



# “Call me Kiran” – ChatGPT as a Tutoring Chatbot in a Computer Science Course

Jaakko Rajala  
Tampere University  
Tampere, Finland  
jaakko.rajala@tuni.fi

Jenni Hukkanen  
Tampere University  
Tampere, Finland  
jenni.j.hukkanen@tuni.fi

Maria Hartikainen  
Tampere University  
Tampere, Finland  
maria.hartikainen@tuni.fi

Pia Niemelä  
Tampere University  
Tampere, Finland  
pia.niemela@tuni.fi

## ABSTRACT

Natural language processing has taken enormous steps during the last few years. The development of large language models and generative AI has elevated natural language processing to the level that it can output coherent and contextually relevant text for a given natural language prompt. ChatGPT is one incarnation of these steps, and its use in education is a rather new phenomenon. In this paper, we study students’ perception on ChatGPT during a computer science course. On the course, we integrated ChatGPT into Teams private discussion groups. In addition, all the students had freedom to employ ChatGPT and related technologies to help them in their coursework. The results show that the majority of students had at least tested AI-powered chatbots, and that students are using AI-powered chatbots for multiple tasks, e.g., debugging code, tutoring, and enhancing comprehension. The amount of positive implications of using ChatGPT takes over the negative implications, when the implications were considered from an understanding, learning and creativity perspective. Relatively many students reported reliability issues with the outputs and that the iterations with prompts might be necessary for satisfactory outputs. It is important to try to steer the usage of ChatGPT so that it complements students’ learning processes, but does not replace it.

## CCS CONCEPTS

• **Applied computing** → **Computer-assisted instruction**; • **Human-centered computing** → *Natural language interfaces*.

## KEYWORDS

artificial intelligence, generative AI, ChatGPT, student perceptions, chatbots, education, tutoring, discussion forum

### ACM Reference Format:

Jaakko Rajala, Jenni Hukkanen, Maria Hartikainen, and Pia Niemelä. 2023. “Call me Kiran” – ChatGPT as a Tutoring Chatbot in a Computer Science Course. In *26th International Academic Mindtrek Conference (Mindtrek ’23)*,



This work is licensed under a Creative Commons Attribution International 4.0 License.

*Mindtrek ’23, October 03–06, 2023, Tampere, Finland*  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0874-9/23/10.  
<https://doi.org/10.1145/3616961.3616974>

October 03–06, 2023, Tampere, Finland. ACM, New York, NY, USA, 12 pages.  
<https://doi.org/10.1145/3616961.3616974>

## 1 INTRODUCTION

"Artificial Intelligence: The New Frontier!", "AI Breakthrough: Machine Learning Defeats Chess Grandmaster!", "AI: Will Robots Steal Your Job?". Just like "The Boy Who Cried Wolf," news outlets have repeatedly heralded the imminent AI revolution over the years. While the AI "wolf" seems more present than ever, many seasoned engineers have grown skeptical due to past exaggerated claims. While it’s debatable whether we are truly on the cusp of an AI revolution, there’s no denying that the advancements in recent years are significant.

Periods of hype have been interspersed with periods of disappointment. One notable period of disappointment occurred in 1973 when James Lighthill’s report provided a critical assessment of the progress made in AI research. This evaluation had a significant impact on the field, resulting in a decrease in funding and a deceleration of advancements. Another significant dry season, commonly referred to as AI Winter, occurred in the late 1990s to early 2000s. During this period, the field experienced reduced interest and funding as many AI projects failed to live up to the high expectations. Enthusiasm waned, and there was a general sense of disillusionment with AI.

Despite these challenges, AI has managed to continue its growth trajectory, driven by advancements in technology. The subsequent decade, the 2010s, witnessed an acceleration of progress. In 2014, Variational Autoencoders (VAEs) [13] were introduced, presenting a probabilistic approach to generative modelling which expanded the capabilities of AI in generating diverse and realistic outputs. Furthermore, in 2018, OpenAI released GPT (Generative Pre-trained Transformer), a groundbreaking large-scale language model capable of producing coherent and contextually relevant text.

Ever since the triumph of generative AI, progress has been breathtaking: ChatGPT 3.0 was released in 2020 as a powerful language model, showcasing remarkable natural language understanding and generation capabilities. It set a new standard for conversational AI models. Building upon the success of 3.0, OpenAI introduced ChatGPT 3.5 in November 2022. This version further refined the model’s performance, addressing some of its limitations and enhancing its abilities. After ChatGPT 3.5, OpenAI introduced ChatGPT 4.0 on March 14, 2023. This upgraded version maintained the path of

enhancement, delivering superior language comprehension and response generation abilities. Users were provided with progressively advanced and efficient language interaction encounters.

The rapid proliferation of ChatGPT in educational settings has been met with a mixture of astonishment and skepticism. Its capabilities, often likened to a blend of advanced technology and magic, have been highlighted in various studies [19]. Educators' responses to this transformative wave range from complete denial to embracing the change and new opportunities. The denial is due to the concerns of impact on students' learning outcomes [14] and creativity [20]. For instance, University of Hong Kong has banned the usage at least temporarily [5], Science Po requests transparent referencing of students' submissions [3], also the Italian data-protection authority expressed privacy concerns regarding a model developed by OpenAI, a US start-up supported by Microsoft. In response to these concerns, the regulator announced an immediate ban on OpenAI's activities in Italy and initiated an investigation into the matter [2]. In contrast, the majority of Scandinavian countries, as noted in a citation from [4], embrace the newfound possibilities offered by ChatGPT. Researchers such as Anders Isaksson from Chalmers University of Technology in Sweden perceive ChatGPT as a catalyst for acquiring innovative skills pertaining to AI prompts; Tampere University has a similar take. To explore the advantages and disadvantages of ChatGPT in higher education in our university, the following research questions were formulated for this study:

- (1) What are students' awareness levels regarding ChatGPT?
- (2) How often students are currently using ChatGPT and what are their main use-cases?
- (3) What are the anticipated effects of ChatGPT on students' learning and creativity?
- (4) How do students perceive the reliability and security of ChatGPT?

## 2 LITERATURE REVIEW

The AI "wolf" has strived to demonstrate its might in the realm of education by huffing and puffing before well-entrenched institutions, such as high schools and universities. Lately, there have been remarkable glimpses of achievement, including the integration of ChatGPT into intelligent tutoring systems and the rise of adaptive and intelligent massive open online courses (MOOC), which have a power to shake the traditional classroom settings [16]. These updated MOOCs employ technology-driven adaptive learning systems to customise the learning experience for each individual participant to align with their abilities, preferences, and progress, while also identifying students at risk of dropping out. Consequently, MOOCs have started to exhibit characteristics resembling those of adaptive learning platforms. These systems analyze and organize vast amounts of logged data to provide advantages for the learner. As a result, the content becomes tailored to the individual, enabling the system to adeptly adjust to the specific requirements of each learner.

Intelligent tutoring system (ITS) incorporates a domain-specific knowledge base that encompasses the subject matter being taught. The focus centers on providing personalized and adaptive instruction to individual learners. ITS employ AI algorithms to analyze

student performance, identify areas of weakness, and deliver targeted feedback and guidance. Illustrative examples of early pioneers in this field include the CMU LISP tutor (1985), which guided students step by step, diagnosed errors, offered corrective feedback, adaptive hints, and progressively advancing exercises [18], as well as AutoTutor (1998), which utilized natural language interaction to achieve similar goals [11].

