INTERNATIONALE HOCHSCHULE

www.iu.de

IU DISCUSSION PAPERS

Business & Management

AI as a Study Partner in Mathematics Education: Evidence from a Multi-Program University Survey

MEHMET EVRIM ALTIN
KIRSTEN JÄGER
SILKE JÜTTE
FLORIAN SCHNEIDER

IU Internationale Hochschule

Main Campus: Erfurt

Juri-Gagarin-Ring 152

99084 Erfurt

Telefon: +49 421.166985.23

Fax: +49 2224.9605.115

Kontakt/Contact: kerstin.janson@iu.org

Autorenkontakt/Contact to the author(s):

Dr. Mehmet Evrim Altin
ORCID iD: 0000-0002-6571-398X
IU Internationale Hochschule – Campus Bad Honnef
Email: mehmet-evrim.altin@iu.org

Prof. Dr. Kirsten Jäger
ORCID iD: 0009-0005-6757-6842
IU Internationale Hochschule - Campus Stuttgart
Email: kirsten.jaeger@iu.org

Prof. Dr. Silke Jütte
ORCID iD: 0000-0001-8960-3016
IU Internationale Hochschule – Virtueller Campus
Email: silke.juette@iu.org

Prof. Dr. Florian Schneider
ORCID iD: 0000-0002-7052-8404
IU Internationale Hochschule - Campus Mannheim
Email: florian.schneider@iu.org

IU Discussion Papers, Reihe: Business & Management, Vol. 7, No. 7 (APR 2026)

ISSN: 2750-0683

DOI: https://doi.org/10.56250/4120

Website: https://repository.iu.org

AI as a Study Partner in Mathematics Education: Evidence from a Multi-Program University Survey

Mehmet Evrim Altin
Kirsten Jäger
Silke Jütte
Florian Schneider

ABSTRACT:

The rapid diffusion of artificial intelligence (AI) tools is fundamentally reshaping learning practices in higher education, particularly in mathematics and statistics courses. This study investigates how undergraduate students across multiple programs at IU International University engage with AI-based mathematics tools, which applications they prefer, and how they evaluate their performance. Drawing on survey data from 174 students enrolled in mathematics lectures during the spring term of 2025, we analyze awareness, usage intensity, and perceived usability, understandability, correctness, and value for money of leading AI tools. The results reveal a highly concentrated market structure: ChatGPT, Photomath, and Gemini dominate student awareness and usage. While ChatGPT is perceived as the most user-friendly and offers strong value for money, Photomath receives the highest ratings for correctness of results. Gemini, in contrast, is evaluated more cautiously across dimensions. Differences in awareness between math-intensive and non-math-intensive programs are small and statistically insignificant, suggesting that AI adoption in mathematics is broadly distributed across disciplines. The findings indicate that students integrate AI tools primarily as complementary learning aids rather than replacements for traditional materials. Overall, the study provides an empirical baseline for understanding how AI functions as a study partner in mathematics education and highlights the growing importance of evaluating not only usage frequency but also perceived reliability and pedagogical value.

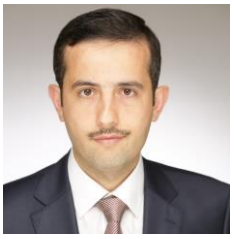
KEYWORDS:

Artificial Intelligence, Mathematics Education, Student Perceptions, Learning Tools

JEL classification:

I21, I23, O33, A22, C83

AUTHORS



Dr. Mehmet Evrim Altin *is a research assistant in mathematics at IU International University of Applied Sciences and also teaches at various institutions focusing on mathematics education and AI integration in learning. His research interests lie in the field of educational mathematics, the impact of artificial intelligence on student learning outcomes, and the pedagogical integration of digital tools in higher education.*

Prof. Dr. Kirsten Jäger *Prof. Dr. Kirsten Jäger is Professor of Business Administration and Department Head Business at IU International University in Stuttgart. Her research focuses on marketing, particularly customer journeys and brand equity in the metaverse. She brings more than 20 years of experience in quantitative data analysis, marketing, and applied econometrics.*

Prof. Dr. Silke Jütte *is professor of quantitative methods at IU International University's virtual campus with a background in consulting, software engineering and logistics. In her research, she focusses on scheduling algorithms, network optimization and specific questions arising at small- and medium-sized businesses.*

Prof. Dr. Florian Schneider *is a professor of mathematics at IU International University and also supervises various engineering projects at the Junior Engineering Academy for Gifted Students at Heinrich Heine Gymnasium in Kaiserslautern. His research interests lie in the field of numerical methods for partial differential equations, mathematics education, and the application of artificial intelligence in a wide range of fields.*

Introduction

The landscape of higher education has undergone a profound transformation in recent years, driven by the rapid advancement and widespread accessibility of artificial intelligence technologies. In mathematics education particularly, this transformation has been both dramatic and multifaceted, affecting not only how students learn but also how educators approach teaching, assessment, and pedagogical design. An initial study, conducted in 2024 with students enrolled in a single undergraduate course at IU International University of Applied Sciences, provided valuable preliminary insights into the relationship between AI adoption to Math courses and student learning behaviors (Altin, 2024). That foundational research examined 34 participants across two semesters, revealing nuanced patterns in how students engaged with AI platforms such as ChatGPT, Gemini, Wolfram Alpha, and mathematics-specific applications alongside traditional resources like YouTube videos.

However, the research was constrained by its narrow scope—limited to a single course, a relatively small sample size, and students at a specific academic level. These limitations raised important questions about the generalizability of our findings and pointed to the need for a more comprehensive investigation. For example, how do students across different mathematical disciplines engage with AI? Which of the artificial intelligence platforms that have entered our lives in a very short time are being used by university students? And how long is this usage limited to? Do the profiles of the students play a role in the preference for artificial intelligence platforms?

Under these critiques, the current expanded research encompasses students from various Bachelor programs at IU International University of Applied Sciences who are enrolled in mathematics and statistics courses. This broader approach allows us to examine AI adoption and impact across a diverse spectrum of academic contexts, student populations, and mathematical content areas. Rather than focusing exclusively on a single mathematics course, we now investigate students studying everything from foundational quantitative methods to advanced statistical modeling and from business mathematics to engineering calculus.

This expansion serves multiple critical purposes. First, it enables us to identify patterns and variations across different student demographics, including age, prior educational background, field of study, and academic level. Second, the larger and more diverse sample provides enhanced statistical power, enabling more robust conclusions and the ability to detect subtle effects that might have been obscured in the smaller initial study.

