Theory and Practice in Teachers Education

Effects of ChatGPT-Assisted Practice Question Answering on Learning, Judgment of Learning, and Metacognitive Monitoring Among Student Teachers

Amin Hossein Pour[1*] | Reza Alam[2] | Seyyed Ali Mosallemi[3] |

1. ***Corresponding author,*** Department of Psychology, Faculty of Humanities, Gorgan Branch, Islamic Azad University, Gorgan, Iran. E-mail: aminteacher1994@gmail.com
2. Department of Educational Sciences, Faculty of Humanities, Azadshahr Branch, Islamic Azad University, Azadshahr, Iran. E-mail: reza.alam72@gmail.com
3. Department of Educational Technology, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: mosallamiedu@gmail.com

Article type:
Research Article

Keywords:
ChatGPT, Generative Artificial Intelligence, Judgment of Learning, Metacognitive Monitoring, Student Teachers' Learning.

ABSTRACT

Background and Objectives: The growing use of generative artificial intelligence in education has raised important questions about its effects on actual learning and metacognitive processes. This study aimed to examine the effect of using ChatGPT for answering practice questions on learning, Judgment of Learning (JOL), judgment bias, and absolute metacognitive monitoring error among student teachers at Farhangian University. **Methods:** This applied study employed a quantitative quasi-experimental pretest–posttest design with a control group. The statistical population consisted of student teachers at Shahid Beheshti Campus of Farhangian University in Gonbad-e Kavus during the 2025–2026 academic year. The final sample included 50 student teachers who were assigned to an experimental group and a control group, with 25 participants in each group. The experimental group used ChatGPT to answer practice questions, whereas the control group answered the same questions independently. The research instruments included a learning achievement test, a Judgment of Learning scale, a judgment bias index, and an absolute metacognitive monitoring error index. Since Judgment of Learning, judgment bias, and absolute metacognitive monitoring error were post-task variables, they had no independent pretest; therefore, the learning pretest was controlled as an indicator of baseline knowledge. The data were analyzed using analysis of covariance (ANCOVA) in SPSS. **Findings:** The results showed that, after controlling for the learning pretest, there was no significant difference between the two groups in actual learning. However, Judgment of Learning was significantly higher in the ChatGPT group. In addition, judgment bias and absolute metacognitive monitoring error were higher in the ChatGPT group, indicating increased overestimation of learning and reduced self-assessment accuracy. **Conclusion:** The findings indicated that, under the conditions of this study, using ChatGPT to answer practice questions affected learners' perceived learning and metacognitive monitoring accuracy more than it improved actual learning. Therefore, educational use of ChatGPT should be guided and structured through self-explanation, critical evaluation of responses, and active cognitive engagement in order to prevent false confidence and superficial learning.

Cite this Article: Hossein Pour, A., Alam, R., & Mosallemi, S. A. (2026). The effect of using generative artificial intelligence (ChatGPT) in answering practice questions and summarizing texts on learning and metacognitive monitoring of Farhangian University students. *Theory and Practice in Teacher Education*, 10(18), 1-16. https://doi.org/10

Extended Abstract

Introduction

Generative artificial intelligence (AI) is rapidly transforming the nature of learning, practice, and assessment in higher education. Tools such as ChatGPT can generate coherent explanations and provide immediate answers to academic questions; however, their growing use has blurred the boundary between learning through active cognitive effort and obtaining ready-made responses. This issue is particularly important in practice-question answering, which traditionally promotes retrieval, identifies knowledge gaps, and guides subsequent study. When learners delegate the generation of answers to AI, the cognitive processes underlying retrieval practice may be weakened, even if the resulting responses appear accurate and well organized.

Beyond actual achievement, using ChatGPT may also influence metacognitive monitoring—that is, learners' ability to assess how much they have learned and how well prepared they are for a test. Inaccurate monitoring can produce false confidence, premature termination of study, and insufficient preparation. This concern is especially significant among student teachers, who will later design learning activities, assignments, and assessments for their own students. Accordingly, the present study examined the effects of using ChatGPT to answer practice questions on actual learning, judgment of learning (JOL), judgment bias, and absolute metacognitive monitoring error among student teachers at Farhangian University.

Research Question(s)

The study addressed the following questions:

- Does using ChatGPT to answer practice questions affect student teachers' actual learning?
- Does using ChatGPT to answer practice questions affect student teachers' judgment of learning?
- Does using ChatGPT to answer practice questions affect student teachers' judgment bias?
- Does using ChatGPT to answer practice questions affect student teachers' absolute metacognitive monitoring error?

Literature Review

The theoretical framework of the study integrated retrieval practice theory, cognitive offloading theory, and models of metacognitive monitoring. Retrieval practice theory proposes that learning becomes deeper and more durable when learners actively retrieve information, generate responses, and reconstruct their knowledge rather than merely observe an available answer (Wilson & Bauer, 2024). Cognitive offloading theory, in contrast, explains how individuals transfer part of their mental processing to external tools. Although offloading can facilitate task completion and reduce cognitive demands, excessive reliance on external resources may reduce active processing and weaken learning (Tao et al., 2024).

Metacognitive models further suggest that judgments of learning can be influenced by superficial cues such as processing fluency, response coherence, and ease of access to information (Carpenter et al., 2020). Consequently, fluent and confident-looking AI-generated responses may create an illusion of competence, leading learners to interpret

ease of processing as evidence of deep learning. Previous research has presented two competing views of generative AI: it may function as an instructional scaffold that provides explanations and feedback, or as a cognitive shortcut that replaces reasoning and reduces productive effort. The effects of AI may therefore depend less on the technology itself and more on whether learners critically evaluate, explain, and reconstruct its responses. However, experimental evidence concerning the cognitive and metacognitive consequences of ChatGPT use remains limited in the Iranian teacher-education context.

