



Bridging the Gap: Promoting Mindful and Reflective Use of Generative AI in Undergraduate Engineering and Computing Education

Ethan Dickey (CS)
Andres Bejarano (CS)
Rhianna Kuperus (IDA+A)
Bárbara Fagundes (CIE)

ENE Seminar, Sep 2025



Department of Computer Science

Empirical work on pedagogy

Generative AI: Yay!

- What's GenAI good for? (2 minutes)

- Here to stay – rapid adoption in industry and education
- Pervasive in student life – $\geq 75\%$ student usage on homework
- Tensions for educators:
 - Academic integrity vs innovation
 - Overreliance risk vs support opportunities
- The “Junior Year Wall” – concern that early dependence will prevent later skill development



Department of Computer Science

2

Generative AI is here to stay. It's no longer a question of if students will use these tools, but how they are using them.

--here to stay

But for us as educators, this brings real tensions. On one hand, there's concern about academic integrity – plagiarism, dishonesty, the shortcut temptation. On the other, there's real potential for innovation in how students learn and create. Similarly, while some worry about overreliance, these same tools can provide meaningful support, especially for students who might otherwise struggle alone.

--junior year wall

The challenge isn't whether AI is good or bad – it's how we design learning experiences that let students use it productively, while still developing durable skills.

“Who cares? Just make written exams 100% of the final grade”

- Oversimplified “solution” to GenAI in coursework
- Ignores diverse learning goals (design, creativity, teamwork)
- Exacerbates inequities (test anxiety, accessibility)
- Misses chance to teach how to use GenAI responsibly



Department of Computer Science

3

Some argue the easy fix is to move everything back to written exams. But that oversimplifies the problem. Exams can test certain skills, but they ignore broader learning outcomes – design, creativity, collaboration. They also amplify inequities, like test anxiety or barriers for students with disabilities. And most importantly, they sidestep the educational responsibility we have to teach students how to use GenAI responsibly, rather than pretend it doesn’t exist.

Proposed Solution: The AI-Lab Intervention

A structured, scaffolded single-lecture intervention. Practical, immediate implementation.

Research Questions:

1. How does structured scaffolding (AI-Lab) affect student use of GenAI?
2. How does it affect student perceptions and comfort?
3. Does it promote mindful, reflective engagement?