The expanded study maintains the quantitative research design of the original investigation while incorporating lessons learned from the initial phase. However, the current research includes refined measurement instruments, expanded questions addressing emerging technologies and applications, and additional variables capturing contextual factors that may influence AI usage in mathematics lectures. The larger sample size also permits more sophisticated statistical analyses, including subgroup comparisons and the examination of interaction effects between different factors.

This expanded investigation explores how students across different mathematical disciplines engage with AI technologies in their learning journey. Central to our inquiry is understanding which AI tools students choose and why—and whether these choices vary systematically between those in math-intensive fields like engineering and data science versus students in business or management

programs. We are particularly interested in how students at different stages of their academic careers, from first-semester undergraduates to advanced learners, differ in their adoption patterns and their capacity to evaluate these tools critically. Beyond simple usage statistics, we examine how students perceive the quality of these AI tools— their ease of use, clarity of explanations, accuracy of results, and overall value—and whether these perceptions align with actual usage behavior. Through this comprehensive approach, we aim to move beyond surface-level observations toward a nuanced understanding of AI's role in mathematics education that can guide both teaching practice and institutional decision-making.

To achieve the goals mentioned above, this study surveyed 174 undergraduate students attending mathematics lectures at IU International University during spring 2025. This study employed a cross-sectional, single-wave survey design conducted at IU International University. The sampling strategy targeted all students enrolled in undergraduate mathematics lectures, with recruitment occurring through in-class administration of a Microsoft Forms questionnaire. Compared to the 2024 cohort, we included students from different study subjects, semesters and study programs. Data was collected on three different aspects:

- Students' association with, prior performance in and self-assessment of mathematics
- Students' awareness and usage time of ten main math AI tools
- Students' rating of the given AI tools with respect to usability, understandability, correctness, and value for money

The overarching objective of this expanded research is to provide a comprehensive, empirically grounded understanding of how AI technologies are reshaping mathematics education in higher education settings. By examining a diverse range of students, courses, and contexts, we aim to move beyond anecdotal observations and initial exploratory findings toward robust conclusions that can inform institutional policy, pedagogical practice, and future research directions.

Literature Review

The rise of AI-powered tools is transforming how students learn and understand mathematics, especially for those who are not math specialists (such as non-majors or out-of-field students in schools, universities and universities of applied sciences). Recent peer-reviewed studies and surveys have begun to shed light on which AI tools students use, in what ways, and how frequently, to support their math learning. Below we summarize key findings from the literature, focusing both on commonly used AI chatbots as well as dedicated math-solving tools and their usage patterns in both school and university contexts.

By late 2023, ChatGPT (OpenAI's conversational AI) emerged as the most favored AI tool among students for assistance with academic work (Garcia et al., 2025; Wardani et al., 2024). In a comparative study of education students, ChatGPT was by far the preferred tool for both math and language assignments (Wardani et al., 2024). One reason is its versatility – ChatGPT can explain concepts, answer questions, and generate step-by-step solutions in natural language. Starting with the GPT-4 and later-on o3 and GPT-5, the models became more capable of providing more coherent and accurate mathematical responses than earlier models (Pepin et al., 2025). Other generative AI chatbots (e.g.

Anthropic's Claude or Google's Gemini) exist, but current education research has focused predominantly on ChatGPT given its widespread adoption.

Aside from chatbots, students frequently use dedicated math-solving apps. Photomath is a prime example – it uses AI (OCR, i.e. optical character recognition, and algebra algorithms) to solve math problems from a photo and provide step-by-step explanations. (Capinding, 2023) demonstrated Photomath's educational value: in a controlled study with 49 high school students, using Photomath in a precalculus course significantly improved students' confidence, self-reliance, interest, and performance on math tasks.

This suggests that AI solvers, when used as learning tools, can enhance understanding and motivation, not just give answers. Other similar tools include Mathway, and Wolfram Alpha, which can perform symbolic calculations or provide worked solutions. While not all of these are "AI" in the modern machine-learning sense (Wolfram Alpha is a computational engine at its core, combined with AI tools), students do turn to them for help with algebraic manipulation, checking homework answers, and exploring alternative solution methods. For instance, Wolfram Alpha has been used in teacher-training settings to illustrate multiple solution strategies for elementary math problems, indicating its usefulness in exploring mathematics from different angles (Abramovich, 2021).

The way students use AI is evolving toward a "partnership" model: rather than simply obtaining answers, many learners engage in a dialogue with AI, using it to test their understanding, generate alternative solutions, and receive feedback on their reasoning. The AI can break down complex problems into simpler steps or offer analogies, helping "non-expert" learners grasp material at their own pace. In the same study, students also used ChatGPT to explore alternative solution strategies for math modeling problems (Garcia et al., 2025).

Because AI tools are available anytime, students can practice math outside of class with guided support. The instant feedback and nonjudgmental nature of AI mean learners can ask "dumb" questions or attempt problems repeatedly without fear. As math anxiety negatively correlates with students' performance (as shown in the meta study of (Barroso et al., 2021)), using AI can act as a game changer in mathematics education: When used as a supportive guide instead of a simple solution provider, AI tools can deepen understanding and reduce anxiety, all while students maintain control of their learning process (Pascale, 2025). This personalized support helps students identify mistakes and correct misconceptions in real time. That said, accuracy is not guaranteed – students must learn to verify AI-provided solutions (as AI can occasionally produce incorrect calculations or "hallucinate" reasonably-sounding answers).

While most studies concentrate on evaluating the use of one specific math AI tool, few national research has so far focused on comparing the use of several tools from the perspective of German students. With this study, we try to shed light on the specific needs of students in using math AI tools and their choice of tools when so many options are available.

Methodology and Demographics

This study was carried out in the setting of undergraduate mathematics lectures at IU International University during the spring term of 2025, hence reflecting the perspective of students who are at the beginning of their academic journey. Students were invited to complete an online survey, designed in

Microsoft Forms, while attending class. Administering the survey during lectures ensured that a broad group of students was reached and that the responses reflected a live academic environment rather than a self-selected sample outside the classroom. Participation was voluntary, and the questionnaire was anonymous. Before starting, students were informed of the purpose of the study and assured that their answers would be used only for research.

In total, 174 valid responses were collected. Roughly 82 per cent of all participants were enrolled in dual study programs. In terms of academic focus, business and management students dominated the sample (55%), but the sample also includes a notable share of technically oriented majors such as computer science (13%) and engineering (6%). This combination makes it possible to compare patterns between math-savvy and non-math-savvy groups.