Methodology

This applied quantitative study employed a quasi-experimental pretest–posttest design with a control group. The statistical population consisted of student teachers enrolled at the Shahid Beheshti Campus of Farhangian University in Gonbad-e Kavus during the 2025–2026 academic year. Two intact classes taking the same Educational Psychology course with the same instructor were selected through convenience sampling. Initially, 52 student teachers participated, but two cases were excluded because of incomplete data. The final sample therefore comprised 50 participants, with 25 in the experimental group and 25 in the control group. Because individual randomization was not feasible, the two classes were randomly assigned to the experimental and control conditions at the class level.

The experimental group used the same version of ChatGPT and a standardized prompt to answer the practice questions. Participants were permitted a maximum of three consecutive interactions with the system and were required to rewrite the final answers in their own words. All interactions were completed during a supervised classroom session. The control group answered the same questions independently, using only the instructional content provided. To maintain cognitive comparability, control-group participants were also required to provide a brief justification for each answer.

The learning instrument was a researcher-developed 12-item multiple-choice achievement test. Its content validity was evaluated by three specialists in educational psychology and research methodology. Two equivalent forms were developed for the pretest and posttest, and the Kuder–Richardson 20 reliability coefficient was .81. Judgment of learning was measured by asking participants to estimate their expected performance on a scale from 0 to 100. Judgment bias was calculated by subtracting actual performance, converted to a percentage, from JOL. Positive scores indicated overestimation. Absolute metacognitive monitoring error was calculated as the absolute difference between JOL and actual performance, with higher scores indicating lower monitoring accuracy.

Both groups completed informed consent, the learning pretest, the instructional material, the practice activity, the JOL assessment, and the learning posttest under similar conditions. Data were analyzed in SPSS 26 using descriptive statistics and analysis of covariance (ANCOVA). The learning pretest was controlled as an indicator of baseline knowledge. Because JOL, judgment bias, and monitoring error were post-task measures, they had no independent pretests. ANCOVA assumptions, including homogeneity of variance and regression slopes, were examined before the main analyses..

Results

The preliminary analyses showed no significant baseline differences between the groups in age, previous experience with ChatGPT, interest in the instructional topic, perceived familiarity with the content, or learning pretest scores. The control and ChatGPT groups obtained pretest means of 9.14 and 9.42, respectively.

Although posttest learning increased in both groups, the difference between the adjusted learning means of the control group (12.26) and the ChatGPT group (12.61) was not statistically significant, $F(1, 47) = 0.45$, $p > .05$. In contrast, JOL was significantly higher in the ChatGPT group than in the control group, $F(1, 47) = 18.10$, $p < .001$. Their adjusted JOL means were 77.98 and 66.74, respectively.

The mean positive judgment bias was also higher in the ChatGPT group (14.28) than in the control group (5.01), indicating greater overestimation of performance. Furthermore, absolute metacognitive monitoring error was significantly higher among ChatGPT users, $F(1, 47) = 18.80$, $p < .001$. The adjusted error means were 24.36 for the ChatGPT group and 20.11 for the control group. Thus, ChatGPT increased perceived learning and confidence without producing a corresponding significant improvement in actual achievement.

Discussion

The findings suggest that ChatGPT functioned more as a cognitive shortcut than as an effective learning scaffold under the conditions of this study. The fluent, coherent, and immediate responses generated by ChatGPT may have served as persuasive external cues, creating a sense of mastery that was not fully supported by test performance. This pattern can be explained through the illusion of competence and processing fluency: learners may mistake the ease of understanding an AI-generated response for their own ability to retrieve and apply the information independently.

The absence of a significant learning advantage may also be explained by reduced retrieval effort. Even though participants rewrote ChatGPT's responses, some of the essential cognitive work of generating, evaluating, and reconstructing answers may have been transferred to the system. The metacognitive findings are particularly important because inflated JOL and reduced monitoring accuracy can distort self-regulated learning decisions, causing students to stop studying before adequate mastery has been achieved..

Conclusion

Using ChatGPT to answer practice questions did not significantly improve student teachers' actual learning, but it increased their judgments of learning, overestimation, and metacognitive monitoring error. Educational effectiveness therefore depends on how ChatGPT is incorporated into learning activities. Students should first generate their own answers and then compare, critique, and revise them using AI. ChatGPT should also provide prompts, hints, or guiding questions rather than complete final answers.

The findings should be interpreted cautiously because the study involved two intact classes, a relatively small sample from one university campus, a single type of learning task, and self-reported judgments of learning. Future studies should include larger samples, more classes, delayed achievement tests, and direct indicators of cognitive engagement. Teacher-education programs should also integrate AI literacy and

metacognitive training to ensure that generative AI supports active learning rather than replacing it.

Ethical Considerations

Compliance with Research Ethics

The authors adhered to all ethical principles in conducting and publishing this study, and all authors have confirmed their compliance with these principles.

Authors' Contributions

All authors contributed equally to the preparation and writing of the manuscript.

Acknowledgments

The authors would like to express their sincere gratitude to all individuals who participated in this study.

Conflict of Interest

The authors declare that they have no conflicts of interest.

تأثیر استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی بر یادگیری، قضاوت یادگیری و پایش فراشناختی دانشجومعلمان دانشگاه فرهنگیان

امین حسین پور*[۱]، رضا علم[۲]، سید علی مسلمی[۳]

۱. ***نویسنده مسئول***، گروه روان‌شناسی، دانشکده علوم انسانی، واحد گرگان، دانشگاه آزاد اسلامی، گرگان، ایران. رایانامه: aminteacher1994@gmail.com

۲. گروه علوم تربیتی، دانشکده علوم انسانی، واحد آزادشهر، دانشگاه آزاد اسلامی، آزادشهر، ایران. رایانامه: reza.alam72@gmail.com

۳. گروه تکنولوژی آموزشی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: mosallamiedu@gmail.com

