**A Pilot Study of Using ChatGPT as a Teaching Tool for Software Engineering Education**

Y.H. Kwok\*,a and K.P. Wong\*\*, b

a Department of IT, Hong Kong Institute of Vocational Education (Sha Tin), Hong Kong

b Centre For Teaching and Teaching, Vocational Training Council, Hong Kong

\*kwokyuho@vtc.edu.hk , \*\*alexander.aw@vtc.edu.hk

**Abstract**

**The aim of this study is to investigate the efficacy of using ChatGPT to support teaching in software engineering education, focusing on its impact on both the efficiency and effectiveness of teaching. The study will employ a mixed-method research design, collecting and analysing both qualitative and quantitative data related to students' academic performance, educators' preparation of teaching and learning material, and students' perceptions of the teaching methods.**

**The participants of this investigation will consist of higher diploma students in software engineering at the Hong Kong Institute of Vocational Education. These students will be allocated randomly to one of two groups: control group shall be provided with human-created learning content, while the experimental group will receive teaching material and interim feedback generated by ChatGPT.**

**ChatGPT-generated teaching and learning materials will encompass a variety of resources to support instruction, such as lecture presentations, tutorial notes, assignments and assessments, and marking schemes. These resources will be thoroughly validated and moderated by the lecturer to ensure alignment between the two groups.**

**Using ChatGPT as a pedagogical tool is expected to amplify the efficiency in teaching software engineering, leading to superior performance by the experimental group over the control group. A supplementary objective of the study is to assess and scrutinise the feasibility of relieving teachers from the burden in preparing instructional materials with the assistance of ChatGPT. The study's results will exemplify the value and usefulness of employing ChatGPT to assist teaching.**

**This research will contribute to the growing body of knowledge on the use of artificial intelligence in education and has implications for the broader field of teaching and learning. Future research can explore the use of ChatGPT in teaching other subjects and investigate students' perceptions of using ChatGPT as a teaching and learning tool.**

**Keywords:** *ChatGPT, teaching tool, EdTech tool, software engineering education, artificial intelligence-assisted teaching, teaching efficiency, teaching effectiveness*

**Introduction**

Software engineering is a rapidly evolving field that requires continuous learning and adaptation to new technologies, methodologies, and best practices. Traditional teaching methods and materials in software engineering education face challenges in keeping up with this fast-paced environment, addressing the diverse learning needs of students, and providing timely and personalised feedback (Bourque & Fairley, 2022). These challenges have motivated researchers and educators to explore the use of AI-based tools to enhance teaching and learning experiences in software engineering education (Chiu et al., 2023).

One such AI-based tool is ChatGPT, a state-of-the-art language model developed by OpenAI (Brown et al. , 2020). It has demonstrated remarkable capabilities in generating human-like text, answering questions, providing explanations, and engaging in conversations across various domains (OpenAI, n.d.). The potential applications of ChatGPT in education have started to gain attention, with studies reporting its effectiveness in generating learning materials, providing tutoring, and facilitating collaborative learning activities (Qureshi, 2023). Though the use of ChatGPT in educational institutes is still in argument, there are both challenges and opportunities for education. (Kirk, 2023) It is vital about integrating promising technology in teaching and learning, finding out and tackling down challenges for ensuring successful implementation and positive student outputs. (Passey et al., 2021)

In this pilot study, we investigate the efficiency and effectiveness of using ChatGPT-generated teaching materials for software engineering education in comparison with traditional human-generated content. We aim to answer the following research questions:

1. Can ChatGPT-generated materials be produced more efficiently (in terms of time and resources) than human-generated content?
2. How do students perform and engage with ChatGPT-generated materials compared to traditional human-generated content in a software engineering course?

By addressing these questions, the researchers   hope to shed light on the potential benefits and limitations of using ChatGPT as a teaching tool in software engineering education and provide insights for future research and development in this area.

