Leveraging ChatGPT to Enhance Computational Thinking Learning Experiences

Abdessalam Ouaazki
Information Management Institute
University of Neuchâtel
Neuchâtel, Switzerland
abdessalam.ouaazki@unine.ch

Kristoffer Bergram
Information Management Institute
University of Neuchâtel
Neuchâtel, Switzerland
kristoffer.bergram@unine.ch

Adrian Holzer
Information Management Institute
University of Neuchâtel
Neuchâtel, Switzerland
adrian.holzer@unine.ch

***Abstract*—Given the pervasive reliance on technology in modern society, teaching Computational Thinking (CT) abilities is becoming increasingly relevant. These abilities, such as modeling and coding, have become crucial for a larger audience of students, not only those who wish to become software engineers or computer scientists. Recent advances in Large Language Models (LLMs), such as ChatGPT, provide powerful assistance to complete computational tasks, by simplifying code generation and debugging, and potentially enhancing interactive learning. However, it is not clear if these advances make CT tasks more accessible and inclusive for all students, or if they further contribute to a digital skills divide, favoring the top students. To address this gap, we have created and evaluated a novel learning scenario for transversal CT skills that leveraged LLMs as assistants. We conducted an exploratory field study during the spring semester of 2022, to assess the effectiveness and user experience of LLM-augmented learning. Our results indicate that the usage of ChatGPT as a learning assistant improves learning outcomes. Furthermore, contrary to our predictions, the usage of ChatGPT by students does not depend on prior CT capabilities and as such does not seem to exacerbate prior inequalities.**

***Index Terms*—Large Language Models, Interactive Learning Environments, Collaborative Learning, Human-Computer Interaction**

I. Introduction

Computational Thinking (CT) skills refer to the capability to model problems and their respective solutions in a manner that allows them to be automated [1] CT skills have been promoted since the 1980s [2], and are now considered an important skill of the 21st century [3]. Importantly, these skills should not be confined to computer scientists, but should be taught as a fundamental competency for everyone [4].

The capacity to model a task so that it can be automated and carried out by a computer could have seemed abstract for the average student a year ago, but this has completely changed with the public release of ChatGPT (https://chat.openai.com) on November 30th, 2022. ChatGPT is an instance of Large Language Models (LLMs). An LLM is described as "a machine-learning system that autonomously learns from data and can produce sophisticated and seemingly intelligent writing after training on a massive data set of text" [5, p.224].

Even though such artificial conversational interactions have been around for a long time, ChatGPT has disrupted the landscape of AI-based conversational models because it is one of the first models that enable one to converse *convincingly* on many topics [5].

Indeed, within a few months if not weeks after its release, it became clear that delegating tasks to computers was going to affect an increasing amount of people. Some jobs that were once believed to be relatively immune from automation have risen to the top of the occupations that are affected by conversational agents [6], [7]. Typically, LLMs can be directly used to automate tasks such as writing emails, generating ideas for marketing campaigns, drafting exam questions, or summarizing research findings [8]. Furthermore, they can be used as powerful tools to generate executable code in order to automate more complex tasks [9].

Nevertheless, these tools are not without limitations, as they can occasionally produce flawed code or misinterpret intended tasks [5], [9]. Furthermore, some universities perceive them as threats to scientific integrity and believe they should be prohibited, as exemplified by the case of Science Po in Paris which reportedly banned their use [10]. Some voices also question whether LLMs will be impediments to learning as reported by the New York Times: "Why learn when a bot does it better?" [11]This is a relevant question from an educational perspective. Questions are also arising about whether LLMs can be used to improve the level of all students, or whether they could exacerbate the gap between students with more or less initial CT abilities, as illustrated by a prominent AI researcher: "good programmers have just become at least twice as productive as they were before November 2022. But to leverage this productivity gain, they already had to be good programmers to begin with [12]. "

As such, it is not yet clear if LLMs are assets or impediments when students are to acquire new skills. This exploratory paper aims to investigate this issue in the context of CT learning, by addressing the following research question:

RQ: Can LLMs improve learning experiences for acquiring computational thinking abilities?