چکیده

نوع مقاله: پژوهشی

پیشینه و اهداف: گسترش استفاده از هوش مصنوعی مولد در آموزش، پرسش‌های مهمی را درباره پیامدهای آن بر یادگیری واقعی و فرایندهای فراشناختی ایجاد کرده است. پژوهش حاضر با هدف بررسی تأثیر استفاده از ChatGPT در پاسخگویی به سؤالات تمرینی بر یادگیری، قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی دانشجومعلمان دانشگاه فرهنگیان انجام شد. **روش‌ها:** این پژوهش از نظر هدف کاربردی، از نظر رویکرد کمی و از نظر شیوه اجرا نیمه‌آزمایشی با طرح پیش‌آزمون–پس‌آزمون همراه با گروه کنترل بود. جامعه آماری پژوهش را دانشجومعلمان دانشگاه فرهنگیان، واحد شهید بهشتی گنبدکاووس، در سال تحصیلی ۱۴۰۴–۱۴۰۵ تشکیل دادند. نمونه نهایی شامل ۵۰ دانشجومعلم بود که در دو گروه ۲۵ نفری آزمایش و کنترل قرار گرفتند. گروه آزمایش برای پاسخگویی به سؤالات تمرینی از ChatGPT استفاده کرد، در حالی که گروه کنترل همان سؤالات را به‌صورت مستقل پاسخ داد. ابزارهای پژوهش شامل آزمون یادگیری، مقیاس قضاوت یادگیری، شاخص سوگیری قضاوت و شاخص خطای مطلق پایش فراشناختی بود. برای متغیرهای قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی، پیش‌آزمون مستقل وجود نداشت و پیش‌آزمون یادگیری به‌عنوان شاخص دانش پایه کنترل شد. داده‌ها با استفاده از تحلیل کوواریانس در نرم‌افزار SPSS تحلیل شدند. **یافته‌ها:** نتایج نشان داد پس از کنترل پیش‌آزمون یادگیری، تفاوت معناداری در یادگیری واقعی دو گروه مشاهده نشد؛ اما قضاوت یادگیری در گروه استفاده‌کننده از ChatGPT به‌طور معناداری بالاتر بود. همچنین سوگیری قضاوت و خطای مطلق پایش فراشناختی در گروه ChatGPT بیشتر مشاهده شد که نشان‌دهنده افزایش بیش‌برآورد یادگیری و کاهش دقت خودارزیابی بود. **نتیجه‌گیری:** یافته‌ها نشان داد استفاده از ChatGPT در پاسخگویی به سؤالات تمرینی، در شرایط این پژوهش، بیش از آنکه موجب ارتقای معنادار یادگیری واقعی شود، بر ادراک یادگیری و دقت پایش فراشناختی اثر گذاشت. بنابراین، بهره‌گیری آموزشی از ChatGPT باید به‌صورت هدایت‌شده، همراه با خودتوضیح‌دهی، ارزیابی انتقادی پاسخ‌ها و الزام به درگیری فعال شناختی طراحی شود تا از شکل‌گیری اعتمادبه‌نفس کاذب و یادگیری سطحی جلوگیری شود.

واژه‌های کلیدی:
پایش فراشناختی، ChatGPT، قضاوت یادگیری، هوش مصنوعی مولد، یادگیری دانشجومعلمان.

استناد به این مقاله: حسین پور، امین، علم، رضا، و مسلمی، سید علی. (۱۴۰۵). تأثیر استفاده از هوش مصنوعی مولد (ChatGPT) در پاسخ‌گویی به سؤالات تمرینی و خلاصه‌سازی متون بر یادگیری و پایشِ فراشناختی دانشجویان دانشگاه فرهنگیان. *نظریه و عمل در تربیت معلمان*، ۱۰(۱۸)، ۱-۱۶.

مقدمه

نظام‌های تعلیم و تربیت در جهان امروز با یک دگرگونی فناورانه روبه‌رو هستند که ماهیت یادگیری، تمرین، و حتی ارزیابی را تغییر داده است (ابراهیمی و ضرابیان، ۱۳۹۷؛ Benatar & Daneman, 2020). ورود هوش مصنوعی مولد[1] به محیط‌های آموزشی، به‌ویژه ابزارهایی که می‌توانند متن تولید کنند (Mehta et al., 2025) و به پرسش‌ها پاسخ دهند (Ahmed et al., 2025)، باعث شده است مرز میان یادگیریِ حاصل از کوشش شناختیِ دانشجو و دریافت پاسخ آماده کمرنگ‌تر شود (Chen et al., 2025). در چنین شرایطی، مسئله فقط دسترسی به یک فناوری جدید نیست؛ بلکه پرسش اصلی این است که این فناوری چگونه بر کیفیت یادگیری و سازوکارهای ذهنیِ دانشجو در مسیر یادگیری اثر می‌گذارد.

یکی از پرکاربردترین شیوه‌های یادگیری در دانشگاه‌ها، پاسخ‌گویی به سؤالات تمرینی و مرور آزمون‌محور برای تثبیت محتواست (Guilding et al., 2021). سؤالات تمرینی معمولاً نقش تمرین بازیابی (Chien et al., 2024)، تشخیص نقاط ضعف، و هدایت مطالعه (Dell & Wantuch, 2017; Logren et al., 2017) را ایفا می‌کنند و همین کارکردها آن‌ها را به قلبِ یادگیریِ مؤثر تبدیل کرده است. با این حال، وقتی دانشجو به جای درگیری شناختی با سؤال و تولید پاسخ، از هوش مصنوعی برای دریافت پاسخ استفاده می‌کند، ممکن است سازوکار اصلی تمرین یعنی تلاش برای بازیابی و استدلال، تضعیف شود یا شکل دیگری پیدا کند. بنابراین، مسئله محوری این پژوهش از همین نقطه آغاز شد: استفاده از هوش مصنوعی در پاسخ‌گویی به سؤالات تمرینی، آیا یادگیری را تقویت می‌کند یا به شکل پنهان، آن را سطحی‌تر می‌سازد؟

پژوهش‌های جدید نشان داده‌اند که استفاده از هوش مصنوعی[2] مولد می‌تواند با کاهش تلاش شناختی و افزایش برون‌سپاری شناختی همراه باشد؛ وضعیتی که در آن یادگیرنده بخشی از فرایند پردازش، بازیابی و تولید دانش را به ابزار بیرونی واگذار می‌کند (Fisher & Oppenheimer, 2021; Risko & Gilbert, 2016). اگرچه این فرایند ممکن است سرعت انجام تکلیف را افزایش دهد، اما هم‌زمان می‌تواند به کاهش پردازش عمیق و شکل‌گیری نوعی توهم تسلط یا خطای فراشناختی منجر شود؛ به‌گونه‌ای که فرد احساس کند مطالب را آموخته، در حالی‌که یادگیری پایدار به‌طور کامل شکل نگرفته است.

