

Case Study

A survey on students' use of AI at a technical university

Fadoua Balabdaoui³ · Nora Dittmann-Domenichini² · Henry Grosse³ · Claudia Schlienger² · Gerd Kortemeyer¹

Received: 15 December 2023 / Accepted: 9 May 2024

Published online: 14 May 2024

© The Author(s) 2024 [OPEN](#)

Abstract

We report the results of a 4800-respondent survey among students at a technical university regarding their usage of artificial intelligence tools, as well as their expectations and attitudes about these tools. We find that many students have come to differentiated and thoughtful views and decisions regarding the use of artificial intelligence. The majority of students wishes AI to be integrated into their studies, and several wish that the university would provide tools that are based on reliable, university-level materials. We find that acceptance of and attitudes about artificial intelligence vary across academic disciplines. We also find gender differences in the responses, which however are smaller the closer the student's major is to informatics (computer science).

Keywords Artificial Intelligence · Student Survey · GPT · Assessment · Teaching and learning

1 Introduction

Following the public availability of ChatGPT [33] in November 2022, the sudden prominence of artificial intelligence (AI), in particular Large Language Models (LLMs), has resulted in considerable public discourse [30]. Technical universities, often at the forefront of technological evolution, are facing unique challenges and opportunities integrating the widespread availability and rapid improvement of AI into their curricula and assessments [6]. Not only do students need to learn *about* AI, but also *with* AI, as the systems can pass introductory STEM courses [23] and even freshman years at elite universities [5].

Before making rash decisions if and how curricula and assessments of learning outcomes need to be adjusted, it is imperative to better understand how students are already incorporating these tools into their learning journeys [13, 40, 44], and thus several studies have already been conducted within the short amount of time that these tools were available [43]. There are definite cultural and area-of-studies differences in the acceptance of these technologies [21]; this study explores how students at a large technical university in the German-speaking part of Switzerland perceive AI. In particular, we investigate patterns of utilizing AI tools, and gather students' insights into how university education should evolve in response to the growing influence of AI.

As the next generation of engineers, scientists, and thought leaders is moving through higher education, it needs to prepare for a workplace where AI might be ubiquitous [26, 36], as well as a society that will be strongly influenced by these tools [37]. While AI will likely augment various educational tools and methodologies [4], it is paramount to understand if students find these advancements beneficial, challenging, or perhaps even superfluous. Furthermore, with AI potentially

✉ Gerd Kortemeyer, kgerd@ethz.ch; Fadoua Balabdaoui, fadoua.balabdaoui@stat.math.ethz.ch; Nora Dittmann-Domenichini, nora.dittmann@let.ethz.ch; Henry Grosse, hgrosse@student.ethz.ch; Claudia Schlienger, claudia.schlienger@let.ethz.ch | ¹Rectorate and AI-Center, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland. ²Educational Development and Technology, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland. ³Seminar for Statistics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland.



redefining job roles and research paradigms, gauging students' preparedness and adaptability can offer critical insights for educational institutions aiming to remain at the vanguard of technological education.

Moreover, as AI increasingly permeates our daily lives and professional realms, it is inevitable for educational paradigms to shift. This survey thus also probes into students' beliefs regarding the direction and magnitude of these changes. Should universities place greater emphasis on interdisciplinary studies, combining AI with traditional disciplines? Is there a growing demand for ethics in AI to be a cornerstone of technical education [31]?

2 Methods

2.1 Setting

ETH Zurich is a large, technical research university, which generally appears prominently in international rankings. It focusses on STEM disciplines, but also offers architecture and humanities. The university has around 25,000 students from 120 countries, about a third of which are female. Admission is highly selective for international students, but open without restrictions to all students with a Swiss high school diploma.

2.2 Data collection

The survey was electronically distributed to the student body in September 2023, coinciding with the start of the fall semester. All submissions were anonymous, and a total of 4798 replies were received, which is approximately one fifth of the student body at ETH.

Recognizing the exploratory nature of the survey, it offered ample opportunity for free-form responses. Students were exceptionally eager to share their insights, concerns, and enthusiasm, as they contributed over 500 pages in 9.5pt proportional font of free-form comments.

2.3 Survey variables

Tables 1 and 2 list the labels, descriptions, ranges, and alternative answers of the binary and Likert-scale items considered in this study.

A new attribute *StudyClass* was created to describe whether a student belongs in one of the groups described in Table 3 based on *StudyProgram*. Note that at ETH Zurich, the Department of Civil, Environmental and Geomatic Engineering ("Bau, Umwelt und Geomatik") is traditionally grouped with architecture instead of the other engineering disciplines. The variables *OpExcl*, *OpGenDev*, *OpGlobal*, *OpOutput*, *OpPotential*, *OpUsage*, and *OpUseTeach* all describe opinions regarding acceptance and adoption of AI, and they were added up into a new variable *OpAI*.

2.4 Representativeness

30.9% of the respondents identified as female, 65.5% as male, and 0.9% as diverse; the remaining percentage preferred not to answer. Overall, the gender ratio of the respondents is representative of that of the student body. Table 4 reports the study respondents by gender; these response rates also reflect the actual distribution of genders within the departments.

2.5 Analysis

Data in this mixed-methods study were analyzed in a variety of ways. Numerical data were analyzed using standard statistical software, in particular R, SPSS, and Excel as appropriate. For free-form data, GPT-4 [34] was used extensively to provide the authors with an overview of student responses, suggestions for data labels, and pre-classification of statements. Due to memory limitations of GPT-4, the data were chunked, summarized, and recompiled as appropriate using Python scripts and direct API-access. Significant portions of the analysis scripts were initially written by GPT-4, but subsequently verified, edited, and adapted by the authors. The preparatory work was verified by the authors, and all responses were read by at least one human; conclusions are the sole responsibility of the authors. DeepL was used for some of the translations of the original language into English, but the results were edited and corrected by the authors as appropriate.

Table 1 The labels, descriptions, ranges, and alternative answers of the binary and Likert-scale items considered in this study, Part 1

Label	Description	Range	Alternative
<i>Alperceive</i>	How should using tools such as ChatGPT be considered in written performance assessments?	As cheating; As plagiarism; As a legitimate tool; It depends	
<i>ComStrat</i>	Have you developed a strategy for using AI-based tools in your studies together with other students (e.g. shared costs for user accounts, task sharing)?	No (0), Yes (1)	
<i>ExpAlTeach</i>	Have you already experienced the use of text- or image-generating AI-based tools in the classroom at ETH?	No (0), Yes (1)	
<i>FamChat</i>	What is the level of your experience and use of text-generating AI-based tools such as ChatGPT?	None (1) - Very high (5)	
<i>FamPresGen</i>	What is the level of your experience and use of presentation-generating AI-based tools such as TOME?	None (1) - Very high (5)	
<i>FamStDiff</i>	What is the level of your experience and use of image-generating AI-based tools such as Stable Diffusion or DALL_E?	None (1) - Very high (5)	
<i>FamTrans</i>	What is the level of your experience and use of AI-based translation tools (e.g. DeepL)?	None (1) - Very high (5)	
<i>Gender</i>	Gender	Male (0), Diverse (0.5), Female (1)	No answer
<i>Grades</i>	My grades are generally...	Below average (1) - Above average (5)	
<i>HelpChat</i>	How helpful has the use of ChatGPT been for you in your studies so far?	Not helpful (1) - Very helpful (5)	Haven't used it
<i>HelpExamPrep</i>	Did the way you prepared for the exam with your "KI Tutor" make you learn better and feel more prepared?	No (1) - Yes (5)	Haven't used it
<i>ImpOff</i>	Do you think it is important that ETH offers learning opportunities for students that promote the use and application of AI-based tools in their studies?	No (0), Yes (1)	
<i>NativeLanguage</i>	What is your native language?		
<i>OpExcite</i>	The buzz around AI and AI-based tools like ChatGPT is, in my view...	Large number of options Exaggerated (1) - Appropriate (5)	Neither
<i>OpExcl</i>	Regarding the impact of AI-based tools on exclusion and discrimination, I am...	Worried (1) - Confident (5)	Neither
<i>OpGenDev</i>	In general, regarding the development and increasing use of AI and AI-based tools I am...	Averse (1) - Approving (5)	Neither
<i>OpGlobal</i>	Regarding the degree of impact of AI-based tools on global challenges such as climate change, poverty or hard-to-cure diseases, I am...	Pessimistic (1) - Optimistic (5)	Neither
<i>OpOutput</i>	In general, regarding the output of AI and AI-based tools I am...	Critical (1) - Trustful (5)	Neither
<i>OpPotential</i>	The potential and possible further developments of AI and AI-based tools make me feel ...	Uneasy (1) - Hopeful (5)	Neither

Table 2 The labels, descriptions, ranges, and alternative answers of the binary and Likert-scale items considered in this study, Part 2