II. Related work

At the time of writing, it has been less than a year since the first mainstream LLM was released, i.e., ChatGPT on the 30th of November 2022. Although research on this class of AI

tools is still nascent, the educational field has been increasingly interested in the investigation of conversational agents. Indeed, a 2021 review of the literature shows an exponentially growing interest since 2016 [13]. These chatbots can assist with content integration, providing quick access to educational information, increasing motivation and engagement, allowing multiple users to access the system at the same time, and providing immediate assistance [13]. Chatbots have been successfully implemented in various educational contexts, aiding in saving time [14], maximizing student learning abilities, and increasing student engagement [15]. Chatbot interactions have also been designed and deployed to function as virtual assistants helping with administrative and academic tasks [16] or answering frequently asked questions (FAQs) to personalize and enrich the user experience [14], [17], [18].

The launch of ChatGPT has resolved several limitations inherent in previous chatbots, such as recalling previous statements in a conversation, understanding corrections made by the user, and rejecting inappropriate requests [19]. As such, it allowed to cater to a wider range of learning scenarios than previous chatbots [20]. Recent research has started investigating how LLM-based chatbots, such as ChatGPT, can be used to help programmers solve some tasks. For instance, [21] created an LLM assistant to help programmers write and debug their code, through the use of conversational interaction. They found that participants interacted with the assistant in two main patterns. Some of them relied on the LLM to solve the *entire* programming challenge. Others used it to *break down* the challenge into smaller tasks, then invoked the LLM to solve each of these. Findings from that study also indicate that the use of the LLM changed the nature of programmers' work process. It allowed many of them to focus on high-level aspects of the work, while it took care of lower-level details. This led to speeding up their work and allowed them to be more focused.

However, literature on chatbot interaction itself is still emerging and several important challenges remain. For instance, it is important to measure and assess chatbots more holistically to determine their actual benefits to users [22]. This evaluation should take into consideration aspects such as usefulness, efficiency, and process support [22], [23], as well as psychological outcomes such as performance, and perception of using chatbot systems [13]. Recent previous research has hinted at the positive feedback provided by students when exposed to a chatbot interaction to learn coding concepts (code style) [24].

In this article, we aim to address some of these open research avenues by designing a novel LLM-augmented learning experience implemented and evaluated in a real-life classroom. In particular, our objective is to contribute to the literature by empirically testing the path model illustrated in Figure 1, in addition to unveiling insights about students' user experiences with ChatGPT, thereby understanding their perceptions of its usage and how they interacted with it as an assistant to achieve their desired tasks. Regarding the proposed path model, it entails three main variables: (1) initial computational thinking capabilities at the beginning of a course, (2) the learning outcomes at the end of the course, i.e., acquired computational thinking capabilities, and (3) the usage of an AI assistant during self-directed learning activities throughout the course. We make three conjectures about the relationships between these variables. First, based on the positive findings of the previous research discussed above, we expect that the use of an AI assistant during self-directed learning activities will help students learn and thus positively influence their learning outcomes [14], [15]. Second, based on previous research findings, we expect learning outcomes to also be positively influenced by initial CT capabilities [25]. Third, based on previous literature that highlights concerns related to increasing the digital divide [5], where more tech-savvy users will better be able to take advantage of AI-powered tools, we expect that initial capabilities will positively influence the use of self-directed learning activities.

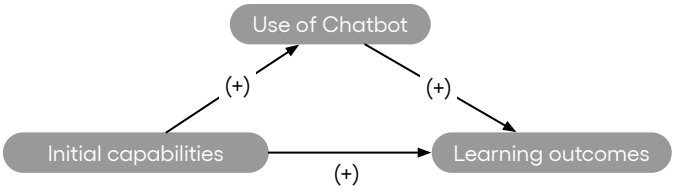

Fig. 1. PLS model. The (+) sign indicates expected positive path coefficients between the constructs.