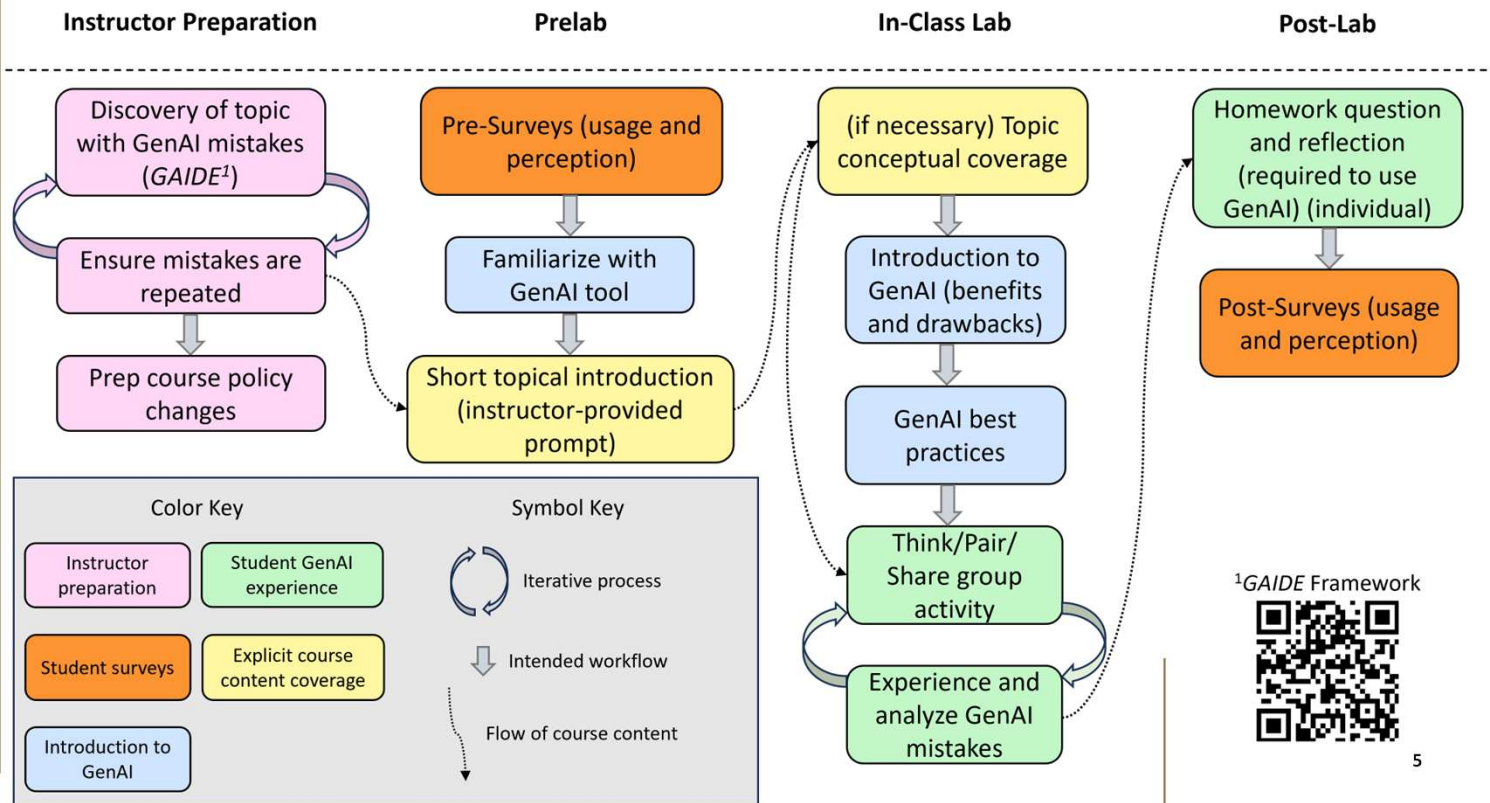
Department of Computer Science

4

Question is not, “is AI good or bad,” but rather, “how can students integrate it with learning.”

Proposed Solution: The AI-Lab Intervention

AI-Lab Framework for Integrating GenAI in Programming Courses



Short – pre/post activity surrounding one lecture. Instructors choose a topic where GenAI fails (admittedly harder to find nowadays), students prepare, then in classtime we cover best practices, benefits and drawbacks, and after gamifying finding errors in course-topical content in student pairs, we share results and strategies. Follow-up homework requires tool usage and critical reflection. This scaffolding is designed to make students both experiment and reflect, bringing in both extremes – causing those who always want to use it to recognize pitfalls and those who never want to to recognize benefits.

STUDY DESIGN AND METHODOLOGY

- Setting: Purdue CS and Engineering undergraduate courses
- Courses:
 - 2x Data Structures (core 2nd-year, separately for CS and DS majors)
 - Competitive Programming (elective, 2nd-year)
 - ENGR130 (FYE, *Transforming Ideas Into Innovations*)
- n=831, 2 semesters (S24, F24)
- Mixed methods:
 - Pre/post surveys (usage/perception)
 - Wilcoxon + effect sizes
 - Focus groups (5-7 weeks later) + thematic analysis
- Data validation:
 - Excluded students who didn't complete both surveys
 - Excluded students who dropped the course



Department of Computer Science

6

We implemented this intervention over two semesters and four courses. During F23, we had a trial run that let us iron out a lot of the kinks. We totaled 831 valid responses and used a mixed-methods design: surveys, statistical analysis, and focus groups. Survey responses were assigned identifiers to allow us to use paired (non-parametric) tests and responses were dropped if the student didn't complete both pre and post.

FINDINGS: QUANT (N=831)

- GenAI usage remained stable

Question	Shapiro-Wilk	Wilcoxon Signed Rank		Rank Biserial Correlation	
	p-value	W	p-value	r_{rb}	Interpretation
How frequently do you use GenAI to get help with programming projects?	0.0000	24834.5	0.9730		
How frequently do you use GenAI to get help with debugging?	0.0000	19618	0.0239	0.9339	Large
How frequently do you use GenAI to get help with homework problems?	0.0000	21427	0.6806		

Table 14. P-values, Rank Biserial Correlation, and interpretations of significance of change between pre- and post- perception and usage surveys for combined S24 and F24 semesters across CP, DSA-CS, and DSA-DSAI.



Department of Computer Science

7

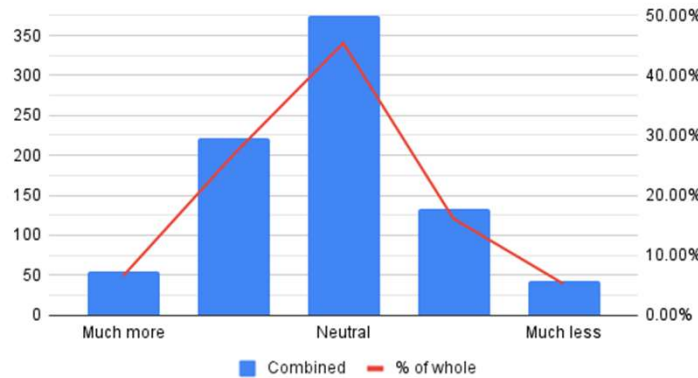
Interestingly, students didn't show an increase in GenAI usage – frequency remained mostly stable. We saw shifts in comfort and openness with large effect sizes, and large shifts in usage for debugging, suggesting students began to see AI as a more intentional educational support tool.

