



43815  
Iranian Educational Technology Association

## Evaluating Students' Learning of the Limit Concept Using Artificial Intelligence

Nourooz Hashemi <sup>1\*</sup>  | Ruhallah Mirzaei Aliarabi <sup>2</sup> 

1. *Corresponding Author*, Professor Assistant, Department of mathematics education, University of Farhangian, Tehran, Iran. E-mail: [h.nourooz@cfu.ac.ir](mailto:h.nourooz@cfu.ac.ir)

2. Lecturer, Department of mathematics education, University of Farhangian, Tehean, Iran. E-mail: [ruhallah37@gmail.com](mailto:ruhallah37@gmail.com)

### Print ISSN:

3060-7167

### Online ISSN:

3060-656X

### Article Type:

Research Article

### Article history:

Received November 20, 2025

Received in revised form December 22, 2026

Accepted February 12, 2026

Published Online March 27, 2026

### Keywords:

Evaluation,  
Limit Concept,  
Artificial  
Intelligence,  
CIIP Model,  
Grok

### ABSTRACT

The limit concept, foundational to calculus, is frequently associated with persistent student misconceptions that can be effectively diagnosed and remediated through rigorous assessment and targeted feedback. This study investigates the advantages and limitations of employing artificial intelligence (AI)—specifically Grok—in evaluating first-year undergraduate mathematics students' mastery of the limit concept. Conducted as an applied mixed-methods investigation, the research involved 37 students from the Mathematics Education Department at Farhangian University of Isfahan. Data were collected via a researcher-designed questionnaire comprising seven items on limits; responses were digitized and submitted to the AI for analysis. Evaluation adhered to the CIPP framework, with emphasis on the process dimension, encompassing answer correction, scoring, ranking, error classification (conceptual, procedural, computational), and individualized feedback for students and the instructor. Quantitative results revealed no statistically significant difference between AI-generated and researcher-assigned scores. Qualitative content analysis identified key strengths—rapid processing, scoring precision, analytical feedback, and resource recommendations—alongside notable weaknesses, including deficiencies in graphical interpretation, overemphasis on final answers at the expense of procedural reasoning, and dependency on grammatically precise input for accurate recognition of mathematical notation. Findings underscore AI's viability as a supplementary assessment tool in mathematics education, contingent upon enhancements in visual processing and process-oriented evaluation.

**Cite this Article:** Hashemi, N., & Mirzaei Aliarabi. R. (2026). Evaluating Students' Learning of the Limit Concept Using Artificial Intelligence. *Trends and Achievements in Learning Technology*, 3(9), 33-68. <https://doi.org/10.22034/jlt.2026.2078437.1064>



© 2026 Author(s)

**Publisher:** Iranian Educational Technology Association

**DOI:** <https://doi.org/10.22034/jlt.2026.2078437.1064>

## Introduction

Bridoux and Grenier-Boley (2025) believe that evaluation plays an important and vital role in learning mathematical concepts, and learning the concept of limit is no exception to this rule. According to Li et al. (2024), in the 21st century, the landscape of the process and method of teaching mathematical concepts has undergone a profound transformation, which is mainly driven by the pervasive influence of digital technology. Evaluation, as one of the most important pillars of the educational process in general and in learning mathematical concepts in particular, has been different in each period according to the facilities, equipment and knowledge of that period and has benefited from current science (Mohamudally-Boolakay and Padachi, 2024). The basic needs of transforming an evaluation into an efficient, high-quality and effective judgment depend on the use of advanced teaching methods and educational tools (Frieder et al., 2023). Today, artificial intelligence, as the most modern educational tool, has transformed learning. This transformation has not only provided mathematics teachers with dynamic, graphical, and interactive tools for classroom teaching, but also the integration of technology into formative assessment increases its effectiveness by facilitating real-time feedback and promoting personalized learning (Awang, Yusop, and Danaee, 2025). Therefore, this paper has pursued three main objectives. First, it has examined the possibility of rationally evaluating students' learning of the important and fundamental concept of limit in the context of artificial intelligence. Then, it has explored how artificial intelligence can provide feedback to students and introduce resources to them to overcome learning difficulties in this concept. It has also considered the final feedback to the instructor, which, based on the results of the students' evaluation, should the instructor make any necessary corrections and actions in his/her teaching process.

## Literature Review

Assessment in mathematics education is a key tool for measuring student mastery, identifying weaknesses, and improving learning (Kistler, 2023). This process enhances motivation, self-regulation, and conceptual understanding, and improves academic performance with continuous feedback (Morris et al., 2021; Khomijani-Farahani et al., 1403). Tyler (1942) defines it as the achievement of educational goals, Staffelbeam (2011) as the provision of information for decision-making, and Scriven (Kus, 2025) as the ongoing judgment of the quality of learning. Assessment is essential in all teaching phases and disciplines,

especially mathematics (Zinn, 2025). Calculus is a fundamental undergraduate course in basic sciences and engineering (González-Martín, 2021). The concept of limit is a prerequisite for derivatives, integrals, continuity, and advanced courses (such as real analysis) and is used in physics and chemistry (Kidron & Zehavi, 2002). Despite its importance, its teaching is often superficial, rushed, and lacking focus on the  $\epsilon$ - $\delta$  relationship (Bokhari & Yushau, 2006). Common misunderstandings resulting from a lack of deep understanding hinder learning in subsequent lessons (Liang, 2016). Careful evaluation of responses reveals the root of these problems and provides corrective solutions (Baptiste, 2024; Ünver & Güzel, 2019).

## Methodology

The present study is an applied study with a mixed approach (quantitative-qualitative) that focuses on investigating the effectiveness of Grok artificial intelligence in evaluating learning of the concept of limit in first-year students of mathematics education. The statistical population included students of this field at Farhangian University of Isfahan and the research sample was 37 people (21 women, 16 men). The data collection tool was a researcher-made questionnaire with 7 conceptual-practical questions about limit (with validity and reliability confirmed by 4 experts). The answer sheets were scanned, converted to PDF, and submitted to Grok. The artificial intelligence was tasked with correcting, scoring (out of 20), ranking, identifying conceptual/process/computational errors, providing individual and collective feedback, and suggesting corrective resources for students and instructors. The theoretical framework of the study was designed based on the CIPP model (Stufflebeam, 2000, 2004) and focusing on the process stage. This stage includes monitoring program implementation, documenting activities, comparing actual performance with planned, and providing corrective feedback. Since the goal was to evaluate the evaluation process itself by AI, this stage was selected (Adib Hajbagheri, 2002). The activities of the evaluation process are depicted in Figure 1 and its indicators are defined in Table 2 (including development, execution, display, feedback). Among the educational chatbots (ChatGPT, DeepSeek, Gemini, etc.), Grok was selected due to its superiority in accepting custom commands, analyzing processes (Souza & Weigang, 2025), and integration with the X platform. Data analysis was performed at two levels. In quantitative analysis, Grok scores were compared with researcher scores through a univariate T-test, and qualitative analysis was conducted using thematic analysis (Pinede, 2023) and coding, Grok

responses were examined. The extracted themes were confirmed by 3 experts and used for coding. Also, 5 randomly selected response sheets were corrected by experts and compared with Grok output to assess the accuracy of feedback, sources, and recommendations.

## Results

Grok AI provided acceptable output in correcting scanned answer sheets of 37 students on the concept of limit. With precise grammar, it provided descriptive answers, scoring (out of 20), ranking and classifying performance (excellent, good, average, poor) with high accuracy and extraordinary speed (less than a minute). The difference in scores with the researcher was not significant. The quantitative (mean, standard deviation) and qualitative (conceptual, procedural, computational errors) feedback was accurate and individual, and a comprehensive matrix table was created to compare errors. Its advantages include speed and accuracy in collective and individual analysis, identification of challenging questions, providing analytical feedback to the instructor (conceptual challenges, weakness in interpreting graphs and mathematical symbols), suggesting resources (workshops, worksheets, books, websites), providing teaching strategies and motivational and personal feedback to students. Also, its limitations such as excessive focus on the final answer (even with an incomplete process, it gave a full score) contrary to the importance of the process in mathematics education (Cuong et al., 2025), inability to draw and analyze graphs, sensitivity to writing mathematical symbols (brackets, absolute values, parentheses) requiring additional explanation in the instruction, based on the CIPP model (process stage) and its themes (development, implementation, display, feedback) can be mentioned. Overall, Grok is an effective complementary tool in mathematical assessment, but for independence it needs improvement in visual analysis, reasoning process assessment and automatic recognition of mathematical symbols. Combination with human supervision is recommended.

## Conclusion

The findings also showed that Grok focuses less on the process-oriented evaluation of answers than on the final answer, while for mathematics teachers, the problem-solving process is more important. This is not just mentioned as a bug, but as a systematic challenge in current AI models of understanding and producing visual mathematical content for analysis. This can be considered as a

weakness in the current design of Grok. However, the speed and accuracy of ranking, the provision of statistical tables and in-depth analyses make this tool a valuable option for wider evaluations. For future research, it is suggested that Grok's capabilities in drawing graphs and visual analysis be strengthened and the ability to more accurately recognize mathematical symbols in input commands be improved. Also, using this tool in larger statistical communities and more diverse mathematical topics can help to better understand its capabilities and limitations. Finally, this study showed that artificial intelligence can be a complementary tool alongside teachers to improve the evaluation process and provide feedback and help improve the quality of mathematics education.

## ارزشیابی یادگیری مفهوم حد با استفاده از هوش مصنوعی

نوروز هاشمی\* | روح‌الله میرزایی علی‌عربی<sup>۲</sup>

۱. نویسنده مسئول، استادیار گروه آموزش ریاضی، دانشگاه فرهنگیان، تهران، ایران. رایانامه: [h.nourooz@cfu.ac.ir](mailto:h.nourooz@cfu.ac.ir)

۲. مدرس گروه آموزش ریاضی، دانشگاه فرهنگیان، تهران، ایران. رایانامه: [ruhallah37@gmail.com](mailto:ruhallah37@gmail.com)

### شاپا چاپی:

۳۰۶۰-۷۱۶۷

### شاپا الکترونیکی:

۳۰۶۰-۶۵۶X

### نوع مقاله:

مقاله پژوهشی

### تاریخچه مقاله

تاریخ دریافت: ۱۴۰۴/۰۸/۲۹

تاریخ بازنگری: ۱۴۰۴/۱۰/۰۱

تاریخ پذیرش: ۱۴۰۴/۱۱/۲۳

تاریخ انتشار: ۱۴۰۵/۰۱/۰۶

### چکیده

مفهوم حد به‌عنوان پایه‌ای برای حساب دیفرانسیل و انتگرال، اغلب با بدفهمی‌های دانشجویان همراه است که ارزشیابی مؤثر و ارائه بازخورد درست می‌تواند آن‌ها را شناسایی و اصلاح کند. این پژوهش به بررسی مزایا و معایب ارزشیابی یادگیری مفهوم حد برای دانشجویان کارشناسی آموزش ریاضی با استفاده از هوش مصنوعی (Grok) می‌پردازد. مطالعه حاضر، کاربردی و ترکیبی (کمی - کیفی)، بر روی ۳۷ دانشجوی سال اول رشته آموزش ریاضی دانشگاه فرهنگیان اصفهان انجام شد. ابزار گردآوری داده، پرسشنامه‌ای محقق‌ساخته با ۷ سؤال درباره مفهوم حد بود که پاسخنامه‌های دانشجویان اسکن و به هوش مصنوعی ارائه گردید. تحلیل بر اساس مدل CIPP (با تمرکز بر مرحله فرایند) انجام گرفت و شامل تصحیح پاسخنامه‌ها، محاسبه نمرات، رتبه‌بندی، شناسایی خطاهای مفهومی، رویه‌ای و محاسباتی، و ارائه بازخوردهای شخصی به دانشجویان و مدرس بود. یافته‌های کمی نشان داد میانگین نمرات داده شده به دانشجویان توسط هوش مصنوعی با نمرات محققان تفاوت معناداری ندارد. یافته‌های کیفی، از طریق تحلیل مضمون، مزایایی مانند سرعت، دقت در رتبه‌بندی، بازخوردهای تحلیلی و پیشنهاد منابع را برجسته کرد، اما معایبی نظیر ضعف در ترسیم نمودارها، تمرکز بیش‌ازحد بر پاسخ نهایی (به‌جای فرایند) و نیاز به دست‌نویسی دقیق برای تشخیص علائم ریاضی را آشکار ساخت. نتایج حاکی از پتانسیل هوش مصنوعی به‌عنوان ابزاری مکمل برای بهبود ارزشیابی آموزش ریاضی است، هرچند نیازمند ارتقاء در تحلیل بصری و فرآیندمحور است.

