Exploring AI-assisted tools in higher education: utility, limits, and student attitudes

Report*

Emily Lu (Bryn Mawr), Emma Yu (Haverford), Clara Fee (Bryn Mawr), Olivia Boyer (Bryn Mawr)

March 2026

Supervisor: Deepak Kumar

*Acknowledgement: This work was supported in part by funds from National Science Foundation Awards 2023505 and 2336929.

Contents

List of Figures

List of Tables

1 Introduction

With the advent of generative artificial intelligence (AI) tools, their use in higher education has prompted widespread institutional debate. Since the public release of ChatGPT in late 2022, adoption of large language models (LLMs) among undergraduate students has grown substantially, with one survey from 2024 estimating that 54% of students now use some form of AI on a weekly basis in their academic work [Cou24]. Since this initial release, we have too seen the release of other large language models and more specialized tools that make use of them, especially for the domains of specific academic topics, with many viewing the integration of generative AI tools into academia as necessary and inevitable.

This report covers investigations into generative AI usage in undergraduate liberal arts education (especially at Bryn Mawr and Haverford Colleges) on three topics: (1) experimentation with AI-assisted coding tools, (2) experimentation with AI tools for academic purposes, and (3) a survey of students at Bryn Mawr and Haverford Colleges on their usage of and opinions towards generative AI. For the purpose of this paper, "AI" will refer to specifically generative AI, often including but not limited to Large Language Models (LLMs) unless otherwise stated.

This report focuses specifically on the utility and student perception of generative AI tools in a liberal arts undergraduate context. It does not investigate several other important dimensions of the generative AI debate. The environmental impact of AI infrastructure, including the substantial energy and water demands of large-scale data centers [Pad26] is a significant concern that falls outside the scope of this report, though it surfaced organically in our survey data (see Section 4). Similarly, this report does not address questions of copyright and intellectual property raised by the training data practices of major AI developers [Gle26], nor does it offer a comprehensive treatment of academic integrity and plagiarism detection, issues that are extensively covered elsewhere in the literature [BP25]. Our focus is narrower: how are these tools being used, what are their practical limits, and what do students at two specific liberal arts institutions think about them?

1.1 Overview of Experiments and Findings

For our experiments in using AI we investigated AI tools for education generally, for data analysis, and for coding. For the use of AI in education, we experimented with NotebookLM [Goo], attempting to use it to explain concepts, review notes, provide citations, and create sample assessments.

For coding application, we experimented with CoPilot in VSCode [Cod] and the potential of using Large Language Models to create simple applications without the need for coding knowledge on the part of the user. We found that both showed promise in specific applications, but they also carried with them some risk, with CoPilot providing full solutions to common kinds of assignments in introductory level computer science courses and concerns about AI's ability to create and maintain more complex applications.

The final part of our investigation was into attitudes towards AI in Bryn Mawr and Haverford (Bi-co) students. The majority of respondents reported using AI in coursework very rarely or not at all. As we had a limited sample, this may not necessarily reflect the full community. Rather than complete generalizations, we examine patterns among those who responded, including looking at the relationship between major and AI attitude, as well as what common reasoning is found on either side. We observed differences in attitudes between students of different disciplines, schools, and class years, but the overall attitude of people who responded was hesitant.

1.2 Related work

1.2.1 AI Tools for Higher Education

A paper from December of 2025 discusses uses of AI among students, indicating that the primary AI tool of choice was ChatGPT [Sch+25]. Students reported using it for a variety of tasks including report writing, explanation of difficult topics, and paraphrasing to improve clarity of writing. Other AI tools cited by students included Quillbot, Grammarly[for grammar checking], Chatsmith [similar to ChatGPT but allegedly simpler to use], SlidesGo [specifically used for presentations], Blackbox [specifically used for coding] and Shaguf Bites.

As we look into the short term effects of AI use in education, there should also be consideration of its effect on skill retention and the degree to which students become reliant on it. In one study researchers demonstrate a relationship between a decrease in learning skills and increased AI tool. The use of LLM had a measurable impact on participants, and while the benefits were initially apparent, as they demonstrated over the course of 4 months, the LLM group's participants performed worse than their counterparts in the Brain-only group at all levels (neural, linguistic, scoring)[Kos+25].