**Literature Review**

Several studies have investigated the efficiency of computer-generated materials in terms of time and resource requirements. AI in education (AIED) adaptive learning and evaluation applications are widely being used to improve educational effectiveness and efficiency (Chassignol et al., 2018; Kurshan, 2016) together with better understanding of student knowledge acquisition (VanLehn et al., 2007; Beal et al., 2010).

Graesser, Conley and Olney (2012) mentioned an Intelligent Tutoring Systems (ITS) is a computer-based learning tool that makes use of AI to create adaptive educational environments that respond both to the learner’s level and needs, and to the instructional agenda. They also highlighted that ITS often incorporated pedagogical, psychological, and other cognitive learning theories into computational models.

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022) presented a case study on the automatic generation of programming exercises, including sample solutions and test cases, and code explanations using large language models, demonstrating the capabilities and efficiency of AI tools in content creation. These studies suggest that the majority of the automatically generated content is both novel and sensible. From their analysis, suggesting that there is significant value in massive generative machine learning models as a tool for instructors, there remains a need for some oversight to ensure the quality of the generated content before it is delivered to students.

In the realm of machine learning, Large Language Models (LLMs) that are trained on voluminous textual data, the ChatGPT, is among the notable examples of LLMs. These models have the remarkable capacity to generate code through natural language prompts. Given their potential, there is a global conversation surrounding the ability of LLMs to produce programming assignments and code explanations. This has the potential to revolutionise the teaching practices of instructors.

The potential of Large Language Models (LLMs) in enhancing educators' pedagogy and research has been evaluated by MacNeil, S., Tran, A., Leinonen, J., Denny, P., Kim, J., Hellas, A., Bernstein, S., & Sarsa, S. (2022), demonstrating LLMs capabilities. Their assessment sought to determine the feasibility and manner of integrating LLMs into the learning programming and have the most positive potential impact on computing education.

In a recent study, Qureshi (2023) performed an experiment to evaluate the efficacy of employing ChatGPT in programming problem-solving. The findings of the study demonstrated that the experimental group, which utilised ChatGPT, outperformed the control group by attaining better scores in a shorter amount of time. It is worth noting, however, that despite achieving higher scores, the experimental group was unable to attain perfect scores, as ChatGPT generated code with some inaccuracies or inconsistencies.

Qureshi, B. (2023) has emphasised the necessity of providing training to faculty members in order to equip them with the skills necessary to incorporate AI tools into their teaching methods. Furthermore, Qureshi, B. (2023) advocates for the importance of imparting knowledge on academic integrity to students, with the aim of ensuring their comprehension of the significance of upholding ethical standards in their academic work.

Wang, Z., Lan, A.S., Nie, W., Waters, A.E., Grimaldi, P.J., & Baraniuk, R. (2018) proposed QG-Net, a type of recurrent neural network model that has been specifically crafted for generating quiz questions based on educational content. Their study has demonstrated that the questions generated by QG-Net are notably distinct from the training data. The authors have concluded that the questions created by QG-Net are coherent, pertinent, and closely resemble those created by humans in comparison to existing models. These positive results suggest that QG-Net has the potential to automate and expand the question generation process for educational purposes, thereby complementing the abundant educational content available.

Understanding how students perform and engage with ChatGPT-generated materials is crucial. Cunningham-Nelson, S., Boles, W.W., Trouton, L., & Margerison, E. (2019) shows that chatbots could provide a variety of enhancements for student learning and educator content delivery. Juanan Pereira (2016) mentioned chatbots can be an ideal teacher assistant. P. Smutny and P. Schreiberova (2020) examined educational chatbots in instant messenger to support learning that can provide a basic level of sending personalised messages to recommending learning content. Wollny, Schneider, Di Mitri, Weidlich, Rittberger, and Drachsler (2021) also highlighted that chatbots can foster active participation, provide immediate feedback, and support individualised learning experiences.