The age distribution shows that our participants are comparably young, which is a result of the fact that our math lectures are very early in the curriculum of all major subjects. Almost half of the students were between 21 and 25, another 43 percent were 20 or younger, and only a few fell above the age of 30. The semester breakdown supports the same conclusion: nearly two thirds were in their second semester, and another quarter in their first semester. Higher semesters were only lightly represented.

Regarding gender, less than half of the students in our sample (39%) were female. Given that the share of female students among all IU dual students is higher in the population (64%), gender-specific results drawn from our sample must be treated with caution.

Potential for Math AI Tools

At the beginning of our survey, students were asked to present one or more spontaneous terms they associate with mathematics. While a majority (69%) gave neutral answers such as "numbers", “calculation”, “formulas”, less than 3% spontaneously had positive associations with mathematics (e.g. “fun” or “exciting”). On the other hand, 28% of the students came up with negative associations such as “stress”, “headache”, ”failing” and even “death”, “suffering” and “horror”. This observation is in line with the findings of Barroso et al (2021).

When asked to rate their own mathematics skills on a scale from 1 (very poor) to 10 (very good), the average rating was just below 6, and the most common score was 6. Very few students placed themselves at the extremes, either at the very bottom or at the very top. Taken together, the pattern suggests moderate confidence with some caution.

At university level, most of our participants had not yet passed a mathematics or statistics class, which fits with the fact that most of the group were first- or second-semester students.

Weekly learning routines are centered on digital resources. Students reported spending on average 1.6 hours with lecture slides and 1.4 hours with further course materials provided by the university. Time spent with textbooks or study groups was much lower, at around 0.5 to 0.75 hours per week. This highlights a strong reliance on official materials provided by instructors and a relatively low use of traditional or collaborative methods. When asked about their use of YouTube, TikTok, or similar platforms for math learning, most students admitted to only engage occasionally. Thirty-one percent said they “rarely” used such videos, 24 percent “rather rarely,” and 21 percent not at all. Only a

minority—about a quarter—reported more regular use, with just 7 percent describing their use as "frequent."

Overall, students' current situation in math learning seems to be promising for the use of AI tools: The subject is commonly considered "difficult", and digital resources are preferred over traditional ones.

Usage of Math AI Tools

Next, we evaluated which AI tools[1] are commonly known and used by students. The overall picture (Figure 1) illustrates that ChatGPT is by far the most widely recognized tool, mentioned by 165 students, making it nearly universal in awareness. In contrast, Photomath (74 mentions), Gemini (61 mentions) and DeepSeek (53 mentions) form a second tier of recognition, familiar to a substantial but clearly smaller share of the cohort. Traditional math tools such as Wolfram Alpha or Desmos remain marginal, despite their long-standing reputation in academic contexts.

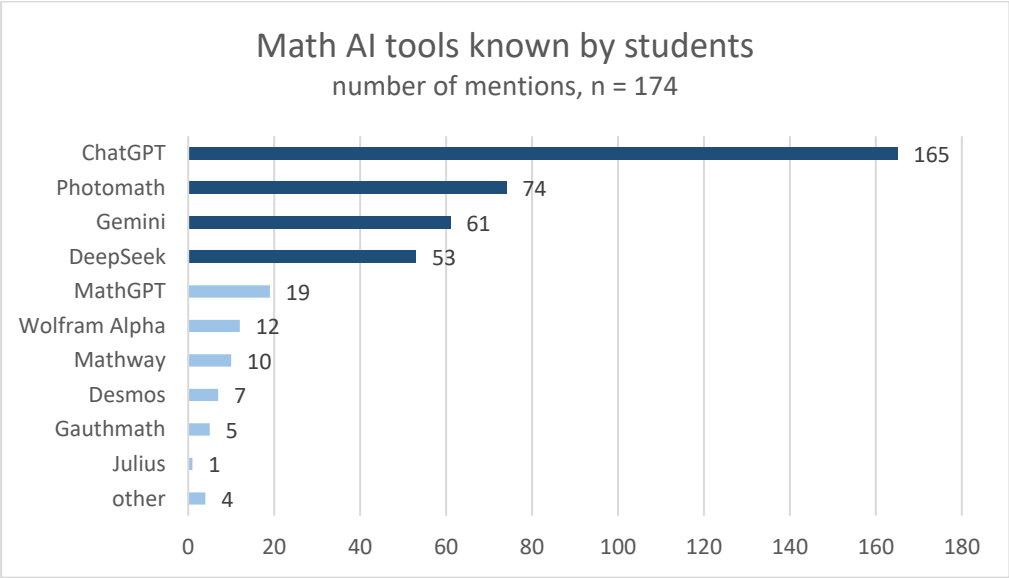


Figure 1

For an in-depth analysis, we distinguished between students from math-savvy subjects (such as engineering and data science) and non-math-savvy subjects (such as management, marketing and social sciences). Additionally, we compared first-semester students (freshers) to those of higher semesters (non-freshers).

[1] At the time the survey was executed, the current versions of the main AI tools listed in Figure 1 were GPT o-series, Photomath 8.43.0, Gemini 2.5 Pro, and DeepSeek-Prover-V2-671B.

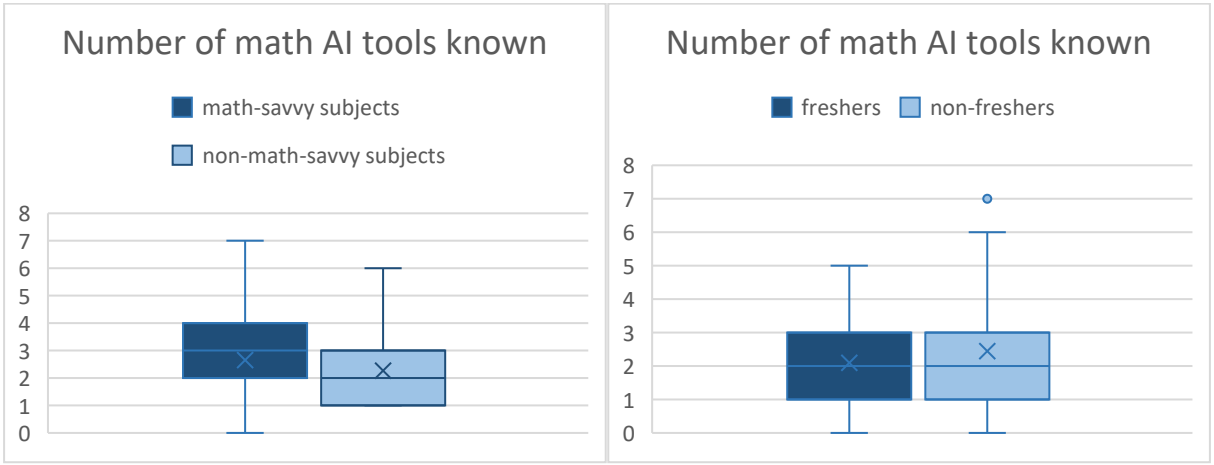


Figure 2

The left two boxplots of Figure 2 show the number of math AI tools that are known by students from math-savvy subjects (n = 43, mean = 2.65, stdev = 1.41) as compared to non-math-savvy subjects (n = 131, mean = 2.26, stdev = 1.20). It turned out that students from math-savvy subjects only know a slightly higher number of math AI tools on average (p = 0.05700).