As we did not see any statistically significant change in frequency of use on programming projects and homework problems, we infer that students were already using these tools with some caution. If there had been large numbers of students at either extreme, overreliance or zero beneficial usage, they would have both shown up in the focus groups and shown up as moving data points in the statistical analysis.

FINDINGS: QUANT

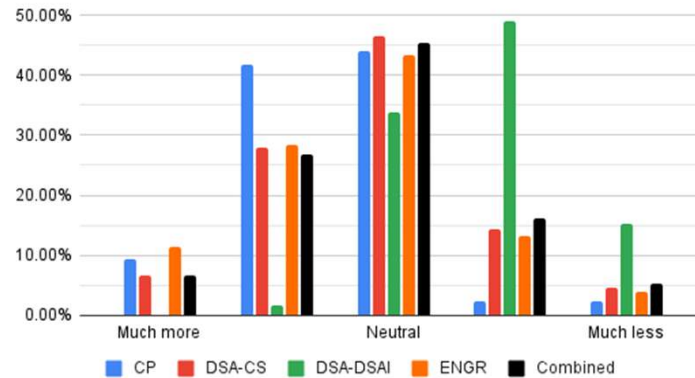
- GenAI usage remained stable
- Desire to use GenAI was different per course

Compared to before the AI Lab, how much do you want to use GenAI?



(a) Combined across all courses and semesters.

Compared to before the AI Lab, how much do you want to use GenAI?



(b) Normalized, shown for each individual course.

Fig. 1. Self-reported change in desire to use GenAI



Department of Computer Science

Each course had a different average perceived shift in their desire to use GenAI. Explicitly, CP, DSA-CS, and ENGR skewed towards more usage, while DSA-DSAI skewed towards less usage.

Based on student interactions and focus groups, we hypothesize that DSA-DSAI, as DS/AI majors, were likely to be overreliant on GenAI coming into the intervention, due to their specific major structure and the perceived difficulty of their classes. On the other hand, CS and ENGR majors were more cautious to begin with, erring on the side of under-use, whereas we wanted to encourage all students to use these tools for their educational benefit.

Future studies are necessary to better analyze differences between courses, as this study was primarily designed to measure the overall impact and success of the AI-Lab intervention on CS undergraduates.

FINDINGS: QUANT

- GenAI usage remained stable
- Desire to use GenAI was different per course
- Large effect sizes: comfort and openness
- Notable shift in usage: debugging

Question	Shapiro-Wilk	Wilcoxon Signed Rank		Rank Biserial Correlation	
	p-value	W	p-value	r_{rb}	Interpretation
How open are you to using GenAI to get help with homework problems?	0.0000	29882.5	0.0140	0.9006	Large
How open are you to using GenAI to get help with debugging?	0.0000	32020	0.6592		
How open are you to using GenAI to get help with conceptual questions?	0.0000	19346	0.0008	0.9357	Large
How comfortable are you using GenAI to help with conceptual questions?	0.0000	25323	0.0000	0.9158	Large
How comfortable are you using GenAI to help with homework problems?	0.0000	35840.5	0.0118	0.8808	Large
How comfortable are you using GenAI to help with debugging?	0.0000	35282.5	0.0009	0.8827	Large
How frequently do you use GenAI to get help with programming projects?	0.0000	24834.5	0.9730		
How frequently do you use GenAI to get help with debugging?	0.0000	19618	0.0239	0.9339	Large
How frequently do you use GenAI to get help with homework problems?	0.0000	21427	0.6806		

Table 14. P-values, Rank Biserial Correlation, and interpretations of significance of change between pre- and post- perception and usage surveys for combined S24 and F24 semesters across CP, DSA-CS, and DSA-DSAI.



Department of Computer Science

9

Throwing a lot at you here, let's break it down.

Interestingly, students didn't show an increase in GenAI usage – frequency remained mostly stable. We saw shifts in comfort and openness with large effect sizes, and large shifts in usage for debugging, suggesting students began to see AI as a more intentional educational support tool.

One especially interesting shift was in debugging. Students reported not just higher comfort but also changes in their frequency of use. This is important because debugging is rarely taught explicitly in CS curricula. Students are expected to learn it on their own, usually by trial and error. Our intervention gave them a structured way to explore GenAI as a debugging partner, which they then carried over to their coursework.

FINDINGS: QUAL (5-7 WEEKS AFTER)

- Casual to strategic prompting

“Usually whenever I try to use ChatGPT, I sort of just give up on it really fast if it doesn’t give me the right answer.”

“Now I give it, for example, everything I do know”

“Before we hadn’t really been given guidelines on how to use it. . . I thought it was helpful in guiding us, giving different examples of how it could be applied.”