Wollny et al. (2021) have pointed to three different pedagogical roles for chatbots in education: a supporting learning role, an assisting role, and a mentoring role. Graham Attwell (2022) conducted a short literature review on using chatbots in education, together with studies of using chatbots for careers guidance, counselling and advice, and concluded that chatbots can be a useful tool in these areas, notably that chatbots are not a replacement for human but should be used as part of a broader strategy.

Holmes, Wayne & Bialik, Maya & Fadel, Charles. (2019) explore the promises and implications of AI in education and highlight the potential of AI tools to support student-centred learning and promote deeper engagement. These studies provide valuable insights into the potential benefits of using computer-generated materials in software engineering education.

ChatGPT-generated materials hold promise in terms of efficiency and student engagement in software engineering education based on the reviewed literature. The studies suggest that ChatGPT can generate materials more efficiently, reducing the time and resources required from educators. Moreover, students may perform better and be more engaged when interacting with ChatGPT-generated materials due to their interactive and personalised nature.

**Methodology**

This study involved a total of 56 higher diploma students enrolled in a software engineering programme at the Department of Information Technology, Hong Kong Institute of Vocational Education (Sha Tin), following the recommendations for sample size in pilot studies. (Whitehead et al., 2015, Lewis et al., 2021). The students were divided into two groups, each composed of 29 and 27  students respectively. Additionally, a lecturer with over ten years of experience in teaching software engineering related programmes participated in the study to provide guidance, supervision, and evaluation.

Two set of teaching materials were prepared for this study, following the guidelines for creating teaching and learning materials of Vocational Training Council:

*Human-generated content:*The lecturer prepared a set of traditional instructional materials including lecture notes, workshop exercises and assessments  according to the syllabus of the module selected for this study.

*ChatGPT-generated materials:*The lecturer provided an outline of topics and learning objectives to be covered in the software engineering programme. Using ChatGPT, a set of instruction material, which is similar to Human-generated content, was generated for the same topics and learning objectives.

The content in both sets of materials was reviewed and approved by the module lecturer to ensure accuracy, relevance and comparability.

*Procedure*

The research context for this project is the learning content on "Micro-services" within the module "Enterprise System Development". This module was selected for this research due to the recent update of its syllabus. Specifically, the addition of "Micro-services" as a new learning content calls for the development of instructional materials to facilitate effective learning. And due to the time constraints, the study was conducted one month, following these steps based on mixed-methods research in education.

*Material generation:*The lecturer generated the human-generated content, then used ChatGPT to generate the ChatGPT-based materials.

*Classroom implementation:* Two classes, 56 students in total, were randomly assigned to two groups. Class 2A (29 students) used the human-generated materials, and class 2C (27 students) used the ChatGPT generated content throughout the semester. Before the implementation of the study, both students received human-generated instructional materials.

1. *Data collection:* Quantitative and qualitative data were collected to evaluate the efficiency and effectiveness of the materials. Quantitative data included students' performance on assignments, surveys and completion rates of workshop exercises.. Qualitative data were obtained through interviews and feedback regarding the materials.
2. *Data analysis:* Quantitative data were analysed using descriptive and inferential statistics to compare the performance of students in both groups (Field, 2013). Thematic analysis was employed to analyse the qualitative data, identifying patterns and themes related to students' experiences and perceptions (Braun and Clarke, 2006).

*Measures*

The following metrics were used to evaluate the efficiency and effectiveness of the ChatGPT-generated materials compared to the human-generated content, based on established evaluation methods in educational research (Fredricks et al., 2004):