در کنار خودِ یادگیری، یک مؤلفه بسیار تعیین‌کننده در موفقیت تحصیلی وجود دارد و آن پایش فراشناختی است (Agrawal et al., 2025; Daher & Hashash, 2022؛ اسماعیلی و همکاران، ۱۳۹۹)؛ یعنی توانایی فرد برای قضاوت درباره اینکه چقدر یاد گرفته و چقدر برای آزمون آماده است (Jaeger & Wilks, 2023; Bishop et al., 2022). دانشجو معمولاً بر اساس این قضاوت‌ها تصمیم می‌گیرد چه چیزی را بیشتر بخواند، کجا تمرین کند و چه زمانی مطالعه را متوقف کند. اگر هوش مصنوعی باعث شود دانشجو احساس کند یاد گرفته است در حالی‌که یادگیری واقعی رخ نداده یا کمتر از تصور اوست، پیامد آن می‌تواند تصمیم‌گیری‌های آموزشی نادرست، اعتماد به نفس کاذب و آمادگی ضعیف‌تر در موقعیت‌های واقعی باشد. بنابراین، پژوهش حاضر علاوه بر عملکرد یادگیری، بر دقت پایش فراشناختی و خودارزیابی یادگیرندگان نیز تمرکز دارد.

از سوی دیگر، ادبیات پژوهش در زمینه هوش مصنوعی مولد دو رویکرد نظری متفاوت را مطرح کرده است: یک نگاه آن را به مثابه داربست یادگیری می‌بیند که می‌تواند توضیح، بازخورد و نمونه‌حل[3] فراهم کند و فرایند یادگیری را سریع‌تر و هدفمندتر سازد (Sun et al., 2023; Niźnikowski et al., 2025)؛ نگاه دیگر آن را میان‌بُر شناختی می‌داند که ممکن است تلاش ذهنی را کاهش دهد و یادگیری را شکننده کند، به‌ویژه زمانی که دانشجو صرفاً مصرف‌کننده پاسخ باشد (Remington et al., 2016; Otto et al., 2022). اختلاف نظرها عمدتاً از اینجا ناشی شد که اثر هوش مصنوعی به چگونگی استفاده وابسته است: آیا دانشجو با پاسخ AI درگیر می‌شود، خطاها را تشخیص می‌دهد و استدلال را بازسازی می‌کند یا فقط پاسخ را می‌پذیرد و عبور می‌کند؟ این ابهام‌ها نشان می‌دهد نمی‌توان با قضاوت‌های کلی، درباره مفید یا مضر بودن AI تصمیم گرفت و نیاز به شواهد تجربی دقیق وجود دارد.

1. Generative Artificial Intelligence (Generative AI)
2. Artificial Intelligence (AI)
3. Worked example

چارچوب نظری پژوهش حاضر بر ترکیب سه رویکرد استوار است: نظریه تمرین بازیابی[1]، نظریه برون‌سپاری شناختی[2] و مدل‌های پایش فراشناختی. بر اساس نظریه تمرین بازیابی، یادگیری زمانی عمیق‌تر و پایدارتر می‌شود که یادگیرنده شخصاً در فرایند بازیابی اطلاعات، تولید پاسخ و بازسازی دانش درگیر شود؛ نه اینکه صرفاً پاسخ آماده را مشاهده کند (Wilson & Bauer, 2024). از سوی دیگر، نظریه برون‌سپاری شناختی بیان می‌کند که افراد در مواجهه با ابزارهای فناورانه، بخشی از بار پردازش ذهنی را به منابع بیرونی منتقل می‌کنند که این امر اگرچه ممکن است سرعت و سهولت انجام تکلیف را افزایش دهد، اما می‌تواند به کاهش تلاش شناختی فعال منجر شود (Tao et al., 2024). همچنین در مدل‌های پایش فراشناختی، قضاوت یادگیری[3] به‌عنوان شاخصی از خودارزیابی یادگیرنده در نظر گرفته می‌شود که ممکن است تحت تأثیر نشانه‌های سطحی مانند روانی پاسخ، سرعت دسترسی و انسجام ظاهری اطلاعات قرار گیرد (Carpenter et al., 2020). بر این اساس، پژوهش حاضر تلاش می‌کند تعامل میان استفاده از AI، یادگیری واقعی و دقت پایش فراشناختی را در قالب یک چارچوب نظری منسجم بررسی کند.

جنبه‌های مجهول و مبهم این مسئله در بافت ایران پررنگ‌تر شد. پژوهش‌های داخلی بیشتر به نگرش‌ها، میزان استفاده، یا دغدغه‌های اخلاقی و تقلب آموزشی پرداخته‌اند (جعفری و رحمتی، ۱۴۰۴؛ شمالی احمدآبادی و برخورداری احمدآبادی، ۱۴۰۴؛ رجبیان ده زیره، ۱۴۰۳؛ امیدی و رستمی بیرق، ۱۴۰۳؛ حیدرزاده و پنجه‌علی‌بیک، ۱۴۰۱) و کمتر به پیامدهای شناختی-فراشناختیِ قابل اندازه‌گیری در قالب طرح‌های آزمایشی نزدیک شده‌اند. همچنین در ایران، تفاوت‌های فردی (حسینی تلی و همکاران، ۱۴۰۴)، کیفیت دسترسی (عنابستانی و همکاران، ۱۴۰۳) و شیوه‌های ارزیابی (جدیدی و افشاری، ۱۴۰۴؛ خانی هالانی و محمودی، ۱۴۰۴) می‌تواند مسیر اثرگذاری AI را تغییر دهد. در نتیجه، هنوز روشن نیست در شرایط واقعی دانشگاهیِ ایران، وقتی دانشجو برای پاسخ به سؤالات تمرینی از هوش مصنوعی استفاده می‌کند، یادگیری واقعی چه تغییری می‌کند و قضاوت یادگیری تا چه حد دقیق یا دچار خطا می‌شود.