The literature shows that students are using AI tools in higher education, and that some studies have found negative impacts on cognitive ability due to this AI use. Further investigation into potential uses of AI-based learning tools is necessary. Through our research we seek to better understand use cases for this technology and student attitudes towards it to help guide Bi-co policy.

1.2.2 Student Opinions on Generative AI

A Turkish study surveyed 540 students on their opinions regarding Generative AI. Similarly to our sample, this study does not claim to represent overall sentiments. Interestingly, the results on how much of the sample used AI were very different than our own, finding a high rate of usage and very few occasional users. Specifically, "Based on the AI usage habits, among 538 valid responses, only 11.0% of respondents (59 individuals) reported never using AI, while 25.7% (138 individuals) reported seldom using it. Only 0.2% (one individual) of respondents said that they sometimes used AI. This shows that AI users mainly fall clearly into infrequent or frequent usage group. In fact, the bulk of respondents reported frequent usage, with 49.1% (264 individuals) indicating that they used it often and 14.1% (76 individuals) reporting using it very often, indicating that AI is in regular use for almost two-thirds of the participants, showing substantial adoption of AI technologies by the sample" [DEÇ25].

Another study on college students from across five Southern US colleges conducted in 2024 found that the majority of students did not use AI for coursework, with reasoning primarily focusing on concerns about cheating and academic integrity. This study also found that AI usage was higher in humanities courses, specifically for use in writing assignments.[Gol+25].

A different US study looked at the differences of age groups in embracing AI and found that first year students were more hesitant that senior and graduate students. The study also found that students overall trended towards optimistic on the future of AI integration into education [Kar25].

As seen here, there is significant prior research on student attitudes and usage of AI. With our study, we seek to analyze sentiments specifically within Bryn Mawr and Haverford colleges, both small liberal-arts institutions. As the current administration and faculty are currently working to create a policy or set of guidelines about AI use in the interest of best preserving academic integrity. To this end we hope to provide more data on student perspectives and the wider context of the matter.

2 Experiments with AI-assisted coding tools

2.1 Code Completion Models

Our group conducted several experiments with AI-assisted coding tools. First, we investigated “code completion” models, which were at the time of research (Spring 2025) state-of-the-art. We primarily investigated using Copilot integrated with VSCode, trying to determine the utility of such a tool in, for example, introductory computer science courses or other higher education settings in which students are learning how to program. We rapidly concluded that such tools could provide assistance with syntax, such as auto-completing class declarations or auto-filling function declarations. Especially with languages such as Java, students often struggle with syntax difficulties which can be a barrier towards their overall conceptual learning of programming [Kad+21].

However, we concluded that these tools were inappropriate for immediate implementation into introductory courses because they would easily and automatically autocomplete full working implementations of basic coding problems like FizzBuzz, isEven, or most sorting algorithms. Critically, these code completion models would often autocomplete the implementations simply based on the *name* of the function or class:

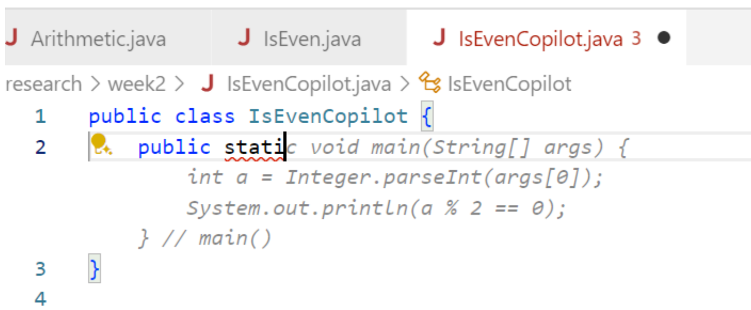

Figure 1: CoPilot autocompletes a working isEven function simply based on the name of the file

These “code completion models” can provide syntax assistance, but they will also immediately provide assistance with the conceptual problem-solving part of programming, without being prompted to do so. There is no way to customize these tools or limit their abilities to syntax assistance only.