The right two boxplots of Figure 2 show the respective analysis when comparing freshers to non-freshers. The number of math AI tools known by non-freshers (n= 132, mean = 2.45, stdev = 1.27) slightly exceeds that of freshers (n = 42, mean = 2.10, stdev = 1.21), though the difference is not significant on a 5% level (p = 0.10878).

When comparing the awareness level for the different tools, we observe that students from math-savvy subjects know a higher variety of tools than those from non-math-savvy subjects (see Figure 3). In total, ChatGPT and Photomath seem to be more prominent among non-math-savvy subjects, while further tools are rather known among math-savvy subjects.

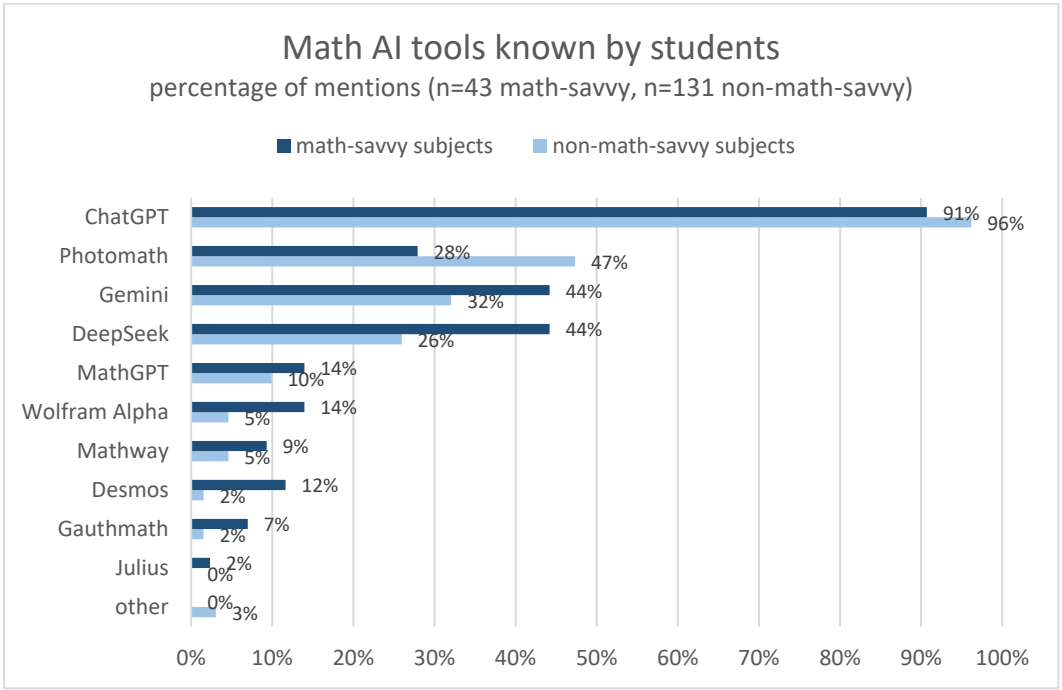


Figure 3

Figure 4 compares the tools' awareness levels for freshers and non-freshers. While the awareness levels of ChatGPT and Photomath hardly change between freshers and non-freshers, Gemini and DeepSeek are much better known among higher semester students.

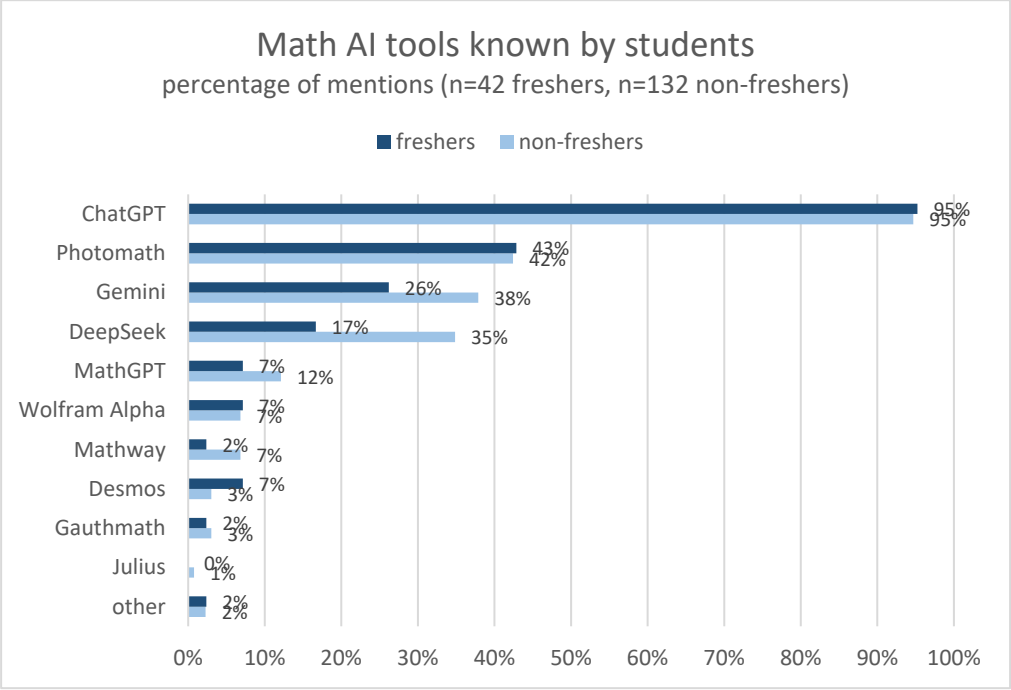


Figure 4

For the analysis of usage times, we will focus on the five most prominent tools—ChatGPT, Photomath, Gemini, DeepSeek, and MathGPT— and hence capture three general-purpose AI applications on the one side and two specialized mathematics-focused solutions on the other side.