### کلیدواژه‌ها:

ارزشیابی،  
مفهوم حد،  
هوش مصنوعی،  
الگوی CIIP،  
Grok

استاد به این مقاله: هاشمی، نوروز، و میرزایی علی‌عربی، روح‌الله. (۱۴۰۵). ارزشیابی یادگیری مفهوم حد با استفاده از هوش مصنوعی. نشریه روندها

و دستاوردها در فناوری یادگیری، ۳(۹)، ۳۳-۶۸. <https://doi.org/10.22034/jlt.2026.2078437.1064>

## مقدمه

اندیشمندان و نظریه‌پردازان یادگیری از قبیل Vygotsky (1935)، Piaget (1970) و Egan (2012) همیشه یکی از باارزش‌ترین علوم مورد نیاز بشر در طول دوره‌های مختلف را علم ریاضی دانسته‌اند و برای یادگیری آن توجه ویژه‌ای قائل بوده‌اند. ریاضیات و اطمینان از یادگیری مفاهیم آن، بستر مناسبی را فراهم می‌کند تا یادگیری علوم دیگر آسان‌تر اتفاق بیفتد (Tashtoush et al., 2025). به علاوه، ریاضی یک علم بر اساس استدلال، یک مهارت شناختی اصلی، برای حل مسائل پیچیده و تصمیم‌گیری آگاهانه است که برای زندگی بشر در تمام اعصار ضروری است (Xia et al., 2025). ریاضی اغلب به‌عنوان یکی از دشوارترین دروس آموزش داده‌شده و توسط یادگیرندگان درک می‌شود (Putri et al., 2024). Lele و همکاران (2024) معتقدند درک و آموزش مفاهیم ریاضی به دلیل ماهیت انتزاعی و تجمعی، نیازمند رویکردی متفاوت در مقایسه با موضوعات دیگر است. از مهم‌ترین چالش‌های آموزشگران و سیاست‌گذاران امر آموزش، آموزش مفاهیم ریاضی در دوره‌های مختلف بوده و باتوجه‌به شرایط آن دوره، از راهبردهای مختلفی برای آموزش آن استفاده شده است (Hashemi & Kashefi, 2025). فرایند آموزش یک مفهوم ریاضی، مبتنی بر آماده‌سازی، اجرا و ارزشیابی اثربخشی روش تدریس است که از مهارت‌های ضروری برای یک آموزشگر است (Kopp et al., 2025). از این رو، در آموزش مفاهیم گوناگون ریاضی، برای پی‌بردن به میزان و کیفیت درک یادگیرندگان، ارزشیابی از آموخته‌ها تکمیل‌کننده فرایند آموزش و یادگیری آن مفهوم است (Pepin et al., 2025). ارزشیابی از یادگیری مفاهیم ریاضی، نقش مهمی در شناسایی زود هنگام خطاهای مفهومی و ارائه بازخورد لازم برای اصلاح بدفهمی‌ها قبل از حرکت یادگیرندگان به مطالب بعدی دارد.

ارزشیابی در آموزش ریاضیات بسیار مؤثر است و به معلمان کمک می‌کند تا میزان تسلط یادگیرندگان بر مطالب تدریس شده را درک کنند، زمینه‌هایی را که نیاز به بهبود دارند شناسایی کنند و دستورالعمل‌ها را برای برآورده کردن نیازهای یادگیرندگان تطبیق دهند (Kistler, 2023). در واقع، ارزشیابی ابزاری قدرتمند برای بهبود یادگیری در ریاضیات است که طیف وسیعی از مزایایی مانند افزایش انگیزه، خودتنظیمی بهتر و افزایش درک مفهومی را ارائه می‌دهد. همان‌طور که توسط مطالعات مختلف، از جمله مطالعات Morris و همکاران (2021)، Lele و همکاران

(2024) و خمیجانی فراهانی و همکاران (۱۴۰۳) بیان شده است ارزشیابی آموزشی به طور قابل توجهی عملکرد تحصیلی را با ارائه بازخورد سازنده و مستمر افزایش می‌دهد. تایلر<sup>۱</sup> ارزشیابی آموزشی را چنین تعریف می‌کند که ارزشیابی آموزشی تعیین می‌کند که هدف‌های آموزشی برنامه درسی و آموزشی تا چه میزانی تحقق یافته‌اند (Tyler, 1942). ارزشیابی آموزشی از دیدگاه استافلیم<sup>۲</sup> فرایندی است برای تعیین کردن، به‌دست آوردن و فراهم کردن اطلاعات مفید آموزشی برای قضاوت و تصمیم‌گیری (Stufflebeam, 2011)، همچنین اسکریون<sup>۳</sup> ارزشیابی را فرایندی تعریف می‌کند که در تمام طول دوره آموزشی یا جلسه تدریس یک مفهوم انجام می‌شود تا درباره میزان و کیفیت یادگیری آن دوره یا مفهوم امکان دآوری صحیح وجود داشته باشد (Kus, 2025). ارزشیابی قبل، حین و در پایان تدریس مرحله‌ای از فعالیت‌های آموزشی معلم محسوب می‌شود که با اجرای آن دیگر فعالیت‌های یادگیری توسط او کامل می‌شود (خزائی و همکاران، ۱۴۰۲). ارزشیابی به‌عنوان یکی از مهم‌ترین مراحل آموزش و یادگیری، در تمام موضوعات آموزشی و در رشته‌های گوناگون درسی از قبیل ریاضیات، علوم پایه، علوم انسانی، هنر، فنی و مهندسی و دیگر رشته‌ها همیشه مورد توجه بوده و یکی از ارکان اصلی برنامه درسی است (Zinn, 2025). علم ریاضی و یادگیری مفاهیم آن از جمله هندسه، جبر، حساب و غیره به‌عنوان پایه علوم دیگر نیز از این قاعده مستثنا نبوده است (Zhang et al., 2025).

حساب دیفرانسیل و انتگرال به‌عنوان یکی از مهم‌ترین دروس ریاضی از گذشته تاکنون، برای دانشجویان مقطع کارشناسی در بسیاری از رشته‌های تحصیلی از اهمیت زیادی برخوردار بوده است (جهانی‌پور و مقصودی، ۱۳۹۸). گونزالو مارتین<sup>۴</sup> بیان داشته است نه تنها در رشته کارشناسی ریاضی، بلکه در رشته‌های گوناگون علوم پایه و فنی و مهندسی، درس حساب دیفرانسیل و انتگرال یکی از بنیادی‌ترین و مهم‌ترین دروس بوده است (González-Martín, 2021). هاشمی و کاشفی، (۱۴۰۳). مفهوم حد یکی از مفاهیم اساسی در حساب دیفرانسیل و انتگرال است که دانشجویان مقطع کارشناسی با استفاده از آن در یادگیری مفاهیمی چون مشتق و انتگرال در سال

---

1. Tyler  
2. Stufflebeam  
3. Scriven  
4. González-Martín

اول با آن مواجه می‌شوند و پیش‌نیاز یادگیری سایر مفاهیم ریاضی و غیر ریاضی مانند فیزیک، شیمی و غیره در رشته‌های مختلف تحصیلی است (Kidron & Zehavi, 2002)؛ ریحانی و شریفی، (۱۳۹۷). حد، مفهومی است که با مفاهیم دیگر از جمله پیوستگی، مشتق و انتگرال ارتباط دارد. شناخت جزئیات و ویژگی‌های این مفهوم کمک زیادی به یاددهی - یادگیری موضوعات دیگر می‌کند (هادیان دهکردی، اسلام‌پور و ریحانی، ۱۳۹۳).

به جنبه‌های اساسی حد از قبیل رابطه  $\epsilon$  و  $\delta$  توجه چندانی نشده است و مدرسان و دانشجویان با این موضوع عجولانه برخورد می‌کنند و نمی‌توانند روی این مفهوم به‌اندازه کافی تأمل کنند و سپس به موضوعات بعدی بروند (Bokhari & Yushau, 2006). برای رشته‌های غیرریاضی، عدم درک مفهوم حد ممکن است فقط مشکل جدی برای مطالعه آنها در دروس ریاضی بعدی ایجاد کند، ولی برای رشته‌های ریاضی موضوع بسیار مهم و اساسی است. درک دانشجویان از حساب دیفرانسیل و انتگرال تا حد زیادی بر توانایی آنها برای مطالعه دروس تجزیه و تحلیل پیشرفته‌تر (مانند متغیر واقعی تابع، تجزیه و تحلیل واقعی، تجزیه و تحلیل تابعی و نظریه اندازه گیری) تأثیر می‌گذارد، زیرا این دروس همه به حساب دیفرانسیل و انتگرال به‌عنوان پیش‌نیاز نیاز دارند و درک مفهوم حد پیش‌نیاز درک بهتر حساب و دیفرانسیل است (Domingos, 2010). منابع زیادی وجود دارد که به مشکلات آموزش و یادگیری این مفهوم کمک می‌کند و با این وجود، اگر از استراتژی‌ها و ابزارهای مناسب استفاده شود، آموزش موفقیت‌آمیز مفهوم حد کار دست‌نیافتنی نیست (Ünver & Güzel, 2019). اگرچه جامعه ریاضیات مدت‌ها مفهوم حد را به‌عنوان پایه حساب و دیفرانسیل امروزی پذیرفته است، اما خود مفهوم حد در آموزش حساب و دیفرانسیل در مقطع کارشناسی به حاشیه رانده شده است و یادگیری این مفهومی‌هایی مواجه است که ضرورت دارد ریشه‌های آن پیدا شود (Liang, 2016). عدم درک درست از مفهوم حد و ناتوانی در حل مسائل مربوط به آن باعث به‌وجود آمدن بدفهمی‌هایی شده است که با ارزشیابی دانشجویان و بررسی پاسخ‌های آنها به سؤالات داده‌شده می‌توان به پیدا کردن ریشه بدفهمی‌ها و رفع آنها و تلاش برای درک صحیح این مفهوم کمک کرد (Baptiste, 2024).

Bridoux and Grenier-Boley (2025) معتقدند در یادگیری مفاهیم ریاضی، ارزشیابی نقش مهم و حیاتی دارد و یادگیری مفهوم حد نیز از این قاعده مستثنا نیست. باتوجه به Li و همکاران (2024) در قرن بیست و یکم، چشم‌انداز فرایند و روش آموزش مفاهیم ریاضیات دستخوش

دگرگونی عمیقی شده است که عمدتاً تحت تأثیر نفوذ فراگیر فناوری دیجیتال هدایت می‌شود. ارزشیابی به‌عنوان یکی از مهم‌ترین ارکان فرایند آموزش، به‌طورکلی و در یادگیری مفاهیم ریاضی به‌طور خاص، در هر دوره‌ای باتوجه به امکانات، تجهیزات و دانش آن دوره متفاوت بوده است و از علم روز بهره برده است (Mohamudally-BooLaky & Padachi, 2024). نیازهای اساسی تبدیل یک ارزشیابی به قضاوتی کارآمد، باکیفیت و اثربخش وابسته به استفاده از روش آموزش و ابزارهای آموزشی پیشرفته است (Frieder et al., 2023). امروزه هوش مصنوعی به‌عنوان مدرن‌ترین ابزار آموزشی، یادگیری را دگرگون کرده است. این دگرگونی نه‌تنها ابزارهای پویا، گرافیکی و تعاملی را برای تدریس در کلاس به معلمان ریاضی ارائه کرده است، علاوه بر این، ادغام فناوری در ارزیابی تکوینی، اثربخشی آن را با تسهیل بازخورد بلادرنگ و ترویج یادگیری شخصی افزایش می‌دهد (Awang, Yusop & Danaee, 2025). ازاین‌رو، مقاله حاضر سه هدف اساسی را دنبال کرده است. ابتدا به بررسی امکان ارزشیابی منطقی یادگیری دانشجویان از مفهوم مهم و زیربنایی حد در بستر هوش مصنوعی پرداخته است. سپس، نحوه‌ی بازخورد هوش مصنوعی به دانشجویان و معرفی منابع به آن‌ها جهت رفع مشکلات یادگیری این مفهوم را مورد کنکاش قرار داده است. همچنین، بازخورد نهایی به مدرس درس را موردتوجه قرار داده است که با توجه به نتایج ارزشیابی دانشجویان، مدرس چه اصلاحات و اقداماتی در فرایند تدریس خود باید انجام دهد.