“I felt like it did much better once it was given clear rules. . . especially if I would explicitly dictate what I wanted it to tell me.”



Department of Computer Science

10

Focus groups confirmed what we observed in the surveys. I’ll show some representative quotes for each theme we found. Students primarily described moving from a trial and error prompting style to more strategic, context-aware prompting. Before the AI-Lab, many students described their use as casual or even frustrating:”” Afterward, they spoke about incorporating context, refining prompts, and treating it as a back-and-forth partner: “”

FINDINGS: QUAL

- Casual to strategic prompting
- Stable use, evolving rationale

“It’s pretty much the same. I just learned better ways to use it.”

“I don’t think I use it more, but I definitely use it better. Now I know how to refine my prompts instead of wasting time rewriting them over and over.”

“At some point, I felt like I was relying too much on ChatGPT. I didn’t want it to replace my own thinking, so I started using it less.”



Department of Computer Science

11

While usage didn’t spike, their rationale evolved and deepened. This first quote captures the essence of what we found in the quant data.

FINDINGS: QUAL

- Casual to strategic prompting
- Stable use, evolving rationale
- Awareness of GenAI's limits

"A lot of the time in certain areas like statistics or linear algebra, it gives incorrect answers. It's obviously incorrect."

"It would make it look really, really right if you lack the sense of judgment. . . I couldn't tell whether you're accurate or not."

"[GenAI] just told me... [but] I cannot access articles through links. It was just funny how... really obvious things beyond its capability."



Department of Computer Science

12

Their awareness of limitations increased:

Many students became more skeptical and critical of GenAI's limitations, especially in areas like math, statistics, or coding. The growing skepticism represented in the quotes is exactly what we hoped for.

FINDINGS: QUAL

- Casual to strategic prompting
- Stable use, evolving rationale
- Awareness of GenAI's limits
- GenAI tools as support, not replacement

"I basically use it to break down hard concepts into simpler terms."

"It's sort of like questions I'd ask a TA. I just ask Chat first."

"I primarily use it for research... it's been really good for taking notes."

"It becomes useful in your programming... I'll ask, what's an example code using this documentation?"



Department of Computer Science

13

Many students described using GenAI as a way to scaffold understanding, clarify unfamiliar concepts, or explore new ideas. Rather than treating it as a shortcut, several participants emphasized that they used GenAI as a starting point, which is something closer to a study partner or TA rather than a solution generator.

FINDINGS: QUAL

- Casual to strategic prompting
- Stable use, evolving rationale
- Awareness of GenAI's limits
- GenAI tools as support, not replacement
- Ethical concerns and tensions

"I don't want it to just generate content for me. I want to make sure I'm still learning."

"I reduced using it... I just feel like I'm getting more dumb... a puppet of ChatGPT."



Department of Computer Science

14

A recurring theme in students' reflections was the tension between leveraging GenAI to support their thinking and maintaining a sense of personal ownership over their learning. While some appreciated the clarity and productivity boost that GenAI offered, others worried that it might compromise their growth or even alter their perception of themselves as learners.

These tensions point to a deeper level of reflection among students, not only about the effectiveness of GenAI but also about what constitutes authentic learning, fairness, and academic identity. While views varied, it was clear that many were still negotiating how to integrate GenAI into their academic lives in a responsible manner.

The amount of discussion and thought we saw from students embodied the full range of opinions and concerns that we've seen from faculty and staff. They are just as aware, as a collective, as we are of the benefits and downsides of AI usage in education.

Together, these themes suggest that students were generally open to using GenAI but approached it with intentionality and care. Their responses revealed a willingness to experiment, a growing understanding of its limitations, and a desire to remain in control of their own learning journey.

KEY INSIGHTS

- Frequency of use remained stable, but quality improved
- Scaffolding led to mindful engagement
- No evidence of increased cheating
- Growth of “AI literacy” and boundaries



Department of Computer Science

15

[frequent concern] If the intervention had spurred widespread “cheating” behaviors, we would have expected significant increases in GenAI usage on graded work. Instead, not only did we fail to observe any statistical rise in frequency of use for homework or projects (graded assessments), the only notable (albeit small) change was actually a slight shift from “moderate” or “very frequent” usage down to “on occasion.” In other words, rather than pushing students toward reliance on AI for answers, the intervention appears to have encouraged a more restrained and reflective approach and reinforced that these tools should supplement rather than replace their own intellectual efforts.

IMPLICATIONS FOR EDUCATORS

- Use short, structured interventions
- Require generating outputs and reflection
- Critical: try to find tasks that reveal limitations
- Normalize transparency and ethical framing (“when should we use it?”)