III. Methodology

We designed a novel LLM-augmented learning experience that enables students to understand how LLMs can help them become more proficient in their CT capabilities. The context for this study was a 3 ECTS Computational Thinking course for master students in business-related masters programs (e.g., Finance, Applied Economics, International Management) at our university. The course took place one week before the spring semester of 2022, but students then had five weeks to work on their assignments. The learning objectives of the course were as follows: (1) Being able to formulate a problem in a computational way, (2) Being able to evaluate an algorithmic solution, and (3) Producing a solution in an algorithmic form. During the course, students were presented with theoretical foundations related to CT concepts as well as with the basics of the Python programming language (conditions and loops, lists, functions, dictionaries) as well as introductions to handling data (graphs, pandas, regex). The format of the course combined theory delivered through semi-active lectures using Jupyter Notebooks, where students were able to modify and execute code during class, with active self-directed learning scenarios for which students were asked to work alone, but were encouraged to contact the teaching staff and use ChatGPT as a learning assistant. The active self-directed learning activities consisted of hands-on lab sessions and an individual assignment. The evaluation of the course was done through three components[1]: self-directed lab sessions

[1]The brackets (...) highlight the proportion of the final course grade

(10%), a self-directed assignment (50%) and an individual pen and paper test in class (40%). The test took place roughly five weeks after the end of the course and served as the deadline for the other two deliverables.

A. Self-directed lab sessions

For Python basics (conditions and loops, lists, functions, dictionaries), we provided students with a set of 15-25 self-directed exercises that they could solve on their own on the Graasp learning platform [26]. Each exercise required students to write simple programs to practice the topic covered in the session. The setting of the exercises allowed students to get feedback if their exercises were correctly solved. An example of an exercise on loops and conditions read as follows:

> A random variable called zebra has been generated in the code block below. Increment the value of zebra by 2, as long as it is strictly less than 100. Finally, store the value of zebra in a variable called result. Run your code, if your result variable is correctly defined you will get a magic code to validate the exercise.

For this type of questions, ChatGPT can easily provide an answer. As such it could be very tempting to quickly solve most exercises without even reading them. To encourage students to actually learn from these exercises, the instructors reduced the pressure for performance and indicated that students would receive the full score for completing 30% or more of the exercises. Accordingly, the instructors informed students that it was more important to use the lab exercises to learn how to solve these problems as practice for the final test, rather than completing all of them correctly by outsourcing the work to ChatGPT. Nevertheless, students were encouraged to use ChatGPT when they were stuck. It should be mentioned that all but two students reached the success threshold and thus had the maximum grade for the lab sessions; one of these two students did not complete any exercise.

B. Self-directed assignment

To go beyond basic coding exercises, we designed an active self-directed assignment for which ChatGPT could be a powerful help, yet could not solve the problem by simply using the assignment instruction as a prompt. The idea of the assignment was to create a histogram that showed the number of news articles per week that were published on the *@celebjets* topic in 2022. The research team compiled a file that contained all news articles scraped from a search on Google News (see Figure 2). Then they provided students with an HTML file (33'061 words) with the data from 274 news articles. Students were asked to write a Python program that parses this file and creates a histogram showing the number of articles published for each week of the year.

