



Beyond the Hype: A Comprehensive Review of Current Trends in Generative AI Research, Teaching Practices, and Tools

James Prather*
Abilene Christian University
Abilene, TX, USA
james.prather@acu.edu

Juho Leinonen*
Aalto University
Espoo, Finland
juho.2.leinonen@aalto.fi

Natalie Kiesler*
Nuremberg Tech
Nuremberg, Germany
natalie.kiesler@th-nuernberg.de

Jamie Gorson Benario
Google
Chicago, IL, USA
jamben@google.com

Sam Lau
University of California San Diego
La Jolla, CA, USA
lau@ucsd.edu

Stephen MacNeil
Temple University
Philadelphia, PA, USA
stephen.macneil@temple.edu

Narges Norouzi
University of California Berkeley
Berkeley, CA, USA
norouzi@berkeley.edu

Simone Opel
FernUniversität in Hagen
Hagen, Germany
simone.opel@fernuni-hagen.de

Vee Pettit
Virginia Tech
Blacksburg, VA, USA
vpettit@vt.edu

Leo Porter
University of California San Diego
La Jolla, CA, USA
leporter@ucsd.edu

Brent N. Reeves
Abilene Christian University
Abilene, TX, USA
brent.reeves@acu.edu

Jaromir Savelka
Carnegie Mellon University
Pittsburgh, PA, USA
jsavelka@cs.cmu.edu

David H. Smith IV
University of Illinois
Urbana, IL, USA
dhsmith2@illinois.edu

Sven Strickroth
LMU Munich
Munich, Germany
sven.strickroth@ifi.lmu.de

Daniel Zingaro
University of Toronto Mississauga
Mississauga, ON, Canada
daniel.zingaro@utoronto.ca

Abstract

Generative AI (GenAI) is advancing rapidly, and the literature in computing education is expanding almost as quickly. Initial responses to GenAI tools were mixed between panic and utopian optimism. Many were fast to point out the opportunities and challenges of GenAI. Researchers reported that these new tools are capable of solving most introductory programming tasks and are causing disruptions throughout the curriculum. These tools can write and explain code, enhance error messages, create resources for instructors, and even provide feedback and help for students like a traditional teaching assistant. In 2024, new research started to emerge on the effects of GenAI usage in the computing classroom. These new data involve the use of GenAI to support classroom instruction at scale and to teach students how to code with GenAI. In support of the former, a new class of tools is emerging that can provide personalized feedback to students on their programming assignments or teach both programming and prompting skills at the same time. With the literature expanding so rapidly, this report

aims to summarize and explain what is happening on the ground in computing classrooms. We provide a systematic literature review; a survey of educators and industry professionals; and interviews with educators using GenAI in their courses, educators studying GenAI, and researchers who create GenAI tools to support computing education. The triangulation of these methods and data sources expands the understanding of GenAI usage and perceptions at this critical moment for our community.

CCS Concepts

• **Social and professional topics** → **Computing education**; • **Computing methodologies** → **Artificial intelligence**.

Keywords

generative AI; GenAI; large language models; artificial intelligence; pedagogical practices; teaching computing; computing education

*Co-leader



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

ITiCSE-WGR 2024, July 8–10, 2024, Milan, Italy
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1208-1/24/07
<https://doi.org/10.1145/3689187.3709614>

ACM Reference Format:

James Prather, Juho Leinonen, Natalie Kiesler, Jamie Gorson Benario, Sam Lau, Stephen MacNeil, Narges Norouzi, Simone Opel, Vee Pettit, Leo Porter, Brent N. Reeves, Jaromir Savelka, David H. Smith IV, Sven Strickroth, and Daniel Zingaro. 2024. Beyond the Hype: A Comprehensive Review of Current Trends in Generative AI Research, Teaching Practices, and Tools. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR 2024)*, July 8–10, 2024, Milan, Italy. ACM, New York, NY, USA, 39 pages. <https://doi.org/10.1145/3689187.3709614>

1 Introduction

Computing education is undergoing a seismic shift due to the advances in generative AI (GenAI) [22, 36, 111]. Beginning in early 2022, computing education researchers showed that these models had incredible accuracy solving programming problems and exam questions in multiple courses and contexts [6, 40, 41, 48, 68, 119, 125, 126]. Other early work focused on the ways in which GenAI can provide support to computing educators [16, 37, 81, 82, 94, 124, 127]. Others were quick to raise concerns about potential threats to education, such as over-reliance, bias in the models, and educational misconduct [17, 80, 111, 114]. Computing instructors at the K-12 level are also struggling with integrating GenAI into their curricula [14, 43, 147]. K-12 teachers outside of computing education are having similar conversations to those occurring in higher education [14, 18, 99, 109, 116, 121]. Public perception of GenAI is mixed, partially because it is a “black box” and that lack of transparency often increases fear [21], which is one reason why GenAI should be designed to increase transparency to end users [142].

However, the discussion has largely moved from threats, challenges, and opportunities [17, 62, 104, 135, 145] to questions of practical adoption [93, 112]. The 2023 ITiCSE working group on GenAI summarized the activity within the computing education community and suggested that the next step is to determine reliable and safe ways to implement it into computing curricula [111]. An essay released along with the 2023 ACM/IEEE Computing Curricula suggested ways in which this could be accomplished based on preliminary data from a few studies and painted an optimistic picture of the future [16]. Indeed, many within the community are calling AI integration just another step in the advancement of educational technology [1]. Others are discussing how GenAI will change programming competencies and skills in the future [65, 132]. New assessment methods [61] and ways to measure user interaction with GenAI [98] are required, and although some are arguing for making assignments “LLM-proof” [20], which seems to be an impractical goal given the rapid improvement in these models.

Although some have advocated for banning GenAI entirely, that also seems impractical given its free availability to students outside the classroom [80]. Furthermore, professional developers are also discussing the role that GenAI will play within their work [8, 44, 75, 84, 141, 155], lending credibility to the idea that students must be prepared for using it after university. Therefore, thoughtful integration and scaffolding appears to be the way forward [36]. However, there is a lack of helpful and clear terminology to discuss the kinds of classroom interventions conducted so far. To this end, our report attempts to distinguish the following use cases: (1) instructors teaching students about using GenAI tools in order to write code, and (2) instructors using GenAI tools to support the teaching of their course via help-seeking bots, code feedback, assignment creation, etc.

The first category is by far the most extensive. Researchers were quick to show that GenAI could automatically enhance programming error messages [82], which was followed with two large-scale replications showing a direct benefit to students [143, 153]. Indeed, GenAI can provide other kinds of advanced and customized help and qualitative feedback to students working on computing assignments [10, 11, 67, 81, 89, 94]. Other more recent work,

such as the Harvard CS50 course, has focused on providing programming tutoring and help to students at scale through TA chatbots [87]. Researchers are only beginning to define how these should be designed, implemented, and evaluated [35]. The same applies to understanding how students actually use GenAI in authentic course settings, for example, in introductory programming courses [69, 128, 129], and advanced computing courses [42].

Instructors can also use GenAI to create customized and unique assignments tailored to student interests [88, 124] as well as generate other educational content [38, 138, 146]. Yet, this second category of GenAI usage in instruction is the least studied to date, possibly because many instructors have not yet thoughtfully integrated GenAI directly into their programming instruction [39, 45, 151]. However, studies thus far are showing mixed results. Some are pointing to its introduction as a way to equip students to move faster through the curriculum than was ever possible before [148]. Others are claiming that using GenAI has no negative effects on student learning outcomes [159]. However, some early work when GitHub Copilot was first released showed that students would flail and wander during programming tasks due to that tool’s constant interruptions [114]. Other more recent work has now shown that GenAI can significantly undercut student critical thinking during code writing and debugging tasks for those students who over-rely on it, decreasing their grades overall [56]. Similarly, in an observational study, researchers found that although a larger percentage of students are able to complete programming tasks with the aid of GenAI tools, students faced new metacognitive challenges [113].

As evidenced by the discussion above and our systematic literature review below, the literature on GenAI in computing education is expanding rapidly. With so much happening so quickly, it is difficult to know what has been done, why it is being done, what works, and where this is all headed. We therefore attempt to capture the zeitgeist of the present moment in computing history by defining terms, ordering and summarizing all of this for the reader.

1.1 Goals

With all of the above in mind, this report addresses the following overarching research goals:

- (1a) How are instructors incorporating GenAI into teaching computing?
- (1b) And why are they making these choices?
- (2a) How have the expectations towards skills in software development changed with the advent of GenAI?
- (2b) Which computing competencies are required in the future, according to teachers and industry professionals?

We address these overarching goals via several fine-grained research questions, and various methods: a systematic literature review, a survey study with educators and software developers in industry, and qualitative interviews with computing educators, computing researchers, and GenAI tool creators.

1.2 Contributions

This working group report describes how and why computing instructors have chosen to integrate (or not integrate) GenAI into their courses, and what they expect with regard to future developments

in curricula and industry usage of GenAI tools. We identify the current state-of-the-art by presenting the following deliverables:

- (1) **Systematic Literature Review (Section 2 and 3):** We review the existing literature on GenAI tools in computing education (through May 23, 2024) and present the studies in which educators report on (a) evidence of GenAI in computing education research, (b) educators using GenAI tools in their teaching practices, and (c) the rationale of educators to incorporate GenAI tools in a certain way.
- (2) **Evaluating Instructional Practices to Teach Students How to Use GenAI Tools (Section 4 and 5):** We gather current integration practices through an international survey of computing instructors. The survey also focused on the tools, policies, motivational aspects, and the impact of GenAI on the competencies students require to succeed – from an educator’s perspective.
- (3) **Evaluating Instructors’ Use of GenAI Tools (Section 4 and 5):** The survey of educators also revealed the use of GenAI tools by these educators, for example, to create tools that would support students.
- (4) **Evaluating Instructors’ Perspectives on Learning Outcomes and Future Developments (Section 4 and 5):** To capture the impact of GenAI tools on actual student outcomes, we conducted semi-structured interviews with instructors. The results reveal the disruptive character of GenAI tools in terms of the potential benefits and limitations of GenAI in computing education.
- (5) **Exploring the Industry’s Experience (Section 4 and 5):** Another contribution of this report is the integration of the industry perspective, their experiences and usage patterns of GenAI tools. This encompasses policies, motivations, and expectations regarding the competencies future developers will need.

1.3 Structure of the Report

The structure of this working group report is as follows. To address related work, the authors present the systematic search and review of prior studies on how and why computing educators have integrated GenAI tools into their teaching practices. The methods of the literature review are presented in Section 2 and the respective results of the systematic literature review are presented in Section 3.

The second major component of this working group report captures the perspective of both educators and software developers. We do so by presenting our methodology (Section 4), consisting of an international survey of computing educators, a series of qualitative interviews with computing educators, and an aligned survey of software developers to gather current industry practices and perspectives. As a part of the methodology, we briefly introduce prior work, fine-grained research questions, the process of developing the survey and interview questions, and the data analysis approaches we applied.

We present the results of our mixed-methods approach (i. e., surveys with educators and developers, and interviews with educators) in Section 5 by triangulating the different data sources for every research question. In Section 6, we discuss the most interesting results, and how they build and extend prior work.

In Section 7, we summarize the threats to validity of the applied methodology, before concluding our work in Section 8, and presenting pathways for future work (Section 9).

2 Systematic Literature Review: Methods

The systematic literature review (SLR) aims to identify how instructors are integrating generative AI into computing classrooms. The goal is to extend the many prior interview and survey studies that have looked at students’ perceptions of how generative AI tools are used [7, 49, 80, 161] to focus instead on what is actually being done within classroom settings. Therefore the goal is to focus on pedagogies, tools, and classroom interventions that feature empirical data about students’ experiences with those classroom interventions. Specifically, we investigate the following three research questions:

SLR-RQ1. How can the reported evidence of Generative AI in CER be summarized?

SLR-RQ2. How is generative AI being incorporated into teaching?

SLR-RQ3. What are the motivations behind incorporating GenAI tools into teaching?

SLR-RQ1 and SLR-RQ2 relate to the overarching research goal (1a), while SLR-RQ3 relates to the overarching research goal (1b).

Figure 1 illustrates the overall process of the whole systematic literature review. We follow the literature review best practices by Kitchenham and Brereton [72]. We first searched through various databases using a search string. Then, we checked for the inclusion of reference papers that should be found to confirm the quality of the search string. This was followed up by doing a title/abstract scan for relevance. After the title/abstract scan, we read the full papers and started extracting relevant information from them. In this stage, some papers were still rejected if they did not pass the inclusion/exclusion criteria. In the end, we had a set of 71 papers that passed the criteria and for which we extracted data. The included papers are listed in Table 1.

2.1 Search String Construction

The research team iteratively constructed a search string aimed at including papers that matched our research interests according to four categories: 1) the *domain* of the work was computer science or computer engineering education, 2) the *topic* was related to the use of Generative AI (GenAI), 3) it aligned with the *working group focus* on the impact of GenAI on pedagogy or teaching tools, and 4) it included the use of empirical *methods*. The literature review team met several times to discuss various keyword permutations and ultimately arrived at the following sub-search string for each of the four categories:

Domain: “Computer science education” OR “Computing education” OR “CS education” OR “CSEd” OR “CER” OR “Computing students” OR “Computing instructors” OR “CS students” OR “CS instructors” OR “Computer engineering education” OR “Programming education” OR “Introductory programming”

Topic: “Large language model” OR “Large language models” OR “LLM” OR “LLMs” OR “Generative AI” OR “ChatGPT” OR “GPT3” OR “GPT4” OR “Multimodal model” OR “Multimodal

Table 1: Papers included in the literature review (listed alphabetically by author, grouped by publication year). Venue shows the status at the time when the article was included.

AUTHOR	TITLE	VENUE	YEAR	CITE
Jonsson and Tholander	Cracking the code: Co-coding with AI in creative programming education	C&C	2022	[55]
Crandall et al.	Generative Pre-Trained Transformer (GPT) Models as a Code Review Feedback Tool in Computer Science Programs	JCSC	2023	[29]
Denny et al.	Can We Trust AI-Generated Educational Content? Comparative Analysis of Human and AI-Generated Learning Resources	arXiv	2023	[32]
Denny et al.	Promptly: Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generator	arXiv	2023	[33]
Dos Santos and Cury	Challenging the Confirmation Bias: Using ChatGPT as a Virtual Peer for Peer Instruction in Computer Programming Education	FIE	2023	[123]
Hajj and Sah	Assessing the Impact of ChatGPT in a PHP Programming Course	ISAS	2023	[3]
Hu et al.	Explicitly Introducing ChatGPT into First-year Programming Practice: Challenges and Impact	TALE	2023	[51]
Karnalim et al.	Plagiarism and AI Assistance Misuse in Web Programming: Unfair Benefits and Characteristics	TALE	2023	[58]
Kazemitabaar et al.	Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming	CHI	2023	[59]
Kumar et al.	Bridging the Gap in AI-Driven Workflows: The Case for Domain-Specific Generative Bots	BigData	2023	[76]
Kuramitsu et al.	KOGI: A Seamless Integration of ChatGPT into Jupyter Environments for Programming Education	SPLASH-E	2023	[78]
Liffiton et al.	CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes	Koli Calling	2023	[86]
MacNeil et al.	Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book	SIGCSE TS	2023	[94]
Markel et al.	GPTEach: Interactive TA Training with GPT-based Students	L@S	2023	[96]
Perry et al.	Do Users Write More Insecure Code with AI Assistants?	CCS	2023	[107]
Prasad et al.	Generating Programs Trivially: Student Use of Large Language Models	CompEd	2023	[110]
Prather et al.	"It's Weird That It Knows What I Want": Usability and Interactions with Copilot for Novice Programmers	TOCHI	2023	[114]
Qureshi	ChatGPT in Computer Science Curriculum Assessment: An analysis of Its Successes and Shortcomings	ICSLT	2023	[115]
Sandoval et al.	Lost at C: A User Study on the Security Implications of Large Language Model Code Assistants	USENIX Security	2023	[122]
Shanshan and Sen	Empowering learners with AI-generated content for programming learning and computational thinking: The lens of extended effective use theory	JCAL	2023	[131]
Shoufan	Can Students without Prior Knowledge Use ChatGPT to Answer Test Questions? An Empirical Study	ACM TOCE	2023	[134]
Speth et al.	Investigating the Use of AI-Generated Exercises for Beginner and Intermediate Programming Courses: A ChatGPT Case Study	CSEE&T	2023	[137]
Valový and Buchalcevo	The Psychological Effects of AI-Assisted Programming on Students and Professionals	ICSM	2023	[149]
Wang et al.	Unleashing ChatGPT's Power: A Case Study on Optimizing Information Retrieval in Flipped Classrooms via Prompt Engineering	IEEE TL	2023	[152]
Wu et al.	Research on the Construction of Intelligent Programming Platform Based on AI-generated Content	ICETC	2023	[157]
Yilmaz and Yilmaz	The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation	Computers and Education: Artificial Intelligence	2023	[160]
Agarwal et al.	Which LLM should I use?: Evaluating LLMs for tasks performed by Undergraduate Computer Science Students"	arXiv	2024	[2]
Arora et al.	Analyzing LLM Usage in an Advanced Computing Class in India	arXiv	2024	[9]
Balse et al.	Evaluating the Quality of LLM-Generated Explanations for Logical Errors in CS1 Student Programs	arXiv	2024	[12]
Barambones et al.	ChatGPT for Learning HCI Techniques: A Case Study on Interviews for Personas	IEEE TL	2024	[13]
Bernstein et al.	"Like a Nesting Doll": Analyzing Recursion Analogies Generated by CS Students using Large Language Models"	arXiv	2024	[19]
Cámara et al.	Generative AI in the Software Modeling Classroom: An Experience Report with ChatGPT and UML	IEEE Software	2024	[30]
Chen et al.	Learning Agent-based Modeling with LLM Companions: Experiences of Novices and Experts Using ChatGPT & NetLogo Chat	CHI	2024	[24]
Choudhuri et al.	How Far Are We? The Triumphs and Trials of Generative AI in Learning Software Engineering	ICSE	2024	[25]
Cipriano and Alaves	"ChatGPT Is Here to Help, Not to Replace Anybody" – An Evaluation of Students' Opinions On Integrating ChatGPT in CS Courses	arXiv	2024	[26]
Cipriano et al.	A Picture Is Worth a Thousand Words: Exploring Diagram and Video-Based OOP Exercises to Counter LLM Over-Reliance	arXiv	2024	[27]
Denny et al.	Explaining Code with a Purpose: An Integrated Approach for Developing Code Comprehension and Prompting Skills	arXiv	2024	[31]
Denny et al.	Prompt Problems: A New Programming Exercise for the Generative AI Era	SIGCSE TS	2024	[34]
Haindl and Weinberger	Students' Experiences of Using ChatGPT in an Undergraduate Programming Course	IEEE Access	2024	[46]
Hou et al.	CodeTailor: Personalized Parsons Puzzles are Preferred Over AI-Generated Solutions to Support Learning	arXiv	2024	[50]
Jacobs and Jaschke	Evaluating the Application of Large Language Models to Generate Feedback in Programming Education	arXiv	2024	[52]
Jeuring et al.	What Skills Do You Need When Developing Software Using ChatGPT? (Discussion Paper)	Koli Calling	2024	[53]
Lin et al.	Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education	CHI	2024	[54]
Jury et al.	Evaluating LLM-generated Worked Examples in an Introductory Programming Course	ACE	2024	[57]
Kazemitabaar et al.	CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs	CHI	2024	[60]
Kimmel et al.	Enhancing Programming Error Messages in Real Time with Generative AI	CHI EA	2024	[71]
Kosar et al.	Computer Science Education in ChatGPT Era: Experiences from an Experiment in a Programming Course for Novice Programmers	Mathematics	2024	[73]
Kuramitsu et al.	Training AI Model that Suggests Python Code from Student Requests in Natural Language	Journal of Information Processing	2024	[77]
Liao et al.	Scaffolding Computational Thinking with ChatGPT	IEEE TL	2024	[85]
Liu et al.	Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Education	SIGCSE TS	2024	[87]
Lyu et al.	Evaluating the Effectiveness of LLMs in Introductory Computer Science Education: A Semester-Long Field Study	arXiv	2024	[90]
Ma et al.	Enhancing Programming Education with ChatGPT: A Case Study on Student Perceptions and Interactions in a Python Course	arXiv	2024	[91]
Manley et al.	Examining Student Use of AI in CS1 and CS2	JCSC	2024	[95]
Moore et al.	Teaching artificial intelligence in extracurricular contexts through narrative-based learnersourcing	CHI	2024	[97]
Nam et al.	Using an LLM to Help With Code Understanding	ICSE	2024	[100]
Neyem et al.	Exploring the Impact of Generative AI for StandUp Report Recommendations in Software Capstone Project Development	SIGCSE TS	2024	[102]
Neyem et al.	Toward an AI Knowledge Assistant for Context-aware Learning Experiences in Software Capstone Project Development	IEEE TL	2024	[101]
Nguyen et al.	How Beginning Programmers and Code LLMs (Mis)read Each Other	CHI	2024	[103]
Pankiewicz and Baker	Navigating Compiler Errors with AI Assistance – A Study of GPT Hints in an Introductory Programming Course	arXiv	2024	[106]
Pesovski et al.	Generative AI for Customizable Learning Experiences	Sustainability	2024	[108]
Roest et al.	Next-Step Hint Generation for Introductory Programming Using Large Language Models	ACE	2024	[120]
Shah et al.	Working with Large Code Bases: A Cognitive Apprenticeship Approach to Teaching Software Engineering	SIGCSE TS	2024	[130]
Sheese et al.	Patterns of Student Help-Seeking When Using a Large Language Model-Powered Programming Assistant	ACE	2024	[133]
Singh et al.	Bridging Learnersourcing and AI: Exploring the Dynamics of Student-AI Collaborative Feedback Generation	LAK	2024	[136]
Sun et al.	Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study	International Journal of Educational Technology in Higher Education	2024	[139]
Tanay et al.	An Exploratory Study on Upper-Level Computing Students' Use of Large Language Models as Tools in a Semester-Long Project	arXiv	2024	[140]
Taylor et al.	gcc --help: Transforming the Role of the Compiler by Generating Context-Aware Error Explanations with Large Language Models	SIGCSE TS	2024	[143]
Wang et al.	A Large Scale RCT on Effective Error Messages in CS1	SIGCSE TS	2024	[153]
Woodrow et al.	AI Teaches the Art of Elegant Coding: Timely, Fair, and Helpful Style Feedback in a Global Course	SIGCSE TS	2024	[156]
Xiao et al.	Exploring How Multiple Levels of GPT-Generated Programming Hints Support or Disappoint Novices	CHI EA	2024	[158]
Zhang et al.	Students' Perceptions and Preferences of Generative Artificial Intelligence Feedback for Programming	AAAI	2024	[162]

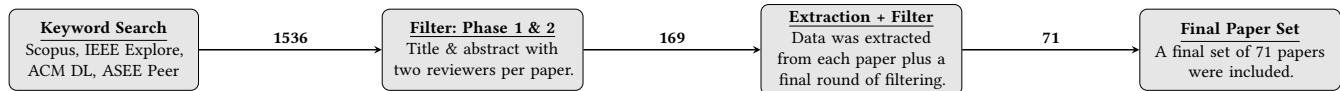


Figure 1: The literature review collection and analysis pipeline.

models” OR “Gemini” OR “Claude” OR “GenAI” OR “GPT-4” OR “GPT-3.5” OR “GPT-3” OR “Copilot” OR “Language model” OR “Language models” OR “Generative pre-trained transformer”

Working Group Focus: “Pedagogy” OR “Pedagogies” OR “Classroom” OR “Student” OR “Students” OR “Teaching approach” OR “Teaching tools”

Method: “Qualitative” OR “Quantitative” OR “Perceptions” OR “Investigating” OR “Exploratory” OR “Survey” OR “Interview” OR “Experiment” OR “Focus group”

The full search string was constructed by combining each of these categories to ensure a paper contained at least one search term per category. As such, the full search string is as follows:

Search String = Domain AND Topic AND Working Group Focus AND Method

We decided to look for papers in a total of five different databases to ensure comprehensive coverage of work. The chosen databases were ASEE Peer, arXiv, Scopus, ACM Digital Library, and IEEE Xplore.

The final search resulted in 1536 papers after duplicate removals. The search was done on May 23rd, 2024, which is the cut-off date for included articles. The breakdown of the number of papers from each database prior to the removals is included in Table 2.

Table 2: The number of papers extracted from each database prior to duplicate removal.

Source	#
ASEE Peer	23
arXiv	47
Scopus	665
ACM Digital Library	721
IEEE Xplore	258

To ensure that no relevant work was missed in our SLR, we broadly sampled the literature. To evaluate our search string for such comprehensive coverage, we selected ten papers that we would expect to find with our search string, shown in Table 3. All ten papers were found with the final search string, which provides evidence that the search string is adequate for finding relevant papers.

2.2 Paper Filtering Process

Balancing the need to broadly include relevant papers while ensuring the central goals of the research remain in focus is a common challenge in SLRs. To address this, we developed exclusion criteria that were applied repeatedly throughout the filtering process, ensuring that only high-quality papers focusing on the use of generative AI in pedagogical practice were included in our final analysis.

A primary criterion for inclusion in the SLR was that the papers prominently featured some form of generative AI. This included studies where students used existing generative AI tools or where such tools were integrated into pedagogical practices. Papers that focused on teaching students about generative AI or its associated ethical aspects were also included.

