A Modern Approach to Teaching Numerical Calculus With Artificial Intelligence

Elisabetta Crescio, Tecnologico de Monterrey, Mexico

The Asian Conference on Education & International Development 2024
Official Conference Proceedings

Abstract
The advent of artificial intelligence in educational institutions has been met with fear and concern due to the ethical implications it carries, especially in the assessment of learning. In this work, results from an experiment aimed at using artificial intelligence as an ally in teaching numerical calculus will be presented. The teaching experience was implemented in an intensive course (five weeks for a total of 24 hours) on numerical methods for the solution of differential equations. The course was directed to undergraduate students of engineering programs in physics and data science. Not all the students attending the course had skills in the required programming language (Python) and, due to the short time, it was not possible to teach them basic computing during the course. The use of tools as chatGPT and Bing for code generation allowed the students to learn and apply different numerical methods of increasing complexity, and, at the same time, improve their skills in python programming. Numerical results of diagnostic exams and a sentiment analysis on the students' perception about the use of artificial intelligence in the course will be presented.

Keywords: Artificial Intelligence, Python Programming, Code Generation, Numerical Methods

Introduction

The advent of artificial intelligence in educational institutions has been met with fear and concern due to the ethical implications it carries, especially in the assessment of learning. Large-scale Language Models (LLMs) offer great opportunities performing several tasks in a brief time. In the experiment reported here, LLMs have been used to generate algorithms and graphs in Python.

The efficiency and accuracy of programming code generation of ChatGPT (OpenAI, 2022) has been reported by several authors. An overall efficiency of 71.875% in generating solutions to proposed programming problems has been reported (Sakib et al., 2021), while an efficiency greater than 90% has been measured for ChatGPT in producing algorithms involving strings and mathematical operations in Python (Buscemi A., 2023). As remarked by (Rahman, Md. M., and Watanobe, Y., 2023), ChatGPT can have a significant impact on programming learning, but also could have a negative impact on critical thinking and problem-solving skills.

The teaching experience was tailored for undergraduate students in engineering programs, specifically those in physics and data science, and it was conducted as part of a numerical calculus course. Not all students in the course possessed the necessary programming skills in Python, and due to time constraints, it was impractical to teach them basic programming during the course. The use of large language models, such as ChatGPT and Bing, for code generation enabled students to learn and apply various numerical methods of increasing complexity without grappling with programming challenges.

The main purpose of the experiment presented here was to assess the benefits of using ChatGPT as a support in programming code writing, as well as to raise awareness on the ethical use of artificial intelligence, on its limitations and its potential threats.

The primary analysis focused on a sentiment analysis to gauge students' perceptions regarding the use of artificial intelligence in the course. Also, a comparative analysis of students' programming skills before and after the experiment has been conducted to analyze whether the students could improve their skills.

Experiment

The experiment spanned the entire five-week duration of the course. Throughout the course, students were tasked with completing various activities in which they had to write algorithms, using Python, for different methods of solving differential equations, including the Euler, modified Euler, implicit Euler, Runge-Kutta of order 4, and finite difference methods. Additionally, they were required to test these methods with numerous examples and create graphs comparing numerical and analytic solutions.

As the final project of the course, students were tasked with coding and solving, from scratch, the system of equations for the SIR epidemiological model (W.O. Kermack and A.G. McKendrick, 1927) and proposing variations to the model. In all these activities, students were required to interact with ChatGPT or Bing to generate the code, run and verify it, using the theoretical results covered in the classroom as a reference.

Sentiment Analysis

At the end of the course, the students were asked to write a short comment about their experience with ChatGPT or Bing. In total, 33 opinions were collected and processed to conduct a sentiment analysis that could summarize the students' opinions.

Firstly, an analysis of the polarity of the comments was performed with the algorithm vaderSentiment (Hutto, C.J. and Gilbert, E.E., 2014). As shown in Figure 1, most students, 89.6%, expressed a positive opinion. A small fraction of students, 7.9% expressed a negative opinion and only 2.6% a neutral opinion.