## روش‌شناسی پژوهش

پژوهش حاضر یک مطالعه تحقق‌کاربردی است که گردآوری و تحلیل داده‌های آن به‌صورت کمی و کیفی انجام شده است. جامعه آن دانشجویان سال اول دوره کارشناسی رشته آموزش ریاضی و نمونه آن ۳۷ نفر از دانشجویان سال اول رشته‌ی آموزش ریاضی دانشگاه فرهنگیان اصفهان بودند که ۲۱ نفر خانم و ۱۶ نفر آقا بودند. فرایند انجام این مطالعه به این صورت بوده است که ابتدا یک آزمون از دانشجویان درباره مفهوم حد و کاربرد آن به عمل آمد و سپس ارزشیابی از فرایند تصحیح و اعلام بازخورد آن در بستر هوش مصنوعی بر اساس مدل سیپ<sup>۱</sup> با تمرکز بر مرحله فرایند موردبررسی قرار گرفت. Stufflebeam (2000) الگوی ارزیابی سیپ

را چارچوبی جامع برای هدایت و ارزیابی‌های برنامه‌ها، پروژه‌ها، کارکنان، محصولات، مؤسسات و سیستم‌های ارزیابی معرفی کرده است. این الگو یا مدل در اواخر دهه ۱۹۶۰ برای کمک به بهبود و دستیابی به مسئولیت‌پذیری برای برنامه‌های مدارس ایالات متحده، به‌ویژه برنامه‌هایی که برای بهبود آموزش و یادگیری در مناطق شهری و مدارس درون شهر کلید خورده‌اند، توسعه یافت. در طول سال‌ها، این مدل بیشتر توسعه یافته و در برنامه‌های آموزشی در داخل و خارج از ایالات متحده به کار گرفته شده است (Stufflebeam, 2004).

متناظر با حروف در مخفف سیپ، مفاهیم اصلی این الگو عبارت‌اند از ارزیابی زمینه<sup>۱</sup>، ورودی<sup>۲</sup>، فرایند<sup>۳</sup> و محصول<sup>۴</sup>. در این روش ارزشیابی، با به‌کارگیری چهار نوع ارزیابی، ارزیاب چندین کارکرد مهم را انجام می‌دهد (Alquraan et al., 2025). ارزیابی زمینه نیازها، مشکلات و فرصت‌ها را در یک محیط تعریف‌شده ارزیابی می‌کند. ارزیابی‌های ورودی، استراتژی‌های مختلف و برنامه‌های کاری رویکردهای انتخاب‌شده برای اجرا را ارزیابی می‌کند. ارزیابی‌های فرایند، فعالیت‌ها را در یک شرایط تعریف‌شده نظارت، مستندسازی و ارزیابی می‌کند. آنها به کاربران ارزیابی کمک می‌کنند تا تلاش‌های روبه‌جلو را انجام دهند و سوابق مسئولیت‌پذیری اجرای برنامه‌های اقدام خود را حفظ کنند. ارزیابی محصول نتایج کوتاه‌مدت، بلندمدت، موردنظر و ناخواسته را شناسایی و ارزیابی می‌کند (Stufflebeam & Coryn, 2014). از آنجائی که هدف اصلی این مطالعه بررسی چگونگی ارزشیابی توسط هوش مصنوعی بود و ارزشیابی نیز خود یک فرایند است، لذا از ارزشیابی فرایند الگوی سیپ استفاده گردید. به بیان دیگر، تمرکز اصلی بر شاخصه‌های مرحله ارزشیابی فرایند الگوی سیپ بود تا فرایند ارزشیابی توسط بسترهای هوش مصنوعی مورد کنکاش قرار گیرد. علاوه بر این، ادیپ حاج باقری (۱۳۸۱) بیان می‌دارد که می‌توان فرایند ارزشیابی را نیز مورد ارزشیابی قرار داد، از این‌رو مطالعه حاضر به ارزیابی کیفیت و چگونگی انجام فرایند ارزشیابی توسط هوش مصنوعی پرداخته است.

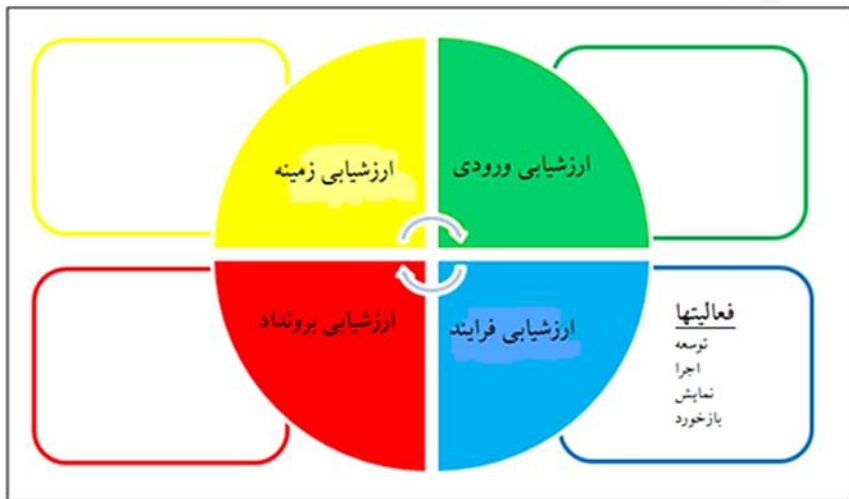
در این مطالعه تمرکز اصلی بر استفاده از مرحله فرایند الگوی ارزشیابی سیپ بوده است، زیرا فرایند استفاده از ارزشیابی یادگیری دانشجوینان با استفاده از هوش مصنوعی هدف اصلی

1. Context
2. Input
3. Process
4. Product

پژوهش است. باتوجه به Alquraan و همکاران (2025)، در مرحله فرایند این الگو، راهنمایی برای اجرای برنامه (بر اساس نظارت و قضاوت فعالیت‌ها و بازخوردهای ارزیابی‌های دوره‌ای)، شرح کامل فرایند واقعی و ثبت رویدادها، مقایسه فرایندها و رویدادهای واقعی از موارد مهم قابل توجه است. به‌علاوه، در اصل ارزشیابی فرایند، بررسی اجرای یک برنامه است و هدف از انجام آن آشکارسازی و یا پیش‌بینی نقایص و معایب برنامه در هنگام اجرا و فراهم‌ساختن بازخورد برای محققان در رابطه با چگونگی هم‌خوانی پیشرفت کار با برنامه تنظیمی است (Stufflebeam, 2004)؛ بنابراین، می‌توان گفت در این مطالعه فرایند ارزشیابی از یادگیری دانشجویان با استفاده از هوش مصنوعی مورد قضاوت قرار گرفته است و باتوجه به Nouraey و همکاران (2020) فعالیت‌های تعریف شده برای آن در شکل ۱ ارائه شده است.

### شکل ۱.

فعالیت‌های تعریف‌شده برای ارزشیابی فرایند بر اساس الگوی سیپ



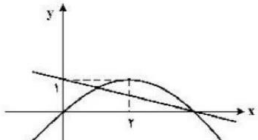
امروزه چت‌بات‌های Grok، Gemini، BlenderBot، YouChat، Google Bard، Copilot، ChatGPT و DeepSeek متداول‌ترین در بستر هوش مصنوعی برای امر آموزش هستند (قشوقی برق و همکاران، ۱۴۰۳؛ احمدآبادی مسرورنیا، ۱۴۰۳). بسترهای هوش مصنوعی که می‌توانستند در مطالعه حاضر مورد استفاده قرار بگیرند چت‌بات‌های Grok، ChatGPT و DeepSeek بوده

است، هرچند ChatGPT دارای قابلیت‌های متعددی است (Manik, 2025)، اما چون در این پژوهش یک کار سفارشی به هوش مصنوعی داده شده است و Souza and Weigang (2025) بر این باورند که Grok برای سفارش گرفتن سطح بسیار بالاتری نسبت به بقیه بسترهای هوش مصنوعی دارد و در تحلیل فرایندها بهتر عمل می‌کند، هوش مصنوعی مورداستفاده در این مطالعه Grok بوده است. Grok به دلیل ادغام آن با پلتفرم X (که قبلاً توییتر بود) بسیار برجسته و کارآمد است و امکان پاسخ‌های سریع، همسو و واقعی‌تر با بحث‌های بداهه را فراهم می‌کند (Wangsa et al., 2024). البته بقیه بسترهای هوش مصنوعی نیز بعضاً کارایی لازم را در تحلیل فرایند ارزشیابی ممکن است داشته باشند، ولی بستر Grok هوش مصنوعی اصلی به‌کاررفته در این پژوهش است و تحلیل و نتیجه‌گیری بر اساس آن انجام شده است.

ابزار گردآوری داده پرسش‌نامه محقق ساخته بود که حاوی ۷ سؤال درباره مفهوم حد است. روش استفاده از هوش مصنوعی به این ترتیب بود که پس از طرح پرسشنامه حاوی سؤالات مفهوم حد که روایی و پایایی آن توسط چهار نفر از متخصصین آموزش ریاضی و ریاضی تأیید گردید و در جدول ۱ آمده است، پاسخنامه دانشجویان به سؤالات پرسشنامه داده شده تک‌به‌تک کاملاً اسکن شده و سپس جمع‌شد و به صورت یک فایل pdf به هوش مصنوعی داده شد. اصل پاسخنامه‌ها توسط محقق تصحیح گردید و نمرات آن از ۲۰ مشخص گردید. در ادامه از هوش مصنوعی خواسته شد ابتدا یک پاسخنامه تشریحی برای این پرسشنامه آماده کند و لیست اسامی و نمره تک‌تک دانشجویان را از ۲۰ ارائه دهد و رتبه آن‌ها در کلاس را تعیین کند و همچنین با توجه به پاسخ دانشجویان به سؤالات پرسشنامه، به صورت فردی به نقاط ضعف آنها اشاره کند. به علاوه، به هوش مصنوعی دستور داده شد تا نقاط ضعف کل دانشجویان و همچنین نقدهای وارد بر مدرس درس را ارائه نموده و برای رفع ایرادات مذکور منابع و توصیه‌های مفید را حاضر کند. این منابع و توصیه‌ها از طرف هوش مصنوعی به صورت فردی نیز به دانشجویان باید ارائه می‌گردید.

جدول ۱.

پرسش نامه ارائه شده به دانشجویان از مفهوم حد

شماره سؤال	متن	بارم
۱	مقدار $\lim_{x \rightarrow \frac{\pi}{6}} [2 \sin x - 1]$ را محاسبه کنید.	۳
۲	نمودار تابع سهمی $f$ و خط راست $g$ در شکل زیر داده شده است. مقدار $\lim_{x \rightarrow 4} \frac{f(x)+g(x)}{4-x}$ را پیدا کنید.	۳
		
۳	هرگاه $f(x)$ به گونه‌ای باشد که برای هر $x$ داشته باشیم $ f(x) + 5  \leq (x - 1)^2$ ، آنگاه مطلوب است محاسبه $\lim_{x \rightarrow 1} f(x)$ .	۳
۴	نمودار تابعی را مثال بزنید که در $x=3$ حد نداشته باشد و دارای پیوستگی چپ باشد (رسم نمودار مهم است).	۲
۵	صحت عبارت زیر را بررسی کنید و استدلال خود را ارائه کنید. $\lim_{x \rightarrow -\infty} \left( x + \sqrt{x^2 - x} \right) = \frac{1}{2}$	۳
۶	با مثالی نشان دهید امکان دارد $\lim_{x \rightarrow a} (f(x) + g(x))$ وجود داشته باشد ولی $\lim_{x \rightarrow a} f(x)$ و $\lim_{x \rightarrow a} g(x)$ وجود نداشته باشد.	۳
۷	نشان دهید برای هر $a > 0$ ، $\lim_{x \rightarrow a} \sqrt{x} = \sqrt{a}$ .	۳

تحلیل داده‌های به دست آمده از پرسش نامه به دو صورت کمی و کیفی انجام گرفت. ابتدا نمرات محاسبه شده دانشجویان توسط هوش مصنوعی با نمرات به دست آمده توسط محقق به صورت کمی مورد مقایسه قرار گرفت. برای مقایسه کمی، از آزمون T تک متغیره استفاده گردید که مشخص شود آیا تفاوت معناداری در اختلاف میانگین نمرات وجود دارد یا خیر. سپس، پاسخ‌های هوش مصنوعی پس از کدگذاری، با استفاده از روش تحلیل مضمون<sup>۱</sup> به عنوان یکی

از روش‌های پژوهش کیفی، موردبررسی قرار گرفت. در این روش، تمرکز اصلی بر کشف، تحلیل و تفسیر الگوهای معنایی موجود در داده‌های کیفی بود (Pinede, 2023). مضمون یا تم به‌عنوان عنصر اصلی این تحلیل، بازتاب‌دهنده معنای کلیدی است که از بخش‌های مختلف داده؛ مانند واژه‌ها، جملات یا پاراگراف‌ها استخراج می‌شود و به‌عنوان مهم‌ترین واحدهای تحلیلی در نظر گرفته می‌شوند. این نوع مضامین پیش‌تر در مطالعه Shoufan (2023) و Zhang and Dong (2025) که درباره تحلیل مضمون در بستر هوش مصنوعی بودند نیز به کار رفته بودند. در جدول ۲، مضامین مرتبط با الگوی سیپ در مرحله فرایند این پژوهش ارائه شده‌اند که باتوجه به تأیید سه نفر متخصص، از همین مضامین برای کدگذاری پاسخ‌های هوش مصنوعی به مسائل مطرح‌شده استفاده شد.