Thus, we conclude that “code completion models” are inappropriate for introductory courses but could be useful tools for, for example, a software engineering course or senior seminar course, in which students have already learned to program but are looking to implement large scale projects. These tools can be useful for boilerplate code, and for quickly implementing simple functions, and thus could help students create larger software products. We posit that further development and investigation of customizable or limited code-completion models could provide a more useful and more suitable AI-assisted coding tool for introductory courses.

2.2 Single-use applications

The rapid improvement in LLMs’ code generation capabilities has opened a new possibility of software creation by non-programmers. Industry commentators have begun describing this use case in terms of ”single-use” applications: lightweight tools built quickly with AI assistance to solve a specific personal problem, without the overhead of traditional software development [Tan25; Bar25]. We were interested in testing the practical limits of this concept.

We prompted ChatGPT (GPT-4o) and Claude (3.7 Sonnet) to generate functional applications with simple prompts, e.g. ”please make me a simple app that does X”, as a non-programmer would. The applications we tested included a web-based birthday reminder tool and a locally hosted password manager. In both cases, the resulting applications were functional and required no manual coding intervention to run. The prompts we used were intentionally simple, closer to a description of a desired outcome than a technical specification, and in both cases the models produced working code that met the stated need.

These results suggest that single-use application development is a genuine and accessible use case for current LLMs. However, our testing also revealed clear limits. When we attempted to extend or modify the generated applications — adding features, fixing edge-case bugs, or integrating multiple components — the process required substantially more technical oversight. This aligns with prior findings that LLMs struggle with legacy codebases and complex, multi-component software projects [Kad+21]. We therefore emphasize that the "single-use app" paradigm is most viable for contained, low-stakes personal tools. Any application serving real users, handling sensitive data, or requiring ongoing maintenance benefits significantly from programming expertise and a human-in-the-loop approach [Str].

2.3 AI tools for data analysis

As part of our exploration of AI-assisted tools, we investigated two platforms for data analysis: JuliusAI, a specialized tool designed specifically for data analysis tasks, and Claude, Anthropic's general-purpose large language model. Using the survey data described in Section 4 as our test dataset, we tasked both tools with identical data visualization and analysis prompts. We evaluated the tools based on ease of use, accuracy, and depth of insight.

Our findings revealed little meaningful difference in output quality between the two tools. Despite JuliusAI's specialization — and possible fine-tuning — for data analysis tasks, Claude produced comparably accurate visualizations and interpretations. This suggests that, at least for straightforward social science data analysis, domain-specific AI tools do not offer a substantial performance advantage over general-purpose LLMs.

That said, we note a potential pedagogical distinction between the two tools. JuliusAI structures its interface around guided prompts and explicit suggestions for which analytical methods to apply, which may benefit students who are new to data analysis and need scaffolding to make methodological decisions. In this respect, a specialized tool's value may lie less in the quality of its outputs and more in how it guides students through the analytical process itself.

3 Experiments with AI tools for education: NotebookLM

3.1 Introduction to NotebookLM

NotebookLM is a research and note-taking online tool developed by Google Labs that uses artificial intelligence, specifically Google Gemini, to assist users in interacting with their documents [Goo]. Users start by uploading documents through a url link or directly copying texts into the chatbox. After NotebookLM processes the document for a moment, various formats of outputs are generated. Video overview is an informal lecture based on the document. Audio overview provides a conversation between two voices and users are able to interact with them. Other features include slide deck, mind map, flashcards, and quiz. The contents are similar across different formats and are all based on the document uploaded.

3.2 Summary of Using NotebookLM as a Learning Tool

Our research group explored NotebookLM for many different classes. Below, we offer a summary of our findings in using NotebookLM as a learning tool.

Spanish: For materials in a 100 level Basic Intermediate Spanish class we found that NotebookLM effectively integrated visuals from lecture slides into its generated video overviews and produced engaging and well-structured quizzes that reinforced understanding. In addition, we used NotebookLM to write a Spanish thesis in a 300 level course, where it generated high quality citations and drew connections between various sources. We first developed some ideas and then backed up the theories by using NotebookLM. This was similar to using Zotero without searching through a large number of sources.