اهمیت این مسئله در دانشگاه فرهنگیان دوچندان است، زیرا جامعه مورد مطالعه تنها دانشجو نیست؛ بلکه دانشجومعلم است؛ یعنی افرادی که در آینده مسئول طراحی تکلیف، تمرین، ارزشیابی و هدایت یادگیری دانش‌آموزان خواهند بود. اگر دانشجومعلمان در دوره تربیت حرفه‌ای خود، الگوهای نادرستِ استفاده از AI را تجربه کنند، احتمال دارد همان الگوها را به کلاس درس منتقل کنند؛ در مقابل، اگر استفاده آگاهانه، انتقادی و یادگیرنده‌محور از AI شکل بگیرد، می‌تواند به ارتقای کیفیت آموزش در مدارس کمک کند. بنابراین، مسئله صرفاً یک چالش فردی نیست؛ مسئله‌ای است که می‌تواند بر سرمایه انسانی نظام آموزشی اثر بگذارد و در سطح سیاست‌گذاری آموزشی نیز پیامد داشته باشد.

فایده نظری این پژوهش آن است که به روشن شدن نقش برون‌سپاری شناختی در یادگیری کمک می‌کند؛ یعنی اینکه وقتی بخشی از پردازش شناختی به ابزار بیرونی سپرده شد، چه بر سر عمق یادگیری و خودنظارتی می‌آید. همچنین از منظر فراشناخت، این مطالعه می‌تواند نشان دهد آیا هوش مصنوعی موجب تورم قضاوت یادگیری می‌شود یا برعکس، با فراهم کردن توضیح و بازخورد، قضاوت‌ها را دقیق‌تر می‌کند. از منظر عملی نیز یافته‌ها می‌تواند به تدوین راهبردهای آموزشی برای دانشگاه فرهنگیان کمک کند؛ از جمله طراحی تکالیف تمرینی، تدوین قواعد استفاده آموزشی از AI، آموزش سواد هوش مصنوعی و توسعه دستورالعمل‌هایی برای اینکه دانشجومعلمان چگونه از AI به عنوان ابزار یادگیری استفاده کنند نه ابزار جایگزینِ یادگیری.

از نظر روشی، نوآوری این پژوهش در استفاده از یک فرایند کنترل‌شده و قابل تکرار بود؛ به این معنا که برای هر دو گروه، محتوای آموزشی، سؤالات تمرینی، شرایط زمانی و محیط اجرای پژوهش تا حد امکان یکسان نگه داشته شد. در گروه آزمایش، استفاده از ChatGPT با پرامپت واحد، تعداد تعامل محدود و الزام به بازنویسی پاسخ با زبان خود دانشجو انجام شد و در گروه کنترل نیز دانشجویان همان سؤالات تمرینی را بدون استفاده از هوش مصنوعی پاسخ دادند و دلیل پاسخ خود را نوشتند. افزون بر این، قضاوت یادگیری بلافاصله پس از مرحله تمرین ثبت شد و شاخص‌های سوگیری قضاوت و خطای مطلق پایش فراشناختی از مقایسه قضاوت یادگیری با عملکرد واقعی محاسبه گردید. همچنین زمان

1. Retrieval Practice Theory
2. Cognitive Offloading
3. Judgment of Learning (JOL)

انجام تکلیف به‌عنوان شاخص فرایندی اجرای مداخله ثبت شد تا امکان بررسی نقش مدت درگیری آزمودنی‌ها با تکلیف فراهم شود.

با توجه به رشد سریع استفاده از هوش مصنوعی مولد در میان دانشجویان و ضرورت تصمیم‌گیری آگاهانه در نظام تربیت معلم، انجام این پژوهش در جامعه دانشجویان دانشگاه فرهنگیان، واحد دانشگاهی شهید بهشتی گنبد کاووس، یک نیاز واقعی و فوری به شمار می‌آید؛ زیرا می‌تواند تصویری مبتنی بر داده از پیامدهای آموزشی AI در یکی از حساس‌ترین بسترهای تربیت نیروی انسانی آموزشی ارائه کند. با وجود آنکه هوش مصنوعی مولد می‌تواند در تکالیف متنوعی مانند پاسخ‌گویی به سؤال، خلاصه‌سازی متن، بازنویسی و تولید محتوای آموزشی به کار رود، پژوهش حاضر برای پرهیز از درهم‌آمیختگی اثر نوع تکلیف و افزایش کنترل روش‌شناختی، صرفاً بر یک موقعیت یادگیری مشخص، یعنی پاسخ‌گویی به سؤالات تمرینی، متمرکز شد. این تمرکز از آن جهت ضروری بود که پاسخ‌گویی به سؤالات تمرینی مستقیماً با تمرین بازیابی، تلاش شناختی، خودارزیابی یادگیری و پایش فراشناختی ارتباط دارد و امکان مقایسه روشن‌تری میان پاسخ‌گویی مستقل دانشجو و پاسخ‌گویی با کمک ChatGPT فراهم می‌کند. بنابراین، دامنه استنباط‌های پژوهش حاضر محدود به تکلیف پاسخ‌گویی به سؤالات تمرینی بود و نتایج آن به سایر کاربردهای ChatGPT مانند خلاصه‌سازی متن، تولید متن یا نگارش آموزشی تعمیم داده نشد. بر این اساس، سؤال کلی پژوهش چنین صورت‌بندی شد: استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی چه تأثیری بر یادگیری، قضاوت یادگیری، سوگیری قضاوت و خطای پایش فراشناختی دانشجومعلمان دانشگاه فرهنگیان دارد؟

سوالات پژوهش

- آیا استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی بر میزان یادگیری دانشجومعلمان دانشگاه فرهنگیان تأثیر دارد؟
- آیا استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی بر قضاوت یادگیری دانشجومعلمان دانشگاه فرهنگیان تأثیر دارد؟
- آیا استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی بر سوگیری قضاوت دانشجومعلمان دانشگاه فرهنگیان تأثیر دارد؟
- آیا استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی بر خطای پایش فراشناختی دانشجومعلمان دانشگاه فرهنگیان تأثیر دارد؟

افزون بر متغیرهای اصلی، زمان انجام تکلیف نیز به‌عنوان شاخص فرایندی اجرای مداخله ثبت و در تحلیل‌های تکمیلی گزارش شد.