Label	Description	Range	Alternative
<i>OpRules</i>	How strongly should politics regulate the use and further development of AI?	If possible, not at all (1) – As much as possible (5)	I don't know
<i>OpUsage</i>	With regard to the benefits of AI and AI-based tools, in my view what outweighs are in general the...	Disadvantages (1) – Advantages (5)	Neither
<i>OpUseTeach</i>	I see the use of text-generating AI-based tools like ChatGPT in teaching as...	Problematic (1) – Desirable (5)	Neither
<i>Problems</i>	Have you encountered any problems or concerns when using ChatGPT when it comes to accuracy, trustworthiness or bias?	No (0), Yes (1)	
<i>ProvAssess^a</i>	While preparing for exams, used AI for providing exercises to check for understanding and for recapitulation	No (0), Yes (1)	
<i>ProvDef^a</i>	While preparing for exams, used AI for providing definitions and examples for learning specialized terminologies	No (0), Yes (1)	
<i>ProvExp^a</i>	While preparing for exams, used AI for providing explanations or examples for better understanding of the learning materials	No (0), Yes (1)	
<i>ProvFlash^a</i>	While preparing for exams, used AI for providing flashcards or similar repetition tools	No (0), Yes (1)	
<i>ProvSum^a</i>	While preparing for exams, used AI for providing summaries of the learning material	No (0), Yes (1)	
<i>StudyProgram</i>	What is your study program?	<i>List of all study programs (majors)</i>	
<i>TeachObs</i>	Will current forms of teaching (face-to-face events, lectures, practice groups, internships, etc.) and assessment soon be obsolete?	Yes; No; They already are	
<i>UseAnaVis^a</i>	I am using AI for data analysis, visualisation or modelling of data	No (0), Yes (1)	
<i>UseCoding^a</i>	I am using AI for programming or simulations	No (0), Yes (1)	
<i>UseDecision^a</i>	I am using AI for decision-making (e.g. by proposing alternatives, arguments, scenarios)	No (0), Yes (1)	
<i>UseExp^a</i>	I am using AI for explanations of subject-specific facts, clarification of questions of understanding	No (0), Yes (1)	
<i>UseIdeas^a</i>	I am using AI for developing concepts or generating ideas (e.g. project designs)	No (0), Yes (1)	
<i>UseLit^a</i>	I am using AI for research or study of literature	No (0), Yes (1)	
<i>UsePrbSolv^a</i>	I am using AI for problem solving (e.g. when solving exercises)	No (0), Yes (1)	
<i>UseTexts^a</i>	I am using AI for text analysis, text processing or text generation (e.g. of summaries, written work, etc.)	No (0), Yes (1)	
<i>UseTranscr^a</i>	I am using AI for language processing (e.g. transcription of lectures, discussions, etc.)	No (0), Yes (1)	
<i>UseTransl^a</i>	I am using AI for translations or overcoming language problems, e.g. simplification of texts in a foreign language	No (0), Yes (1)	

^aPart of one question regarding exam preparation with multiple checkbox options

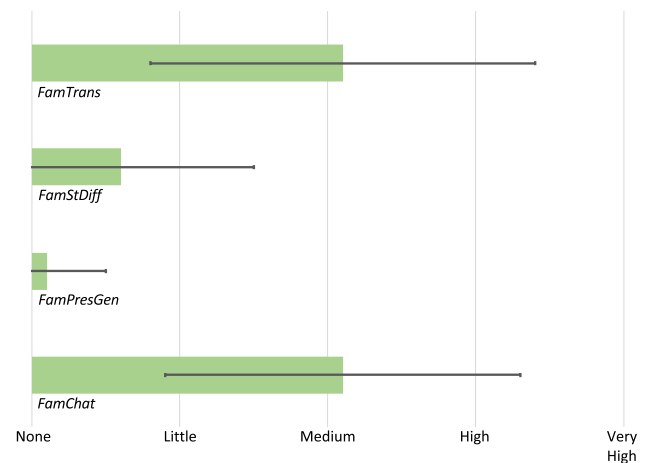
Table 3 Classification of the study programs (*StudyClass*)

<i>StudyClass</i>	Departments	#Programs	Samples
Informatics	Informatics	4	711
Engineering	Engineering without Informatics	17	1291
Arch/Civil	Architecture and Civil Engineering	12	569
NatSci	Natural Sciences	22	1164
SysSci	System-oriented Natural Sciences	14	777
–	Management and Social Sciences	6	87
–	Other	3	9

The latter two groups had too few respondents for further consideration

Table 4 Gender distribution for the departments (*StudyClass*) listed in Table 3

	Informatics (%)	Engineering (%)	Arch/civil (%)	NatSci (%)	SysSci (%)
Male	84.4	82.7	62.6	64.4	36.4
Female	15.6	17.3	37.4	35.6	63.6

Fig. 1 Familiarity of students with AI tools, showing the averages and standard deviations (attributes from Table 1)

The students did not always address particular issues within the context of the corresponding free-form questions, likely because the survey was administered in multiple, separate pages, and they did not know if they had a chance to address their concerns and suggestions later.

3 Results

3.1 Familiarity with and usage of AI

Students were asked how familiar they were with certain classes of AI tools; Fig. 1 shows the result.

On the average, students claim medium familiarity with chat and translation tools, while on the average they claim little to no familiarity with image or presentation generating tools. There are large standard deviations on these self-assessments of familiarity, indicating a wide spectrum of experience and comfort level among the students. In a free-form field, students were asked which *other* tools they were using, and 770 students responded (ironically, based on the writing style, at least one respondent likely used ChatGPT to answer this question).

This question about *other* AI tools was apparently understood by a large number of students to list and describe *any* tool they used, which students were eager to share; Fig. 2 shows the result, where several specific tools would have fallen into the general categories shown in Fig. 1. In addition to general-purpose tools, such as Google Bard [17] and ChatGPT [33] for writing tasks (both human and computer languages), Wolfram Alpha for mathematical

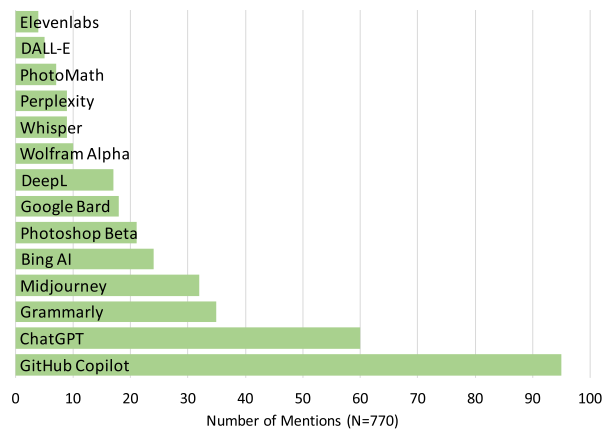
operations, Whisper for transcribing interviews, Grammarly [18] for grammar checking, DeepL [8] for language translation, and GitHub Copilot [16] for programming tasks, they are also using more specialized tools, such as ATLAS.ti [3] for qualitative data analysis, Rayyan [38] for literature reviews, and Quillbot [27] for paraphrasing. Besides Python code, students also generate Matlab code, as well as LaTeX for document typesetting and TikZ [45] for embedding plots into LaTeX documents. Finally, as these are students from a technical university, several of them remark that they train their own models, and in this context mention PyTorch [35] and CUDA [32] as tools, and BERT [9] as base. In addition to these academic uses, students also state that they are using Midjourney [29] and Elevenlabs [10] for generating social media content, and GPT for getting ideas what to cook.

Overall, students make individual choices about which tools they are using. The question if they have common strategies among peers (*ComStrat*) was denied by 96.8% of the students. The remaining 3.2% had common strategies that were mostly about sharing costs for Grammarly and GPT-4 [34] access (including the associated plugins such as Wolfram Alpha). One student stated that he or she shares a GPT-4 account with seven peers (at the time of writing, a \$20/month expense), but this means that the conversation history always needs to be cleared out.

In the context of account-sharing, there was a frequent call for ETH to purchase an organization-wide license for GPT-4 as a means of leveling the playing field. Also, several students encouraged ETH to train and provide its own Large Language Model, based on lecture materials, as a study aid.

3.2 Current usage of AI in teaching and learning

Only 17.2% of the students stated that they had experienced AI in a teaching situation (*ExpAITeach* in Table 1). The free-form responses mentioned only two instances of instructors using AI for teaching purposes: one example was explicit guidance and instruction on how to use AI for computer programming, and the other was an instructor who generated explanatory essays with ChatGPT and had the students look for the mistakes (note that providing guidance



Elevenlabs: text-to-speech generator.

DALL-E: image generation.

PhotoMath: solving math equations through image recognition.

Perplexity: search engine and research assistant for finding and summarizing academic papers.

Whisper: speech-to-text transcription.

Wolfram Alpha: computational knowledge engine for solving mathematical problems and providing answers to various queries.

DeepL: translation of text.

Google Bard: chatbot.

Photoshop Beta: photo editor with AI-based tools.

Bing AI: AI-enhanced search engine.

Midjourney: image generation and modification.

Grammarly: grammar correction and writing enhancement.

ChatGPT: chatbot.

GitHub Copilot: programming assistant for generating code suggestions and completions.

Fig. 2 Tools explicitly mentioned by students

for how to use AI is different from teaching about how AI works). However, the students had a wide range of ideas on how it could and should be used, as evident by their associated free-form statements ($N = 1701$).

Many students use it for common language tasks: overcoming initial writer's block, as well as enhancing, translating, or correcting texts in natural languages. For example, they draft the main points they want to make in simple language or bullet points and then have ChatGPT write "nicely" formulated paragraphs in the required style. Several students mentioned also writing emails to faculty this way. Foreign students also stated they it helps them overcome language barriers.

- "I am not a native English speaker, so I always write my texts and then ask ChatGPT to rewrite the whole text, paragraph, or specific sentences in a certain way. For example, more scientific and professional."

Several students used it as a "better Google" or an "entry point" to Google. It would provide helpful and targeted first responses to specific questions for which a regular Google search only provides very general results. Details of those answers can again be looked up using Google, once that ChatGPT made it clearer what to look for.

Besides operating on natural language, many students highlighted its ability with computer languages and the impact on basic programming teaching and learning:

- "In the form of Github Copilot, ChatGPT is very useful because it allows you to make, for example, Python plots for data analysis and do so entirely without prior knowledge. Furthermore, repetitive tasks in programming become writing the first line or comment and then the rest is completed automatically."
- "Programming in languages that you understand well but do not master."
- "ChatGPT can give the rough structure for code that you can then revise and refine yourself."

The same student goes on to explain that this still needs understanding of the underlying concepts:

- "From my experience, however, for a meaningful use of ChatGPT you still need the competence that is taught in their studies, because the product rarely meets the requirements of the task and must always be understood and revised."

The tool is frequently used for debugging and finding the errors in existing source code. Many students stated that they also use it to explain provided code:

- "To get documentation for source code, which is not documented otherwise."
- "Very useful for coding, particularly code understanding."
- "A lab provided a lengthy and intricate code for me to use, so I asked ChatGPT to break down the code in smaller parts and explain to me what it was doing. Based on that I was able to implement it in my own code."