Department of Computer Science

16

What does this mean for educators? First, interventions don’t need to be long to be impactful. This took up 30-40 minutes of one lecture, plus short pre-lecture and homework activities. Required reflection gave students permission to explore good and “bad” usage of the tools, and an educational space where experimenting and talking with peers about what would be normally considered “cheating” is ok.

Critically, it is important to find a task where LLMs make mistakes. [180 example]

Again, creating a space to try and talk about these things allows students to develop richer understandings of tool usage.

FUTURE DIRECTIONS

- Longitudinal outcomes (“Junior-Year Wall”)
- Cross-institutional studies
- Differences between disciplines
 - Math GTAs
- Comparison with other interventions



Department of Computer Science

17

-- Directions

Conclude with this:

GenAI is not inherently harmful – how you integrate it matters

The AI-Lab provided mindful, reflective engagement

Students ended up more intentional, not more dependent

COLLABORATORS



Andres Bejarano
Assistant Teaching
Professor
CS Department



Rhianna Kuperus
Senior Data Analyst,
IDA+A



Bárbara Fagundes
Postdoc with CIE

This work was supported by the Purdue Innovation Hub (H-AI-23002) and conducted in accordance with protocol IRB-2023-1079.



Department of Computer Science

THANK YOU! QUESTIONS?

AI-Lab Instructor Resources (slides, guides, handouts) available on
Ethan Dickey's Purdue website:
<https://www.cs.purdue.edu/homes/dickeye/index.html#resources>



Department of Computer Science

Appendix

Y'all had a lot of good questions

Student Population Breakdown

Perception	
Course	
CP	43
DSA-CS	676
DSA-DSAI	59
Total	778
ENGR	53
Grand Total	831

Table 1. Perception survey totals, combined courses and semesters.

Usage	
Course	
CP	43
DSA-CS	671
DSA-DSAI	59
Total	773
ENGR	53
Grand Total	826

Table 2. Usage survey totals, combined across courses and semesters

Perception			
Course	S24	F24	Grand Total
CP	-	43	43
DSA-CS	129	547	676
DSA-DSAI	59	-	59
Total	188	590	778
ENGR	-	53	53
Grand Total	188	643	831

Table 17. Perception survey totals, broken down per course and semester.

Usage			
Course	S24	F24	Grand Total
CP	-	43	43
DSA-CS	127	544	671
DSA-DSAI	59	-	59
Total	186	587	773
ENGR	-	53	53
Grand Total	186	640	826

Table 18. Usage survey totals, broken down per course and semester.

We also analyzed the differences across demographics like gender, race/ethnicity, residency, and classification(sophomore/junior/senior) and found no significant differences between these groups.

“AI LITERACY”

- “A set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace.” [1]
- “AI literacy includes the knowledge and skills that enable people to critically understand, evaluate, and use AI systems and tools to safely and effectively participate in an increasingly digital world.” [2]
- “AI literacy represents the technical knowledge, durable skills, and future-ready attitudes required to thrive in a world influenced by AI. It enables learners to engage, create with, manage, and design AI, while critically evaluating its benefits, risks, and ethical implications.” Also, “AI literacy empowers learners to understand AI and make decisions about its use in meaningful and ethical ways. When integrated thoughtfully into learning, AI provides new opportunities to exercise critical thinking and creative expression.” [3]

[1] Long, Duri, and Brian Magerko. "What is AI literacy? Competencies and design considerations." Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.

[2] Mills, Kelly, et al. "AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology." Digital Promise (2024).

[3] European Commission and OECD. 2025. Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education. OECD (in collaboration with the European Commission and Code.org). Draft homepage accessed on 2025-08-04.

F24 WITH AND WITHOUT ENGR

Question	F24 With ENGR				F24 Without ENGR			
	Wilcoxon Signed Rank		Rank Biserial Corr.		WSR		RBC	
	W	p-value	r_{rb}	Interpretation	W	p-value	r_{rb}	Interp.
Open to using for homework problems?	20420	0.0011	0.9014	Large	16905.5	0.0047	0.9030	Large
Open to using for debugging?	15433.5	0.0001	0.9255	Large	14131	0.0012	0.9189	Large
Open to using for conceptual questions?	13353.5	0.0000	0.9355	Large	10819	0.0002	0.9379	Large
Comfortable using for conceptual questions?	17965	0.0000	0.9132	Large	15549	0.0002	0.9114	Large
Comfortable using for homework problems?	24981.5	0.0098	0.8793	Large	21701	0.0592		
Comfortable using for debugging?	20500	0.0000	0.9010	Large	18270.5	0.0000	0.8952	Large
Frequency of use on programming projects?	20098	0.6582			17090.5	0.9096		
Frequency of use on debugging?	14054	0.0045	0.9315	Large	12310.5	0.0370	0.9287	Large
Frequency of use on homework problems?	15276	0.7413			13443	0.4879		

Table 16. P-values, Rank Biserial Correlation, and interpretations of significance of change between pre- and post- perception and usage surveys for F24 semester across CP, DSA-CS, and showing with and without ENGR. Shapiro-Wilk p-values are omitted for space as they are all 0.0000.

