.com/blog/chatgpt>

- 2) Learn and work with state-of-the-art tools for doing effective text mining and analytics including various statistical and machine learning methods for gaining insight from text.
- 3) Understand the various sources for obtaining data for text mining and analysis.
- 4) Work with various API's and state-of-the-art methods for extracting and analyzing text from social media applications.
- 5) Develop an understanding in conducting research, project research development(with code repository), and research paper creation.
- 6) Identify potential ethical, political, legal, or social issues.
- 7) Contribute constructive feedback in the form of peer review for up to three classmate papers.
- 8) Leverage Generative AI to assist in the research development process.

C. Integrating Generative AI

The incorporation of Generative AI complimented the course CURE project in Fall 2023 in various ways, which included using ChatGPT to assist with the literature review process, coding and technical aspects of the project, and obtaining clarification on concepts related to the research project. Based on output from a previous inclusion of Generative AI in a computer science course (Database Systems) [16], a class session on prompt engineering was incorporated to give students an overview of ChatGPT and guidance for writing effective prompts [17]. Students were also made aware of the importance of using Generative AI responsibly by especially being critical of the output due to the issue of factual inconsistencies where output appears plausible but is not correct.

III. SURVEY DESIGN AND ANALYSIS

A. Survey Design

The survey was designed to evaluate the effectiveness of CURE and the impact of Generative AI on students' research experiences across two semesters: Fall 2022 and Fall 2023. The survey employed a Likert scale format, which allows respondents to express their level of agreement with various statements related to the course and Generative AI. The scale ranged from 1 (Strongly Disagree) to 5 (Strongly Agree).

The survey consisted of 15 questions for Fall 2023 and 10 questions for Fall 2022. Questions Q1 to Q10 were common to both semesters and focused on various aspects of the course, such as the clarity of topics, relevance of literature review materials, and understanding of research methods. For Fall 2023, additional questions (Q11 to Q15) were included to specifically assess the impact of Generative AI on different aspects of the research project, including literature review, coding, and understanding complex concepts. Details on the survey questions can be found in the Appendix. The survey was distributed to students enrolled in the course during Fall 2022 and Fall 2023. The survey aimed to measure students' satisfaction, perceived benefits, and overall experience with the

CURE project, particularly focusing on the use of Generative AI introduced in the Fall 2023 course. Students voluntarily participated in the survey, providing their feedback anonymously to ensure honest and unbiased responses. Table 1 summarizes the student enrollment and survey response.

Semester	Total Student Enrollment	Survey Response
Fall 2023	15	15
Fall 2022	13	10

TABLE I
STUDENT COUNT PER COURSE AND RESPONSE RATE OF SURVEY.

B. Analysis Methods

To analyze the survey data and compare responses between the Fall 2022 and Fall 2023 cohorts, the following statistical methods were utilized:

1) *Descriptive Statistics*: Mean scores were calculated for each question in both semesters to provide an overview of central tendencies in the responses. These mean scores offer a straightforward comparison of student responses across the two semesters and across different aspects of the course.

2) *Statistical Tests*: Given the nature of the data—Likert scale responses with small and unequal sample sizes ($n=10$ for Fall 2022, $n=15$ for Fall 2023)—both non-parametric and parametric tests were utilized. The non-parametric test was primarily used, with the parametric test serving as a comparison:

- Mann-Whitney U test: This non-parametric test was chosen as the primary analysis method because it is designed for ordinal data like Likert scale responses and does not assume normality in the data distribution as well as robust to outliers and unequal sample sizes between groups.
- Independent Samples T-test: Despite the ordinal nature of Likert scale data, t-tests are frequently used in educational research and included this parametric test for comparison.

Both tests were used to compare responses between Fall 2022 and Fall 2023 for questions Q1-Q10, which were common to both semesters. The statistical tests used a significance level of $\alpha = 0.05$. This means that p-values less than 0.05 were considered statistically significant, indicating strong evidence against the null hypothesis of no difference between the semesters.

3) *Correlation Analysis*: Correlation analysis was incorporated to examine the relationship between students' responses of the traditional literature review process and their use of Generative AI for literature review. Specifically, Spearman's rank correlation coefficient between Q3 (Literature Review Session) and Q12 (Use of Generative AI in Literature Review) for the Fall 2023 cohort.