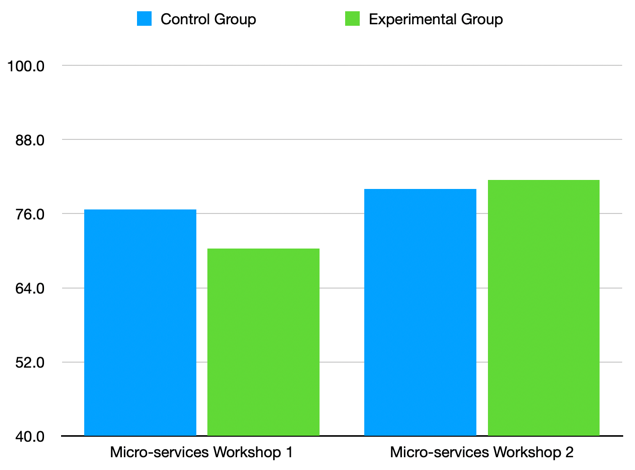
1. *Efficiency:* The time and resources required to generate each set of materials were recorded and compared.
2. *Student performance:* Assignment grades, and in-class activity completion rates were compared between the two groups to assess the effectiveness of the materials.
3. *Student perceptions:* Surveys and interviews were conducted to gather students' feedback on the quality, relevance, and effectiveness of the materials, as well as their overall learning experience (Ryan & Deci, 2000).

These measures allowed us to address the research questions and evaluate the potential of ChatGPT as a teaching tool in software engineering education.

**Findings**

*Test Results and Completion Rate*

The study conducted two micro-services workshops with a control group and an experimental group. *Figure 1* presents the completion rates of the workshop exercises. The completion rates for the first workshop were 76.7% for the control group and 70.4% for the experimental group. For the second workshop, the completion rates were 80.0% for the control group and 81.5% for the experimental group. These completion rates suggest that both groups were actively engaged in the workshops, with the experimental group showing slightly higher completion rates than the control group in the second workshop.

****

*Figure 1. Completion rate of human-generated and ChatGPT generated workshop exercises*

In addition, a quiz was conducted to assess the learning outcomes of the micro-services learning content. The average marks for the quiz were 56.7 for the control group and 56.4 for the experimental group, indicating that both groups had similar levels of understanding of the concepts covered in the workshops.

 The findings suggest that both the control and experimental groups were equally successful in completing the workshops and had similar levels of understanding of the concepts covered in the quiz.

*Student Survey Responses*

A total of 46 of 56 research subjects’ survey responses were collected and are presented in *Figure 2.*

****

*Figure 2. Results of the post-experiment student survey*

The survey results provided interesting insights into students' perceptions of the teaching materials. In the first question, 86% of the control group and 92% of the experimental group students believed that their lecture notes were human-generated. This suggests that the majority of both groups perceived the content to be of human origin, regardless of whether it was AI-generated or not.

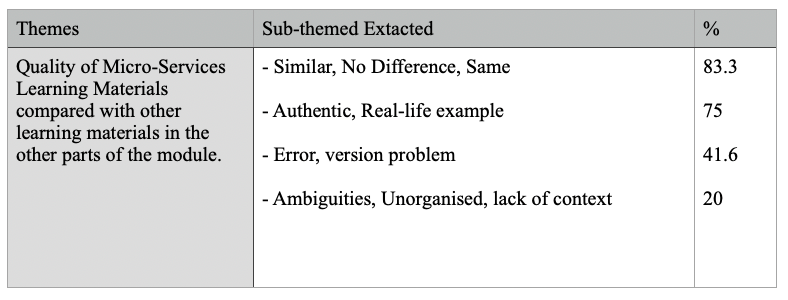
In the second question, 86% of the control group and 83% of the experimental group students believed that their workshop exercises were human-generated. The results are similar to the first question, further emphasising that students in both groups could not clearly differentiate between human-generated and AI-generated content.

The third question compared the learning materials to other content within the module. In the control group, 78% of students felt that the learning materials were no different from other learning materials, and 10% thought they were better. In the experimental group, 67% believed that the learning materials were no different from others, while a higher proportion (25%) thought the learning materials were better. This indicates that a considerable number of students in the experimental group found the AI-generated content to be of higher quality compared to other learning materials in the module.