Considering the intensity of weekly use across the five most well-known AI tools, we observe clear differences: In a 100% stacked bar chart, Figure 5 breaks down the number of students familiar with a tool to their weekly use of the tool. ChatGPT stands out: while 52 students reported using it for only one hour per week, a considerable group (35 students) indicated heavy use of four or more hours weekly, making it both the most widely and the most intensively employed tool. MathGPT shows a similar pattern, though the share of heavy users (four or more hours weekly) is much smaller. Note, however, that the total number of students knowing MathGPT is much smaller than that of ChatGPT. Photomath appears primarily as a light-use application, with nearly half of the group (31 students) engaging for just one hour per week and very few extending beyond three hours. Gemini and DeepSeek are the least used overall- most students who stated they know the tool do not use it at all (41 students each). This suggests that Gemini and DeepSeek have not yet achieved regular integration into students' math study routines.

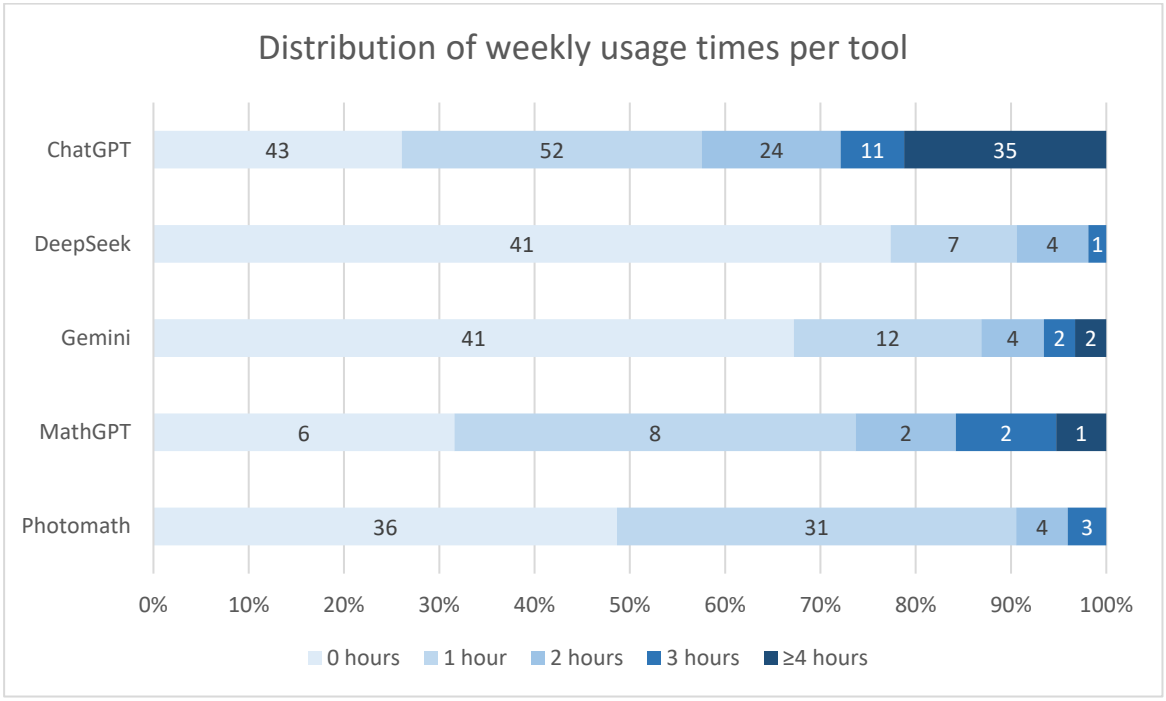


Figure 5

When calculating average usage times per tool (with a conservative transformation of the "≥ 4 hours" answer to a numeric value of 4), the above-mentioned pattern gets even more prominent (Figure 6): Recall that students' awareness about the different tools showed large variation. Hence, when calculating the average usage times per tool across all students, non-surprisingly, the average usage time of ChatGPT by far exceeded that of all other tools (light blue bars of Figure 6). However, only calculating mean usage times across those students knowing the respective tool draws a different picture (dark blue bars of Figure 6), with especially MathGPT catching up in usage time.

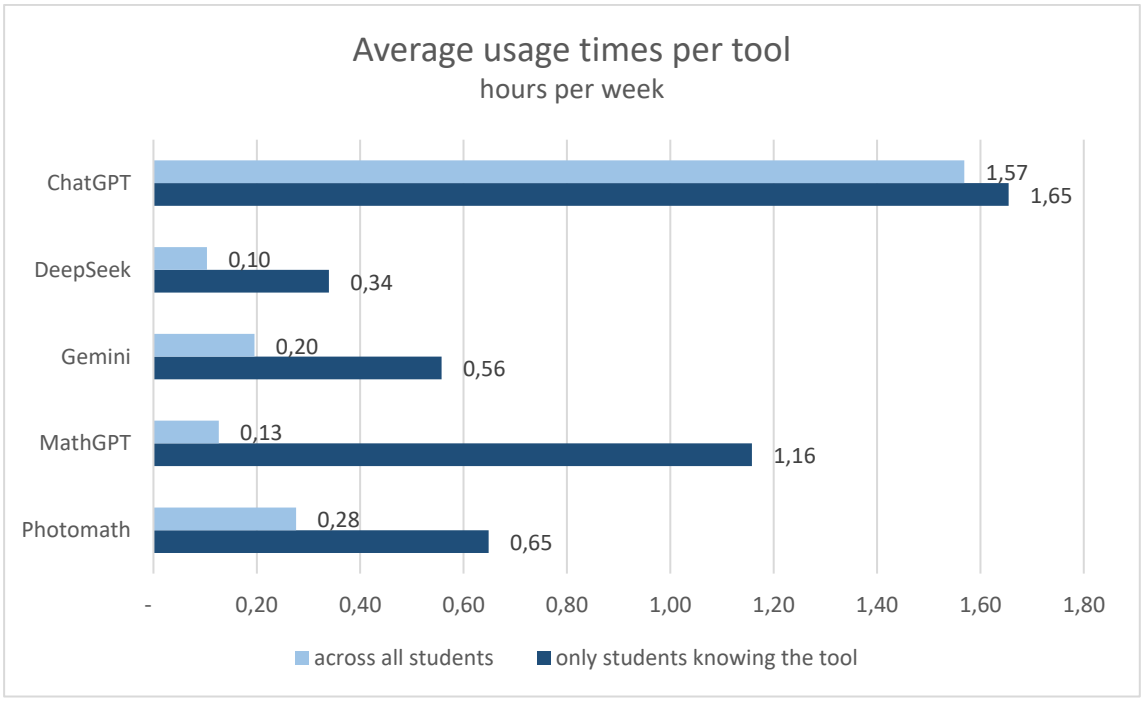


Figure 6

Rating of Math AI Tools

Based on the math AI tools known by students (Figure 1), the subsequent analysis focuses exclusively on the three most widely recognized AI tools—ChatGPT, Photomath, and Gemini—as these account for the vast majority of mentions and therefore represent the tools most relevant to students' everyday learning practices. The remaining tools were excluded from further analysis because their very small number of mentions at the rating questions does not allow for statistically meaningful or reliable comparisons.