## جدول ۲.

شاخصه‌ها و مضامین ارزشیابی فرایند بر اساس الگوی سیپ در بستر هوش مصنوعی

مضمون	شاخصه
ارتباط با دیگر مفاهیم، تعمیم به موارد دیگر، استفاده از بستر دیگر	توسعه
راهنمایی برای اجرا، شرح کامل فرایند، مقایسه فرایند	اجرا
نمایش نتیجه در قالب نمودار یا جدول، خلاصه‌کردن نتایج	نمایش
نقد صحیح پاسخ‌ها، نمره‌دهی منطقی، ارائه منابع مناسب	بازخورد

پس از تحلیل فرایند ارزشیابی توسط هوش مصنوعی با مضامین جدول ۲ که کیفیت و چگونگی انجام این تحلیل‌ها دوباره توسط متخصصان موردبررسی قرار گرفت و اصلاح گردید، ۵ پاسخ‌نامه به‌صورت تصادفی از بین همه پاسخ‌نامه‌ها انتخاب گردید و تصحیح گردید و با بازخورد و تصحیح هوش مصنوعی مقایسه شد تا مشخص شود آیا بازخوردها، توصیه‌ها و منابع معرفی‌شده توسط هوش مصنوعی منطقی بوده است یا نیاز به اصلاحاتی دارد. قابل‌ذکر است که درستی روش تصحیح پاسخ‌نامه‌های انتخابی و مقایسه آن با نتایج هوش مصنوعی مورد تأیید متخصصان (سه نفر) قرار گرفت.

## دستورنویسی (پرامپت‌نویسی) در بستر هوش مصنوعی برای سنجش دانشجویان

باتوجه به جدول ۲ و شاخصه اجرا، در بستر هوشی ابتدا باید یک دستور مدون و دقیق برای معرفی موضوع، تشریح فرایند و درخواست بازخورد برای هوش مصنوعی نوشت تا بر اساس آن پاسخ دهد (Zhang and Dong, 2025). همچنین باید نقش هوش مصنوعی را تعریف کرد و سپس پاسخ‌های آن مورد ارزیابی قرار گیرد. در ادامه دستور داده شده به هوش مصنوعی ارائه می‌شود تا ابعاد استفاده از آن بیشتر مشخص گردد.

نقش شما: شما یک متخصص ارزیابی آموزشی و استاد ریاضیات باتجربه هستید. تخصص شما در تحلیل عملکرد دانشجویان، شناسایی چالش‌های یادگیری در مباحث پیچیده ریاضی (مانند مبحث حد) و ارائه بازخوردهای سازنده و دقیق به هر دو گروه مدرسان و دانشجویان است. هدف شما فراتر از نمره دادن است؛ شما به دنبال بهبود فرایند تدریس و یادگیری هستید. شما توانایی خوانش و تفسیر اسناد دست‌نویس<sup>۱</sup>، تحلیل الگوهای خطا، و تولید گزارش‌های چندرسانه‌ای (متن، جدول و نمودار) را دارید و مانند یک مربی دلسوز و تحلیل‌گر دقیق عمل می‌کنید.

**هدف اصلی:** بر اساس فایل‌های ورودی، بر اساس فایل سؤالات امتحان و فایل پاسخ‌نامه‌های دانشجویان که ارائه می‌شود، وظایف زیر را به صورت دقیق و مرحله‌بندی شده انجام دهید:

- ایجاد یک پاسخ‌نامه تشریحی کامل برای امتحان.
- تصحیح دقیق هر پاسخ‌نامه و تعیین نمره نهایی.
- ایجاد فهرستی از نمرات به تفکیک نام دانشجو.
- ارائه یک تحلیل جامع و گزارش بازخورد برای معلم.
- تولید یک بازخورد شخصی‌سازی شده و آموزشی برای هر دانشجو.

### ورودی‌ها:

- فایل ۱ (الزامی): سؤالات امتحان ریاضی مبحث حد به صورت Pdf.
- فایل ۲ (الزامی): مجموعه‌ای از پاسخ‌نامه‌های دانشجویان (اسکن یا فایل متنی).

- فایل ۳ (اختیاری): فایل راهنمای نمره‌دهی ۱، در صورتی که معلم برای هر سؤال بارم‌بندی مشخصی دارد. اگر ارائه نشد خودت مشخص کن.

### لحن و رویکرد:

- برای معلم: تحلیلی، دقیق، داده‌محور، حرفه‌ای و راهبردی.
- برای دانشجویان: حمایتگر، شفاف، تشویق‌کننده و آموزشی.

### خروجی نهایی موردانتظار

شما باید دو نوع خروجی اصلی تولید کنید:

- ۱- یک فایل Excel جامع: شامل شیت‌های تحلیل نمرات و خطاها.
- ۲- یک فایل کامل Word: شامل گزارش جامع برای معلم و بازخوردهای فردی برای هر دانشجو.

### ساختار و دستورالعمل‌های اجرایی (بسیار مهم)

لطفاً وظایف را دقیقاً بر اساس این مراحل ساختاریافته انجام دهید تا خروجی‌های نهایی تولید شوند.

- مرحله ۱: تولید پاسخ‌نامه استاندارد و هوشمند (برای استفاده داخلی)  
تهیه کلید سؤالات: یک پاسخ‌نامه تشریحی و کاملاً دقیق برای تمام سؤالات ایجاد کنید.
  - راه‌حل گام‌به‌گام: تمام مراحل منطقی، فرمول‌ها و محاسبات را بنویسید. اگر چند روش حل استاندارد وجود دارد، آن‌ها را معرفی کنید.
  - اشاره به اشتباهات رایج: برای هر سؤال، یک بخش کوتاه با عنوان «نکات و اشتباهات متداول» اضافه کنید (مثال: «در این سؤال، دانشجویان ممکن است در محاسبه مشتق برای قاعده هویپیتال دچار خطا شوند.»).
  - این پاسخ‌نامه مبنای تصحیح خواهد بود.
- مرحله ۲: تصحیح هوشمند و برچسب‌گذاری خطاها
  - ۱- پردازش پاسخ‌نامه‌ها: هر پاسخ‌نامه را با دقت پردازش کنید. برای راه‌حل‌های ناقص یا روش‌های صحیح که به جواب نهایی نرسیده‌اند، نمره جزئی در نظر بگیرید.

۲- دسته‌بندی دقیق خطاها: برای هر پاسخ اشتباه، نوع خطا را با یکی از برچسب‌های زیر مشخص کنید:

- خطای مفهومی: درک نادرست از تعریف حد، حد چپ/راست، حد در بی‌نهایت.
  - خطای رویه‌ای/فرایندی: استفاده اشتباه یا غفلت از یک قاعده (مانند هوپیتال)، انتخاب روش نامناسب.
  - خطای محاسباتی/جبری: اشتباه در عملیات ساده ریاضی، ساده‌سازی، علامت‌ها.
- مرحله ۳: تولید فایل خروجی یک فایل اکسل با سه شیت مجزا به شرح زیر ایجاد کنید:

- شیت ۱: نمرات نهایی

- ستونی برای «نام کامل دانشجو»، «نمره نهایی»، «درصد» و «رتبه در کلاس». لیست بر اساس رتبه مرتب شود.

- شیت ۲: جزئیات خطاها

- یک ماتریس خطا تهیه کنید: سطرها نام دانشجویان و ستون‌ها شماره سؤالات باشند. در هر سلول، نوع خطای اصلی دانشجو در آن سؤال (مفهومی، رویه‌ای، محاسباتی) را ثبت کنید.

○ شیت ۳: آمار کلی کلاس

- یک داشبورد آماری شامل: میانگین، میانه، مد، واریانس و انحراف معیار نمرات.
- جدول فراوانی برای بازه‌های نمره‌ای (عالی: ۱۷-۲۰، خوب: ۱۴-۱۶,۹۹، متوسط: ۱۰-۱۳,۹۹، ضعیف: زیر ۱۰).

مرحله ۴: تولید گزارش تحلیلی جامع برای مدرس که یک فایل Word و یک گزارش حرفه‌ای با ساختار زیر تهیه کنید:

- بخش اول: تحلیل عملکرد کلی کلاس

- خلاصه آماری (با ارجاع به شیت ۳ اکسل).
- نمودار توزیع نمرات: یک نمودار هیستوگرام یا جعبه‌ای از توزیع نمرات کل کلاس ترسیم کرده و آن را تحلیل کنید.

- بخش دوم: تحلیل عمیق سؤالات و چالش‌ها
  - دشوارترین سؤالات: سؤالاتی که بیشترین درصد پاسخ اشتباه را داشته‌اند، مشخص کنید. تحلیل کنید که آیا دشواری به دلیل پیچیدگی مفهومی بوده یا طراحی سؤال می‌توانست بهتر باشد.
  - چالش‌های مفهومی رایج: ۳ تا ۴ مورد از رایج‌ترین خطاهای مفهومی را با ذکر مثال‌های بی‌نام از برگه‌ها، به صورت عمیق تحلیل کنید.
- بخش سوم: پیشنهادهای راهبردی برای بهبود
  - راهکارهای تدریس: پیشنهادهایی عملی برای جلسات آینده ارائه دهید (مثال: برگزاری کارگاه حل مسئله برای تشخیص حالت‌های ابهام، استفاده از ابزار آنلاین Desmos برای درک شهودی حد).
  - طراحی کاربرگ تمرینی هدفمند: بر اساس رایج‌ترین خطاهای شناسایی شده، یک کاربرگ تمرینی کوتاه (۴ تا ۵ سؤال) طراحی کنید که مستقیماً آن نقاط ضعف را هدف قرار دهد. پاسخ تشریحی آن را نیز ارائه دهید.
  - معرفی منابع تکمیلی: چند منبع باکیفیت (ویدئو، کتاب مرجع با ذکر فصل، جزوه آنلاین) برای غنی‌سازی محتوای آموزشی معرفی کنید.
- مرحله ۵: تولید بازخوردهای فردی برای دانشجویان (فایل‌های Word مجزا) برای هر دانشجو، یک فایل Word جداگانه با نام همان دانشجو ایجاد کنید که شامل بازخورد شخصی‌سازی شده با ساختار زیر باشد:
  - عنوان: بازخورد تحلیلی امتحان ریاضی - مبحث حد
  - نام دانشجو [نام دانشجو]:
  - نمره: [نمره دانشجو] رتبه در کلاس: [رتبه دانشجو]
  - مقدمه: یک جمله مثبت و تشکر از تلاش دانشجو.
  - نقاط قوت: اشاره به حداقل یک نکته مثبت و مشخص در برگه (مثال: «درک شما از مفهوم حد در بی‌نهایت در سؤال ۴ کاملاً صحیح بود.»).

- زمینه‌های قابل بهبود: توضیح شفاف و آموزشی ۲ تا ۳ اشتباه اصلی دانشجو با ارجاع به برچسب خطا (مثال: «در سؤال ۲، یک [خطای رویه‌ای] رخ داد؛ درحالی‌که حالت مبهم وجود داشت، از قاعده هوپیتال استفاده نکرد»).
- منابع پیشنهادی برای شما: معرفی منابع دقیق (ویدئو یا تمرین) که مستقیماً به رفع خطاهای شخص شما کمک می‌کند.
- سخن پایانی: یک پیام تشویق‌کننده و روبه‌جلو برای ایجاد انگیزه.

