EXPLORING STUDENTS' USE OF AI SYSTEMS IN UNIVERSITY MATHEMATICS EDUCATION

Larissa Tröbelsberger[1], Alissa Fock[2], Reinhard Oldenburg[1], Hans-Stefan Siller[2]

[1]University of Augsburg, [2]Julius-Maximilians-University Würzburg

Artificial Intelligence (AI) is increasingly integrated into students' everyday learning practices, yet little is known about how mathematics students actually use AI systems in their university studies. This study investigates patterns of AI use among 112 mathematics students in their first four semesters at two German universities. A quantitative survey comprising five theoretically derived scales was administered and analysed. The results indicate that AI plays a substantial role in clarifying mathematical content, while lecture notes remain a very important learning resource. Students show a high level of error awareness, but this awareness does not influence their usage patterns. Strong associations were found among productive and substitutive use, leaning-supportive use and optimizing AI-Interaction, forming a triangular structure of mutual implications.

INTRODUCTION

Artificial Intelligence (AI) has advanced rapidly in the recent years, driving profound transformations across multiple domains and became an integral part of everyday life, particularly within education (OECD, 2023). As technological developments continuously offer new opportunities, the future professional and personal lives of students will increasingly be shaped by AI technologies. Therefore, the competent and reflective use of technology, including the use of AI-technologies, will be essential for an active participation in the future digital society (Ergas et al., 2022). Despite this, there remains a considerable lack of research on AI in education that involves human participants (Fock & Siller, 2025). Most existing studies conceptualise AI primarily as an object of study rather than examining its role as a tool for learning and knowledge construction. To advance research in mathematics education, this contribution captures the current state of how students make use of AI systems as learning tools within their studies.

AI IN UNIVERSITY MATHEMATICS EDUCATION

The term "Artificial Intelligence" is used in diverse ways across research and practice (see, e.g., Russell & Norvig, 2022). One rapidly developing branch of AI that has gained particular attention since the release of ChatGPT by OpenAI in late 2022 is generative AI. For the purpose of this paper, *AI* only refers to generative AI, which includes systems capable of producing new content, such as texts, images or videos, based on patterns leaned from extensive training data (Taulli, 2023).

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 1483-1490. PME.

Recent studies on AI performance in mathematics indicate that current systems still tend to produce incorrect or low-quality solutions (see, e.g. Fang et al., 2024; Sun et al., 2024). However, over time, substantial improvements have already been achieved and turned AI into a tool that is appealing even to students of university level math. In higher education research, several studies have begun to explore how AI tools interact with students' learning behaviours and cognitive processes. Gerlich (2025) identified a negative impact of the use of AI tools on critical thinking, suggesting that increased reliance on AI can lead to cognitive offloading and a decrease in critical engagement with the AI. This highlights the importance of examining how AI is used during the learning process in order to assess whether its use supports learning-oriented engagement. Furthermore, Noster et al. (2024) highlight prompting as an important AI-user skill; this may help to improve output quality, which is of considerable importance for learning. In addition, to ensure the quality of AI-supported work, Wardat et al. (2023) recommend strategies such as cross-checking and additional verification of answers provided by AI. This practice supports the correctness of AI-generated outcomes, which is of course an essential component of meaningful knowledge acquisition. Closely related to this facet of student behaviour is the aspect that students (and generally all AI users) need to develop error awareness while working with AI in order to recognize and evaluate AI-generated errors. This affects the perception of AI and students' perception can play a major role in whether students accept or reject AI tools in their learning process (Zhuang, 2025).

RESEARCH QUESTIONS

The study aims to provide an empirical foundation for the further development of university-level mathematics education in the context of increasing AI integration. We want to get an insight into how students use AI in their learning processes. To this end, we focus on the following research questions:

(RQ1) *In what way do mathematics students use available AI systems to support their university learning?*

(RQ2) *To what extent do different modes of AI use relate to each other?*

METHOD

A questionnaire was conducted among mathematics students and related study programs in their first four semesters at the universities of Augsburg and Würzburg. The choice of this sample by academic experience was motivated by the idea to have users that had some time to adopt their learning strategies to university level math programs while not yet working in a specialised area of math that might interact with tool use. In total, 112 students (43 female, 67 male, one non-binary, one without specification) participated the study.