Intriguingly, some students stated that they are using GPT to learn PyTorch and TensorFlow [1]; in other words, they are using AI to learn how to build and train AI solutions.

A large number of students used ChatGPT to summarize papers or lecture materials, and they stated that the summary is frequently better understandable than the original. This does not rely on pre-trained knowledge, but on the tool's ability to "calculate with words." While several students stated that ChatGPT was bad in math, and that proofs and derivations tended to be incorrect, they also stated that is helpful in explaining proofs and derivations provided by the instructors.

A frequent notion was that ChatGPT was helpful in gaining a quick overview for a new topic, functioning as an entry point for further research elsewhere.

- "ChatGPT is particularly useful in the first and last steps of a paper, i.e. initial knowledge acquisition/research and then revising/rewriting the text at the end. In between, the questions that need to be answered are usually too specific."

Another frequent notion was that ChatGPT was like a "study buddy," but without the feeling of embarrassment when asking "dumb questions;" also, it was appreciated that ChatGPT would "not feel offended when not taking its advice." Students asked for quick summaries of lecture materials "in easy words," and oftentimes found the dialogue more helpful than any attempt at engineering "the perfect prompt." In addition, several students used ChatGPT to generate practice questions for exams based on the lecture script, and they used other AI-tools to generate flashcards. In the same memorization context, one student used ChatGPT to invent mnemonics based on first letters of terms.

Overall, students found GPT most useful in computer programming, followed by biology and chemistry, and least useful in physics and mathematics.

3.3 Usage of AI for exams and written assessments

3.3.1 Legitimate tool or fraud?

Students were asked how the use of AI-based tools such as ChatGPT should currently be considered in written performance assessments, Fig. 3 shows the result.

The prevalence of “depends” is not surprising, and in the associated free-form comments many students emphasized that it depends on the rule set by the instructor:

- “Are we talking about open-book only here? Then, of course, it’s legitimate. Otherwise, of course, it’s an unauthorized aid and therefore cheating.”

Many students agreed that the use of AI-tools, like ChatGPT, is legitimate for tasks such as proofreading, correcting grammar, generating ideas, and aiding with research.

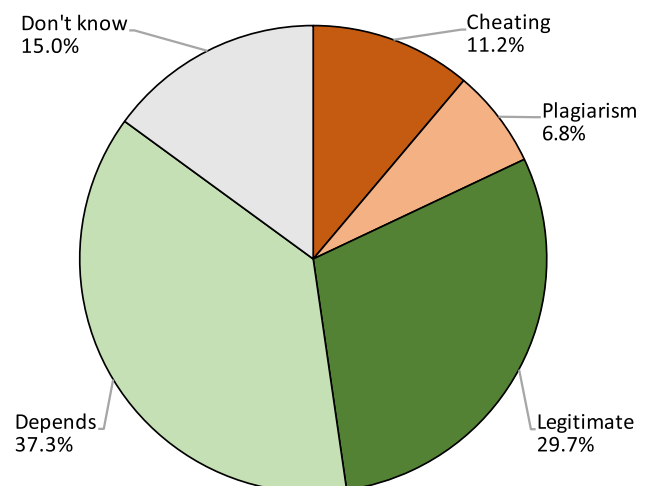
- “The use of AI-tools for written performance assessments is legitimate if it helps in brainstorming, proofreading or as a discussion partner, as most of the work is still done by the student and not just copy-pasted.”
- “It would depend on how much is human-made and how much AI-generated. If someone uses a generated text as a first draft and develops it to have a perfect answer, this is a good example of a person using ChatGPT effectively.”
- “It would be good to trust that students are not so stupid as to simply let ChatGPT write texts and hand them in. Most of us are well aware of the limitations of these tools.”

The use as a programming aid is also included in this notion:

- “For tedious tasks (especially code tasks) when AI based tools come up with code to visualize, extract data, the tool is not plagiarism.”
- “Legitimate: If you use the AI tool to create the code needed for plotting in Python. Not legitimate: If you use the AI tool to write the actual work or simply copy theoretical results. Gray area: If the heart of the work (the actual task) was creating a program and you would partially use ChatGPT for it.”

The latter delineation is similar to remarks about human language:

Fig. 3 How should the use of AI-based tools be considered in written performance assessments (*AIperception*)?



- “When the purpose of the written performance assessments is to examine language skills, AI-based tools should be viewed as cheating. But on the other hand, for works that focus on the writing itself... it’s reasonable to claim the usage of AI tools as cheating or even plagiarism.

However, they believe the primary ideas and content should come from the student. Most agreed that copying text from AI output verbatim is cheating or plagiarism.

- “When the complete text is written by ChatGPT, it would be cheating. However, to help find studies, explain terms, and revise text, I think it’s okay.”
- “It depends on what it is specifically used for. Using AI tools to write an entire report or scientific paper should be considered fraud. Using AI tools to formulate ideas is valid. As long as the generated text is read by the author and edited to the extent that what is written is the author’s intention, it is legitimate.”

Some students, though, are less concerned:

- “I am not worried that students will be able to cheat their way through their studies with AI, because they still have to work out and understand the big connections and concepts themselves.”

They also concurred on the need for transparency in the use of AI-tools with several students suggesting they should be cited.

- “ChatGPT is a simplified Google for me. The internet can be used with the source, so ChatGPT should also be OK, ideally mentioning the exact chat log or problem definition.”
- “If you use it as an information source and cite it, then it’s like having interviewed another person and thus it’s legitimate.”

One student sarcastically remarked that it is a great tool to generate senseless verbiage when content does not matter, but the assignment demands to simply put down some words.

Despite supporting the use of AI-tools, students raised concerns about over-reliance, potential data privacy issues, and the difficulty of distinguishing AI-generated work.

- “If one needs an AI crutch for their studies, one should just drop out instead.”

Another concern was being wrongly accused of using AI:

- “Last semester, unfortunately, I was mistakenly asked by a professor for an interview because she thought I had used ChatGPT for bonus assignments. [...] Of course, this is a big difficulty for teachers, but with the advent of these new technologies, I wish there was better communication when problems like this arise before wrong conclusions are drawn.”
- “What is absolutely to be avoided (for the current time) is relying on AI detector tools to judge the originality of written text. I have seen many times my own text comes up as ‘80% AI generated’ when actually it was entirely my own creation. I talked to fellow students from other Swiss universities that have been penalized for this.”

Overall, students advocate context-based use and recommend clear guidelines from educational institutions and instructors.

- “Clear rules on what is allowed. These rules should be pragmatic and embrace the fact that AI tools are an integral part of many productive workflows.”
- “I think there needs to be clear rules for the use of AI in graded homework or projects. There is a big difference between, for example, using ChatGPT as a kind of tutor or more efficient search engine to help you solve a problem on your own, and having an AI tool write entire essays for you without any input from you.”
- “Clear communication of what is allowed and what is not (has often been done well so far).”

3.3.2 Consequences for future assessments

Many students stated that they do not believe it will influence the examinations themselves, the majority of which at ETH use a traditional, paper-and-pencil type format; traditional closed-book exams would remain unaffected as digital resources are usually restricted.

- “ChatGPT should not have a direct impact on the examination process. The results are too unreliable for this and the extent of applicability is generally too poorly known.”
- “For exams, I think the pen/paper approach is still the best.”
- “ChatGPT and Co do not yet have a place in an exam (generally speaking). In 10 years this will be different, AI will then belong to everyone’s life (like today’s mobile phone), so the exam formats may need to change accordingly.”

Much more influence is expected in long-form written assessments such as theses, reports, and projects.

- “For exams I don’t see a big impact but for writing papers I do. There I think it should be defined for how much AI tools can be used.”
- “Essays and similar exercises are useless because they can be easily handed over to ChatGPT. Verbal-based exercises are moot now, and students must be provided projects/assignments that involve creating and building rather than compiling/summarizing already existing information.”

Many students stated that AI might be effective during exam preparation and assignments, and that it should be integrated in the learning process.

- “For me, these tools are a helpful platform to improve my work or even my learning.”

They see AI as a tool for human work, not a replacement of it.

- “These tools are to be regarded like a pocket calculator. Closing yourself off is not the solution.”
- “These tools are out there, and there is no going back.”

However, significant concerns were raised about potential risks and challenges associated with AI. Many are worried about its misuse leading to plagiarism and the disruption of academic integrity. They fear it could bring an unfair advantage to those who use it, or maybe can afford it, fostering an uneven academic playing field. They pointed out that students that do not want to use AI tools for various reasons including labor conditions during fine-tuning and possible copyright violations in the text corpus should not be penalized for their conscientious choices.

Students also expressed concerns about the reliability and accuracy of AI-generated content, especially for more complex academic tasks that require a higher level of expertise and specific data. Students frequently expressed the need for new regulatory guidelines concerning the use of AI, particularly to preserve fairness.

- “Basically, I am intimidated by the fact that, in the worst case, it can give you an unfair advantage, and thus diminish the performance of others who do not use the tools, especially in terms of semester performance.”
- “Written semester projects not make any sense to me anymore with ChatGPT. Students can have the assignment ready in a matter of minutes, and only need to check if the machine did a good job or not. This is not fair compared to students who refuse to use such a tool and do it ‘the old way’”

They also emphasized the need for a possible shift in exam formats to a larger focus on understanding, critical thinking, and larger application of knowledge relevant to real-world scenarios.

- “Examinations that rely on repetitive calculation and rote learning are likely to face challenges but examinations where a degree of thinking is required will likely only benefit.”
- “Exams requiring memorization can be killed off by AI, exams must be more values and insights-based.”

- “For me it means that exams have to be more aware that you either have to consider the use of AI tools (e.g. the questions have to consider the use of ChatGPT, but still require the student to be able to think), or you have to adjust the exams in a way that even a ban of ChatGPT and AI tools in general comes close to reality.”
- “This means that the assignments and theses will be a lot more polished and in-depth. Also, simple formula-based questions will be rendered ineffective and the questions have to be framed from a more conceptual standpoint.”
- “More emphasis must be placed on ensuring that the knowledge learned is networked and applied to *new* problems across departments, for example as part of individual projects. In my opinion, “standard tasks” or multiple choice for facts that can be solved with AI will bring little added value in the future.”