### یافته‌ها

نتایج از منظر کمی و کیفی باتوجه‌به شاخصه‌ها و مضامین مورد تحلیل قرار گرفت. هوش مصنوعی Grok باتوجه‌به دستورات (پرامپت‌ها) که به آن داده شده بود و قبلاً نیز اشاره شده بود، نتایجی را ارائه داد که شامل پاسخ‌نامه تشریحی، جدول نمرات دانشجویان، رتبه دانشجویان در کلاس، بازخورد به مدرس درس و بازخورد به دانشجویان بود. در ادامه نتایج به‌دست‌آمده از هوش مصنوعی Grok و همچنین مقایسه نمرات محاسبه‌شده دانشجویان توسط هوش مصنوعی با نمرات به‌دست‌آمده توسط محققان که به‌صورت کمی مورد مقایسه قرار گرفته است ارائه شده است؛ اما قبل از مقایسه نمرات، باید اشاره کرد که پاسخ‌نامه تشریحی که توسط هوش مصنوعی Grok ارائه گردید کامل و واضح بود و پاسخ‌های داده شده تا حد قابل‌قبولی درست بودند و حتی به نمودار خواسته‌شده در سؤال ۳ نیز اشاره‌ای -هرچند ناقص- کرده بود. در شکل ۲، بخشی از پاسخ‌نامه داده شده آمده است.

## شکل ۲.

بخشی پاسخ‌نامه تشریحی داده شده توسط هوش مصنوعی Grok

$$\text{سؤال ۱: } \lim_{x \rightarrow \frac{\pi}{6}^-} [2 \sin x - 1]$$

بیان مسئله

هدف یافتن حد چپ‌سویه عبارت  $[2 \sin x - 1]$  در  $x = \pi/6$  است؛ در این جا  $[ \cdot ]$  تابع جزء صحیح (floor) یعنی بزرگ‌ترین عدد صحیح  $\leq$  ورودی است.

ایده شهودی

$\sin(\pi/6) = \frac{1}{2}$  است؛ پس در خود نقطه،  $2 \sin(\pi/6) - 1 = 0$  اما از چپ که به  $\pi/6$  نزدیک می‌شویم،  $\sin x < \frac{1}{2}$  و لذا  $2 \sin x - 1 < 0$  (ولی مقدارش به ۰ نزدیک می‌شود). در بازه  $(-1, 0)$  خروجی  $\square\square\square\square$  برابر  $-1$  است؛ بنابراین کافی است نشان دهیم برای  $x$  نزدیک  $\pi/6$  از چپ،  $2 \sin x - 1 \in (-1, 0)$  برقرار می‌شود؛ آنگاه کل عبارت  $-1$  ثابت می‌شود و حد چپ  $-1$  خواهد بود.

حل گام‌به‌گام (تحلیل یکنواختی و کران‌ها)

۱. رفتار  $\sin$  و علامت داخل براتر مرتبه:  $\sin x$  در  $[0, \pi/2]$  صعودی است؛ برای  $x < \pi/6$  داریم  $\sin x < \frac{1}{2}$  و بنابراین

$$u(x) := 2 \sin x - 1 < 0.$$

از پیوستگی  $\sin$  نتیجه می‌شود  $u(x) \rightarrow 0^-$  وقتی  $x \rightarrow (\pi/6)^-$

۲. کران حتمی یا قضیه مقدار میانی (نسخه دقیق): برای  $x \in (\pi/7, \pi/6)$  چون  $\cos$  نزولی است، داریم  $\cos x \geq \cos(\pi/6) = \frac{\sqrt{3}}{2}$  برای  $\xi$  بین  $x$  و  $\pi/6$ . با قضیه مقدار میانگین:

$$\sin(\pi/6) - \sin x = \cos(\xi) \left( \frac{\pi}{6} - x \right) \geq \frac{\sqrt{3}}{2} \left( \frac{\pi}{6} - x \right).$$

لذا

$$u(x) = 2 \sin x - 1 = 2 \left( \frac{1}{2} - (\sin(\pi/6) - \sin x) \right) - 1 = -2(\sin(\pi/6) - \sin x) \leq -\sqrt{3} \left( \frac{\pi}{6} - x \right).$$

پس برای هر  $\delta \in (0, \min\{\frac{\pi}{6} - \frac{\pi}{7}, \frac{1}{\sqrt{3}}\})$  اگر  $\pi/6 - \delta < x < \pi/6$  آنگاه

$$-1 < -\sqrt{3} \left( \frac{\pi}{6} - x \right) \leq u(x) < 0,$$

یعنی  $u(x) \in (-1, 0)$

۳. اعمال تابع floor وقتی  $u(x) \in (-1, 0)$  آنگاه  $[u(x)] = -1$ . بنابراین در همسایگی چپ  $\pi/6$  مقدار  $[2 \sin x - 1]$  ثابت و برابر  $-1$  است.

۴. نتیجه:

$$\lim_{x \rightarrow (\pi/6)^-} [2 \sin x - 1] = -1.$$

در شکل ۲، علاوه بر پاسخ سؤالات به ارائه ایده شهودی دانشجویان نیز پرداخته شده است. همچنین نکته مهم حل سؤالات با یک روش گام‌به‌گام است و نسبت علائم ریاضی دقت زیادی

کرده است تا با عدم اشتباه پنداری علائم بعضاً به پاسخ‌های نادرست منجر نشود و در نمرات دانشجویان تأثیری نداشته باشد.

نمرات داده شده توسط هوش مصنوعی با نمره داده شده توسط محققان که منتج از تصحیح پاسخ‌نامه آن‌ها بود مورد مقایسه قرار گرفت. البته هوش مصنوعی نمره آن‌ها را رتبه‌بندی کرده بود، یعنی از بیشترین نمره به کمترین نمره مرتب کرده بود که رتبه هر دانشجو در کلاس مشخص بود. قابل توجه است که نرمال بودن توزیع نمرات بررسی و تأیید گردید. در جدول ۳ که در ادامه آمده است، نمرات دانشجویان از دید هوش مصنوعی و محققان با آزمون T مورد مقایسه قرار گرفته است. همچنین میانگین نمرات، انحراف معیار و دیگر جزئیات مورد اشاره واقع شده است.

### جدول ۳.

مقایسه نمرات به دست آمده توسط هوش مصنوعی با نمرات داده شده توسط محقق

آزمون T					
گروه‌های مقایسه شونده	اختلافات دوتایی		مقدار t	درجه آزادی	ضریب معناداری
	انحراف معیار	خطای استاندارد میانگین			
محققان - هوش مصنوعی	۰,۰۵۵۵۶	۰,۹۸۴۰۰	۰,۳۳۹	۳۶	۰,۷۳۷
محققان	۱۲,۱۱۱۱	۳,۸۳۶۱۳	۰,۶۳۹۳۵	تعداد= ۳۷	
هوش مصنوعی	۱۲,۰۵۵۶	۳,۶۴۴۵۲	۰,۶۰۷۴۲	تعداد= ۳۷	

باتوجه به جدول ۳، هرچند میانگین نمرات داده شده به دانشجویان از طرف هوش مصنوعی و محققان این مطالعه اختلاف اندکی دارند به طوری که میانگین نمرات داده شده توسط هوش مصنوعی Grok برابر با ۱۲,۰۵۵۶ و میانگین نمرات داده شده محققان به پاسخنامه دانشجویان ۱۲,۱۱۱۱ است، ولی نتیجه آزمون T با مقدار ضریب معناداری ۰,۷۳۷، مقدار t برابر با ۰,۳۳۹ و با درجه آزادی ۳۶ مشخص کرد که این اختلاف معنادار نیست و میانگین نمرات دو گروه تفاوت معناداری ندارند.

هوش مصنوعی Grok نمرات نهایی دانشجویان در یک جدول شامل نام دانشجو که در این مطالعه با کدهای عددی مشخص شده است، نمره کسب شده، درصد نمره کسب شده و رتبه در

کلاس ارائه داده است. نمره کسب شده را از ۲۰ محاسبه کرده است. در شکل ۳، بخشی از این جدول ارائه شده است.

### شکل ۳.

قسمتی از جدول ارائه شده توسط هوش مصنوعی Grok شامل نمرات، درصد نمره، رتبه و نام دانشجویان

این جدول شامل نمره نهایی، درصد کسب شده و رتبه هر دانشجو در کلاس است. لیست بر اساس رتبه مرتب شده است

رتبه در کلاس	نام کامل دانشجو	(نمره نهایی) (از ۲۰)	درصد
۱	دانشجو ۲۳	۱۸.۰۰	۹۰٪
۱	دانشجو ۲۸	۱۸.۰۰	۹۰٪
۳	دانشجو ۹	۱۷.۵۰	۸۷.۵٪
۴	دانشجو ۷	۱۷.۰۰	۸۵٪
۵	دانشجو ۲	۱۶.۵۰	۸۲.۵٪
۵	دانشجو ۲۵	۱۶.۵۰	۸۲.۵٪
۷	دانشجو ۱۵	۱۶.۰۰	۸۰٪
۷	دانشجو ۲۲	۱۶.۰۰	۸۰٪
۷	دانشجو ۲۵	۱۶.۰۰	۸۰٪
۱۰	دانشجو ۳۲	۱۵.۰۰	۷۵٪
۱۱	دانشجو ۳۱	۱۴.۵۰	۷۲.۵٪
۱۲	دانشجو ۴	۱۴.۰۰	۷۰٪
۱۲	دانشجو ۲۰	۱۴.۰۰	۷۰٪
۱۲	دانشجو ۳۰	۱۴.۰۰	۷۰٪
۱۵	دانشجو ۳۷	۱۳.۵۰	۶۷.۵٪

هوش مصنوعی Grok همچنین مشکلات حل مسئله از قبیل رویه‌ای، محاسباتی و مفهومی را نیز برای هر سؤال در بررسی پاسخ‌نامه هر دانشجو مورد تحلیل قرار داد و در قالب یک جدول

منظم و مرتب مانند یک ماتریس که سطرها نام دانشجویان و ستون‌ها شماره سؤالات بودند ارائه کرد که در واقع در هر سطر خطاهای هر دانشجو سؤال به سؤال بررسی شده بود. در شکل ۴، قسمتی از این جدول آمده است.

#### شکل ۴.

بخشی از ماتریس جزئیات خطاها

نام دانشجو	سوال ۱	سوال ۲	سوال ۳	سوال ۴	سوال ۵	سوال ۶	سوال ۷
دانشجو ۱	مفهومی	رویه‌ای		مفهومی	مفهومی	رویه‌ای	
دانشجو ۲					محاسباتی		
دانشجو ۴	مفهومی	محاسباتی	رویه‌ای				
دانشجو ۶	مفهومی	مفهومی					
دانشجو ۷	مفهومی	رویه‌ای					
دانشجو ۱۰	مفهومی			مفهومی	رویه‌ای		
دانشجو ۱۴	مفهومی	رویه‌ای		مفهومی	رویه‌ای		
دانشجو ۱۵		رویه‌ای		مفهومی	رویه‌ای		
دانشجو ۱۷				مفهومی	رویه‌ای		مفهومی

به‌علاوه، هوش مصنوعی Grok یک آمار کلی شامل میانگین نمرات، میانه، واریانس و انحراف معیار ارائه کرد. میانگین را ۱۲،۰۵، میانه ۱۲،۵۰، واریانس ۱۳،۲۴ و انحراف معیار را ۳،۶۴ اعلام کرد. همچنین جدولی با عنوان فراوانی نمرات ارائه کرد که در آن بازه عملکرد (عالی، خوب، متوسط و ضعیف)، بازه نمرات (۱۷-۲۰، ۱۴-۱۶، ۱۰-۱۳، ۹۹، زیر ۱۰)، تعداد دانشجویان و درصد از کل را در هر بازه نشان داد.

