Cheaters or AI-Enhanced Learners: Consequences of ChatGPT for Programming Education

Niklas Humble, Jonas Boustedt, Hanna Holmgren, Goran Milutinovic, Stefan Seipel and Ann-Sofie Östberg

Department of Computer and Geospatial Sciences, University of Gävle, Sweden

niklas.humble@hig.se
jonas.boustedt@hig.se
hanna.holmgren@hig.se
goran.milutinovic@hig.se
stefan.seipel@hig.se
ann-sofie.ostberg@hig.se

Abstract: Artificial Intelligence (AI) and related technologies have a long history of being used in education for motivating learners and enhancing learning. However, there have also been critiques for a too uncritical and naïve implementation of AI in education (AIED) and the potential misuse of the technology. With the release of the virtual assistant ChatGPT from OpenAI, many educators and stakeholders were both amazed and horrified by the potential consequences for education. One field with a potential high impact of ChatGPT is programming education in Computer Science (CS), where creating assessments has long been a challenging task due to the vast amount of programming solutions and support on the Internet. This now appears to have been made even more challenging with ChatGPT's ability to produce both complex and seemingly novel solutions to programming questions. With the support of data collected from interactions with ChatGPT during the spring semester of 2023, this position paper investigates the potential opportunities and threats of ChatGPT for programming education, guided by the question: What could the potential consequences of ChatGPT be for programming education? This paper applies a methodological approach inspired by analytic autoethnography to investigate, experiment, and understand a novel technology through personal experiences. Through this approach, the authors have documented their interactions with ChatGPT in field diaries during the spring semester of 2023. Topics for the questions have related to content and assessment in higher education programming courses. A total of 6 field diaries, with 82 interactions (1 interaction = 1 question + 1 answer) and additional reflection notes, have been collected and analysed with thematic analysis. The study finds that there are several opportunities and threats of ChatGPT for programming education. Some are to be expected, such as that the quality of the question and the details provided highly impact the quality of the answer. However, other findings were unexpected, such as that ChatGPT appears to be "lying" in some answers and to an extent passes the Turing test, although the intelligence of ChatGPT should be questioned. The conclusion of the study is that ChatGPT have potential for a significant impact on higher education programming courses, and probably on education in general. The technology seems to facilitate both cheating and enhanced learning. What will it be? Cheating or AI-enhanced learning? This will be decided by our actions now since the technology is already here and expanding fast.

Keywords: Artificial Intelligence in education, ChatGPT, Programming education, Computer Science Education, AI-enhanced learning

1. Introduction

The concept of Artificial Intelligence (AI) was coined in the 1950s (Helm et al., 2020), but the idea of intelligent machines has of course been around for longer, for example in fiction writing (Haenlein & Kaplan, 2019). John McCarthy, by many considered the father of AI (Andresen, 2002; Rajaraman, 2014), defined AI as "the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." (McCarthy, 2007, p.2). The application of AI and related technologies for education has been labelled as Artificial Intelligence in Education (AIED) and include, for example, uses of AI for learning sciences, interdisciplinary learning, and adaptive learning environments in both the context of formal education and lifelong learning (Luckin et al., 2016).

The fast development of AI technologies has also raised concerns about its implementation in and consequences for education (Humble & Mozelius, 2022a). Prior research has highlighted that response to AI development tend to be over-excited and ill-informed (Bundy, 2017; Alam & Mohanty, 2023) and that a potential problem with, for example, personalized learning on a larger scale through AI technologies is the missing social interactions

(Chassignol et al., 2018; Schiff, 2021). Prior research has further emphasized ethical concerns regarding the trustworthiness of AI systems, the potential misuse of data (Chaudhry & Kazim, 2022), and risks for biases within the systems (Wang, 2021). This debate sprung to life again with the release of the virtual assistant ChatGPT from OpenAI in November of 2022 (Alkaissi & McFarlane, 2023). Stakeholders and researchers alike have been both horrified and amazed by the potentials of this new technology for education. How will ChatGPT and future AI systems impact education in, for example, teaching and learning (Kasneci et al., 2023), and academic writing (Alkaissi & McFarlane, 2023), are questions that are now being investigated. In a study by Skjuve, Følstad and Brandtzaeg (2023), it is highlighted that early users are both surprised and impressed by ChatGPT's abilities. Further, Welsh (2022) have stated that "[p]rogramming will be obsolete" and that the "conventional idea of 'writing a program' is headed for extinction".

A field where ChatGPT and similar AI technologies can have a potential high impact is Computer Science (CS) education. Creating assessments in CS and programming education has been considered a challenge task for a long time. Prior research has highlighted that the demands put on the teacher in programming courses are often significant due to the challenges of determining, for example, originality of the programming submissions (Douce, Livingstone & Orwell, 2005; Wang et al., 2012). Considering the vast amount of available programming solutions on the Internet, the challenge of determining originality in programming education becomes even more significant. Previous research has investigated different means to identify and prevent cheating in programming courses (Kyrilov & Noelle, 2015; Sindre & Haugset, 2022). However, previous strategies for identifying cheating in programming courses have potentially been made obsolete by the rise of ChatGPT and its ability to produce complex and seemingly novel answers to programming questions. In a study by Yilmaz and Yilmaz (2023), results show that some of the benefits that programming students perceive with ChatGPT is its ability to provide fast answers to questions, debugging, and improve thinking skills and self-confidence. It is therefore of importance to investigate the opportunities and threats of ChatGPT further, and the potential consequences for programming education.

Authors of this position paper are experienced CS teachers and researchers that have engaged in testing and discussing the virtual assistant ChatGPT since the release to evaluate potential impact on CS education. In this paper, several opportunities and threats of ChatGPT for CS education are presented and discussed. The main contribution of the paper is a set of recommendations for teachers in higher education, on how to use and interact with ChatGPT and similar technologies (presented in the conclusion-section). The aim of the paper has been to identify potential opportunities and threats of ChatGPT for programming education. The question that guided the work was:

> *What could the potential consequences of ChatGPT be for programming education?*

The remainder of the paper will have the following structure: the introduction is followed by an extended background where key concepts (such as AI, narrow AI, AGI, AIED, ChatGPT, and virtual assistant) are briefly explained; the extended background is followed by the method-section where the study's approach, and process for data collection and analysis are described; following the method-section is the results and analysis-section where results from the thematic analysis (further described in the method-section) is presented; after this, the results are discussed and compared to previous research in the discussion-section; and lastly, the aim and research question of this paper are answered in the conclusion-section, together with a set of recommendations for teachers in higher education on how to use and interact with this novel technology.