Considering future uses of AI, students expressed eagerness for a deep technical understanding of AI technology. Taking into account the possible AI influences, some argued for transparency in AI use and for the need to adapt educational processes, such as changing traditional exams to more task-oriented or problem-based assessments that reflect workplace realities. While there are concerns, AI’s inevitable role in the future of academics and professions was acknowledged by most students.

3.4 Trust in AI

Students are well-aware of potential trust issues surrounding AI. When asked if they had encountered any problems or concerns when using ChatGPT when it comes to accuracy, trustworthiness or bias (attribute *Problems* from Table 2), 80.5% stated that they had. When asked to rate their attitude toward the output of AI tools on a Likert Scale (attribute *OpOutput* from Table 1) from Critical (1) to Trustful (5), their response was 2.3 ± 1.0 , that is, essentially ranging from critical to neutral.

The free-form responses mainly highlighted concerns about the accuracy, reliability, trustworthiness, and consistency of outputs from AI tools like ChatGPT. With regard to accuracy, students frequently reported incorrect or inconsistent answers, particularly for complex or specialized tasks, including math and coding problems.

- “ChatGPT makes up code libraries or mixes up versions of them.”
- “When I use it to write code, it tends to use some non-existing functions.”
- “Errors in solving calculus and linear algebra problems. For example, ChatGPT could not even calculate the eigenvalues of a 4x4 matrix without me pointing out three times that it made errors.”
- “Because it is a AI, the result are always just approximations and stem from a black box.”

Probably the most frequently mentioned concern, however, was around wrong calculations: wrong numerical and wrong symbolic calculations.

In this context, students also mentioned that information might be out-of-date, and specifically referred to the fact that training for GPT at the time of the survey ended in 2021.

- “ChatGPT is only valid until 2021.”
- “Sometimes the answers are not updated, one should never trust ChatGPT blindly without any previous knowledge on the matter.”
- “ChatGPT has partially outdated information about specific subject areas, which is why I check most of it.”

The AI often “hallucinated” or generated wrong responses, sometimes even providing different answers for the same question.

- “ChatGPT hardly says ‘I don’t know,’ but invents some nonsense that usually sounds quite plausible.”
- “I only use it when I already have a basic understanding or intuition for the topic so that I can recognize when it is feeding me hallucinations.”
- “ChatGPT suffers from the Dunning-Kruger effect.”
- “ChatGPT itself cannot check if what it says is true. Its level of trustworthiness is equivalent to that of a puppy.”
- “ChatGPT is great for things that are difficult to find, but easy to verify.”

Some students remarked that ChatGPT can be so convincing about wrong answers that it makes them question and rethink concepts and sample solutions, and that in the end, they learn more.

- “Many times it stimulates critical thinking and fuels doubt and the desire to confirm or assure the answers he generated.”

Many students highlighted issues of incorrect or invented sources, and the lack of verifiable references. A lack of source traceability could lead to misinformation propagation and eroded confidence in the AI’s realm of capabilities.

- “When I asked ChatGPT to list some publications regards to a topic, it turns out the papers given by ChatGPT do not exist at all.”
- “ChatGPT gives no sources and has already made several assertions that were wrong. ChatGPT repeats content from the internet without checking its correctness or topicality.”

Similarly, AI’s performance was less trusted when handling nuanced tasks or less-common topics. Bias was another issue; while bias is usually seen as an unintentional, some students may have suspected that the system was manipulated during training. Furthermore, students raised concerns about data privacy and security, affirming the need for a more transparent, reliable, unbiased, and up-to-date information system.

- “ChatGPT has already swallowed half the internet and thus also the associated bias.”
- “The generative text is only based on the training set, if there is bias or errors, these are taken over into the generative text (garbage in, garbage out).”

Some students suspected a political agenda or at least strong bias:

- “It’s very obviously politically left-leaning. This bias did not affect me in any way but it becomes clear quite quickly. Maybe that’s for the best, don’t know.”
- “ChatGPT seems to heavily rely on politically left-wing assumptions and premises, thus consequently defending biased ideas, even without mentioning explicitly left-wing figures or parties.”

One student also criticized that ChatGPT would not answer questions about controversial topics (likely the result of fine-tuning), criticizing that this would patronize the user.

A common notion is that AI is a tool and thus also needs to be demystified:

- “In addition, the limitations of ChatGPT will be highlighted in order to ‘demystify’ it a little. If you look at the start-up scene, there seems to be an explosion of platforms and devices that are basically just integrating a ChatGPT API into existing products and selling this as an innovation. It’s just a big, very complex text generator that has been fed with a gigantic data set.”

Despite their criticism, many also acknowledged the utility of the AI tool as a platform for quick summaries and overviews, albeit emphasizing the importance of cross-verifying the information:

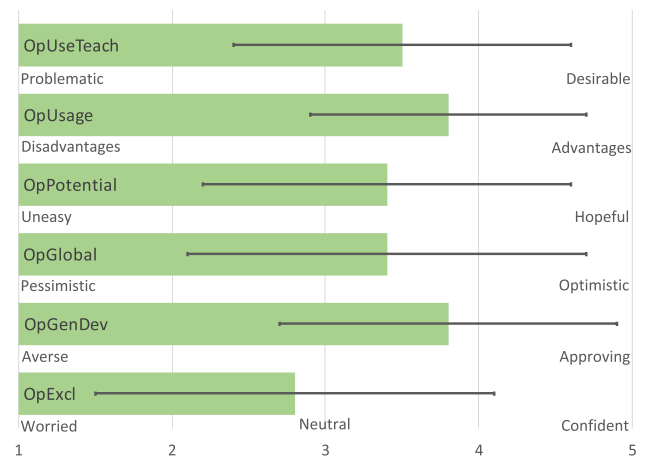
- “ChatGPT can be a good discussion partner when you are alone. You simply cannot believe everything.”
- “It offers ideas, but everything it says need verification from the user side, which is a nuisance.”

3.5 Attitudes toward AI

Several questions on the survey were designed to assess the students’ attitudes toward AI. As shown in Fig. 4, with the notable exception of concerns about exclusion and discrimination, students are moderately optimistic about the use of AI. Particularly strong is the support for continued development of AI (*OpGenDev*) and the belief that advantages of this technology outweigh disadvantages (*OpUsage*).

Together with the variable *OpOutput* evaluated in Sect. 3.4, these variables form the summative variable *OpAI*. Figure 5 shows how these opinions vary by discipline and gender, sorted by the overall average scores (note that gender ratios vary between programs; thus for example Biosystems and Engineering has a lower attitude score than Technology Management, even though the male score is higher). The study program on Technology Management shows the most positive attitude toward AI; the program covers topics of entrepreneurship and commercialization

Fig. 4 Attitudes towards the potential of AI



of technology. The wide spread of opinions, indicated by the bars, limits claims derived from the data, however, a tendency can be observed that women are more skeptical about AI than men. Engineering and biological sciences tend to be more accepting of AI than non-engineering and system-oriented sciences, with Mathematics and Physics in the middle.

As it turns out, by adding the perceived helpfulness *HelpChat* as a covariate and doing an ANCOVA, some of these differences between disciplines and departments decrease; in other words, disciplinary differences in attitude might depend on how helpful AI is in that discipline. Using *HelpChat* as a covariant also closes some of the gender gap.

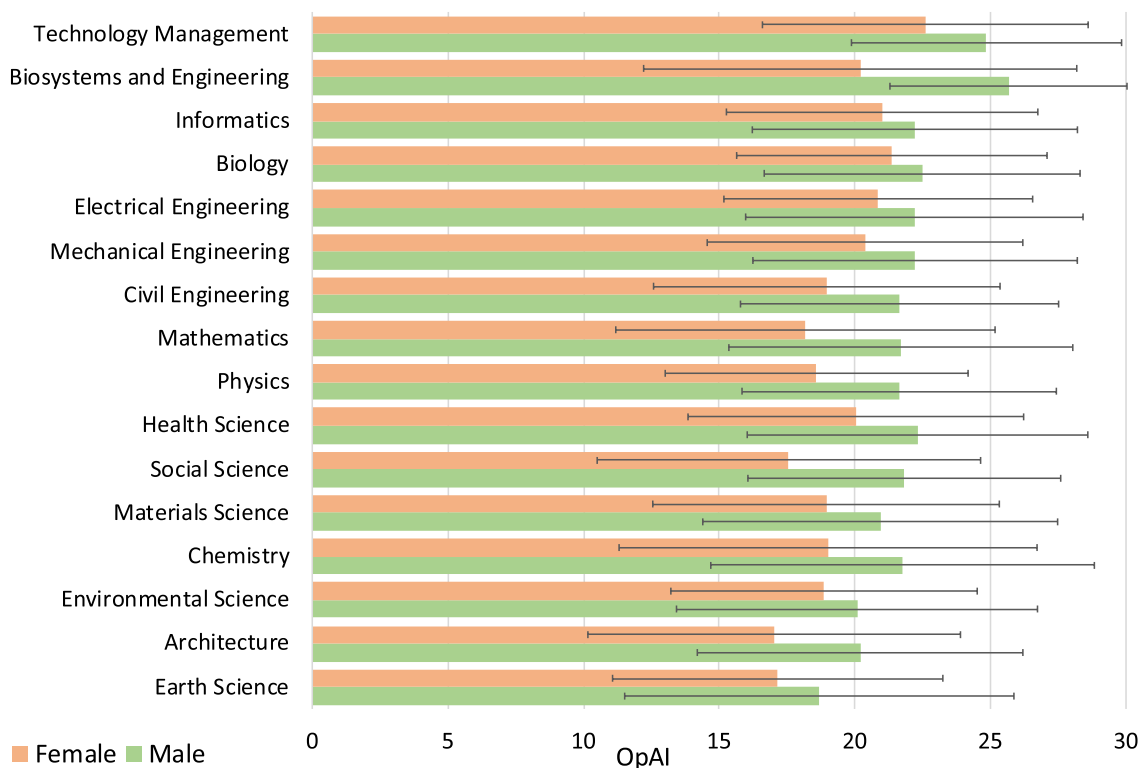


Fig. 5 Summative attitudes *OpAI* towards the potential of AI by discipline

3.6 The future of teaching and learning with AI

Overall, the students do not believe that current forms of teaching and assessment will be outdated any time soon. Figure 6 shows the answer distribution of the variable *TeachObs*, where over 3/4th (76.6%) of the students state that current techniques will not be obsolete.