Introduction to Artificial Intelligence: We used NotebookLM for preparation for exams in this 300 level class, where the materials included lots of dates, people, and conceptual theories. We started by uploading the class lecture slides. NotebookLM generated summaries, flash cards, and quizzes that helped with clarification of ideas.

Algorithms for Big Data: In the Big Data Algorithms class (a graduate level class at U. Pennsylvania), we used NotebookLM to help understand notes from the class. The notes contained detailed mathematical proofs. However, the model tended to stay at a conceptual level and ignored the proofs, even when it was explicitly prompted to explain them.

Computational Neuroscience: NotebookLM was used as preparation for quizzes in this 200 level class. However, because it was not clear what content would be covered from the readings, mock quizzes that NotebookLM generated were often not very helpful. They primarily reinforced a basic understanding of the readings, whereas the actual quizzes sometimes focused on specific details or required a deeper understanding of particular theories.

Behavioral Neuroscience: In the context of writing a paper for a 300 level behavioral neuroscience lab, NotebookLM was not very helpful with the readings. This was probably because the paper instructions included very detailed questions about the readings, but oftentimes NotebookLM only offered a brief understanding about the papers. In addition, NotebookLM also made mistakes when explaining specific aspects of the readings at a deeper level.

Iranian Cinema: Lastly, We experimented with NotebookLM for a 300 level Iranian Cinema class, which is a classical liberal arts class with discussion, reading, and papers. One member initially used this tool to analyze readings and draw connections between different lectures. However, they concluded that NotebookLM did not add any significant learning that the student was missing, and actually discouraged the member from paying full and close attention to each reading. Thus, the member discontinued the use of NotebookLM for this class.

In summary, NotebookLM performs especially well in courses that focus on conceptual learning and broad information synthesis, and it was helpful for summarizing lecture slides, providing good sources for thesis, and organizing and differentiating conceptual theories. One of the best use case that we recommend is making mock exams for these types of classes. NotebookLM turns out to be not so helpful in explaining mathematical reasoning for some STEM classes. It is of limited use for seminar style classes. This was not limited to humanities, where it may be important for students to struggle through their readings on their own. Additionally, NotebookLM could potentially make mistakes in terms of details and deeper comprehension about the readings. It would be hard for students to identify mistakes without doing the readings themselves. We also note that sometimes NotebookLM failed to generate output based on multiple sources that were uploaded simultaneously. NotebookLN could be enhanced further if it could save their chat history in order to provide continuing context.

4 Survey of undergraduate students on AI use in higher education

In exploration of AI in higher education settings, it is critical to gather qualitative and quantitative data regarding students' attitudes towards, and usage of, AI-assisted tools. To this end, we administered a month-long survey approved by the Institutional Review Board. We sought to gain insight into Bi-Co students' usage of and attitudes towards generative AI in academic settings. In the survey, "generative AI" refers to tools such as chatGPT, Claude, Gemini, or other large language model (LLM). It also refers to generative AI–based tools such as Cursor, Windsurf, NotebookLM, etc. We were particularly interested in differences in attitudes and usage between the colleges, as well as between different departments.

Out of 184 students who disclosed their AI usage, the overall split is nearly even: approximately 48.4% identify as AI users, while 51.6% identify as non-users. However, institutional differences were significant: while Haverford respondents were evenly split, Bryn Mawr students made up nearly three quarters (72.6%) of the non-user group.

Looking within the schools, we also see that the majority of Bryn Mawr respondents reported having never used AI, whereas the Haverford student majority reported using it in some capacity.