We included only papers that focused on computing education at the tertiary level. Papers must have featured student participants of some kind and could not focus solely on K-12 or professional developers. However, papers that featured both tertiary students and K-12 or professional developers were included. The reason to focus on tertiary students was that these students are being trained to become computing professionals in the near future.

In keeping with the goal of evaluating robust interventions within computing education, we excluded shorter papers, such as posters, demos, and other shorter formats. Since the typical length of papers in the field of computing education is six pages or more in a double-column format, we used this lower limit to exclude less comprehensive studies.

Based on these goals, we applied the following exclusion criteria at each stage of the filtering process:

- (1) **Not GenAI:** Papers that did not include a generative AI component were excluded. Papers that simply provide implications for generative AI were also not included.
- (2) **Not Computing Education:** Papers that were not related to computing education were excluded. For example, papers primarily focused on professional developers were excluded.
- (3) **No Human Evidence:** The paper did not contain empirical data or included only an expert evaluation.
- (4) **No Intervention:** Papers that did not feature a classroom intervention were excluded. However, user studies with students were included.
- (5) **Exclusively K-12:** Papers where the participants were exclusively K-12 were excluded.
- (6) **Too Short:** Papers were excluded if they were under 5 pages for double-column articles and under 8 pages for single-column articles.
- (7) **Not an Article:** We excluded conference proceedings’ front matters, white papers, or other non-research content.

In the first phase, we had two reviewers evaluate the title and abstract of the paper for inclusion. If either of the reviewers thought that the paper was relevant, it was chosen for the next stage – extraction.

2.3 Extraction

For all the papers that were not excluded in the title/abstract scanning phase, we thoroughly read the full paper. A paper could still be rejected at this stage if it did not pass the inclusion/exclusion criteria. To consistently systematically extract the content from

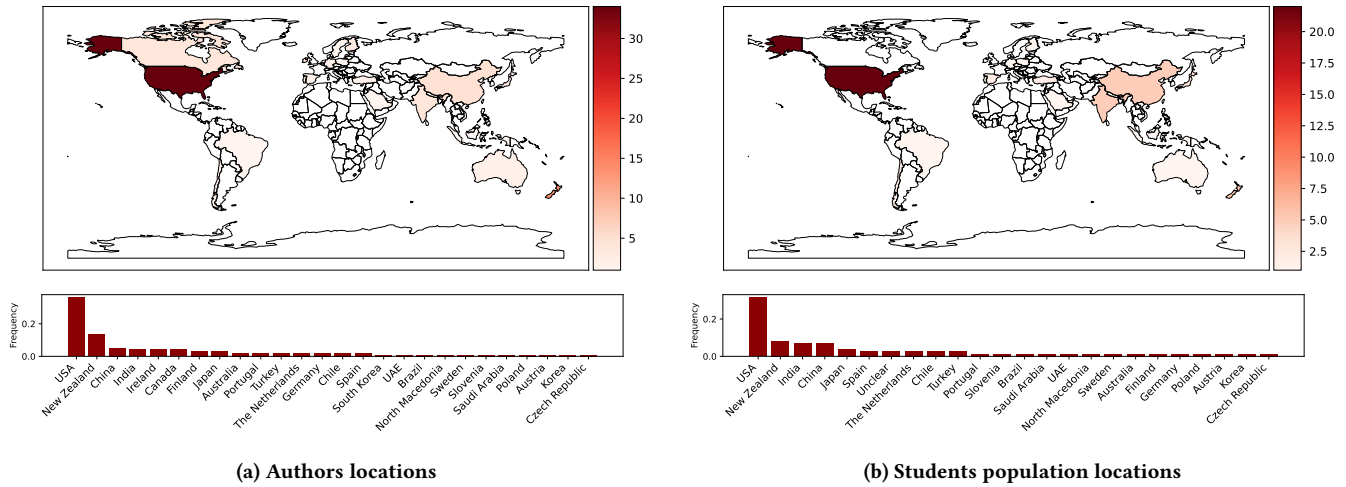


Figure 2: Heatmaps of where authors publishing are located (Figure 2a) and where the student populations they are investigating are located (Figure 2b)

Table 3: The reference papers used to evaluate the quality of the search string.

Paper title	Citation
Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Education	[87]
Prompt Problems: A New Programming Exercise for the Generative AI Era	[34]
CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes	[86]
Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book	[94]
Evaluating LLM-generated Worked Examples in an Introductory Programming Course	[57]
Next-Step Hint Generation for Introductory Programming Using Large Language Models	[120]
ChatGPT for Learning HCI Techniques: A Case Study on Interviews for Personas	[13]
CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs	[60]
“ChatGPT Is Here to Help, Not to Replace Anybody” – An Evaluation of Students’ Opinions On Integrating ChatGPT In CS Courses	[26]
Explaining Code with a Purpose: An Integrated Approach for Developing Code Comprehension and Prompting Skills	[31]

each paper, we developed a structured form shown in Appendix F to extract information for all papers that were included in the review. Four researchers extracted the content from the final set of 71 papers.

3 Systematic Literature Review: Results

In this section, we present the results of the systematic literature review. They are based on the information extracted from the final set of 71 papers that resulted from the process described in the previous section.

3.1 Descriptive Statistics

From the evaluation form we collected a variety of descriptive data that relates to: 1) the authors and the students they evaluated, 2) the characteristics of the studies they conducted, 3) the types of courses in which these studies took place, and 4) the custom tools that have been developed.

Course Information. A variety of courses were used to study AI tools. We saw 26 upper division courses including three masters’ level courses. CS1 was studied in 20 papers. Eight papers included more than one course. Examples of other courses were Human Computer Interaction, Software Modeling, and Embedded Systems.

Author and Student Information. As shown in Figure 2a, the locations of the authors’ institutions varied widely with the largest proportion of articles having authors from the United States (34 %) and New Zealand (14 %). Additionally, they were primarily from academic institutions ($n=67$) with very few coming from industry or involving industry-academic collaborations ($n=2$), or just from industry ($n=1$). As might be expected, the majority of studies took place at one or more of the authors’ institutions which results in the distribution of student populations that were investigated looking quite similar to the one for authors (Figure 2b).

Table 4: Characteristics of the studies and their methods.

Methodology	#
Both	36
Quantitative	25
Qualitative	9

(a) Methodologies Used

Study Type	#
Unsupervised study	41
Supervised study	24
Both	2
Unclear	3

(b) Level of Supervision

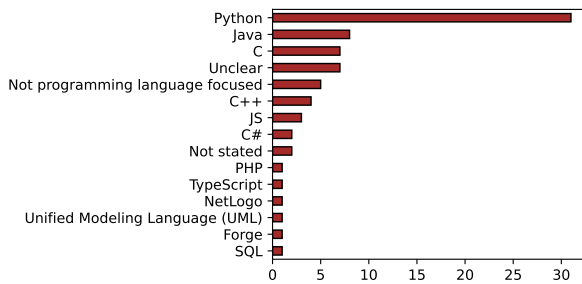


Figure 3: The programming languages reported on by the studies.

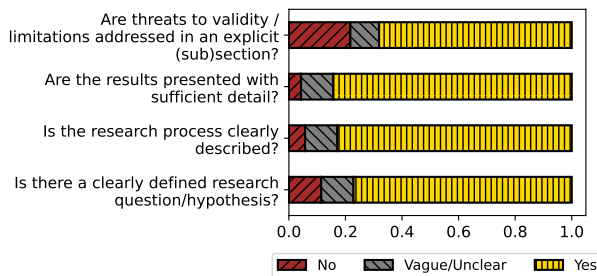


Figure 4: Quality of work as evaluated using the paper quality metrics of Hellas et al. [47].

Study Context and Quality. The majority of the evaluations used a mixed methods approach, followed by quantitative methods (Table 4a). With respect to the student populations under investigation both mixed-methods (median=52.0 IQR=(24.0-105.0)) and quantitative (median=56.0 IQR=(49.0-160.0)) studies investigated student populations of similar sizes. Purely qualitative studies, as might be expected, had the smallest number of participants (median=21.0 IQR=(12.0-72.0)). Additionally, the majority of studies conducted were done in an unsupervised (i. e., uninvigilated) manner (Table 4b) and Python was by far the most common language being used (Figure 3).

Using the quality metrics presented by Hellas et al. [47], we evaluated each paper according to its presentation of: 1) addressed limitations, 2) descriptions of the presented results, 3) description of the research process, and 4) presentation of the research questions. Overall, we find the majority of the empirical work we evaluated were sufficient across all four of these dimensions (Figure 4). The dimension that saw the largest degree of vagueness (20 %) or exclusion (11 %) was that of clearly defined limitations.

Custom Tools. In the course of our literature review we uncovered a wide variety of custom tooling that has been developed for the support of instructors and students alike. These included: CodeAid [60], CodeHelp [86, 133], WorkedGen [57], LLM Hint Factory [158], Tutor Kai [52], CodeTutor [90], GPTeach [96], Charlie the Coding Cow [103], KOGI [78], NetLogo Chat [24], Promptly [33, 34], IPSSC [85], CS50 Duck [87], Gilt [100], CodeTailor [50], and a variety of others that lacked names [76, 97, 101, 156].

Using Generative AI for Teaching vs Teaching Students to Use Generative AI. In relation to the research goals of the group, we categorized the literature into two broader categories: studies where generative AI was *used* for some pedagogical purpose and studies where the main goal was to *teach* students about generative AI specifically or to use a specific generative AI tool. Of the papers reviewed, 57 fell into the *use* category and the remaining 14 were categorized as *teach*.

3.2 How Is Generative AI Being Incorporated into Teaching?

In the extraction form, we had an open-ended question on “How Instructors Incorporate Generative AI into Teaching Computing?”, which directly maps to our SLR-RQ2.

To analyze the results for this question, three authors thematically analyzed the open-ended responses and came up with initial tags. They discussed the initial tags and then combined them into definitive tags along three axes: tool type, purpose from the student’s point of view, and whether students received guidance on how to use generative AI.

Our tags for each axis are below.

- Tool type: general purpose (e. g., ChatGPT), task-specific (e. g., GitHub Copilot), instructor-provided guardrails (e. g., CodeHelp [86], Promptly [33], CodeAid [60], CodeTailor [50]).
- Purpose: hints, debug, learning resources, writing code, teacher training, code comprehension, code review, teach GenAI, motivation, UML, multiple, not specified.
- Guidance on GenAI use: yes, no, unclear.

For tool type, we categorized tools as general purpose (such as ChatGPT), whether the tool is task-specific (i. e., the tool is meant for a specific task, but not education-focused, such as GitHub Copilot), and whether there are instructor-provided guardrails (i. e., there were pedagogical guardrails or other constraints in the tool).

For purpose from the student’s point of view, we looked at the tasks that students worked on with the support of GenAI. During the thematic analysis, we came up with the following categories.

- **Writing code** – GenAI was used for code writing support. For example, CodeTailor helps students write code and interact with Parsons problems [50].

- **Code comprehension** – GenAI was used to teach code comprehension. For example, Gilt provides scaffolding contextualized to select sections of students’ code [100].
- **Hints** – GenAI was used to produce hints for students. For example, next-step hints [120].
- **Learning resources** – GenAI was used to create or improve learning resources that could be used by other students too. For example, students were instructed to generate analogies using LLMs [19].
- **Teach generative AI** – the article describes an approach or tool to teach students how to use GenAI. For example, Promptly provides scaffolding for students to learn how to prompt GenAI [33].
- **Multiple tags** – GenAI was used for multiple purposes. For example, Choudhuri et al. investigated students’ use of GenAI for multiple different software engineering tasks such as debugging and refactoring [25].
- **Debug** – GenAI was used for debugging help, such as by explaining error messages [131].
- **Code review** – GenAI was used for code review. For example, by integrating it into an assignment submission system [29].
- **UML** – GenAI was used to assist in generating UML diagrams. This was done by Cámara et al. [30].
- **Teacher training** – GenAI was used for teacher or teaching assistant training. GPTeach creates LLM agents that act as students to train new TAs [96].
- **Motivation** – GenAI was used to increase student motivation. This was done by Moore et al. in a narrative-based learnersourcing platform [97].
- **Not specified** – the purpose of GenAI was not specified in the paper or was unclear.

For guidance on GenAI use, we categorized each paper as a “yes”, “no”, or “unclear” (i. e., whether students received guidance or instructions on how to use generative AI).

The results of this analysis are presented in Table 5. Based on the findings, most studies focused on using generative AI for writing code, code comprehension, hints, and generating learning resources. Slightly under half (32/71) of the studies used tools with instructor-provided guardrails, for example, a custom tool with pedagogical guardrails. However, most commonly (34/71), studies used a general purpose generative AI tool such as ChatGPT. In the majority of studies (49/70), it was not reported that students would have been instructed on how to use generative AI.

We also looked at the type of evidence used, categorizing it into perceptual (e. g., opinions of activity) or behavioral (e. g., correctness of produced code). This was contrasted with the nature of the findings, which were categorized as positive, negative, mixed, or neutral. The results of this analysis are presented in Table 6. The results suggest that the majority of studies found positive results, with studies using behavioral evidence slightly more likely to report positive findings. 54 % of studies with perceptual evidence reported positive findings whereas 69 % of studies with behavioral and 70 % of studies with both types of evidence reported positive findings.

Table 5: The characteristics of studies in terms of (a) their intended purpose, (b) the type of the tool, and (c) guidance provided on how to use GenAI.

Task	#	Response	Count
writing code	26	general purpose	34
code comprehension	10	instructor	32
hints	8	guardrails	
learning resources	7	task-specific	4
teach GenAI	6	unclear	1
multiple	5	(b) Tool Type	
debug	2		
code review	2		
UML	1		
teacher training	1		
motivation	1	no	49
not specified	2	yes	21
		unclear	1
(a) Intended Purpose		(c) Guidance on GenAI Use	

Table 6: Type of evidence versus nature of findings. Note that one study had positive findings, but unclear type of evidence, leading to the numbers only summing to 70.

	Positive	Negative	Mixed	Neutral	Total
Perceptual	13	1	9	1	24
Behavioral	11	1	1	3	16
Both	21	3	5	1	30
Total	45	5	15	5	70

3.3 What Are the Motivations Behind Incorporating GenAI Tools into Teaching?

In the extraction form, we had an open-ended question on “And why do they incorporate GenAI tools that way?”, which directly maps to our SLR-RQ3. In the previous section, we outlined *how* instructors are incorporating generative AI into their teaching. Part of the *why* is overlapping – the ways it is incorporated often tell about the *why*. Thus, we here focus on *in benefit of whom* it is being integrated. We tagged each of the included papers with “students”, “instructors” or “both”. Out of the 71 papers, in 59 GenAI was incorporated in benefit of students. In 5 papers it was in benefit of teachers. In 7 papers it was in benefit of both students and teachers.

3.4 Recommendations for Incorporating GenAI

To analyze the effectiveness of incorporating generative AI into computing classrooms, we cross-tabulated the type of tool and guidance on GenAI use provided to students with the nature of the findings (positive, negative, mixed or neutral). The results of this analysis are reported in Table 7. The main result that can be seen from the table is that when there is no guidance on using GenAI from the instructor, it is recommended to use a tool that includes instructor-provided guardrails, e. g., pedagogical guardrails. When students are not provided guidance and use a general purpose generative AI model (e. g., ChatGPT), only 55 % (12/22) of studies

Table 7: Comparing the types of tools and level of guidance provided.

Guidance	Tool type	Positive	Negative	Mixed	Neutral	Total
no	general	12	3	4	3	22
no	task-specific	1				1
no	guardrailed	19		5	2	26
yes	general	7	1	4		12
yes	task-specific	2		1		3
yes	guardrailed	4	1	1		6

Table 8: Comparing outcomes based on the task type.

	Positive	Negative	Mixed	Neutral	Total
Writing code	15	2	8	1	26
Code comprehension	8	1	1		10
Hints	4		2	2	8
Learning resources	6		1		7
Teach GenAI	4	1	1		6
Multiple	2	1	2		5
Debug	1			1	2
Code review	2				2
UML	1				1
Teacher training	1				1
Motivation	1				1
Not specified	1			1	2
Total	46	5	15	5	71

found positive results. However, even without guidance, if the tool included instructor-provided guardrails, positive results were found in 73 % (19/26) of studies. When students are given instructions on how to use generative AI, whether the tool has built-in pedagogical guardrails does not seem to matter as much – studies that used general purpose tools found positive results in 58 % (7/12) of cases and studies that used instructor-guardrailed tools found positive results in 67 % (4/6) of cases.

In a similar vein, we cross-tabulated the nature of the findings with the task that GenAI was used for to examine what computing education tasks might benefit most from generative AI. The results of this analysis are reported in Table 8. There are differences between tasks in whether findings of the studies have been positive or not. For code writing, 58 % (15/26) studies reported positive results, whereas for code comprehension, 80 % (8/10) of studies reported positive results. According to our results, hint generation is an area where generative AI could still improve, since only half of the studies (4/8) reported positive results. Better results have been observed for generating learning resources, where 86 % (6/7) studies found positive results.

4 Educator and Developer Views: A Mixed-Methods Approach

Computing educators want to prepare students for a successful career in software development. It is thus important to understand the experiences of software developers in addition to educators’ perspectives to evaluate how different the two perspectives are. For

this reason, we conducted a survey study with both target groups, and an interview study with educators.

Through the educator survey study, we aimed to gather a large sample of educator perspectives on whether (and how) they incorporate GenAI in their classroom, their motivations for this decision, and how they see competencies changing for programming education. We also surveyed developers to build a landscape of how GenAI tools are currently being used in industry settings and how developers view the changing competencies required for programming. The larger sample size of the survey also allows us to begin to explore equity-related questions related to student access and exposure to GenAI tools.

Through the interview study, we aimed to gather first-hand accounts, and deeper insights from educators on changes to their classroom as a result of emerging GenAI tools. We interviewed educators who teach students how to use GenAI tools to engage with class materials as well as how to use such tools in their professional careers. We also included educators who took a deliberate (explicit) stand to disallow the use of GenAI tools in their classes. Further, we interviewed developers of LLM-powered educational tools as well as researchers focused on the use of GenAI in computing education.

In this section, we first summarize prior studies that similarly gathered perspectives of educators and developers. Then we present the ten research questions that we answer through both the surveys and interviews. Next, we describe the applied methodology.

4.1 Prior Studies of Educators' and Developers' Perceptions

Computing educators' perspectives on GenAI tools have been reported in several prior studies. For example, Chan and Lee [23] surveyed 184 educators and 399 students, primarily from Hong Kong but across different disciplines. They focused on the perceptions, experience, and knowledge about GenAI, and compared educator and student responses. They found that the educators seemed more concerned about students' over-reliance and ethical issues and were skeptical towards GenAI tools and their capabilities. The need for policies and guidelines ensuring academic integrity and equitable learning conditions was yet another outcome [23].

Amani et al. [5] also investigated students' and instructors' perceptions towards GenAI through two surveys at Texas A&M university. The goal of the instructor survey was to capture how GenAI affected their recent courses and how they think students should use respective tools. One important finding is that the responses from 243 staff members emphasize the need for teaching practices to adapt [5]. Another survey of educators was conducted by Prather et al. [111]. The 57 respondents elaborated on their perceptions, experience, usage, course policies, expectations, and beliefs, indicating that educators should stay abreast of the technological developments. The results also highlight the need to provide guidance for students regarding the ethical use of GenAI.

In addition to these surveys, a number of recent research studies used interviews to gather the instructor perspective on the impact of GenAI on Computing education research and practice [80, 111, 118, 132, 154, 161]. For example, Lau and Guo conducted interviews with 20 instructors about their intentions to adapt their teaching to emerging GenAI tools (e. g., ChatGPT and Copilot) [80]. Wang et al. [154] also interviewed 11 instructors about their perceptions. Despite sharing concerns about over-reliance and misuse of GenAI tools, the instructors did not have plans to adapt their courses due to the lack of effective strategies at the time. Another example is the interview study with 40 instructors by Rajabi et al. [118], which showed that educators were cautious about banning AI tools, because students would find ways of using them regardless. At the same time, instructors seemed concerned about increasing anxiety among students by focusing too heavily on exams, which is a well-known issue [64, 79, 92]. A recent interview study with educators was conducted by Sheard et al. [132]. They focused on educators' current practices, concerns, and planned adaptations relating to these tools. They found, for example, that educators appreciate that the tools can be a source of support for students, but that these tools may lead to students missing out on learning. Finally, Zastudil et al. [161] conducted an interview study comparing the perspectives of 6 instructors and 12 students, finding alignment in their concerns about over-reliance and misalignment in their motivations and relative knowledge about generative AI.

While there has been some work in studying how developers can use GenAI tools to enhance their workflow, little has been done to understand the perceptions of how software developers see the practice changing. Specifically, one study found that software developers are using GenAI tools at many points in the programming process, including creating, modifying, debugging, and explaining

code [15]. Recent work reports that GenAI automated code generation increases developer productivity [83], and GenAI tools are particularly useful for assisting in resolving technical issues [28]. Typically, productivity is measured in terms of the acceptance rates of coding suggestions [163].

In summary, we identified several survey and interview studies about the perceptions of educators regarding the *prospective* adoption of courses, learning objectives, assessments, or institutional policies, but not the *actual* adoption. Moreover, we found fewer studies focused on developers. In our survey study, we focus on both instructors and developers, to understand the perspectives of experts both within the learning environment and within the industry.

4.2 Research Questions

Based on the overarching research goals (see Section 1), the systematic literature review, and the identified gap on actual integration practices in education and industry, we constructed the following research questions guiding our work:

- RQ1 Policies:** What are the existing policies and practices around using GenAI in Computing courses? (1b)
- RQ2 Instruction-use:** How are instructors teaching students about using GenAI tools in order to program? (1a)
- RQ3 Instruction-support:** How are instructors using generative AI tools to support the teaching of their courses? (1a)
- RQ4 Motivations:** Why are instructors making these choices around using GenAI? (1b)
- RQ5 Tools:** What kinds of GenAI tools are emerging in Computing education? (1a)
- RQ6 Perceived Outcomes:** How has GenAI impacted instructor perceptions of student outcomes (compared to before)? (2a)
- RQ7 Industry usage:** How are industry developers using GenAI tools? (2a)
- RQ8 Competencies:** How has GenAI impacted desired student competencies? (2a, 2b)
- RQ9 Equity:** How has GenAI impacted student equity? (1a, 1b, 2b)
- RQ10 Future:** How is GenAI shaping the future of Computing education? (2b)

We intentionally divided RQ2 and RQ3 into two distinct research questions as the Working Group believes it is valuable to emphasize that GenAI tools are used in the classroom and also impacting course learning outcomes. Specifically, RQ2 focuses on investigating how educators prepare students for the future of programming where GenAI tools are available. This may include instructing them about GenAI capabilities for programming or allowing GenAI tools to be used by students as a method to teach students how to effectively incorporate tools into their workflow. RQ2 is in contrast to RQ3 which explores how educators incorporate GenAI tools into the classroom and their workflow. This could include using GenAI tools for grading, assignment creation, student feedback, or providing help (like a TA-bot).

4.3 Mixed-Methods

To address the research questions regarding educators' and software developers' viewpoints, we employed a mixed-methods approach.

To capture the breadth of GenAI usage and perspectives in computing education, we conducted two surveys, one targeting educators and the other targeting developers. To more deeply understand current views and potential future uses for GenAI, we carried out an interview study involving tool creators, educators studying GenAI, and educators using GenAI.

The subsequent sections describe the design of both the surveys and interview study. Since our studies were structured around our research questions, we have organized the results section accordingly. Rather than presenting survey and interview results separately, we integrate the findings from both methods for each research question. For clarity, we present the mapping of research questions and survey questions in Table 9.

4.4 Survey Development

To address our research questions, we developed surveys to quantify the opinions and behaviors of a broad range of CS educators and professional programmers. Because CS educators have different responsibilities and may have different perceptions than professional programmers, we created two separate surveys: an educator survey and a developer survey.

To design the questionnaires, we collaborated within our group to draft survey questions for each overarching research question, including closed-ended, open-ended, and rating scale question types. As a part of this process, we reviewed the literature (e.g., [111]) and questions from existing surveys. However, few questions could be reused due to our focus on the rationale and actual integration of GenAI tools rather than expectations or anticipated use.

We then reviewed and edited these questions to increase clarity, relevance, neutrality, and specificity, following best practices from survey design [74]. When appropriate, questions were shared between the educator and developer surveys to allow us to compare responses between the two audiences. As part of our development process, we also piloted our survey with educators and developers to increase alignment with our target audiences.

4.4.1 Educator Survey. The educator survey was developed to address the two overall goals of the working group, and most of the research questions mentioned earlier (see Section 4.2 and Table 9 for the precise mapping of survey questions and research questions). As the previous year’s working group had surveyed educators [111], we also drew from their survey questions where possible to facilitate potential comparisons. We expected educators to have a wide variety of views on the topic of GenAI and designed the survey to be multi-pathed based on their earlier responses. The full survey comprises 30 questions. It can be found in Appendix A. The precise mapping of survey questions to research questions is shown in Table 9.