The advancements that lead to a jump from ITS to the next category, Adaptive Assessment Systems, include improved data collection: AI-enhanced learning management systems (LMS) that gather more comprehensive data on student performance, such as real-time tracking of progress, learning patterns, and interaction data; enhanced natural language processing: AI models have improved in understanding and processing human language, enabling more effective interaction and personalised feedback in tutoring systems; and finally advances in machine learning: algorithms have become more sophisticated, allowing for better analysis of learner data and more accurate predictions of student needs and performance, and the improved assessment. Adaptive Assessment Systems offer a natural progression to incorporate evaluation and diagnostics into the learning process and use AI to dynamically tailor assessments to each learner's abilities and needs. These systems adjust the difficulty and content of questions based on the learner's performance. An illustrative example is Khan Academy, which adjusts difficulty and content based on learner performance, providing targeted feedback and practice exercises.

The shift to Adaptive Learning Platforms is driven by technological disruptions. Advancements in statistical models and psychometrics enhance Item Response Theory, improving learner ability estimation and item selection in adaptive assessments [10]. Algorithms now instantly identify knowledge gaps and create adaptive assessments. Moreover, AI's ability to interpret diverse data types, like text, images, audio, and video, enriches assessment experiences. Adaptive Learning Platforms may tailor learning experiences to individual learners. They leverage AI to provide customised content, resources, and learning paths based on learner needs and exemplified by Smart Sparrow and its capability to create interactive and adaptive learning experiences "education through exploration" interactive simulations, and branching pathways allowing learners to explore concepts at their own pace [1, 6].

Technological disruptions that pave way to learning analytics (LA) systems, include improvements in content recommendation systems: AI algorithms can analyse learner preferences, behaviour, and performance to recommend relevant learning materials and activities. Progressing towards LA was prompted also by a vast amount of data gathered on learner behaviour, progress, and engagement. LA systems enables educators and institutions to derive valuable insights from this data to inform instructional design, monitor learning outcomes, and improve overall learning experiences. LA employs AI and data analysis techniques to gain actionable insights from learner data. They focus on extracting meaningful patterns and trends to inform decision-making and improve educational outcomes. Technological disruptions in this category include Big Data and cloud computing that necessitate the availability of large-scale data storage and processing capabilities. LA in Moodle is not the most convincing example of LA, but the system is open-source and allows anyone to extend the system to get better

insights into learner behaviour and to enable data-driven instructional decision-making, such as VeLA and GISMO [8].

All of these categories utilise natural language processing (NLP) to capture the users' behaviour and knowledge level. The field of NLP has developed with enormous steps during the last years.

## 2.1 AI-based large language models for chatbots

NLP witnessed a significant advancement in 2017 with the introduction of transformers [21]. A transformer is a deep learning model, which process the input entirely all at once. Transformers allow increased levels of parallel operations, which have allowed larger training data sets and reduced the required training times. The increased training capabilities stimulated the development of pre-trained systems, such as GPTs (generative pre-trained transformers) [7, 17] and Bert (bidirectional encoder representations from transformers) [9].

ChatGPT's ability to process and produce natural language has reached the level which has astonished many and it has enormous consequences for multiple fields. ChatGPT can produce long texts that are grammatically correct [7]. It also process and analyse automatically the given input text for being able to answer coherently. In addition, ChatGPT understands the given instructions for the processing so that the instructions can be given as a form of natural language. Therefore, ChatGPT can be used for automatizing tasks, such as for generating texts and replies on chatbots and summarising long texts. ChatGPT can also be utilised for expanding the existing ideas and giving additional sparring for the ideas and thoughts. The issues with large language models include their non-public development and training, leading to vulnerabilities in terms of biases, concerning e.g. gender and age. These biases mainly stem from the inherent biases within the training datasets. In addition, their interpretability and explainability are poor, i.e., it is difficult to explain how the model ends up to a specific output, and what is the exact reasoning behind it.

## 2.2 Chatbots and ChatGPT in education

The advancement of large language models is significantly influencing the conventional methods of education, teaching, and assessment. This progress has introduced numerous valuable possibilities, as well as a few notable challenges. Next, we will shortly touch some main aspects regarding AI-powered chatbots and especially ChatGPT in education.

Kasneci et al., in their paper [12], provides a summary of the opportunities and challenges associated with the utilization of language models in education, examining these aspects from both teacher and student perspectives. ChatGPT opens doors in education for personalized and effective learning experiences that match each student's style and needs. ChatGPT and other large language models are developed using large text corpuses and they are able to produce grammatically and syntactically correct text. They are also able to make translations and to use different writing styles and tones. Hence, large language model based systems are especially suited for developing reading and writing skills [12]. Large language models' ability to find writing errors and propose corrections for the given text are valuable scaffolding aids. Similarly for computer science courses, large language models' ability to detect and

produce code of different coding languages are highly useful; and the models can be used for generating code and showing example code snippets, proposing commonly used tools and functions for each specific use-case, and finding bugs from the code.

Many of the challenges regarding ChatGPT for education are the same as are in general, such as bias, fairness and security issues and the lack of transparency. Kasneci et al. [12] mention copyright and plagiarisms issues that should be taken into account: The output of large language models might contain sentences or longer sets of text in a form that were used in training the model. They propose to interfere the issues by using open-source content, or asking the permissions for the training data sets, and importantly informing the users of the risks.

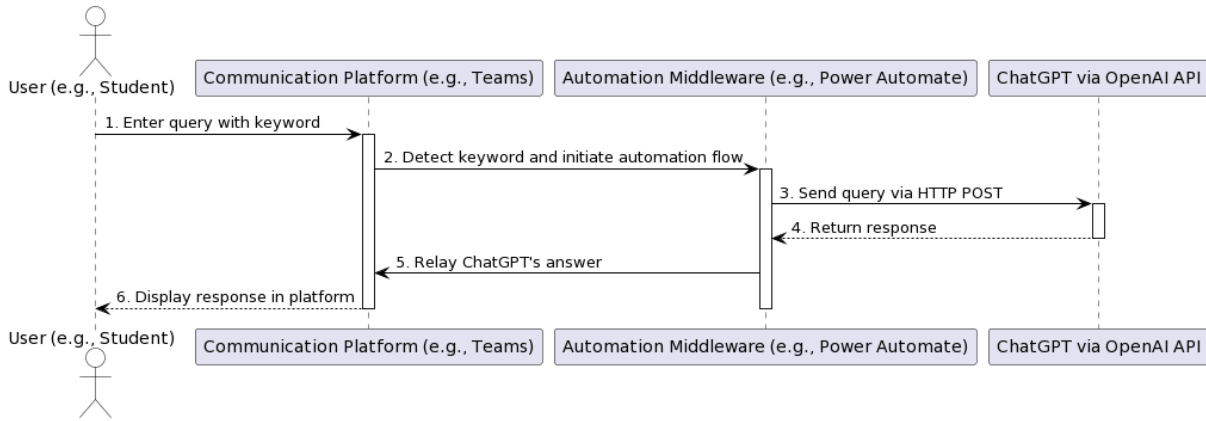
For teachers, new large language models creates new challenges in the form that how to assess students knowledge and skills reliably. Large language models have eased students access to produce concise explanations to difficult theoretical questions without deep understanding on the topic. Therefore, there are needs to discuss how to support deep understanding whilst not losing all the great opportunities large language models are offering.