Survey participants were asked to rate the performance of the math AI tools they know in four different categories: usability, understandability, correctness, and value for money. Figure 7 provides an overview of the results, showing the average ratings our students assigned to their top 3 known AI tools in the four categories (note that a rating was only included in the statistics if the participant indicated that he or she knows the tool):

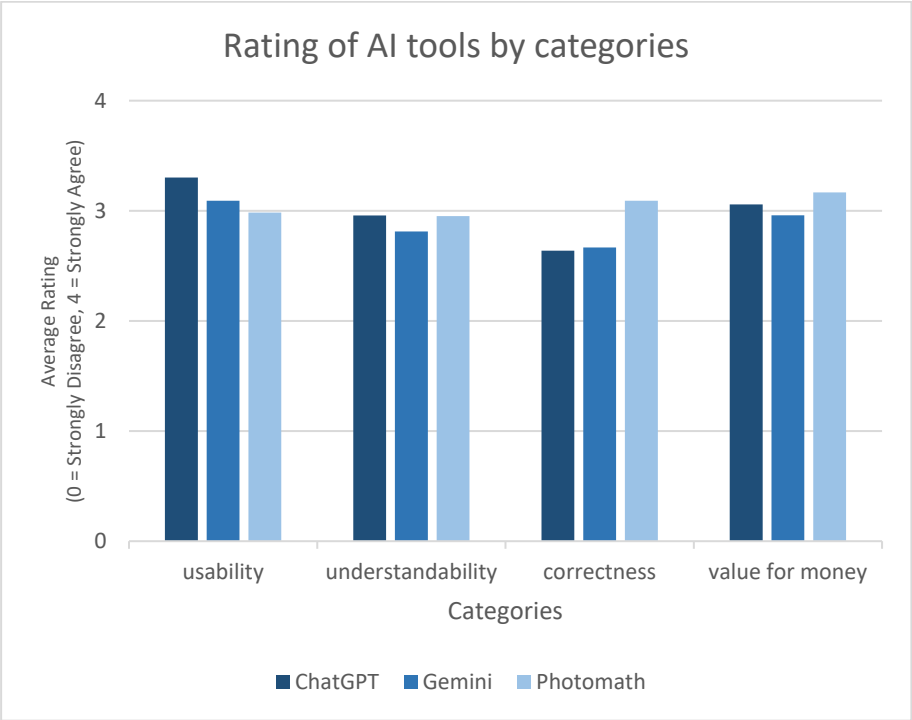


Figure 7

While average scores for the understandability and value-for-money categories are very close, the mean differences between the ChatGPT and Photomath ratings turn out to be significant on a 10% level for both usability ($n = 55$, $p = 0.00465$) and correctness ($n = 42$, $p = 0.07989$). Below, we investigate the usability and correctness categories in more detail.

USABILITY

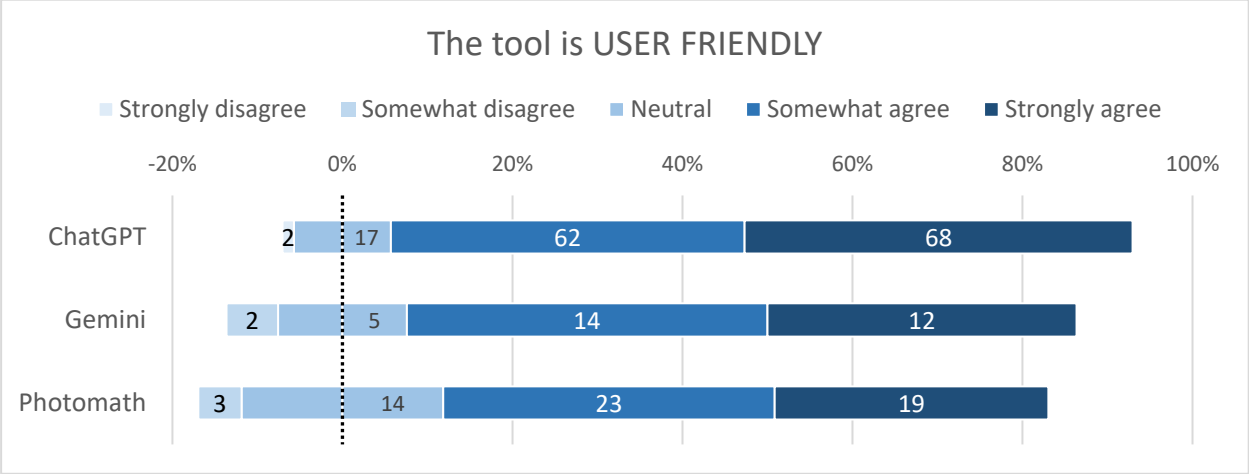


Figure 8

Figure 8 illustrates how students perceive the user-friendliness of the three AI tools ChatGPT, Gemini, and Photomath. Some differences emerge across the tools. ChatGPT receives overwhelmingly positive evaluations: the vast majority of respondents either somewhat agree (62) or strongly agree (68) that the tool is easy to use. Only a very small number expressed disagreement (2 strongly disagree, 17 neutral), indicating a broad consensus regarding ChatGPT's intuitive and accessible interface.

Gemini occupies an intermediate position. While a noticeable share of students evaluates the tool positively (14 somewhat agree and 12 strongly agree), a non-negligible group remains neutral (5). In addition, small numbers of respondents expressed disagreement (2 somewhat disagree).

Photomath shows the least favorable profile among the three tools. A substantial proportion of students rated the tool positively (23 somewhat agree and 19 strongly agree), while a smaller group remained neutral (14). Negative assessments were present but limited (3 somewhat disagree), indicating that Photomath is generally regarded as user-friendly, though not with the same level of confidence observed for ChatGPT.

Overall, the results underline ChatGPT as the clear benchmark for user-friendliness, with Gemini perceived as accessible but less consistently intuitive. Photomath, meanwhile, appears to be evaluated with greater reservation, reflecting a more heterogeneous user experience and less established usability among students.