Prior studies conducted in the very early days of LLMs emergence and general availability necessarily needed to focus on short- and long-term concerns of stakeholders, as many educators were unsure of how to proceed. Moreover, educators had broadly taken one of the two approaches to GenAI: either they do not use GenAI and disallow students from using it, or they actively include it in their teaching or at least allow students to use it [80]. There are still open questions, of course, but now the community has had more time to process the upheaval and more research to inform their

teaching. As such, our survey shifts to focus on specific reasons, opportunities, and constraints for why educators and industry professionals use or do not use GenAI. To increase specificity and to capture a diverse range of perspectives, our educator survey also has separate branches for instructors that allowed or did not allow their students to use GenAI tools. We are also asking about concrete teaching approaches and assessments.

We are also particularly interested in the ways that learning to program, and programming in industry, are changing. To that end, we have asked several questions around whether educators believe that programming skills are shifting, how competencies (meaning knowledge, skills and dispositions in context of a task [117]) have or have not changed, and which of them may become more or less relevant. The latter has been subject to discussion ever since GenAI tools have emerged [16]. Another ongoing concern is around equitable access to these tools. As such, in our educator survey, we ask which tools students are expected to use and how they will access those tools (including whether there is a cost).

4.4.2 Developer Survey. The developer survey was created for three purposes: (1) to capture an overarching view on how GenAI tools are being used in professional software development, (2) to capture developer viewpoints on how the tools are impacting their workflow, and (3) to capture developer viewpoints on the expected changes in competencies required to develop software. Another aspect we are addressing is whether and how developers have experienced issues regarding ethical concerns, or harm being done by the use of GenAI tools.

Gathering the developers’ perspectives helps us develop a grounded understanding of how software development is changing in the field. Moreover, we can compare their viewpoint with the educators’ perspective. This enables a more realistic evaluation of the current state-of-the-art in industry, which may eventually contribute to adapting teaching practices and curricula. All 19 questions of the developer survey are available in Appendix B. It was designed to address both users and non-users. The mapping of survey questions to research questions is shown in Table 9.

4.5 Distribution of the Surveys

To recruit instructors, we sent emails to colleagues, public mailing lists, and private mailing lists. We also asked recipients to forward the survey links to peers to increase its distribution. We sent emails that contained recruitment for the educators survey to the following groups and mailing lists:

- SIG “Computer Science Education” (ACM)
- SIG “Educational Technology” of the German Computer Science society
- SIG “Human-Computer-Interaction” of the German Computer Science Society
- Participants at ITiCSE 2024 conference
- Contacts at Minority Serving Institutions (MSIs)
- A Slack group for teaching faculty who are predominantly in the United States
- Departmental mailing lists of the authors
- Colleagues of the authors
- LinkedIn networks of the authors

Table 9: Mapping of the research questions (1st column) and respective survey questions (2nd column) in the Educator Survey (ES) and Developer Survey (DS).

Research Questions	Survey Questions
RQ1: Policies	ES-1: Do you explicitly disallow students to use GenAI tools for all of your computing courses (within the last 12 months)? ES-2: Are you incorporating GenAI tools (e.g., actively integrating it into the curriculum or exercises) into your recent courses (within the last 12 months)? ES-10: Are you doing anything to prevent GenAI tools' use in your course? ES-11: What are you doing to prevent GenAI tools' use in your course? ES-21: Please describe any changes you have made to your teaching approaches in courses you are teaching as a result of GenAI tools. ES-22: Please describe any changes you have made to your assessment approaches in courses you are teaching as a result of GenAI tools. DS-14: What is the main reason that you do not use GenAI tools for professional software development?
RQ2: Instruction-use and RQ3: Instruction-support	ES-12: We ask you to think of a recent course (within the last 12 months) that you teach that is most influenced by GenAI tools. ES-13: Select the size of the recent course that you teach that is most influenced by GenAI tools ES-14: Who uses (or is expected to use) GenAI tools in your course(s)? ES-15: If students are allowed to use GenAI tools, how do you expect students to access them? ES-16: Which type of GenAI tools are you incorporating into your recent course that is most influenced by GenAI tools? ES-17: In what ways have you incorporated GenAI tools into your recent course that is most influenced by GenAI tools?
RQ4: Motivations	ES-9: Why don't you allow GenAI tools in your courses? ES-18: Why have you incorporated GenAI tools into your recent course? ES-20: Why have you not incorporated GenAI tools (e.g., actively integrating it into the curriculum or exercises) into your recent courses (within the last 12 months)?
RQ5: Tools	ES-16: Which type of GenAI tools are you incorporating into your recent course that is most influenced by GenAI tools?
RQ6: Perceived Outcomes	Interviews only
RQ7: Industry usage	ES-7: How often do you believe professional software engineers are using GenAI tools as part of their professional role? ES-8: For which tasks or contexts are industry professionals using GenAI tools? DS-1: Do you use GenAI? DS-2: How often do you use GenAI? DS-3: What types? DS-4: Describe how you use them DS-5: Select the tasks for which you use GenAI DS-6: How not useful or useful have GenAI tools been to your software development? DS-7: Have GenAI tools made your software development more or less efficient? DS-8: How not harmful or harmful have GenAI tools been to your software Development? DS-9: If you consider GenAI tools harmful, please describe a situation you have experienced, e.g., what were you doing, what did you expect, why was the use of the GenAI tools harmful and to whom? DS-14, DS-15: What is the main reason that you do not use GenAI tools for professional software development?
RQ8: Competencies	ES-3: Do you believe the skills to create software have changed after the advent of GenAI tools? ES-4: Please elaborate on your last response why skills have not changed. ES-5: In what ways do you think the skills needed to create software have changed with the introduction of GenAI tools? ES-6: When using GenAI tools to create (parts of) software, which skills become the most important (select up to 3)? ES-19: Have you changed any of the learning objectives of your recent course based on the capabilities of GenAI tools? DS-10: Did the competencies (i.e., knowledge, skills, dispositions in context of a task) required to professionally develop software change with the availability of GenAI tools? DS-11: If you have seen changes, from your experience with GenAI tools, what do you believe are new relevant competencies to professionally develop software with GenAI tools? DS-12: If you have seen changes, from your experience with GenAI tools, what do you believe are competencies that are no longer or less relevant to professionally develop software with GenAI tools?
RQ9: Equity	ES-26: Do you teach at an institution that serves a minority population in your country?
RQ10: Future	Interviews only
All	DS-13, DS-16: What advice would you give to novice programmers regarding the use of GenAI tools?

To recruit developers in industry, we sent emails to personal contacts, public and private mailing lists. We also asked recipients to forward the survey links to peers to increase its distribution. We sent emails with the developer's survey link to the following mailing lists and social networks:

- SIG "Educational Technology" of the German Computer Science society

- SIG "Human-Computer-Interaction" of the German Computer Science Society
- Personal contacts through online networks (e. g. LinkedIn)
- Miscellaneous company lists
- CS department alumni groups

In addition, the German "Fraunhofer Gesellschaft", a non-profit organization for research and development, helped share the survey

among their scientific staff and industrial partners. The same applies to the authors' network with the Leibniz Association—a connection of 96 research institutes in Germany.

4.6 Survey Data Analysis

The survey data were analyzed using a mixed-methods approach to accommodate both closed-ended and open-ended questions. Quantitative responses from closed-ended items were processed using descriptive statistics and statistical data visualizations.

Qualitative data from open-ended questions underwent a thematic analysis [144]. Two members of the research team coded responses, identified recurring themes, and categorized them into broader conceptual groups. To improve reliability, the two researchers independently coded a subset of responses, compared their coding schemes, and resolved discrepancies through discussion until reaching an agreement.

For the educator survey, we received 209 responses (cutoff date July 29, 2024). However not all responses were used. We did not use the responses for which the participants did not agree to the consent form or which were incomplete (100 responses). In addition, we had some test data that were also not included (33 responses). Based on these criteria, we have $N = 76$ fully complete responses for the present analysis.

For the developer survey, we received 94 responses (cutoff date 29 July 2024). Again, not all answers were used: we did not use the responses for which the participants did not agree to the consent form or which contained no data. Based on these criteria, we analyze the resulting $N = 39$ response sequences, 29 of which contain responses to all survey questions (and are thus fully complete).

4.7 Interviews Method

Before conducting the interviews, we compiled a list of individuals who fit into one of the following three categories of interest:

- **Tool creators:** This group comprises faculty, graduate students, software developers, and tech leads who design, build, and refine GenAI systems for educational use. These tools focus on enhancing coding efficiency, supporting educational needs, and integration into various computing environments. Insights from this group provide a critical understanding of the technical challenges and opportunities associated with the adoption of LLMs in computing education.
- **Educators studying GenAI:** Educators studying GenAI are primarily researchers and academic professionals who investigate these technologies' implications, efficacy, and educational potential. This group includes faculty members, educational researchers, and curriculum developers who are exploring how LLMs can be leveraged to enhance learning outcomes, transform teaching methodologies, and address the evolving needs of computing education. They examine the theoretical underpinnings, practical applications, and ethical considerations of using AI in educational settings.
- **Educators using GenAI:** Educators using GenAI are instructors actively incorporating LLMs into their teaching practices. This group spans faculty members from computer science and related disciplines, teaching both majors and

non-majors. These educators use AI tools to facilitate learning, provide personalized support, and improve the overall educational experience. They bring practical perspectives on the benefits and challenges of integrating AI into the classroom, including its impact on student engagement, assessment, and skill development. Their experiences offer a practical understanding of how GenAI can be effectively implemented to enhance teaching and learning in computing education.

The interviewees have been active in the computing education community in at least one of the areas identified above. The research team completed 4, 5, and 7 interviews in each of the respective areas (in the order as listed above), and one interview with a thoughtful non-user. We additionally interviewed the thoughtful non-user to obtain broad perspectives.

We used the research questions outlined in Section 4 to pursue our overall research goals. We used these research questions to guide our analysis of the transcripts.

4.7.1 Interview Question Development. Developing interview questions was a collaborative and iterative process involving a team of four members. This team was responsible for designing a set of questions associated with each category of individuals being interviewed: tool creators, educators studying GenAI, and educators using GenAI. The primary aim was to ensure that the questions would elicit key insights from these individuals while addressing the main research questions of the study.

Initially, the team drafted a comprehensive list of questions for each category, focusing on capturing the unique perspectives and experiences of the interviewees. The questions were designed to explore various aspects of using LLMs and GenAI in computing education, including their development, implementation, and impact.

Once the initial set of questions was prepared, it was presented to the larger group of researchers for feedback. During this revision phase, several modifications were made to refine the questions further. Modifications include:

- **Content Overlap:** Some questions were removed due to redundancy and overlap in content to ensure that each question addressed distinct aspects of the research.
- **Duration Management:** To keep the interviews within a manageable duration of one hour, certain questions were cut. This decision was made to respect the interviewees' time and maintain the focus and depth of the discussions.
- **Improving Quality and Directedness:** The wording of several questions was revised to enhance clarity, specificity, and directedness. This helped ensure the questions were straightforward and elicited detailed, relevant responses.
- **Adding Follow-Up Questions:** To extract more context and deeper insights, follow-up questions such as "Why?" and "Can you elaborate more?" were incorporated. These prompts encouraged interviewees to provide more detailed explanations and examples.

4.8 Interview Process

The interview process was designed and implemented to gain comprehensive insights. To ensure a diverse and representative sample from each category, we selected individuals for an interview based

on the results of our literature review, i. e., who is writing about this topic from the perspectives of educators using GenAI in their classes or researching GenAI in computing or creating GenAI tools for computing classrooms. A total of 17 individuals were selected to interview, with 4, 5, and 7 individuals from each of the previously identified categories, respectively. The one remaining participant could be classified as a thoughtful non-user because we felt it important to hear from someone who was intentionally and thoughtfully trying to avoid using it in their computing classroom. These participants were invited to participate in one-on-one interviews, and they accepted and signed the consent form.

4.8.1 Conducting Interviews. A single team member conducted each interview via Zoom. The scheduling of these interviews ensured that each session lasted for no more than one hour to adhere strictly to the allocated time frame. Using Zoom facilitated a convenient and efficient way to conduct these interviews remotely.

4.8.2 Recording and Transcription. All Zoom sessions were recorded with participant permission to create accurate transcriptions. To comply with Institutional Review Board (IRB) protocol requirements, these recordings were deleted once the transcriptions were completed. We used secure transcription services to create the transcripts and cleaned them by removing typographic errors.

4.8.3 Interview Structure. The interview questions were posed to each interviewee sequentially. It is important to note that these questions were not shared with the participants prior to the interview. This approach aimed to elicit spontaneous and genuine responses to provide more authentic and valuable data for the study.

4.9 Interview Data Analysis

This section describes the process used to analyze the interview transcripts. The analysis was conducted systematically, involving multiple team members to ensure consistency and reliability.

4.9.1 Initial Review and Consensus Building. Three team members jointly reviewed one transcript to establish a consistent tagging methodology. During this review, they tagged sections (which could be one sentence, multiple sentences, one paragraph, or multiple paragraphs) with relevant categories. The predefined categories used for tagging were: Instruction-use, Instruction-teach, Tools, Policies, Motivations, Competencies, Outcomes, Future, Industry, and Equity, aligned with the research questions described in Subsection 4.2. This collaborative step ensured that all team members were aligned on the tagging criteria and approach.

4.9.2 Individual Tagging and Data Compilation. After reaching a consensus on tagging methodology, the remainder of the transcripts were independently tagged by one researcher, who was not the interviewer. This approach helped maintain objectivity and consistency in the tagging process. The tagged segments from all transcripts were then compiled into an Excel sheet. This compilation step included all identified segments categorized by theme.

Each segment entry in the Excel sheet was summarized into one or a few sentences. This step was used to capture the essence of the responses without losing the context or the main points expressed by the interviewees.

4.9.3 Thematic Analysis. With the summarized segments, we conducted a thematic analysis across all interviews. This analysis involved 1) grouping summaries by category, 2) identifying patterns and themes, and 3) synthesizing community thoughts.

Through this process, we aim to derive meaningful conclusions from the interview data and understand the community's views on the various aspects covered in the interviews.

5 Educator and Developer Views: Results

We now present the results of our surveys and interviews. We begin with the participants' demographics from both surveys and the interviews. Following this, we present the aggregated survey and interview results for each research question.

5.1 Demographics of the Samples

First, we present the characteristics of the surveyed educators and developers by providing details on their demographics. We also present the demographics of the interviewed educators.

5.1.1 Educator Survey Demographics. In the educator survey (N = 76), educators from the USA (28), Germany (13), Canada (11), UK (5), and more countries participated, as shown in Table 10. Overall, we gathered educators' perspectives from North America, Europe, and Australia. However, 6 respondents did not indicate the country of their institution.

Table 10: Educator Survey – Country Institution is Located

Country	Percent Respondents
USA	36.8 %
Germany	17.1 %
Canada	14.5 %
UK	6.6 %
Sweden	5.3 %
Australia	1.3 %
Finland	1.3 %
France	1.3 %
Iceland	1.3 %
Ireland	1.3 %
Netherlands	1.3 %
Poland	1.3 %
Spain	1.3 %
Switzerland	1.3 %
Ukraine	1.3 %
No country given	6.6 %

31.6% (24 out of 76) of the educators identify themselves as females, 57.9% (45 out of 76) as males, 2.6% (2 out of 76) as non-binary or gender diverse, and 5.3% (4 out of 76) prefer not to disclose (ES-29). The majority of the educator respondents described their institutions (ES-24) as a university that grants graduate degrees (52 out of 76), and the others teach at colleges or other types of institutions, as shown in Table 11.

Most of the educators teach *CS1 – Introduction to programming* (50.6%, 39 out of 76; cf. Table 12), which is one of the courses we assume would be most influenced by the abilities and potential of GenAI. At the same time, it should be noted that the respondents teach various other crucial computing courses, a total of 193 classes

Table 11: Educator Survey – Institutional Characteristics (ES-24)

Institution	Percent Respondents
Secondary	11.84 %
2-year college (associates)	1.3 %
Vocational school	2.6 %
College (bachelor’s degree granting)	11.8 %
University (graduate degree granting)	68.4 %
Other (research institutes)	2.6 %
No institution given	1.3 %

as displayed in Table 12. Among the courses listed as “other” are the following: Data Science (5 responses), Data Base Systems (3 responses), Computing Education, and Web Technologies and Development (2 responses each). The diversity of courses also reflects upon the diversity of the responding educators.

Table 12: Educator Survey – Course Characteristics (ES-12); 193 responses from 76 educators

Primary area	Percent Responses
Algorithms and Complexity	7.3 %
Architecture and Organization	2.6 %
Artificial Intelligence/ML	5.2 %
Computational Science	2.6 %
CS 1 – Introduction to Programming	20.2 %
CS 2 – Introduction to Data Structures	10.4 %
Discrete Structures	1.6 %
Human-Computer Interaction	5.7 %
Information Assurance and Security	0.5 %
Graphics and Visualization	3.6 %
Information Management	2.1 %
Networking and Communications	1.6 %
Operating Systems	2.6 %
Parallel and Distributed Computing	1.6 %
Platform-based Development	0.5 %
Programming Languages	4.7 %
Robotics	0.5 %
Social Issues and Professional Practice	2.1 %
Software Development Fundamentals	5.7 %
Software Engineering	7.3 %
Systems Fundamentals	1.6 %
Teacher Preparation (age 5–18)	2.1 %
Other	8.3 %

We furthermore asked the educators how many years they have been teaching (ES-28). Most of the surveyed educators proved to have more than a decade of teaching experience, as the average number of years is 15.8, with a median of 14.5 years (cf. Table 13). The largest groups of educators have been teaching for 11 to 15 years (16 out of 76) and from 16 to 25 years (20 out of 76), so they can be described as experienced on average.

5.1.2 Developer Survey Demographics. Based on our selection process, $N = 39$ developers provided reasonably relevant responses to our survey. As mentioned, 29 of them provided answers to all questions (i. e., finished the survey). In this section, we present the results of survey questions DS-17–19 (in the developer survey) to characterize our sample.

Table 13: Educator Survey – Years of Teaching (ES-28)

Years	Number	Percent Respondents
0–2	7	9.2 %
3–5	9	11.8 %
6–10	14	18.4 %
11–15	16	21.1 %
16–25	21	27.6 %
> 25	8	10.5 %
no answer	1	1.3 %

18 developers provided the name of the country in which they are currently employed (DS-17). Most developers who participated are employed in the United States (66.7 %, 12 out of 18), or Germany (22.2 %, 4 out of 18). Others work in France (5.6 %, 1 out of 18), or the United Kingdom (5.6 %, 1 out of 18).

Moreover, we asked the developers about their job title (DS-18, $n=23$). Most developers identified their job title as *software developer* (70 %, 16 out of 23) followed by *research engineers* (13 %, 3 out of 23), (*scientific*) *researchers* (13 %, 3 out of 23), and a *software engineer* (4 %, 1 out of 23).

The companies of the developers (DS-19, $n=27$) can be characterized as summarized in Table 14. The largest percentage of developers (59.3 %, 16 out of 27) work in a large company with more than 500 employees. A smaller number of developers work at a research institute (11.1 %, 3 out of 27), a small or medium software company (7.4 % each, 2 out of 27), a non-software focused company (7.4 %, 2 out of 27), or for the government (3.7 %, 1 out of 27). One respondent (3.7 %) selected the option “other” without specifying the type of company any further. None of the participants are employed in a start-up or non-profit organization.

Table 14: Developer Survey – Company Characteristics

Corporation Type	Percent Resp.
Start-up (10 engineers or less)	0
Small Software Company (11–50 engineers)	7.4
Medium Software Company (51–500 engineers)	7.4
Large Software Company (more than 500 engineers)	59.3
Non-profit	0
Non-software focused company	7.4
Government	3.7
Research Institute	11.1
Other (not specified)	3.7

5.1.3 Interview Sample Demographics. The authors initially built a list of potential interviewees based on personal knowledge of who was actively teaching with AI, publishing research on its use in computing education classrooms, or building tools for use in computing education. We then augmented that list based on the results of the literature review, personal networking at the ITiCSE 2024 conference, and an email to a large mailing list. Next, we selected people from the list to interview based on what would provide a diverse set of experiences and ideas in terms of interviewee position, seniority, nationality, gender, and race. Not everyone agreed to be interviewed, so we made some substitutions with others on the list that were not initially selected. Finally, we also intentionally

sought out some interviews from people who are less likely to be selected, such as K-12 teachers, professors who teach computing outside of computer science, and thoughtful non-users who have intentionally decided to resist generative AI in all forms.

5.2 RQ1 – Policies

This research question was addressed by the survey for educators, as well as the interview study. Policies applicable for software developers were also gathered through the developer survey.

5.2.1 Educator Policies preventing, tolerating, or integrating GenAI. In the survey study, we asked educators about their classroom policies regarding the use of GenAI tools by students (ES-1, $n = 76$). The majority of faculty do not explicitly disallow students from using GenAI tools (59 out of 76, 77.6%). Roughly a third of faculty actively incorporate GenAI into their courses (ES-2; 27 out of 76, 35.5%). This is particularly interesting as the majority of faculty (57 out of 76, 75%) report believing the skills to create software have changed after the advent of GenAI (ES-3, see Section 5.8 for more details on how the competencies have changed). Thus, incorporating GenAI seems to be lagging behind the percentage of faculty who think that skills have changed.

Among faculty who are disallowing students from using GenAI, we learned from ES-10 (“Are you doing anything to prevent GenAI tools’ use in your course?”) that 58.8% (10 out of 17) are working to prevent their use.

Among those faculty trying to prevent the use of GenAI, they reported, in response to ES-11, (“What are you doing to prevent GenAI tools’ use in your course?”) on different actions they had taken. For example, 36.0% (9 out of 25) reported that, together with the students, they carry out code reviews in various forms on assignments and tests, to see whether the students have understood their own code or whether typical AI markers can be found. Another 28.0% (7 out of 25) of the responses describe changes they have made to assignments and exams. Another point mentioned several times is the appeal to the students’ social responsibility (20.0%; 5 out of 25). Technical measures and the introduction of workflows to be documented were rarely mentioned (8.0%; 2 each).

The following quote in response to ES-11 is an example of a detailed explanation of an educator regarding their attempts to prevent students from using GenAI tools in their courses:

“We tell the students it’s not allowed, and make sure they see that by putting a question on our course rules quiz. We, for now, have a few patterns we look for that are typical of AI code but which we don’t teach in our class that we automatically scan for as students submit assignments. When we see these we do a human check of whether we think the code was probably written by an AI or not. Last semester, in 13/14 cases, accused students admitted they used AI (the other person learned some extra concepts from alternate resources and was able to demonstrate understanding of them). For next semester, we’re also introducing a citation policy: If students use stuff they learned outside of our teaching resources, they are required to cite where they learned it. This is good practice in writing code anyway, but it also gives us an

extra way to distinguish AI coding from people using third-party resources legitimately.”

Based on responses to question ES-2 (“Are you incorporating GenAI tools (e. g., actively integrating it into the curriculum or exercises) into your recent courses (within the last 12 months)?”) we found that 35.5% (27 out of 76) of instructors are integrating GenAI into their courses. In the follow-up question ES-21, we asked those educators to elaborate on how their teaching approaches have changed in an open-question format. Two members of the working group performed a thematic analysis of the 60 open-ended responses. The following five themes summarize how educators changed their **teaching approaches**:

- (1) **Focus on different programming skills:** This theme captured responses from the educators who have refocused the skills that they teach. Some respondents refer to existing skills that are nonetheless now more important than before, such as reading code (8 respondents), modifying code (1), testing code (1), and problem decomposition (2). Others refer to entirely new skills that did not exist prior to the use of GenAI, such as writing prompts (4) and incorporating code from GenAI into larger projects (1). Still others refer to a lessening importance of existing skills; for example, that syntax is now less important than before and that the focus should be on a higher level of abstraction (3).
- (2) **Policy on when and how to use GenAI.** Respective respondents report setting expectations to students of tasks where they should use GenAI and when they should not (7 participants). In addition, if students are allowed to use these tools, these respondents require that students describe how they used the tools including clear attribution to GenAI (3), e. g., *“I actively encourage students to use GenAI tools and set expectations where they should be used and where they should not. I have created a policy that when in doubt students are allowed to use the tools but they have to clearly and fully describe the use.”*
- (3) **Change lecture to adapt to GenAI.** Participants report demonstrating working with GenAI in class (12), incorporating GenAI into in-class activities (2), using more active learning in class (2), and showing non-deterministic responses from the GenAI in class (1).
- (4) **Teach students to use GenAI effectively.** Some participants reported encouraging students to use GenAI (2) and others reported trying to teach students how to be more effective when working with GenAI beyond the demonstrations in class (2). The following quote illustrates: *“Focusing on teaching students prompt design techniques like tree of thoughts, etc. that utilize the underlying architecture of transformer models.”*
- (5) **Not yet / need to at some point.** Some participants responded that they had not changed in class activities yet but recognized they may need to in the future (8 participants). *“None, so far. I think their negative impact on a course is, at this point, less than other types of illicit support that students might seek”, “Not much at the moment, but I feel I need to do more here.”*

Yet another follow-up question for educators integrating GenAI asked how assessment approaches have changed as a result of GenAI tools (ES-22). Again, two members of the working group performed a thematic analysis of the 63 open-ended responses to

question ES-22 and any responses to question ES-21 that related to changes in assessments. We uncovered six themes representing **how assessment approaches have changed**:

- (1) **Process over product.** Six participants described assessing students on their process of creating software more than on the correctness of the final result. One participant is “... *shifting things to watching the process (ongoing use of repos, iteration, trackable code growth, etc.) vs final artifact evaluation*”. Another focuses on “*More incremental checkpoints and questioning the process before completion... the end work is less weighted and the process towards the solution is more emphasized.*”
- (2) **Invigilated exams.** Thirteen participants reported placing more emphasis on the proctoring of assessments. Six described more emphasis on proctoring, for example, “*programming assignments are now preparation for the proctored weekly quizzes.*” Three participants spoke of the importance of giving exams on paper. Seven spoke of adding or placing more emphasis on oral exams: “*Oral exams are now a vital part of student assessment.*”
- (3) **Increased weighting on exams.** Ten participants decreased the weight of unproctored assessments and increased the weight of proctored exams in assigning grades. One participant wrote, “*Minimized unsupervised assessment weight (e. g., assignments and projects) ... Increased supervised assessment weight (e. g., tests and exams).*” One participant described instituting “*a minimum cutoff for exams*”.
- (4) **No change (yet).** Nine participants said they had not made any changes with an additional six participants saying they had not made changes yet but had plans to in the future.
- (5) **Confuse the LLM.** Five participants chose to create assignments where the questions were designed to make the LLMs do poorly. One participant wrote, “*Homework assignments have been customized so that LLMs can't solve them well.*”
- (6) **Describe AI use.** Three participants are asking students to describe how they use GenAI on their assignments and provide chat logs with the LLM, as exemplified in the following quote: “*We require students to disclose if they used GenAI (as part of any outside help), and if they do, they are required to provide the prompts that they used to get their solution.*”

In addition, interview results revealed concerns of educators and researchers around AI policy. As GenAI tools are appearing more and more in the classroom, both institutions and individual instructors are adding policies around GenAI use. One concern repeated throughout the data was one of user privacy. There is the concern that students' private data will be leaked through the use of publicly available AI models or that it may be stored in a country that has different data privacy laws than the institution's host country. Some also expressed concern that these models will learn off of private course assignments and teaching materials. Then there are academic concerns, which primarily take three forms. First, some institutions will attempt to ban or limit their use. This was particularly apparent at the K-12 level where many schools and school districts have set policies in place to ban AI tools like ChatGPT entirely. Second, some individual instructors are adding policies to their courses either banning or limiting the use of generative AI tools. Third, some instructors are allowing or even encouraging generative AI usage, but requiring students

to use specific tools that will not reveal the answer or requiring that a student must be able to explain the code that they submit. Sometimes these tools are built by the university and housed locally, which also addresses the privacy concerns.

