International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 5, Sep-Oct 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

AI Tools in College Mathematics Education: A Survey on Current Usage, Challenges, and Student Needs

He Yuxuan[1], Yan Chuhan[2], Yang Junyi[3], Deng Hao[4], Yang Xingyu[5]

[1]School of Systems Science and Statistics, Beijing Wuzi University, Beijing, China
[2,3,4,5]School of Information, Beijing Wuzi University, Beijing, China

ABSTRACT

With the rapid development of large language models, AI-powered tutoring tools have been widely adopted in college mathematics learning. However, students often face difficulties in selecting appropriate tools due to a lack of systematic understanding of different AI models' capabilities. This study conducted a questionnaire survey among 100 college students who have taken mathematics courses, investigating their usage patterns, tool preferences, perceived accuracy, and major concerns when using AI for mathematics problem-solving. The results show that 97% of students have used AI tutoring tools, with 61.86% preferring general-purpose large language models (e.g., Doubao) and 38.14% preferring subject-specific search-based AI (e.g., Kuaidui). Major concerns include the inability to identify AI-generated errors (68%), overly general explanations (48%), and over-reliance on AI leading to reduced independent thinking ability (41%). Students' primary demands are complete solution steps (58%), answer accuracy (54%), and textbook-aligned explanations (54%). These findings reveal the existing gaps in current AI mathematics tutoring tools and provide empirical evidence for future tool selection and optimization.

KEYWORDS: *Artificial Intelligence, AI Tutoring, Mathematics Education, Large Language Models, Educational Technology.*

How to cite this paper: He Yuxuan | Yan Chuhan | Yang Junyi | Deng Hao | Yang Xingyu "AI Tools in College Mathematics Education: A Survey on Current Usage, Challenges, and Student Needs" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-5, October 2026, pp.90-98, URL: www.ijtsrd.com/papers/ijtsrd142197.pdf

IJTSRD142197

Copyright © 2026 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0)

1. INTRODUCTION

The rapid advancement of large language models has significantly transformed the landscape of educational technology. AI-powered tutoring tools have become increasingly prevalent in higher education, particularly in mathematics learning, where students frequently seek additional support beyond traditional classroom instruction. These tools range from general-purpose large language models such as Doubao and DeepSeek to subject-specific search-based AI platforms like Kuaidui and Zuoyebang, each offering different approaches to mathematical problem-solving assistance.

Despite the growing popularity of these AI tools, a significant gap exists between student adoption and their understanding of these tools 'capabilities and limitations. Many students lack systematic knowledge about which types of AI tools are most suitable for different mathematical tasks, leading to inefficient tool selection and, in some cases, negative learning outcomes. Students may blindly trust AI-generated solutions without critical evaluation, potentially compromising their mathematical reasoning skills and conceptual understanding.

The integration of AI tools into mathematics education presents both opportunities and challenges. On one hand, these tools provide immediate feedback, step-by-step solutions, and access to vast knowledge bases that can support student learning outside the classroom. On the other hand, the accuracy and pedagogical quality of AI-generated solutions vary significantly across different tools and problem types. The variability in performance raises important questions about how students can effectively leverage these tools without undermining their own learning and critical thinking development.

Furthermore, the rapid proliferation of AI tools has outpaced the development of guidance on their appropriate use in educational settings. Students often receive little to no instruction on how to

critically evaluate AI outputs, how to identify potential errors, or how to use AI as a learning aid rather than a substitute for their own reasoning. This lack of guidance exacerbates the risks associated with AI tool usage, particularly in subjects like mathematics where conceptual understanding and reasoning skills are paramount.

This study addresses these gaps by investigating the current usage patterns, challenges, and needs of college students using AI tutoring tools for mathematics learning. Specifically, we aim to: (1) understand the prevalence and frequency of AI tool usage among college mathematics students, (2) identify students' tool preferences and their perceptions of AI accuracy, (3) uncover the major problems and concerns students encounter when using AI for mathematical problem-solving, and (4) determine students 'primary demands for AI tutoring tools. The findings provide empirical evidence for guiding students in more effective AI tool selection and inform future developments in AI-assisted mathematics education.