The relevant datetime data for each article was embedded within a substantial amount of HTML code. The snippet below demonstrates the structure of the markup:

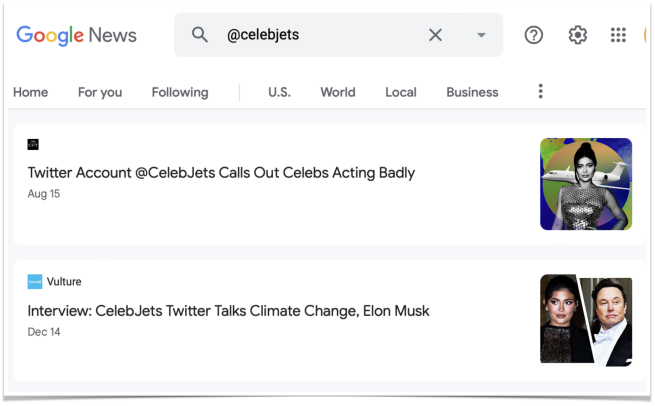

Fig. 2. Screenshot of the results for articles containing the word "@celebjets" on Google News.

```
...
<a href="...">
    Interview: CelebJets Twitter Talks
    Climate Change, Elon Musk
</a>
...
<time datetime="2022-12-14T08:00:00Z">
    Dec 14
</time>
...
```

The assignment required students to go through the following steps: (a) import the text file, (b) extract the dates using regEx, (c) format the dates correctly, (e) store the dates in a list, and (f) create a plot based on the list. The following code illustrates a working solution to solve the assignment:

```
import re
from datetime import datetime
import matplotlib.pyplot as plt
import matplotlib.dates as mdates

#Import the text file
with open('news-celebjets.txt', 'r') as file:
   f = file.read()

#Extract the date strings using regEx
s_dates = re.findall(r'datetime="(.{10})T', f)

#Convert string dates to datetime type
dates=[]
for s_date in s_dates:
  date = datetime.fromisoformat(s_date)
  #Put the dates in a list
  dates.append(date)

#Create a plot based on the list
plt.hist(dates, bins=52)
plt.show()
```

For the deliverable, students were asked to hand in their code and write a two-page report explaining their process, and linking their work to CT concepts and ChatGPT usage.

C. Metrics

We used three main data sources for this study: pre- and post-survey answers provided by students, interaction log data from ChatGPT, and course evaluation results.

1) Pre- and post-surveys: At the beginning of the course, students were asked to fill in a pre-survey, where they could give their consent to participate in the research and answer questions related to the study. The pre-survey consisted of a set of eight programming questions that students were asked to answer without the use of any technology. This test was designed to measure the initial programming capabilities of students. These questions were designed specifically for the course. The post-survey contained questions about the students' use of ChatGPT as an assistant. Questions included behavioral information such as the number of labs for which they used ChatGPT, and their attitudes towards ChatGPT usage. The attitudes were measured using the Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) [27], as well as three custom Likert scale questions about the usefulness of ChatGPT for learning. We asked students to evaluate the following statements using a five-point Likert scale (where 1 = strongly disagree and 5 = strongly agree): (1) Overall, using ChatGPT has improved my ability to solve problems, (2) ChatGPT has helped me reach my goals for this course, (3) ChatGPT has helped me learn new skills for this course. Finally, students were asked to discuss their experience with ChatGPT in an open question.

2) ChatGPT interaction data: We used log data from the student conversation with ChatGPT during the take-home assignment to investigate interaction patterns. We asked students to provide a log of their interactions with ChatGPT while they were solving the assignment.

3) Course evaluation results: Course evaluation results were used as a measure of learning outcomes. In particular, we used the final grade as a metric, which combines the score of the lab sessions (10%), the score of the self-paced assignment (50%), and the score of the written test (40%).

IV. Results

This section presents the results of our analysis. Overall, 25 students (15 Male, 10 Female), aged 22-34 (M=24), gave their consent for the study and answered the pre- and post-surveys.

A. Student User Experience

Table I presents the results of the SUPR-Q scores, including usability, trust, appearance, loyalty, and the overall total score. The table also shows the results of single sample t-tests that were performed on the different scores compared to the reference means extracted from the original paper presenting the SUPR-Q scale [27].