## 3 RESEARCH CONTEXT

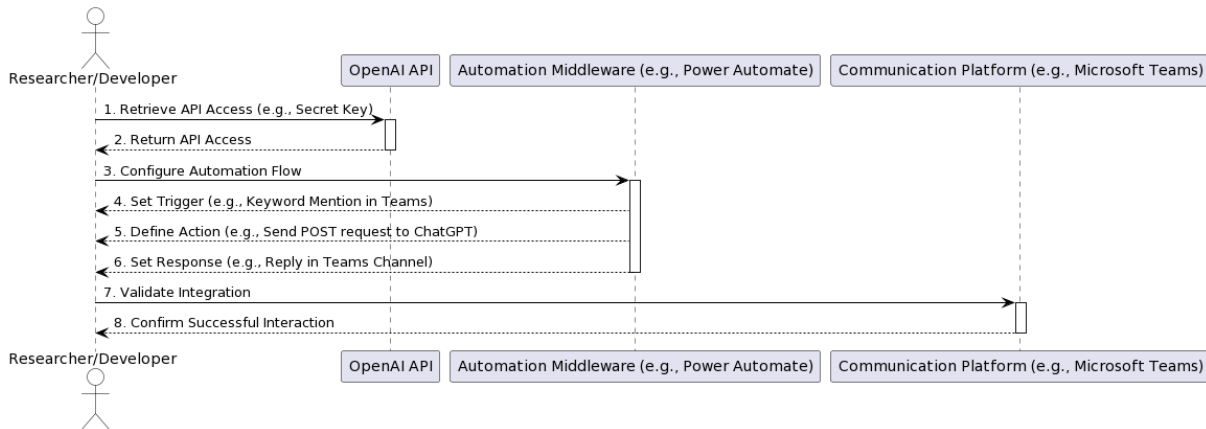
The research was performed during the basic Web Development course at Tampere University spring 2023. The studied course is targeted for the first- and second-year students, and the course focuses on frontend web technologies, such as HTML, CSS, and JavaScript. During the study period, 233 students enrolled in the Plussa LMS.

The course was conducted using a hybrid learning approach, wherein the initial portion of the lectures were delivered in a physical lecture hall and simultaneously recorded for later access; the latter part of the lectures was exclusively presented online. The exercises covered various question formats, such as multiple-choice questions and programming tasks. These tasks were submitted through the Plussa LMS, which automatically graded them. Notably, around one-third of the exercises were submitted by small groups composed of 3 to 7 individuals. Over time, more gamified exercises have been added to the course curriculum. Even the exam predominantly centered on gamified elements, allowing students the chance to progressively enhance their answers.

Sample solution sessions were hold after the deadline of the respective course module; their purpose was to deepen the learned topics by allowing students to describe the solutions to fellow students and by discussing the related topics further. The sessions were hold online, and there were four similar sessions each week to choose from, so that there were 20–40 students participating in each session. In order to stimulate the discussion and learning, the students were divided into small groups and all groups searched for answers to questions on topics related to the exercises. These small groups were the same which were also used in the group exercises, and in the optional course project. At the end of the Sample solution session, all the answers for the exercises and questions were discussed together.



**Figure 1: Sequence diagram illustrating the process of querying ChatGPT via Microsoft Teams, facilitated by automation middleware.**



**Figure 2: Workflow illustrating the integration of ChatGPT with Microsoft Teams via middleware for automated responses.**

### 3.1 Implementing ChatGPT to discussion groups, introducing "Kiran"

In the pedagogical approach adopted for the course, students were segmented into small discussion groups. Within these confines, half of these groups were given the opportunity to harness the capabilities of ChatGPT on Microsoft Teams. This allowed students to pose questions directly to ChatGPT during sample solution sessions, experimenting with a variety of prompts. ChatGPT started functioning when triggered by the keyword "Kiran." This particular keyword was selected because it was the nickname the system had chosen for this purpose. While the course also permitted the use of other AI-driven language models for coursework, instructors refrained from offering additional training or guidelines to optimize these tools.

The integration of ChatGPT into Microsoft Teams was realized through the Power Automate Cloud Flow, as illustrated in Fig. 1. When a student enters a query embedded with the specific keyword, in our case "Kiran", into Microsoft Teams, the communication

platform detects this keyword, activating the automation flow. The query is then dispatched to ChatGPT via an HTTP POST request. Upon processing, ChatGPT sends a response which, after preprocessing in Power Automate, is relayed back to the communication platform.

**Scenario:** Imagine a student in the Teams discussion group typing: "Can you Kiran elucidate Dijkstra’s algorithm for me?" Upon detecting the keyword "Kiran", the automation flow directs the query to ChatGPT. After processing, a response is generated by ChatGPT. This response undergoes preprocessing in Power Automate to ensure the absence of the triggering keyword before being relayed back to the student within Teams.

The successful deployment of this integration required a meticulous setup process to ensure fluid communication between Microsoft Teams, Power Automate, and ChatGPT. This process is depicted in Fig. 2. Initially, the developer procured API access from OpenAI. Once OpenAI provided the API access, the developer orchestrated the automation flow within Power Automate. This flow was designed to detect mentions of the keyword "Kiran" in Teams

and dispatch a POST request to ChatGPT. Before relaying the response from ChatGPT back to the Teams channel, it was preprocessed in Power Automate to ensure that the keyword "Kiran" did not appear in the response, preventing potential looping calls. The final step involved validating the integration within Microsoft Teams.

## 4 METHODOLOGY

### 4.1 Data collection

The data was collected by two electronic surveys on the basic web course at Tampere University during spring 2023. The surveys consisted of open-ended questions and Likert-scale [1...5] questions. The first survey had 5 open-ended questions, and the second survey had 8 open-ended questions. Both surveys had eight Likert-scale questions. The first survey was performed during the first weeks of the course, and the second survey at the end of the course. Answering to the surveys was voluntary but for answering each open-ended question, student gained one exercise point, and answering to all the Likert-scale questions gave them two points in total. The required number of exercise points for passing the course was more than 400, so the effect of gained points for student's overall performance on the course was rather symbolic.

### 4.2 Data analysis methods

Student's responses are analysed using both quantitative and qualitative methodologies with experimental variables of initial classification of students to experimental and control groups, and also considering the split between pre- and post-surveys.

**Experimental variables:** the responses are split into two sets of variables. The first set of experimental variables considers initial group classification based on access to Kiran (the chatbot): the experimental and the control group. The experimental group consists of students who attended the exercise sessions, where Kiran was available. On the other hand, students who did not have access to Kiran belong to the control group. This group still had access to other language models outside our platform and therefore our experimental setting for the control group was not pure.