- “I expect AI to take on a background role and lectures will still be in the foreground. Otherwise there will be one less reason to attend lectures.”

Many students express a desire for these tools to be incorporated into their learning experience, believing that AI could enhance their academic and professional lives. This goes along with their answer to the question if ETH should offer learning opportunities to promote the use and application of AI-based tools in their studies ($\approx 2/3$ rd, 65% answering “yes” for *ImpOff*); the question was certainly formulated exuberantly (“promote the use”), but students have a realistic view on the buzz and hype around AI: on a scale of exaggerated (1) to appropriate (5), they rated the buzz around AI 3.2 ± 1.1 (*OpExcite*).

Some students voiced the suggestion that AI should only be used in classes that teach about AI:

- “That they should stay limited to classes studying them.”

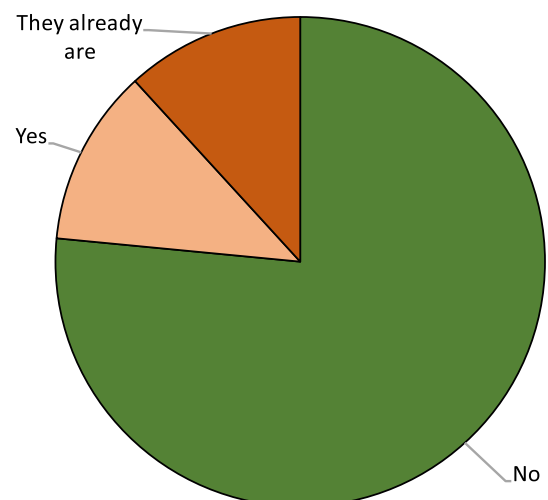
However, the following comments are representative of the majority opinion, characterizing the incorporation of AI as inevitable:

- “It is inevitable that this will be integrated into our lives, and the lives of students yet to come. It is unlikely that banning this technology will prevent its use and prevalence in students’ work, whether intentionally or not. Instead, as a leading educational institution, it is important that this is factored in for assignments and learning.”
- “Good integration, it should not be used as an obstacle but as an opportunity to improve your studies.”

This was, of course, not unanimous. Some students expressed how AI could elevate the level of what is being taught and tested:

- “Not ban, but meaningful integration. Especially in computer science, memorizing code for exams should be dropped and higher concepts such as planning or debugging should be introduced.”
- “[I wish] that we are taught things that can’t be solved with ChatGPT and one minute time; it seems like we are learning mental arithmetic while the first calculators appear, and back then the argument was, you won’t always carry a

Fig. 6 Will current forms of teaching and assessment soon become obsolete (*TeachObs*)?



calculator in your pocket? We all know how that went. So why not be proactive and teach something that ChatGPT can't do?"

- "That we have tasks at ETH where it is not possible for us to find the solutions with such tools. The standardization and commoditization of such tools could lead to us students not experiencing enough challenges because the solutions are found by a tool and not by our heads."

They recognize the increasing significance of AI and subsequently emphasize the importance of understanding its capabilities and limitations. To this end, students express a need for educational guidance on effective and responsible AI usage and its potential abuse:

- "I want to learn about the possible danger and how to use it wisely. I think these AI things scare me greatly with AI being able to reproduce voice, and so I don't want to use it for fear of further increasing its capacities."
- "Employ them, teach (or encourage) students to use them and at the same time discuss their downsides. These tools are here to stay, and I think they should be viewed as such-tools. They can help students (or people in general) achieve tasks more effectively, often skipping over the boring parts of work."

While the majority of students advocated for themselves to be able to use the tools, some argued that instructors should not use them:

- "I very much hope that the quality of teaching will not suffer in the future because lecturers try to generate lecture structures, slides or exercises using AI-supported tools."
- "I don't see how AI-based tools can replace the work of lecturers without the teaching suffering as a result."

At the extreme end of the spectrum, some students go as far as calling for a complete ban:

- "Prohibit, prohibit, prohibit, ban from ETH, set up a block in the ETH network that makes access impossible. Insert in the terms of use for ETH-IT as forbidden abuse."
- "Please don't. Simply no."

They point out several drawbacks like error detection deficiencies, the potential for work and learning opportunities being taken away, and concerns over possible biases introduced by these tools. They are also worried that their institution might be degraded if AI is overused:

- "Most professors and assistants I know are quite a bit smarter than AI tools in their field. I expect an institution such as ETH not to degrade itself following trends; there has to be a very clear, concrete, and solid reasoning behind the implementation of AI, and it needs to be justified in a comprehensive way. Not a generic technobabble."

While several students suggest getting a campus license for GPT-4, some students suggest creating an institutional AI:

- "The key is to find the right balance, possibly developing an ETH AI that is legal for learning and teaching by limiting it to not substitute the role of teacher and student. What is absolutely to be avoided (for the current time) is relying on AI detector tools to judge the originality of written text."
- "I wish ETH students would have their own AI created for the purpose of studying."
- "A more professionalized version of ChatGPT is needed for it to be reliable in academia. ChatGPT in itself is too unreliable to be source of information."
- "Why not give students a lecture-specific GPT where they can ask their questions?"
- "Short term: Experiments with evaluation on how to improve teaching, e.g., Retrieval Augmented Generation for course materials such as PDFs. Long term: Uniform chat interface to most subject-specific courses."
- "A text AI tool sponsored by ETH. In my opinion, this would alleviate the frustration when topics are unclear. Furthermore, the workload of the assistants would be reduced and any questions during exercises could be partially answered in this way."
- "Maybe a personalized TA would be super helpful."

- “ETH’s own subject-specific AI tools (‘additional TA’) would be exciting. ChatGPT is still very error-prone, depending on the subject area.”

However, despite the identified issues, the majority of students argues that the tools should be embraced rather than banned, but they should be used as a supplement, not a substitute for learning.

3.7 Relationship between survey responses

3.7.1 Correlations

Figure 7 shows a Fruchterman-Reingold representation [12, 14] of the statistically significant correlations between the items in Tables 1 and 2 (Spearman correlations, $p < 0.01$ [19]). The vertices denote the attributes; green edges denote positive correlations, while red edges denote negative ones; the thickness and saturation of these edges denote the correlation strength. Mutually closely correlated or anticorrelated vertices tend to cluster together, while unrelated vertices are farther apart. The rotation and handedness of the graphs are random.

Figure 8 shows a heat map representation [15] of the absolute values of the statistically significant correlations between the items in Tables 1 and 2 (Spearman correlations, $p < 0.01$ [19]). The strength of the correlation is indicated along a color spectrum, where blue indicates little to no correlation, while red indicated a strong correlation.

It is apparent that blocks of items are interconnected. For example, the items *OpGenDev*, *OpPotential*, *OpUsage* and *OpUseTeach* form such a cluster of positively correlated items ($r > 0.42$), indicating general skepticism or favoritism toward using AI. Also closely positively correlated are *ProvExpl* and *ProvDefi* ($r = 0.57$), as well as *HelpChat* and *HelpExamPrep* ($r = 0.57$).

The attribute *Gender* is negatively correlated with several of the attributes expressing confidence in and enthusiasm about AI; as already seen in Fig. 5, there is a gender gap across all study programs. Students identifying as female also express stronger support for developing rules.

Fig. 7 Fruchterman-Reingold [12, 14] representation of the significant ($p < 0.01$) correlations [19] between the binary and Likert-scale attributes in Tables 1 and 2