Data for ENGR alone was not available in digital format in time for this presentation.

DIFFERENCES BETWEEN SEMESTERS

Perception

Course	S24	F24	Grand Total
CP	–	43	43
DSA-CS	129	547	676
DSA-DSAI	59	–	59
Total	188	590	778
ENGR	–	53	53
Grand Total	188	643	831

Table 17. Perception survey totals, broken down per course and semester.

Usage

Course	S24	F24	Grand Total
CP	–	43	43
DSA-CS	127	544	671
DSA-DSAI	59	–	59
Total	186	587	773
ENGR	–	53	53
Grand Total	186	640	826

Table 18. Usage survey totals, broken down per course and semester.

How frequently do you use GenAI to get help with homework problems?

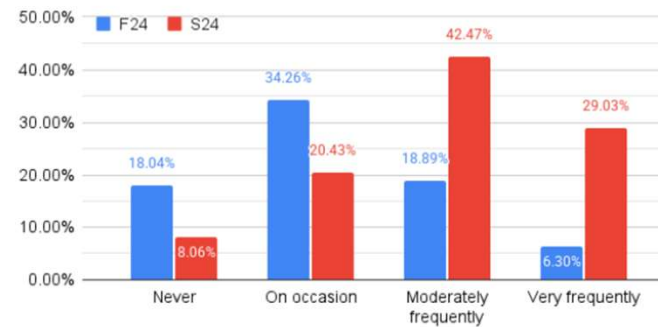
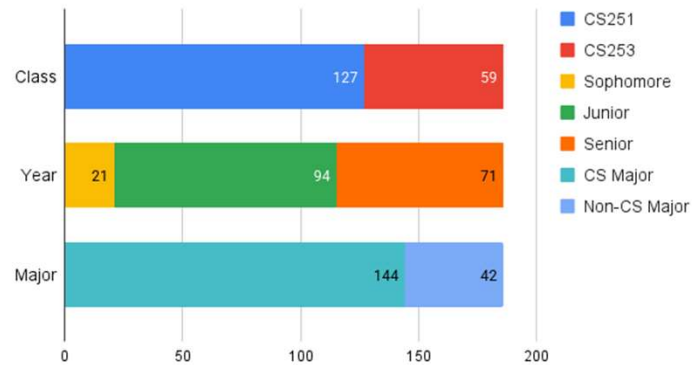


Fig. 5. A stark difference between Fall 2024 and Spring 2024, in percentage of students who used GenAI moderately or very frequently compared to never or on occasion. F24 shows much higher percentages in never or on occasion (52.30%) compared to S24 (28.49%). Note that populations were primarily composed of DSA-CS students in both semesters. See Section 4.2 for analysis.

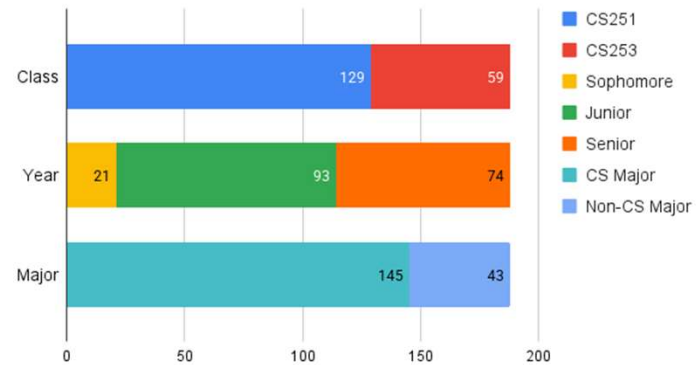
See Section 4.2: “Differences between semesters” in our “Evaluating the AI-Lab...” paper for reflection on this slide (preprint on arXiv, submitted for publication).

MISC CHARTS

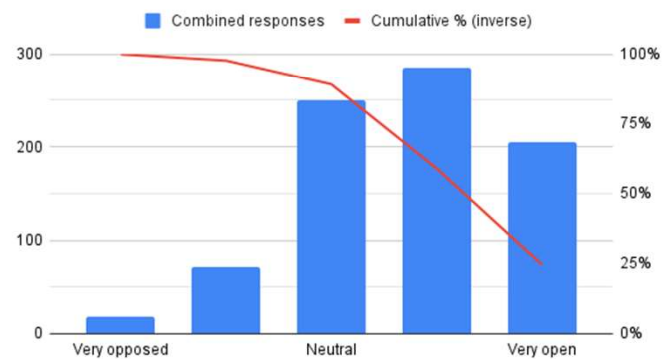
Class, Year, and Major (Usage Survey, n=186)