The survey results demonstrate that students in both the control and experimental groups were largely unable to differentiate between human-generated and AI-generated content, and many of them perceived the AI-generated materials to be of similar or better quality compared to other learning materials in the module.

*Students Perception on Instruction Materials*

The thematic analysis of the extracted theme focuses on the comparison of ChatGPT-generated learning material in micro-services with human-generated learning materials. The data is derived from a semi-opened interview involving 12 students in an experimental group. The analysis has identified four sub-themes, each with its respective prominence in the data. And the summary of the thematic analysis is presented in *table 1*.

****

*Table 1. Summary of Thematic Analysis*

*Similarity (83.3%):* The most prominent sub-theme suggests that a significant majority of the students found the ChatGPT-generated learning material to be similar or indistinguishable from human-generated learning materials. This indicates that the quality and content of the ChatGPT-generated material are comparable to those created by humans, at least from the perspective of the students in the experimental group.

*Authenticity (75%):* The second most prominent sub-theme reveals that the students found the ChatGPT-generated learning material to provide authentic, real-life examples. This suggests that the AI-generated material is relevant and practical, offering value to the learning process by being grounded in real-world situations.

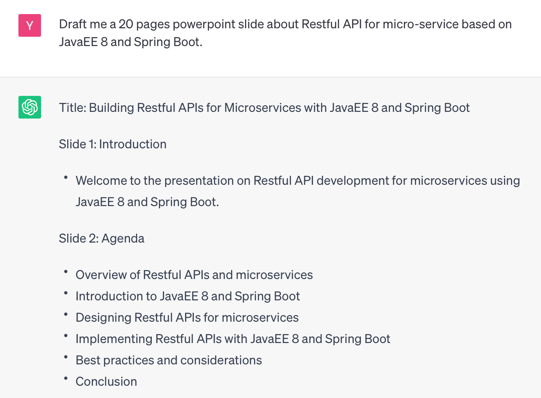
*Errors and Version Problems (41.6%):* This sub-theme indicates that some students encountered issues with errors or version problems in the ChatGPT-generated learning material. While this does not necessarily imply that the AI-generated material is inferior, it does highlight potential areas for improvement in terms of content accuracy and consistency.

*Ambiguities and Lack of Context (20%):* The least prominent sub-theme suggests that a minority of the students found the ChatGPT-generated learning material to be ambiguous, unorganised, or lacking context. This may indicate that some of the AI-generated content requires further clarification or additional context to enhance its usefulness and comprehensibility for learners.

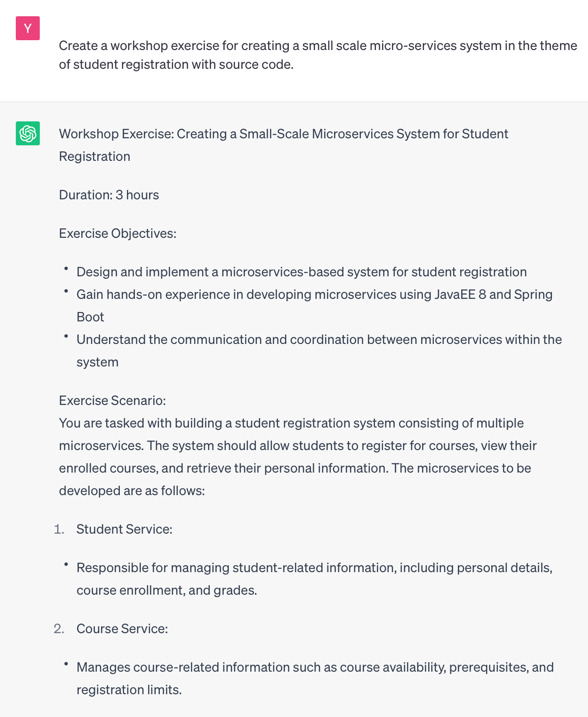
*Lecturer Preparation Time*

The findings indicate that the lecturer dedicated a significant amount of time to prepare for lectures and workshop exercises. For lecture preparation, the lecturer spent around 2-3 hours on tasks such as searching for materials, drafting and finalising the lecture notes. Additionally, the lecturer spent approximately 4-5 hours planning for a theme, program, and validating the code for workshop exercises.