The data were collected through an online questionnaire via SoSciSurvey. Participation was voluntary and anonymous. The questionnaire comprised five item groups – each answered on a 7-point Likert scale – representing distinct modes of AI use derived

from the scientific findings of Gerlich (2025), Noster et al. (2024), Wardat et al. (2024) and Zhuang (2025): *Productive and Substitutive Use, Learning- and Understanding-Supportive Use, Optimizing Interaction with AI, Quality Ensuring Use* and *Error Awareness*. By the first two item groups, participants were also asked to indicate the extent of their offloading tendency. *Productive and Substitutive* Use refers to students' use of AI to generate content, such as texts of solutions of exercise sheets, as well as to the use to replace conventional learning resources and activities. This includes using AI instead of textbooks, explanatory videos or search engines. *Learning- and Understanding-Supportive Use of AI* refers to the use of AI systems as a cognitive aid to support comprehension and conceptual understanding in mathematics learning. It involves employing AI to explain unclear terms or statements, to provide and work through examples, to deepen understanding of lecture content or to check one's own understanding. In this usage mode, AI functions primarily as an explanatory and learning companion rather than as a substitute for students' own work. The other three item groups refer to how students improve and verify AI-generated answers and perceive AI errors.

Table 1: Sample Items

Scale	**Sample Item**
Productive and Substitutive Use	I use AI to generate texts for seminar papers or solutions for home assignments.
Learning- and Understanding-supportive Use	I use AI to explain terms or sentences that I did not understand.
Optimizing Interaction with AI	When formulation prompts, I sometimes include extra guidance on the required approach, such as requesting step-by-step reasoning.
Quality Ensuring Use	For challenging questions, I use multiple AI systems and compare their answers.
Error Awareness	I have occasionally received AI-generated answers that I was certain were incorrect.

The data analysis was carried out using R. To examine the internal consistency of the scales, Cronbach's alpha coefficients were calculated. To answer the RQ1, means und standard derivations were calculated. RQ2 was investigated by a correlation analysis as well as structural equation modelling. This enables to determine which usage modes were interrelated. Finally, an entropic implicative analysis (SIA, Gras et al. 2008) was performed to explore whether and how significantly correlated forms of AI engagement imply or depend on one another.

RESULTS

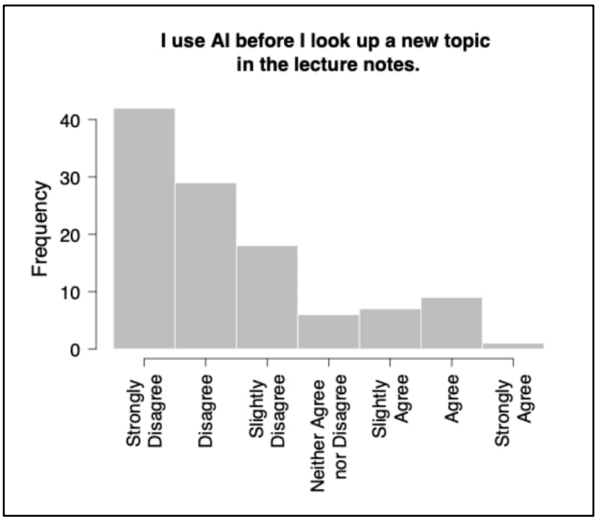
The resulting Cronbach's alpha coefficients ranged from 0.60 to 0.90 (see Tab. 2). All except one lie within the accepted range (Nunnally & Bernstein, 1994), indicating a good level of internal consistency across the items of these four scales. This suggests that the targeted constructs were measured reliably. The comparatively low Cronbach's alpha of the scale *Optimizing Interactions with AI* suggests limited internal consistency, which may be due to the heterogeneous nature of the items capturing different facets of interactions with AI. But because of the exploratory nature of this study, the scale was retained. However, conclusions drawn from this scale remain tentative. Table 2 also presents the means and standard derivations of all scales.

Table 2: Results of the reliability analysis

Scale	Cronbach's Alpha	Mean	SD
Productive and Substitutive Use	0.81	3.39	1.44
Learning- & Understanding-supportive Use	0.90	4.06	1.83
Optimizing Interaction with AI	0.60	3.82	1.42
Quality Ensuring Use	0.71	4.34	1.42
Error Awareness	0.86	6.57	0.9

One central finding of the survey is the relevance of AI for supporting students' understanding of mathematical content. Many students reported using AI-systems to have new concepts explained or to receive worked examples, as reflected in the mean of scale 2 (see Tab. 2). Some of the items related to this finding ask about the replacement of traditional resources. Despite the reported intensive use of AI, many students indicated that AI does not replace the lecture notes as their central learning resource (see Fig. 1).