The remainder of this paper is organized as follows. Section 2 describes the research methodology, including survey design, participant characteristics, and data analysis procedures. Section 3 presents the detailed results of the survey across all dimensions. Section 4 discusses the implications of the findings in relation to existing research and educational practice. Section 5 concludes the paper with a summary of contributions, limitations, and directions for future work.

2. METHODOLOGY

2.1. Survey Design and Procedure

A questionnaire titled "Survey on the Current Usage of AI in College Mathematics Learning" was designed to collect data on students' AI tool usage behaviors. The survey covered multiple dimensions, including demographic information, tool usage frequency, specific tools used, perceived accuracy of AI-generated solutions, common problems encountered, functional requirements for AI tutoring tools, and open-ended feedback on overall experiences.

The questionnaire was developed through an iterative process that included literature review, expert review by the supervising professor, and pilot testing with a small group of students. The final version consisted of 12 questions organized into four sections: demographic information, AI usage behavior, problems and concerns, and functional requirements.

The questionnaire was distributed online through the Wenjuanxing platform, a widely used survey platform in China. The survey link was disseminated through class WeChat groups, university forums, and course learning platforms to reach a broad range of students across different majors and academic years. A total of 132 responses were collected over a two-week period. After data cleaning, which excluded invalid responses from students who had not taken college mathematics courses and incomplete responses, 100 valid responses were retained for analysis.

All respondents were undergraduate students who had completed or were currently taking mathematics courses. Among them, 97 students reported using AI tutoring tools, representing a usage rate of 97.00%. The high response rate and the comprehensive coverage of the survey dimensions provide a solid foundation for the subsequent analysis.

2.2. Participants

The 100 valid responses covered various academic years and majors. Table 1 presents the detailed demographic characteristics of the sample.

Table 1. Demographic Characteristics of Participants

Variable	Category	Frequency	Percentage
Academic Year	Freshman	18	18.00%
	Sophomore	35	35.00%
	Junior	20	20.00%
	Senior	18	18.00%
	Other	9	9.00%
Major	Science & Engineering	46	46.00%
	Economics & Management	39	39.00%
	Humanities & Arts	11	11.00%
	Other	4	4.00%

The largest group was sophomore students, accounting for 35 respondents (35.00%). This is consistent with the curriculum structure at the university, as sophomore students are typically enrolled in core mathematics courses including Calculus, Linear Algebra, and Probability Theory, leading to higher demand for AI-assisted learning tools. Junior students represented the second largest group with 20 respondents (20.00%), while freshmen and senior students each contributed 18 respondents (18.00%). The remaining 9 respondents (9.00%) were from other academic categories, including students in extended programs.

In terms of academic disciplines, science and engineering majors represented the largest group with 46 respondents (46.00%). Economics and management majors comprised 39 respondents (39.00%), which is consistent with the university's focus on both STEM and business education. Humanities and arts majors accounted for 11 respondents (11.00%), and other majors comprised the remaining 4 respondents (4.00%). This distribution provides a comprehensive representation of the student population at the university and allows for meaningful comparisons across different disciplinary backgrounds.

2.3. Data Analysis

Data analysis was conducted using SPSSPRO, a statistical analysis platform that provides a comprehensive suite of tools for social science research. Descriptive statistics, including frequency analysis and percentage calculations, were employed to characterize the basic demographic and usage patterns.

For multiple-response questions, including AI-related problems, usage concerns, and functional requirements, multiple response analysis and chi-square goodness-of-fit tests were applied. The multiple response analysis allowed for the examination of response patterns across multiple categories, while the chi-square goodness-of-fit tests determined whether the distribution of responses across categories was statistically significant.

Open-ended responses were coded and categorized through manual text analysis to extract qualitative insights. A thematic coding approach was used, with initial codes derived from the responses and subsequently grouped into broader categories. Both quantitative and qualitative data were integrated to form comprehensive conclusions that capture both the breadth and depth of students' experiences with AI tutoring tools.