C. Statistical Analysis of Survey Results

1) *Mean Scores between Cohorts*: Figure 1 shows the mean scores for both Fall 2022 and Fall 2023. Fall 2022 overall mean (Q1-Q10) is 4.08 with Fall 2023 overall mean at

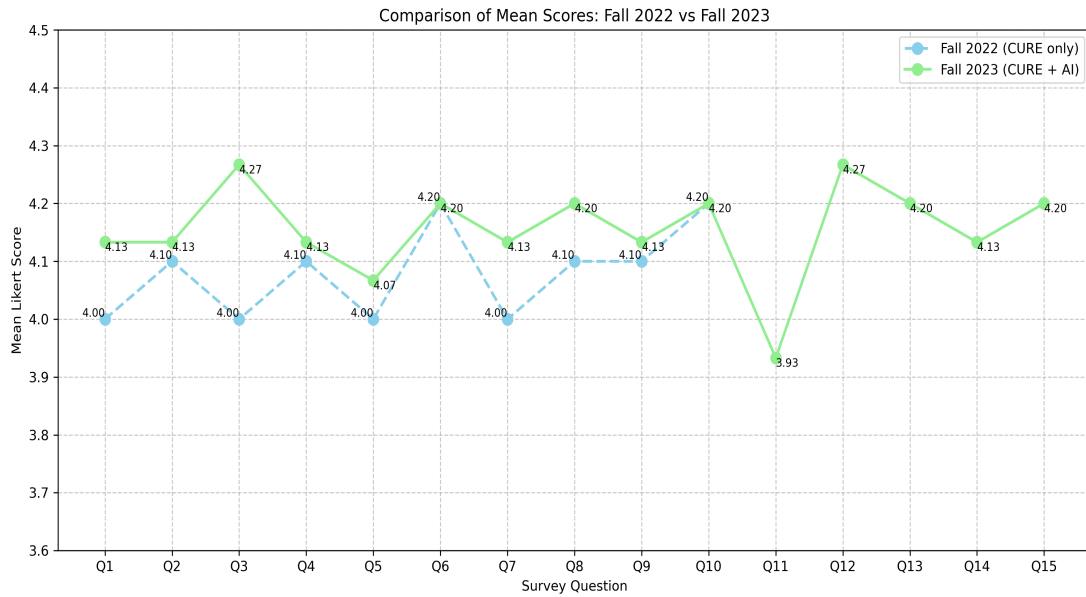


Fig. 1. Comparative plot illustrating mean responses from survey for Fall 2023 and Fall 2022. Questions 1-10 CURE specific and 11-15 questions specific to integrating CURE with Generative AI.

4.15 (Q1-Q15) and Fall 2023 mean at 4.16 (Q1-Q10 only). The observational slight improvement suggests a perceived improvement with Generative AI integration.

2) *Comparison of Fall 2022 and Fall 2023 Responses:* Table II presents the results of the Mann-Whitney U tests and independent samples t-tests to compare responses between Fall 2022 and Fall 2023 cohorts for questions Q1-Q10. None of the questions (Q1-Q10) show statistically significant differences between Fall 2022 and Fall 2023 in either the t-test or Mann-Whitney U test (all p-values > 0.05). While not reaching statistical significance, Q3 shows the largest difference between semesters ($p = 0.246$ for t-test, $p = 0.294$ for Mann-Whitney U test).

Question	T-statistic	T-test p-value	U-statistic	Mann-Whitney U p-value
Q1	-0.563	0.578	67.0	0.613
Q2	-0.152	0.880	73.0	0.915
Q3	-1.189	0.246	59.0	0.294
Q4	-0.152	0.880	71.5	0.832
Q5	-0.236	0.814	70.5	0.804
Q6	0.000	1.000	73.0	0.921
Q7	-0.810	0.425	66.0	0.459
Q8	-0.349	0.729	64.5	0.512
Q9	-0.152	0.880	73.0	0.915
Q10	0.000	1.000	82.0	0.677

TABLE II
T-TEST AND MANN-WHITNEY U TEST RESULTS COMPARING FALL 2023 AND FALL 2022 RESPONSES.