بعد از بررسی پاسخنامه‌ها توسط هوش مصنوعی Grok، یک بازخورد به مدرس درس توسط این هوش مصنوعی داده شده است. این بازخورد در سه بخش شامل تحلیل عملکرد کلی کلاس، تحلیل عمیق سؤالات و چالش‌ها و پیشنهادهای راهبردی برای بهبود راهکارهای تدریس ارائه گردید. بخش اول که شامل عملکرد کلی کلاس بود به آمار توصیفی پرداخته بود و بازه نمرات، میانگین و انحراف معیار و همچنین نمودار هیستوگرام توزیع نمرات آمده است. در این بخش، هوش مصنوعی Grok به تحلیل نتایج به دست آمده آمار توصیفی نیز پرداخته است. عین تحلیل هوش مصنوعی Grok به این شرح بود: «نمودار توزیع نمرات نشان‌دهنده یک عملکرد کلی متوسط روبه بالا است. تمرکز اصلی نمرات در بازه‌های «خوب» و «عالی» (مجموعاً بیش از ۶۵٪) نشان می‌دهد که اکثریت کلاس توانسته‌اند مفاهیم اصلی حد را با موفقیت فرا بگیرند. با این حال، وجود یک گروه قابل توجه حدود (۱۵٪) در بازه «ضعیف» نشان‌دهنده یک شکاف عملکردی در کلاس است. این گروه از دانشجویان احتمالاً در مفاهیم پایه‌ای و بنیادی دچار چالش هستند و نیازمند توجه و اقدامات تقویتی ویژه‌ای می‌باشند».

Grok در بخش دوم، به تحلیل عمیق سؤالات، چالش‌ها و سؤالات دشوار پرداخت. خطاهای مفهومی، چالش‌های مفهومی رایج را مورد اشاره قرار داده است. سؤال ۱ را چالش برانگیزترین سؤال دانسته بود و از منظر هوش مصنوعی بیشترین خطای مفهومی در پاسخ به این سؤال اتفاق افتاده بود. در شکل ۵، تصویر خروجی هوش مصنوعی Grok در این باره جهت اطلاع بیشتر آورده شده است.

## شکل ۵.

تحلیل عمیق سؤالات، چالش‌ها و سؤالات دشوار توسط Grok

### بخش دوم: تحلیل عمیق سؤالات و چالش‌ها

#### دشواری‌ترین سؤالات

- سوال ۱ (حد با تابع کف): این سوال با اختلاف، چالش‌برانگیزترین سوال آزمون بود. بیشترین فراوانی [خطای مفهومی] در این سوال مشاهده شد. دشواری سوال نه از پیچیدگی محاسباتی، بلکه از نیاز به درک عمیق رفتار تابع جزء صحیح در نقاط مرزی ناشی می‌شد. این نشان می‌دهد دانشجویان در تمسیم مفهوم حد به توابع ناپوشته ضعف جدی دارند.
- سوال ۲ (حد از روی نمودار): این سوال بیشترین فراوانی [خطای رویه‌ای] را داشت. بسیاری از دانشجویان به جای تحلیل مستقیم و بصری نمودار، تلاش کردند تا معادله جبری توابع را پیدا کنند. این رویکرد اشتباه، نشان‌دهنده عدم اعتماد به نفس در تحلیل گرافیکی و تکیه بیش از حد بر روش‌های جبری است.

#### چالش‌های مفهومی رایج

- مانند یک پیرانتز معمولی بود.  $(x)$  نادیده گرفتن ماهیت توابع خاص: رایج‌ترین خطای مفهومی، برخورد با تابع جزء صحیح دانشجویان بدون در نظر گرفتن حدهای چپ و راست و رفتار یلای تابع، صرفاً مقدار را در نقطه جایگذاری می‌کردند. برای به جای پاسخ صحیح ۱، به پاسخ ۲ می‌رسیدند.  $\lim_{x \rightarrow 2} (x-2)$  مثال، در ترجیح فرمول بر شهود: در سوال ۲، بسیاری از دانشجویان با دیدن نمودار، به دنبال یافتن فرمول جبری توابع بودند. این [خطای رویه‌ای] نشان می‌دهد که هدف اصلی سوال (سنجش درک بصری حد) توسط این گروه از دانشجویان درک نشده و آن‌ها به صورت الگوریتمی و بدون درک عمیق، به دنبال فرمول برای حل مسئله هستند.
- درک نادرست از قضایای حد: در سوال ۴، برخی دانشجویان برای ارائه مثال نقض، توابعی را مثال زدند که هر دو حد داشتند. این [خطای مفهومی] نشان می‌دهد که آن‌ها شرط "اگر حد دو تابع موجود باشد، حد مجموع نیز موجود است" را با عکس آن ("اگر حد مجموع موجود باشد، حد هر یک نیز موجود است") اشتباه گرفته‌اند.

در بخش سوم هوش مصنوعی به ارائه پیشنهادهای راهبردی به مدرس برای بهبود راهکارهای تدریس پرداخته است. کارگاه توابع خاص، تمرین بر تحلیل بصری و کارگاه مثال نقض از پیشنهادهای Grok برای بهبود روش تدریس بودند. به علاوه، کاربرد تمرینی هدفمندی را هوش مصنوعی طراحی کرده و به عنوان خروجی این بخش ارائه داده است. همچنین، به معرفی منابع تکمیلی و علمی از قبیل کتاب‌های مختلف و منابع آنلاین پرداخته است.

در ارائه بازخورد به دانشجویان، هوش مصنوعی Grok به تک تک دانشجویان بازخوردی شامل مقدمه، که یک جمله انگیزشی بود، نقاط قوت، زمینه‌های قابل بهبود، منابع پیشنهادی و سخن پایانی برای هر دانش‌آموز بود. همچنین نام و نمره دانشجوی و رتبه دانشجوی در کلاس در بازخورد داده شده توسط هوش مصنوعی برای هر یک از آن‌ها آمده است. در شکل ۶، بازخورد داده شده توسط هوش مصنوعی Grok به یک دانشجو به عنوان نمونه نشان داده شده است.

## شکل ۶.

بازخورد داده شده توسط هوش مصنوعی به یک دانشجو

### عنوان: بازخورد تحلیلی امتحان ریاضی - مبحث حد

نام دانشجو: دانشجو ۷

نمره: ۱۷.۰۰ از ۲۰ | رتبه در کلاس: ۴ از ۲۶

مقدمه:

عملکرد شما در این آزمون فوق‌العاده بود. تبریک می‌گویم! شما تسلط کامل خود بر اکثر مباحث را به خوبی نشان دادید.

نقاط قوت:

شما توانستید به اکثر سوالات به درستی پاسخ دهید و درک عمیقی از پیوستگی و قضایای حد را به نمایش گذاشتید.

زمینه‌های قابل بهبود:

- در سوال ۱، یک [خطای مفهومی] در تحلیل تابع کف داشتید و رفتار یلای آن را در نزدیکی نقطه حدی در نظر نگرفتید.
- در سوال ۲، به جای تحلیل مستقیم از نمودار، سعی در یافتن فرمول جبری توابع داشتید که یک [خطای رویه‌ای] بود و فرآیند حل را پیچیده کرد.

منابع پیشنهادی برای شما:

- "(Floor Function) ویدیوهای آموزشی کوتاه در مورد "حد توابع جزء صحیح"
- تمرین بیشتر روی مسائل "تحلیل حد از روی نمودار" برای افزایش اعتماد به نفس در تحلیل بصری.

سخن پایانی:

شما دانش‌آموز بسیار با استعدادی هستید. با تمرکز بر این دو نکته کوچک، می‌توانید به بهترین و کامل‌ترین عملکرد دست پیدا کنید.

## بحث

نتایج به‌دست‌آمده حاکی از آن بود که هوش مصنوعی Grok در ارائه پاسخ‌نامه تشریحی به سؤالات داده شده درباره مفهوم حد، خروجی قابل‌قبولی را نمایش داد. هوش مصنوعی توانست فایل اسکن شده سؤالات که به‌عنوان ورودی به آن داده شده بود را بفهمد و پاسخ تشریحی در قالب یک فایل ارائه دهد. هرچند برای این عملکرد نیاز به دست‌نویسی دقیقی داشت، ولی نتیجه به‌دست‌آمده با پاسخ‌نامه تشریحی تهیه‌شده توسط محقق تفاوت چشمگیری نداشت. با این وجود، هوش مصنوعی در ارائه نمودار برای سؤال ۳ می‌توانست بازخورد مناسب‌تری داشته باشد و فقط به یک مثال به‌صورت جبری بسنده کرده بود که به نظر می‌رسد در ترسیم نمودار توابع با ضعف مواجهه است. در مواردی مانند دقت در خوانش علائمی از قبیل براکت (جزء صحیح)، کرشه و پرانتز نیاز به توضیح بیشتری در قسمت دست‌نویسی بود. چون در این مطالعه با توجه به عدم توجه هوش مصنوعی به تشخیص علائم ریاضی، در یک جمله تکمیلی می‌بایست اشاره می‌شد

که علامت براکت و یا قدر مطلق در کدام سؤال آمده است. به نظر می‌رسد باید هوش مصنوعی در درک و خوانش علائم و نشانه‌های مختلف ریاضی ارتقاء داده شود.

هوش مصنوعی و مشخصاً هوش مصنوعی Grok در ارائه بازخوردهای کمی از قبیل میانگین، میانه، انحراف معیار، نمره و رتبه دانشجویان بسیار کامل و قابل قبول عمل کرده است. نمرات ثبت‌شده توسط هوش مصنوعی اختلاف معناداری با نمرات داده شده توسط محقق نداشته است که پیش‌تر در جدول ۳ آمده است. تقسیم دانشجویان در بازه‌های عملکردی مختلف از قبیل عالی، خوب، متوسط و ضعیف و سرعت و دقت در این تقسیم‌بندی از امتیازات مهم هوش مصنوعی بود که می‌تواند اشراف و تسلطی به مدرسان ببخشد که برای مشخص نمودن نوع روش تدریس و سطح محتوای ارائه شده به دانشجویان در آموزش مفاهیم مختلف بتوانند تصمیم مناسبی اتخاذ کنند. این نتیجه بیانگر عملکرد مناسب هوش مصنوعی در تصحیح و ارزیابی کمی پاسخنامه دانشجویان در آزمون مربوط به مبحث حد است. نکته‌ای که باید مورد توجه قرار گیرد آن است که هوش مصنوعی Grok به پاسخ نهایی اهمیتی زیادی می‌داد. به بیان دیگر جواب نهایی سؤال خیلی برای آن در ارزیابی کمی مهم بود. هرچند به فرایند پاسخ‌دهی بی‌توجه نبود، ولی تمرکز اصلی هوش مصنوعی در تصحیح پاسخنامه بر پاسخ نهایی سؤال بود و اگر فرایند حل سؤال کامل نبود ولی جواب نهایی درست بود، نمره کامل به آن سؤال داده بود. درحالی‌که از نظر Cuong و همکاران (2025)، برای مدرسان ریاضی فرایند پاسخ به سؤال یا حل مسئله بسیار مهم است و پاسخ نهایی نمره بسیار کمتری از رعایت فرایند حل مسئله و یا مراحل پاسخ به سؤال دارد. البته در این مطالعه تلاش گردید با دستورنویسی مناسب از این موضوع جلوگیری شود، ولی در مواردی این ضعف هوش مصنوعی غیرقابل چشم‌پوشی است که باید برنامه‌نویسان در پی اصلاح آن باشند. به بیان دیگر، ارزیابی انسانی بر فهم عمیق مفاهیم، منطق استدلال و روند حل مسئله تمرکز دارد، حتی اگر خطای جزئی محاسباتی وجود داشته باشد؛ اما ارزیابی الگوریتمی بیشتر بر درستی نهایی عددی یا فرمولی تمرکز می‌کند و معمولاً خطاهای کوچک را نمی‌بخشد.