The use of both quantitative and qualitative methods in this study follows a complementary approach, where the survey data provide statistical evidence of patterns and trends, while the open-ended responses offer rich contextual information that helps explain the quantitative findings. This mixed-methods approach enhances the validity and completeness of the research conclusions.

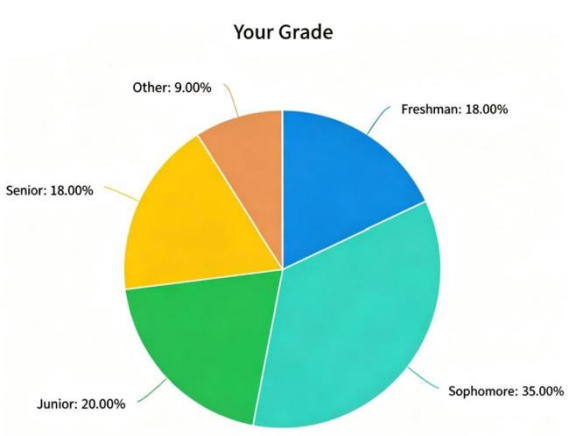
3. RESULTS

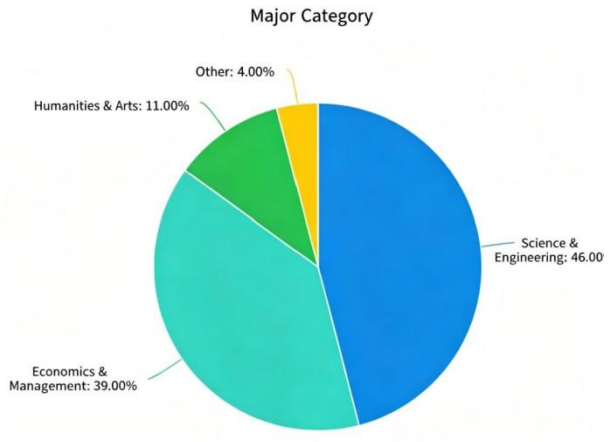
3.1. AI Tool Usage Frequency

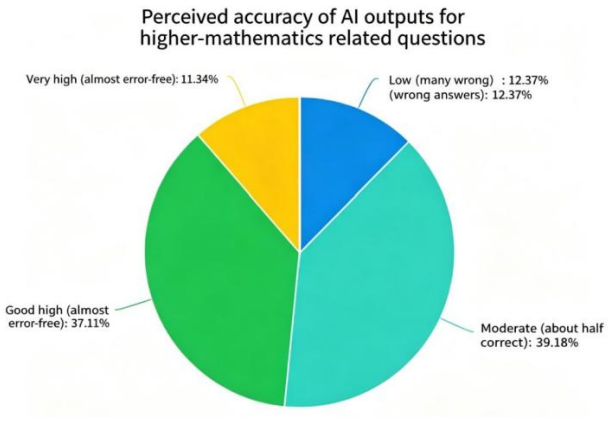
Among the 97 students who had used AI tutoring tools, 40 students (41.24%) reported using AI tools frequently, defined as 1-2 times per week, representing the dominant usage pattern. 30 students (30.93%) reported high-frequency usage, using AI tools almost daily. 27 students (27.84%) reported occasional usage, no more than 4 times per month.

The high proportion of students using AI tools on a weekly or daily basis indicates that AI has become a routine supplementary tool in students'study habits, rather than an occasional resource. This consistent usage suggests that students perceive tangible benefits from these tools, whether in terms of immediate problem-solving assistance or longer-term learning support.

The frequent usage pattern is particularly notable among sophomore students, who are simultaneously enrolled in multiple mathematics courses. This suggests that course load and academic demands may be significant drivers of AI tool adoption.