3) *Correlation Analysis on Generative AI impact on Literature Review:* Correlation analysis conducted between Q3 (Literature Review Session) and Q12 (Use of Generative AI in Literature Review) for the Fall 2023 cohort:

- Spearman's rank correlation coefficient: -0.2417

- P-value: 0.3854

This weak negative correlation is not statistically significant, suggesting no strong relationship between students' perceptions of the literature review session and their use of Generative AI for literature review. The perceived improvement in the literature review session, although not statistically significant, aligns with the positive reception of AI for literature review (Q12, mean score 4.27). This suggests that the integration of Generative AI may have contributed to a perceived enhancement of the literature review process, even if this improvement doesn't reach the threshold for statistical significance with the sample size.

The primary limitation of this study is the small sample size (10 responses for Fall 2022 and 15 responses for Fall 2023). Small sample sizes can lead to low statistical power, increasing the likelihood of Type II errors, which is a failing to detect a true effect. A power analysis was conducted to estimate the sample size required to detect a medium effect size (Cohen's $d = 0.5$) with a significance level α of 0.05 and power of 0.80. The analysis indicated that approximately 64 participants per group are needed for robust statistical conclusions.

IV. DISCUSSION

In this study, AI had been utilized in a CURE-based course for conducting literature review, addressing coding and technical issues, and clarifying complex concepts. However, there are challenges and limitations with incorporating Generative AI in a CURE-based course.

A. Challenges and Limitations of Generative AI in CURE

Instructors need training to effectively incorporate AI into their teaching and to help students use AI tools responsibly and effectively. There is a risk of students becoming overly

reliant on AI, potentially undermining the development of fundamental skills and critical thinking, and traditional assessment methods may not fully capture the benefits of AI-enhanced learning experiences. Accessibility issues may arise because not all students or institutions have equal access to advanced AI technologies, potentially exacerbating educational inequalities. Access to the latest models such as ChatGPT 4.0 require a monthly subscription. A major limitation is the rapid pace of AI advancements resulting in tools and methods that can quickly become outdated, necessitating continuous learning and adaptation. It's important to address ethical issues related to AI, such as data privacy, algorithmic bias, and the responsible use of AI technologies. These limitations highlight the need for a thoughtful and careful management to ensure the effective and equitable implementation of AI-enhanced CUREs.

B. Benefits and Advantages of Generative AI in CURE course

The integration of Generative AI with CURE would represent a novel innovation in educational practices. In the context of course-based research the expedition of the research process could be greatly enhanced with the integration of AI. Generative AI can provide personalized feedback and guidance, catering to individual learning paces and styles. This is a significant advantage over traditional one-size-fits-all educational models. Most important of all is providing students with exposure to cutting-edge AI technology enables students to prepare for future challenges in a tech-driven world, making them more competitive in the job market and more capable at adapting to technological advancements.

V. CONCLUSION

This work-in-progress study entailed a preliminary integration of Generative AI with a course modified to incorporate course-based research experiences. Students gained hands-on application and understanding of the responsible use of Generative AI in the classroom ensuring alignment with educational goals. The results suggest a positive trend in student perceptions of the course content and the use of Generative AI. However, the lack of statistical significance indicates that these differences may not be robust enough to draw definitive conclusions. The survey relied solely on quantitative data, which, while useful, may not have captured the full depth of student experiences and perceptions.

VI. FUTURE WORK

While the current study's results are not statistically significant, they provide a foundation for further research. This study shows the potential for transforming CURE-based courses in computer science by the incorporation of Generative AI into the course-based research experience. Future work will include open-ended questions to ensure richer qualitative insights to complement the quantitative data. Also, the small sample size is a major limitation, and future studies with larger participant groups are necessary to validate these findings and provide more reliable insights.