By using ChatGPT, the lecturer followed a different approach. The lecturer spent about 10 minutes entering the prompt and then around 50 minutes validating and transferring the result to PowerPoint. *Figure 3* presents the ChatGPT prompt input by the lecturer to generate the structure for the lecture note. The lecturer then invested 10-20 minutes using ChatGPT to generate workshop exercises with a theme. Finally, the lecturer spent around 2 hours validating the course code and transferring the content to a word document. *Figure 4* presents the prompt used to plan the structure of the workshop. Fig*ure 5* presents the sample code generated by ChatGPT for a given theme about building a student registration system.

****

*Figure 3. Screen Capture of ChatGPT for generating Lecture Note*

****

*Figure 4. Screen Capture of ChatGPT for generating Workshop Exercise*

|  |
| --- |
| @Path("/students")  public class StudentResource {      @GET      @Produces(MediaType.APPLICATION\_JSON)      public Response getAllStudents() {          // Logic to fetch all students from the database          List<Student> students = // Fetch students from the database          return Response.ok(students).build();      }      @GET      @Path("/{studentId}")      @Produces(MediaType.APPLICATION\_JSON)      public Response getStudentById(@PathParam("studentId") String studentId) {          // Logic to fetch student by ID from the database          Student student = // Fetch student from the database based on the ID          if (student != null) {              return Response.ok(student).build();          } else {              return Response.status(Response.Status.NOT\_FOUND).build();          }      }      // some source code are trimmed  } |

*Figure 5. Source code extract generated by ChatGPT*

It is worth noting that the results obtained from ChatGPT were satisfactory at the level of higher diploma. Additionally, the use of ChatGPT to create learning materials significantly reduced the time needed for preparation. The lecturer spent only a fraction of the time on creating exercises compared to other tasks such as planning and validation.

*Quality of ChatGPT-Generated Instructional Materials*

Despite the promising results in terms of efficacy and preparation time, ChatGPT-generated workshop exercises were not without issues. Some exercises contained errors, such as incomplete programs or source code that could not be compiled. This highlights the need for further refinement and improvement of ChatGPT-generated materials to ensure their reliability and accuracy in educational settings. For example, ChatGPT may generate deprecated source code from the old version of Java Enterprise Edition.

**Discussion**

The findings of this study provide valuable insights into the efficacy and efficiency of using ChatGPT-generated teaching materials in the context of computer science education. The results demonstrated that students using AI-generated materials performed similarly to those using human-generated content, indicating that ChatGPT-generated materials can be an effective alternative to traditional teaching resources. Furthermore, students were unable to distinguish between the two types of materials, suggesting that the quality of AI-generated content is comparable to that of human-generated materials.

A significant advantage of using ChatGPT-generated materials was the reduced preparation time for the lecturer. The time saved in preparing lectures and workshop exercises could be allocated to other tasks, such as student support, research, or administrative duties. Additionally, the ease of generating themed workshop exercises using ChatGPT has the potential to enhance the variety and relevance of learning materials, offering students a more engaging and diverse educational experience.

However, the study also revealed some limitations of ChatGPT-generated materials, particularly in the context of workshop exercises. The occasional errors found in the generated source code, such as incomplete programs or code that could not be compiled, highlight the need for further development and refinement of AI-generated content. To ensure the reliability and quality of AI-generated materials, it is crucial to address these issues and implement strategies for detecting and correcting errors before deploying them in an educational setting.

As AI and NLP technologies continue to advance, enhanced versions of ChatGPT might offer improved accuracy and reliability, making it an even more viable option for generating educational materials. Future research should explore the long-term efficacy of using AI-generated content in various educational contexts and investigate potential strategies for addressing the limitations identified in this study. Additionally, the integration of AI-generated materials with other educational technologies, such as learning management systems, adaptive learning platforms, and collaborative learning tools, could further enhance the learning experience for students.