CORRECTNESS

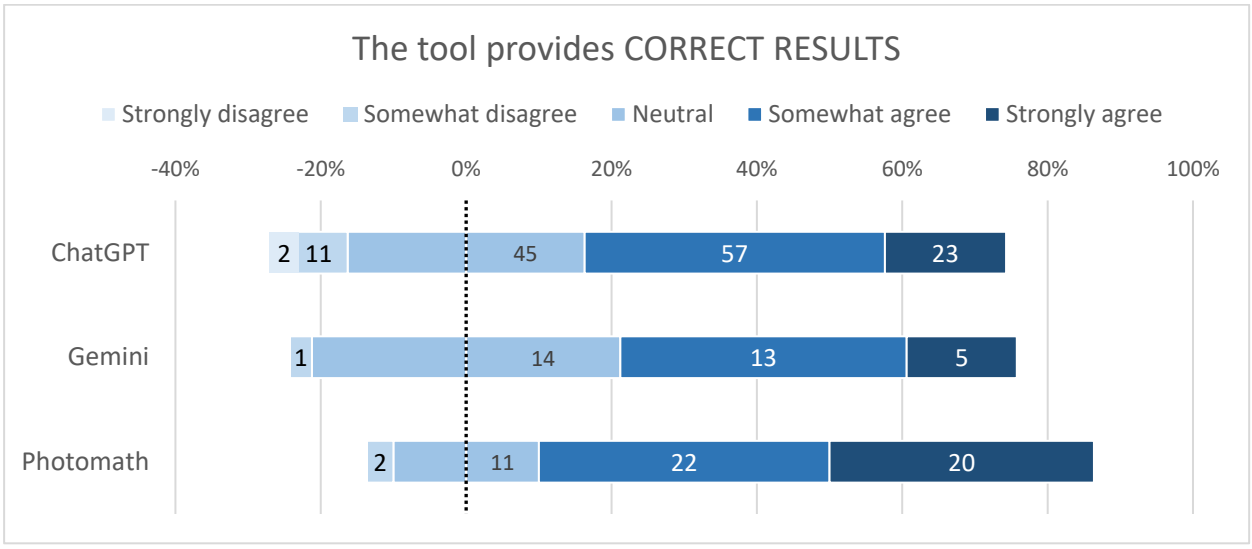


Figure 9

Figure 9 presents students' views on whether the three AI tools—ChatGPT, Gemini, and Photomath—deliver correct results. Overall, the distribution suggests that Photomath is perceived as the most reliable tool, whereas ChatGPT and Gemini show more cautious evaluations. Photomath receives the strongest endorsement, with 22 students selecting somewhat agree and 20 strongly agree, while only 11 remained neutral and disagreement was rare. This pattern is consistent with Photomath's specialized design as a mathematics-focused problem-solving tool.

In comparison, ChatGPT and Gemini show a more reserved trust profile. Although ChatGPT still receives more positive than negative ratings (57 somewhat agree, 23 strongly agree), the large neutral group (45) indicates that many students remain uncertain about the correctness of its mathematical output. A similar tendency appears for Gemini, where positive ratings (13 somewhat agree, 5 strongly agree) are accompanied by a substantial neutral share (14), suggesting limited confidence and less established credibility. Taken together, the results indicate that, while students acknowledge the usefulness of generative AI tools, they attribute the highest accuracy to Photomath, whereas ChatGPT and Gemini show greater hesitation regarding correctness.

Discussion

This study examines how artificial intelligence applications, which we use in almost every area of our lives in recent years, have changed students' use of mathematics at the university level. Through quantitative analysis and surveys, IU International University students studying in different departments were asked which artificial intelligence applications they used in quantitative courses such as mathematics or statistics. Students were asked to evaluate the tools of 10 leading companies in the field and assess these tools from certain perspectives such as usability, understandability, correctness, and value for money.

We observed that the students' current situation in math learning appears promising for the use of AI tools: The subject is commonly considered "difficult," and digital resources are preferred over traditional ones. When examining the students' awareness of tools, we found out that students from math-savvy subjects are more familiar with AI tools than non-math-savvy subjects. However, the familiarity with math AI tools did not differ between students who had just started university education and those who had not. Regarding the choice of tools, our survey results indicated that ChatGPT, Gemini, and Photomath applications dominate the market from almost every perspective and receive positive feedback from students from different perspectives. In addition, ChatGPT outperforms other applications in terms of usability, while Photomath emerges as the most appreciated application in terms of correctness. Despite this difference, when examining which product is used more, ChatGPT surpasses its competitors with approximately 1.5 hours of usage. This difference may stem from Photomath only being used by taking photos and not having a conventional prompt structure.

From a usage time perspective, the product that comes closest to ChatGPT is, surprisingly, MathGPT, with a weekly average of 1.16 hours among students knowing the tool. In contrast to ChatGPT, however, MathGPT is only known by approximately 10 percent of our survey participants. At this point, the most divergent result was observed with Gemini. Although familiar to a large group of students, Gemini lags far behind in usage times, with a majority of students knowing the tool stating that they do not use it in quantitative courses at all. This suggests that Gemini is used in areas other than mathematics.

As we proceed with this expanded investigation, we are aware that the landscape continues to evolve. The findings we report today may require revision as AI technologies advance and educational practices adapt. Nevertheless, by establishing a robust empirical baseline and identifying key patterns and relationships, this research contributes to an ongoing conversation about the future of mathematics education and, more broadly, the role of artificial intelligence in shaping how we teach and learn in the 21st century.

Outlook

With the capabilities of math AI tools increasing rapidly, all studies should be understood as flashlights on the current situation. Future changes to our findings are likely, underlining the need for ongoing research on the matter.

Our empirical results cover only the frontier models that were available up to mid-2025 (e.g., Gemini 2.5 Pro and OpenAI's o-series), and the central limitation our students observed was *correctness*: even when solutions looked coherent, subtle logical slips and unchecked arithmetic frequently led to wrong final answers. This has been a known limitation for many models for many years (Glazer et al., 2025). Since then, publicly reported evaluations suggest that newer generations have substantially improved mathematical correctness and reliability on harder, answer-verified benchmarks (most notably FrontierMath) where reported accuracy on Tiers 1-3 increased from roughly ~10% for Gemini 2.5 Pro preview to 40.3% for GPT-5.2 Thinking (*FrontierMath | Epoch AI*, n.d.; *Introducing GPT-5.2*, 2026)- see Figure 10.