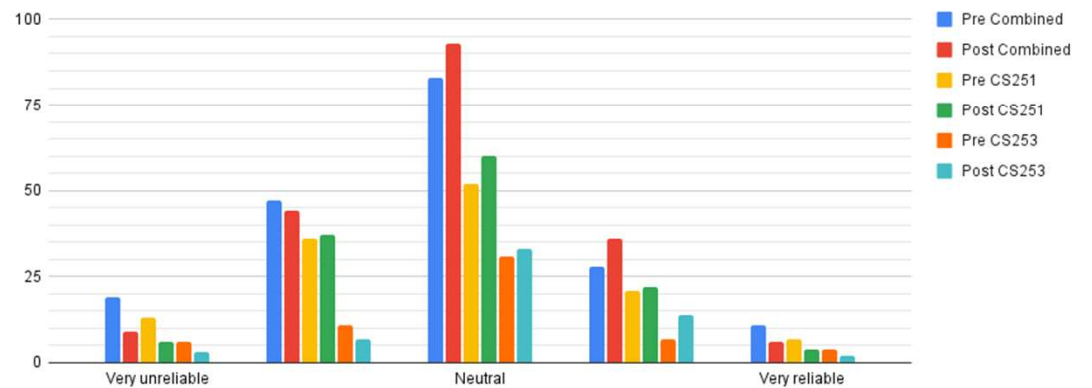
Class, Year, and Major (Perception Survey, n=188)



In general, how open are you to GenAI being allowed in a collegiate class?



How reliable do you think GenAI is for most problems you have in programming classes?



QUESTIONS

- “Because of the potential for viewing AI usage as academic dishonesty, how did you handle this potential bias for self-reporting in your survey?”
 - 2 surveys, more likely to lie on the actual usage survey, limitations discussed as such.
- “How do we balance the ethical dilemma that is AI in education? And how do you feel about the comparison of AI to when the calculator was introduced?”
 - The comparison is decent on a surface levels, but fails on deeper levels. The calculator couldn’t apply its calculations to problems, LLMs can. As for the ethical dilemma, you’ll have to be more specific.
- “Could you provide a specific example from your student focus groups of what “mindful and deliberate engagement” looks like in practice?”
 - See focus group quotes. Specifically, we suggest using the tools in a way that compliments their learning (e.g., “here’s my problem, help me find the issue by asking me guiding questions only” or “provide suggestions for how to start without solving my problem”)

QUESTIONS

- “... For my degree [20 years ago], the tool professors contended with was parametric CAD modeling and related structural simulation tools. A similar approach was taken, to make sure we understand the fundamentals before using the tools. My question is about the early foundational phases of learning. I recently read that a Stanford professor, with the buy-in of the students to prevent cheating, actually required all work to be hand written. Similarly, when I studied mechanical engineering, most problem sets and exams were hand written. Do you think the C.S. department might implement a manual paper and pencil component to the curriculum possibly due to the challenge professors might encounter to continually design assignments that generate suboptimal responses from the rapidly advancing Gen-AI field? Thanks!”
- Short answer: no. Longer answer: The CS department has been really suffering under the load of new students and the rate of growth of the department (3000 undergraduate CS majors, freshmen and sophomore classes of sizes 800-1000+). Automation has been key to attempting to keep up with this, and grading handwritten assignments is a huge burden with so many students. Furthermore, manually checking source code for errors is very difficult, as each solution can look very different (there has been cyber security research done on “fingerprinting” people based on the style of their code; it’s very unique). What we’ve been seeing instead is a growth of reliance on in-person exams.

QUESTIONS

- “Given your finding that structure AI-Lab scaffolding shifts students toward more mindful GenAI use, without increasing frequency, how will you move beyond self reports to behavioral evidence (e.g. prompt logs, error fix traces, code provenance)?”
 - Great question. There has been some work in this space, namely Prather et al (2024, “The widening gap...”), who used observation, interview, and eye tracking; Ghimire and Edwards (2024, “Coding with AI...”), who used key logging, prompts, and AI replies; Denny et al. (2026, unpublished), who has a lot of work in this area; and several other unpublished works that will appear in SIGCSE 2026. In short, it’s fairly difficult to track this kind of behavioral data above 30-40 students. It requires fundamental software changes to major systems currently in place, and usually those are done by companies seeking to sell their programming homework software to universities (and therefore are inaccessible easily to researchers). Other assignment types (among them code traces, submitting prompts instead of work, and “find the AI bug”) are being explored by a large number of CSEd researchers. Usually, findings are hindered by lack of a common “student usage pattern” – students write code in many different ways, especially novices.