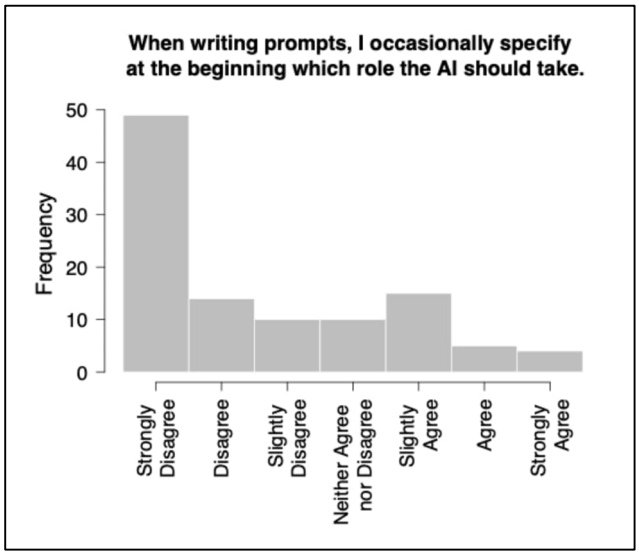
Figure 1: AI replaces lecture notes



A second descriptive finding concerns the scale *Error Awareness*, which consisted of three items capturing students' awareness of the potential inaccuracy and fallibility of AI systems. The high mean and low standard derivation of this scale (see Tab. 2) clearly indicate that students consistently demonstrate a well-developed awareness of possible AI errors. Participants strongly agreed that AI systems can produce incorrect or misleading answers, and many reported having previously encountered such erroneous AI-generated outputs in their own use. Overall, the results demonstrate that students approach AI-generated output with a considerable degree of caution, reflecting a generally strong awareness of the limitations of current AI systems.

In addition, students' strategies for interacting with AI systems were examined. As illustrated in Figure 2, their strategies to improve the process-related quality of the AI output are used only rarely by students, despite the fact that many of them reported having received incorrect AI responses.

Figure 2: Optimizing Interaction with AI



It is noteworthy that cross-checking with analogue and non-AI strategies, e.g. by using lecture notes or manual calculations, are considerably more common than digital cross-checking using additional AI systems or tools. A confirmatory factor analysis with a measurement model for all five constructs was conducted with the R package lavaan (Rosseel, 2024). The model shows a good fit (CFI = 0.90, TLI = 0.86, RMSEA = 0.07) so that its estimates can be interpreted with confidence. The results of the structural equation modelling together with the correlation matrix reveal substantial associations between several usage modes. Given the high level of error awareness reported by the students, this scale showed no correlation higher than 0.15. In contrast, strong correlations were found between *Productive and Substitutive Use* and *Learning- and Understanding-Supportive Use* ($r = 0.85$, $p < 0.001$), between *Optimizing Interaction with AI* and *Productive and Substitutive Use* ($r = 0.69$, $p < 0.001$), and between *Quality-Ensuring Use* and *Optimizing Interaction with AI* ($r = 0.65$, $p < 0.001$).

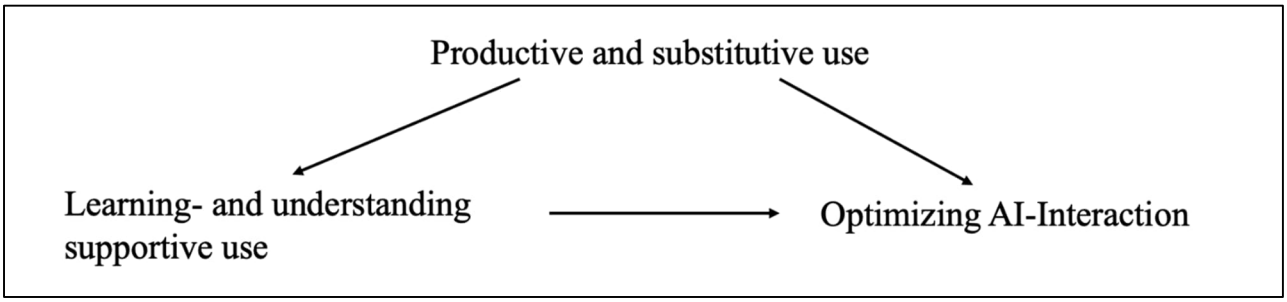
Since these three scales showed correlations, an implicative analysis was conducted to determine which of the three usage modes imply or depend on one another. The entropic implication intensities ϕ are presented in Table 3.

Table 3: Results of the SIA

Implication	ϕ
Productive and Substitutive Use → Optimizing AI-Interaction	0.61
Productive and Substitutive Use → Learning- and Understanding-Supportive Use	0.60
Learning- and Understanding-Supportive Use → Optimizing AI-Interaction	0.54

The graphical representation of the values (see Fig. 3) shows a triangular structure of implications, in which *Productive and Substitutive Use* spears to function as a superordinate usage pattern.