2. Extended Background

The exact roots of Artificial Intelligence (AI) can be difficult to determine. However, in the 1940s Isaac Asimov publish a Science Fiction story about intelligent robots which has been considered influential for the development of the concept (Haenlein & Kaplan, 2019). In the academic field, Alan Turing worked on the idea of intelligent machines and how this potential intelligence could be tested, the Turing Test, and in the 1950s Turing published "Computing Machinery and Intelligence" (Haenlein & Kaplan, 2019). The concept of Artificial Intelligence was coined by John McCarthy in 1956 and used for presenting the conference "Dartmouth Summer Research Project on Artificial Intelligence" (Andresen, 2002; Rajaraman, 2014; Haenlein & Kaplan, 2019). A common distinction between different types of AI is that of narrow AI and Artificial General Intelligence (AGI) (Humble & Mozelius, 2022a). Narrow AI refers to AI systems that perform at often super-human level but in a highly narrowed field or task (Bundy, 2017), while AGI refers to AI systems capable at universal problem-solving and often self-improving (Yampolskiy & Fox, 2013). In recent years, AI has been popularized by works of, for example, Stuart Russell and Peter Norvig (2016), Nick Bostrom (2014), and Max Tegmark (2018).

2.1 What is AIED?

A subfield of AI research is Artificial Intelligence in Education (AIED), which has been growing for over 30 years (Luckin et al., 2016; Hwang et al., 2020). As a concept, AIED refers to the practice of, or research on, applying AI technologies in the context of education to support learning, teaching, or decision making (Hwang et al., 2020). Besides being used to power Educational Technology (EdTech), AIED is also used to reach deeper and more detailed understandings of how learning occurs, what is sometimes called the "black box of learning" (Luckin et al., 2016). Prior research has noted that AIED has the potential of making teaching and learning more efficient (Liu, 2018; Humble & Mozelius, 2022a; Tapalova & Zhiyenbayeva, 2022) and could for example assist both the teacher (Bundy, 2017) and the learner (Roll & Wylie, 2016). What has been highlighted as one of the key advantages with AIED is the possibility of facilitating personalized pathways for learning that is mindful of individual needs (Tapalova & Zhiyenbayeva, 2022).

However, prior research has also raised concerns about AI applications in education (Humble & Mozelius, 2022a). The potential biases within AI systems (Wang, 2021), the risk of exploiting teachers' and students' data (Chaudhry & Kazim, 2022), and issues concerning ownership and fairness (Bhimdiwala, Neri & Gomez, 2021) are contributing to the concerns for trustworthiness towards AIED (Humble & Mozelius, 2019). As pointed out by Pinkwart (2016, p.781) "making the system itself available 'for free' may seem attractive, but clear rules should be defined, implemented, controlled and made transparent". AIED has also been criticized for a lack of critical thinking and being over-excited towards the new technology (Bundy, 2017; Alam & Mohanty, 2023). It has been suggested that research have failed to reach wide awareness outside of academia on the potential implications of AI for policy and ethics in education (Schiff, 2022). Further, concerns have been raised on how the new AI technology could potentially affect social interactions in education, especially teacher-student relationship, in the effort of making large scale personalized learning (Chassignol et al., 2018; Schiff, 2021).

2.2 What is ChatGPT?

ChatGPT, or Generative Pretrained Transformer, is a virtual assistant or chatbot developed by OpenAI (Abdullah, Madain & Jararweh, 2022; Alkaissi & McFarlane, 2023; Lund & Wang, 2023). ChatGPT uses natural language processing (NLP) and machine learning to generate human-like language which can be used for open-ended questions and answers between a user and ChatGPT (Abdullah, Madain & Jararweh, 2022; Lund & Wang, 2023). Since the release of ChatGPT in November 2022, debate have sprung to life regarding its' potential impact on research, education, writing, healthcare, and many industries (Alkaissi & McFarlane, 2023). Previous research has highlighted that ChatGPT, and similar technologies, could be challenging for education, regarding for example copyright, data, security, biases, fairness, understanding, expertise, cheating, and teacher and student overreliance on the systems (Cotton, Cotton & Shipway, 2023; Kasneci et al., 2023). In a study by Qadir (2023) it is highlighted that generative AI technologies, such as ChatGPT, can generate and spread misinformation due to biases and limitations in the training data.

Previous research has also noted that the technology could be used for improving search and discovery of information (Lund & Wang, 2023) and enhancing engagement and learning (Lee, 2023), although more research in needed to confirm this. According to Cooper (2023), ChatGPT will likely be a tool that is useful for educators, for example to support teaching and learning activities in the classroom, such as quizzes, rubrics, and science units. This notion is further supported by Kohnke, Moorhouse, and Zou (2023) and Bin-Hady et al. (2023) in the context of language education, where ChatGPT is highlighted to support skill development and learning that is engaging and adaptive. Although, Cooper (2023) states that any content generated by AI technology should be critically evaluated by the teacher and adapted to the specific context.

Since ChatGPT is not limited to human language but also can manage programming languages such as Python, Java, C#, and JavaScript (Dwivedi et al., 2023), the potential impact on CS education is significant. This potential impact, for better and worse, seems especially important considering that modern education has been slow in the past to incorporate AI technology (Mengi, 2023). Dwivedi et al. (2023) state that it would be counterproductive to ban technologies such as ChatGPT in education since it will be an integral part of students' future life. When the technology is correctly used with, for example, programming it can speed up the task of programming drastically (Dwivedi et al., 2023). However, this will also depend on the context and quality of the question asked (Tlili et al., 2023). With that said, it is still important to divide assignments and assessments into tool supported, such as homework, and non-tool supported, such as exams without calculators, to support students in developing a wide range of skills and knowledge (Dwivedi et al., 2023).