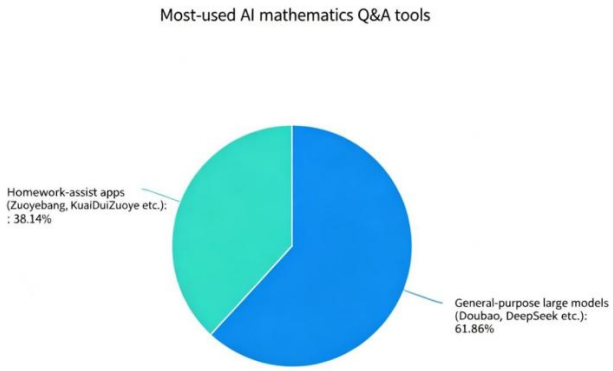


3.2. Tool Preferences

In terms of tool preferences, 60 students (61.86%) preferred general-purpose large language models, such as Doubao and DeepSeek, making these the primary choice among students. The remaining 37 students (38.14%) preferred subject-specific search-based AI tools, such as Zuoyebang and Kuaidui.

The preference for general-purpose models was notably higher, suggesting that students value versatility and broader functionality. General-purpose models offer the advantage of handling a wide range of queries beyond mathematics, potentially making them more convenient for students who use AI for multiple subjects. The lower preference for search-based AI tools, despite their specialized focus on problem-solving, may reflect limitations in their perceived capabilities or user experience.

However, both tool categories had sufficient user bases to warrant further comparative investigation. The nearly 40% preference rate for search-based tools indicates a substantial segment of students who find specialized tools more suitable for their needs. The factors driving these different preferences warrant further investigation.

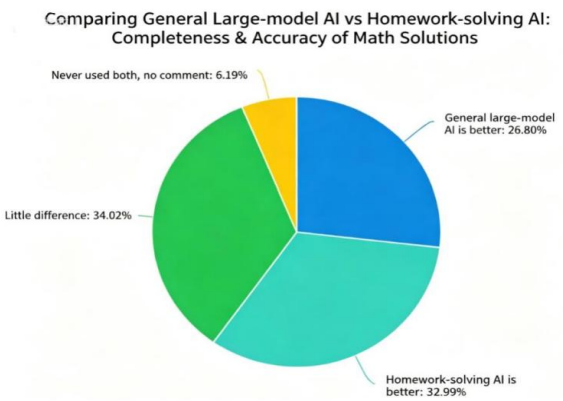


3.3. Perceived Accuracy of AI Solutions

When asked to evaluate the accuracy of AI-generated solutions, 38 students (39.18%) rated the accuracy as moderate, indicating that approximately half of the AI solutions were correct. This moderate rating suggests that students perceive AI tools as neither consistently reliable nor consistently unreliable, but rather dependent on the specific problem and context.

36 students (37.11%) perceived accuracy as relatively high, with most solutions being correct. This group likely represents students who have developed effective strategies for using AI tools, including appropriate prompting techniques and careful evaluation of outputs. 12 students (12.37%) reported low accuracy with many errors, while only 11 students (11.34%) considered accuracy to be very high with almost no errors.

Combined, over 50% of students perceived the accuracy of AI mathematics solutions as insufficient, highlighting a significant reliability concern. This perception aligns with findings from prior research indicating that AI models often struggle with complex mathematical reasoning and symbolic manipulation. The high proportion of students expressing concerns about accuracy suggests that tool developers should prioritize improving mathematical reasoning capabilities.



3.4. Comparison of Tool Types

Among students who had experience with both tool types, their subjective evaluations were notably divided. 33 students (34.02%) perceived little difference between the two types of tools in terms of solution step clarity. 32 students (32.99%) believed that search-based AI provided more understandable solution steps. 26 students (26.80%) favored general-purpose large language models for better solution clarity. 6 students (6.19%) had not used both types and could not provide a comparison.

The divided opinions indicate that neither tool type has an absolute advantage in students' subjective judgment, suggesting the need for objective, systematic comparisons. The near-even split between the two preferences suggests that the relative effectiveness of each tool type may depend on factors such as problem type, student background, and specific usage contexts.