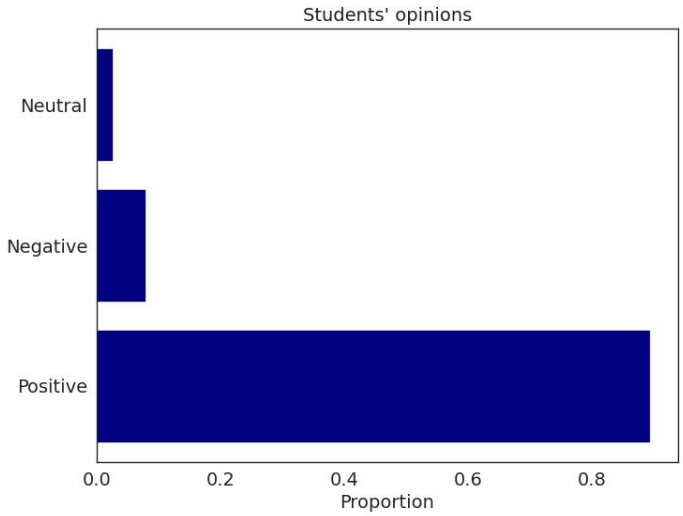

Figure 1. Result of the analysis of polarity.

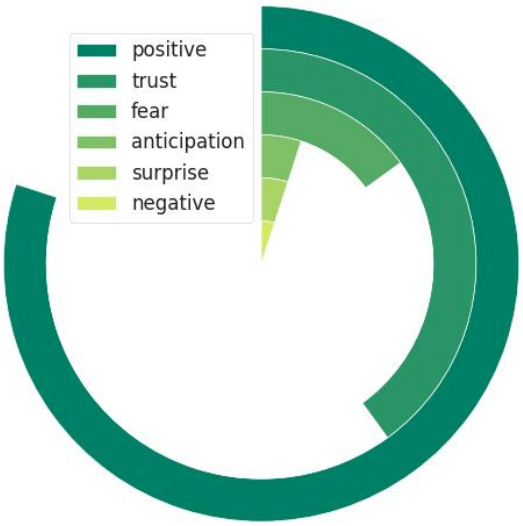

Figure 2. Result of the analysis of emotions.

Subsequently, the emotions of the students were extracted using the algorithm NRCLex (Mark M. Bailey, 2019). As shown in Figure 2, five principal emotions were found. In order of importance, the first was a positive feeling (53.3%), then trust (26.6%), fear (10%), anticipation (3.3%), surprise (3.3%) and a negative feeling (3.3%). It is noteworthy that 10% of the students expressed fear regarding the use of artificial intelligence. This fear may be

associated with the potential for unethical use of artificial intelligence, as well as concerns about the reliability of the content it generates.

Comparative Analysis

Even though the primary focus of the course was to teach numerical methods for solving differential equations rather than Python programming, a diagnostic test was administered at both the beginning and end of the course to assess whether students had improved their programming skills. One of the main reasons for doing so was the potential controversy surrounding the use of ChatGPT or Bing. Allowing students to rely solely on artificial intelligence for quickly generating solutions to class activities and homework might impact their actual understanding and knowledge development.

Due to the limited number of students and class planning, formal study with test and control groups was not feasible, preventing the assessment of causal relations. A comparative analysis was undertaken, encompassing both qualitative and quantitative aspects of the results obtained from the diagnostic test. The test comprised 20 multiple choice questions related to Python programming, focusing specifically on syntax and data structures. The individual results of each student were assessed based on the number of correct answers.