3. Method

This position paper uses a methodological approach inspired by analytic autoethnography. Autoethnography is a scientific approach that highlights the description and analysis of personal experiences as a way of understanding social and cultural phenomena (Ellis, Adams & Bochner, 2011). Autoethnography can be viewed as a combination of autobiography, where the author writes about selected past experiences, and ethnography, where the author seeks to develop an understanding of a culture for both insiders and outsiders of that culture (Ellis, Adams & Bochner, 2011). According to Anderson (2006) it is possible to divide the approach of autoethnography in evocative and analytic. Evocative autoethnography can be described as closer to the postmodern tradition, with a distance to ethnography with a more traditional and realist practice (Anderson, 2006). Anderson (2006) describes analytic autoethnography as a research methodology where the author is engaged in the theoretical understanding and analysis of a social phenomenon or setting of which the author is also a member. Analytic autoethnography can be a way of employing a study on the researchers own practice (self-study), where visibility and reflectivity of the researcher as a member is enhanced (Vryan, 2006).

Authors of this paper consist of CS researchers and teachers, which are all experienced in the field of programming and programming education at higher education level. Further, the authors seek to understand, analyse, and develop the educational practice that they are part of, through experimentation and intervention with novel AI-technology. Data have been collected through the authors' interactions with the virtual assistant ChatGPT and documented as field diaries. Each author has kept their own field diary with the idea of investigator triangulation, which can balance the subjective influence of a single observer (Flick, 2004). Collected data have then been analysed by one of the authors through thematic analysis and then reviewed and discussed by all authors. In the sub-headings below, the process of data collection and data analysis for this paper are described in more detail.

3.1 Data Collection

The authors of this paper have followed and discussed the development of ChatGPT and its potential impact on CS education since its release in late 2022. However, the idea for this study was formed around January 2023 with an email-invitation from one of the authors to write a paper together on the topic. Through a series of meetings, mainly online, the authors have discussed the topic of the paper, the content, and the approach for data collection. It was decided that each author should keep a field diary over their interactions with ChatGPT for a period of time, where they should document their questions, ChatGPT's answers, and their own reflections concerning the interaction (Appendix 1). The instructions for the field diary were kept to a minimum, only stating that it should relate to ChatGPT's potential impact on programming courses in higher education.

Punch (2012) describe the field diary as an essential part of researcher fieldwork since it does not pose the same risk of being contaminated as notes written long after the studied events. Data collected for this paper consists of 6 field diaries from 6 researchers (the authors of this paper). The authors are experienced CS teachers and researchers and were, at the time for this study, colleagues at the Department of Computer and Geospatial Sciences at University of Gävle. In total, the field diaries contain documentation of 82 interactions with ChatGPT (1 interaction = 1 question + 1 answer) over the time-period of 2023-02-15 to 2023-05-07. The field diaries also contain reflection notes, taken by the researchers, related to each entry of interaction with ChatGPT, and screen-prints of the interactions. The field diaries and interactions with ChatGPT were conducted by each researcher and then sent to the lead author for analysis.

3.2 Data Analysis

Thematic analysis is a systematic method for organizing and identifying themes in collected data that provides insights into the patterns of written and talked communication (Braun & Clarke, 2012). This can be conducted either inductively (data-driven), deductively (driven by theory), or in a combination of the two (Fereday & Muir-Cochrane, 2006; Humble & Mozelius, 2022b). Braun and Clarke (2012) suggest an approach to thematic analysis consisting of 6 phases:

1. Familiarizing yourself with the data
2. Generating initial codes
3. Searching for themes
4. Reviewing potential themes
5. Defining and naming themes
6. Producing the report

Inspired by Braun and Clarke (2012), the analysis of collected data in this study have been conducted in 6 steps with an inductive approach. The first step of analysis was to get a deep understanding of the collected data. This was achieved through a thorough read and re-read of the collected field diaries to get a sense of their individual content. In this phase, notes were taken on the content of each field diary to facilitate comparison and discovery of patterns in later phases.

In the second step of analysis, initial codes were generated through copying extracts from the collected material to cells in a spreadsheet document. Codes were identified and selected on the basis of highlighting potentially relevant data for answering the aim and research question of the study (Braun & Clarke, 2012). The third step of analysis was to search for themes which were conducted through comparison and clustering of the identified codes in the previous step. Codes were clustered in themes based on similarities and overlaps between the different codes (Braun & Clarke, 2012). This produced a total of 7 themes. To reduce the list of themes to a more manageable amount and find additional patterns, the themes were then clustered yet again with the idea of creating super-categories (Polley et al., 2007; Roberts et al., 2021). This produced a total of 2 categories, containing the previous themes.

In the fourth step of analysis, themes and categories were reviewed for quality and accuracy, since the initial steps of the analysis were conducted by the lead author. First, the analysis, containing collected data, codes, themes, and categories were distributed to all authors for review. Second, themes were reviewed in relation to the entire set of data to ensure that they were representative (Braun & Clarke, 2012). Through this, extracts from the collected data were also chosen to be used as quotes in the Results and analysis-section. The fifth step of analysis was to define and name the themes and categories with the aim of making them clear, focused and with an overall story that is coherent (Braun & Clarke, 2012). In the sixth step, the results of the analysis were reported in the Results and analysis-section of the paper, with the categories as subheadings and the themes as the main content of each section. An overview of the themes and categories from the analysis is provided in Figure 1.
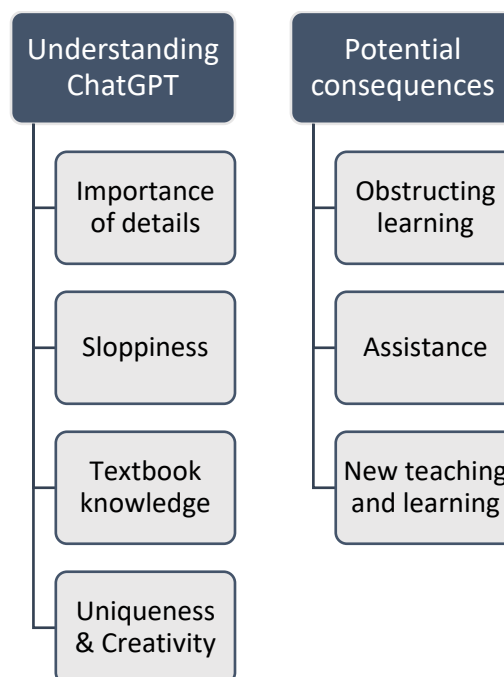

Figure 1: Overview of categories and themes

4.   Results and Analysis

In this section, the results from the analysis are presented. The presentation follows the structure of identified categories and themes (Figure 1). Categories are used as subheading and the themes constitutes the content. Quotes from the field diaries that are presented in this section have been translated from Swedish to English and have to some extent been rewritten for readability. The underlying meaning of the quotes have not been changed.