Thinking about privacy policies and its impact on instruction, one instructor said:

“I know at our university we are concerned about [privacy], and are trying to get certain tools with basically a university wide license for our students. I think we were trying to get a university-wide license for [GitHub] Copilot.”

Speaking on their own course policies, another instructor said:

“The bottom line is you need to demonstrate that you understand the material and anything that you turn in. It needs to be something that you know you've created and can fully understand what's going on”

One K-12 teacher said:

“One of the major changes that's gonna happen is that [anonymous state] has a whole bunch of new policy regulation guidance and support around generative AI and schools and teachers are expected to follow that. Now, teachers are going to have to know about those policy pieces, and we're going to have to figure out how to include that in the learning that they're doing on top of all of the other policy work that they're doing. In addition to just the classroom work, [teachers] now have to also be aware of all of this extra policy and regulation about it, too.”

5.2.2 Industry Policies. Eight developers stated that they do not use GenAI tools (cf. Subsection 5.7). Regarding industry policies around this non-use (subset of DS-14 and DS-15), one developer answered that “*it's currently not allowed to use GenAI tools but we expect it will be in the near future*”. Interestingly, none of the other replies concerned their company's policies.

5.3 RQ2 and RQ3 – Instruction-use and Instruction-support

The second and third research question pertain to teaching students how to use GenAI tools effectively, and how to support them doing so. For example, students may be taught explicitly how to use GenAI to help them write software. Of the educators who participated in our survey, 27 were actively incorporating GenAI into their courses. This section reports on these instructors and their replies, before reporting on the educators who were interviewed.

Educators incorporating GenAI teach a variety of courses (ES-12). Most of the surveyed instructors either taught introduction to programming courses (11 out of 27, 40.7%) or software engineering courses (9 out of 27, 33.3%). The remaining instructors (7 out of 27, 25.9%) taught more advanced courses such as databases, data visualization, and artificial intelligence.

Course sizes varied widely (ES-13), as shown in Figure 5—half the courses were 50 students or smaller (13 out of 27, 48.1%), while the other half were larger than 50 students (14 out of 27, 51.9%).

The large majority of instructors who incorporated GenAI into their courses reported that both instructors and students were

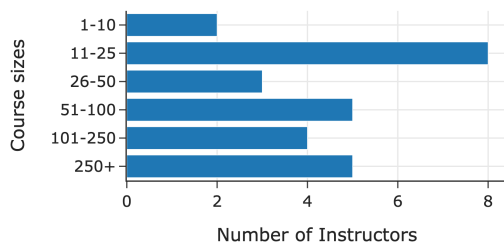


Figure 5: Instructors integrated GenAI into a wide variety of course sizes (ES-13).

expected to use GenAI for their courses (20 out of 27, 74.1%), as summarized in Table 15 (ES-14). However, students using the tools does not necessarily mean they are being trained in how to use them, as the instructor may have used the GenAI tools as a tutor.

Table 15: Most instructors who integrated GenAI expected that both instructors and students would use GenAI (ES-14).

Who uses GenAI	Number of Respondents
instructors only	3
students only	4
instructors and students	20

To explore classroom use of GenAI in more depth, we asked instructors how they expect students to access GenAI tools (ES-15). As shown in Figure 6, when students were expected to use GenAI tools, instructors incorporated a tool that was free to use, like a free publicly available tool (16 out of 27, 59.3%) or a paid tool that allowed students to use it for free (4 out of 27, 14.8%).

In response to a question asking which types of GenAI tools instructors incorporated (ES-16), all educators reported that they used standard industry GenAI tools like ChatGPT and Copilot (27 out of 27, 100%), while a few instructors also used customized tools that the instructors had created themselves (2 out of 27, 7.4%).

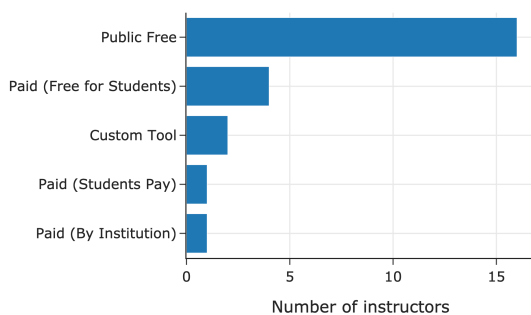


Figure 6: Instructors preferred to use tools that did not require students to pay (ES-15).

Question ES-17 provided additional insight, finding instructors primarily incorporated GenAI tools to teach students about using GenAI tools (20 out of 27, 74.1%) and as an educational content generator for teaching material (16 out of 27, 63.0%). Other educators

used GenAI for feedback, correcting student work, validating quality of assignments, and to support grading, as shown in Table 16.

Table 16: Ways in which instructors incorporated GenAI (ES-17).

Answer	Count
to teach students about using GenAI tools	20
as educational content generator for teaching material	16
to automatically provide feedback to students using a custom tool	5
to support the correction of student work	4
to validate the quality of assignments	3
to support grading	2

The interviews revealed that while many instructors are using generative AI to provide feedback to student submissions or increase access to student help-seeking, many are still not directly using it in class to support their teaching. What we did find still seems rather rudimentary and is one piece of this that we expect will continue to evolve. First, some instructors are using generative AI to help students understand bias in the models and planned in-class activities around it. Others will use it to generate code or proofs and then have students attempt to pick it apart, determining if it is correct or incorrect, and why. For example, one instructor detailed an in-class activity on generating code and then having students write tests as a first homework assignment. Finally, several educators mentioned using generative AI coding tools in upper-division courses or courses with heavy software development tasks as a way to prepare students for industry.

Regarding bias in the models, one instructor said:

“And then these models can perpetuate bias because they’re trained on data that’s biased. And we try to look at some examples of that. It does seem like students are a little bit surprised about some of these bias issues. I think it’s important to talk about how these tools can have bias in them.”

When talking about an upper division course that an instructor teaches, they said:

“So for upper division classes, where the point is coding as a professional would: Hell yeah, like we’ll [use it for] everything. It’ll be Gen AI from beginning to end.”

Regarding using it for large software development project courses, one instructor said:

“I’m about to teach a software development course that I’ve taught a lot before. Well, AI and generative AI is a big part of software development. So I’m going to be working on teaching generative AI as a tool, as part of the software development process as well. And so working with students on: okay, what can it do? How do we use it effectively and integrate it as a tool? Just like we have IDEs and debuggers and linters and everything else.”

In the interviews, the participants were predominantly mentioning using the GenAI tools to directly interface with the students

and help them with their work. A common theme was tweaking such tools as to provide meaningful learning experience.

“I added the functionality of having an instructor be able to specify keywords or concepts that we didn’t want to have show up in a response to be able to kind of better tailor the responses for the courses.”

Another theme was using the GenAI tools in specific contexts to support students while not interfering with their learning.

“So it is something we’ve built into an office hour queuing system. So we have a website students log into when they want to join an office hour. Just join the office hour queue and get TA support. And when the students type in what their question is, they’re gonna ask the TA here, or why they’re in office hours, then to our tool and it basically takes and adds some context to the question. And says this question is from a student with this level of knowledge, and who knows these languages and so on. And give them an answer, and don’t provide a complete working program.”

Finally, the participants often mention they would use the GenAI tools to create learning materials or brainstorm lesson plans.

“And so, we’ve used it, me and [CS teacher] have used it. And so is our other teacher a little bit. I think we’re probably the only ones, but it’s it’s exclusively to help us with our lesson planning. It’s phenomenal.”

5.4 RQ4-Motivations

To better understand the educators’ motivations behind their decision to prevent, tolerate or actively integrate GenAI tools, we asked three questions in the respective survey:

- ES-9 – Why don’t you allow GenAI tools in your courses?
- ES-18 – Why have you incorporated GenAI tools into your recent course?
- ES-20 – Why have you not incorporated GenAI tools (e. g., actively integrating it into the curriculum or exercises) into your recent courses?

The reasons for not allowing the use of GenAI are very diverse (ES-9): 39.3% (11 out of 28) of the statements are about didactic aspects and curricular approaches, with a common theme being that students should first acquire a solid foundation in programming. A further 21.4% (6 out of 28) name tasks or topics in which students can use GenAI and how they should document their use. 17.9% (5 out of 28) of the statements express an attitude towards the students: students would misuse GenAI, and they do not (yet) have the skills to use GenAI or to understand and interpret the results. It is also interesting to note some statements (10.7%; 3 out of 28) reflecting negativity towards GenAI in general; for example, that GenAI is not capable of solving the course assignments anyway. Another concern is that *“GenAI will almost certainly be criminalized (as stealing code)”*, or that only humans should write code, not an AI. Organizational and legal aspects were mentioned in only 7.1% (2 out of 28) of the statements.

The following quote from the survey summarizes a general view of these educators on the use of GenAI tools in their courses:

“Introductory CS students need to learn basic skills before they can move on. It is a similar situation to calculators in K-12, where usage is not generally permitted in the primary grades where basic skills and number sense are being taught.”

Educators also provided their motivation and rationale for incorporating GenAI tools into their recent courses (ES-18), which asked: “Why have you incorporated GenAI tools into your recent course?”. 24 open-ended responses were collected, and a thematic analysis by two of the authors revealed the following reasons:

- (1) **Preparation for industry and career readiness.** Some educators believe students will use GenAI tools in their career, so it is a matter of career-readiness to successfully use these tools: *“I believe that nearly everyone will use AI assistants like Copilot and ChatGPT to program in the future. I want to prepare my students for their careers and hence wish to train them in how to use those tools.”* Another related response is: *“In work life, students will use AI, too. So, we try to work under ‘normal’ work conditions in the course.”*
- (2) **Ethical use and responsibility.** Another reason educators mentioned was their responsibility as a teacher. Due to the sheer potential and assumed impact of GenAI, but also its bias and challenges, educators think it is responsible to teach about GenAI: *“Because to not do so would be irresponsible, and because GenAI engages students.”*. Another educator highlights respective concerns: *“I have observed the technology’s usage with disastrous learning outcomes, easily shown in a few iterations of two different courses where students did very well on homework that is code-oriented and were unable to do even the most basic things on the same topic in written or oral context.”*
- (3) **Efficiently supporting student learning.** In some cases, educators use LLMs to compensate for a lack of teaching resources, or to generate teaching and learning material. For example, *“For scalability reasons (my teaching team is too small for the large quantity of students we have). To support students in their moment of need and better support their learning.”* Several educators mentioned how they themselves use GenAI: *“It makes me faster at generating materials, like brainstorming the setup for exam questions.”*
- (4) **Adapting to recent technological trends.** Another motivating factor for educators to use GenAI tools was to educate themselves about recent technologies. For example, *“I wanted to get a personal sense of the capabilities of models, and to then showcase effective ways to use it in the class should students elect to use it.”* For courses with an AI focus, integrating GenAI was a given: *“I teach an introduction to artificial intelligence. Not using GenAI tools would be a great disservice to the students.”*

We were also interested in why educators do not incorporate GenAI into their courses (ES-20). We identified five areas of underlying motivations within 55 statements which we categorized and summarized.

- (1) **Lack of educator resources.** 24 of the statements were related to the educators. Educators cited a lack of time, lack of skills, or lack of didactic competency in particular.
- (2) **Students’ motivation and abilities.** Another important aspect mentioned by 10 educators was students’ perceived skills

and motivation. For example, educators assume that students already know and use GenAI, that they teach themselves how to use it anyway – but also that students lack the skills to use it properly.

- (3) **Doubts regarding usefulness.** Attitudes towards GenAI also appear to be a motivating factor in 8 responses. Educators are cautious about the technology and its benefits, especially in basic programming education.
- (4) **External factors.** Institutional factors such as a lack of support, unresolved legal issues or privacy concerns were also mentioned in 6 responses.
- (5) **Other reasons.** The remaining 7 statements concerned general teaching practice issues, such as not teaching a suitable course at the moment or explaining the use of GenAI to students but not integrating the tools further.

In the interviews, the participants often stressed the importance of designing and providing solutions that focus on education. A common concern was that general tools are focused on providing complete solutions as opposed to helping students learn.

“The tools are really good at giving complete working program functions. But what we want with it, with a student who has a code problem is some kind of hint to get them going and get them unstuck. So a lot of times the tools give too much help.”

Another motivation for solutions that are educationally-focused is centered around potential over-reliance of students on these always available assistants.

“they can become very reliant on just asking questions as soon as they get stuck. And I think the idea of being able to guide students much in the way that a human teaching assistant would. I mean, if you go and ask a TA for help. They would typically try and use a kind of Socratic method to maybe ask a question back at you to guide you to a solution rather than just giving you the answer.”

Another common theme often mentioned by participants was that GenAI tools often provide a significant help to the instructors, easing their work load and allowing them to focus their attention where it is needed the most.

“The idea was that basically I get another teaching assistant, another person in the room that can give the students tailored feedback. I do think there’s a small subset of students that actually prefer getting feedback from the AI. Another goal of mine has been to increase the breadth of students that are getting through introductory CS courses.”

A very important motivation for using GenAI in computing education often mentioned by the participants was providing immediate (timely) help to the students. This would otherwise not be possible in many contexts.

“in the age of AI for me philosophically, and technologically, this is strictly a net positive, because we now have the ability to provide students with far more real time and iterative feedback, not only for our undergraduates on campus at Harvard and Yale but open courseware

audience around the world, who never had access to TA’s office hour sections, and that human support structure, and they still don’t have a human support structure, but arguably an increasingly good approximation thereof.”

On a related note, another person said,

“[We] leverage GenAI to provide help at scale, which is one of the things we’re always trying to do and modify how we teach computer science to better help more people succeed. That’s always my goal. Help more people. My diversity of people succeed in computing and find interest in it, you know. Use it to do interesting things.”

Some participants explained their motivation along the lines of students needing to encounter GenAI tools during their studies because they will be using them in practice.

“I’m kind of pushing them hard to adopt the view that it’s almost academic malpractice to not teach people how to write with these things, because once you’re out in practice you’re going to have access to these and your competitor [...] will have access to these things. And so you need to know how to use them responsibly, how to best use them. How to check them for errors like, what types of things can go wrong with this, etc, etc.”

Among the participants studying GenAI in computing education, a common motivation was the need to understand how these are changing the classroom:

“I think one main point for me would be, how the introduction of GenAI into introductory programming courses are affecting students, learning and also affecting the concept that the students need to learn and then how the problem can be this become the solution itself, how we can utilize the affordances of GenAI. For helping students with learning what they need to learn.”

5.5 RQ5 – Tools

As mentioned in Section 5.3, all educators reported using standard industry GenAI tools like ChatGPT and Copilot (ES-16, 27 out of 27, 100%) via the online survey. Only two instructors also used customized tools that they had created themselves (2 out of 27, 7.4%).

Interviews revealed a varying range of tools from AI-powered assistants to specific applications designed to enhance learning experiences in programming and other courses. One notable category of tools is Retrieval-Augmented Generation (RAG), which not only provides information but does so within the context of a given course. Another category of tools tracked student interactions and completion of the course assessment elements, which helps educators understand how students are learning so that they can adapt course curriculum and delivery accordingly. These tools initially were used and deployed in computing courses but now are being increasingly adopted in other disciplines such as business and economics.

The usage and impact of these tools are considerable. GenAI tools are extensively used in writing problem sets, assisting with assignments, and providing explanations and help in coding tasks. There are tools specifically developed to facilitate functionalities

for asking questions, explaining code, writing code, and fixing code. A common trend is to put guardrails in place so these tools cannot directly return generated code. Many of these tools are built on top of existing AI models and are trained on specific course material.

“So it’s essentially like a sandbox version of ChatGPT. Now, that doesn’t mean, I think, that it’s completely safe. But I think that what it means is that the responses that students get from the model are going to be based only on the material that the instructor has trained the model on.”

An interesting observation by educators who deployed and used GenAI assistants in their courses is the noticeable reduction in the use of office hours by students following the deployment of these tools, indicating a shift in how students seek help and interact with course material.

In terms of design and feedback mechanisms, the tools are crafted to be simple, avoiding complex options for students.

“I think AI tools in CSed should be designed to ensure users remain cognitively engaged in their coding tasks, preventing situations where users can offload the necessary cognitive efforts onto the AI without thought, while also not causing frustration.”

The responses provided by these tools are also simplified, and students have the ability to offer feedback on the responses they receive. This feedback loop is critical for continuous improvement.

“If the student doesn’t know what the AI is doing and how it’s coming up with something, they cannot debug this help seeking tool.”

The use of GenAI in education extends beyond direct student support. GenAI tools are employed to analyze student data, including test case-driven auto-graders, knowledge component-based tutors, and adaptive feedback mechanisms.

5.6 RQ6 – Perceived Outcomes

In the interviews, instructors shared perceived outcomes of GenAI use that they consider to be both negative and positive. The negative perceptions mostly revolve around students using GenAI to write code for them. This has raised concern in multiple ways. First, there is concern that the students are able to circumvent work by having the AI do the work for them. Multiple instructors have reported a sizable increase in academic dishonesty violations in the last few semesters.

“The bar to cheating is lower because of these GenAI tools.”

Second, it is perceived that this use of GenAI is resulting in lack of actual learning. For example, a scenario shared by professors is that students will use GenAI to complete open assignments and therefore not actually understand the learning objectives. Professors are saying this scenario causes students to be unprepared for assessments.

“I know I don’t think we’ve ever had such large numbers of students who go into a test and simply can’t do anything.”

In some cases, it appears that students think they are learning by using the AI but they are mistaken. Even when using GenAI

in ways supported by instructors, it can be problematic for students. Instructors shared that it is very easy for students to be over-trusting of the AI-generated suggestions and waste their time going down incorrect paths. In some extreme cases, students have been discouraged from continuing their education due to the fear that AI will make their careers obsolete.

However, there are still many positive perceived outcomes shared by instructors. First, students who use GenAI are able to create work that is more complicated and of a higher quality than they would without assistance. This includes both coding projects and writing assignments. Used as a tutor, AI can help students by explaining difficult concepts, spotting logical errors, and even providing better compiler error messages than the actual compiler. Use of GenAI in this capacity has even been perceived to lower the load on office hours held by instructors and teaching assistants.

The interviews also revealed some larger structural changes reported by instructors. One common theme is the idea that the very concept of being a programmer will change. It is predicted that students will graduate with fewer fundamental coding skills, but with the ability to produce more advanced work. Some instructors predict a branching into two fields: computer scientists and conversational programmers. Conversational programmers will focus much less on writing code and much more on understanding how to write code and use AI tools to generate code for them. One professor went so far as to say that this shift in computing education is impactful on the same scale as the invention of the printing press.

5.7 RQ7 – Industry Usage

In this section, the perspectives of industry developers towards GenAI tools are presented. Moreover, we compare the developers’ views and their use of GenAI with educators’ beliefs of the industry’s use of GenAI.

The majority of surveyed developers (31 out of 39, 79.5 %) use GenAI tools in their professional role in developing software (DS-1). Eight developers claimed not to use GenAI. Two of these stated that they were concerned about ethical issues as their main reason not to use it (DS-14). The reasons *“My company does not let me”*, *“I do not believe they will help me code better”*, *“Haven’t gotten around to it!”* and *“I tried it and did not work for my needs”* were stated once each. Three developers did not provide a main reason why they do not use GenAI. Elaborated reasons were entered by three developers. One developer stated that they did not receive the “needed code”, maybe they were trying on “not so standard” problems (DS-15). Other reasons were mentioned in the open-ended response field: *“In theory AI generated code could be a derivative work”* and *“it’s not allowed to use GenAI tools but we expect it will be in the near future”* (cf. Subsection 5.2).

In the following, we present the results of the responses from developers who are users of GenAI (n=31). We asked developers about the frequency of their GenAI use in their professional role (DS-2). Most of the developers in our survey reported that they use GenAI at least once a day, with 52.2 % reporting using it several times a day and 4.3 % using it once a day (see Figure 7). A similar question (ES-7) was provided to educators, asking them to estimate how often they believe professional developers use GenAI in their

work (see Figure 7). Many educators in our survey expect professional developers to use GenAI at least routinely (36.4%). But the percentage of developers in our survey reporting to use GenAI every day is 56.5% (sum of responses to “once a day” and “several times a day”). This indicates that educators’ expectations seem to underestimate the use of GenAI tools by professional developers.

Asked to select the types of tools being used (DS-3), developers reported that autocompletion tools are the most frequently used type (52%, 16 out of 31), followed by Chatbots (48%, 15 out of 31). Nine developers (29%) did not provide any answer to question DS-3.

We asked developers to identify the tasks for which they use GenAI from a preset list of programming tasks (DS-5). Eight developers did not answer this question. On average the developers selected 4.4 tasks (median 4). The most often selected tasks were *generating code* and *autocompleting code* (both selected by 17 developers, 70.8%). Furthermore, the tasks *modifying code* and *creating documentation/comments* were selected by the majority. The least frequently selected task is *modeling algorithms*, i. e., one developer selected this option. Only one other task, “research”, was mentioned for *Other, please specify* by one developer.

Just prior to the closed-ended DS-5, we provided an open-ended question asking the developers to describe how they use GenAI tools in their professional work (DS-4). It was intentionally placed there to avoid influencing the developers by the options offered in DS-5. The open-ended question was answered by 17 developers mostly on a high abstraction level and was coded using the existing options of DS-5 as a starting point. In general, the explicitly named tasks match the selected tasks in DS-5, but most developers selected more tasks in response to DS-5 than they had selected in DS-4. On average, 2.6 tasks were tagged in the open-ended answers to DS-4 and 4.4 tasks were selected in the closed-ended question DS-5.

In the following, some interesting examples of the open-ended responses are provided: Two developers explicitly stated that they use GenAI for repetitive tasks, or to “*automate boring stuff (i. e., code that no human should ever write)*”. Three developers stated they use GenAI for generating code or examples for languages they are not comfortable or less familiar with. A common theme was to use GenAI to avoid reading documentation (mentioned 4 times) or to “*get help with undocumented features*” (mentioned once). For the task *modifying code*, refactoring or cleaning code were mentioned explicitly by two developers. Interestingly, one developer stated they use GenAI to get “*code reviewed to make it more clear, concise and maintainable*” but did not select *modifying code* and one developer claimed to use GenAI to “*explain errors or exceptions for frameworks that I rarely use*” but did not select “Debugging.” Further notable examples that did not fit into the existing categories were “*Generating workshop material,*” helping “*with unusual tasks (e. g., I recently had to convert timestamps stored in a funny text format into a number of minutes ... in Excel)*”, “*summarizing large text*”, and “*writing quality updates*”. The latter two seem to be focused on prose text instead of program code.

We asked educators to report their perceptions of which tasks developers use GenAI tools for (ES-8). 74 educators answered this question. Educators, on average, selected 6.3 tasks that they assume developers use GenAI for (median 6). Most frequently, the educators selected *generating code* (79.2%), followed by *autocompleting code* (75.3%), and least frequently *modeling algorithms* (20.8%). Only the

tasks *modeling algorithms*, *modifying code*, *generating ideas*, and *debugging* were not selected by the majority of the educators. Five educators selected *Other, please specify*. The following statements by educators should be noted: “*Most of the industry [sic!] people I know are just experimenting, not using it for official tasks*”, “*all of the above*”, “*exploring alternatives, explaining code, explaining downsides/upside of approaches*”, and “*developers I’ve talked use ai differently depending on their context*.” The comparison of the results from educators and developers is summarized in Figure 8.