REFERENCES

- [1] L. C. Auchincloss, S. L. Laursen, J. L. Branchaw, K. Eagan, M. Graham, D. I. Hanauer, G. Lawrie, C. M. McLinn, N. Pelaez, S. Rowland *et al.*, "Assessment of course-based undergraduate research experiences: a meeting report," 2014.
- [2] F. e. a. delogu, "A systemic transformation of an arts and sciences curriculum to nurture inclusive excellence of all students through course-based research experiences," *Frontiers in Education*, vol. 8, 10 2023.
- [3] G. Bangera and S. E. Brownell, "Course-based undergraduate research experiences can make scientific research more inclusive," *CBE—Life Sciences Education*, vol. 13, no. 4, pp. 602–606, 2014.
- [4] C. D. Shaffer, C. J. Alvarez, A. E. Bednarski, D. Dunbar, A. L. Goodman, C. Reinke, A. G. Rosenwald, M. J. Wolyniak, C. Bailey, D. Barnard *et al.*, "A course-based research experience: how benefits change with increased investment in instructional time," *CBE—Life Sciences Education*, vol. 13, no. 1, pp. 111–130, 2014.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [6] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
- [7] S. Bauschard, A. Rao, P. Shah, and C. Shryock, *Chat (GPT): Navigating the impact of Generative AI on Education Theory and Practice*. Pedagogy Ventures, 2023.
- [8] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja *et al.*, "“so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *International Journal of Information Management*, vol. 71, p. 102642, 2023.
- [9] T. Adiguzel, M. H. Kaya, and F. K. Cansu, "Revolutionizing education with ai: Exploring the transformative potential of chatgpt," *Contemporary Educational Technology*, vol. 15, no. 3, p. ep429, 2023.
- [10] J. Willems, "Chatgpt at universities—the least of our concerns," *Available at SSRN 4334162*, 2023.
- [11] J. Rudolph, S. Tan, and S. Tan, "Chatgpt: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023.
- [12] X. Zhai, "Chatgpt user experience: Implications for education," *Preprint: <https://ssrn.com/abstract=4312418> or <http://dx.doi.org/10.2139/ssrn.4312418>*, 2022.
- [13] P. P. Ray, "Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, 2023.
- [14] H. Peng, S. Ma, and J. M. Spector, "Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment," *Smart Learning Environments*, vol. 6, no. 1, pp. 1–14, 2019.
- [15] P. Lauren, "A study of course-based undergraduate research experiences and the challenges and opportunities for computer science," in *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*. American Council on Science and Education, 2023, pp. 1008–1015.
- [16] P. Lauren and P. Watta, "Work-in-progress: Integrating generative ai with evidence-based learning strategies in computer science and engineering education," in *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2023, pp. 1–5.
- [17] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.

APPENDIX - SURVEY QUESTIONS

Questions 1-10 were administered to Fall 2022 course and Questions 1-15 administered to Fall 2023 course. The slight variation of the question is noted to align with the project designated for the course during each semester. As noted earlier, the Likert scale had been utilized where students indicate their level of agreement with the following statements:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neutral
- 4 = Agree
- 5 = Strongly Agree

- Q1** *Fall 2023:* The introduction to emotion detection session was clear and informative.
Fall 2022 The introduction to gender bias detection session was clear and informative.
- Q2** *Fall 2023:* The session on the overview to research enhanced my understanding of Large Language Models (LLMs).
Fall 2022: The session on the overview to research enhanced my understanding of topic modeling.
- Q3** *Fall 2023:* The session on the introduction to the literature review along with the preliminary papers for the literature review were highly relevant to our research topic on emotion detection.
Fall 2022: The session on the introduction to the literature review along with the preliminary papers for the literature review were highly relevant to our research topic on gender bias detection.
- Q4** *Fall 2023:* I clearly understood the research question on addressing imbalance using LLM text generation in the emotion detection classification task.
Fall 2022: I clearly understood the research question on using topic modeling in the gender bias detection classification task.
- Q5** *Fall 2023/Fall 2022:* The session on research methods was clear and helped me understand the process of conducting research.
- Q6** *Fall 2023/Fall 2022:* The preliminary codebase provided was useful for understanding the baseline of our research project.
- Q7** *Fall 2023/Fall 2022:* The session on LaTeX² significantly helped in the construction of my research paper.
- Q8** *Fall 2023/Fall 2022:* The overview to the peer review process provided effective criteria for evaluating papers.
- Q9** *Fall 2023/Fall 2022:* The four-week CRE format was effective in delivering comprehensive knowledge and skills in conducting research in text-based machine learning.
- Q10** *Fall 2023/Fall 2022:* I am satisfied with my learning experience from the course-based research experience as part of the course.
- Q11** *Fall 2023:* Generative AI provided useful assistance with coding and technical aspects of my research project.
- Q12** *Fall 2023:* Generative AI was beneficial for the literature review process in my research project.
- Q13** *Fall 2023:* Generative AI was effective in clarifying complex concepts and answering queries related to my research project.
- Q14** *Fall 2023:* The session on prompt engineering helped me formulate more effective prompts to get better results using Generative AI.
- Q15** *Fall 2023:* Using Generative AI improved the efficiency of my work for the research project.

²<https://www.latex-project.org>