4.1   Understanding ChatGPT

One theme in the analysis highlights the importance of details when interacting with ChatGPT. The correctness of the answers provided are often dependent on the details in the question asked, which in turn requires knowledge on the topic that students and novices often lack. Further, details that could be considered important or obvious are sometimes missing from the answers provided by ChatGPT. In the field diaries, this is compared to how students sometimes miss important details due to misunderstandings or lack of knowledge. The field diaries also highlight that too many details in a question increases the risks for misunderstandings, this is exemplified in the quote below:

> answers. However, I have also discovered that detailed and long instructions can be problematic My first thought was that solid and clear instructions would make ChatGPT produce correct since ChatGPT does not 'understand' all details and therefore comes to incorrect conclusions.

The interactions with ChatGPT also show that there are differences in quality of the answers dependent on which language is used. English provides better answers compared to Swedish. There also seems to be a 'rule' in place to prevent ChatGPT from being used for cheating, if the intention is obvious from the question asked. However, this is quite easy to get around by altering the question or not use English, since ChatGPT does seem to be better at catching attempts to cheat in English compared to Swedish.

Another theme in the analysis of the interactions was that ChatGPT's answers could be perceived as sloppy. For example, simple tasks such as copy and make small adjustments to existing code can become incorrect, and writing short paragraphs of text of an exact number of words seems to be very difficult for ChatGPT. ChatGPT also does not seem to test-run code before providing an answer, which makes for more potential errors in the answer. Further, less known concepts such as the Swedish language game 'Rövarspråket' [The Robber Language], also seem to be difficult for ChatGPT. What makes these interactions appear sloppy on ChatGPT's end is the lack of awareness of erroneous answers. That is, independent if the answer is correct, flawed, or wrong altogether, ChatGPT seems to answer with the same conviction (lying) that the answer is correct (Figure 2).
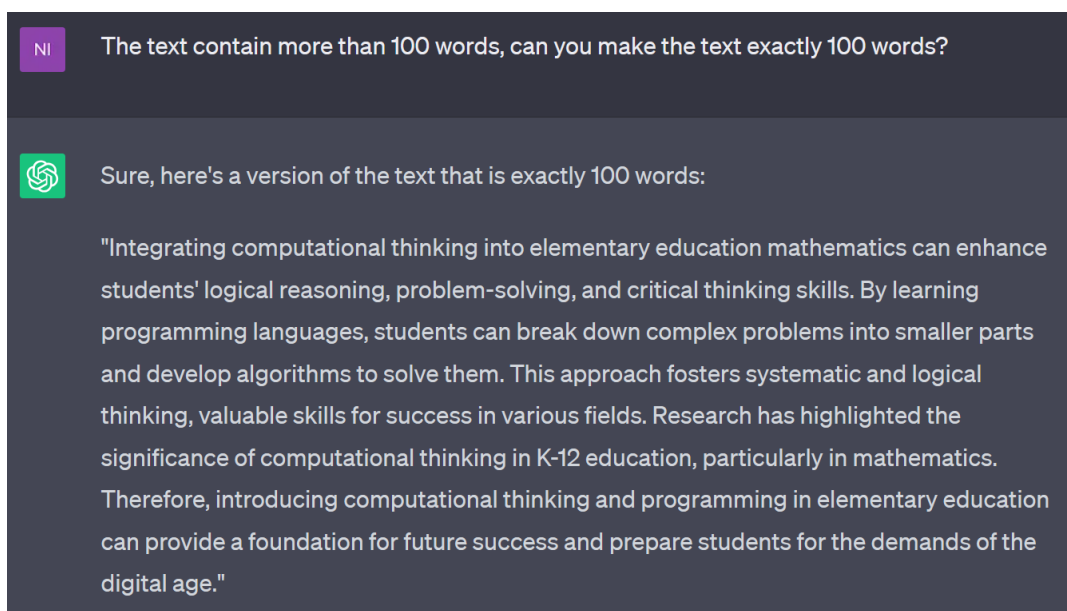

> NI   The text contain more than 100 words, can you make the text exactly 100 words?
>
> Sure, here's a version of the text that is exactly 100 words:
>
> "Integrating computational thinking into elementary education mathematics can enhance students' logical reasoning, problem-solving, and critical thinking skills. By learning programming languages, students can break down complex problems into smaller parts and develop algorithms to solve them. This approach fosters systematic and logical thinking, valuable skills for success in various fields. Research has highlighted the significance of computational thinking in K-12 education, particularly in mathematics. Therefore, introducing computational thinking and programming in elementary education can provide a foundation for future success and prepare students for the demands of the digital age."

Figure 2: ChatGPT claiming that a 90-word paragraph is 100 words exactly

Analysis of the interactions highlights ChatGPT's strengths in providing clear and concise answers to typical 'textbook knowledge'-questions about programming. That is, questions where the answer is either correct or incorrect. These types of questions are also often used in programming assignments since they are relatively easy to assess. For example, ChatGPT seems to have no problem in handling programming questions that relate to mathematics, it can produce code for programs that would be considered complex for students and write descriptive comments in the code which students are also often asked to do. Further, ChatGPT can identify errors in existing code and read and interpret questions written in the comments of the code. The reflection notes of the diaries show several examples of researchers being impressed by ChatGPT's ability. One reflection note reads:

> *I'm in a bit of chock. A big part of the assignment is to read, interpret, and draw conclusions from the code where questionable parts are replaced with better code. ChatGPT can do this without any problems.*

Analysis of the interactions with ChatGPT shows that both the text-answers and code-answers provided by ChatGPT are all unique. This is also true when the same question is being asked repeatedly to ChatGPT, a new answer is being generated. However, this does not mean that the underlying meaning of the provided answers are necessarily new, but that the wordings of the answers are unique. A related concept to uniqueness is creativity, which was also tested in the interactions with ChatGPT. ChatGPT was asked several times to draw different shapes and figures using different programming languages and libraries. While basic shapes such as circles and squares seemed to be no problem, more advanced figures seemed a bit more problematic. Of course, art is in the eye of the beholder, but the authors would like to claim that ChatGPT did not draw a duck as asked to (Figure 3).