The frequencies of educators estimating developers’ GenAI usage and actual developers’ GenAI usage seem to approximately match for the tasks *generating code*, *autocompleting code*, and *providing code examples*. There is a particularly higher percentage of educators who identified *getting started with a problem*, *modeling algorithms*, *generating ideas*, *generating test cases*, and *finding resources/documentation/libraries*. Interestingly, more developers selected *modifying code* than educators. In all other cases, the expectation of the educators were higher.

Figure 9 shows the ratings of the usefulness of GenAI as reported by the developers (n=21, DS-6). Almost all developers who answered (95%, 20 out of 21) find using GenAI at least a little useful. Six developers out of 21 (29%) find it very useful and only one developer did not find it useful.

At the same time, the majority of developers (81%, 17 out of 21) who answered the subsequent question (DS-7) think GenAI makes their work more efficient. Three developers (14%) saw no change, and only one developer found GenAI to make their work much less efficient.

When asked about the harmfulness of GenAI (n=21, DS-8), the majority of our surveyed developers found it *not harmful* (62%, 13 out of 21), one third *a little harmful* (33%, 7 out of 21), and one developer *moderately harmful* (5%). Six of the developers who found GenAI (a little) harmful elaborated on the harmfulness (DS-9). Two themes emerged from the thematic analysis. The first one was **wrong results or bugs in the code** (even for simple cases, n=5) and the second represents **concerns regarding code quality** (which is reported to be “*generally worse than humans*”, n=1). To present more specific examples, one developer reported on an issue with an invalid SQL statement and another developer said that GenAI invented features. A developer summarized that every line of code needs to be rechecked – even though errors may only occur in rare cases.

5.8 RQ8 – Competencies

5.8.1 Educators’ Perspectives. The majority of educators (57 out of 76) who participated in the survey believe the skills needed to create software have changed after the advent of GenAI (ES-3). Among educators who have changed their courses to integrate GenAI, 25 out of 27 believe the skills to create software have changed. In contrast, among educators who have not changed their courses to integrate AI, 32 out of 49 believe the skills have changed.

Examining educators’ opinions based on whether they believe the skills to create software have changed, we found that among faculty who do not believe the skills have changed, only 2 out of 19 have changed their courses to incorporate GenAI. However,

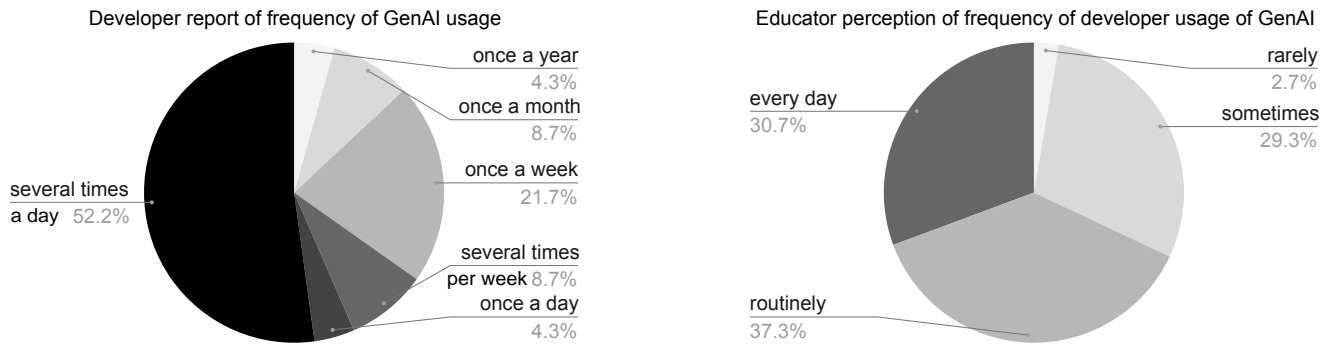


Figure 7: Left: Frequencies of developers (n=23) reporting on their use of GenAI tools as part of their professional role. Right: Perspective of educators (n=76) about the frequent use of GenAI tools by professional developers.

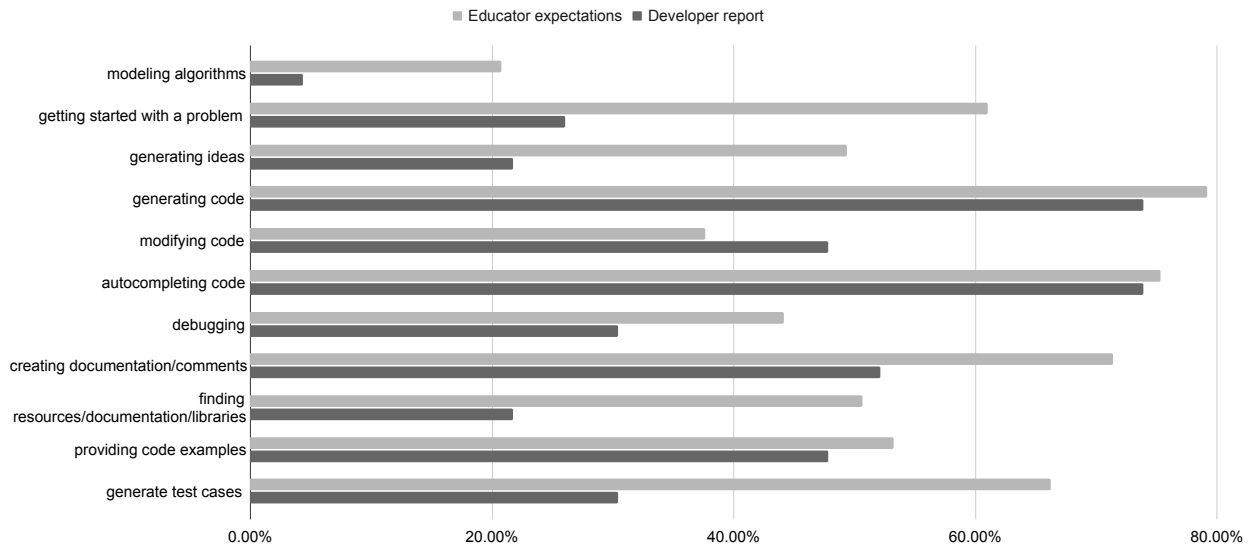


Figure 8: Comparison of tasks developers use GenAI for and educators think developers use it for.

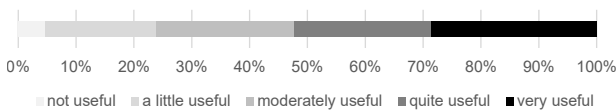


Figure 9: Ratings of the usefulness of GenAI as reported by the developers (N=21)

among faculty who believe the skills have changed, 25 out of 57 have changed their courses to incorporate GenAI.

Question 4 of the educator survey (ES-4) was given to faculty who responded “No” to whether the skills needed to write software have changed. The question was “Please elaborate on your last response why skills have not changed. (open-ended)”. Two members of the working group performed a thematic analysis of the 16 open

ended responses on **why skills have not changed** uncovering the following themes:

- **Same Skills/Shifting Importance.** The largest group of respondents, seven, believe that the skills are the same. However, the importance of these skills has changed. For example, multiple participants reported believing that understanding code is still crucial, but code evaluation has become more important. “Software skills remain a mixture of design, implementation, static analysis, and debugging. Changing the balances does not change the core skills involved.” Their responses seem to indicate that evaluating code generated by AI is not different than the code evaluated in the past.
- **Accelerating skilled programmers.** Two respondents believe that GenAI will only serve to speed up programmers who are already skilled on their own. “GenAI is only a tool which can

speed up the development process for those already skilled and knowledgeable.”

- **Know everything.** One participant believe that programmers still need to know everything at all levels below them regardless of GenAI: *“Using gen ai is a lot like being a team lead. The team leads and project managers still need the skills of the programmers under them”*. However, it was unclear how deep they felt these skills should go.
- **Don’t know yet.** One participant believe that GenAI is changing too rapidly for us to know if or how these skills will change. *“Too little time to know (from my limited point of view)”*.

Question 5 on the educator survey was given to participants who replied “Yes” to the question of whether the skills to write software have changed with the advent of GenAI (ES-5). Question 5 asked: “In what ways do you think the skills needed to create software have changed with the introduction of GenAI tools?”. Two members of the working group performed open-coding on the 50 open-ended responses followed by a thematic analysis to uncover the following themes related to **why educators think skills have changed**:

- **Less writing, more reading code.** The most common responses from participants were that reading and understanding code is more important (21 participants): *“It’s becoming more important to be able to read code that you have not written”* and writing syntactically correct code from scratch is less important (11 participants), *“There is less need to write code.”*
- **Higher-level skills are needed.** Fifteen participants said that writing software with GenAI makes higher level skills more important. Examples of higher level skills included creativity, problem understanding, problem specification, problem decomposition, software design, and software architecture. *“The focus needs to be more on the algorithmic thinking and problem solving, not the actual syntax/coding”*, and *“Problem decomposition becomes essential in a way it wasn’t in the past.”*
- **Evaluating and fixing code.** Eleven participants reported that evaluating code (testing) from the GenAI is an important skill and nine participants said debugging the code from GenAI is important as well: *“Code testing and debugging, if not previously emphasized, should be now.”*
- **Prompting as a new skill.** Seven participants reported that being able to prompt the GenAI to receive desired code is a new and important skill: *“There are new skills like prompting AI.”*

The educators who believe skills have changed (ES-6) chose *problem understanding* as the most important skill for programming with GenAI. They see *reading code* as the second most important competency and *problem decomposition* as the third most important.

Only three people who allowed GenAI in their classes changed any of the learning objectives of their recent courses (ES-19); six people reported not changing them at all.

5.8.2 Developers’ Perspectives. In question 10 of the developer survey (DS-10), we asked the developers whether the competencies required to professionally develop software are changing due to the availability of GenAI tools. 21 developers responded: 12 of them (57%) indicated a *slight change*, 4 of them a *moderate change*, and 5 of them *no change*.

The 16 participants who recognized a change in the competencies in DS-10 were of particular interest for the follow-up questions DS-11 and DS-12. We received 11 open-ended responses to DS-11, which asked for the **new relevant competencies to professionally develop software** with GenAI tools. Within these responses, we identified several themes. It should be noted that some of these themes had occurred within the responses of educators.

- **Same skills/Shifting importance.** Not necessarily new competencies, but shifting focus on understanding of concepts, e. g., *“there’s a bigger need to understand underlying concepts”*
- **Accelerating skilled programmers.** Successfully use GenAI tools to increase productivity, at least for simple tasks, e. g., *“leverage GenAI tools to improve productivity”*, *“using the speed for solving simple/generic problems.”*
- **Knowing about general limitations.** Know the strengths and weaknesses of GenAI tools, e. g., *“accounting for hallucinations from GenAI.”*
- **Evaluating GenAI use and its output.** Evaluate, if it is appropriate to (not) use GenAI tools for a specific problem, e. g., *“knowing the sorts of problems that GenAI can be useful to solve is important”*, *“Being able to determine if autocompletions are actually useful.”*
- **Prompting as a new skill.** Prompting GenAI to generate useful and appropriate responses, e. g., *“prompting AI with appropriate context to receive relevant/valid responses.”*
- **Meticulousness.** Paying attention to detail (being meticulous) when checking generated code, e. g., *“I need to carefully examine the output of Copilot [...] and I often have to make small edits to its output.”*

The second competency-related follow-up question for developers (DS-12, n=9) asked for **competencies that were perceived as no longer or less relevant** to professionally develop software with GenAI tools. We identified the following themes in the responses:

- **Know everything.** Know syntax and other elements of programming languages by heart, e. g., *“Less need for specific knowledge; eg language specific nuances.”*
- **Know and use other sources.** Using other external resources, e. g., *“searching stack overflow or other non-official help or documentation (official docs still has some usage).”*
- **Tolerating high levels of frustration.** Being purpose-driven or persistent despite frustration, e. g., *“The competence not to be easily frustrated if the first hit on Google does not provide the answer to the problem.”*
- **None.** None of the previous acquired competencies are perceived as irrelevant e. g., *“I see no competencies that are less relevant”*, *“It’s a good starting point, nothing else.”*

Interestingly, when asked about what advice developers would give to novice programmers regarding the use of GenAI tools (DS-13 and DS-16, n=18), many responses related to the responses to DS-11. The analysis of the open-ended responses revealed the following themes, which repeat some of the competencies mentioned before:

- **Evaluating GenAI use and its output.** Evaluate, if it is appropriate to (not) use GenAI tools for a specific problem, e. g., *“Find the right use cases for GenAI but don’t trust it. Verification is crucial.”*, *“Don’t simply trust them.”*

- **Meticulousness.** Paying attention to detail (being meticulous), e. g., *“always read back over the code the ai recommends: sometimes it is not entirely correct”*
- **Prompting as a new skill.** Prompting GenAI to generate useful and appropriate responses, e. g., *“learn how to speak AIs language but also learn enough domain knowledge that lets you adequately explain your needs to an AI.”*

The recommendations by developers are somewhat divided between using GenAI tools to learn and get comfortable with them – and not using them:

- **Resist using GenAI.** Do not use GenAI tools as a novice programmer, e. g., *“Don’t.”*, *“Please take the time to learn to be an algorithmic thinker yourself. Try the problem on your own”*, *“Learn the basics of programming without GenAI.”*
- **Rarely use GenAI.** Use GenAI, but rarely, e. g., *“learn to work without them as much as you can”*, *“There’s value in not needing it”*, *“Don’t lean on it too heavily.”*
- **Critically use GenAI.** Use GenAI tools, but always be critical of the output *“Use it to start out, but still need to read over the code that you write to understand it.”*, *“Use it to learn languages (explain what existing code is doing) but know that it’s nowhere near perfect.”*
- **Excel at using GenAI.** Use GenAI tools, and become more productive, e. g., *“Start using it early and get comfortable with leveraging it as a way to code faster.”*

It should be noted though that the last comment was the exception. The majority of the comments highlighted the need for a critical use of GenAI tools, and many comments were in favor of learning the basics of programming without GenAI to avoid over-reliance early on respective tools.

The interviews further reveal that educators are grappling with how GenAI tools are reshaping students’ competencies in computing education. One of the most commonly discussed shifts is the reduced emphasis on writing code from scratch and the growing importance of reading and understanding pre-existing code (or code written by GenAI tools). Educators believe that as GenAI tools provide students with ready-made solutions, the ability to critically evaluate and modify this code becomes more essential. Furthermore, higher-level cognitive skills, such as problem decomposition, task specification, and computational thinking, are emphasized as essential competencies that students must develop to navigate the evolving landscape of AI-assisted programming. Several interviewees also highlight the necessity of integrating communication skills, such as explaining one’s code, into course objectives to ensure that students genuinely understand what they create rather than merely generating functional code.

“I think potentially, students will start knowing less about how computers work. Right? As these tools evolve, we’re potentially having students that come in with less knowledge at that level, but more capability at higher levels.”

“I think with these tools, people are going to spend more time reading because they’re gonna get help. They’re gonna get assistance from these tools. It’s gonna give them code. People are gonna need to read it and understand it.”

Moreover, prompt engineering is emerging as a new core competency. Educators are increasingly aware that these models can do much of the “heavy lifting” in coding, but students must learn how to ask the right questions and interpret AI-generated outputs to succeed. Some express concerns that students might bypass key stages of the problem-solving process, such as task decomposition, which could lead to gaps in their understanding. While many educators see these shifts as an opportunity to teach more advanced and meta-cognitive skills, there is also a concern that students may lose touch with foundational computing concepts. Ultimately, the consensus is that GenAI is pushing educators to rethink their course objectives and better prepare students for a world where interacting with AI is a critical part of the programming process.

“At the end of the day, we want students to practice task decomposition. So how should AI be used? Hopefully, it should help them develop those skills.”

5.9 RQ9 – Equity

As a part of the survey study, educators were asked whether they teach at an institution that serves a minority population in their country (ES-26). 23.7 % (18 out of 76) of educators report teaching at an institution serving a minority population. However, 30.3 % (23 out of 76) are not sure whether it applies to their institution (one participant did not answer this question). Due to these relatively low response rates and insecurity among the participants, we did not analyze the data any further with regard to GenAI teaching practices and instructional approaches and whether and how they differ depending on the educators’ institutions.

In the interviews, educators raised significant concerns regarding the equity implications of GenAI tools in education. Many pointed out that access to these tools is often tied to financial resources, both at the individual and institutional levels. For instance, students from wealthier backgrounds or universities may have access to premium tools such as Microsoft Copilot, while others may have limited or no access. This disparity could exacerbate existing inequalities, as students with better tools may accelerate their learning, leaving behind those with fewer resources.

“ChatGPT-like technologies are not a silver bullet to help the poor students get better.”

The issue extends beyond financial access; educators also highlighted that not all students possess the meta-cognitive skills required to use these tools effectively. Those who are more capable of self-regulation may benefit from AI tools, while students lacking these skills might misuse them, producing correct outputs without learning the underlying processes.

“[N]ot all students are equipped with the Metacognitive skills to use these sorts of unconstrained tools productively.”

Additionally, linguistic and cognitive barriers were flagged as potential equity concerns. Non-native English speakers and students with lower literacy skills might struggle with AI-generated responses, which are often complex to read and comprehend. As such, AI tools could exacerbate difficulties in comprehension and critical thinking.

“For students whose first language isn’t English, they may struggle more with interpreting the AI-generated responses, which could create another layer of disadvantage.”

Furthermore, students with disabilities, particularly those relying on assistive technologies like text-to-speech tools, may face challenges in interacting with GenAI.

“we’ll have students with various disabilities in classes who will be watching their peers use it in some senior-level software engineering or capstone course, and they’ll be charging forward much, much faster to do this. But like, we’ll have text-to-speech issues for people who have motor impairments who can’t type where they’re doing speech to text to do text entry for some large language model, or they have a speech impediment. And so the speech recognition for the model doesn’t work very well, and then they can’t actually just get precise enough prompts into the system.”

While some educators noted efforts to mitigate these issues, such as providing institutional licenses or designing tools for broader access, the overall sentiment was that AI tools, at present, risk amplifying existing inequities rather than alleviating them.

5.10 RQ10 – Future

The future of GenAI in education, particularly in computer science, is seen as both an opportunity and a challenge. Interviewees generally agree that AI will continue to reshape the landscape, but the extent of its influence remains uncertain.

“I think if we really were honest with ourselves about new learning outcomes and wanting students to demonstrate proficiency. Working with these models. It would be good if we actually had some assessments where they were allowed to do that.”

Many educators foresee GenAI becoming an integral part of the classroom, assisting in various ways, from automating tedious tasks such as grading and creating assessments, to dynamically analyzing student code and providing real-time feedback. Some interviewees suggested that AI could cluster student submissions to provide more tailored feedback, saving time for instructors while offering more personalized learning experiences.

“One interesting project... is to make a conversational agent that takes students’ programs as input and then ensures they have learned what they need to learn.”

However, the consensus is that while these advancements have the potential to greatly improve education, careful planning and continuous adaptation are necessary to ensure that AI is used effectively and equitably.

Several interviewees also envision a future where learning outcomes and assessments shift to accommodate AI’s growing role. Instead of focusing purely on coding, students might be evaluated on how well they interact with AI tools to solve complex problems, emphasizing prompt engineering and problem decomposition. The challenge lies in balancing AI’s capabilities with the need to preserve essential skills, such as critical thinking and deep learning,

which are at risk if students rely too heavily on AI. Many interviewees expressed hope that AI would free up educators to focus on higher-level teaching tasks, fostering stronger teacher-student relationships and creating more meaningful, engaging learning environments. At the same time, they acknowledged that this future is still unfolding, and much of the educational community is cautiously experimenting with how best to integrate these new tools.

6 Discussion

In this section, we discuss some of the most interesting findings of the present study in an integrated manner. The discussion is thus led by themes and findings, and not by the research methods used to gather them.

Changing Competencies. Several interesting insights emerged from the educator survey. The majority of faculty (77 %) believe the skills to write software have changed because of the advent of LLMs, but less than half (36 %) have changed their courses to include GenAI. Perhaps those who feel the skills have changed, but have not changed their courses to include GenAI, have changed their courses in other ways. Regardless, an increasing number of instructors are adopting (or at least not banning) GenAI in their courses compared to prior studies [49, 111]. In the present study, we see about 30 % of educators actively integrating GenAI, and 75 % acknowledge it in their class. It is not a total change but a clear trend.

We also found that educators who are changing their courses to incorporate LLMs are focusing on the new competencies, such as prompting (ES-5, see Subsection 5.8). Based on the results of the literature review, there are differences in how well generative AI can support different tasks in computing courses. For example, studies that used it for code comprehension reported more positive findings compared to studies that used it for hint generation or code writing.

Lastly, recognizing that LLMs can solve take-home assignments, faculty are allocating more course grades on proctored assessments and creating more, or new, proctored assignments.

In the educator survey, educators are divided about how competencies are changing. Most educators report that they believe the skills to write software are either changing or shifting in some way as a result of LLMs and GenAI. In contrast, a minority of educators report believing either that there has been no change to competencies, or that students still need to develop certain programming competencies and skills before they should start to work with/use GenAI tools. Educators who see a shift in competencies feel we are shifting away from code writing to code reading; toward higher level skills like creativity, problem understanding, problem specification, problem decomposition, and software architecture; and toward evaluating/testing and debugging.

Reasons for Disallowing GenAI. Educators disallowing the use of GenAI provide two main reasons, (1) their own (at least perceived) lack of skills and competencies regarding the concept and use of GenAI, and (2) their attitude that the students would not have the necessary skills and should, therefore, learn the basics of programming first without any support of GenAI. Further research is needed to determine whether this attitude is pedagogically sound

or whether using GenAI can make it easier for beginners to learn programming – and if so, in what ways it may even enhance learning (these ways may relate to the shift in competencies discussed earlier).

Gap between Educators Expectations and Developers Actual Use of GenAI. We further noted some differences between the use of GenAI tools among developers and how educators think developers are using GenAI tools. More educators expected developers to use GenAI for certain tasks than developers reported. This applied to the following tasks: Modeling algorithms, getting started with problem solving, generating ideas, debugging, creating documentation/comments, finding resources/documentation/libraries, providing code examples, and generation of test cases (cf. Figure 8). Modifying code was the only task where developers exceeded educators’ expectations of GenAI use. It is thus important for computing education to regularly connect to industry practices and to not lose track of the competencies needed in the workplace. Despite these differences regarding expectations and actual use, the survey results alone show that GenAI tools are indeed used by developers for a great number of different tasks.

Developers’ Use of and Trust in GenAI Tools. The survey with professional software developers further revealed that GenAI is used for getting started, debugging, code cleanup, and not having to read documentation (cf. Subsection 5.7, question DS-4).

Even though only 39 professional developers took part in our survey, the results are in line with larger surveys, e. g., the StackOverflow 2024 survey with professional developers and their AI usage (see [105], conducted in May 2024, filtered for professional developers). It should be noted that the number of responses greatly varied among questions which is why we also provide the absolute numbers of responses in the following. In the StackOverflow survey, about 63 % of about 46,000 professional developers state that they use GenAI (vs. 79.5 % in our survey) and about 23 % state that they do not use GenAI and do not plan to do so in the future. About 83 % of about 28,500 developers see an increase in productivity and 58 % a greater efficiency. This number matches the 81 % of developers in our survey who see an increase of efficiency when using GenAI tools.

The StackOverflow survey further shows that developers are split about whether they trust AI output: about 41 % have (some) trust in AI and about 31 % have (some) distrust in the accuracy. Furthermore, about 45 % of the developers in the StackOverflow study think AI tools are bad or very bad on complex tasks. Both aspects are also reflected in comments in our survey. Hence, programming skills are still required to spot such issues and creating complex software. The top ethical concerns that the majority of developers stated in the StackOverflow survey are “misinformation and disinformation in AI results” (79 %) and issues with “source attribution” (65 %). This may contradict our result that about 62 % of the developers in our study found GenAI *not harmful*, but may also be caused by the way the ethical concerns were collected in the StackOverflow survey (i. e., via a closed question).

Equity Concerns. With standard industry tools being used by all educators who participated in our survey, we are concerned about how these tools may negatively impact equity among students. For example, we cannot neglect the cost of using GenAI tools. Considering a standard three- or four-year program (undergraduate

degree in Europe or the US) and a monthly cost of 20 US dollars per month, this may amount to about one-thousand US dollars. Access alone may thus be another financial burden for students with an already tight budget.

According to the educator survey (ES-15), educators seem to expect students to use the free version of standard industry tools. However, we did not ask them whether they have any knowledge of students using the paid versions regardless, and how that may have had an impact on their outcomes or exam results. When being used for take-home tasks or assignments, educators have no control whatsoever on students’ tool use. Even in the classroom, some students will have the resources to access GenAI easily via their smartphone, while others will not. For this reason, we recommend educators to carefully consider how and for which tasks they incorporate GenAI tools from an equity lens to not disadvantage any of their students. In related work, it has been shown that students with various prior knowledge and education use GenAI tools differently in introductory programming classes, and they have varying success rates [69]. Similarly, we are only beginning to understand the persisting inequities between GenAI technologies and the accessibility needs of people with disabilities [4].