Students who preferred search-based AI often noted the structured, step-by-step presentation of solutions as a key advantage. In contrast, students who favored general-purpose models appreciated the flexibility and the ability to engage in conversational follow-up questions. These differences in user experience suggest that the optimal tool choice may vary depending on the specific learning goal and task.

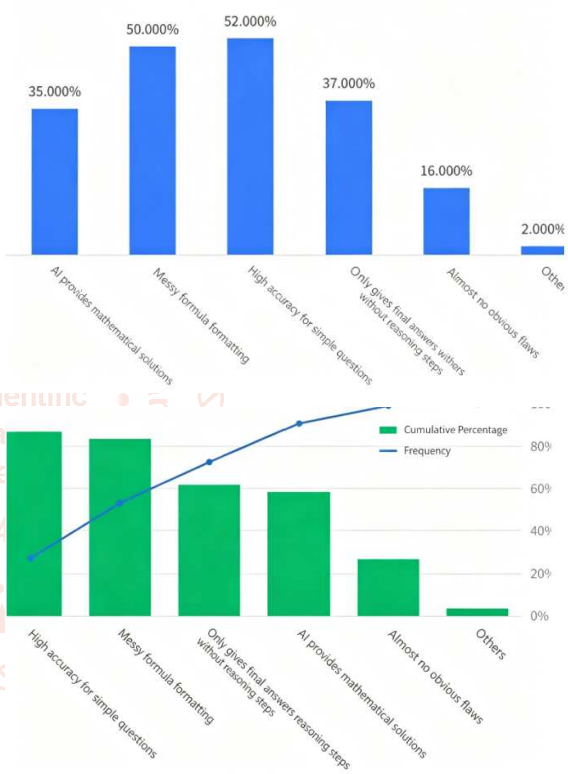
3.5. Common Problems with AI-Generated Mathematics Solutions

The multiple-response analysis revealed several significant problems with AI-generated mathematics solutions. The most frequently reported issue was that AI performed well on simple problems but frequently made errors on complex or comprehensive problems, reported by 52 students (52% prevalence). This pattern suggests that current AI tools may rely on pattern recognition and retrieval of common solutions, but struggle with multi-step reasoning and novel problem-solving scenarios.

Formula formatting errors and mathematical symbol confusion were reported by 50 students (50%). These formatting issues are particularly problematic in mathematics education, where the accurate representation of symbols, equations, and mathematical structures is essential for comprehension. Students who receive poorly formatted solutions may be unable to follow the reasoning or correctly interpret the solution.

37 students (37%) reported that AI provided only final answers without explaining the underlying principles. This lack of explanation limits the pedagogical value of AI tools, as students may not understand the reasoning behind the solution. 35 students (35%) noted that even when the final answer was correct, the solution steps were often skipped and difficult to understand, further reducing the learning value.

Only 16 students (16%) reported no significant problems with AI solutions. The chi-square goodness-of-fit test confirmed that these differences were statistically significant ($p < 0.001$). These results indicate that technical shortcomings are concentrated in three main areas: complex problem-solving, mathematical notation rendering, and incomplete solution derivation.



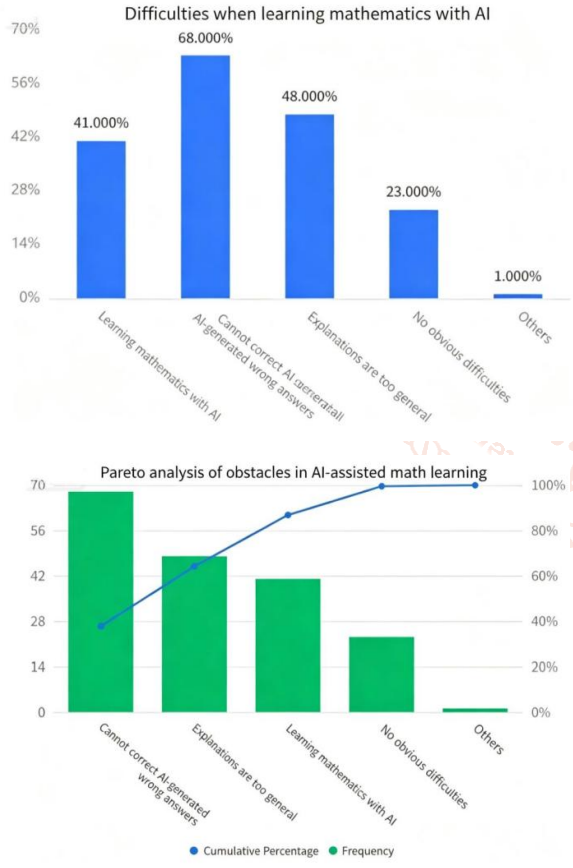
3.6. Core Concerns in Using AI for Mathematics Learning