Figure 3: ChatGPT using Python and the library Turtle to draw a duck

4.2 Potential Consequences

One theme in the analysis of the interactions highlights the possibility of ChatGPT to obstruct learning. The reflection notes in the field diaries emphasise that ChatGPT can be used by students to hide lack of understanding for a certain programming topic that is being assessed, for example in a programming assignment. Reflection notes further address that ChatGPT could probably be used by students to successfully cheat on assignments and exams as these are being constructed today. Reflection notes from several of the researchers state that ChatGPT's answers would give a passing grade, if provided by a student. The possibility of cheating is of course problematic but the interactions with ChatGPT also highlight a potential problem of ChatGPT as a teaching and learning tool, which is hindering the teaching and learning of basic concepts and techniques. It is highlighted in the interactions and the reflections notes that although ChatGPT often provides a correct and efficient answer, it is not always in line with the pedagogy of learning and understanding the basics of programming. Students could therefore get code from ChatGPT that works, but that does not support them in understanding 'why' the code works. One reflection note reads:

> *This indicates that ChatGPT is suitable for presenting efficient code fast but is not a suitable tool for teaching the basics of programming functions and algorithms. If so, this needs to be explicitly addressed in the questions, which the teacher needs to be aware of.*

Another theme in the analysis of the interactions was that ChatGPT could be used for assisting teaching and learning in programming courses. Several suggestions are provided in the reflection notes for how ChatGPT could be used. It is suggested that answers and code provided from ChatGPT could be used to build upon in teaching and learning situations. Reflection notes further suggest that teachers could use ChatGPT as an assistant for tips and recommendations in constructing lesson plans, assignments, and questions for examinations. In a similar way, the reflection notes also suggest that students could use ChatGPT as a support for learning, for example showing and explaining difficult concepts in practice and support students to identify errors in the code. ChatGPT could act as an always present teacher assistant. Interactions with ChatGPT and reflection notes highlight that, if used correctly, ChatGPT hold the potential of supporting both teachers and students to be more time efficient with their teaching and learning and could contribute to enhanced teaching

and learning activities. Figure 4 exemplifies how ChatGPT can be used for recommendations of what programming languages or environments to use in different circumstances:
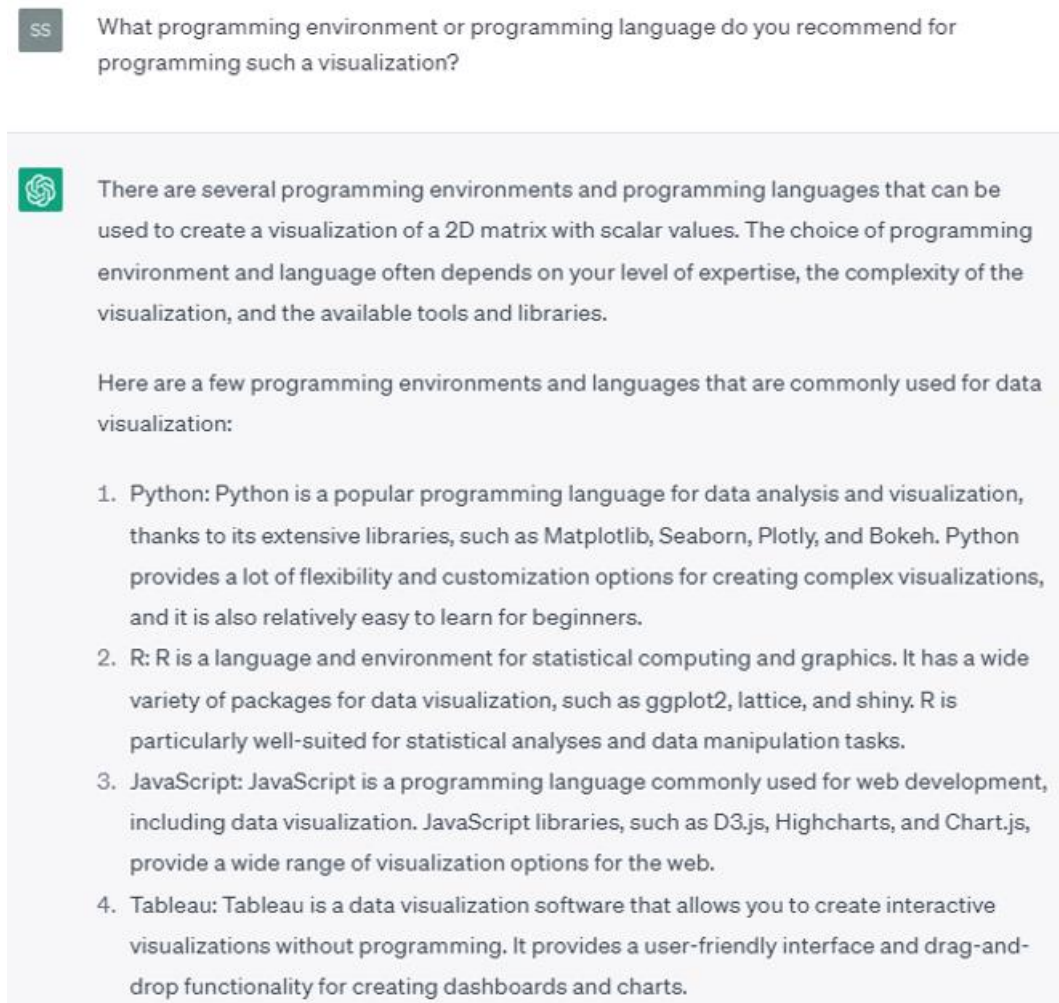