QUESTIONS

- “The development and mainstream adoption of certain tools often seems to lead to changes in how students are taught certain subjects. For example, the widespread availability of more powerful calculators (graphing calculators) led some math teachers to ban them in the classroom or on tests. Similarly, using digital encyclopedias was for a short time banned, an actual book reference was needed. (both of these examples age me lol) My one question for Mr. Dickey is this: In what critical ways do you feel that the emergence of students using AI tools, and the response by educators, differs from tools in the past that affected the performance of students, created an over-dependence, or lack of in depth understanding?”
- Yes and no. There are critical ways in which students using AI are different, but I do not believe they include performance of students, over-dependence, or lack of depth (they are all similarly bad to previous iterations). What will change as we move forward is what we value in student education. Do we really care that a student can program in assembly and can accurately implement a Red-Black tree? No. In computing education, we care about the problem solving process. The path to get to those solutions requires an in-depth understanding of how computers work, and teaches them to think algorithmically. Fundamentally, that is what we are teaching, which is why so many recruiters from other fields (finance, etc.) recruit CS students with no understanding of finance. Returning to the question, previously, students could get around some assignments by using a calculator (for example). Now, students can get around at least two whole years of collegiate education by using an LLM. The impact, and therefore educator responsibility and urgency to address, is much deeper. It allows students a lack of independent thinking.
As Leslie Lamport said, “coding is to programming what typing is to writing.”¹

¹Magazine Q. The Man Who Revolutionized Computer Science With Math. Accessed on: August 14, 2023. <https://www.youtube.com/watch?v=rkZzg7Vowao>. 2022.

QUESTIONS

- “This week’s question focuses on the future of teaching and the role that Generative AI will play, especially since these tools have not yet been widely adopted in applied settings such as the factory floor. As Generative AI becomes common in programming education, how can teachers adjust their methods so that students benefit from these tools while still building strong problem-solving and algorithmic thinking skills for advanced courses and future jobs?”
 - One answer is to address the source of the problem, rather than the effect. The AI-Lab is designed to address this source: that students are using these tools without restraint, nuance, or structured thought about if it’s right, if it’ll harm them, and how they could adjust their use. There are many other answers and proposed solutions to this. The short version is that we don’t really know what will or will not work as we move forward. Especially as these tools evolve into everything we dream and fear.

QUESTIONS

- “Should colleges leverage GenAI in teaching student? As we know GenAI is very popular now and most of the homework can be done very quickly with GenAI. So my question is that should college shift from heavy typical homework to how to use GenAI to find the correct answers?”
 - I don’t think this is an exclusive situation. We can maintain quality standards and homework load while integrating when and how to use GenAI. This decision also shouldn’t be at the college level, it should be at the instructor level, as all other course structure decisions are. (That doesn’t mean institutions can’t encourage and support one way or another, however.)
- “... Some of the framework seems to be designed around understanding there are shortcomings in most current GenAI. However, it seems the terminal velocity of how far this technology can go is not even close to where it will likely land. Do you feel that it is possible that GenAI will be able to understand when it has shortcomings, and not just commands, but intent, and will be able to prompt a user for guidance, or even try to help the user understand shortcoming in the wording of their own request in order to ensure it is responding correctly? ... it seems like targeting the shortfalls of GenAI might be like trying to hit a moving target.”
 - First question, yes, absolutely. We are already seeing this in Deep Research mode on ChatGPT, and it will get much better in the next one year. Second question, your completely right, this will probably be irrelevant in 5-10 years. That doesn’t mean we should sit back and wait for the technology to stop moving, or wait for student problems to become obvious. The AI-Lab is focused on being practical and immediately implementable, and we’ve shown it can help being students back to balanced tool usage.

QUESTIONS

- “I’m particularly interested in your discussion of The Junior Year Wall. Given the relatively recent emergence of generative AI to the mainstream (i.e less time than that of a 4-yr program), how is it being demonstrated that impacts such as this truly exist? Are the impacts significant to a degree that they show up this immediately?”
 - In short, we found it unacceptable to wait for such an event to occur, as anecdotally, we found students becoming reliant on it and even saw slippage of learning in our own usage. If the risk to graduate students and faculty exists and has to be carefully avoided, it is very likely at least some students will fall into its trap. Temptation to rely on free answers has always existed. Students who use Chegg to cheat on Mastering Physics are unlikely to do well in higher-level physics classes (or even exams in the same course). This is similar to that, but on a larger scale. Instead of having the answers to a single course (or two), LLMs can provide the answers to two years’ worth of classes, both major and non-major requirements. As a graduate student, the benefits are very tempting – I have to be disciplined about not using these tools to help on homework because I know that I need to be able to take a problem from start to finish without anyone else’s help. Students frequently lack the discipline or investment in their own learning to maintain such an approach.

In short, yes, the impacts are significant enough to show up immediately. They are more likely to show up now than in 5 years when students have had GenAI since the 6th grade (and therefore are more likely to know how and when to use it for their benefit).