Need for Educators’ Professional Development. An important finding of the educator survey is that many educators have not yet considered the potential risks of GenAI in depth or the benefits that GenAI tools can bring to their courses or students. On the one hand, educators stated that they have not yet familiarized themselves with GenAI, usually due to a lack of time, and therefore do not incorporate or allow GenAI in their courses. On the other hand, they seemed to have fixed attitudes about whether GenAI is beneficial to students or not. If we assume that they have hardly worked with GenAI, this attitude is based less on facts than opinions. This lack of familiarity with GenAI tools may also help explain why some of the responding educators had no clear ideas about using GenAI in their courses in a pedagogically meaningful way. The following quote in response to survey question ES-20 represents this challenge:

“I haven’t yet investigated the Generative AI tools. I’m observant and beginning to research and use the tools currently myself first to try to understand the realm of what can be done and how it can be integrated. I want to see how to scale the integration of the tools. I want to preserve the integrity of learning but still keep new and developing technologies viable because they are now in society, but I need examples for the classroom.”

Another consequence of not having explored GenAI tools is that educators may not be fully aware of the risks and potential harm, meaning they are unable to discuss it critically with their students. These barriers to adoption should be addressed more intensively by professional development courses and pedagogical training opportunities at all levels of computing education.

Future Developments. As we consider the future, we also speculate on potential use cases for generative AI (GenAI) tools. Currently, most GenAI tools primarily rely on text-based interactions. However, current GenAI tools can already generate additional modalities, such as images, audio, and video. A future tool might,

for example, be able to analyze a plot created by a student and identify if certain data points have been omitted.

We also expect that GenAI tools will become more personalized. A GenAI tool tailored to a specific course could recognize that students are developing new skills as they progress and adjust its generated results accordingly, avoiding advanced language features until they have been introduced by the instructor. Additionally, a GenAI tool personalized for an individual student might identify specific content areas where the student needs further assistance and direct them to relevant course materials to address any misconceptions, or even generate such personalized materials on the fly, targeted towards the individual student's educational needs (see also [63, 88, 124]).

7 Threats to Validity

The present work has some limitations that need to be considered. In the following, we outline the threats to validity of the respective methodologies, the data and its analysis and interpretation.

7.1 Systematic Literature Review

Regarding the systematic literature review, it should be noted that we did not check every available database in the context of computing education. However, we are confident that the applied methodology suffices in representing the state-of-the-art literature. Due to the cut-off date, some publications appearing around that date may also be missing, as arXiv, for example, can have delays in the process due to its moderation system. Another limitation related to the arXiv search is that it only allowed the title and abstract to be used for the search.

Related to the analysis on the nature of findings, we acknowledge that there is likely publication bias, which should be taken into account. Studies with positive findings could more likely be accepted for publication (or written up by the authors in the first place), which could inflate the number of included studies that reported positive results.

7.2 Educator and Developer Surveys

We invited the participants of both surveys through several mailing lists to draw a broad sample of educators and computing professionals. However, it is possible that some educators on these mailing lists are more involved in CS education than their peers and, given the acute impact of GenAI over the past few year(s), may have spent more time thinking through the implications of GenAI. To mitigate this, we also encouraged recipients of our emails to share the survey with others. Moreover, we noted that many potential respondents dropped out of the survey after seeing the consent form. It might have had a discouraging effect as participants had to sign a form to proceed.

Regarding the developer survey, it is possible that the respondents are not typical software developers, as they could be more interested in developments in academia and the development of tools. Moreover, we recognize that we are computing education researchers who do not have access to developers within large tech companies worldwide, which explains the low number of (full) responses. To mitigate this issue, we correlated our results with other research initiated by industry.

A limitation that applies to both surveys is that of self-reporting, meaning participants may have exaggerated, omitted information, or expressed thoughts they believe are socially desired. To somewhat mitigate this limitation, we included both open-ended and closed response options, and triangulated the survey data with the data gathered via the interviews with educators.

7.3 Interview Study

Regarding the interviews with educators, it should be noted that the sample may not be indicative of the computing education community. For example, interviewees were recruited in English, and the interviews were also conducted in English only.

Another limitation is, again, related to self-reporting, which may involve subjective representations, exaggerations, omissions, or socially desirable statements.

8 Conclusions

The present working group report aimed to address two overarching goals: (1) identifying how and why instructors incorporate GenAI tools into their teaching, and (2) outline how the competencies and skills in software developments changed and will change further in the future due to GenAI. To address these goals, we applied a mixed-methods design comprising a systematic literature review (SLR), a survey for computing educators and developers in the software industry, and semi-structured interviews with computing educators.

8.1 Systematic Literature Review

The SLR focused on the reported evidence of GenAI in CER, more specifically on the types of class interventions used and whether findings have been positive or not. We then chose the search strings and databases (ASEE PEER, arXiv, Scopus, ACM Digital Library, and IEEE Xplore). After filtering of the papers based on inclusion and exclusion criteria, the main characteristics of the papers was extracted. This led to the final set of 71 papers that were included in the literature review. The findings of the literature review suggest that thus far, generative AI has mostly been studied in unsupervised conditions, such as having students use it for homework. Generative AI has mostly been used for writing code, code comprehension, automatic hints and generating learning resources. Most commonly, students were not instructed on how to use generative AI, and were directed to use general tools such as ChatGPT. However, we did find that there are many custom tools available. These often include some sort of pedagogical guardrails aimed to make the use of generative AI more productive for learning.

Related to the nature of the findings, we found that studies that used custom tools that included some instructor scaffolding (e. g., pedagogical guardrails) more often reported positive findings. However, this was only the case when students were not given explicit instructions or guidance on how to use generative AI. Thus, based on the findings of the literature review, it can be recommended that instructors should either guide students on how to use generative AI if the tools used are general purpose (such as ChatGPT or Gemini), or alternatively use custom tools that include scaffolding (such as guardrails that, e. g., guide the model to provide more educationally appropriate responses).

8.2 Educator and Developer Views

We designed a survey for educators and software developers to capture perspectives on competencies and skills required for future graduates of computing. Precisely, we wanted to gather the educator’s perspective on using and teaching the use of GenAI tools to their students, which tools are used, policies that need to be considered or developed, and their motivations to integrate GenAI (RQ1-RQ6). Moreover, we wanted to identify the impact of GenAI on students’ competencies (RQ8), equitable conditions for learning (RQ9), and future perspectives (RQ10).

Another goal was gathering the industry perspective (RQ7) of generative AI usage, so we developed a survey emphasizing the experiences and reflections of software developers regarding GenAI tools. This includes both their use pattern but also potential harm being done by using GenAI. In addition, we also compared educator perspectives on industry usage with industry reports. The main conclusions related to the educators’ and developers’ perspectives are summarized in the grey box.

To elaborate on the educator perspective, we further used semi-structured interviews with tool creators and experienced educators studying or using GenAI on how they developed or integrated respective tools. The interviews addressed the research questions regarding actual student outcomes (RQ6), changing competencies (RQ8), and how GenAI shapes the future of computing education (RQ10). We found that educators are already seeing both the negative and positive impacts of GenAI from student use of these tools. Despite intentional planning for integration into their courses, instructors are noticing a large influx in both cheating and student unpreparedness for exams. However, students who use these tools responsibly are able to accomplish more than traditionally possible, especially in introductory courses. Instructors also highlighted the importance of GenAI tools in expanding assistance to students through help-seeking tools, allowing them to ask questions without the social risk of doing so in front of peers. While some might lament these uses as a degradation of student knowledge, many instructors who are already using GenAI in their courses see it as something that deprecates certain previously required competencies and enables new ones. Whether or not that’s true, most instructors shared the belief that competencies are currently in flux and that computing education is rapidly shifting to meet the occasion.

9 Future Work

Building upon the presented results, there are several pathways for future work. For example, we could replicate the systematic literature review in other domains beyond computing education and compare the CER community’s perspective to those of other domains, e.g., teacher education, other engineering disciplines, or even less technical domains. The same applies to the survey and interviews with educators. It would be interesting to identify differences and overlaps in how educators across domains integrate and perceive GenAI tools in their teaching practices.

The systematic literature review shows that research has focused on developing GenAI tools to help computer science students with tasks like writing code, understanding code, and receiving feedback and hints for programming problems. These tools are a natural

Key takeaways from our results include:

- **GenAI impacts educators and developers:** 80 % of developers use GenAI tools in their professional roles (DS-1) with those not using it citing either ethical concerns or company limitations in its use. At the same time, 75 % of educators acknowledge that the skills to program are changing as a result of GenAI (ES-3) and 30 % of educators are integrating it into their classes (ES-2). In contrast, only 22 % of educators are explicitly disallowing GenAI use (ES-1). Given the wide availability of GenAI tools for students, we are heartened that instructors are adapting to the new reality and not attempting to ban what effectively cannot be banned.
- **GenAI changes assessments:** Educators are increasing the weighting of exams (ES-22), exploring alternative invigilated exam techniques (including oral exams), and are emphasizing assessing the process of learning over correctness of answers.
- **GenAI changes programming competencies:** The majority of surveyed educators (75 %) believe program competencies have changed as a result of GenAI (ES-3). Of the remaining 25 %, many felt that the skills to program have remained the same, however, the importance of some skills have shifted (ES-4). Educators believe code reading has become more important than writing code from scratch, and that higher level skills like code testing, problem decomposition, problem understanding, and debugging have become more essential (ES-4, ES-5). When given options to rank, educators chose *problem understanding*, *reading code*, and *problem decomposition* as the most important skills for programming with GenAI. Developers pointed out a similar shift (DS-10), but added the need to *critically evaluate GenAI use and its output*, *writing prompts*, and *meticulousness* as new relevant competency components.
- **Need to train students to use GenAI in industry:** GenAI tools are almost ubiquitous in industry. Educators need to train students for industry so they are successful in their career. To achieve this goal, it is crucial to keep an open eye on recent developments regarding GenAI and job ads, and cooperate with industry partners so that educators and institutions can align their curricula, competency expectations, and study programs [66, 70, 150].
- **Guide students on how to use GenAI or use custom tools:** Based on the results of the literature review, studies reported more positive results when students were provided with guidance on how to use GenAI tools or used custom tools that include pedagogical guardrails.
- **Teach (educators) and students about GenAI challenges:** Critically using GenAI is crucial according to educators and developers. It is therefore important to teach students about the limitations and issues of current GenAI tools (e.g., ethical aspects, biases, plagiarism, etc.) so they can make informed decisions. For that to happen, educators need to receive relevant training. Our educator survey showed that those educators who did not integrate GenAI cited the lack of time, skills, pedagogical training, support, or resources as reasons for non-use (ES-20).

extension of existing GenAI tools like ChatGPT, which excel at solving problems in introductory computer science courses but struggle with guiding students to find answers on their own without giving away the solution entirely. Looking forward, we should consider how GenAI could address other challenges in learning to program. One possible area is debugging. Expert tutors often guide students to use debugging techniques like print statements, debuggers, or searching Stack Overflow. A key question is how future GenAI tools can effectively incorporate best practices, and

how they can actually support learning processes, scaffolding, or, for example, mastery learning. The role of human tutors is also important as GenAI tools become more capable of using teaching methods like the Socratic method without giving away solutions. This raises the question of what effective teaching looks like for instructors and how we should train them to help students in a GenAI-supported environment.

Another crucial aspect of future work is the impact of GenAI tools on the job market, and, for example, the qualifications of graduates. Are there going to be fewer entry-level software engineering jobs? To date, we know from studies that CS students and developers are using GenAI at a great scale, and even educators are increasingly integrating it into their courses. It is thus important to regularly investigate and align these recent developments.



Figure 10: Working Group 09 in Milan, Italy.

Data Availability

The data of the systematic literature review is published on OSF: https://osf.io/wpjxb/?view_only=cebf366db7f5423792b39de754972400.

Acknowledgments

This research was supported by the Research Council of Finland (Academy Research Fellow grant number 356114) and in part by the U.S. National Science Foundation IUSE Award #2417374.

We also acknowledge the support by the Google Award for Inclusion Research Program.

Finally, we want to say “Thank You” to all of our interview and survey participants. In particular, we acknowledge those educators who participated in our interviews who agreed to be named (in alphabetical order by surname):

- Bita Akram
- Kevin D. Ashley
- Brett Becker
- Yan Chen
- Paul Denny
- Barbara Ericson
- Majeed Kazemitabaar
- Amy Ko
- Lauren Margulieux
- Tereza Novotna
- Wesley Oliver
- Ray Pettit
- Brad Sheese

We acknowledge and thank the participation of several others who participated in the interviews who requested to remain anonymous.

References

- [1] Amy Adair. 2023. Teaching and Learning with AI: How Artificial Intelligence is Transforming the Future of Education. *XRDS* 29, 3 (apr 2023), 7–9. <https://doi.org/10.1145/3589252>
- [2] Vibhor Agarwal, Madhav Krishan Garg, Sahiti Dharmavaram, and Dhruv Kumar. 2024. “Which LLM should I use?”: Evaluating LLMs for tasks performed by Undergraduate Computer Science Students. arXiv:2402.01687 [cs.CY] <https://arxiv.org/abs/2402.01687>
- [3] Jana Al Hajj and Melike Sah. 2023. Assessing the Impact of ChatGPT in a PHP Programming Course. In *2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*. IEEE, New York, NY, USA, 1–10. <https://doi.org/10.1109/ISAS60782.2023.10391549>
- [4] Bedour Alshaigy and Virginia Grande. 2024. Forgotten Again: Addressing Accessibility Challenges of Generative AI Tools for People with Disabilities. In *Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction (Uppsala, Sweden) (NordCHI '24 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 68, 6 pages. <https://doi.org/10.1145/3677045.3685493>
- [5] Sara Amani, Lance White, Trini Balart, Laksha Arora, Dr. Kristi J. Shryock, Dr. Kelly Brumbelow, and Dr. Karan L. Watson. 2023. Generative AI Perceptions: A Survey to Measure the Perceptions of Faculty, Staff, and Students on Generative AI Tools in Academia. arXiv:2304.14415
- [6] Sihem Amer-Yahia, Angela Bonifati, Lei Chen, Guoliang Li, Kyuseok Shim, Jianliang Xu, and Xiaochun Yang. 2023. From Large Language Models to Databases and Back: A Discussion on Research and Education. *SIGMOD Rec.* 52, 3 (nov 2023), 49–56. <https://doi.org/10.1145/3631504.3631518>
- [7] Matin Amoozadeh, David Daniels, Daye Nam, Aayush Kumar, Stella Chen, Michael Hilton, Sruti Srinivasa Ragavan, and Mohammad Amin Alipour. [n.d.]. Trust in Generative AI among Students: An Exploratory Study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (New York, NY, USA, 2024-03-07) (SIGCSE 2024)*. Association for Computing Machinery, 67–73. <https://doi.org/10.1145/3626252.3630842>
- [8] Karen Anewalt and Jennifer Polack. 2023. Industry Trends in Software Engineering: Alumni Perspectives. *Journal of Computing Sciences in Colleges* 39, 3 (2023), 159–170. <https://doi.org/10.5555/3636988.3637014>
- [9] Chaitanya Arora, Utkarsh Venaik, Pavit Singh, Sahil Goyal, Jatin Tyagi, Shyama Goel, Ujjwal Singhal, and Dhruv Kumar. 2024. Analyzing LLM Usage in an Advanced Computing Class in India. arXiv:2404.04603 <https://arxiv.org/abs/2404.04603>
- [10] Imen Azaiz, Oliver Deckarm, and Sven Strickroth. 2023. AI-enhanced Auto-Correction of Programming Exercises: How Effective is GPT-3.5? *International Journal of Engineering Pedagogy (IJEP)* 13, 8 (Dec. 2023), 67–83. <https://doi.org/10.3991/ijep.v13i8.45621>
- [11] Imen Azaiz, Natalie Kiesler, and Sven Strickroth. 2024. Feedback-Generation for Programming Exercises With GPT-4. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1 (Milan, Italy) (ITiCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 31–37. <https://doi.org/10.1145/3649217.3653594>
- [12] Rishabh Balse, Viraj Kumar, Prajish Prasad, and Jayakrishnan Madathil Warriem. 2023. Evaluating the Quality of LLM-Generated Explanations for Logical Errors in CS1 Student Programs. In *Proceedings of the 16th Annual ACM India Compute Conference (Hyderabad, India) (COMPUTE '23)*. Association for Computing Machinery, New York, NY, USA, 49–54. <https://doi.org/10.1145/3627217.3627233>
- [13] Jose Barambones, Cristian Moral, Angélica de Antonio, Ricardo Imbert, Loïc Martínez-Normand, and Elena Villalba-Mora. 2024. ChatGPT for Learning HCI Techniques: A Case Study on Interviews for Personas. *IEEE Transactions on Learning Technologies* 17 (2024), 1486–1501. <https://doi.org/10.1109/TLT.2024.3386095>
- [14] Erik Barendsen, Violetta Lonati, Keith Quille, Rukiye Altin, Monica Divitini, Sara Hooshangi, Oscar Karnalim, Natalie Kiesler, Madison Melton, Calkin Suero Montero, and Anna Morpurgo. 2024. AI in and for K-12 Informatics Education. Life after Generative AI. In *Proceedings of the 2024 ACM Virtual Global Computing Education Conference V.2*. <https://doi.org/10.1145/3649409.3691073>
- [15] Shraddha Barke, Michael B James, and Nadia Polikarpova. 2023. Grounded Copilot: How Programmers Interact with Code-Generating Models. *Proceedings of ACM on Programming Languages* 7, OOPSLA1 (2023), 85–111. <https://doi.org/10.1145/3586030>
- [16] Brett A Becker, Michelle Craig, Paul Denny, Hieke Keuning, Natalie Kiesler, Juho Leinonen, Andrew Luxton-Reilly, James Prather, and Keith Quille. 2024. Generative AI in Introductory Programming. In *Computer Science Curricula 2023*, Amruth N. Kumar, Rajendra K. Raj, Sherif G. Aly, Monica D. Anderson, Brett A. Becker, Richard L. Blumenthal, Eric Eaton, Susan L. Epstein, Michael Goldweber, Pankaj Jalote, Douglas Le, Michael Oudshoorn, Marcelo Pias, Susan Reiser, Christian Servin, Rahul Simha, Titus Winters, and Qiao Xiang (Eds.). Association for Computing Machinery, New York, NY, USA, 438–439.

- [17] Brett A. Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. 2023. Programming Is Hard - Or at Least It Used to Be: Educational Opportunities and Challenges of AI Code Generation. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) (*SIGCSE 2023*). Association for Computing Machinery, New York, NY, USA, 500–506. <https://doi.org/10.1145/3545945.3569759>
- [18] Yasmine Belghith, Atefeh Mahdavi Goloujeh, Brian Magerko, Duri Long, Tom Mcklin, and Jessica Roberts. 2024. Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Article 276. <https://doi.org/10.1145/3613904.3642332>
- [19] Seth Bernstein, Paul Denny, Juho Leinonen, Lauren Kan, Arto Hellas, Matt Littlefield, Sami Sarsa, and Stephen Macneil. 2024. "Like a Nesting Doll": Analyzing Recursion Analogies Generated by CS Students Using Large Language Models. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1* (Milan, Italy) (*ITIcSE 2024*). Association for Computing Machinery, New York, NY, USA, 122–128. <https://doi.org/10.1145/3649217.3653533>
- [20] Chris Bopp, Anne Foerst, and Brian Kellogg. 2024. The Case for LLM Workshops. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (*SIGCSE 2024*). Association for Computing Machinery, New York, NY, USA, 130–136. <https://doi.org/10.1145/3626252.3630941>
- [21] Philipp Brauner, Alexander Hick, Ralf Philipsen, and Martina Ziefle. 2023. What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. *Frontiers in Computer Science* 5 (2023), 1113903. <https://doi.org/10.3389/fcomp.2023.1113903>
- [22] Doga Cambaz and Xiaoling Zhang. 2024. Use of AI-driven Code Generation Models in Teaching and Learning Programming: a Systematic Literature Review. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (*SIGCSE 2024*). Association for Computing Machinery, New York, NY, USA, 172–178. <https://doi.org/10.1145/3626252.3630958>
- [23] Cecilia Ka Yuk Chan and Katherine K. W. Lee. 2023. The AI Generation Gap: Are Gen Z Students More Interested in Adopting Generative AI Such as ChatGPT in Teaching and Learning Than Their Gen X and Millennial Generation Teachers? arXiv:2305.02878
- [24] John Chen, Xi Lu, Yuzhou Du, Michael Rejtig, Ruth Bagley, Mike Horn, and Uri Wilensky. 2024. Learning Agent-based Modeling with LLM Companions: Experiences of Novices and Experts Using ChatGPT & NetLogo Chat. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 141, 18 pages. <https://doi.org/10.1145/3613904.3642377>
- [25] Rudrajit Choudhuri, Dylan Liu, Igor Steinmacher, Marco Gerosa, and Anita Sarma. 2024. How Far Are We? The Triumphs and Trials of Generative AI in Learning Software Engineering. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering* (Lisbon, Portugal) (*ICSE '24*). Association for Computing Machinery, New York, NY, USA, Article 184, 13 pages. <https://doi.org/10.1145/3597503.3639201>
- [26] Bruno Pereira Cipriano and Pedro Alves. 2024. "ChatGPT Is Here to Help, Not to Replace Anybody" – An Evaluation of Students' Opinions On Integrating ChatGPT In CS Courses. arXiv:2404.17443
- [27] Bruno Pereira Cipriano, Pedro Alves, and Paul Denny. 2024. A Picture Is Worth a Thousand Words: Exploring Diagram and Video-Based OOP Exercises to Counter LLM Over-Reliance. arXiv:2403.08396 <https://arxiv.org/abs/2403.08396>
- [28] Mariana Coutinho, Lorena Marques, Anderson Santos, Marcio Dahia, Cesar França, and Ronnie de Souza Santos. 2024. The Role of Generative AI in Software Development Productivity: A Pilot Case Study. In *Proceedings of the 1st ACM International Conference on AI-Powered Software* (Porto de Galinhas, Brazil) (*AIware 2024*). Association for Computing Machinery, New York, NY, USA, 131–138. <https://doi.org/10.1145/3664646.3664773>
- [29] Aaron S. Crandall, Gina Sprint, and Bryan Fischer. 2023. Generative Pre-Trained Transformer (GPT) Models as a Code Review Feedback Tool in Computer Science Programs. *Journal of the Consortium for Computing Sciences in Colleges* 39, 1 (oct 2023), 38–47.
- [30] Javier Cámara, Javier Troya, Julio Montes-Torres, and Francisco J. Jaime. 2024. Generative AI in the Software Modeling Classroom: An Experience Report With ChatGPT and Unified Modeling Language. *IEEE Software* 41, 6 (2024), 73–81. <https://doi.org/10.1109/MS.2024.3385309>
- [31] Paul Denny, David H. Smith IV au2, Max Fowler, James Prather, Brett A. Becker, and Juho Leinonen. 2024. Explaining Code with a Purpose: An Integrated Approach for Developing Code Comprehension and Prompting Skills. <https://doi.org/10.48550/arXiv.2403.06050> arXiv:2403.06050 [cs.HC]
- [32] Paul Denny, Hassan Khosravi, Arto Hellas, Juho Leinonen, and Sami Sarsa. 2023. Can We Trust AI-Generated Educational Content? Comparative Analysis of Human and AI-Generated Learning Resources. arXiv:2306.10509 <https://arxiv.org/abs/2306.10509>
- [33] Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A. Becker, and Brent N. Reeves. 2023. Promptly: Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generators. arXiv:2307.16364 [cs.HC] <https://arxiv.org/abs/2307.16364>
- [34] Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A. Becker, and Brent N. Reeves. 2024. Prompt Problems: A New Programming Exercise for the Generative AI Era. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (*SIGCSE 2024*). Association for Computing Machinery, New York, NY, USA, 296–302. <https://doi.org/10.1145/3626252.3630909>
- [35] Paul Denny, Stephen MacNeil, Jaromir Savelka, Leo Porter, and Andrew Luxton-Reilly. 2024. Desirable Characteristics for AI Teaching Assistants in Programming Education. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1* (Milan, Italy) (*ITIcSE 2024*). Association for Computing Machinery, New York, NY, USA, 408–414. <https://doi.org/10.1145/3649217.3653574>
- [36] Paul Denny, James Prather, Brett A. Becker, James Finnie-Ansley, Arto Hellas, Juho Leinonen, Andrew Luxton-Reilly, Brent N. Reeves, Eddie Antonio Santos, and Sami Sarsa. 2024. Computing Education in the Era of Generative AI. *Commun. ACM* 67, 2 (Jan. 2024), 56–67. <https://doi.org/10.1145/3624720>
- [37] Paul Denny, Sami Sarsa, Arto Hellas, and Juho Leinonen. 2022. Robosourcing Educational Resources – Leveraging Large Language Models for Learnersourcing. <https://doi.org/10.48550/ARXIV.2211.04715>
- [38] Jacob Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang, Juran Zhang, Yujia Zheng, Aidan Doyle, Pragnya Sridhar, Arav Agarwal, Christopher Bogart, Eric Keylor, Can Kultur, Jaromir Savelka, and Majd Sakr. 2024. A Comparative Study of AI-Generated (GPT-4) and Human-crafted MCQs in Programming Education. In *Proceedings of the 26th Australasian Computing Education Conference (ACE '24)*. Association for Computing Machinery, New York, NY, USA, 114–123. <https://doi.org/10.1145/3636243.3636256>
- [39] Amanda S. Fernandez and Kimberly A. Cornell. 2024. CS1 with a Side of AI: Teaching Software Verification for Secure Code in the Era of Generative AI. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 345–351. <https://doi.org/10.1145/3626252.3630817>
- [40] James Finnie-Ansley, Paul Denny, Brett A. Becker, Andrew Luxton-Reilly, and James Prather. 2022. The Robots Are Coming: Exploring the Implications of OpenAI Codex on Introductory Programming. In *Proceedings of the 24th Australasian Computing Education Conference (ACE '22)*. Association for Computing Machinery, New York, NY, USA, 10–19. <https://doi.org/10.1145/3511861.3511863>
- [41] James Finnie-Ansley, Paul Denny, Andrew Luxton-Reilly, Eddie Antonio Santos, James Prather, and Brett A. Becker. 2023. My AI Wants to Know If This Will Be on the Exam: Testing OpenAI's Codex on CS2 Programming Exercises. In *Proceedings of the 25th Australasian Computing Education Conference*. 97–104. <https://doi.org/10.1145/3576123.3576134>
- [42] Virginia Grande, Natalie Kiesler, and Maria Andreina Francisco R. 2024. Student Perspectives on Using a Large Language Model (LLM) for an Assignment on Professional Ethics. In *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 1*, 478–484. <https://doi.org/10.1145/3649217.3653624>
- [43] Shuchi Grover, Deborah Fields, Yasmin Kafai, Shana White, and Carla Strickland. 2024. Enduring Lessons from 'Computer Science for All' for AI Education in Schools. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, 1533–1534. <https://doi.org/10.1145/3626253.3631656>
- [44] Philip J Guo. 2023. Six Opportunities for scientists and engineers to learn programming using AI Tools such as ChatGPT. *Computing in Science & Engineering* 25, 3 (2023), 73–78. <https://doi.org/10.1109/MCSE.2023.3308476>
- [45] Tonia Haikal and Robert Harold Lightfoot. 2024. Enhancing Education Through Thoughtful Integration of Large Language Models in Assigned Work. In *2024 ASEE-GSW*. American Society for Engineering Education, Washington, DC, USA, 9. <https://doi.org/10.18260/1-2-45377>
- [46] Philipp Haindl and Gerald Weinberger. 2024. Students' Experiences of Using ChatGPT in an Undergraduate Programming Course. *IEEE Access* 12 (2024), 43519–43529. <https://doi.org/10.1109/ACCESS.2024.3380909>
- [47] Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Predicting academic performance: a systematic literature review. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 175–199. <https://doi.org/10.1145/3293881.3295783>
- [48] Irene Hou, Owen Man, Sophia Mettelle, Sebastian Gutierrez, Kenneth Angelikas, and Stephen MacNeil. 2024. More Robots are Coming: Large Multimodal Models (ChatGPT) can Solve Visually Diverse Images of Parsons Problems. In *Proceedings of the 26th Australasian Computing Education Conference*. 29–38. <https://doi.org/10.1145/3636243.3636247>
- [49] Irene Hou, Sophia Mettelle, Owen Man, Zhuo Li, Cynthia Zastudil, and Stephen MacNeil. 2024. The Effects of Generative AI on Computing Students' Help-Seeking Preferences. In *Proceedings of the 26th Australasian Computing Education Conference*. ACM, Sydney NSW Australia, 39–48. <https://doi.org/10.1145/3636243.3636248>
- [50] Xinying Hou, Zihan Wu, Xu Wang, and Barbara J. Ericson. 2024. CodeTailor: LLM-Powered Personalized Parsons Puzzles for Engaging Support While Learning Programming. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale*. ACM, Atlanta GA USA, 51–62. <https://doi.org/10.1145/3657604.3662032>