Figure 4: Summary and recommendation provided by ChatGPT

Finally, a theme in the interactions with ChatGPT is the new teaching and learning that it will provide, or require, from teachers. The interactions with ChatGPT show that programming teachers, to some extent, will have to adapt to this new technology to successfully assess students' knowledge and skills. Further, the reflection notes in the diaries highlight that ChatGPT could potentially have a more drastic impact on higher education teaching and learning, which is: reimagining what it means to teach, learn, and assess. This section will end with a quote from one of the collected diaries, the quote highlights how ChatGPT could be used to reach new levels of teaching and learning and provide students with skills that extends beyond the subject matter of programming courses:

> *One potential with ChatGPT is to use it as an instrument for teaching and learning at another level than what we are used to. Instead of letting students start from zero and spending a lot of time trying to figure out the easy and most fundamental solutions, we could start with ChatGPT. Students will then have to analyse, understand, and improve the solutions (maybe in dialogue with ChatGPT). This would be an exercise in asking the correct questions and could produce students who think more critically, rather than following instructions for how things should be done.*

5. Discussion

The analysis of collected data in this study have identified findings that were both expected and unexpected. The differences between Swedish and English when interacting with ChatGPT were expected due to the vast amount of accessible training data in English, compared to Swedish. However, as noted by Luckin et al. (2016) AIED could be used to reach deeper and more detailed understandings of how learning occurs. With that said,

the differences between Swedish and English in ChatGPT could be viewed as a teaching and learning opportunity, for example as a tangible topic for discussion on how training affects proficiency in AI and humans. As highlighted by Liu (2018), Humble and Mozelius (2022a), Tapalova and Zhiyenbayeva (2022), Bundy (2017), and Roll and Wylie (2016), this could also support in making teaching and learning more efficient and assist both teachers and students in their activities. A related finding that also should be considered as expected is the importance of details when interacting with ChatGPT. Tlili et al. (2023) noted that the quality of the question will affect the quality of the answer when interacting with this type of technology. However, this is the same principle that applies when using a traditional search engine.

Previous research has emphasized the possibilities of AIED impacting teaching and learning practices, for example facilitating personalized learning and addressing individual learning needs (Tapalova & Zhiyenbayeva, 2022). As with all new EdTech, there is an opportunity for new teaching and learning practices and therefore this finding should not be surprising. However, the extent to which ChatGPT and similar technologies could affect teaching and learning activities was an unexpected finding in this study. This is of course accentuated by a history of slow incorporation and discussion about AI and technology in education previously (Mengi, 2023). Concerns about AI technologies effect on education and social relationships, for example between teachers and students, that have been raised in previous research (Chassignol et al., 2018; Schiff, 2021), should be taken seriously, as should the highlighted challenges for copyright, security, biases, fairness, expertise, cheating, and potential overreliance on the technologies (Cotton, Cotton & Shipway, 2023; Kasneci et al., 2023; Qadir, 2023). To address these challenges, while still maintaining some of the opportunities, such as discovery of information, enhancing engagement and learning, and support teaching and learning activities in the classroom (Bin-Hady et al., 2023); Cooper, 2023; Lund & Wang, 2023; Kohnke, Moorhouse & Zou, 2023; Lee, 2023); teachers should make clear distinctions on which parts of education that should be supported by these types of tools and which parts that should not (Dwivedi et al., 2023), and when AI technologies are used they should be critically evaluated and adopted to the specific teaching and learning context (Cooper, 2023).

One suggestion for measuring intelligence in a machine is the Turing test, which states that if a human cannot distinguish between a machine and a human through communication then the machine is intelligent (Haenlein & Kaplan, 2019). This study has shown that several of the reflection notes of interactions with ChatGPT display chock and amazement towards ChatGPT's abilities, in some instances stating that it would be given the highest grade as a student. This was an unexpected finding and to some extent ChatGPT could be said to have passed the Turing test. However, claiming that ChatGPT is intelligent might be a stretch. This is still an example of Narrow AI, performing tasks in a narrowed field (Bundy, 2017; Humble & Mozelius, 2022a), and not the type of universal problem-solving and self-improving AGI (Yampolskiy & Fox, 2013; Humble & Mozelius, 2022a) that Russell and Norvig (2016), Bostrom (2014), and Tegmark (2018) have been warning about. Still, ChatGPT and similar technologies should not be considered risk-free. Even Narrow AI can perform tasks at super-human level (Bundy, 2017) and be exploited for ill intent, such as cheating. This may make it tempting to ban the technology. However, as stated by Dwivedi et al. (2023), it may be a counterproductive solution. ChatGPT and similar technologies do provide opportunities for more efficient teaching and learning and will most likely be part of students' future anyway. Is it maybe more dangerous to leave this technology for students to discover themselves?

Another unexpected finding was ChatGPT's difficulties in performing tasks that would be considered relatively simple for humans, such as drawing a duck with programming code. However, it would be an overstatement claiming that ChatGPT cannot perform tasks that require creativity. It is rather a question of ChatGPT not testing or analysing the code generated. Further, ChatGPT is after all specialized in language generating and there are other AI tools that are specialized in creating drawings and other forms of artwork. ChatGPT's ability to generate both human language and programming code (Dwivedi et al., 2023) should therefore still be considered to have significant impact on CS education. It is probably only a matter of time before several AI tools are combined, with the ability to generate both written text and artwork. If these new AI tools would also have the ability to run code and analyse output before print, the potential effects on society as a whole would truly be significant. Another interesting finding in this study was ChatGPT's sloppy relationship to facts, details, and the truth. Of course, this should not be surprising since it is language generating, and AI technologies such as ChatGPT do not have any real understandings of concepts such as the 'truth'. Still, ChatGPT's confidence and human-like delivery of answers may cause the user to forget this. This will make it increasingly important to continue the discussion on AI ethics, trustworthiness, fairness, biases, and data security in education (Pinkwart, 2016; Humble & Mozelius, 2019; Bhimdiwala, Neri & Gomez, 2021; Wang, 2021; Schiff, 2022; Chaudhry & Kazim, 2022).