The second set of variables are: 'pre', 'pre (post exists)', 'post'. Here, the data has been split according to the time when the surveys were held. The terms 'pre' and 'post' refer to the data collected at the pre-survey and the post-survey respectively. The term 'pre (post exists)' represents pre-survey data from students who completed both the pre- and the post-surveys. This variable helps in understanding the effect of dropout rates on the results, ensuring that comparisons are made between consistent sets of respondents. There were 190 respondents in the pre-survey. 'Pre (post exists)' is the reference class that contains those 109 students' pre-survey responses that had replied also in the post-survey. Of the 111 students who completed the post survey, 2 did not partake in the pre-survey.

**Quantitative analysis:** same seven Likert-scale questions were asked both in pre- and post-surveys. The answers of these questions were analysed using quantitative measures and represented as horizontal bar charts.

**Qualitative content analysis:** Student replies to the open-ended questions are examined using qualitative content analysis [15]. Each individual response serves as the unit of analysis. It's

important to note that a single student's reply could pertain to multiple topics. The team agreed upon the coding scheme and the level of abstraction. The summarizing phase included breaking down the research material. While each researcher determined the categories separately, a certain degree of alignment emerged during the review process. The analysis involved noting the sentiment, i.e., the way the person was talking about the topic of some specific category. These sentiments were divided into negative, neutral, and positive. The categories and sentiments were quantified and represented as horizontal bar charts, which helps to quickly depict the different categories (topics) that were discussed and the general mood constructed by the sentiments.

## 5 RESULTS

The dataset comprises two surveys: a pre-survey and a post-survey. The pre-survey was completed by 190 students, while the post-survey was completed by 111 students. Notably, 109 students responded to both surveys.

### 5.1 Students' awareness levels on AI-based chatbots

In the pre-survey, we opened with the following question:

- Q0: "How did you hear about chatGPT?"

All but two students responded that they had already heard about ChatGPT, it being introduced in multiple channels during spring 2023. The main sources of hearing of ChatGPT were news, social media (YouTube, TikTok, Reddit, LinkedIn), and friends. Notably, one student expressed a sense of astonishment, questioning how anyone could have remained unaware of ChatGPT's existence.

### 5.2 Analysis of AI-powered learning tool usage

We divided students into two distinct groups: the experimental group, who had exposure to the AI tool through teams, and the control group.

**5.2.1 Prior to the Course.** Our initial exploration revolved around the frequency of AI-powered learning tool usage before the commencement of the course. Students responded to the pre-survey question:

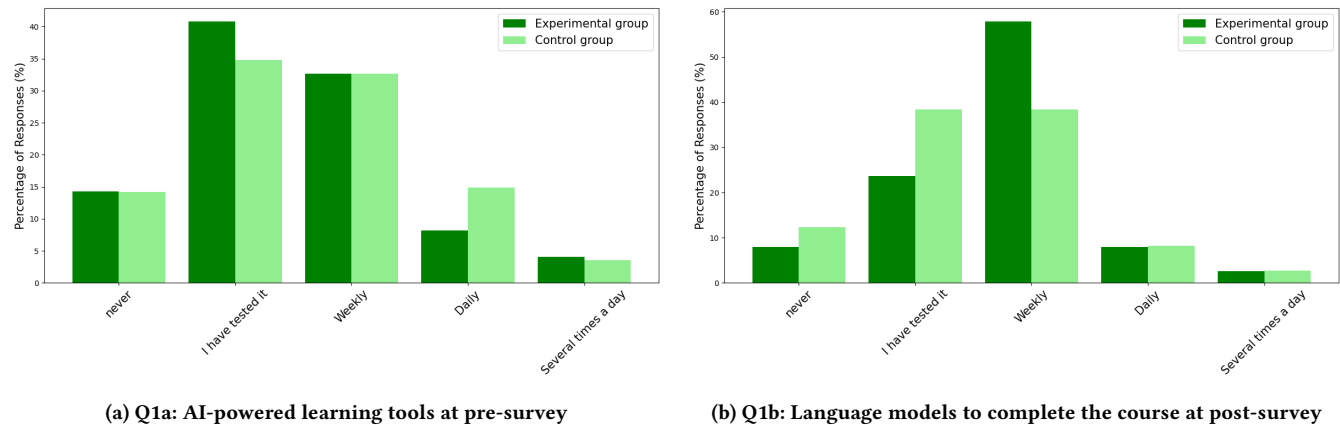
- Q1a: "I use AI-powered learning tools"

The possible responses spanned 'never', 'tested', 'weekly', 'daily', and 'several times a day'. The results are illustrated in Fig. 3a.

In the experimental group, 40.82% had "tested" the tools, while the control group registered 34.75%. Furthermore, daily usage was more prevalent among users from the control group at 14.89%, compared to experimental group users' 8.16%. Other categories like 'never' and 'weekly' drew close parallels between the groups.

Following a series of statistical assumption tests, we deduced the absence of compelling evidence to indicate a significant disparity in the median frequency of AI-powered learning tool usage between the two groups.

**5.2.2 During the course.** Our focal point during the course was the extent to which these AI-powered tools, specifically language models, were utilized. Students were prompted with the post-survey question:



**Figure 3: Q1 a and b: frequency of usage. Bar chart depicting the distribution of responses between the experimental (dark green) and the control group (light green).**

- Q1b: "How often did you use chatGPT or any language model to complete this course?"

The response landscape mirrored the pre-survey, with options like 'never', 'tested', 'weekly', 'daily', and 'several times a day'. Fig. 3b delineates the findings.

In the post-survey, the experimental group predominantly used the tools on a weekly basis, accounting for 58%, while the control group stood at 38%. Conversely, 38% of the control group had "tested" the tools, in contrast to the experimental group's 24%.

Subsequent statistical assumption tests reflected a lack of evidence to assert a significant distinction in the median frequency of ChatGPT usage between the two groups.

**5.2.3 Implications.** Despite observable variations in the usage patterns across both pre-survey and post-survey groups, our statistical analyses did not substantiate these differences as significant. To uphold the integrity of our research, we abstained from drawing stark distinctions between the groups in subsequent sections of this research paper. This ensures our conclusions are firmly rooted in statistical rigor, effectively mitigating potential biases.

### 5.3 Experiences on ChatGPT while collaborating

In order to get insight on students' experiences on ChatGPT while collaborating, we set the following question to all the students in the post-survey:

- Q2: "Did you collaborate with others while using ChatGPT or other language models? How did that impact your experience?"

Only small fraction of all the respondents claimed that they had collaborated with others while using ChatGPT. Of those students that said 'yes' for collaborating while using ChatGPT, some students said that it was useful for writing and finding explanations, helped them to see good examples of use-cases, or they discussed the outputs of the ChatGPT. On the other hand, some answers considered that it was difficult to collaborate this way, or it was confusing to use. A couple of students replied that they did not

understand the question, as they felt that using ChatGPT while collaborating does not differentiate from using Google search and none would ask on that.