Due to the necessity of performing a paired t-test, the sample size of usable answers was reduced to 26. The mean number of correct answers in the first test at the beginning of the course was 14.04, while at the end of the course, it increased to 14.81. A pooled two tailed t-test, with a significance level of 0.05, was performed, finding a p-value of 0.51. The hypothesis test suggests that there is not a significant difference between the two values, so, on average, there was no improvement in the students' programming skills. The individual results of each student were also analyzed, and Figure 3 illustrates the comparison between the number of correct answers in the two tests for each student. Each bar, labeled with a number, represents a student, with the score obtained in the first test marked by a black circle. The colored bars represent the new score. Fifty percent of the students showed an improvement in their score after the course, with a mean improvement of 3 points. This suggests that half of the students may have benefited from the assistance of the code generated by artificial intelligence in enhancing their knowledge.

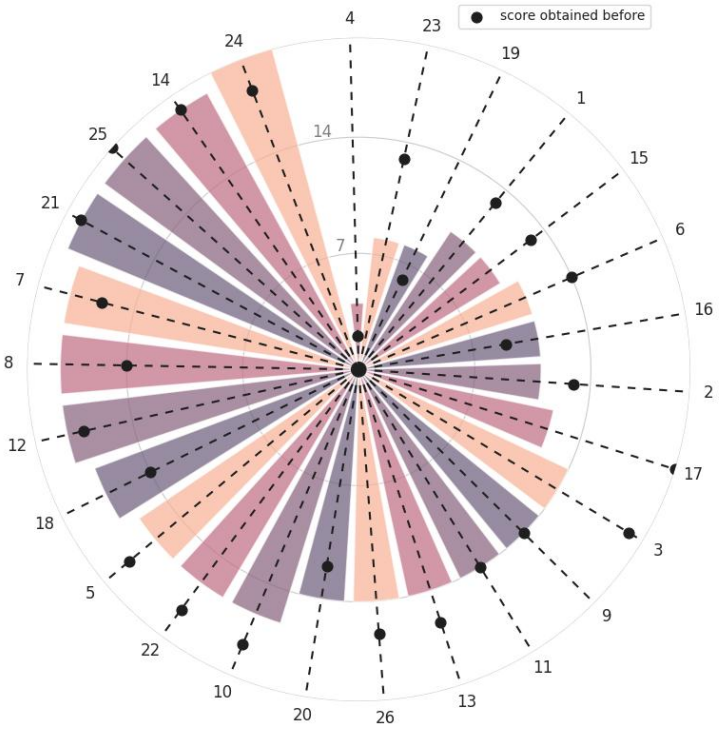

Figure 3. Comparison of the individual results of each student in the two tests.

Conclusion

The teaching experience of utilizing LLMs as a support for students in Python programming has been well-received by most students (89.6%), eliciting positive emotions (53.3%) and trust (26.6%). However, around 10% of students expressed fear, indicating potential concerns about the ethics or accuracy of content generated by artificial intelligence. Diagnostics conducted on students before and after the course revealed that, on average, there was no significant improvement in their programming skills. Upon examining individual results, it was found that 50% of students may have benefited from the experience, demonstrating an average improvement of three points in their scores.

Acknowledgements

I would like to express my gratitude to the Science Department of Tecnologico de Monterrey, Campus Estado de Mexico, for the financial support.

References

Bailey, M. M. (2019). https://pypi.org/project/NRCLex/

Buscemi, A. (2023). A Comparative Study of Code Generation using ChatGPT 3.5 across 10 Programming Languages. *arXiv:2308.04477*. https://doi.org/10.48550/arXiv.2308.04477

Hutto, C. J. & Gilbert, E.E. (2014). https://pypi.org/project/vaderSentiment/

Kermack, W. O. & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond,* 115, 700-721.

OpenAI, (2022). *ChatGPT: Optimizing language models for dialogue.* https://openai.com

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for Education and research: Opportunities, threats, and strategies. *Applied Sciences, 13(9)*, 5783. https://doi.org/10.3390/app13095783

Sakib, F. A., Khan, S. H., & Karim, A. H. M. R. (2023). Extending the Frontier of ChatGPT: Code Generation and Debugging. *arXiv:2307.08260*. https://doi.org/10.48550/arXiv.2307.08260

Contact email: elisabetta@tec.mx