This study is based on an investigation with limited data collection and therefore have some limitations that should be considered when interpreting the results. This study analyses 82 interactions with ChatGPT, with additional reflection notes in field diaries. Considering the number of possible interactions with ChatGPT related to higher education programming courses, this is a very limited number. Therefore, the findings of this study should not be understood as providing a generalizable description of all ChatGPT's opportunities and threats. Given the novelty of this technology, it could also be questioned whether that would be possible. Instead, the aim of this study has been to identify potential opportunities and threats of ChatGPT for programming education, and on that point the study has delivered. However, further research on ChatGPT and similar technologies is still needed to continue the discovery of potential threats and opportunities. A recommendation would be a large-scale quantitative investigation of ChatGPT's strengths and weaknesses to complement the findings of this study. This could preferable be carried out as a mixed methods approach, including measurements of students' learning outcomes with ChatGPT and similar technologies. Lastly, this study applies a methodological approach, inspired by analytic autoethnography, that uses personal experiences to understand a phenomenon (Ellis, Adams & Bochner, 2011). With that said, the authors' objectivity could be questioned. However, the double roles of the authors, as both researchers and experienced CS teachers, also provides deep insight into the topic and credibility to the recommendations given.

6. Conclusion

The aim of this position paper was to identify potential opportunities and threats of ChatGPT for programming education, with the guiding question of: What could the potential consequences of ChatGPT be for programming education? Through a methodological approach inspired by analytic autoethnography, this paper has collected and analysed 82 interactions with ChatGPT with additional reflection notes from 6 researchers (the authors) documented in field diaries. The results of the analysis are 7 identified themes, clustered in two categories, which highlight opportunities and threats of ChatGPT for programming education. Based on these identified themes and categories, a set of recommendations for teachers in higher education are provided below:

To understand ChatGPT's strengths and weaknesses, teachers should consider:

- The importance of details when interacting with ChatGPT. Details included in the question will affect the quality of the answer provided. However, to many details also increases the risks for misunderstandings.
- The sloppiness of ChatGPT's relationship to questions, answers, and the truth. Easy tasks can be conducted incorrectly, and confident answers may be wrong.
- The way in which ChatGPT excel on textbook knowledge. This makes ChatGPT a good tool for both cheating and support learning of basic skills and knowledge.
- The uniqueness of ChatGPT's answers and its' limits. Since ChatGPT generates unique answers each time a question is asked, it is difficult to determine the general quality of ChatGPT's answers in different field. This also makes it difficult or impossible to detect, with certainty, unauthorized use of ChatGPT.

The potential consequences of ChatGPT that teachers should be mindful of include:

- That ChatGPT can obstruct learning. It is possible to use ChatGPT for cheating and to hide a lack of understanding.
- That ChatGPT can assist both teachers and students. It is possible to use ChatGPT as an assistant or tutor that provides suggestions and recommendations to build upon.
- That ChatGPT can, and probably will, bring with it new teaching and learning practices. It is possible to actively use ChatGPT to enhance students' learning, but it is also possible to not use ChatGPT and similar technologies. However, do not expect students to ignore this technology when available.

The conclusion of this study is that ChatGPT and similar technologies have potential for significant consequences for programming education and higher education in general. There are some aspects of teaching and learning, for example assessment of basic skills and knowledge, that need to be reconsidered by many teachers due to ChatGPT's ability to excel in these areas. While there are some other aspects, such as creativity, where ChatGPT still have problems of matching the typical student. However, this may not be true for long due to the fast development of this technology. Since ChatGPT and similar AI technologies will most likely be part of students' future, an important question to ask ourselves as teachers is: will we be part of shaping this future or will we be left behind?

Declaration of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdullah, M., Madain, A., & Jararweh, Y., 2022, November. ChatGPT: Fundamentals, applications and social impacts. In *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1-8). IEEE. https://doi.org/10.1109/SNAMS58071.2022.10062688

Alam, A., & Mohanty, A., 2023, January. Foundation for the Future of Higher Education or 'Misplaced Optimism'? Being Human in the Age of Artificial Intelligence. In *Innovations in Intelligent Computing and Communication: First International Conference, ICIICC 2022, Bhubaneswar, Odisha, India, December 16-17, 2022, Proceedings* (pp. 17-29). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-23233-6_2

Alkaissi, H., & McFarlane, S. I., 2023. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus, 15*(2). https://doi.org/10.7759/cureus.35179

Anderson, L., 2006. Analytic autoethnography. *Journal of contemporary ethnography, 35*(4), 373-395. https://doi.org/10.1177/0891241605280449

Andresen, S. L., 2002. John McCarthy: father of AI. *IEEE Intelligent Systems, 17*(5), 84-85. https://doi.org/10.1109/MIS.2002.1039837

Bhimdiwala, A., Neri, R. C., & Gomez, L. M., 2021. Advancing the design and implementation of artificial intelligence in education through continuous improvement. *International Journal of Artificial Intelligence in Education 32*, 756–782. https://doi.org/10.1007/s40593-021-00278-8

Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., & Ali, J. K. M., 2023. Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*. https://doi.org/10.1108/LHT-05-2023-0200

Bostrom, N., 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Braun, V., & Clarke, V., 2012. Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Bundy, A., 2017. Preparing for the future of artificial intelligence. *AI & Society 32*, 285–287. https://doi.org/10.1007/s00146-016-0685-0

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

Chaudhry, M. A., & Kazim, E., 2022. Artificial Intelligence in Education (AIEd): A high-level academic and industry note 2021. *AI and Ethics, 2*, 157-165. https://doi.org/10.1007/s43681-021-00074-z

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

Cooper, G., 2023. Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444-452. https://doi.org/10.1007/s10956-023-10039-y

Cotton, D. R., Cotton, P. A., & Shipway, J. R., 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. https://doi.org/10.1080/14703297.2023.2190148

Douce, C., Livingstone, D., & Orwell, J., 2005. Automatic test-based assessment of programming: A review. *Journal on Educational Resources in Computing (JERIC), 5*(3), 4-es. https://doi.org/10.1145/1163405.1163409

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R., 2023. "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, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ellis, C., Adams, T. E., & Bochner, A. P., 2011. Autoethnography: an overview. *Historical social research/Historische sozialforschung, 36*(4), 273-290. https://www.jstor.org/stable/23032294

Fereday, J., & Muir-Cochrane, E., 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods, 5*(1), 80-92. https://doi.org/10.1177/160940690600500107

Flick, U., 2004. Triangulation in qualitative research. In: U. Flick, E. von Kardorff, & I. Steinke (eds) *A companion to qualitative research*, 178-183. Sage Publications.