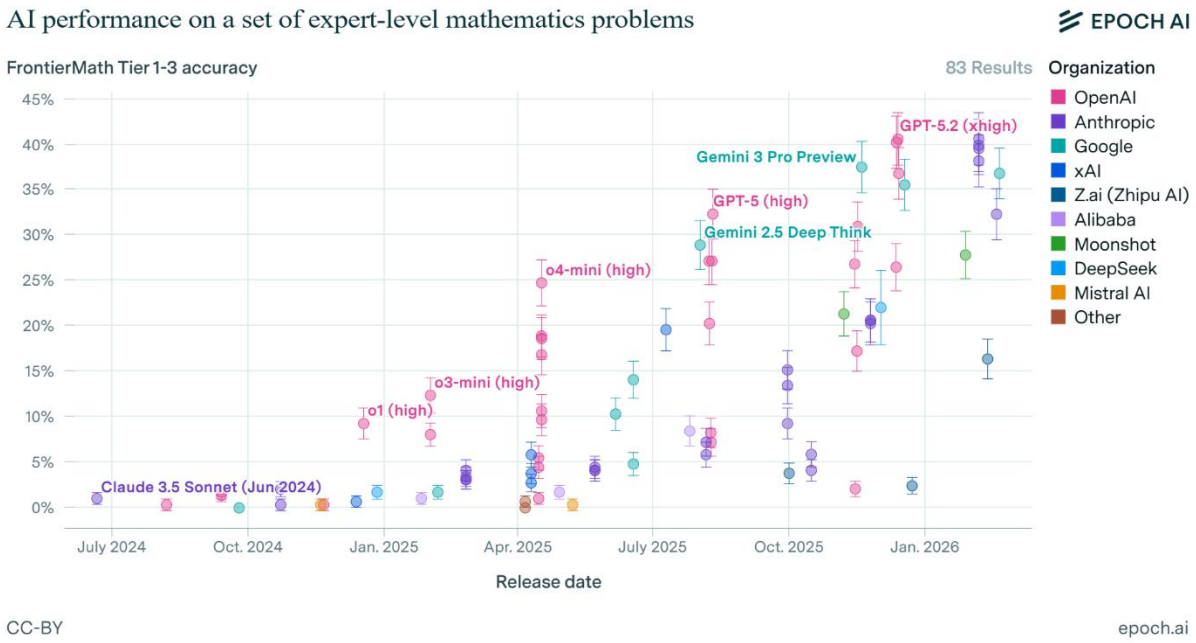


Figure 10: Taken from (*FrontierMath | Epoch AI*, n.d.).

In parallel, mathematical reasoning performance has also advanced, with Gemini Deep Think reported to achieve gold-medal standard on IMO 2025 problems (*Gemini 3 Deep Think*, 2026; Mirrokni, 2026). Taken together, these developments indicate that the correctness bottleneck documented in our mid-2025 study is being rapidly reduced, motivating new investigations on post-2025 models as well as newly-emerging tools (and updated protocols for verification, tool-use, and reproducibility). Compared to our current cross-sectional setup, a longitudinal study design appears promising both for tracing technological advancements and changes in the way students interact with math AI tools.

A second field of interest for future studies addresses the relationship between students' rating of math AI tools and their choice and usage times of the tools. Our current survey results across the rating categories of usability, understandability, correctness, and value-for-money suggest that students weigh these categories differently when it comes to choosing a tool. Future research should elaborate on and further refine rating categories, followed by carefully tracing their impact on tool usage.

Literaturverzeichnis:

Abramovich, S. (2021). Using Wolfram Alpha with Elementary Teacher Candidates: From More Than One Correct Answer to More Than One Correct Solution. *Mathematics*, *9*(17), 2112. https://doi.org/10.3390/math9172112

Altin, M. (2024). *Exploring the Influence of AI and Technical Innovations on Mathematics Students in Higher Education*. https://www.researchgate.net/publication/401615836_Exploring_the_Influence_of_AI_and_Technical_Innovations_on_Mathematics_Students_in_Higher_Education

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A Meta-analysis of the Relation Between Math Anxiety and Math Achievement. *Psychological Bulletin*, *147*(2), 134–168. https://doi.org/10.1037/bul0000307

Capinding, A. T. (2023). Revolutionizing Pre-Calculus Education: Photomath's AI-Powered Mathematics Tutorship. *Problems of Education in the 21st Century*, *81*(6), 758–775.

FrontierMath | Epoch AI. (n.d.). Retrieved February 25, 2026, from https://epoch.ai/benchmarks/frontiermath

Garcia, K. F., Ong, A. K. S., Gumasing, Ma. J. J., & Delos Reyes, C. R. V. (2025). Engineering students' perceptions and actual use of AI-based math tools for solving mathematical problems. *Acta Psychologica*, *256*, 105004. https://doi.org/10.1016/j.actpsy.2025.105004

Gemini 3 Deep Think: Advancing science, research and engineering. (2026, February 12). Google. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-deep-think/

Glazer, E., Erdil, E., Besiroglu, T., Chicharro, D., Chen, E., Gunning, A., Olsson, C. F., Denain, J.-S., Ho, A., Santos, E. de O., Järviniemi, O., Barnett, M., Sandler, R., Vrzala, M., Sevilla, J., Ren, Q., Pratt, E., Levine, L., Barkley, G., … Wildon, M. (2025). *FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI* (arXiv:2411.04872). arXiv. https://doi.org/10.48550/arXiv.2411.04872

Introducing GPT-5.2. (2026, February 13). https://openai.com/de-DE/index/introducing-gpt-5-2/

Mirrokni, T. L. and V. (2026, February 11). *Gemini Deep Think: Redefining the Future of Scientific Research*. Google DeepMind. https://deepmind.google/blog/accelerating-mathematical-and-scientific-discovery-with-gemini-deep-think/

Pascale, M. (2025, April 7). More Than Half of High School Students Say AI Use Can Reduce Math Anxiety, Global Survey Reveals. *M3 Challenge*. https://m3challenge.siam.org/newsroom/artificial-intelligence-math-anxiety/

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A Scoping Survey of ChatGPT in Mathematics Education. *Digital Experiences in Mathematics Education*, *11*(1), 9–41. https://doi.org/10.1007/s40751-025-00172-1

Wardani, H. K., Mazidah, E. N., & Hidayah, B. (2024). Comparing the Use of AI Tools in Mathematics and English Education: The Potential and Challenges of AI as Learning Assistant for FKIP UQ Students in Completing Academic Tasks. *Qomaruna Journal of Multidisciplinary Studies*, *1*(2), 10–24. https://doi.org/10.62048/qjms.v1i2.45