The results show that the total score (M=4.03) and the usability dimension (M=4.22) are generally good and are comparable with the reference websites (M=3.93 and M=4.06). Furthermore, appearance (M=4.30) and loyalty (M=4.35) are significantly better for ChatGPT than reference websites (M=3.88 and M=3.91). Within the items in these constructs, students have expressed their willingness to encourage others to use ChatGPT. However, students rated ChatGPT significantly worse on the trust dimension (M=3.24) than reference websites (M=3.80).

TABLE I
Single-Sample t-test Results for SUPR-Q sub-scales

SUPR-Q score	sample mean (std)	*reference mean	t-statistic	p-value
Usability	4.22 (0.84)	4.06	0.949	.352
Trust	3.24 (0.71)	3.80	-3.950	.001
Appearance	4.30 (0.74)	3.88	2.853	.009
Loyalty	4.35 (0.81)	3.91	2.705	.012
Total score	4.03 (0.55)	3.93	0.879	.388

*Based on the following SUPR-Q benchmarks [27, p. 83].

Table II shows the results of student responses to the questions about ChatGPT's usefulness. Furthermore, it shows the results of single-sample t-tests comparing the means to the neutral response (3.0). All tests are significant indicating that students found ChatGPT helpful in enhancing their problem-solving skills. Furthermore, they found that it helped them reach their goals for the course and learn new skills.

TABLE II
Single-Sample T-Tests Results for Learning Aspects

Aspect	sample mean (std)	reference mean	t-stat	p-value
Problem Solving	4.32 (0.69)	3	9.560	.000
Reaching Goals	4.12 (0.72)	3	7.716	.000
Learning New Skills	4.2 (0.87)	3	6.928	.000

The responses to the open user experience questions shed some further light on the quantitative results above. One common theme that emerged from the comments is the perceived usefulness of ChatGPT in helping students get started with their tasks. For instance, a student mentioned: "ChatGPT was very nice to use and very easy to understand. It helped me in the most difficult point of this class by giving me quick and smart direction to accomplish the task". Another stated: "Even if sometimes I was given wrong answers for asked questions it would still provide me a general idea from where to start." In terms of learning, a student noted: "It's beneficial for beginners if you have a computational concept in mind and you want to express it in Python codes or if you don't understand the functionality of code and you need an explanation."

Another theme was the potential risks associated with delegating too much to ChatGPT. For instance, a student mentioned: "I tried to do all exercises by myself first and used ChatGPT in exceptional cases when I couldn't solve the problem for a long time. It's more interesting for me to solve problems without any external assistance". Another stated: "For this course, since I wanted to learn Python basics and be able to rely on myself I didn't use it for the TPs. That is why I needed ChatGPT for only 20% of my assignment. Still, it was really helpful and having the opportunity to discover this tool is for me a big asset." Another student highlighted a potential risk associated with learning: "It can be damaging if people only copy the task description as prompt and use the

code without understanding the functionality." Finally, another student commented: "I think ChatGPT is good. But sometimes, he can give some wrong answers."

B. Learning outcomes

To validate the path model, we conducted a partial least squares (PLS) analysis technique with SmartPLS. PLS is a variance-based structural equation modeling (SEM) analysis technique increasingly used in Information Systems to better understand phenomena outcomes [28]. In our model, we measured our three variables with the following metrics: (1) initial capabilities were measured by the score on the pre-survey Python test, (2) learning outcomes were measured by the score of the final grade, and (3) the use of ChatGPT was measured by the number of self-directed activities in which ChatGPT was used. We ran the PLS analysis by bootstrapping the existing data with 1000 re-samples. Figure 3 shows the outcome of the analysis. More specifically, the results show that the initial capabilities do indeed influence the learning outcomes positively. Moreover, the use of ChatGPT positively influences learning outcomes. Finally, no statistically significant relationship was found between the initial capabilities and the use of ChatGPT.