**Future Research Directions**

To expand on the findings of this study, we suggest several directions for future research. Investigating the long-term impact of using AI-generated teaching materials on student performance, engagement, and retention rates across various educational settings and disciplines could provide valuable insights (Seo et al., 2021). Developing and evaluating methods for automatically detecting and correcting errors in AI-generated content, particularly in the context of code examples and workshop exercises, would help address the limitations identified in this study (Combefis, 2021).

Assessing the effectiveness of integrating AI-generated materials with other educational technologies such as learning management systems, adaptive learning platforms, and collaborative learning tools can provide a more comprehensive understanding of the potential benefits of AI-generated content (Siemens and Baker 2012). Exploring the potential of AI-generated materials to provide personalised and adaptive learning experiences tailored to individual students' needs, learning styles, and progress could enhance the educational experience (Younes, 2021).

Investigating the effectiveness of AI-generated materials when combined with various pedagogical approaches, such as problem-based learning, flipped classrooms, and peer instruction, can help determine the optimal conditions for their implementation (Brame, 2013). Additionally, examining the experiences and perspectives of instructors using AI-generated materials, including their perceptions of quality, utility, and the impact on their teaching practice, can provide further insights into the practical applications of AI-generated content in education (Zawacki-Richter et al., 2019).

By pursuing these research directions, future studies can contribute to a deeper understanding of the potential benefits and challenges associated with using AI-generated teaching materials in education. This knowledge can ultimately help optimise their use and enhance learning outcomes for students.

**Conclusions**

This study has demonstrated that ChatGPT-generated teaching materials can serve as an effective and efficient alternative to human-generated content in computer science education (OpenAI, n.d.). The findings indicate that students' performance is comparable when using AI-generated materials, and they are unable to distinguish between human-generated and AI-generated content. Furthermore, the reduced preparation time for instructors offers a significant advantage in terms of resource allocation. However, the occasional errors in workshop exercises emphasise the necessity for further refinement of AI-generated materials to ensure their reliability and quality.

**References**

Beal, C.R., Arroyo, I., Cohen, P.R., & Woolf, B.P. (2010). Evaluation of AnimalWatch: An Intelligent Tutoring System for Arithmetic and Fractions. *Journal of Interactive Online Learning, 9*, 64-77.

Benotti, L., Martínez, M. C., & Schapachnik, F. (2014). Engaging high school students using chatbots. In *Proceedings of the 2014 conference on Innovation & technology in computer science education* (pp. 63-68).

Brame, C. J. (2019). Flipping the classroom. *Science Teaching Essentials*, 121–132. https://doi.org/10.1016/b978-0-12-814702-3.00009-3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. & others (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165, .

Bourque, P., & Fairley, R. (2022). *Software Engineering. Guide to the Software Engineering Body of Knowledge (SWEBOK)*. https://doi.org/10.3403/30314312u

Chames, J. & Lieberman, L. (1965). Differences between normal and clinical groups in judging, evaluating and associating needs. *Journal of Clinical Psychology*, 21, 145-156.

Chassignol M, Khoroshavin A, Klimova A, Bilyatdinova A. Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science. 2018;136:16–24*. doi: 10.1016/j.procs.2018.08.233.

Combéfis, S. (2022). Automated code assessment for education: Review, Classification and perspectives on techniques and tools. *Software*, *1*(1), 3–30. https://doi.org/10.3390/software1010002

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cunningham-Nelson, S., Boles, W., Trouton, L., & Margerison, E. (2019). A review of chatbots in education: practical steps forward. In *30th annual conference for the australasian association for engineering education (AAEE 2019): educators becoming agents of change: innovate, integrate, motivate* (pp. 299-306). Engineers Australia.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage Publications.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Graesser, A. C., Conley, M. W., & Olney, A. (2012). *Intelligent tutoring systems. APA educational psychology handbook, Vol 3: Application to learning and teaching*., 451-473.