### 5.4 Students' responses to Likert-scale questions

Seven Likert-scale questions were given to students in both the pre- and post-surveys. Students' responses are for the pre-survey (N=190) are shown in Fig. 4, and for the post-survey (N=111) are shown in Fig. 5. The summary of statistics showing the averages for the 'pre', 'pre (post exists)', and 'post' are shown in Table 1. It can be seen that there has not been statistically significant changes in students' responses between 'pre', 'pre (post exists)' and 'post'. Fig. 4 and Fig. 5 show that students do not trust that AI-powered chatbots would protect their personal information. Although, the amount of trust has slightly decreased from the pre-survey to post-survey, it is not statistically significant. Students are not that much concerned about the possibility of surveillance and do not fear of extra control as they were showing lack confidence on chatbots protection on personal information. Higher amount of students would continue using chatbots if conversations were being monitored or recorded than would not continue using. Students do not think that ChatGPT would be a reliable source of information. Students do not think that AI-powered chatbots would hinder their ability to persevere when facing challenging academic tasks, and even higher amount of students are not concerned about AI-powered chatbots impact to develop critical thinking and analytical skills. Most students would prefer interacting with a human teaching assistant to an AI-powered chatbot.

### 5.5 Responses to open-ended questions on learning and creativity

Both pre- and post-surveys included open-ended questions, which are handled in the following subsections by using content analysis that was introduced in Chapter 4.2.

**5.5.1 Implications for understanding and learning.** The primary focal point of concern pertains to the influence exerted by ChatGPT



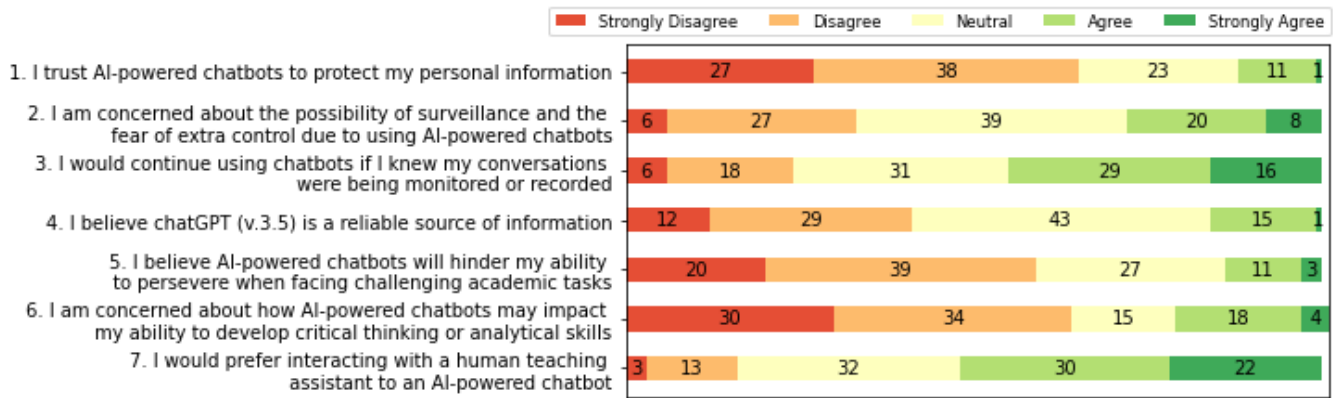


Figure 4: 190 students' responses on pre-survey to Likert-scale questions on percentages.

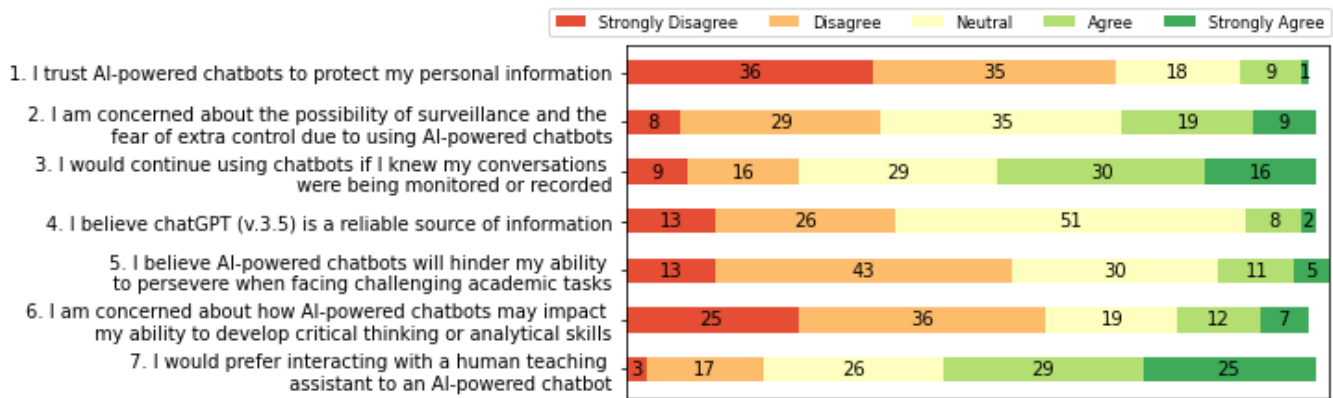


Figure 5: 111 students' responses on post-survey to Likert-scale questions on percentages.

Table 1: Summary of statistics for the Likert-scale questions.

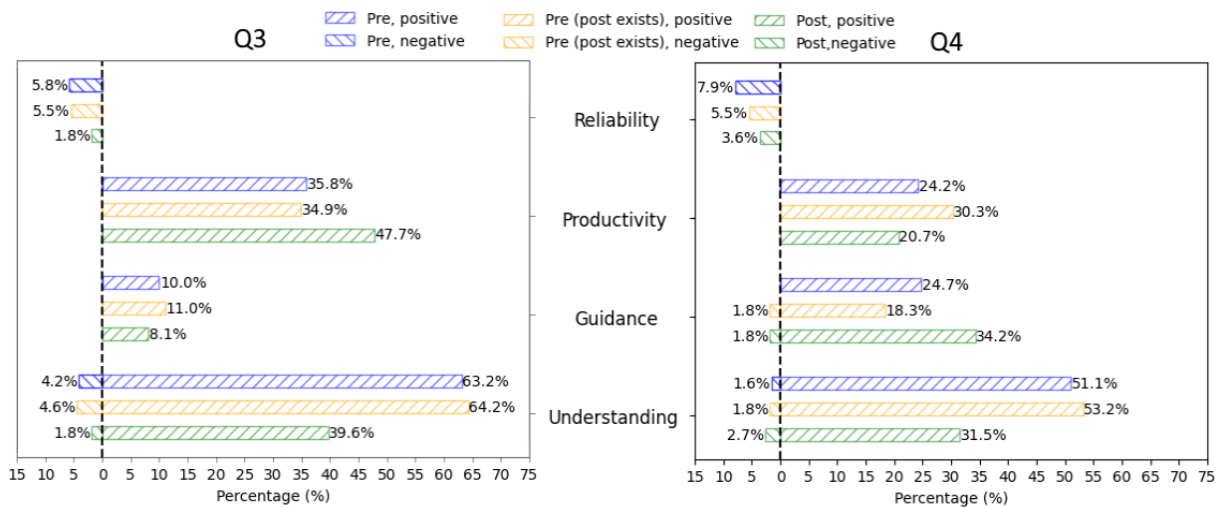
	'pre'	Averages 'pre' (post exists)'	'post'	average diff. of 'post' and 'pre (post exists)'	std of the diff.
1. protect personal information	2.21	2.29	2.03	-0.25	1.00
2. surveillance and extra control	2.96	2.93	2.92	-0.02	1.23
3. would continue despite monitoring	3.32	3.35	3.28	-0.06	1.08
4. reliable information source	2.65	2.78	2.60	-0.17	0.80
5. hinder ability to persevere	2.36	2.47	2.52	0.06	1.11
6. critical thinking or analytical skills	2.32	2.49	2.39	-0.14	1.29
7. prefer teaching assistants	3.55	3.61	3.57	-0.06	1.03

on students' comprehension and learning processes. In light of this concern, we posed the following questions in both pre- and post-survey phases:

- Q3: In what ways do you think ChatGPT can assist you in understanding and retaining information covered in lectures or course materials?