- [51] Minjie Hu, Tony Assadi, and Hamid Mahrooian. 2023. Explicitly Introducing ChatGPT into First-year Programming Practice: Challenges and Impact. In *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. IEEE, New York, NY, USA, 1–6. <https://doi.org/10.1109/TALE56641.2023.10398297>
- [52] Sven Jacobs and Steffen Jaschke. 2024. Evaluating the Application of Large Language Models to Generate Feedback in Programming Education. In *2024 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, New York, NY, USA, 1–5. <https://doi.org/10.1109/EDUCON60312.2024.10578838> arXiv:2403.09744 [cs]
- [53] Johan Jeuring, Roel Groot, and Hieke Keuning. 2024. What Skills Do You Need When Developing Software Using ChatGPT? (Discussion Paper). In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research (Koli Calling '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3631802.3631807>
- [54] Hyoungwook Jin, Seonghee Lee, Hyungyu Shin, and Juho Kim. 2024. Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–28. <https://doi.org/10.1145/3613904.3642349>
- [55] Martin Jonsson and Jakob Tholander. 2022. Cracking the Code: Co-coding with AI in Creative Programming Education. In *Proceedings of the 14th Conference on Creativity and Cognition (C&C '22)*. Association for Computing Machinery, New York, NY, USA, 5–14. <https://doi.org/10.1145/3527927.3532801>
- [56] Gregor Jošt, Viktor Taneski, and Sašo Karakatič. 2024. The Impact of Large Language Models on Programming Education and Student Learning Outcomes. *Applied Sciences* 14, 10 (2024), 4115.
- [57] Breanna Jury, Angela Lorusso, Juho Leinonen, Paul Denny, and Andrew Luxton-Reilly. 2024. Evaluating LLM-generated Worked Examples in an Introductory Programming Course. In *Proceedings of the 26th Australasian Computing Education Conference (Sydney, NSW, Australia) (ACE '24)*. Association for Computing Machinery, New York, NY, USA, 77–86. <https://doi.org/10.1145/3636243.3636252>
- [58] Oscar Karnalim, Erico Darmawan Handoyo, Hapnes Toba, Yehezkiel David Setiawan, Meliana Christianti Johan, and Josephine Alvina Luwia. 2023. Plagiarism and AI Assistance Misuse in Web Programming: Unfair Benefits and Characteristics. In *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. IEEE, 1–5.
- [59] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J. Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the Effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Article 455. <https://doi.org/10.1145/3544548.3580919>
- [60] Majeed Kazemitabaar, Runlong Ye, Xiaoning Wang, Austin Zachary Henley, Paul Denny, Michelle Craig, and Tovi Grossman. 2024. CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 650, 20 pages. <https://doi.org/10.1145/3613904.3642773>
- [61] Tyson Kendon, Leanne Wu, and John Aycok. 2023. AI-Generated Code Not Considered Harmful. In *Proceedings of the 25th Western Canadian Conference on Computing Education*. Article 3. <https://doi.org/10.1145/3593342.3593349>
- [62] Krishnamurthy Kethapadi, Himabindu Lakkaraju, and Nazneen Rajani. 2023. Generative AI meets Responsible AI: Practical Challenges and Opportunities. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5805–5806. <https://doi.org/10.1145/3580305.3599557>
- [63] Hieke Keuning, Andrew Luxton-Reilly, Claudia Ott, Andrew Petersen, and Natalie Kiesler. 2024. Goodbye Hello World - Research Questions for a Future CS1 Curriculum. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*. <https://doi.org/10.1145/3699538.3699591>
- [64] Arshia Khan and Janna Madden. 2018. Active Learning: A New Assessment Model That Boosts Confidence and Learning While Reducing Test Anxiety. *International Journal of Modern Education and Computer Science* 10, 12 (2018), 1.
- [65] Natalie Kiesler. 2023. Beyond the Textbook: Rethinking Students' Competencies in the LLM Era. Generative AI: Implications for Teaching and Learning. <https://doi.org/10.13140/RG.2.2.28355.37922/1>
- [66] Natalie Kiesler and John Impagliazzo. 2023. Industry's Expectations of Graduate Dispositions. In *2023 IEEE Frontiers in Education Conference (FIE)*. 1–5. <https://doi.org/10.1109/FIE58773.2023.10343406>
- [67] Natalie Kiesler, Dominic Lohr, and Hieke Keuning. 2023. Exploring the Potential of Large Language Models to Generate Formative Programming Feedback. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, College Station, TX, USA, 1–5. <https://doi.org/10.1109/FIE58773.2023.10343457>
- [68] Natalie Kiesler and Daniel Schiffrer. 2023. Large Language Models in Introductory Programming Education: ChatGPT's Performance and Implications for Assessments. <https://doi.org/10.48550/arXiv.2308.08572> arXiv:2308.08572 [cs.SE]
- [69] Natalie Kiesler, Ingo Scholz, Jens Albrecht, Friedhelm Stappert, and Uwe Wienkop. 2024. Novice Learners of Programming and Generative AI - Prior Knowledge Matters. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*. <https://doi.org/10.1145/3699538.3699580>
- [70] Natalie Kiesler and Carsten Thorbrügge. 2023. Socially Responsible Programming in Computing Education and Expectations in the Profession. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (Turku, Finland) (ITiCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 443–449. <https://doi.org/10.1145/3587102.3588839>
- [71] Bailey Kimmel, Austin Lee Geisert, Lily Yaro, Brendan Gipson, Ronald Taylor Hotchkiss, Sidney Kwame Osae-Asante, Hunter Vaught, Grant Winger, and Chase Yamaguchi. 2024. Enhancing Programming Error Messages in Real Time with Generative AI. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 608, 7 pages. <https://doi.org/10.1145/3613905.3647967>
- [72] Barbara Kitchenham and Pearl Brereton. 2013. A systematic review of systematic review process research in software engineering. *Information and software technology* 55, 12 (2013), 2049–2075.
- [73] Tomaž Kosar, Dragana Ostojić, Yu David Liu, and Marjan Memik. 2024. Computer Science Education in ChatGPT Era: Experiences from an Experiment in a Programming Course for Novice Programmers. *Mathematics* 12, 5 (2024). <https://doi.org/10.3390/math12050629>
- [74] Jon A Krosnick. 2018. Questionnaire design. *The Palgrave handbook of survey research* (2018), 439–455.
- [75] Mohammad Amin Kuhail, Sujith Samuel Mathew, Ashraf Khalil, Jose Berengueres, and Syed Jawad Hussain Shah. 2024. "Will I be replaced?" Assessing ChatGPT's effect on software development and programmer perceptions of AI tools. *Science of Computer Programming* 235 (2024), 103111.
- [76] A. Kumar, M. Lakshmi Devi, and J. S. Saltz. 2023. Bridging the Gap in AI-Driven Workflows: The Case for Domain-Specific Generative Bots. In *2023 IEEE International Conference on Big Data (BigData)*. IEEE Computer Society, Los Alamitos, CA, USA, 2421–2430. <https://doi.org/10.1109/BigData59044.2023.10386894>
- [77] Kimio Kuramitsu, Momoka Obara, Miyu Sato, and Yuka Akinobu. 2024. Training AI Model that Suggests Python Code from Student Requests in Natural Language. *Journal of Information Processing* 32 (2024), 69–76. <https://doi.org/10.2197/ipsjip.32.69>
- [78] Kimio Kuramitsu, Yui Obara, Miyu Sato, and Momoka Obara. 2023. KOGI: A Seamless Integration of ChatGPT into Jupyter Environments for Programming Education. In *Proceedings of the 2023 ACM SIGPLAN International Symposium on SPLASH-E (Cascais, Portugal) (SPLASH-E 2023)*. Association for Computing Machinery, New York, NY, USA, 50–59. <https://doi.org/10.1145/3622780.3623648>
- [79] Celine Latulipe, N Bruce Long, and Carlos E Seminaro. 2015. Structuring Flipped Classes With Lightweight Teams and Gamification. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. 392–397.
- [80] Sam Lau and Philip Guo. 2023. From "Ban It Till We Understand It" to "Resistance is Futile": How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools such as ChatGPT and GitHub Copilot. <https://doi.org/10.1145/3568813.3600138>
- [81] Juho Leinonen, Paul Denny, Stephen MacNeil, Sami Sarsa, Seth Bernstein, Joanne Kim, Andrew Tran, and Arto Hellas. 2023. Comparing Code Explanations Created by Students and Large Language Models. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*. 124–130. <https://doi.org/10.1145/3587102.3588875>
- [82] Juho Leinonen, Arto Hellas, Sami Sarsa, Brent Reeves, Paul Denny, James Prather, and Brett A. Becker. 2023. Using Large Language Models to Enhance Programming Error Messages. 563–569. <https://doi.org/10.1145/3545945.3569770>
- [83] Hongxin Li, Jingran Su, Yuntao Chen, Qing Li, and Zhaoxiang Zhang. 2024. SheetCopilot: bringing software productivity to the next level through large language models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (New Orleans, LA, USA) (NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 220.
- [84] Jenny T. Liang, Chenyang Yang, and Brad A. Myers. 2024. A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*. Article 52. <https://doi.org/10.1145/3597503.3608128>
- [85] Jian Liao, Linrong Zhong, Longting Zhe, Handan Xu, Ming Liu, and Tao Xie. 2024. Scaffolding Computational Thinking With ChatGPT. *IEEE Transactions on Learning Technologies* 17 (2024), 1668–1682. <https://doi.org/10.1109/TLT.2024.3392896>
- [86] Mark Liffiton, Brad E Sheese, Jaromir Savelka, and Paul Denny. 2024. CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research (Koli, Finland) (Koli Calling '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 11 pages. <https://doi.org/10.1145/3631802.3631830>
- [87] Rongxin Liu, Carter Zenke, Charlie Liu, Andrew Holmes, Patrick Thornton, and David J. Malan. 2024. Teaching CS50 with AI: Leveraging Generative

- Artificial Intelligence in Computer Science Education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. 750–756. <https://doi.org/10.1145/3626252.3630938>
- [88] Evanfiya Logacheva, Arto Hellas, James Prather, Sami Sarsa, and Juho Leinonen. 2024. Evaluating Contextually Personalized Programming Exercises Created with Generative AI. In *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1* (Melbourne, VIC, Australia) (*ICER '24*). Association for Computing Machinery, New York, NY, USA, 95–113. <https://doi.org/10.1145/3632620.3671103>
- [89] Dominic Lohr, Hieke Keuning, and Natalie Kiesler. 2024. You're (Not) My Type – Can LLMs Generate Feedback of Specific Types for Introductory Programming Tasks? *Journal of Computer Assisted Learning* (2024). <https://doi.org/10.48550/arXiv.2412.03516> Preprint available.
- [90] Wenhan Lyu, Yimeng Wang, Tingting Rachel Chung, Yifan Sun, and Yixuan Zhang. 2024. Evaluating the Effectiveness of LLMs in Introductory Computer Science Education: A Semester-Long Field Study. *arXiv preprint arXiv:2404.13414* (2024).
- [91] Boxaun Ma, Li Chen, and Shin'ichi Konomi. 2024. Enhancing Programming Education with ChatGPT: A Case Study on Student Perceptions and Interactions in a Python Course. *arXiv preprint arXiv:2403.15472* (2024). <https://doi.org/10.48550/arXiv.2403.15472>
- [92] Stephen MacNeil, Celine Latulipe, Bruce Long, and Aman Yadav. 2016. Exploring Lightweight Teams in a Distributed Learning Environment. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (Memphis, Tennessee, USA) (*SIGCSE '16*). Association for Computing Machinery, New York, NY, USA, 193–198. <https://doi.org/10.1145/2839509.2844577>
- [93] Stephen MacNeil, Juho Leinonen, Paul Denny, Natalie Kiesler, Arto Hellas, James Prather, Brett A. Becker, Michel Wermelinger, and Karen Reid. 2024. Discussing the Changing Landscape of Generative AI in Computing Education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*. 1916. <https://doi.org/10.1145/3626253.3635369>
- [94] Stephen MacNeil, Andrew Tran, Arto Hellas, Joanne Kim, Sami Sarsa, Paul Denny, Seth Bernstein, and Juho Leinonen. 2023. Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) (*SIGCSE 2023*). Association for Computing Machinery, New York, NY, USA, 931–937. <https://doi.org/10.1145/3545945.3569785>
- [95] Eric D. Manley, Timothy Urness, Andrei Migunov, and Md. Alimoor Reza. 2024. Examining Student Use of AI in CS1 and CS2. *Journal of Computing Sciences in Colleges (JCSC)* 39, 6 (May 2024), 41–51.
- [96] Julia M Markel, Steven G Opferman, James A Landay, and Chris Piech. 2023. GPTEach: Interactive TA Training with GPT-based Students. In *Proceedings of the tenth ACM conference on learning@scale*. 226–236.
- [97] Dylan Edward Moore, Sophia R. R. Moore, Bansharee Ireen, Winston P. Iskandar, Grigory Artazyan, and Elizabeth L. Murnane. 2024. Teaching artificial intelligence in extracurricular contexts through narrative-based learnersourcing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 270, 28 pages. <https://doi.org/10.1145/3613904.3642198>
- [98] Hussein Mozannar, Gagan Bansal, Adam Fournay, and Eric Horvitz. 2024. Reading Between the Lines: Modeling User Behavior and Costs in AI-Assisted Programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Article 142. <https://doi.org/10.1145/3613904.3641936>
- [99] Emiliana Murgia, Maria Soledad Pera, Monica Landoni, and Theo Huibers. 2023. Children on ChatGPT Readability in an Educational Context: Myth or Opportunity?. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 311–316. <https://doi.org/10.1145/3563359.3596996>
- [100] Daye Nam, Andrew Macvean, Vincent Hellendoorn, Bogdan Vasilescu, and Brad Myers. 2024. Using an LLM to Help With Code Understanding. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering* (Lisbon, Portugal) (*ICSE '24*). Association for Computing Machinery, New York, NY, USA, Article 97, 13 pages. <https://doi.org/10.1145/3597503.3639187>
- [101] Andrés Neyem, Luis A. González, Marcelo Mendoza, Juan Pablo Sandoval Alcocer, Leonardo Centellas, and Carlos Paredes. 2024. Toward an AI Knowledge Assistant for Context-Aware Learning Experiences in Software Capstone Project Development. *IEEE Transactions on Learning Technologies* 17 (2024), 1639–1654. <https://doi.org/10.1109/TLT.2024.3396735>
- [102] Andres Neyem, Juan Pablo Sandoval Alcocer, Marcelo Mendoza, Leonardo Centellas-Claros, Luis A. Gonzalez, and Carlos Paredes-Robles. 2024. Exploring the Impact of Generative AI for StandUp Report Recommendations in Software Capstone Project Development. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (*SIGCSE 2024*). Association for Computing Machinery, New York, NY, USA, 951–957. <https://doi.org/10.1145/3626252.3630854>
- [103] Sydney Nguyen, Hannah McLean Babe, Yangtian Zi, Arjun Guha, Carolyn Jane Anderson, and Molly Q Feldman. 2024. How Beginning Programmers and Code LLMs (Mis)read Each Other. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 651, 26 pages. <https://doi.org/10.1145/3613904.3642706>
- [104] Lars Oestreicher. 2023. New Perspectives on Education and Examination in the Age of Artificial Intelligence. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.
- [105] Stack Overflow. 2024. Stack Overflow Developer Survey 2024. <https://survey.stackoverflow.co/2024/ai>.
- [106] Maciej Pankiewicz and Ryan S Baker. 2024. Navigating Compiler Errors with AI Assistance—A Study of GPT Hints in an Introductory Programming Course. *arXiv preprint arXiv:2403.12737* (2024).
- [107] Neil Perry, Megha Srivastava, Deepak Kumar, and Dan Boneh. 2023. Do Users Write More Insecure Code with AI Assistants?. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (Copenhagen, Denmark) (*CCS '23*). Association for Computing Machinery, New York, NY, USA, 2785–2799. <https://doi.org/10.1145/3576915.3623157>
- [108] Ivica Pesovski, Ricardo Santos, Roberto Henriques, and Vladimir Trajkovik. 2024. Generative AI for Customizable Learning Experiences. *Sustainability* 16, 7 (2024), 3034.
- [109] Carrie Anne Philbin. 2023. Impact of Generative AI on K-12 Students' Perceptions of Computing: A Research Proposal. In *Proceedings of the 18th WiPSCe Conference on Primary and Secondary Computing Education Research*. Article 28. <https://doi.org/10.1145/3605468.3609775>
- [110] Siddhartha Prasad, Ben Greenman, Tim Nelson, and Shirram Krishnamurthi. 2023. Generating Programs Trivially: Student Use of Large Language Models. In *Proceedings of the ACM Conference on Global Computing Education Vol 1* (Hyderabad, India) (*CompEd 2023*). Association for Computing Machinery, New York, NY, USA, 126–132. <https://doi.org/10.1145/3576882.3617921>
- [111] James Prather, Paul Denny, Juho Leinonen, Brett A. Becker, Ibrahim Albluwi, Michelle Craig, Hieke Keuning, Natalie Kiesler, Tobias Kohn, Andrew Luxton-Reilly, Stephen MacNeil, Andrew Petersen, Raymond Pettit, Brent N. Reeves, and Jaromir Savelka. 2023. The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*. 108–159. <https://doi.org/10.1145/3623762.3633499>
- [112] James Prather, Juho Leinonen, Natalie Kiesler, Jamie Gorson Benario, Sam Lau, Stephen MacNeil, Narges Norouzi, Simone Opel, Virginia Pettit, Leo Porter, Brent N. Reeves, Jaromir Savelka, David H. Smith, Sven Strickroth, and Daniel Zingaro. 2024. How Instructors Incorporate Generative AI into Teaching Computing. In *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 2*. 771–772. <https://doi.org/10.1145/3649405.3659534>
- [113] James Prather, Brent Reeves, Paul Denny, Brett Becker, Juho Leinonen, Stephen MacNeil, Solofoarisina Arisoa Randrianasolo, Bailey Kimmel, Jared Wright, and Ben Briggs. 2024. The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers. In *Proceedings of the ICER 2024*.
- [114] James Prather, Brent N. Reeves, Paul Denny, Brett A. Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett Powell, James Finnie-Ansley, and Eddie Antonio Santos. 2023. 'It's Weird That it Knows What I Want': Usability and Interactions with Copilot for Novice Programmers. <https://doi.org/10.48550/arXiv.2304.02491>
- [115] Basit Qureshi. 2023. ChatGPT in Computer Science Curriculum Assessment: An analysis of Its Successes and Shortcomings. In *Proceedings of the 2023 9th International Conference on E-Society, e-Learning and e-Technologies* (Portsmouth, United Kingdom) (*ICSLT '23*). Association for Computing Machinery, New York, NY, USA, 7–13. <https://doi.org/10.1145/3613944.3613946>
- [116] Md Mostafizer Rahman and Yutaka Watanobe. 2023. ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences* 13, 9 (2023), 5783.
- [117] Rajendra Raj, Mihaela Sabin, John Impagliazzo, David Bowers, Mats Daniels, Feliene Hermans, Natalie Kiesler, Amruth N. Kumar, Bonnie MacKellar, Renée McCauley, Syed Waqar Nabi, and Michael Oudshoorn. 2021. Professional Competencies in Computing Education: Pedagogies and Assessment. In *Proceedings of the 2021 Working Group Reports on Innovation and Technology in Computer Science Education*. 133–161. <https://doi.org/10.1145/3502870.3506570>
- [118] Parsa Rajabi, Parnian Taghipour, Diana Cukierman, and Tenzin Doleck. 2023. Exploring ChatGPT's Impact on Post-secondary Education: A Qualitative Study. In *Western Canadian Conference on Computing Education (WCCCE'23), May 04–05, 2023*. Simon Fraser University.
- [119] Brent Reeves, Sami Sarsa, James Prather, Paul Denny, Brett A. Becker, Arto Hellas, Bailey Kimmel, Garrett Powell, and Juho Leinonen. 2023. Evaluating the Performance of Code Generation Models for Solving Parsons Problems With Small Prompt Variations. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*. 299–305. <https://doi.org/10.1145/3587102.3588805>