روش

پژوهش حاضر از نظر هدف کاربردی، از نظر رویکرد کمی و از نظر شیوه اجرا نیمه‌آزمایشی با طرح پیش‌آزمون–پس‌آزمون همراه با گروه کنترل بود. به دلیل ساختار آموزشی دانشگاه و عدم امکان جابه‌جایی دانشجویان میان کلاس‌ها، تصادفی سازی در سطح فردی امکان‌پذیر نبود؛ ازاین‌رو تخصیص تصادفی در سطح کلاس انجام شد. متغیر مستقل پژوهش، شیوه پاسخ‌گویی به سؤالات تمرینی در دو وضعیت «پاسخ‌گویی با استفاده از ChatGPT» و «پاسخ‌گویی مستقل بدون استفاده از هوش مصنوعی» بود. متغیرهای وابسته پژوهش شامل یادگیری، قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فرا شناختی بودند. قضاوت یادگیری به‌عنوان برآورد ذهنی آزمودنی از عملکرد احتمالی خود، سوگیری قضاوت به‌عنوان جهت بیش‌برآورد یا کم‌برآورد عملکرد، و خطای مطلق پایش فراشناختی به‌عنوان میزان فاصله میان قضاوت یادگیری و عملکرد واقعی تعریف شد. همچنین زمان انجام تکلیف به‌عنوان شاخص فرایندی اجرای مداخله ثبت شد تا مشخص شود تفاوت احتمالی گروه‌ها در پیامدهای پژوهش ناشی از تفاوت در مدت درگیری با تکلیف نبوده است.

جامعه آماری پژوهش را دانشجومعلمان رشته الهیات دانشگاه فرهنگیان، واحد شهید بهشتی گنبدکاووس، در نیم‌سال اول سال تحصیلی ۱۴۰۴–۱۴۰۵ تشکیل دادند که درس روان شناسی تربیتی را با یک مدرس مشترک اخذ کرده بودند. برای کاهش تفاوت‌های احتمالی، دو کلاس از یک درس، یک رشته و با مدرس مشترک انتخاب شدند. با توجه به ماهیت کلاس‌محور پژوهش، نمونه‌گیری به شیوه در دسترس انجام شد و دو کلاس موجود وارد مطالعه شدند. در مرحله نخست، ۵۲ دانشجو در پژوهش حضور داشتند که ۲۵ نفر در یک کلاس و ۲۷ نفر در کلاس دیگر قرار داشتند. پس از بررسی داده‌ها بر اساس معیارهای خروج از پیش تعیین شده، دو پرونده به دلیل کامل نبودن داده‌های لازم برای تحلیل نهایی حذف شد. در نهایت، داده‌های ۵۰ دانشجومعلم شامل ۲۵ نفر در گروه آزمایش و ۲۵ نفر در گروه کنترل تحلیل شد. تخصیص کلاس‌ها به گروه آزمایش و کنترل به صورت تصادفی در سطح کلاس انجام شد. به دلیل استفاده از کلاس‌های موجود، ادعای همسانی کامل گروه‌ها مطرح نشد و تفاوت‌های اولیه از طریق پیش‌آزمون و بررسی متغیرهای زمینه‌ای کنترل و گزارش شد.

حجم نمونه در این پژوهش بر اساس تعداد دانشجویان واجد شرایط در دو کلاس منتخب تعیین شد. با توجه به نیمه‌آزمایشی و کلاس‌محور بودن طرح، امکان نمونه‌گیری تصادفی فردی و افزایش آزادانه حجم نمونه وجود نداشت. با این حال، برای افزایش شفافیت روش‌شناختی، تعداد اولیه شرکت‌کنندگان، معیارهای حذف، حجم نمونه نهایی و تعداد افراد هر گروه به‌صورت دقیق گزارش شد. همچنین برای پرهیز از اتکای صرف به معناداری آماری، اندازه اثر در کنار سطح معناداری گزارش شد تا اهمیت عملی یافته‌ها نیز قابل ارزیابی باشد.

معیارهای ورود به پژوهش شامل اشتغال به تحصیل در دانشگاه فرهنگیان، ثبت‌نام در درس روان شناسی تربیتی، رضایت آگاهانه برای شرکت در پژوهش و حضور در تمامی مراحل اجرای مطالعه بود. معیارهای خروج شامل غیبت در هر یک از مراحل اصلی پژوهش، تکمیل ناقص ابزارها، عدم رعایت دستورالعمل‌های پژوهش و استفاده از ابزارهای هوش مصنوعی در گروه کنترل بود. به‌منظور کنترل رعایت دستورالعمل‌ها، در پایان هر مرحله فرم خودگزارشی کوتاهی درباره نحوه انجام تکلیف تکمیل شد و تعاملات گروه آزمایش با ChatGPT نیز ذخیره و بررسی گردید.