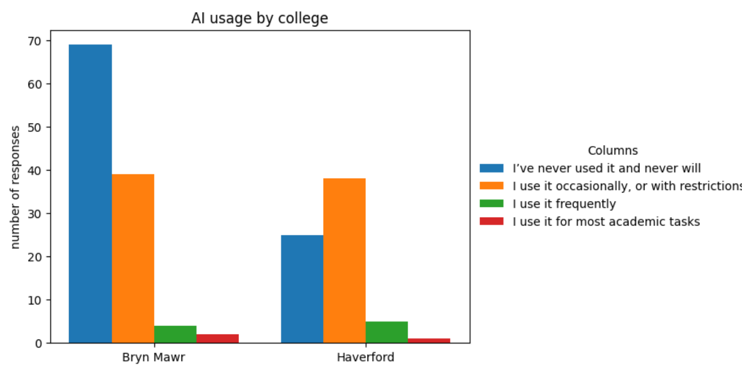

Figure 2: Bar chart displaying reported AI usage by college

To look at differences across majors we split respondents' majors into three categories: STEM, humanities/social-science, and interdisciplinary. Humanities and social-sciences were grouped together as both domains focus more on qualitative data and writing than STEM courses. Interdisciplinary refers to respondents' who were double majors in STEM and humanities or social science fields, or those who reported a minor in a different area than their major.

We found that pure STEM majors reported the highest AI usage with most(52.46%) selecting the category "I use it occasionally, or with restrictions (e.g. for only certain tasks)." Notably, the next highest category for STEM majors was "I've never used it and never will" at 36.07% with very few respondents(11.46%) citing frequent usage. The majority (57.14%) interdisciplinary majors reported never using AI, though the divide was still close to a 50/50 split. 63.79% of pure humanities and social-science majors reported never using AI. Only STEM majors ever reported using AI "for most academic tasks," and even then it was only 3.28% of respondents.

The following chart displays this data:

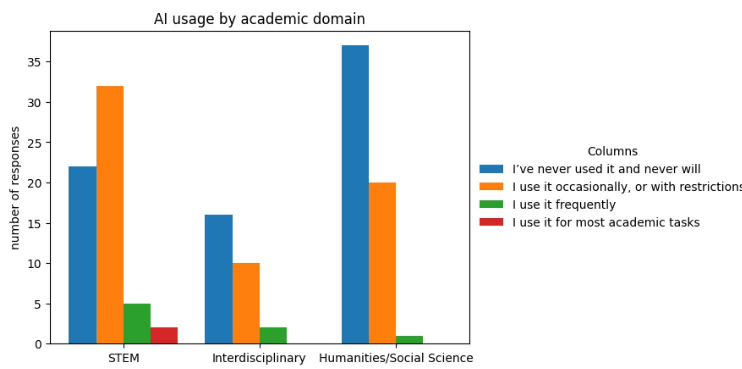

Figure 3: Bar chart displaying reported AI usage by academic discipline

The responses of students to the open-ended questions reveal that there is a deep divide between cautious acceptance and complete rejection of generative AI (Notably, 3 out of 145 respondents explicitly admitted to using AI to complete this survey). Among self-identified AI users, the majority restrict its use to occasional tasks. For example, a mere 1.7% of responding students (3 out of 176) utilize AI for creative or reasoning tasks such as drafting full essays. Instead, a more substantial 15.9% (28 out of 176) use AI like a conceptual tutor to navigate complex topics, while 8.5% (15 out of 176) use it merely to initiate tasks, such as brainstorming outlines or generating pseudocode.

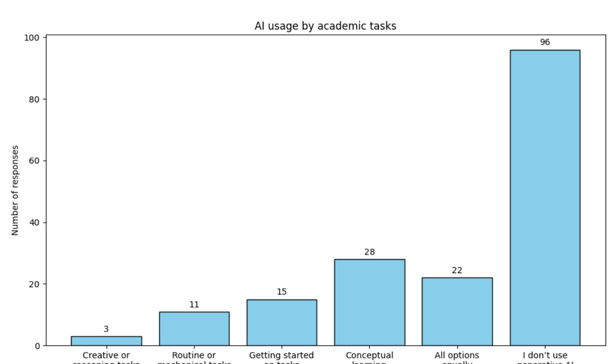

Figure 4: Bar chart displaying reported AI usage by academic task types

For these users, AI functions as a supplementary tool rather than a replacement for foundational cognitive work. As one student put it,

> "I think it is a helpful tool, but we should be very selective and intentional about how we use it."