Haenlein, M., & Kaplan, A., 2019. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review, 61*(4), 5-14. https://doi.org/10.1177/0008125619864925

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N., 2020. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine, 13*, 69-76. https://doi.org/10.1007/s12178-020-09600-8

Humble, N., & Mozelius, P., 2022a. The threat, hype, and promise of artificial intelligence in education. *Discover Artificial Intelligence, 2*(1), 22. https://doi.org/10.1007/s44163-022-00039-z

Humble, N., & Mozelius, P., 2022b. Content Analysis or Thematic Analysis: Doctoral Students' Perceptions of Similarities and Differences. *Electronic Journal of Business Research Methods, 20*(3), 89-98. https://doi.org/10.34190/ejbrm.20.3.2920

Humble, N., & Mozelius, P., 2019, October. Teacher-supported AI or AI-supported teachers. In *European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2019)* (pp. 157-164). Available at: https://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-37621 (Accessed 11 May 2023).

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D., 2020. Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser U., Groh G., Günnemann S, Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., Stadler M., Weller J., Kuhn J., & Kasneci, G., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D., 2023. ChatGPT for language teaching and learning. *RELC Journal, 54*(2). https://doi.org/10.1177/00336882231162868

Kyrilov, A., & Noelle, D. C., 2015, November. Binary instant feedback on programming exercises can reduce student engagement and promote cheating. In *Proceedings of the 15th Koli Calling Conference on Computing Education Research* (pp. 122-126). https://doi.org/10.1145/2828959.2828968

Lee, H., 2023. The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*. https://doi.org/10.1002/ase.2270

Liu, M., 2018. The application and development research of artificial intelligence education in wisdom education era. In *2nd International Conference on Social Sciences, Arts and Humanities (SSAH 2018)* (pp. 95-100). https://doi.org/10.25236/ssah.2018.021

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B., 2016. *Intelligence unleashed: An argument for AI in education*. Pearson Education. Available at: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed-v15-Web.pdf (Accessed 10 May 2023).

Lund, B. D., & Wang, T., 2023. Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library Hi Tech News 40*(3), 26-29. https://doi.org/10.1108/LHTN-01-2023-0009

Mengi, G., 2023. Accessible and Individualized Learning: MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) Cambridge, MA. *XRDS: Crossroads, The ACM Magazine for Students, 29*(3), 58-59. http://dx.doi.org/10.1145/3589653

McCarthy, J., 2007. What is artificial intelligence. Available at: http://www-formal.stanford.edu/jmc/whatisai.pdf (Accessed 27 April 2023).

Pinkwart, N., 2016. Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *International journal of artificial intelligence in education, 26*, 771-783. https://doi.org/10.1007/s40593-016-0099-7

Polley, M. J., Seers, H. E., Cooke, H. J., Hoffman, C., & Paterson, C., 2007. How to summarise and report written qualitative data from patients: a method for use in cancer support care. *Supportive care in cancer, 15*, 963-971. https://doi.org/10.1007/s00520-007-0283-2

Punch, S., 2012. Hidden struggles of fieldwork: Exploring the role and use of field diaries. *Emotion, space and society, 5*(2), 86-93. https://doi.org/10.1016/j.emospa.2010.09.005

Qadir, J., 2023, May. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE. https://doi.org/10.1109/EDUCON54358.2023.10125121

Rajaraman, V., 2014. JohnMcCarthy—Father of artificial intelligence. *Resonance, 19*, 198-207. https://doi.org/10.1007/s12045-014-0027-9

Roberts, C., Montgomery, E., Richens, Y., & Silverio, S. A., 2021. (Re) activation of survival strategies during pregnancy and childbirth following experiences of childhood sexual abuse. *Journal of Reproductive and Infant Psychology 41*(2), 152-164. https://doi.org/10.1080/02646838.2021.1976401

Roll, I., & Wylie, R., 2016. Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*, 582-599. https://doi.org/10.1007/s40593-016-0110-3

Russell, S., & Norvig, P., 2016. *Artificial Intelligence: A Modern Approach*. Pearson Education.

Schiff, D., 2022. Education for AI, not AI for Education: the role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education, 32*(3), 527-563. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., 2021. Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI & society, 36*(1), 331-348. https://doi.org/10.1007/s00146-020-01033-8

Sindre, G., & Haugset, B., 2022, October. Techniques for detecting and deterring cheating in home exams in programming. In *2022 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE. https://doi.org/10.1109/FIE56618.2022.9962547

Skjuve, M., Følstad, A., & Brandtzaeg, P. B., 2023, July. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (pp. 1-10). https://doi.org/10.1145/3571884.3597144

Tapalova, O., & Zhiyenbayeva, N., 2022. Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. *Electronic Journal of e-Learning, 20*(5), 639-653. https://doi.org/10.34190/ejel.20.5.2597

Tegmark, M., 2018. *Life 3.0: Being human in the age of artificial intelligence*. Penguin Books.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B., 2023. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Vryan, K. D., 2006. Expanding analytic autoethnography and enhancing its potential. *Journal of Contemporary Ethnography, 35*(4), 405-409. https://doi.org/10.1177/0891241606286977

Wang, Y., 2021. When artificial intelligence meets educational leaders' data-informed decision-making: A cautionary tale. *Studies in Educational Evaluation, 69*, 100872. https://doi.org/10.1016/j.stueduc.2020.100872

Wang, Y., Li, H., Feng, Y., Jiang, Y., & Liu, Y., 2012. Assessment of programming language learning based on peer code review model: Implementation and experience report. *Computers & Education, 59*(2), 412-422. https://doi.org/10.1016/j.compedu.2012.01.007

Welsh, M., 2022. The end of programming. *Communications of the ACM, 66*(1), 34-35. https://doi.org/10.1145/3570220

Yampolskiy, R., & Fox, J., 2013. Safety engineering for artificial general intelligence. *Topoi, 32*, 217-226. https://doi.org/10.1007/s11245-012-9128-9

Yilmaz, R., & Yilmaz, F. G. K., 2023. Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans, 1*(2), 100005. https://doi.org/10.1016/j.chbah.2023.100005

Appendix 1: Field diary

Name:

Date	Question	ChatGPT's answer	Reflections	Screen-prints

Date	Question	ChatGPT's answer	Reflections	Screen-prints