- Q4: How do you think ChatGPT can be utilised to further enhance your learning experience outside of traditional classroom settings?

Q3 primarily emphasises reflections on the language models abilities in enhancing learning and understanding in school environment, while Q4 prompts students to consider didactic applications of language models beyond the confines of traditional classrooms. However, certain students encountered difficulty in addressing both



**Figure 6: Q3/Q4: Enhancing traditional/non-traditional learning experience. 'Pre' (N=190), 'pre (post exists)' (N=109) and 'post' (N=111) groups' responses categorised.**

questions; they perceived the questions as indistinguishable and had already exhausted their feedback in response to Q3. As a result, the main four categories addressed by students in these questions overlap; nonetheless, the responses differ in quiddity.

Fig. 6 showcases the feedback for ChatGPT concerning Q3 and Q4, respectively. The data has been segmented into primary categories, and is sourced from the pre-survey, post-survey, and 'pre (post exists)' subsets. To determine the statistical significance of differences in proportions across the datasets, Z-tests for two proportions were conducted. A p-value less than 0.05 was considered indicative of statistical significance. Significant findings are found in Tables 2 and 3. We first examine the positive feedback expressed by the students.

#### Positive feedback highlights for Q3:

**Understanding:** ChatGPT's capacity to amplify comprehension drew commendation from most in the pre-survey, with 63.2% of respondents providing positive feedback. However, in the post-survey results, that number had decreased to only 39.6% of the 111 respondents. This variation in feedback was observed in both comparisons with post data to be statistically significant, as indicated in Table 2 with  $p < 0.001$ .

**Productivity:** participants lauded ChatGPT's summarising prowess. The uptick from 35.8% to 47.7% in positive feedback from the pre-survey to post-survey was significant, corroborated by the data in Table 2.

**Guidance:** A minority of students highlighted ChatGPT's ability to provide new perspectives and tutoring. There were no statistically significant changes between the pre- and post-survey respondent mentions in this aspect.

**Positive feedback highlights for Q4:** Understanding: Participants acknowledged ChatGPT's role in deepening subject understanding. However, much like in Q3, what followed was a significant decline in student confidence in this regard. This is indicated in

Table 3 with  $p < 0.001$  between the pre-survey and post-survey, (ii) and between the 'pre (post exists)' and 'post' groups with  $p < 0.01$ .

**Productivity:** ChatGPT's utility in coding-related tasks was emphasised. Such tasks included debugging and creation of code, as well searching for answers more proficiently. Positive feedback was observed in 24.2% of responses in the pre-survey, 30.3% in the 'pre (post exists)' data, and 20.7% in the post-survey. The statistical analysis revealed no significant difference.

**Guidance:** the tool's round-the-clock accessibility and non-judgmental nature were highlighted. 24.7% of pre-survey responses were positive, compared to 18.3% in the 'pre (post exists)' data and 34.2% in the post-survey. Feedback between the 'pre (post exists)' and post-survey showed a significant difference, corroborated by the data in Table 3 indicating  $p < 0.01$ . A notable response described ChatGPT in the following manner: *A helpful, kind of know-it-all person to have a chat with* (Response 18)

#### Negative feedback - unreliability and safety concerns :

In the pre-survey data, approximately 9.74% of respondents provided negative feedback. This percentage slightly decreased in the post-survey data to 5.86% and was nearly consistent in the 'pre (post exists)' data at 9.63%. The feedback was predominantly centered around the following themes:

- **Reliability:** concerns were raised regarding ChatGPT's inconsistency or unreliability in responses, as well as uncertainties about its underlying mechanisms.
- **Understanding:** respondents emphasised the value of learning through personal errors and individual effort rather than relying on ChatGPT's ready-made answers.
- **Guidance:** A segment of participants leaned towards traditional classroom settings, expressing reservations about increasing reliance on remote teaching methods and chatbot utilities.

The issue of poor reliability emerged as the most frequent concern raised by users, which was often accompanied by otherwise



**Table 2: Q3: Enhancing traditional learning experience. Statistical significance of positive feedback across different comparisons.**

Comparison	Understanding	Productivity	Guidance	Reliability
'pre' vs 'post'	p < 0.001	p < 0.05		
'pre' vs 'pre (post exists)'				
'pre (post exists)' vs 'post'	p < 0.001			

**Table 3: Q4: Enhancing non-traditional learning experience. Statistical significance of positive feedback across different comparisons.**

Comparison	Understanding	Productivity	Guidance	Reliability
'pre' vs 'post'	p < 0.001			
'pre' vs 'pre (post exists)'				
'pre (post exists)' vs 'post'	p < 0.01		p < 0.01	

positive evaluations of ChatGPT. For certain students, the unreliability of the system constituted a significant obstacle that prevented them from using it:

*As I can't trust if the information is correct, I don't think it really can (Response 87)*

.. or finding it not fit for serious use:

*Well maybe in the (near) future, but right now its still way too unreliable to be used in for example corporate learning environment. Right now its better as a creatative tool .*

In addition, students noticed novel possibilities or advantages that arose from its sporadic unreliability: *And what I see as a pro is that since its answers can never be trusted, I need to check them and test them before I'm sure they work. In my opinion that's almost like peer reviewing something which is almost always beneficial.* (Response 139)

In terms of safety concerns, the system's lack of transparency was considered as a distinct perspective. Students expressed the need for introductory tutoring on ChatGPT and the principles underlying large language models. Furthermore, they advocated for broader societal deliberations on regulations and legislation pertaining to the usage of chatbots.

Several respondents expressed concerns that relying too heavily on ChatGPT might disrupt genuine learning processes, emphasising the importance of learning from one's own mistakes:

*..I would learn better when I will make mistakes (Response 44, in Q3)*

*I don't believe it can assist if thinking is outsourced (Response 110, Q3)*

Furthermore, students expressed discontent with the current trajectory of change, perceiving it as overly rapid, and advocated for the preservation of traditional teaching methods involving face-to-face classroom interactions.

**5.5.2 Implications to creativity.** Another important factor is the potential effects on creativity that were exhibited while completing university assignments. The question that we will tackle next is:

- Q5: "How do you think ChatGPT will affect your creativity when completing university assignments?"