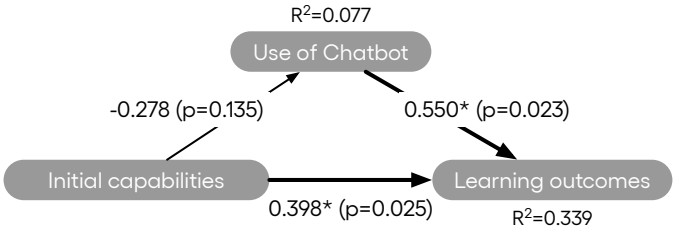

Fig. 3. PLS results

C. LLM Conversations

We conducted an exploratory content analysis [29], [30] of the interactions with ChatGPT during the self-directed assignment, in order to understand how students interacted with the LLM. Four students did not use ChatGPT for their assignment and thus were excluded from this analysis. 21 students made 520 prompts to ChatGPT in total, with an average of 24.76 prompts per student ($SD = 24.57, min = 3, max = 85$). We coded these prompts using an inductive deductive approach [31]. First, we used deductive codes such as instructions and debugging, and inductive codes were added after inspecting the data (context, explanation request, other). Results of this coding are shown in Figure 4. Additionally, the figure shows the proportion of students who used each category of prompts at least once.

Over half of the prompts fell into the *instruction* category, with every student utilizing this type of prompt. This suggests that students primarily interacted with ChatGPT by issuing directives to guide the LLM toward solving their assignment. The *debugging* prompts, where students identified errors or bugs in ChatGPT's output and asked for corrections, accounted

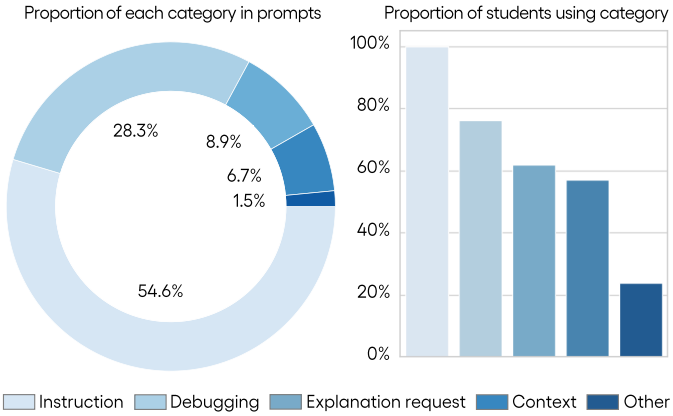

Fig. 4. Proportion of prompt categories, and proportion of students who used a given category at least once.

for about 30% of prompts. Approximately three-quarters of the students provided at least one *debugging* prompt during their interaction with the model.

The *context* prompts, where students provide the LLM with the context of their request, and *explanation request* prompts, where students ask for an explanation, made up smaller proportions of the prompts, with about 8.8% and 6.7% respectively. Yet, more than half of the students engaged with each of these categories, suggesting that these aspects, although less frequent, were important in the conversations. Lastly, the *other* category, encompassing prompts that did not fit into the prior categories represented around 1% of prompts and included casual greetings or irrelevant information (see Table III for prompt examples).

V. Discussion and Conclusion

In this paper, we investigated how an LLM can be leveraged to improve CT learning experiences. To that end, we designed and evaluated self-directed course activities in the context of a real 3 ECTS master-level computational thinking course in which students were encouraged to use ChatGPT as a learning assistant. We conducted both quantitative and qualitative analyses of the learning context. A first important finding of our study showed that the use of ChatGPT was positively associated with learning outcomes, conveying the fact that the more students used ChatGPT for their self-directed learning activities the better they performed in the final exam. This novel finding supports the idea that AI tools such as ChatGPT can successfully be integrated into learning experiences for computational thinking to improve learning rather than diminishing it. This expands upon previous research that showed positive results when using conversational agents in education [13]–[15]. Our finding seems to indicate that students used ChatGPT to better understand the topic. This is reflected through the self-reported usage of ChatGPT where students overwhelmingly agreed that ChatGPT helped them learn new skills. The investigation into the actual prompts used by students during the individual assignment showed that almost two-thirds of the students (61.90%) explicitly