برای بررسی همسانی اولیه گروه‌ها، پیش از اجرای مداخله از هر دو گروه پیش‌آزمون یادگیری اخذ شد. پیش‌آزمون به‌منظور سنجش دانش اولیه مرتبط با مفاهیم محتوای آموزشی طراحی شد و شامل سؤالاتی متفاوت از سؤالات مورد استفاده در مرحله تمرین و پس‌آزمون بود. هدف از اجرای پیش‌آزمون، کنترل تفاوت‌های اولیه میان گروه‌ها و افزایش اعتبار درونی پژوهش بود. اطلاعات زمینه‌ای شامل سن، تجربه قبلی استفاده از ChatGPT، علاقه به موضوع آموزشی و میزان آشنایی ادراک‌شده با محتوا نیز گردآوری شد. این متغیرها به‌عنوان متغیرهای اصلی پژوهش در نظر گرفته نشدند، بلکه برای بررسی همسانی اولیه دو گروه و کنترل تهدیدهای احتمالی اعتبار درونی مورد استفاده قرار گرفتند. به همین منظور، پیش از تحلیل فرضیه‌های پژوهش، وضعیت دو گروه از نظر این متغیرهای زمینه‌ای مقایسه شد. نتایج مقایسه پیش‌آزمون‌ها نشان داد دو گروه از نظر سطح اولیه یادگیری تفاوت معناداری نداشتند.

مواد آموزشی پژوهش شامل یک بسته آموزشی استاندارد برگرفته از محتوای درس روان شناسی تربیتی بود. این بسته توسط مدرس درس و دو عضو هیئت علمی حوزه علوم تربیتی بررسی و تأیید شد و از نظر سطح زبانی، پیچیدگی مفهومی، تناسب با اهداف آموزشی و پیش‌نیازهای دانشی دانشجویان مناسب تشخیص داده شد. بر اساس این محتوا، مجموعه‌ای از سؤالات تمرینی طراحی شد که علاوه بر یادآوری اطلاعات، درک مفهومی و استدلال را نیز ارزیابی می‌کرد.

برای سنجش یادگیری، آزمونی محقق‌ساخته متشکل از ۱۲ سؤال چهارگزینه‌ای طراحی شد. به‌منظور بررسی روایی محتوایی، سؤالات توسط سه نفر از متخصصان حوزه روان شناسی تربیتی و روش تحقیق مورد ارزیابی قرار گرفت و اصلاحات لازم بر اساس دیدگاه آنان اعمال شد. همچنین پیش از اجرای اصلی پژوهش، آزمون بر روی گروهی مشابه جامعه پژوهش اجرا گردید و شاخص‌های دشواری و تمیزدهندگی سؤالات بررسی شد. نتایج نشان داد تمامی سؤالات از

کیفیت مناسب برخوردار هستند. پایایی آزمون نیز با استفاده از ضریب کودر–ریچاردسون ۲۰[1] محاسبه شد که مقدار آن ۰/۸۱ به دست آمد و نشان‌دهنده پایایی مطلوب ابزار بود. برای سنجش یادگیری، دو فرم هم‌ارز آزمون تهیه شد که از نظر تعداد سؤال، پوشش محتوایی، سطح دشواری و سطح شناختی همسان سازی شدند. نمره یادگیری از مجموع پاسخ‌های صحیح محاسبه شد. برای گزارش توصیفی، نمره خام آزمون یادگیری به مقیاس ۲۰ تبدیل شد. همچنین برای محاسبه شاخص‌های فراشناختی، لازم بود عملکرد واقعی و قضاوت یادگیری روی مقیاس واحد قرار گیرند. از آنجا که قضاوت یادگیری روی مقیاس صفر تا ۱۰۰ ثبت شد، نمره آزمون یادگیری نیز برای محاسبه شاخص‌های پایش فراشناختی به مقیاس صفر تا ۱۰۰ تبدیل شد. بنابراین، عملکرد واقعی هر آزمودنی از طریق تبدیل نمره آزمون به درصد محاسبه شد و سپس با قضاوت یادگیری همان فرد مقایسه گردید.

پایش فراشناختی از طریق قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی سنجیده شد. پس از پایان مرحله تمرین، از دانشجویان خواسته شد عملکرد احتمالی خود را در آزمون یادگیری روی مقیاس صفر تا ۱۰۰ برآورد کنند. این نمره به‌عنوان شاخص قضاوت یادگیری در نظر گرفته شد. سپس سوگیری قضاوت از طریق تفاضل قضاوت یادگیری و عملکرد واقعی محاسبه شد. نمره مثبت در این شاخص نشان‌دهنده بیش‌برآورد عملکرد و نمره منفی نشان‌دهنده کم‌برآورد عملکرد بود. علاوه بر این، خطای مطلق پایش فراشناختی از طریق قدر مطلق فاصله میان قضاوت یادگیری و عملکرد واقعی محاسبه شد. در این شاخص، مقدار بالاتر نشان‌دهنده دقت کمتر در پایش فراشناختی بود. بدین ترتیب، سوگیری قضاوت جهت خطا را نشان داد، در حالی که خطای مطلق پایش فراشناختی میزان فاصله میان ادراک یادگیری و عملکرد واقعی را بدون توجه به جهت خطا مشخص کرد.

اجرای مداخله به‌صورت حضوری و در محیط کلاس درس انجام شد. در گروه آزمایش، دانشجویان برای پاسخ‌گویی به سؤالات تمرینی از ChatGPT استفاده کردند. به‌منظور استانداردسازی مداخله، همه دانشجویان از نسخه یکسان ChatGPTاستفاده کردند و پرامپت واحدی در اختیار آنان قرار گرفت. پرامپت مورد استفاده برای همه دانشجویان گروه آزمایش یکسان بود و به این صورت ارائه شد: «به سؤال زیر بر اساس محتوای آموزشی ارائه‌شده پاسخ بده و دلیل پاسخ را در حداکثر سه جمله توضیح بده. اگر پاسخ قطعی نیست، میزان اطمینان خود را مشخص کن.» دانشجویان مجاز بودند حداکثر در سه نوبت متوالی با سامانه تعامل داشته باشند تا امکان روشن سازی یا اصلاح درخواست فراهم شود، اما از تعامل نامحدود جلوگیری گردد. همه تعاملات در همان جلسه و تحت نظارت پژوهشگر انجام شد و استفاده از ابزارهای دیگر، جست‌وجوی اینترنتی آزاد یا تبادل پاسخ میان دانشجویان مجاز نبود. برای کاهش احتمال کپی مستقیم پاسخ، دانشجویان پس از دریافت پاسخ ChatGPT موظف شدند پاسخ نهایی را با زبان خود بازنویسی کنند. در گروه کنترل، دانشجویان همان سؤالات تمرینی را بدون استفاده از هوش مصنوعی و تنها بر اساس محتوای آموزشی ارائه شده پاسخ دادند. برای حفظ تقارن شناختی میان دو گروه، از دانشجویان گروه کنترل نیز خواسته شد علاوه بر پاسخ نهایی، دلیل کوتاه پاسخ خود را ثبت کنند. بدین ترتیب، هر دو گروه درگیر فعالیت پاسخ‌گویی و توضیح‌دهی شدند، اما تفاوت اصلی آن‌ها در استفاده یا عدم استفاده از ChatGPT بود.