From the responses, we extracted four main topics, namely "Understanding", "Productivity", "Idea Generation", and "Problem Solving". These encapsulate in essence the majority of the student answers. Besides these topics, some students expressed positive feedback regarding ChatGPT's ability to inspire and be overall useful. These topics were not included in the visual depiction due to a low number of respondents discussing them. Furthermore, the subset of students who did not report any experiences or awareness of ChatGPT were excluded from the analysis as they did not provide relevant insights at this particular time.

Fig. 7 draws upon the main four topics, presenting a visual comparison between them, the two questionnaires as well as the positive and negative sentiments contained. This depiction aids in understanding the dynamic evolution of student perceptions as they gained more experience with ChatGPT during the course.

#### **Understanding concepts:**

*ChatGPT will give me versatile answers which enlarge my viewpoint about the subject.* (Pre-survey, Student Answer 118, positive feedback)

*It will help to solve harder problems and then help to learn further.* (Student Answer 53, post-survey, positive feedback)

#### **Idea generation:**

*It can give me ideas and give me new standpoints and thoughts, thus affecting my creativity.* (Pre-survey, Student Answer 155, positive feedback)

*It works as an idea generator so it helps in that sense. On the other hand, it makes it less necessary to use own thinking in some challenges.* (Post-survey, student answer 105, positive and negative feedback)

#### **Productivity:**

*It acts as a faster Google, finding answers and solutions to errors and bugs more efficiently.* (Pre-survey, Student Answer 5, positive feedback)

*ChatGPT could expand vocabulary or make a student more lazy depending on how the student is using it.* (Pre-survey, student Answer 46, positive and negative feedback)

#### **Problem Solving:**

*It will allow me to solve problems in new ways that I could not do before. Some of them may be less or more efficient and some will be technically on another level that I would have had to study much longer to be able to implement.* (Pre-survey, student Answer 75, positive and negative feedback on problem solving)

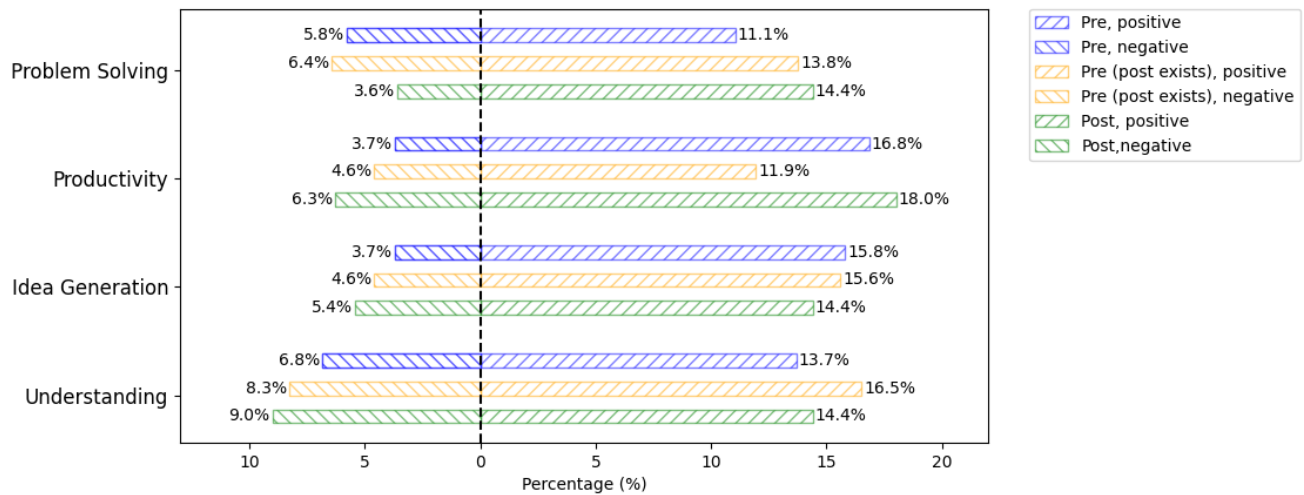


Figure 7: Q5: creativity topics. 'Pre' (N=190), 'pre (post exists)' (N=109) and 'post' (N=111) survey responses categorised.

*It will give me some more ideas so that I can start with my study easily. However, if I overuse it, then I may also become too lazy to do things on my own.* (Post-survey, Student Answer 13, positive and negative feedback)

Fig. 7 shows some visual disparity between the pre- and post-surveys. However, the analysis of Q5 data demonstrates no significant differences between the pre- and post-surveys ( $p > 0.05$ ) in the creativity categories. In the pre-survey, a predominant presence of positive sentiments was observed across all topics, with a specific emphasis on productivity, as reported by 39 participants. The post-survey data may reveal a discernible increase in concerns regarding productivity and comprehension of concepts, while the positive comment rates were maintained at the pre-survey level. Nonetheless, positive sentiments still outweigh negative sentiments, suggesting the language model's potential impact on students' creativity to be overall positive. Students appreciate ChatGPT for enhancing idea generation and productivity, as over 80% of mentions convey positivity.

## 5.6 Encountered challenges while using ChatGPT and how students did overcome those

In the post-survey, the following question was set to students:

- Q6. Did you encounter any challenges while using chatGPT or other language models? How did you overcome them?

The amount of students that described encountering some challenges while using ChatGPT was around 70%. Most of those with challenges reported that ChatGPT did not understand the question or the context of the issue and it was giving false answers or did not answer the question itself and explained phenomena around it. Those students overcame the challenge by reforming their inputs into more broader or specific questions. Students also noted that as they used ChatGPT more, their understanding of its limitations and

proficiency in prompting improved. Students also replied that if they were not able to get satisfactory outputs after some iterations of the question, they completed their tasks in other ways. Around 15% of students' responses considered that one should not blindly believe the outputs of the ChatGPT. One student said that since he was unsure about the correctness of the answer he needed to contact a real person to verify it. Three students mentioned that ChatGPT is a large language model and different answers are given at different times. Also, one student mentioned that he got contradictory examples. Three students said that they were able to use it only for simple and general answers.

We set the following question to students on the post-survey:

- Q7: "How would you like to get trained to use ChatGPT or other language models?"

Around one half of the students would prefer some training, and many mentioned teaching or training either for prompting, efficient usage or critical and ethical usage of ChatGPT that takes into consideration also the drawbacks of ChatGPT and similar systems. The other half of the students replied that either they don't need training or don't know.

## 6 DISCUSSION

Students' feedback regarding ChatGPT's capabilities presented a mixed picture. Many initially appreciated its ability to enhance comprehension, but this sentiment waned in the post-survey phase. This shift suggests that while the initial impressions of ChatGPT were positive, prolonged exposure might have led to a more nuanced evaluation of its capabilities. Conversely, ChatGPT's summarisation skills received increased positive feedback, indicating its potential in assisting students with information processing.

One drawback of large language models is the generation of non-existent data or misinformation. Students' ability to verify and validate the outputs of the model, such as testing the code snippet

or referring to existing peer-reviewed academic literature, should be prioritised in education. This was also highlighted in some of the students' answers, emphasising the need for training in the critical, efficient, and ethical use of ChatGPT. Security and privacy issues are well-known drawbacks of large language models. The findings in this paper indicate that students lack confidence in AI-powered chatbots to safeguard their personal information. Implementing large language models locally could be a solution to concerns linked to external systems collecting students' data.