asked ChatGPT for an explanation of the code. The qualitative analysis of student comments on ChatGPT seems to confirm a certain level of awareness from students when it comes to the limits and dangers of misusing LLMs for educational purposes. This finding offers potential reassuring perspectives on how students make use of LLMs to improve their learning rather than just using them as tools that perform their tasks on their behalf. Future research could further investigate under which conditions students learn best using LLMs, and how these models can be conditioned through prompt engineering to provide more useful answers in that context.

A second important finding is the fact that, although initial CT capabilities positively influenced learning outcomes (in line with previous studies [25]), they did not appear to have an influence on ChatGPT usage. This means that ChatGPT was not mainly used by good programmers to become even better programmers. This result could indicate that, in certain conditions, tools such as ChatGPT do not exacerbate pre-existing gaps when it comes to acquiring CT capabilities. Since being a source of additional inequalities is seen as a major potential issue linked with LLMs [5], this finding should be further investigated in follow-up studies.

Specifically, more high-powered and extensive studies could aim at highlighting the pre-conditions that could lead some groups of students to interact more or less with LLMs.

TABLE III
EXAMPLES OF STUDENT PROMPTS FOR EACH PROMPT CATEGORY

Category	**Examples**
Instruction	"write python code to scrape articles from google news with the tag 'celebjet'". "to this code, add the number of the weeks that have the most number of articles, displayed like you did for the mean" "create an array of bin edges from 0 to 54" "write python code to locally save the json files from the code above"
Debugging	"start_date = datetime.strptime(dates[0], '%Y-%m-%d') there is an error here" "I got an error output: Traceback (most recent call last)..." "now I get the following error: ValueError: Invalid isoformat string: '2022-12-14T08:00:00Z'"
Explanation request	"Explain how plt.hist works" "what does it mean Parse the HTML content using Beautiful Soup" "what is the difference between object and string in python"
Context	"Context: We would like to know when news articles on the @celebjets twitter account were written in 2022." "I will give you a sample of the data so you can know in which format it is"
Other	"hi chat !" "Ok, thank you. I want something else. I will look on other websites. Thank you."

REFERENCES

[1] J. Wing, "Research notebook: Computational thinking—what and why," *The link magazine*, vol. 6, 2011.

[2] S. Papert, *Children, computers, and powerful ideas*. Harvester, 1980.

[3] D. M. Mohaghegh and M. McCauley, "Computational thinking: The skill set of the 21st century," 2016.

[4] J. M. Wing, "Computational thinking," *Communications of the ACM*, vol. 49, no. 3, pp. 33–35, 2006.

[5] E. A. Van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, "Chatgpt: five priorities for research," *Nature*, vol. 614, no. 7947, pp. 224–226, 2023.

[6] E. Felten, M. Raj, and R. Seamans, "How will language modelers like chatgpt affect occupations and industries?" *arXiv preprint arXiv:2303.01157*, 2023.

[7] ——, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, vol. 42, no. 12, pp. 2195–2217, 2021.

[8] P. P. Ray, "Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, 2023.

[9] H. Tian, W. Lu, T. O. Li, X. Tang, S.-C. Cheung, J. Klein, and T. F. Bissyandé, "Is chatgpt the ultimate programming assistant–how far is it?" *arXiv preprint arXiv:2304.11938*, 2023.