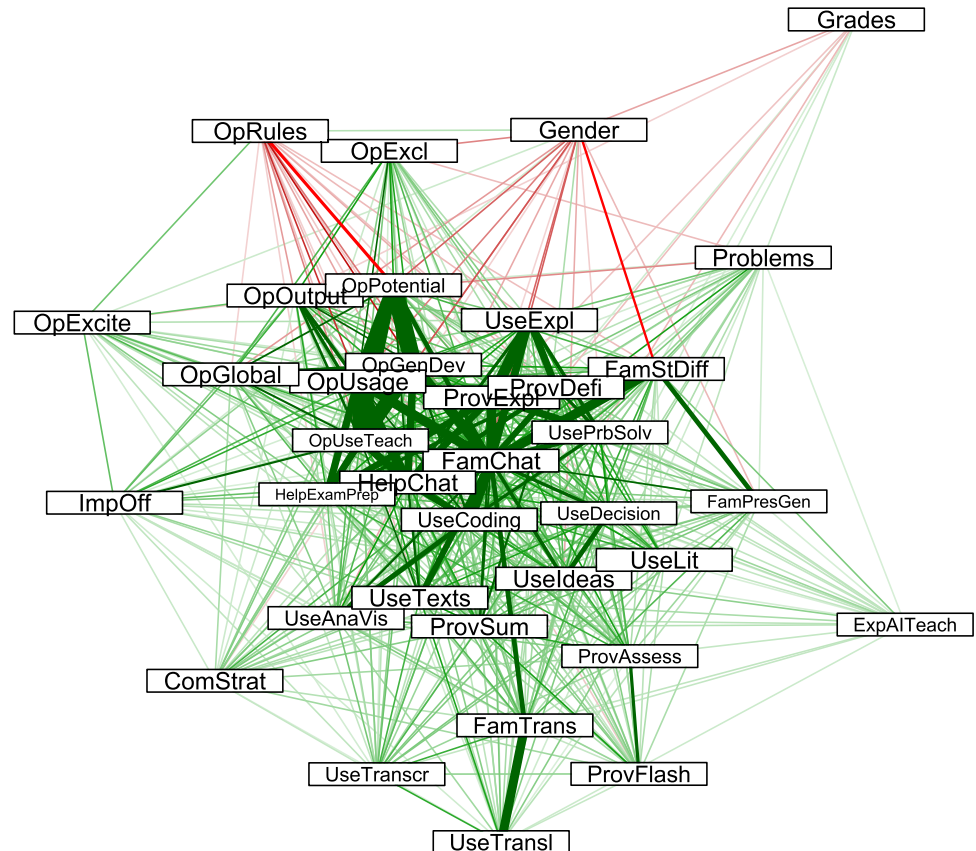
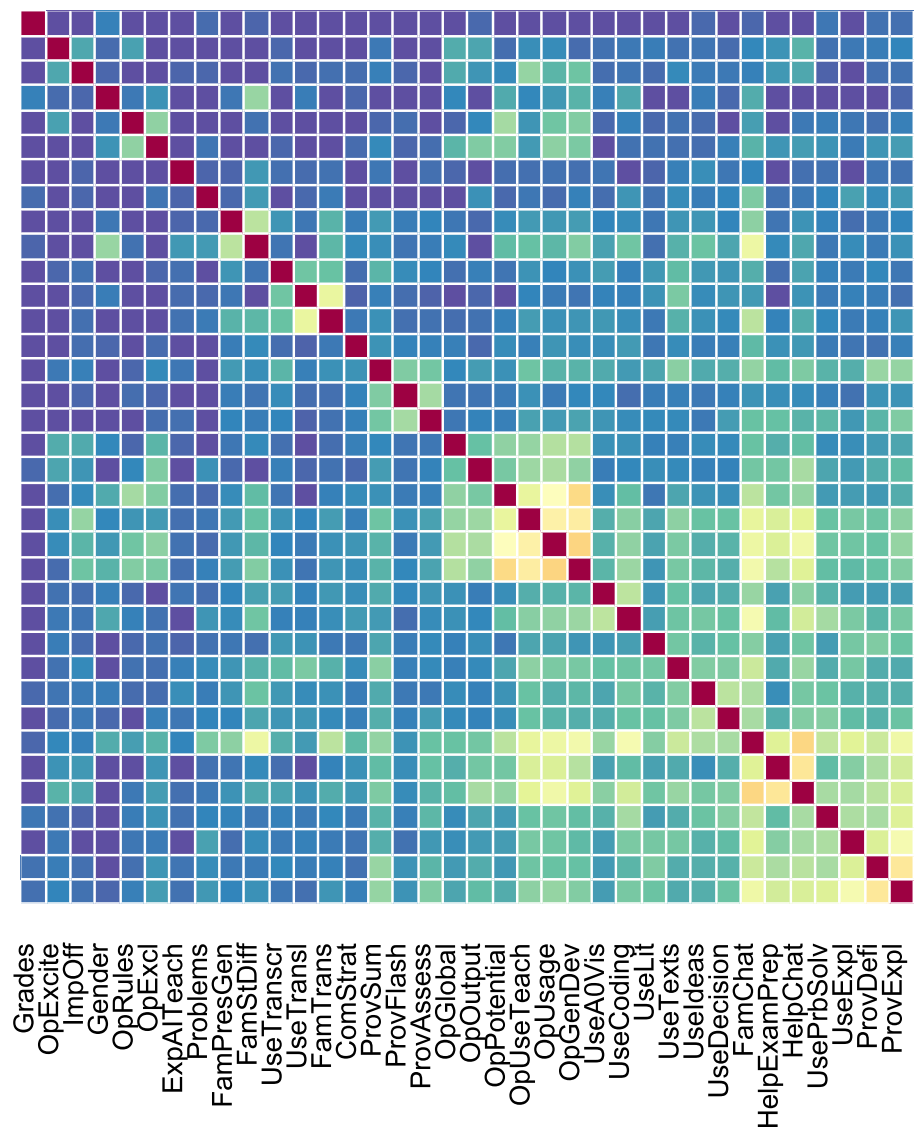


Fig. 8 Heat map [15] representation of the absolute values of the significant ($p < 0.01$) correlations [19] between the items in Tables 1 and 2



3.7.2 Differences by gender and study program

Students could enter four different gender categories at the end of the survey: “Male” (3088 observations), “Female” (1454 observations), “Diverse” (42 observations), and “No Answer” (214 observations). For the purposes of the following analysis, we will set aside “Diverse” and “No answer” responses, since it turns out these populations do not amount to a large enough sample size to yield statistically significant results. To start with a preparatory analysis, we want to test which variables are independent of gender. To this end, we apply a χ^2 -test (including the Bonferroni correction [46]) to the variable *Gender* against each one of the numeric variables (categorical and continuous), with the null hypothesis

$H_{0,n}$: The female and male distributions of question n are identical.

On one hand, we rejected $H_{0,n}$ (that is, signs of dependence) for 21 out of the 43 numeric variables in the dataset (see Table 5). For instance, we observe statistically significant dependence with the variable *AIperception*. Same with both *Grades* and *StudyProgram*.

On the other hand, we could not reject $H_{0,n}$ (i.e., no sign of dependence) for the remaining 16 variables. We notice the seeming lack of dependence with the variable *HelpExamPrep*, suggesting that male and female students answer similarly to the question “Did the way you prepared for the exam with your ‘AI Tutor’ make you learn better and feel more prepared?”.

Table 5 Result of a χ^2 -test to the variable *Gender* against binary variables (left table) and multi-valued variables (right table)

	Gender		Gender
<i>ComStrat</i>		<i>Alperception</i>	×
<i>ExpAIteach</i>		<i>FamChat</i>	×
<i>ImpOff</i>		<i>FamPresGen</i>	×
<i>Problems</i>		<i>FamStDiff</i>	×
<i>ProvAssess</i>		<i>FamTrans</i>	
<i>ProvDefi</i>		<i>Gender</i>	×
<i>ProvExpl</i>		<i>Grades</i>	×
<i>ProvFlash</i>		<i>HelpChat</i>	×
<i>ProvOther</i>		<i>HelpExamPrep</i>	
<i>ProvSum</i>		<i>NativeLanguage</i>	
<i>UseAnaVis</i>	×	<i>OpExcite</i>	×
<i>UseCoding</i>	×	<i>OpExcl</i>	×
<i>UseDecision</i>		<i>OpGenDev</i>	×
<i>UseExpl</i>		<i>OpGlobal</i>	×
<i>UseLit</i>		<i>OpOutput</i>	
<i>UseOther</i>		<i>OpPotential</i>	×
<i>UsePrbSolv</i>		<i>OpRules</i>	×
<i>UseTexts</i>		<i>OpUsage</i>	×
<i>UseTranscr</i>		<i>OpUseTeach</i>	×
<i>UseTransl</i>	×	<i>StudyClass</i>	×
<i>UseIdeas</i>	×	<i>StudyProgram</i>	×
		<i>TeachObs</i>	

Marked with an × are all tests that suggest dependence

Note that the variable *Grades* is self-perceived. This would explain why grades appear to be dependent on it, while previous studies have shown that at ETH, actual exam grades are independent of gender [25]; in this study, students identifying as “Female” tended to underestimate their grades.

Students could enter one of 78 different study programs, each part of one of the groups described in Table 3. Similarly to variable *Gender*, we will set aside study programs part of Groups 6 and 7 to keep a relatively consistent sample size across groups.

Again, we perform a recurring analysis, but this time for *StudyClass* and *Grades*. The results are shown in Tables 6 and 7.

We will take a look at two population variables at a time, namely *Gender* and *StudyClass*, and investigate the variable *Alperception*, see Fig. 3 for the overall answer distribution across all respondents. Table 8 shows how these percentages vary across *StudyClass*. For each of the five possible answers and each of the five department groups, we computed the difference between that group’s rate and the whole population’s rate. Note, in particular, that informatics students tend to be a lot more sure about their perception of the use of AI (in informatics, 5.2% fewer students answered “Don’t know” compared to the ETH average). They also tend to think AI should be less considered cheating than legitimate, compared to the overall population (in informatics, 3.9% more students answered “Legitimate” compared to the overall average).

Before considering *Gender*, a highly influential consideration is that the gender ratios vary greatly between the departments listed in Table 3, see Table 4.

As seen in the univariate analysis in Table 5, there is a statistically significant dependence between the variables *Gender* and *Alperception*. A subsequent question would be to ask about the dependence between *Gender* and *Alperception* if we start conditioning on *StudyClass*. In other words, we test the null hypothesis

$H_{0, StudyClass}$: The female and male distributions for group *StudyClass* are identical

for each of the five groups. We report in Table 9 the p -values of each of these respective tests, along with the underlying p -value that was found when investigating Table 7 for the whole dataset.