Many users compare AI to a calculator. They believe that AI can be highly useful only after core academic skills have been mastered, but it should not replace the process of learning the basics. Furthermore, more than half of the users generally perceive themselves as capable evaluators of AI outputs: 41.9% reported being "moderately confident" and an additional 14.0% felt "very confident" in their ability to identify hallucinations.

On the other hand, non-users reject the technology much more strongly. Critical open-ended responses frequently characterize generative AI as "plague" or "soulless." Some worry that it is undermining people's memory, critical thinking, and overall capabilities. Beyond these concerns, the environmental impact of AI came up frequently. Nearly 40 distinct responses showed concern about the excessive energy consumption and water demands of AI data centers, which was the primary reason they rejected the use of AI.

Beyond students' personal attitude towards the use of generative AI, the survey also reveals that the faculty guidelines for using AI are inconsistent across disciplines. While most professors have strict rules, the specific guidelines vary widely. In fields such as Computer Science and Mathematics, expectations can change from class to class: one professor may strictly prohibit AI on problem sets, whereas another explicitly encourages its use for reviewing concepts or debugging code syntax. Students reported that this lack of college-wide consistency fosters frustration as the boundaries of academic integrity seem ambiguous. Several respondents suggested that a clear and unified college-wide guideline would help reduce this confusion.

When asked how classes should adapt, students proposed two main approaches. Students more open to AI argue that the rise of technology is unavoidable. Rather than banning it, they believe that educators should teach students how to use it responsibly. Suggestions include shifting evaluation metrics towards learning process itself, such as requiring progressive drafts, outlines, or version history tracking via Google Docs. On the other hand, students opposed to AI support stricter restrictions. They suggest enforcing complete bans, such as returning to handwritten assignments and placing greater emphasis on in-person exams.

Despite students having differing perspectives on the practical utility of AI, there is a consensus regarding the coursework. Both users and non-users strongly value in-class, discussion-based, and oral assessments. Whether the goal is to support academic integrity or to promote deeper intellectual learning, students agree that assignments requiring real-time thinking and unassisted reasoning are an essential part of a liberal arts education.

5 Discussion

5.1 Summary of findings

In our experiments with AI assisted coding, we tested both Copilot code auto-completion and use of LLMs to generate code for single use applications. We found that auto-complete could be helpful, but it is unsuitable for introductory Computer Science courses as it will often automatically fill in the common, simple programs that tend to make up early assignments, thus stopping students from actually learning these fundamentals. For coding single use apps, we found that using the AI models to code them was quite successful, however human input would be needed to create more complicated applications.

For our experiments with NotebookLM, we found it was most helpful for explaining general concepts for courses with exams. It failed to adequately explain complex topics and go in depth to specific detail even when specifically prompted. We also saw success in using it for creating citations in thesis writing, where it helped to streamline the process of finding papers and creating citations for them. Overall, we saw that it could be helpful as a tool in some instances, but it cannot be relied upon, especially in higher level courses.

With a sample size of 228 out of over 2,000 Bi-Co students and a survey distribution that allowed for selection bias, we cannot claim to know the full makeup of opinions on generative AI, but we have identified patterns of usage and common reasoning and concerns among those who did respond. The relatively high number of AI users in STEM fields could be due to a variety of reasons, including but not limited to lower restrictions from professors, the nature of work seeming more conducive to AI use, or simply mindset trends among STEM majors. Further study should be done to investigate the mechanisms influencing AI usage across subject areas. We also see concerns around the ethicality of AI technology and academic integrity among students. This uncertainty and confusion could be decreased through the development of college-wide guidelines that incorporate student input and ethical concerns in the conversation.

5.2 Recommendations

From our assessment of student attitudes towards AI, it would seem that a formal college or bi-college AI policy or set of guidelines would be beneficial. Within this guideline departmental and individual class policies can be developed. The variety of responses regarding reasoning behind AI use or lack thereof demonstrate a level of engagement in the topic from students, indicating that policy discussions and decisions would benefit greatly from direct student input and collaboration.