When asked about their primary concerns when using AI for mathematics learning, 68 students (68%) cited the inability to identify AI errors, indicating that incorrect solutions can directly mislead student learning. This represents the most significant concern by a considerable margin, highlighting the challenge of detecting errors in AI outputs. Students who cannot distinguish between correct and incorrect solutions may inadvertently learn incorrect concepts or develop flawed reasoning strategies.

48 students (48%) reported that the explanations provided were too general and difficult to understand. This concern reflects the pedagogical

limitations of current AI tools, which often generate explanations that are not tailored to students 'existing knowledge or specific areas of confusion. 41 students (41%) expressed concerns about over-reliance on AI tools leading to a decline in independent thinking ability. This concern reflects students 'awareness of the potential negative effects on learning, including the erosion of critical thinking and problem-solving skills.

Only 23 students (23%) reported no significant concerns. The chi-square goodness-of-fit test confirmed significant differences in the distribution of these concerns ($p < 0.001$). The most prominent risk was the difficulty in detecting AI errors, followed by poor pedagogical adaptability and the potential erosion of students' problem-solving skills. These findings suggest that addressing error detection and pedagogical quality should be top priorities for tool developers.



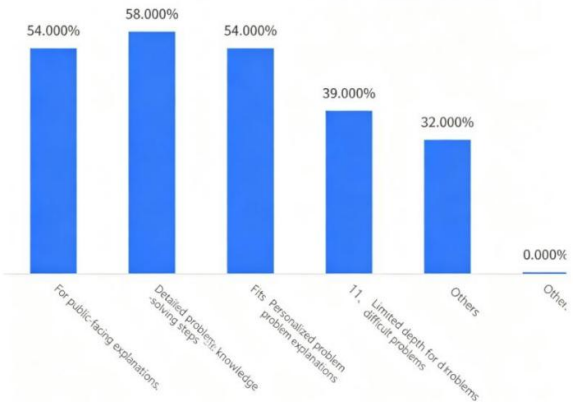
3.7. Student Functional Requirements for AI Tutoring Tools

Students 'functional needs were also assessed through multiple-response analysis. The most highly prioritized requirement was complete and detailed solution steps, selected by 58 students (58%). This finding indicates that students value the learning process itself, not merely obtaining the final answer. Detailed step-by-step solutions can serve as learning resources, helping students understand the logic and reasoning behind each step.

Answer accuracy was the second most valued requirement, chosen by 54 students (54%). Despite the importance of pedagogical features, students still prioritize getting correct answers, underscoring the need for improved accuracy in AI tools. Textbook-aligned explanations were also prioritized by 54 students (54%), suggesting that students want AI tools to match the language and approaches used in their formal coursework.

Personalized error review was selected by 39 students (39%), while in-depth analysis of challenging problems was selected by 32 students (32%). The chi-square goodness-of-fit test confirmed that these requirements were significantly differentiated ($p < 0.001$).

The results indicate that students 'core demand is not simply obtaining answers, but rather receiving complete solution steps, high answer accuracy, and explanations that align with textbook content. This pattern suggests that students view AI tools primarily as learning aids rather than as answer providers, and they expect these tools to support their conceptual understanding in ways consistent with their formal mathematics instruction.