- [120] Lianne Roest, Hieke Keuning, and Johan Jeuring. 2024. Next-Step Hint Generation for Introductory Programming Using Large Language Models. In *Proceedings of the 26th Australasian Computing Education Conference*. 144–153.
- [121] Pati Ruiz, Alessandra Rangel, and Merijke Coenraad. 2024. Using Generative AI to Support PK-12 Teaching and Learning: Developing Sample Lessons and More. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*. 1800–1801. <https://doi.org/10.1145/3626253.3635522>
- [122] Gustavo Sandoval, Hammond Pearce, Teo Nys, Ramesh Karri, Siddharth Garg, and Brendan Dolan-Gavitt. 2023. Lost at C: A User Study on the Security Implications of Large Language Model Code Assistants. In *32nd USENIX Security Symposium (USENIX Security 23)*. 2205–2222.
- [123] O. Dos Santos and D. Cury. 2023. Challenging the Confirmation Bias: Using ChatGPT as a Virtual Peer for Peer Instruction in Computer Programming Education. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–7. <https://doi.org/10.1109/FIE58773.2023.10343247>
- [124] Sami Sarsa, Paul Denny, Arto Hellas, and Juho Leinonen. 2022. Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. In *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1 (Lugano and Virtual Event, Switzerland) (ICER '22)*. Association for Computing Machinery, New York, NY, USA, 27–43. <https://doi.org/10.1145/3501385.3543957>
- [125] Jaromir Savelka, Arav Agarwal, Marshall An, Chris Bogart, and Majd Sakr. 2023. Thrilled by Your Progress! Large Language Models (GPT-4) No Longer Struggle to Pass Assessments in Higher Education Programming Courses. In *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 1 (Chicago, IL, USA) (ICER '23)*. Association for Computing Machinery, New York, NY, USA, 78–92. <https://doi.org/10.1145/3568813.3600142>
- [126] Jaromir Savelka, Arav Agarwal, Christopher Bogart, Yifan Song, and Majd Sakr. 2023. Can Generative Pre-trained Transformers (GPT) Pass Assessments in Higher Education Programming Courses?. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*. 117–123. <https://doi.org/10.1145/3587102.3588792>
- [127] Jaromir Savelka, Paul Denny, Mark Liffiton, and Brad Sheese. 2023. Efficient Classification of Student Help Requests in Programming Courses Using Large Language Models. *arXiv preprint arXiv:2310.20105* (2023).
- [128] Andreas Scholl and Natalie Kiesler. 2024. How Novice Programmers Use and Experience ChatGPT when Solving Programming Exercises in an Introductory Course. <https://doi.org/10.48550/arXiv.2407.20792> arXiv:2407.20792 [cs.AI] accepted at 2024 IEEE ASEE Frontiers in Education Conference.
- [129] Andreas Scholl, Daniel Schiffner, and Natalie Kiesler. 2024. Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT. In *Proceedings of DELFI 2024*, Sandra Schulz and Natalie Kiesler (Eds.), 63–79. https://doi.org/10.18420/delfi2024_05
- [130] Anshul Shah, Jerry Yu, Thanh Tong, and Adalbert Gerald Soosai Raj. 2024. Working with Large Code Bases: A Cognitive Apprenticeship Approach to Teaching Software Engineering. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. 1209–1215. <https://doi.org/10.1145/3626252.3630755>
- [131] Shang Shanshan and Geng Sen. 2024. Empowering learners with AI-generated content for programming learning and computational thinking: The lens of extended effective use theory. *Journal of Computer Assisted Learning* (2024). <https://doi.org/10.1111/jcal.12996>
- [132] Judy Sheard, Paul Denny, Arto Hellas, Juho Leinonen, Lauri Malmi, and Simon. 2024. Instructor Perceptions of AI Code Generation Tools - A Multi-Institutional Interview Study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. 1223–1229. <https://doi.org/10.1145/3626252.3630880>
- [133] Brad Sheese, Mark Liffiton, Jaromir Savelka, and Paul Denny. 2024. Patterns of Student Help-Seeking When Using a Large Language Model-Powered Programming Assistant. In *Proceedings of the 26th Australasian Computing Education Conference (Sydney, NSW, Australia) (ACE '24)*. Association for Computing Machinery, New York, NY, USA, 49–57. <https://doi.org/10.1145/3636243.3636249>
- [134] Abdullhadi Shoufan. 2023. Can Students without Prior Knowledge Use ChatGPT to Answer Test Questions? An Empirical Study. *ACM Trans. Comput. Educ.* 23, 4, Article 45 (dec 2023), 29 pages. <https://doi.org/10.1145/3628162>
- [135] Carlos Alexandre Gouvea da Silva, Felipe Negrelle Ramos, Rafael Veiga de Moraes, and Edson Leonardo dos Santos. 2024. ChatGPT: Challenges and benefits in software programming for higher education. *Sustainability* 16, 3 (2024), 1245.
- [136] Anjali Singh, Christopher Brooks, Xu Wang, Warren Li, Juho Kim, and Deepti Wilson. 2024. Bridging Learnersourcing and AI: Exploring the Dynamics of Student-AI Collaborative Feedback Generation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference (Kyoto, Japan) (LAK '24)*. Association for Computing Machinery, New York, NY, USA, 742–748. <https://doi.org/10.1145/3636555.3636853>
- [137] Sandro Speth, Niklas Meißner, and Steffen Becker. 2023. Investigating the Use of AI-Generated Exercises for Beginner and Intermediate Programming Courses: A ChatGPT Case Study. In *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, 142–146.
- [138] Pragnya Sridhar, Aidan Doyle, Arav Agarwal, Christopher Bogart, Jaromir Savelka, and Majd Sakr. 2023. Harnessing LLMs in Curricular Design: Using GPT-4 to Support Authoring of Learning Objectives. *arXiv preprint arXiv:2306.17459* (2023).
- [139] Dan Sun, Azzeddine Boudouaia, Chengcong Zhu, and Yan Li. 2024. Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 14.
- [140] Ben Arie Tanay, Lexy Arinze, Siddhant S. Joshi, Kirsten A. Davis, and James C. Davis. 2024. An Exploratory Study on Upper-Level Computing Students' Use of Large Language Models as Tools in a Semester-Long Project. arXiv:2403.18679 <https://arxiv.org/abs/2403.18679>
- [141] Steven L. Tanimoto. 2023. Five Futures with AI Coding Agents. In *Companion Proceedings of the 7th International Conference on the Art, Science, and Engineering of Programming*. 32–38. <https://doi.org/10.1145/3594671.3594685>
- [142] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Article 680. <https://doi.org/10.1145/3613904.3642902>
- [143] Andrew Taylor, Alexandra Vassar, Jake Renzella, and Hammond Pearce. 2024. dcc-help: Transforming the Role of the Compiler by Generating Context-Aware Error Explanations with Large Language Models. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (Portland, OR, USA) (SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 1314–1320. <https://doi.org/10.1145/3626252.3630822>
- [144] Gareth Terry, Nikki Hayfield, Victoria Clarke, Virginia Braun, et al. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2, 17–37 (2017), 25.
- [145] Andreas Tolk, Philip Barry, Margaret L Loper, Ghaith Rabadi, William T Scherer, and Levent Yilmaz. 2023. Chances and Challenges of ChatGPT and Similar Models for Education in M&S. In *2023 Winter Simulation Conference (WSC)*. IEEE, 3332–3346.
- [146] Andrew Tran, Kenneth Angelikas, Egi Rama, Chiku Okechukwu, David H Smith, and Stephen MacNeil. 2023. Generating multiple choice questions for computing courses using large language models. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–8.
- [147] Minh Tran. 2023. Prompt Engineering for Large Language Models to Support K-8 Computer Science Teachers in Creating Culturally Responsive Projects. In *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 2*. 110–112. <https://doi.org/10.1145/3568812.3603453>
- [148] Annapurna Vadaparty, Daniel Zingaro, David H. Smith IV, Mounika Padala, Christine Alvarado, Jamie Gorson Benario, and Leo Porter. 2024. CS1-LLM: Integrating LLMs into CS1 Instruction. In *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 1*. 297–303. <https://doi.org/10.1145/3649217.3653584>
- [149] Marcel Valový and Alena Buchalceva. 2023. The Psychological Effects of AI-Assisted Programming on Students and Professionals. In *2023 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. 385–390. <https://doi.org/10.1109/ICSME58846.2023.00050>
- [150] Sander Valstar, Caroline Sih, Sophia Krause-Levy, Leo Porter, and William G. Renswold. 2020. A Quantitative Study of Faculty Views on the Goals of an Undergraduate CS Program and Preparing Students for Industry. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 113–123. <https://doi.org/10.1145/3372782.3406277>
- [151] Yoshija Walter. 2024. Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 15.
- [152] Mo Wang, Minjuan Wang, Xin Xu, Lanqing Yang, Dunbo Cai, and Minghao Yin. 2024. Unleashing ChatGPT's Power: A Case Study on Optimizing Information Retrieval in Flipped Classrooms via Prompt Engineering. *IEEE Transactions on Learning Technologies* 17 (2024), 629–641. <https://doi.org/10.1109/TLT.2023.3324714>
- [153] Sierra Wang, John Mitchell, and Chris Piech. 2024. A Large Scale RCT on Effective Error Messages in CS1. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. 1395–1401. <https://doi.org/10.1145/3626252.3630764>
- [154] Tianjia Wang, Daniel Vargas-Diaz, Chris Brown, and Yan Chen. 2023. Towards Adapting Computer Science Courses to AI Assistants' Capabilities. *arXiv preprint arXiv:2306.03289* (2023).
- [155] Justin D. Weisz, Michael Muller, Steven I. Ross, Fernando Martinez, Stephanie Houde, Mayank Agarwal, Kartik Talamadupula, and John T. Richards. 2022. Better Together? An Evaluation of AI-Supported Code Translation. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 369–391. <https://doi.org/10.1145/3490099.3511157>

- [156] Juliette Woodrow, Ali Malik, and Chris Piech. 2024. AI Teaches the Art of Elegant Coding: Timely, Fair, and Helpful Style Feedback in a Global Course. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (*SIGCSE 2024*). Association for Computing Machinery, New York, NY, USA, 1442–1448. <https://doi.org/10.1145/3626252.3630773>
- [157] Hao Wu, Daidong Fa, Xiaoling Wu, Weijun Tan, Xiwen Chang, Ying Gao, and Jinta Weng. 2023. Research on the Construction of Intelligent Programming Platform Based on AI-generated Content. In *Proceedings of the 15th International Conference on Education Technology and Computers*. 9–15.
- [158] Ruiwei Xiao, Xinying Hou, and John Stamper. 2024. Exploring How Multiple Levels of GPT-Generated Programming Hints Support or Disappoint Novices. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 142, 10 pages. <https://doi.org/10.1145/3613905.3650937>
- [159] Yuankai Xue, Hanlin Chen, Gina R. Bai, Robert Tairas, and Yu Huang. 2024. Does ChatGPT Help With Introductory Programming? An Experiment of Students Using ChatGPT in CS1. In *Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training*. 331–341. <https://doi.org/10.1145/3639474.3640076>
- [160] Ramazan Yilmaz and Fatma Gizem Karaoglan Yilmaz. 2023. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence* 4 (2023), 100147.
- [161] Cynthia Zastudil, Magdalena Rogalska, Christine Kapp, Jennifer Vaughn, and Stephen MacNeil. 2023. Generative AI in Computing Education: Perspectives of Students and Instructors. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.
- [162] Zhengdong Zhang, Zihan Dong, Yang Shi, Thomas Price, Noboru Matsuda, and Dongkuan Xu. 2024. Students' perceptions and preferences of generative artificial intelligence feedback for programming. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23250–23258. <https://doi.org/10.48550/arXiv.2312.11567>
- [163] Albert Ziegler, Eirini Kalliamvakou, X Alice Li, Andrew Rice, Devon Rifkin, Shawn Simister, Ganesh Sittampalam, and Edward Aftandilian. 2024. Measuring GitHub Copilot's impact on productivity. *Commun. ACM* 67, 3 (March 2024), 54–63.

Appendix

A Educator Survey Questions

- Q1 Do you explicitly disallow students to use GenAI tools for your computing courses (within the last 12 months)?
- Yes, I explicitly disallow students to use GenAI tools (enables Q9–11)
 - No, I do not explicitly disallow students to use GenAI tools (condition on Q20)
- Q2 Are you incorporating GenAI tools (e. g., actively integrating it into the curriculum or exercises) into your recent courses (within the last 12 months)?
- Yes (enables Q12–Q19)
 - No (condition on Q20)
- Q3 Do you believe the skills to create software have changed after the advent of GenAI tools?
- Yes (proceed with Q5)
 - No (proceed with Q4)
- Q4 Please elaborate on your last response why skills have not changed. (open question, depends on Q3=No)
- Q5 In what ways do you think the skills needed to create software have changed with the introduction of GenAI tools? (open question, depends on Q3=Yes)
- Q6 When using GenAI tools to create (parts of) software, which skills become the most important? (please drag and order your top 3)? (depends on Q3=Yes)
- Integration skills
 - Testing
 - Reading code
 - Problem decomposition
 - Problem solving
 - Modifying code
 - Prompt engineering
 - Problem understanding
 - Developing algorithms
 - Debugging
 - Understanding error messages
 - Other, please specify (open question)
- Q7 How often do you believe professional software engineers are using GenAI tools as part of their professional role?
- Never
 - Rarely
 - Sometimes
 - Routinely
 - Everyday
- Q8 Please select the tasks or contexts you think industry professionals are using GenAI tools.
- Modeling algorithms
 - Getting started with a problem
 - Generating ideas
 - Generating code
 - Modifying code
 - Autocompleting code
 - Debugging
 - Creating documentation/comments
 - Finding resources/documentation/libraries
 - Providing code examples
 - Generate test cases
 - Other, please specify: (open question)
- Block condition (for Q9–Q11): depends on Q1=Yes
- Q9 Why don't you allow for GenAI tools use in your courses?
- Q10 Are you doing anything to prevent GenAI tools' use in your course?
- No
 - Yes (enables Q11)
- Q11 If yes, what are you doing? (open question, depends on Q10=Yes)
- Block condition (for Q12–Q19): depends on Q2=Yes
- Q12 We ask you to think of a recent course (within the last 12 months) that you teach that is most influenced by GenAI tools, and respond to the following questions based on this course. Please enter the course name. (open question)
- Q13 Select the size of the recent course that you teach that is most influenced by GenAI tools:
- 1–10
 - 11–25
 - 26–50
 - 50–100
 - 101–250
 - 250+
- Q14 Who uses (or is expected to use) GenAI tools in your course(s)?
- Only instructor/instructional staff (proceed with Q16)
 - Students (proceed with Q15)
 - Both (proceed with Q15)
- Q15 If students are allowed to use GenAI tools, how do you expect students to access them (depends on Q14=Students or Q14=Both)
- N/A
 - Publicly free version
 - Paid version (free for students)
 - Paid version (students are expected to pay)
 - Paid version (paid by institution)
 - Custom tool available to everyone
- Q16 Which type of GenAI tools are you incorporating into your recent course that is most influenced by GenAI tools?
- Standard industry tools, e.g., ChatGPT, Copilot
 - Customized tools created by others
 - Customized tool created by myself
- Q17 In what ways have you incorporated GenAI tools into your recent course that is most influenced by GenAI tools?
- To support grading
 - To automatically provide feedback to students using a custom tool
 - To support the correction of student work
 - To teach students about using GenAI tools
 - As educational content generator for teaching material
 - To validate the quality of assignments
 - Other, please specify: (open question)
- Q18 Why have you incorporated GenAI tools into your recent course? (open question)
- Q19 Have you changed any of the learning objectives of your recent course based on the capabilities of GenAI tools?
- Yes
 - No
- Q20 Why have you not incorporated GenAI tools (e.g., actively integrating it into the curriculum or exercises) into your recent courses (within the last 12 months)? (open question, depends on Q1=No and Q2=No)
- Q21 Please describe any changes you have made to your teaching approaches in courses you are teaching as a result of GenAI tools. (open question)

Q22 Please describe any changes you have made to your assessment approaches in courses you are teaching as a result of GenAI tools. (open question)

Demographic Questions for all participants:

Q23 In which country are you employed? (select from dropdown list with all countries of the world)

Q24 How would you characterize your institution?

- Primary
- Secondary
- 2-year college (Associates)
- Vocational School
- College (bachelor's degree granting)
- University (graduate degree granting)
- Other, please specify: (open question)

Q25 Please provide the name of the institution you are currently teaching. (open question)

Q26 Do you teach at an institution that serves a minority population in your country?

- Yes
- No
- Unsure/doesn't apply in my country

Q27 What course/area do you primarily teach or identify with?

- CS 1 – Introduction to Programming
- Software Engineering
- Artificial Intelligence (ML/Intelligent Systems)
- Human-Computer Interaction
- Networking and Communications
- Architecture and Organization
- CS 2 – Introduction to Data Structures
- Information Assurance and Security
- Graphics and Visualization
- Information Management
- Software Development Fundamentals
- Parallel and Distributed Computing
- Platform-based Development
- Operating Systems
- Computational Science
- Teacher Preparation (for teaching CS to ages 5–18)
- Programming Languages
- Systems Fundamentals
- Discrete Structures
- Social Issues and Professional Practice
- Robotics
- Algorithms and Complexity
- Other

Q28 How many years have you been teaching for? (open question)

Q29 What is the gender you identify yourself with?

- Female
- Male
- Non-binary or gender diverse
- Prefer not to disclose
- Prefer to self-describe (open question)

Q30 Would you be willing to be interviewed in more detail?

- No
- Yes - please enter your email (open question)

B Developer Survey Questions

Q1 Do you use GenAI tools in your professional role developing software?

- Yes (proceed with Q2)
- No (proceed with Q14)

Q2 How often do you use GenAI tools as part of your professional role on average?

- Several times a day
- Once a day
- Several times per week
- Once a week
- Once a month
- Once a year

Q3 What types of GenAI tools do you use?

- Chatbot (e.g., Chatgpt, Gemini)
- Autocomplete code (e.g., Copilot)
- Other, please specify: (open question)

Q4 Can you please describe how you use GenAI AI tools in your professional work as a software developer? (e.g., give 2-3 examples of situations in which you apply them) (open question)

Q5 Please select the tasks or contexts for which you generally use GenAI tools.

- Modeling algorithms
- Getting started with a problem
- Generating ideas
- Generating code
- Modifying code
- Autocompleting code
- Debugging
- Creating documentation/comments
- Finding resources/documentation/libraries
- Providing code examples
- Generate test cases
- Other, please specify: (open question)

Q6 How not useful or useful do you feel GenAI tools have been to your software development?

- Not useful
- A little useful
- Moderately useful
- Quite useful
- Very useful

Q7 Have GenAI tools made your software development more or less efficient?

- Much less efficient
- Less efficient
- No change
- More efficient
- Much more efficient

Q8 How not harmful or harmful do you feel GenAI tools have been to your software development?

- Not harmful
- A little harmful
- Moderately harmful
- Quite harmful
- Very harmful

Q9 If you consider GenAI tools harmful, please describe a respective situation you have experienced, e.g., what were you doing, what did you expect, why was the use of the GenAI tools harmful and to whom? (open question)

Q10 Did the competencies (i.e., knowledge, skills, dispositions in context of a task) required to professionally develop software change with the availability of GenAI tools?

- No change
- Slight change
- Moderate change
- Extreme change

- Q11 If you have seen changes, from your experience with GenAI tools, what do you believe are new relevant competencies to professionally develop software with GenAI tools? (open question)
- Q12 If you have seen changes, from your experience with GenAI tools, what do you believe are competencies that are no longer or less relevant to professionally develop software with GenAI tools?
- Q13 What advice would you give to novice programmers regarding the use of GenAI tools? (open question)
- Q14 What is the reason that you do not use GenAI tools for professional software development?
- My company does not let me
 - I do not believe they will help me code better
 - I am concerned about ethics issues (e.g., privacy)
 - Other, please specify: (open question)
- Q15 Please feel free to elaborate on your reasoning. (open question)
- Q16 What advice would you give to novice programmers regarding the use of GenAI tools? (open question)

Demographic Questions for all participants:

- Q17 In which country are you employed? (select from dropdown list with all countries of the world)
- Q18 What is your job title?
- Software developer
 - Product Manager
 - Research Engineer
 - Other, please specify: (open question)
- Q19 What is the type of company where you are employed?
- Start-up (5 engineers or less)
 - Small software company (10 engineers or less)
 - Medium software company (50 engineers or less)
 - Large software company (more than 50 engineers)
 - Non-profit
 - Non-software focused company
 - Government
 - Research Institute
 - Other, please specify: (open question)

C Tool Creators Interview Questions

- Q1a Please give us a description of the GenAI tool you created.
- Q1b Who is using it?
- Q1c What did you learn throughout the process of deployment, and what modifications did you make along the way?
- Q1d What are the privacy and security policies required by your institution for the deployment of the tool?
- Q2 What outcomes have you seen from usage of the tool?
- Q3 What data are you collecting from the tool?
- Q4 What learning objectives are you hoping to reinforce through your tool?
- Q5 What expertise is required from educators to be able to deploy and use your tools in their courses?
- Q6 Please share your future development plans.
- Q7a How do you think AI should be used in CS education to improve teaching and learning?
- Q7b What tools need to be created?
- Q8 Is there anything else you want to discuss about GenAI in your courses?

D Educators Studying GenAI Interview Questions

- Q1 Describe your understanding of categories of GenAI research in CS education and its landscape.
- Q2 What do you hope to learn from studying GenAI in CS education?

- Q3 What AI technologies are currently used by educators for what?
- Q4 Have you encountered methods to prevent the negatively perceived aspects of GenAI? If so, please elaborate.
- Q5 Have you seen inequities in utilizing GenAI in CS education?
- Q6 Have you seen skills/competencies change ever since LLMs emerged? (positively or negatively) And if so, how?
- Q7 Where do you think GenAI in CS education is going to take us next?

E Educators using GenAI Interview Questions

- Q1 What course(s) are you using GenAI in?
- Q2a How are you using GenAI tools in your courses? What's the rationale/why?
- Q2b Are AI tools included in the syllabus (or planned for the next one)?
- Q2c Are you transparent with students about the use of GenAI in the course?
- Q3 Have you changed the learning objectives of course(s)? How?
- Q4a What were the outcomes (positive/negative) of GenAI use in your courses so far?
- Q4b Have you seen any equity issues due to GenAI? If so, can you provide an example? Are you generally concerned about these issues?
- Q5 Are you planning on expanding the use of GenAI in your course(s)? What about others at your institution?
- Q6 Is there anything else you want to discuss about GenAI in your courses?
- Q7 Is there a particular memorable moment in your course regarding GenAI?

F Extraction Survey Questions

Include paper?

- Description: This paper has been originally classified for inclusion. However, now that you have read the paper in detail if you think this is incorrect (and the paper does not meet the inclusion criteria) then there is no need to extract the data.
- Options:
 - Yes
 - No, too short
 - No, K-12
 - Not genAI
 - Not computing education
 - No intervention
 - No empirical human evidence

Bibtex entry (note, use authorYEARword format)

- Description: Please provide a complete bibtex entry – if possible, please use the entry from the ACM DL as this is typically very complete.

Data source

- Description: A “Supervised” study is a study that is conducted in a research lab (i.e. a highly controlled environment), while an “Un-supervised” study is a study that is conducted in a less restricted environment (such as online or where participants are not supervised).
- Options:
 - Supervised study (lab study, observations, etc.)
 - Unsupervised study (no human overseeing use)
 - Other

Author affiliation (type)

- Options:
 - Academic
 - Industry

Author affiliation (country)

- Description: Comma separated list of institutions' countries

Human participants (country)

- Description: If data is collected from human participants, provide a comma separated list of countries of where participants were located (if not explicitly mentioned, put "unclear")

Human participants (level)

- Options:
 - Tertiary education (e.g. college, university)
 - Informal education (e.g. MOOCs)
 - K-12 (in addition to tertiary)
 - Other

Number of human participants from whom data was collected (if available, type into other)

- Options:
 - Unclear
 - Other

Description of participants

- Description: A copy-paste (or paraphrased) description of participants from whom data is collected which may be useful to a more detailed thematic analysis. This information can often be found at the beginning of a Methods section.

How do the authors motivate the work?

- Description: A copy-paste (or paraphrased) description of the motivation for the work as expressed by the authors.

What LLM / tool is used?

- Options:
 - GPT-3
 - GPT-3.5
 - GPT-4
 - GPT-4o
 - Codex
 - GitHub Copilot
 - Gemini
 - Claude
 - Multimodal model
 - Open-source model
 - Unclear
 - Other

What are the explicit research questions / research goals / hypotheses in the article?

- Description: A copy-paste (or paraphrased) description of the RQs, goals, hypotheses

What programming languages are involved in the study?

- Options:
 - Java
 - Python
 - C
 - C++
 - Not programming language focused
 - Other

How does the article evaluate the data collected?

- Options:
 - Qualitatively
 - Quantitatively

Quality assessment

- Description: An assessment of the research "quality". For the last question, on threats to validity / limitations: code as "Yes" if there is an explicit (sub)section, "Vague" if they are mentioned as part of some other section (i.e. Discussion, Conclusions), or "No" if they are not mentioned.
- Sub-questions:
 - Is there a clearly defined research question/hypothesis?
 - * Yes
 - * No
 - * Vague / Unclear
 - Is the research process clearly described?
 - * Yes
 - * No
 - * Vague / Unclear
 - Are the results presented with sufficient detail?
 - * Yes
 - * No
 - * Vague / Unclear
 - Are threats to validity / limitations addressed in an explicit (sub)section?
 - * Yes
 - * No
 - * Vague / Unclear

What is the contribution / what are the key results of the article?

- Description: Provide a short summary of the main findings.

How Instructors Incorporate Generative AI into Teaching Computing?

- Description: Provide information on how generative AI was incorporated into teaching in the article (if applicable).

And why do they incorporate GenAI tools that way?

- Description: Provide information on why generative AI was incorporated into teaching in the article (if applicable).

How have the expectations towards skills in software development changed with the use of Generative AI?

- Description: Provide any potential changes to expectations towards skills in software development that could result from this work or were discussed in this work. When answering this question, please feel free to provide some of your own commentary – it is fine to mention changes which the paper prompted you to think about, even if they aren't explicitly mentioned by the paper authors.

Which computing competencies are required in the future?

- Description: Provide anything interesting for computing competencies of the future that could result from this work or were discussed in this work. When answering this question, please feel free to provide some of your own commentary – it is fine to mention changes which the paper prompted you to think about, even if they aren't explicitly mentioned by the paper authors.

Limitations of the study

- Description: Add any important limitations of the study (potentially mentioned by the authors, or just noticed by you).

Additional notes

- Description: Can be used to note any interesting aspects of the paper or anything else relevant that isn't captured in the extraction fields above.