[10] G. De Clercq, "Top french university bans use of chatgpt to prevent plagiarism," Reuters, 2023, retrieved June 26, 2023. [Online]. Available: https://www.reuters.com/technology/top-french-university-bans-use-chatgpt-prevent-plagiarism-2023-01-27/

[11] The Learning Network, "What students are saying about chatgpt," The New York Times, Feb 2023, retrieved June 20, 2023. [Online]. Available: https://www.nytimes.com/2023/02/02/learning/students-chatgpt.html

[12] M. Salathé, "Artificial intelligence: What to worry about," Substack, May 2023, retrieved June 22, 2023. [Online]. Available: https://engineeringprompts.substack.com/p/artificial-intelligence-what-to-worry

[13] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100033, 2021.

[14] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for university related faqs," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 1525–1530.

[15] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "Chatbot: An education support system for student," in *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29–31, 2018, Proceedings 10*. Springer, 2018, pp. 291–302.

[16] S. Sinha, S. Basak, Y. Dey, and A. Mondal, "An educational chatbot for answering queries," in *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*. Springer, 2020, pp. 55–60.

[17] K. Peyton and S. Unnikrishnan, "A comparison of chatbot platforms with the state-of-the-art sentence bert for answering online student faqs," *Results in Engineering*, vol. 17, p. 100856, 2023.

[18] F. Sethi, "Faq (frequently asked questions) chatbot for conversation," *Authorea Prepr*, vol. 8, 2020.

[19] J. Jeon and S. Lee, "Large language models in education: A focus on the complementary relationship between human teachers and chatgpt," *Education and Information Technologies*, pp. 1–20, 2023.

[20] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[21] S. I. Ross, F. Martinez, S. Houde, M. Muller, and J. D. Weisz, "The programmer's assistant: Conversational interaction with a large language model for software development," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 2023, pp. 491–514.

[22] A. Følstad, T. Araujo, E. L.-C. Law, P. B. Brandtzaeg, S. Papadopoulos, L. Reis, M. Baez, G. Laban, P. McAllister, C. Ischen *et al.*, "Future directions for chatbot research: an interdisciplinary research agenda," *Computing*, vol. 103, no. 12, pp. 2915–2942, 2021.

[23] A. Rapp, L. Curti, and A. Boldi, "The human side of human-chatbot interaction: A systematic literature review of ten years of research on

text-based chatbots," *International Journal of Human-Computer Studies*, vol. 151, p. 102630, 2021.

[24] J. C. Farah, B. Spaenlehauer, K. Bergram, A. Holzer, and D. Gillet, "Challenges and opportunities in integrating interactive chatbots into code review exercises: A pilot case study," in *EDULEARN22 Proceedings*. IATED, 2022, pp. 3816–3825.

[25] A. De Santo, J. C. Farah, M. L. Martínez, A. Moro, K. Bergram, A. K. Purohit, P. Felber, D. Gillet, and A. Holzer, "Promoting computational thinking skills in non-computer science students gamifying computational notebooks to increase student engagement," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 394–405, 2022.

[26] D. Gillet, A. Vozniuk, M. J. Rodríguez-Triana, and A. Holzer, "Agile, Versatile, and Comprehensive Social Media Platform for Creating, Sharing, Exploiting, and Archiving Personal Learning Spaces, Artifacts, and Traces," in *The World Engineering Education Forum*, 2016.

[27] J. Sauro, "Supr-q: A comprehensive measure of the quality of the website user experience." *Journal of usability studies*, vol. 10, no. 2, 2015.

[28] J. Hair, C. L. Hollingsworth, A. B. Randolph, and A. Y. L. Chong, "An updated and expanded assessment of PLS-SEM in information systems research," *Industrial Management and Data Systems*, pp. 442–458, 2017.

[29] J. Saldaña, *The Coding Manual for Qualitative Researchers*, ser. Core textbook. London: SAGE, 2021. [Online]. Available: https://books.google.nl/books?id=iQK1zQEACAAJ

[30] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[31] J. Fereday and E. Muir-Cochrane, "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development," *International journal of qualitative methods*, vol. 5, no. 1, pp. 80–92, 2006.