3.8. Open-Ended Feedback

The questionnaire included an open-ended question to collect qualitative feedback on students 'experiences with AI tools. Six valid responses were received, which, although limited in number, provided valuable qualitative insights.

Four main themes emerged from the analysis of these responses. First, reliability concerns: students noted that AI answers often contain errors and should not be blindly trusted, requiring repeated adjustments to queries to obtain correct results, with certain problem types consistently yielding incorrect answers. Second, problem-type limitations: graphical and visual math problems had

significantly higher error rates than computational questions, suggesting that AI models may struggle with geometric reasoning and visual interpretation. Third, usability requirements: students reported that effective use of AI tools requires skill in prompt engineering and iterative refinement, with significant practice needed to achieve good results. Fourth, positive evaluations: a small number of students reported generally positive experiences with AI tools, particularly for straightforward computational tasks and routine problem-solving.

These qualitative findings reinforce the quantitative results from the survey, confirming the core challenges of AI tool reliability, pedagogical quality, and effective usage. The alignment between qualitative and quantitative evidence strengthens the validity of the overall conclusions.

4. DISCUSSION

The findings of this study reveal several important insights into the current state of AI tutoring tool usage in college mathematics education. First, the remarkably high adoption rate (97%) indicates that AI tools have become deeply integrated into students 'mathematics learning routines. This widespread use underscores the need for systematic guidance to help students choose appropriate tools and use them effectively.

The high usage rate reflects broader trends in educational technology adoption, where students are increasingly turning to digital resources to support their learning. However, the rapid adoption of AI tools has not been accompanied by commensurate development of guidance on their effective use. Many students may lack the critical evaluation skills necessary to assess AI outputs, potentially leading to the inadvertent reinforcement of incorrect concepts or flawed reasoning.

Second, the significant concerns students reported regarding AI accuracy and reliability highlight critical limitations in current AI tutoring tools. The finding that over half of students perceived insufficient AI accuracy, coupled with the top-cited concerns about identifying errors, suggests that current AI tools may inadvertently reinforce incorrect mathematical understanding if used without critical evaluation. This aligns with prior findings in educational technology research, which caution against the uncritical use of generative AI in educational contexts.

The most serious risk identified—students 'inability to distinguish correct from incorrect solutions—presents a fundamental pedagogical challenge. Without proper verification skills or external validation, the same tool designed to support learning could potentially undermine it. This concern is particularly acute in mathematics, where errors in early problem-solving steps can compound and lead to incorrect conclusions.

Third, students' prioritization of detailed solution steps, accuracy, and textbook alignment over simply obtaining final answers suggests that students seek pedagogical value from AI tools, not merely transactional answer retrieval. This presents an opportunity for AI developers to differentiate their products through educational design rather than purely technical performance. The fact that students value step-by-step explanations and alignment with curricular materials indicates that the ideal AI tutor should function more like a patient teaching assistant than a solution calculator.

The emphasis on pedagogical features suggests that students are not content with simply obtaining correct answers but want to understand the reasoning behind them. This finding aligns with constructivist theories of learning, which emphasize the importance of understanding the process of problem-solving, not just the final product.

Fourth, the divided subjective evaluations between general-purpose and subject-specific AI tools suggest that no single solution currently satisfies all student needs. This finding supports the argument for differentiated tool selection based on task type, learning objectives, and student prior knowledge. Simple problem-solving tasks might benefit from different tools than complex conceptual exploration tasks, and different learner profiles may derive varying benefits from different tool designs.

The preference for general-purpose models among a majority of students suggests that versatility remains a key factor in tool adoption. However, the substantial minority who prefer specialized tools indicates that there is still a significant role for subject-specific solutions. The optimal approach may involve a combination of both, with students using general-purpose tools for broad exploration and specialized tools for focused problem-solving.

Fifth, the challenges identified in this study related to error detection and verification skills align with broader trends in AI-assisted education research, further underscoring the need for targeted pedagogical interventions. Students may benefit from explicit training in how to evaluate AI outputs, when to trust AI suggestions, and how to use AI as a thinking partner rather than an answer provider.