Hodges, D.A. & Jackson, H.G. (1983). *Analysis and design of digital integrated circuits.*  New York: McGraw.

Holmes, Wayne & Bialik, Maya & Fadel, Charles. (2019). Artificial Intelligence in Education. Promise and Implications for Teaching and Learning.

Johnson, A. (1996).  *The changing face of education.* Retrieved from http://www.stemnet.inf.ca/Community/Prospects/V3n3/tcfoe.htm.

Kirk, T. (2023, April 5). CHATGPT: Opportunities and challenges for Education. University of Cambridge. Retrieved April 13, 2023, from https://www.cam.ac.uk/stories/ChatGPT-and-education

Kurshan B. The future of artificial intelligence in education. Forbes Magazine; 2016.

Lewis, M., Bromley, K., Sutton, C. J., McCray, G., Myers, H. L., & Lancaster, G. A. (2021). Determining sample size for progression criteria for pragmatic pilot RCTS: The Hypothesis Test Strikes Back! *Pilot and Feasibility Studies*, *7*(1). https://doi.org/10.1186/s40814-021-00770-x

Luckin, R., Holmes, W., Grifths, M., & Forcier, L. B. (2016). Intelligence Unleashed. An argument for AI in education. Pearson.

MacNeil, S., Tran, A., Leinonen, J., Denny, P., Kim, J., Hellas, A., Bernstein, S., & Sarsa, S. (2022). Automatically Generating CS Learning Materials with Large Language Models. *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2.*

OpenAI. (n.d.). GPT-4.  https://openai.com/product/gpt-4

Passey, D., Brinda, T., Cornu, B., Holvikivi, J., Lewin, C., Magenheim, J., Morel, R., Osorio, J., Tatnall, A., Thompson, B., & Webb, M. (2021). Computers and education – recognising opportunities and managing challenges. *IFIP Advances in Information and Communication Technology*, 129–152. https://doi.org/10.1007/978-3-030-81701-5\_5

Pereira, J. (2016). Leveraging chatbots to improve self-guided learning through conversational quizzes. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*.

Qureshi, B. (2023, April 16). *Exploring the use of CHATGPT as a tool for learning and assessment in Undergraduate Computer Science Curriculum: Opportunities and challenges*. arXiv.org. https://arxiv.org/abs/2304.11214

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066x.55.1.68

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. In ICER 2022 - Proceedings of the 2022 ACM Conference on International Computing Education Research (pp. 27-43). ACM. https://doi.org/10.1145/3501385.3543957

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00292-9

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, 151, 103862.

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. https://doi.org/10.1145/2330601.2330661

VanLehn K, Graesser AC, Jackson GT, Jordan P, Olney A, Rose CP. When are tutorial dialogues more effective than reading? *Cognitive Science. 2007;31:3–62.* doi: 10.1080/03640210709336984.

Wang, Z., Lan, A.S., Nie, W., Waters, A.E., Grimaldi, P.J., & Baraniuk, R. (2018). QG-net: a data-driven question generation model for educational content. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale.*

Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2015). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research*, *25*(3), 1057–1073. https://doi.org/10.1177/0962280215588241

Willis, P. (1983).  Cultural production and theories of reproduction. In L. Barton & S. Walker (Eds).  *Race, class and education* (pp. 40-43).  San Francisco: Jossey-Bass.

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet?-A systematic literature review on chatbots in education. *Frontiers in artificial intelligence*, 4, 654924.

Younes, S. S. (2021). Examining the effectiveness of using adaptive AI-enabled e-learning during the pandemic of covid-19. *Journal of Healthcare Engineering*, *2021*, 1–14. https://doi.org/10.1155/2021/3928326

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on Artificial Intelligence Applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0171-0