Figure 3: Triangular structure of implications



DISCUSSION

The findings provide important insights into how students currently integrate AI systems into their learning process in university mathematics. First, AI holds a high level of relevance for clarifying mathematical content. Nonetheless, lecture notes remain the primary learning medium for most students and are not perceived as being replaced by AI. Second, despite autodidactic use, students' usage mode is not primarily driven by dishonest intentions or a desire to reduce effort. According to their self-reports, strategic misuse or shortcut-oriented use plays only a minor role. These findings indicate that cognitive offloading does not appear to be the dominant form of AI use, which, following Gerlich (2025), may be considered beneficial for learning. Moreover, the productive and substitutive usage mode appears to function as a superordinate pattern of AI engagement, which suggests that it serves as a valuable starting point for developing students' AI-related skills. Additionally, the usage mode *Quality Ensuring Use* is relatively pronounced, indicating that students apply strategies such as cross-checking and additional verification as recommended by

Wardat (2023). This suggests a correctness-oriented approach to handling AI-generated results. Beyond this, students' engagement with AI is not uncritical. They are aware of potential AI errors, likely due to prior encounters with flawed responses. Yet, the insignificant correlations show that this awareness has no measurable impact on productive or substitutive use, learning-supportive use, or optimizing strategies, although such an effect would be desirable. This is not consistent with the finding by Zhuang's (2025) study, which suggests that a high perception of errors conveys a negative image of AI and leads to rejection, thereby influencing usage behaviour. Furthermore, clear and precise prompt formulation as well as explicit role assignments are fundamental strategies for optimizing AI-generated responses. Such strategies may help to enhance the output quality (Noster et al., 2024), however, these approaches are rarely applied by students. A plausible explanation for both findings could be that students use AI pragmatically under time pressure and often lack explicit strategies for checking AI outputs or other user skills. Without structured guidance on reflective AI use, they struggle to turn awareness into consistent quality-assuring practices, precisely where university teaching must intervene.

When interpreting these results, several limitations must be considered. The relatively small sample restricts generalisability, and the self-report data may be shaped by subjective bias or social desirability. Moreover, similar scale scores can mask divergent intentions – for instance, both convenience-driven avoidance and a desire for deeper understanding may prompt students to use AI as an explanatory tool, despite fundamentally different motivations.

The findings open several important pathways for future research. A longitudinal perspective could provide insights into how mathematics students' use of AI develops over the course of their studies or, generally, over time. Further research is also needed to clarify how different forms of AI engagement relate to learning outcome and academic success. In addition, intervention studies could help identify which instructional designs and task formats effectively foster reflective AI use (Fock & Siller, 2025) and the development of optimizing strategies. Another interest of future work concerns the persistence of AI-related behaviour, particularly the question of why students continue to rely on AI systems even if they are aware of the potential inaccuracies.

References

Ergas, O., Gilead, T., Eilam, E., Kyza, E., Akar, B., Buck, M. F., & Kumar, S. (2022). A curricular framework for flourishing in education. In *The International Science and Evidence-based Education Assessment:ISEE Assessment Working Group 1* (S. 180–250). UNESCO.

Fang, M., Wan, X., Lu, F., Xing, F., & Zou, K. (2024). *MathOdyssey: Benchmarking Mathematical Problem-Solving Skills in Large Language Models Using Odyssey Math Data* (No. arXiv:2406.18321). arXiv. https://doi.org/10.48550/arXiv.2406.18321

Fock, A., & Siller, H.-S. (2025b). Generative artificial intelligence in secondary STEM education in the light of Human Flourishing: A scoping literature review. *International Journal of STEM Education*, *12*(1), 67. https://doi.org/10.1186/s40594-025-00589-5

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Gras, R., Suzuki, E., Guillet, F., & Spagnolo, F. (Hrsg.). (2008). *Statistical Implicative Analysis: Theory and Applications* (Bd. 127). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78983-3

Noster, N., Gerber, S., & Siller, H.-S. (2024). Pre-Service Teachers' Approaches in Solving Mathematics Tasks with ChatGPT. *Digital Experiences in Mathematics Education*, *10*(3), 543–567. https://doi.org/10.1007/s40751-024-00155-8

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3. ed., [Nachdr.]). McGraw-Hill.

OECD. (2023). *Is Education Losing the Race with Technology?: AI's Progress in Maths and Reading*. OECD Publishing. https://doi.org/10.1787/73105f99-en

Rosseel, Y. (2024). lavaan: Latent variable analysis (R package Version 0.6-20). https://CRAN.R-project.org/package=lavaan

Russell, S. J., & Norvig, P. (with Chang, M., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I., Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J.). (2022). *Artificial intelligence: A modern approach* (Fourth edition, global edition). Pearson.

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, *11*(1), 1278. https://doi.org/10.1057/s41599-024-03811-x

Taulli, T. (2023). *Generative AI: How ChatGPT and Other AI Tools Will Revolutionize Business*. Apress. https://doi.org/10.1007/978-1-4842-9367-6

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Zhuang, Y. (2025). Lessons from Using ChatGPT in Calculus: Insights from Two Contrasting Cases. *Journal of Formative Design in Learning*, *9*(1), 25–35. https://doi.org/10.1007/s41686-025-00098-2