The finding that students 'top functional requirements—detailed solution steps, answer accuracy, and textbook alignment—correspond precisely to the areas where students report the most frequent problems suggests that addressing these specific issues could significantly enhance the educational value of AI tutoring tools. Tool developers should focus on improving not only the accuracy of generated solutions but also their pedagogical quality and alignment with educational standards.

Finally, the qualitative feedback confirms that students' experiences with AI tools are not uniformly positive or negative but depend on usage context, problem characteristics, and individual student factors. This complexity suggests that blanket recommendations for tool use may be less effective than differentiated guidance based on specific learning objectives and student needs.

5. CONCLUSION AND FUTURE WORK

This study provides empirical evidence regarding the current usage patterns, challenges, and needs of college students using AI tutoring tools for mathematics learning. The findings demonstrate that AI tools have achieved near-universal adoption among mathematics students, yet significant concerns remain regarding accuracy, comprehensibility, and pedagogical effectiveness. Students 'primary demands for detailed solution steps, answer accuracy, and textbook alignment highlight the importance of pedagogical design in AI tutoring tools.

The practical implications of these findings extend to multiple stakeholder groups. For students, the results provide a framework for evaluating and selecting AI tools based on their specific learning needs. For educators, the findings suggest the need for guidance on the appropriate use of AI tools in mathematics learning, as well as the integration of AI literacy into the curriculum. For tool developers, the study identifies key areas for improvement—particularly the accuracy of complex problem-solving, the quality of step-by-step explanations, and the alignment with standard educational materials.

However, this study has several limitations that must be considered when interpreting the findings. First, the sample was drawn from a single institution, which may limit the generalizability of the findings. The specific characteristics of the student population at this institution, including the distribution of majors and academic programs, may not be representative of all college mathematics students. Second, the open-ended question section had relatively few qualitative responses, limiting the depth of qualitative analysis. Third, this survey relied on students 'subjective perceptions rather than objective measures of AI accuracy. While subjective perceptions are important for understanding user experience, they do not necessarily align with objective performance measures.

Future work should expand the survey scope to include multiple institutions and a broader range of student demographics. This would allow for more robust conclusions about general patterns and the identification of institution-specific factors that may influence AI tool adoption and usage. Additionally, the development and validation of objective measures of AI accuracy and pedagogical quality would complement subjective perceptions with performance-based data.

In our ongoing research, we have developed a three-dimensional evaluation framework—covering answer accuracy, solution completeness, and logical interpretability—and are conducting standardized comparative tests using 60 mathematics problems covering calculus, linear algebra, and probability theory. The results of these controlled experiments will further validate the subjective findings reported here and provide a more comprehensive understanding of AI tools 'capabilities and limitations.

Ultimately, these investigations aim to provide concrete tool selection guidelines for students and to inform the development of more effective AI-assisted mathematics learning experiences. As AI tools continue to evolve, ongoing empirical research will be essential to ensure that these resources support rather than undermine student learning and to maximize their potential to enhance mathematical education.

ACKNOWLEDGMENT

This work was supported by the Beijing Wuzi University Student Innovation and Entrepreneurship Training Program.

REFERENCES

[1] Wang, C., & Liu, L. (2010). "Goodness-of-Fit Test Statistics Setting Method." Statistics and Decision, (5), 154-156.

[2] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11) [Online Application Software]. Retrieved from https://www.spsspro.com.

[3] Ahn, J. (2024). "Generative AI in Education: Pedagogical Challenges and Opportunities."

Educational Technology Research, 72(3), 245-263.

[4] Chen, L., & Zhang, Y. (2023). "Large Language Models for Mathematics Problem Solving: A Comprehensive Review." Journal of Educational Technology, 58(4), 112-128.

[5] Liu, H., & Wang, F. (2024). "College Students 'Perceptions of AI-Assisted Learning: Benefits and Risks." Journal of Computing in Higher Education, 36(1), 89-106.