Note that (quite surprisingly, one might add), we only reject the null hypothesis that *Gender* and *Alperception* are independent for the whole population. In other words, much in the spirit of Simpson’s paradox [41] (see Appendix),

Table 6 Result of a χ^2 -test to the variables *StudyClass* and *Grades* against binary variables

	<i>StudyClass</i>	<i>Grades</i>
<i>ComStrat</i>	×	
<i>ExpAllTeach</i>	×	
<i>ImpOff</i>	×	×
<i>Problems</i>	×	
<i>ProvAssess</i>		
<i>ProvDefi</i>		
<i>ProvExpl</i>	×	
<i>ProvFlash</i>		
<i>ProvOther</i>		
<i>ProvSum</i>		
<i>UseAnaVis</i>		×
<i>UseCoding</i>	×	×
<i>UseDecision</i>	×	
<i>UseExpl</i>	×	
<i>UseIdeas</i>	×	×
<i>UseLit</i>		
<i>UseOther</i>		
<i>UsePrbSolv</i>	×	
<i>UseTexts</i>	×	
<i>UseTranscr</i>	×	
<i>UseTransl</i>	×	×

Marked with an × are all tests that suggest dependence

Table 7 Result of a χ^2 -test to the variables *StudyClass* and *Grades* against binary variables

	<i>StudyClass</i>	<i>Grades</i>
<i>Alperception</i>	×	
<i>FamChat</i>	×	×
<i>FamPresGen</i>		×
<i>FamStDiff</i>	×	×
<i>FamTrans</i>	×	
<i>Gender</i>	×	×
<i>Grades</i>		×
<i>HelpChat</i>	×	
<i>HelpExamPrep</i>	×	
<i>NativeLanguage</i>	×	×
<i>OpExcite</i>	×	×
<i>OpExcl</i>	×	×
<i>OpGenDev</i>	×	×
<i>OpGlobal</i>	×	×
<i>OpOutput</i>		
<i>OpPotential</i>	×	×
<i>OpRules</i>	×	×
<i>OpUsage</i>	×	×
<i>OpUseTeach</i>		×
<i>StudyClass</i>	×	×
<i>StudyProgram</i>	×	×
<i>TeachObs</i>		

Marked with an × are all tests that suggest dependence

Table 8 Answering rates of *AI perception* for each study groups, compared each time to the overall answer rates

	Informatics (%)	Engineering (%)	Arch/civil (%)	NatSci (%)	SysSci (%)
Cheating	-2.1	-0.5	-1.5	2.8	-0.4
Plagiarism	0.5	-0.1	-0.9	0.2	0.2
Legitimate	3.9	3.6	2.3	-5.9	-2.4
Depends	2.8	-1.2	-0.9	-0.2	0.4
Don't know	-5.2	-1.8	1.0	3.1	2.3

For instance, the first cell corresponds to the rate at which an informatics student will answer "Cheating" compared to the overall population (here, 2.1% less chance to answer "Cheating")

Table 9 χ^2 -test *p*-values associated to each of the five major department groups *StudyClass*

	Whole Pop.	Informatics	Engineering	Arch/civil	NatSci	SysSci
<i>p</i>	0.001	0.160	0.627	0.761	0.424	0.079

conditioning on the group *StudyClass* seems to make the dependence between those two variables vanish, regardless of the group of departments *StudyClass*.

4 Discussion

4.1 Academic integrity

In late 2022, educators were getting concerned by the capabilities of ChatGPT, and many of them immediately saw the potential for abuse and undermining academic integrity. As this survey shows, what came as a wake-up call to educators has long been commoditized by the students, at least those at a technical university. Section 3.1 shows that they are using specialized tools for tasks such as qualitative data analysis, language translation, math calculations, code generation, text summarization, image generation, and more. Particularly prevalent is usage for assisting in programming tasks, where GitHub Co-Pilot features most prominently. We thus conclude that these tools are not primarily used to undermine assessment, but to increase productivity. This wide range of usage has been found at other universities within the German higher-education tradition [44]. It should be remarked, though, that within this tradition, exercises and projects during the running semester have very little to no influence on grades, since those are largely determined by high-stake summative exams at the end or after the semester. Thus, many of these in-semester assessments are mostly learning opportunities.

High-stake exams will remain on campus under supervision, which is a policy the university enforced even during the COVID-19 pandemic. We are of the opinion that the integrity of these high-stake exams cannot be guaranteed in off-campus ("at home") settings, even more so now with the wide availability of AI-tools [24].

4.2 AI proficiency

While chat and translation tools are moderately familiar to students, image or presentation generating tools are not as well-known. This could be attributed to the more widespread use and accessibility of chat and translation tools in daily life, such as messaging apps and online translation services. On the other hand, image or presentation generating tools might be more specialized and not as commonly encountered by students at a technical university, where they seem to be mostly used in the context of producing social media content. The use of AI tools for social media content generation, like Midjourney and Elevenlabs, reflects the integration of AI in personal and social spheres.

The fact that students are training their own models and using advanced tools like PyTorch and CUDA suggests a high level of technical proficiency among the students who learn *about* AI; this needs to be distinguished from learning *with* AI [42] (which includes more than just the Large Language Models that have gained most of the recent attention [2]). Students do not foresee a complete overhaul of traditional teaching and assessment methods due to AI. However, there's a strong desire for AI integration in the learning process.

4.3 AI usage in teaching

The results show that a minority of students have experienced AI of any kind in a teaching situation. This could be due to the nascent stage of AI integration in educational settings. However, students have a plethora of ideas on how AI can be utilized in education, especially in programming and language-related tasks. The mention of teaching assistants providing guidance on using AI tools like GPT and Copilot suggests that there's some institutional support for AI integration, but not yet coherent guidance and recommendations. We need to better understand AI usage by instructors and thus conducted a separate survey among faculty members regarding their perceptions and usage of AI in teaching.

Based in part on findings of these studies, a committee of faculty members and administrators will convene to formulate recommendations and guidelines. However, it is to be expected that the authority to make decisions about AI usage, as well as the responsibility for providing meaningful assessments and preserve their integrity, will largely remain with the individual faculty member, and that it may differ from assignment to assignment based on learning objectives.

4.4 "Study buddy"

Students' perception of ChatGPT as a "study buddy" underscores the potential of AI tools to provide personalized learning support. The use of AI tools for exam preparation, like generating practice questions and flashcards, indicates the adaptability of students in leveraging technology for academic success. In spring semester 2024, ETH Zurich started piloting custom chatbots for courses, which answer questions based on reference materials provided by the instructors (Retrieval Augmented Generation [28]). While our current chatbots still use a commercial system as their conversational component, we hope to tune an open-weight Large Language Model over the course of the year to serve as backend.

4.5 Shifting landscape of perceptions

The results indicate a nuanced view of students regarding the use of AI in assessments. While many students see the benefits of using AI for tasks like proofreading and idea generation, they also emphasize the importance of originality and academic integrity. A strong compounding factor is that there is no reliable way to detect the use of AI [11, 22], and in the end, any attempt would be a probabilistic AI vs. AI arms race [39]. The mention of transparency and the need to cite AI tools reflects a mature understanding of ethical considerations. Students expect that instructors give clear policies for allowed and prohibit usage of these tools for their assignments.

In past, doomsday scenarios regarding AI have been the realm of science fiction [20], but today's concerns are much more specific. The concerns about over-reliance on AI, data privacy issues [37], and distinguishing AI-generated work highlight the challenges of integrating AI in academic settings. The need for clear guidelines from educational institutions is evident.

The results show that students are critical of AI outputs, especially in terms of accuracy and reliability, consistent with other findings [7]. Concerns about outdated information, invented sources, and biases indicate a discerning user base that is aware of the limitations of AI. The mention of potential political agendas or biases in AI training suggests a deeper understanding of the socio-political implications of AI.

The results indicate a moderate optimism among students regarding AI. The support for continued AI development and the belief in its advantages suggest a positive outlook. However, concerns about exclusion and discrimination indicate awareness of potential societal implications.

The gender differences in attitudes toward AI, especially in the context of study programs, highlight the need for more inclusive AI education and outreach.

5 Conclusions

The survey results provide insight into students' familiarity with, usage of, and attitudes towards artificial intelligence (AI) tools in an academic setting. We found a disparity in familiarity, which underscores the diverse range of experiences and comfort levels among students. While only a minority have experienced AI in formal teaching situations, they have a plethora of ideas on its potential applications, especially in programming and language-related tasks.

The sentiment towards AI is generally optimistic, with students recognizing its potential benefits in enhancing their academic and professional endeavors. However, concerns about trustworthiness, accuracy, and potential biases in AI outputs are prevalent. The majority believe that while AI tools can be beneficial supplements and should be integrated into teaching scenarios, they should not replace traditional learning methods.

Furthermore, the responses suggest a nuanced perspective on the use of AI in assessments. While many students deem AI tools legitimate for tasks like proofreading and idea generation, there's a consensus that primary content should originate from the student, with verbatim copying from AI outputs considered unethical. The need for transparency and clear guidelines from educational institutions is emphasized, with students advocating for a context-based use of AI. Despite the potential advantages, there are concerns about over-reliance on AI, data privacy issues, and the challenge of distinguishing AI-generated work. Several students asked for institution-provided AI tools to overcome these concerns.

Acknowledgements We would like to thank the students who participated in this study, and who provided such rich insights into their usage and evaluation of AI.

Author contributions All authors contributed according to ETH's authorship regulations.

Funding No funding was received for this study.

Data availability In observance of the approved human subject research protocol ETH EK-2023-N-189, raw data from this study is confidential and cannot be publicly disclosed. Researchers interested in working with our data should contact Gerd Kortemeyer (kgerd@ethz.ch) to submit an amendment to the protocol for follow-up or replication studies.

Declarations

Ethics approval and consent to participate The protocol was approved by the ETH Zurich Ethics Commission in accordance with the EU guidelines for Ethics in Social Science and Humanities (Sect. 4), the "swissethics" guidelines, and the particular guidelines for Educational Research at ETH Zurich (approved by the commission in November 2021) as study ETH EK-2023-N-189. Research was carried out in accordance to the approved protocol. Participation in the survey was voluntary, and Informed Consent was obtained from all participants in this study.

Consent for publication All authors agreed to the participation in this study.

Competing interests The authors have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Simpson's paradox

In the following, let A be the answer to the question:

How should the use of AI-based tools such as ChatGPT currently be considered in written performance assessments?

Recall that the possible answers are "Cheating," "Plagiarism," "Legitimate," "Depends," and "Don't know." For simplicity's sake, a possible answer is any integer $a \in \{1, 2, 3, 4, 5\}$. Also, let G be the gender variable restricted to being female or male and hence is binary. Finally, let D be the department variable which takes values as described in the first column of Table 3.