References

[Kad+21] Rozita Kadar et al. "A Study of Difficulties in Teaching and Learning Programming: A Systematic Literature Review". In: *International Journal of Academic Research in Progressive Education and Development* 10 (Aug. 2021). DOI: `10.6007/IJARPED/v10-i3/11100`.

[Cou24] Digital Education Council. 2024. URL: `https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024` (visited on 04/23/2026).

[BP25] Himendra Balalle and Sachini Pannilage. "Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity". In: *Social Sciences Humanities Open* 11 (2025), p. 101299. ISSN: 2590-2911. DOI: `https://doi.org/10.1016/j.ssaho.2025.101299`. URL: `https://www.sciencedirect.com/science/article/pii/S2590291125000269`.

[Bar25] Alistair Barr. *Disposable apps' are the hot new thing in tech, as AI makes coding easier and quicker.* Oct. 31, 2025. URL: `https://www.businessinsider.com/disposable-apps-ai-makes-coding-easier-vercel-2025-10` (visited on 04/16/2026).

[DEÇ25] Volkan Duran, Ercümend Ersanlı, and Hurinur Çelik. "Unveiling student sentiment dynamics toward AI-based education through statistical analysis and Monte Carlo simulation". In: *British Educational Research Journal* (May 2025). DOI: `10.1002/berj.4188`.

[Gol+25] Jonathan M. Golding et al. "Generative AI and College Students: Use and Perceptions". In: *Teaching of Psychology* 52.3 (2025), pp. 369–380. DOI: `10.1177/00986283241280350`. URL: `https://doi.org/10.1177/00986283241280350`.

[Kar25] Betsy Barre Karen Singer-Freeman Kristi Verbeke. "Generative AI Usage Among University Students Depends on Academic Level and Task". In: *Higher Learning Research Communications* (2025), pp. 1–25. DOI: `10.18870/hlrc.v15i2.1616`.

[Kos+25] Nataliya Kosmyna et al. *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task.* 2025. DOI: `https://doi.org/10.48550/arXiv.2506.08872`. arXiv: `2506.08872 [cs.AI]`. URL: `https://arxiv.org/abs/2506.08872`.

[Sch+25] Dusana Alshatti Schmidt et al. "Integrating artificial intelligence in higher education: perceptions, challenges, and strategies for academic innovation". In: *Computers and Education Open* 9 (2025), pp. 100–274. ISSN: 2666-5573. DOI: https://doi.org/10.1016/j.caeo.2025.100274. URL: https://www.sciencedirect.com/science/article/pii/S2666557325000333.

[Tan25] Jason Tanner. *The Resurgence of Disposable Apps: A New Era in Software Monetization*. Mar. 26, 2025. URL: https://appliedframeworks.com/blog/the-resurgence-of-disposable-apps-a-new-era-in-software-monetization (visited on 04/16/2026).

[Gle26] Stephanie Schmidt Marc B. Collier Logan Woodward Ethan Glenn. "AI in litigation series: An update on AI copyright cases in 2026". In: *Norton Rose Fulbright* (Mar. 2026). URL: https://www.nortonrosefulbright.com/en/knowledge/publications/ce8eaa5f/ai-in-litigation-series-an-update-on-ai-copyright-cases-in-2026 (visited on 04/16/2026).

[Pad26] Laura Paddison. "Scientists have found an alarming environmental impact of vast data centers". In: *CNN Climate* (Mar. 30, 2026). URL: https://www.cnn.com/2026/03/30/climate/data-centers-are-having-an-underrported (visited on 04/16/2026).

[Cod] Visual Studio Code. *GitHub Copilot in VS Code*. URL: https://code.visualstudio.com/docs/copilot/overview (visited on 04/16/2026).

[Goo] Google. *NotebookLM*. URL: https://notebooklm.google/ (visited on 04/16/2026).

[Str] Cole Stryker. *What is human-in-the-loop?* URL: https://www.ibm.com/think/topics/human-in-the-loop (visited on 04/16/2026).