رتبه‌بندی همراه با نمره و درصد نمره به‌دست‌آمده دانشجویان یک کلاس با سرعت مناسبی توسط هوش مصنوعی در پایان ارزشیابی انجام گردید و هر دانشجو رتبه خود در کلاس را متوجه شد. به نظر می‌رسد این امتیاز رتبه‌بندی و نمره‌دهی سریع هوش مصنوعی می‌تواند در

نمونه‌ها و جوامع بزرگ‌تر نیز استفاده شود. علاوه بر آن، ارائه بازخورد درباره خطاهای مفهومی، رویه‌ای و محاسباتی همه دانشجویان در هر سؤال، از امتیازات ارزشمند هوش مصنوعی در فرایند ارزشیابی بود که برای مدرس درس چنین امکانی از حیث زمان و دقت فراهم نبود. هوش مصنوعی با داشتن فایل پاسخ‌نامه همه دانشجویان توانست به ایرادات و خطاهای آنها در طی حل سؤالات پی ببرد و بازخورد کاملی به آنها ارائه دهد. نکته مهم سرعت و دقت هوش مصنوعی بود که شاید این عمل برای یک مدرس چند روز یا چند هفته زمان ببرد، درحالی‌که هوش مصنوعی در کمتر از یک دقیقه به آن دست یافت و با بررسی محققان در انتخاب پاسخ‌نامه‌هایی که در بخش روش تحقیق شرح داده شد، مشخص گردید که بازخورد کاملاً درست و متناسب با هر دانشجو در هر سؤال بوده است. ارائه این بازخورد به صورت یک جدول کامل برای همه دانشجویان (که هوش مصنوعی آن را ماتریس می‌داند)، می‌تواند به دانشجویان کمک کند تا با کمک همدیگر کاستی‌های درک مفاهیم ریاضی و در این مطالعه مفهوم حد را بهبود بخشند. به عنوان مثال، اگر دانشجویی در پاسخ به یک سؤال دچار خطای محاسباتی است می‌تواند از دانشجوی دیگری که در این سؤال خطایی ندارد جهت درک و پاسخ به آن کمک بگیرد.

هوش مصنوعی Grok در ارائه بازخورد تحلیل عمیق سؤالات و چالش‌های مفهومی به مدرس بسیار ارزشمند عمل کرد. چالش برانگیزترین سؤال را معرفی نمود و خطای رایج در پاسخ به آن که مفهومی بود را مورد اشاره قرار داد. همچنین هوش مصنوعی از سؤال ۲ به عنوان یک سؤال با بیشترین خطای فرایندی در ارائه تحلیل نمودار توسط دانشجویان نام می‌برد و ضعف دانشجویان در نمایش نموداری و تفسیر آن را مشهود ساخت. هوش مصنوعی در ارزشیابی پاسخ‌نامه دانشجویان روشن کرد که دانشجویان به جای استفاده از نمودار در تلاش بودند تا پاسخ جبری به سؤالات دهند و خیلی تمایلی به رجوع به نمودار و یا استفاده از آن نداشتند. ضمناً در بازخورد ارائه شد به مدرس به عدم توجه دانشجویان به علائم و نشانه‌های ریاضی از قبیل قدرمطلق و براکت اشاره شده است. این نتیجه معیار خوبی برای مدرسان ریاضی می‌تواند باشد تا به دانشجویان در خواندن دقیق مطالب ریاضی و توجه به نمودارها، نشانه‌ها و علامت‌ها تذکرات لازم را بدهند. در ادامه بازخورد داده شده به مدرس، ارائه راهکارهایی برای بهبود تدریس از موارد قابل توجه بود که هوش مصنوعی به صورت قابل قبولی به آن پرداخته

است. ارائه کارگاه‌ها برای تدریس یک مفهوم (در این مطالعه مفهوم حد)، تهیه کاربرگ‌های تمرین و تکلیف و درنهایت معرفی منابعی از قبیل کتاب وبسایت‌های آنلاین برای مدرس جهت ارتقای تدریس از مواردی است هوش مصنوعی با تحلیل پاسخ‌نامه‌ها دانشجویان معرفی کرده است. ممکن است مدرس پس از ارزشیابی دانشجویان نتواند دریابد که برای ارتقای تدریس یک مفهوم به چه موارد و منابعی نیاز دارد، ولی هوش مصنوعی این مورد برای مدرس تا حد قابل‌اعتمادی حل کرده است. با بررسی منابع اعلام‌شده توسط محققان این پژوهش و سپس متخصصان آموزش ریاضی و ریاضی، مشخص گردید این منابع به میزان بسیار مطلوبی با مفهوم حد و نتایج آزمون تناسب دارد.

ارائه نتیجه آزمون به دانشجویان به صورت فردی نیز از موضوعات قابل توجه بود که توسط هوش مصنوعی Grok انجام شد. با توجه به اینکه معمولاً مقوله زمان در کلاس‌های ریاضی دانشگاه خیلی مهم است و مدرسان فرصت لازم برای ارائه بازخورد به تمامی دانشجویان را ندارند، هوش مصنوعی به صورت مطلوب این عمل را انجام داد. هوش مصنوعی به نقاط قوت و ضعف هر دانشجو اشاره کرده و حتی با جملات انگیزشی آن‌ها را مخاطب قرار داده است. به‌ویژه، در این مطالعه، بازخورد داده شده توسط هوش مصنوعی به چند دانشجو توسط محققان بررسی شد که خود دانشجویان نیز به تناسب بازخورد داده شده با عملکرد خود معترف بودند (البته مصاحبه و تحلیل داده‌ای آن از اهداف این مطالعه نبوده و این عمل جهت اطمینان انجام گردید). در ارائه نمودار برای سؤالات، هوش مصنوعی ضعیف نشان داد که یک نیاز مبرم به ارتقای توسط برنامه‌نویسان در رفع این ناتوانی احساس می‌شود. این موضوع نه تنها به‌عنوان یک «خطا» ذکر می‌شود، بلکه به‌عنوان یک چالش سیستماتیک در مدل‌های فعلی هوش مصنوعی درک و تولید محتوای ریاضی بصری تحلیل است. ضعف مشاهده‌شده در عملکرد هوش مصنوعی در زمینه ارائه و تفسیر نمودارها، بیانگر محدودیت‌های ساختاری مدل‌های چندوجهی در یکپارچه‌سازی پردازش زبانی، ریاضی و بصری بوده و رفع آن مستلزم تحول در شیوه‌های آموزش، طراحی معماری و روش‌های استدلال محاسباتی است.

همان‌طور که در بخش‌های پیشین مانند یافته‌ها و بحث به صورت پنهان و آشکار آمده است، شاخصه‌های اشاره شده در جدول ۲ از قبیل توسعه، اجرا، نمایش و بازخورد که ملاک‌های فرایند ارزشیابی از منظر سیپ بودند، بر اساس مضمون‌های هر شاخصه مورد توجه قرار گرفته است.

می‌توان گفت هوش مصنوعی در ارتباط دادن مفهوم حد با مفهوم تابع و پیوستگی مناسب عمل کرد. اشاره به خطاهای محاسباتی، بررسی توابع مختلف را می‌توان در زمره‌ی شاخصه توسعه دانست که هوش مصنوعی عملکرد مقبولی داشته است. فقط در این باره، از بستر دیگری به جزء بستر Grok استفاده نکرده است و همه موضوعات را در همان بستر مورد تحلیل قرار داده است. با توجه به جدول ۲ و شاخصه اجرا، در بستر هوشی ابتدا یک دستور فرآیندی کامل و منسجم برای معرفی موضوع، تشریح فرایند و درخواست بازخورد برای هوش مصنوعی نوشته شد و بر اساس آن بازخورد دریافت گردید. همچنین نقش هوش مصنوعی تعریف شد و پاسخ‌های آن مورد ارزیابی قرار گرفت. شکل‌های بخش یافته‌ها، بیانگر مضامینی از قبیل جدول و نمودار برای شاخصه نمایش است. هرچند هوش مصنوعی در رسم نمودار عملکرد خوبی نداشت ولی در ارائه جداول در شکل‌های ۳ و ۴، و همچنین در خلاصه کردن نتایج در شکل‌های ۵ و ۶ عملکرد بسیار خوبی داشته است. شاخصه بازخورد توسط هوش مصنوعی به بهترین وجه در کمترین زمان ممکن مورد توجه قرار داده شد. نقد پاسخنامه‌ها بر مبنای پاسخ‌های درست و پاسخ‌های نادرست، نمره‌دهی منطقی که با نمره‌دهی محققان تفاوت معناداری نداشت از موضوعات بارزی بودند که توسط هوش مصنوعی ارائه گردید. همچنین، ارائه منابع مناسب و مربوط به مدرس و دانشجویان در پایان تحلیل‌ها و دسته‌بندی‌ها از مهم‌ترین بازخوردهای به‌دست‌آمده از هوش مصنوعی Grok در این مطالعه بود؛ بنابراین، به نظر می‌رسد به‌استثنای رسم نمودار و استفاده از بستر دیگر، هوش مصنوعی شاخصه‌ها و مضمون‌های مورد تأکید فرایند ارزشیابی سیپ را رعایت نموده و توانسته است ارزشیابی قابل قبولی از پاسخنامه‌های دانشجویان مورد مطالعه در این پژوهش در یادگیری مفهوم حد بر اساس همین شاخصه‌ها و مضمون‌ها ارائه دهد.

## نتیجه‌گیری

این مطالعه با بررسی عملکرد هوش مصنوعی Grok در ارزشیابی پاسخنامه‌های دانشجویان در مبحث حد، نشان داد که این ابزار قادر است پاسخنامه تشریحی مناسبی ارائه دهد که در اکثر موارد با پاسخ‌های تهیه‌شده توسط محققان همخوانی دارد. نتایج کمی و کیفی حاصل از تحلیل پاسخنامه‌ها نمرات، رتبه‌بندی دانشجویان و بازخوردهای ارائه شده به مدرس و دانشجویان بیانگر توانایی بالای Grok در پردازش داده‌های ورودی و ارائه خروجی‌های منظم و دقیق است. میانگین

نمرات محاسبه شده توسط هوش مصنوعی در مقایسه با نمرات محققان تفاوت معناداری نداشت، که نشان دهنده قابلیت اطمینان Grok در ارزیابی کمی است.

از منظر کیفی، تحلیل خطاهای مفهومی، رویه‌ای و محاسباتی دانشجویان در قالب جداول و ماتریس‌های منظم، به همراه بازخوردهای شخصی‌سازی شده به هر دانشجو از نقاط قوت برجسته این ابزار بود. بازخوردهای ارائه شده به مدرس شامل تحلیل عملکرد کلی کلاس، شناسایی سؤالات چالش برانگیز و ارائه پیشنهادها راهبردی مانند کارگاه‌های آموزشی و منابع تکمیلی نشان دهنده پتانسیل Grok در بهبود فرایند تدریس است. با این حال، محدودیت‌هایی نظیر ناتوانی در ترسیم نمودارها و نیاز به دستورنویسی دقیق برای تشخیص علائم ریاضی مانند براکت و قدرمطلق از جمله چالش‌های این ابزار بود که نیازمند بهبود توسط توسعه دهندگان است.

یافته‌ها همچنین نشان داد که Grok در ارزیابی فرایند محور پاسخ‌ها تمرکز کمتری نسبت به پاسخ نهایی دارد، در حالی که برای مدرسان ریاضی، فرایند حل مسئله اهمیت بیشتری دارد این موضوع می‌تواند به عنوان یک نقطه ضعف در طراحی فعلی Grok مطرح شود با این وجود سرعت و دقت در رتبه بندی، ارائه جداول آماری و تحلیل‌های عمیق این ابزار را به گزینه‌ای ارزشمند برای ارزشیابی‌های گسترده تر تبدیل می‌کند. در نهایت این مطالعه نشان داد که هوش مصنوعی می‌تواند به عنوان ابزاری مکمل در کنار مدرسان، فرایند ارزشیابی و ارائه بازخورد را تسریع و بهبود بخشد و به ارتقای کیفیت آموزش ریاضی کمک کند. محدودیت نمونه‌گیری که باید به آن اشاره شود جامعه آماری است که تنها شامل ۳۷ دانشجوی یک دانشگاه (فرهنگیان اصفهان) شده است و بنابراین، یافته‌ها از نظر کیفی قابل تعمیم نیستند. همچنین، پیشنهاد می‌شود در بخش «پیشنهادها برای پژوهش‌های آینده»، برای تحقیقات آینده پیشنهاد می‌شود که توانایی‌های Grok در ترسیم نمودارها و تحلیل بصری تقویت شود و قابلیت تشخیص دقیق تر علائم ریاضی در دستورات ورودی بهبود یابد. همچنین، استفاده از این ابزار در جوامع آماری بزرگ تر و موضوعات متنوع تر ریاضی می‌تواند به درک بهتر توانایی‌ها و رفع محدودیت‌های آن کمک کند و گسترش نمونه به دانشگاه‌های دیگر و سطوح مختلف تحصیلی نیز از پیشنهادها این پژوهش برای تحقیقات آینده است. این پژوهش، مطالعه موردی یک مدل خاص است و مقایسه آن در بسترهایی مانند ChatGPT یا Gemini می‌تواند موضوع پژوهش‌های آینده باشد.