In case the answer A is independent of G conditionally on the department D , then we should have the equality

$$\begin{aligned} \mathbb{P}(A = a, G = g \mid D = d) \\ = \mathbb{P}(A = a \mid D = d)\mathbb{P}(G = g \mid D = d) \end{aligned}$$

for all $a \in \{1, \dots, 5\}$, $g \in \{0, 1\}$ and $d \in \{1, \dots, 5\}$, which is what a χ^2 -test tests for. Now, considering the whole population of respondents, we use the Law of Total Probability – and the aforementioned independence assumption – to write that

$$\begin{aligned}
 \mathbb{P}(A = a, G = g) &= \sum_d \mathbb{P}(A = a, G = g \mid D = d)\mathbb{P}(D = d) \\
 &= \sum_d \mathbb{P}(A = a \mid D = d)\mathbb{P}(G = g \mid D = d)\mathbb{P}(D = d) \\
 &= \sum_d \mathbb{P}(A = a \mid D = d) \left(\mathbb{P}(G = g \mid D = d) - \mathbb{P}(G = g) \right) \mathbb{P}(D = d) \\
 &\quad + \sum_d \mathbb{P}(A = a \mid D = d)\mathbb{P}(G = g)\mathbb{P}(D = d).
 \end{aligned}$$

Using again the Law of Total Probability we see that the second term in the last sum is equal to

$$\mathbb{P}(A = a)\mathbb{P}(G = g).$$

It follows that at the whole-population level, we have that

$$\begin{aligned}
 \mathbb{P}(A = a, G = g) - \mathbb{P}(A = a)\mathbb{P}(G = g) &= \sum_d \mathbb{P}(A = a \mid D = d)\mathbb{P}(D = d) \\
 &\quad \cdot \left(\mathbb{P}(G = g \mid D = d) - \mathbb{P}(G = g) \right) \\
 &= \sum_d \mathbb{P}(A = a, D = d) \left(\mathbb{P}(G = g \mid D = d) - \mathbb{P}(G = g) \right).
 \end{aligned}$$

Thus, it is clear that if $\mathbb{P}(G = g \mid D = d) - \mathbb{P}(G = g) = 0$ for all d and g , we get independence of A and G at the whole-population level. This also means that the conclusion will be coherent with what is observed at the level of the departments in case the gender distribution is independent of the department. In case this independence does not hold true, a different conclusion is to be expected at the level of the whole population. From the collected data, it seems that the distribution of male and female students depends heavily on which department they study in. To give an example, there are 82.7 % of males in Engineering while only 36.4% in System Oriented Natural Sciences (see Table 4). Thus, while the p -values obtained with the χ^2 -test for independence does not allow to reject the independence assumption for each of the departments taken individually, the same assumption is rejected with very significant p -value at the level the whole population, as seen in Table 9.

References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous systems; 2015. <https://www.tensorflow.org/> (Software available from tensorflow.org).
2. Adiguzel T, Kaya MH, Cansu FK. Revolutionizing education with AI: exploring the transformative potential of chatgpt. *Contemp Educ Technol.* 2023;15(3):ep429.
3. ATLAS.ti Scientific Software Development GmbH (n.d.). ATLAS.ti. <https://atlasti.com>.
4. Baidoo-Anu D, Ansah LO. Education in the era of generative artificial intelligence (ai): understanding the potential benefits of chatgpt in promoting teaching and learning. *J AI.* 2023;7(1):52–62.
5. Bodnick M. GPT-4 can already pass freshman year at Harvard. *The Chronicle of Higher Education*, July 26 2023. <https://www.chronicle.com/article/gpt-4-can-already-pass-freshman-year-at-harvard>
6. Chan CKY. A comprehensive AI policy education framework for university teaching and learning. *Int J Educ Technol High Educ.* 2023;20(1):1–25.
7. Chan CKY, Hu W. Students' voices on generative ai: Perceptions, benefits, and challenges in higher education. arXiv preprint; 2023. [arXiv: 2305.00290](https://arxiv.org/abs/2305.00290)
8. DeepL Inc (n.d.) DeepL Translator. <https://www.deepl.com/>. Accessed 10 May 2024
9. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint; 2018. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
10. ElevenLabs Inc (n.d.) ElevenLabs AI Voice Generator & Text to Speech. <https://elevenlabs.io>. Accessed 10 May 2024
11. Elkhatat AM, Elsaid K, Almeer S. Evaluating the efficacy of ai content detection tools in differentiating between human and AI-generated text. *Int J Educ Integr.* 2023;19(1):17.
12. Epskamp S. R package: qgraph [Computer software manual]; 2022. <https://cran.r-project.org/web/packages/qgraph/qgraph.pdf>.
13. Farhi F, Jeljeli R, Aburezeq I, Dweikat FF, Al-shami SA, Slamene R. Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Comput Educ: Artif Intell.* 2023;5:100180. <https://doi.org/10.1016/j.caeai.2023.100180>.
14. Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp.* 1991;21(11):1129–64.
15. Galili T. R package: gplots [Computer software manual]; 2022. <https://cran.r-project.org/web/packages/gplots/gplots.pdf>.

16. GitHub, Inc. (n.d.). GitHub Copilot. <https://github.com/features/copilot>. Accessed 10 May 2024.
17. Google (n.d.). Bard (replaced by Gemini). <https://bard.google.com/>. Accessed 10 May 2024.
18. Grammarly Inc. (n.d.). Grammarly. <https://www.grammarly.com>. Accessed 10 May 2024.
19. Harrell F. R package: Hmisc [Computer software manual]; 2023. <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf>.
20. Hermann I. Artificial intelligence in fiction: between narratives and metaphors. *AI & Soc.* 2023;38(1):319–29.
21. Ibrahim H, Liu F, Asim R, Battu B, Benabderrahmane S, Alhafni B, et al. Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Sci Rep.* 2023;13(1):12187.
22. Khalil M, Er E. Will ChatGPT get you caught? Rethinking of plagiarism detection. arXiv preprint; 2023. [arXiv:2302.04335](https://arxiv.org/abs/2302.04335).
23. Kortemeyer G. Could an artificial-intelligence agent pass an introductory physics course? *Phys Rev Phys Educ Res.* 2023;19(1): 010132.
24. Kortemeyer G, Bauer W. Cheat sites and artificial intelligence usage in online introductory physics courses: what is the extent and what effect does it have on assessments? arXiv preprint; 2023. [arXiv:2312.15334](https://arxiv.org/abs/2312.15334).
25. Kortemeyer G, Dittmann-Domenichini N, Schlienger C, Spilling E, Yaroshchuk A, Dissertori G. Attending lectures in person, hybrid or online—how do students choose, and what about the outcome? *Int J Educ Technol High Educ.* 2023;20:19.
26. Lane M, Williams M, Broecke S. The impact of ai on the workplace: Main findings from the oecd ai surveys of employers and workers. *OECD Social, Employment and Migration Working Papers*; 2023. p. 289.
27. Learneo, Inc. (n.d.). QuillBot. <https://quillbot.com>. Accessed 10 May 2024.
28. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrievalaugmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst.* 2020;33:9459–74.
29. Midjourney (n.d.). Midjourney. <https://www.midjourney.com/>. Accessed 10 May 2024.
30. New York Times (n.d.). ChatGPT Spotlight. <https://www.nytimes.com/spotlight/chat-gpt>. Accessed 10 May 2024.
31. Nguyen A, Ngo HN, Hong Y, Dang B, Nguyen B-PT. Ethical principles for artificial intelligence in education. *Educ Inf Technol.* 2023;28(4):4221–41.
32. NVIDIA (n.d.). GPT-4. <https://developer.nvidia.com/cuda-toolkit>. <https://www.nytimes.com/spotlight/chat-gpt>. Accessed 10 May 2024.
33. OpenAI (n.d.-a). ChatGPT. <https://chat.openai.com>. Accessed 10 May 2024.
34. OpenAI (n.d.-b). GPT-4. <https://openai.com/gpt-4>. Accessed 10 May 2024.
35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. In: *Proceedings 33rd conference on advances in neural information processing systems 32 (NeurIPS 2019)*. Curran Associates, Inc.; 2019. ISBN: 9781713807933.
36. Pereira V, Hadjielias E, Christofi M, Vrontis D. A systematic literature review on the impact of artificial intelligence on workplace outcomes: a multi-process perspective. *Human Resour Manag Rev.* 2023;33(1): 100857.
37. Rayhan S. Ethical implications of creating AGI: impact on human society, privacy, and power dynamics. https://www.researchgate.net/profile/Shahana-Rayhan/publication/372752216_Ethical_Implications_of_Creating_AGI_Impact_on_Human_Society_Privacy_and_Power_Dynamics/links/64c63dcb880b81123c5f5bf9/Ethical-Implications-of-Creating-AGI-Impact-on-Human-Society-Privacy-and-Power-Dynamics.pdf. Accessed 10 May 2024.
38. RAYYAN (n.d.). rayyan. <https://www.rayyan.ai>. Accessed 10 May 2024.
39. Santra PP, Majhi D. Scholarly communication and machine-generated text: is it finally ai vs ai in plagiarism detection? *J Inf Knowl.* 2023;60:175–83.
40. Smolansky A, Cram A, Radulescu C, Zeivots S, Huber E, Kizilcec RF. Educator and student perspectives on the impact of generative ai on assessments in higher education. In: *Proceedings of the tenth ACM conference on learning@ scale*; 2023. p. 378–82.
41. Sprenger J, Weinberger N. Simpson's Paradox. In: Zalta EN, editor. *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University; 2021. <https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/>.
42. Tang K-Y, Chang C-Y, Hwang G-J. Trends in artificial intelligence-supported e-learning: a systematic review and co-citation network analysis (1998–2019). *Interact Learn Environ.* 2023;31(4):2134–52.
43. Vargas-Murillo AR, de la Asuncion INM, de Jesús Guevara-Soto F, et al. Challenges and opportunities of ai-assisted learning: A systematic literature review on the impact of chatgpt usage in higher education. *Int J Learn Teach Educ Res.* 2023;22(7):122–35.
44. von Garrel J, Mayer J. Artificial intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. *Hum Soc Sci Commun.* 2023;10(1):1–9.
45. Walczak Z. Graphics in latex using tikz. *TUGboat*, 20. p. 176–9; 2008; <https://tug.org/TUGboat/tb29-1/tb91walczak.pdf>.
46. Weisstein EW. Bonferroni correction; 2004. <https://mathworld.wolfram.com/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.