A primary constraint of our study was the absence of quantitative usage data. The lack of concrete metrics on ChatGPT usage left gaps in our understanding. We could not gauge the frequency or depth of student interactions with ChatGPT, the nature of their queries, or their engagements. Such data would have been pivotal for a more holistic understanding and helped interpret their qualitative replies. Another essential aspect to consider is the potential bias introduced by the phrasing of some survey questions. Questions such as "In what ways do you think ChatGPT can assist you in understanding and retaining information covered in lectures or course materials?" and "How do you think ChatGPT can be utilised to further enhance your learning experience outside of traditional classroom settings?" presuppose positive contributions of ChatGPT to the learning experience. Such phrasing could have inadvertently led respondents to primarily consider and report positive aspects, overshadowing any reservations or criticisms they might have had.

In wrapping up, the integration of ChatGPT and similar language models into educational settings is a double-edged sword. They hold promise in terms of enhancing comprehension and productivity, but their implementation requires careful consideration to ensure they augment the learning process rather than detract from it. The observed patterns of ChatGPT usage and the mixed feedback from students underscore the need for a deeper exploration of the long-term effects of integrating such language models into educational settings. While our study lays the groundwork, future research should delve into the nuances of student interactions with these tools over extended periods and in diverse learning contexts.

## 7 CONCLUSIONS

This paper explores the use of ChatGPT, a language model, as a tutoring chatbot in a computer science course. The study examines students' perception of ChatGPT and its impact on their learning. The results show that students have used AI-powered chatbots for various tasks, such as debugging code and enhancing comprehension. While there were some reliability issues with the outputs, the positive implications of using ChatGPT outweighed the negative ones.

However, it is already evident that these advancements will have a profound impact on various fields. In education, instructors are forced to re-evaluate their teaching and especially testing methods in order to be able to reliably measure students' knowledge levels and skills. Students have already shown creative approaches of using the new large language technologies to simplify and explain large text sets and by implementing applications to automate the easy tasks. This paper studied advantages and disadvantages of ChatGPT in higher education from students' perspective, and the answers to the formulated research questions are as follows:

- **RQ1: Awareness level:** Students have at least heard about it and use it for multiple tasks.
- **RQ2: Frequency of usage, use cases:** Over half of the students reported using AI-powered chatbots on a weekly basis, while many who initially indicated they had only 'tested' the technology later revealed more frequent usage beyond preliminary trials. Students used ChatGPT mainly for theory questions, code generation, debugging, and tutoring. The other main use-cases included searching after examples and requesting summaries for long texts. Also, few students had some innovative approaches such as whisper:speech-to-text, which allowed students to concentrate fully on lectures. Students were hesitating in using ChatGPT in group settings.
- **RQ3: Learning implications, implication to creativity** The vast majority emphasises the positive impacts to learning, such as getting the material interpreted in alternative ways, quick replies, helpful summaries and examples. Challenges comprise the unreliability of ChatGPT's output and students needing to iterate their prompts to improve the outputs. For a few students, the prompting challenges and reliability issues were show-stoppers for using ChatGPT. In addition, a few students were concerned about laziness, and the missed opportunities to learn from mistakes. While some participants see ChatGPT as a helpful tool that can enhance creativity, provide guidance, and save time, others have reservations about the potential for misuse and its impact on autonomy and learning. It is important for users to strike a balance and use ChatGPT responsibly, ensuring that it complements their learning process rather than replacing it.
- **RQ4: Security and privacy concerns:** Students remain sceptical about ChatGPT's ability to safeguard their personal information, yet this doesn't deter them from utilising it.

## REFERENCES

- [1] [n. d.]. About Smart Sparrow. <https://www.smartsparrow.com/about/>. (Accessed on 06/08/2023).
- [2] [n. d.]. ChatGPT banned in Italy over privacy concerns - BBC News. <https://www.bbc.com/news/technology-65139406>. (Accessed on 06/05/2023).
- [3] [n. d.]. Sciences Po bans the use of ChatGPT without transparent referencing - Espace Presse Sciences Po. <https://newsroom.sciencespo.fr/sciences-po-bans-the-use-of-chatgpt/>. (Accessed on 06/05/2023).
- [4] [n. d.]. Universities adjust to ChatGPT, but the 'real AI' lies ahead. <https://www.universityworldnews.com/post.php?story=20230301105802395>. (Accessed on 06/05/2023).
- [5] [n. d.]. University of Hong Kong temporarily bans students from using ChatGPT, other AI-based tools for coursework | South China Morning Post. <https://www.scmp.com/news/hong-kong/education/article/3210650/university-hong-kong-temporarily-bans-students-using-chatgpt-other-ai-based-tools-coursework>. (Accessed on 06/05/2023).
- [6] Ariel D. Anbar, Chris Mead, Donald Bratton III, Lev Horodyskyj, Jacqui Hayes, David Schonstein, Sara Watt, Keith Watt, Dror Ben-Naim, and Alfredo Leon. 2017. Demonstrating the value of education through exploration as a theory of digital design. In *AGU Fall Meeting Abstracts*, Vol. 2017. ED41B-0275.
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33 (2020), 1877–1901.
- [8] Miguel Ángel Conde, Ángel Hernández-García, Francisco J. García-Peñalvo, and María Luisa Sain-Echaluze. 2015. Exploring student interactions: Learning analytics tools for student tracking. In *Learning and Collaboration Technologies: Second International Conference, LCT 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings 1*. Springer, 50–61.

- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [10] Susan E Embretson and Steven P Reise. 2013. *Item response theory*. Psychology Press.
- [11] Arthur C. Graesser, Patrick Chipman, Brian C. Haynes, and Andrew Olney. 2005. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education* 48, 4 (2005), 612–618. <https://doi.org/10.1109/TE.2005.856149>
- [12] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, Stepha Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [13] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
- [14] Lingyao Li, Zihui Ma, Lizhou Fan, Sanggyu Lee, Huizi Yu, and Libby Hemphill. 2023. ChatGPT in education: A discourse analysis of worries and concerns on social media. *arXiv preprint arXiv:2305.02201* (2023).
- [15] Philipp Mayring. 2014. Qualitative content analysis: theoretical foundation, basic procedures and software solution. (2014).
- [16] Alexandros Papadimitriou. 2023. Adaptive and Intelligent MOOCs: How they Contribute to the Improvement of the MOOCs' Effectiveness. *International Journal of Research in E-learning* (2023), 1–21.
- [17] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
- [18] Brian J. Reiser, John R. Anderson, and Robert G. Farrell. 1985. Dynamic Student Modelling in an Intelligent Tutor for LISP Programming. In *IJCAI*, Vol. 85. 8–14.
- [19] Kevin Roose. 2022. The brilliance and weirdness of ChatGPT. *The New York Times* (2022).
- [20] Andrej Thurzo, Martin Strunga, Renáta Urban, Jana Surovková, and Kelvin I Afrashtehfar. 2023. Impact of artificial intelligence on dental education: a review and guide for curriculum update. *Education Sciences* 13, 2 (2023), 150.
- [21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30 (2017).

Received 15 June 2023; revised 31 August 2023