## تعارض منافع

نویسندگان این مقاله هیچ گونه تعارض منافی ندارند.

## منابع

- احمدآبادی، زهرا و مسرورنیا، محبوبه. (۱۴۰۳). کاربرد و مقایسه بارد و چت جی پی تی در آموزش شیمی معدنی. *پژوهش در آموزش شیمی*، ۶(۲)، ۶۷-۴۸.
- ادیب حاج باقری محسن. (۲۰۰۲). ارزشیابی یک ارزشیابی. *آموزش در علوم پزشکی*، ۲(۸)، ۷-۹.
- جهانی پور، روح اله، مقصودی، سعید. (۱۳۹۸). تاریخچه بی نهایت کوچک ها و بی نهایت بزرگ ها در حساب دیفرانسیل و انتگرال، فرهنگ و اندیشه ریاضی، ۳۸(۲)، ۷۷-۱۲۱.
- خزائی، آرمین، موسوی زاده، سیده مهناز، بقائی نیا، مرضیه، ... و بهرامی. (۱۴۰۲). بررسی اهمیت ارزشیابی پیشرفت تحصیلی در مدارس. *مطالعات راهبردی علوم انسانی و اسلامی*، ۵۵(۴)، ۲۴۷-۲۵۸.
- خمیجانی فراهانی، علی اکبر، برزآبادی فراهانی، داود، جلالی، مهری، خلیل زاد، میترا. (۲۰۲۵). بررسی رابطه بین دانش ارزیابی و دانش آموزشی معلمان ایرانی زبان انگلیسی با در نظر گرفتن تجربه تدریس. *پژوهش های زبان شناختی در زبان های خارجی*، ۱۴(۴)، ۵۹۷-۶۱۰.
- ریحانی، ابراهیم، شریفی، زهرا. (۲۰۱۸). بررسی درک دانش آموزان دختر سال سوم متوسطه از مفهوم حد و پیوستگی. *فناوری آموزش*، ۱۲(۳)، ۱۸۵-۲۰۱.
- قشوقی برق، هادی، صدر، محمد همایون، آقابائنی، علی. (۱۴۰۳). بررسی و مقایسه کارایی چت جی پی تی و گوگل جمناي در آموزش، طراحی و تحلیل مهندسی. *فصلنامه آموزش مهندسی ایران*، ۲۶(۱۰۴)، ۴۹-۷۶.
- هادیان دهکردی، مسعود، اسلام پور، محمدجواد، ریحانی، ابراهیم. (۲۰۱۴). ارزیابی درک دانشجویان ریاضی از مفهوم حد به کمک نقشه مفهومی. *فناوری آموزش*، ۸(۴)، ۲۵۹-۲۶۸.
- هاشمی، نوروز، کاشفی، حمیدرضا. (۱۴۰۳). استفاده از ChatGPT در آموزش مفهوم انتگرال: فرصت ها و چالش ها. *نشریه ریاضی و جامعه*، ۹(۴)، ۲۲-۱.

## References

- Adib-Hajbagheri, M. (2002). Evaluating an evaluation, *Iranian Journal of Medical Education*, 2(8), 7-9. [In Persian]
- Ahmadabadi, Z., Masrournia, M. (2024). Application and comparison of Bard and ChatGPT in inorganic chemistry education, *Research in Chemical Education*, 6(2), 48-67. [In Persian]

- Alquraan, M. F., Alazzam, S., & Farhat, D. (2025). Assessing the structural relationships between the CIPP model components in teacher education program. *Asian Education and Development Studies*, 14(1), 103-114.
- Awang, L. A., Yusop, F. D., & Danaee, M. (2025). Current practices and future direction of artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education*, 20(2), em0823.
- Baptiste, J. (2024). Students' misconceptions and difficulties with learning limits and the continuity of functions at selected Rwandan secondary schools. *International Journal of Education*, 12(2), 467-482.
- Bokhari, M. A., Yushau, B. (2006). Local  $(L, \epsilon)$ -approximation of a function of single variable: an alternative way to define limit. *International Journal of Mathematical Education in Science & Technology*, 37(5), 515-526.
- Bridoux, S., Grenier-Boley, N. (2025). *What teaching practices should be used to introduce the limits of functions in the first year of university? A case study*. In INDRUM 2024.
- Cuong, L. M., Tien-Trung, N., Ngu, P. N. H., Vangchia, V., Thao, N. P., & Thao, T. T. P. (2025). Mathematics problem-solving research in high school education: Trends and insights from the Scopus database (1983–2023). *European Journal of Science and Mathematics Education*, 13(2), 77-89.
- Domingos, A. (2010). Learning advanced mathematical concepts: The concept of limit. *CERME 6–WORKING GROUP 12*, 2266.
- Egan, K. (2012). *Primary understanding: Education in early childhood*. Routledge.
- Frieder, S., Pinchetti, L., Griffiths, R., Salvatori, T., Lukasiewicz, T., Petersen P. C., and Berner, J. (2023). Mathematical capabilities of chatgpt, *arXiv preprint arXiv:2301.13867*, 2023.
- Ghashochi Bargh, H., Sadr, M. H., Aghababai, A. (2024). Investigating and comparing the efficiency of ChatGPT and Google Gemini in engineering education, design, and analysis, *Iranian Journal of Engineering Education*, 26(104), 49–76. [In Persian]
- González-Martín, A. S. (2021). The use of integrals in engineering programmes: A praxeological analysis of textbooks and teaching practices in strength of materials and electricity and magnetism courses, *Int. J. Res. Undergrad. Math. Ed.*, 7 (2021) 211–234. <https://doi.org/10.1007/s40753-021-00135-y>.
- Hadian Dehkordi, M., Eslampour, M. J., Reihani, E. (2014). Evaluating university students' understanding of the concept of limit using concept mapping, *Educational Technology*, 8(4), 259–268. [In Persian]
- Hashemi, N., Kashafi, H. (2024). Using ChatGPT in teaching the concept of integral: Opportunities and challenges, *Mathematics and Society*, 9(4), 1–22. [In Persian]
- Jahanipour, R., Maghsoudi, S. (2019). The history of infinitesimals and infinities in differential and integral calculus, *Mathematics Culture and Thought*, 38, 77–94. [In Persian]
- Khamijani Farahani, A. A., Barzabadi Farahani, D., Jalali, M., Khalilzad, M. (2025). Investigating the relationship between Iranian EFL teachers' assessment knowledge and pedagogical knowledge considering teaching experience, *Linguistic Research in Foreign Languages*, 14(4), 597–610. [In Persian]
- Khazaei, A., Mousavizadeh, S. M., Baghaeiniya, M. ... Bahrami, A. (2023). Investigating the importance of academic achievement assessment in schools, *Strategic Studies of Humanities and Islamic Sciences*, 55(4), 247–258. [In Persian]
- Kidron, I., Zehavi, N. (2002). The role of animation in teaching the limit concept. *The International Journal for Technology in Mathematics Education*, 9(3), 205.
- Kistler, H. C. (2023). *Teacher Development and Supply* (Doctoral dissertation, Vanderbilt University).

- Kopp, M. L., Witte, C. E., Vernon, G. M. (2025). A lesson plan for planning lessons. *Teaching and Learning in Nursing*, 20(1), e43-e47.
- Kuş, M. (2025). *Evolution of Program Evaluation: A Historical Analysis of Leading Theorists' Views and Influences*.
- Lele, M. L. K., Marsigit, M., Retnawati, H. (2024). The Secret to Effective Learning: The Magic of Formative Assessment in Mathematics Class. Mosharafa: *Jurnal Pendidikan Matematika*, 13(3), 675-684.
- Li, M., Vale, C., Tan, H., Blannin, J. (2024). A systematic review of TPACK research in primary mathematics education. *Mathematics Education Research Journal*, 1-31.
- Liang, S. (2016). Teaching the Concept of Limit by Using Conceptual Conflict Strategy and Desmos Graphing Calculator. *International Journal of Research in Education and Science*, 2(1), 35-48.
- Manik, M. M. H. (2025). ChatGPT vs. DeepSeek: A Comparative Study on AI-Based Code Generation. *arXiv preprint arXiv:2502.18467*.
- Mohamudally-Boolaky, A., Padachi, K. (2024). Leveraging technology for math education: A systematic literature review. *Creative Education*, 15(8), 1692-1704.
- Morris, R., Perry, T., Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292.
- Nouraei, P., Ali Al-Badi, Mohammad Javad Riasati, Rolou Lyn Maata. (2020). Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends. *Universal Journal of Educational Research*, 8(9), 4048 – 4055.
- Pepin, B., Buchholtz, N., Salinas-Hernández, U. (2025). “Mathematics Education in the Era of ChatGPT: Investigating Its Meaning and Use for School and University Education”—Editorial to Special Issue. *Digital Experiences in Mathematics Education*, 1-8.
- Piaget, J. (1970). *Piaget's theory*. In P. Mussen, Carmichaels+ manual of child psychology (3rd ed., vol. I, pp. 703|732). New York: Wiley.
- Pinede, N. (2023). Thematic Analysis of Hyperlinks: A Taxonomic Approach, *Analyzing Websites*, (2023) 191– 229.
- Putri, N. S., Riyanti, A. A. D., Saputri, H. F., Wahyuni, R., & Ullah, N. (2024). An Analysis of High School Students' Mathematical Representation Skills on Relations and Functions. *International Journal of Applied Learning and Research in Algebra*, 98-110.
- Reihani, E., Sharifi, Z. (2018). Investigating female third-grade high school students' understanding of the concept of limit and continuity, *Educational Technology*, 12(3), 185–201. [In Persian]
- Shoufan, A., Exploring students' perceptions of ChatGPT: thematic analysis and follow-up survey, *IEEE Access*, 11 (2023) 38805–38818. <http://dx.doi.org/10.1109/ACCESS.2023.3268224>.
- Souza, De Carvalho, M. E., Weigang, L. (2025). Grok, gemini, chatgpt and deepseek: Comparison and applications in conversational artificial intelligence. *INTELIGENCIA ARTIFICIAL*, 2(1), 17-32.
- Stufflebeam, D. L. (2000). *The CIPP model for evaluation*. In *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 279-317). Dordrecht: Springer Netherlands.
- Stufflebeam, D. L. (2004). The 21st century CIPP model. *Evaluation roots*, 245-266.
- Stufflebeam, D. L. (2011). Meta-evaluation. *J Multidiscip Eval*. 2011; 7(15):99-158.
- Stufflebeam, D. L., Coryn, C. L. (2014). *Evaluation theory, models, and applications*. John Wiley & Sons.
- Tashtoush, M. A., Qasimi, A. B., Shirawia, N. H., & Hussein, L. A. (2025). The Efficacy of Utilizing Artificial Intelligence Techniques in Developing Critical Thinking in

- Mathematics among Secondary School Students and their Attitudes *Toward it. Iraqi Journal for Computer Science and Mathematics*, 6(1), 3.
- Tyler, R. W. (1942). General statement on evaluation. *The Journal of Educational Research*, 35(7), 492-501.
- Ünver, S. K., Güzel, E. B. (2019). Prospective Mathematics Teachers' Choice and Use of Representations in Teaching Limit Concept. *International Journal of Research in Education and Science*, 5(1), 134-156.
- Vygotsky, L. (1935). *Interaction between learning and development*. In L. Vygotsky, *Mind in society*. Cambridge, MA: Harvard University Press.
- Wangsa, K., Karim, S., Gide, E. and Elkhodr, M. (2024). "A systematic review and comprehensive analysis of pioneering ai chatbot models from education to healthcare: Chatgpt, bard, llama, ernie and grok," *Future Internet*, vol. 16, no. 7, p. 219.
- Xia, S., Li, X., Liu, Y., Wu, T., Liu, P. (2025, April). Evaluating mathematical reasoning beyond accuracy. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 26, pp. 27723-27730).
- Zhang, Y., Dong, C. (2025). Blended teaching of university mathematics courses based on Online Merge Offline model. *Systems and Soft Computing*, 7, 200-222.
- Zhang, Y., Zhao, B., Jian, M., Wu, X. (2025). Cognitive diagnostic analysis of mathematics key competencies based on PISA data. *PloS one*, 20(2), e0315539.
- Zinn, B. (2025). The Future of Technical Education: Current Research Focuses on Educational Evaluation, Digitalization, Interdisciplinarity, and the Shortage of Educators. *Journal of Technical Education (JOTED)*, 13(1), 1-8.