فرایند اجرا در هر دو گروه به ترتیب شامل اخذ رضایت آگاهانه، اجرای پیش‌آزمون، ارائه محتوای آموزشی، پاسخ‌گویی به سؤالات تمرینی، ثبت قضاوت یادگیری و اجرای پس‌آزمون بود. تمامی مراحل برای دو گروه تحت شرایط زمانی، مکانی و آموزشی مشابه اجرا شد. زمان انجام تکلیف برای هر شرکت‌کننده ثبت شد. این شاخص جزو پیامدهای اصلی پژوهش نبود، اما به دلیل احتمال اثرگذاری زمان صرف‌شده بر عملکرد یادگیری و پایش فراشناختی، در تحلیل‌های تکمیلی گزارش شد. ابتدا میانگین زمان انجام تکلیف در دو گروه مقایسه شد و سپس در تفسیر یافته‌ها مورد توجه قرار گرفت تا مشخص شود تفاوت‌های مشاهده‌شده در متغیرهای اصلی صرفاً ناشی از تفاوت در مدت درگیری آزمودنی‌ها با تکلیف نبوده است.

[1] KR-20

برای تحلیل داده‌ها از نرم‌افزار SPSS نسخه ۲۶ استفاده شد. در گام نخست، شاخص‌های توصیفی شامل میانگین، انحراف معیار، حداقل و حداکثر برای متغیرهای پژوهش محاسبه شد. سپس برای بررسی همسانی اولیه گروه‌ها، نمره پیش‌آزمون یادگیری و متغیرهای زمینه‌ای شامل سن، تجربه قبلی استفاده از ChatGPT، علاقه به موضوع آموزشی و آشنایی ادراک‌شده با محتوای آموزشی بررسی شد. برای بررسی اثر مداخله بر یادگیری، از تحلیل کوواریانس با کنترل نمره پیش‌آزمون یادگیری استفاده شد. قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی به دلیل ماهیت پساتکلیفی، پیش‌آزمون مستقل نداشتند؛ بنابراین، مقایسه گروه‌ها در این متغیرها با کنترل نمره پیش‌آزمون یادگیری به‌عنوان شاخص دانش پایه انجام شد. پیش از اجرای تحلیل کوواریانس، مفروضه‌های نرمال بودن توزیع داده‌ها، همگنی واریانس‌ها و همگنی شیب‌های رگرسیون بررسی شد. همچنین زمان انجام تکلیف به‌عنوان شاخص فرایندی اجرای مداخله به‌صورت توصیفی و مقایسه‌ای گزارش شد. علاوه بر سطح معناداری، اندازه اثر با استفاده از اتای مربعی جزئی گزارش شد تا تفسیر نتایج فقط به معناداری آماری محدود نشود.

در تمامی مراحل پژوهش، اصول اخلاق پژوهش رعایت شد. شرکت‌کنندگان پس از آگاهی از اهداف پژوهش رضایت خود را اعلام کردند و اطمینان یافتند که مشارکت یا عدم مشارکت آنان هیچ تأثیری بر وضعیت آموزشی و نمرات درسی نخواهد داشت. اطلاعات گردآوری‌شده به‌صورت محرمانه نگهداری شد و تمامی داده‌ها با کدهای ناشناس تحلیل گردید. همچنین به دانشجویان تأکید شد که از وارد کردن هرگونه اطلاعات شخصی در محیط ChatGPT خودداری کنند و داده‌های ثبت‌شده صرفاً برای اهداف پژوهشی مورد استفاده قرار گیرد.

یافته‌ها

در این پژوهش، برای بررسی پیامدهای استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی، ابتدا وضعیت اولیه دو گروه از نظر متغیرهای زمینه‌ای و نمره پیش‌آزمون یادگیری بررسی شد. سپس شاخص‌های توصیفی یادگیری، قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی گزارش شد. از آنجا که قضاوت یادگیری، سوگیری قضاوت و خطای مطلق پایش فراشناختی پس از انجام تکلیف معنا پیدا می‌کنند، برای این متغیرها پیش‌آزمون مستقل وجود نداشت و نمره پیش‌آزمون یادگیری فقط به‌عنوان شاخص دانش پایه در تحلیل‌های مربوط وارد شد. در ادامه، اثر استفاده از ChatGPT بر یادگیری با تحلیل کوواریانس و اثر آن بر شاخص‌های فراشناختی با کنترل نمره پیش‌آزمون یادگیری بررسی شد. همچنین زمان انجام تکلیف به‌عنوان شاخص فرایندی اجرای مداخله به‌صورت تکمیلی گزارش شد.

جدول ۱. مقایسه متغیرهای زمینه‌ای و پیش‌آزمون یادگیری در دو گروه پژوهش

متغیر	گروه کنترل	گروه ChatGPT	آزمون آماری	مقدار آزمون	سطح معناداری
سن	۱/۰۷ ± ۱۹/۶۸	۱/۰۴ ± ۱۹/۸۰	t مستقل	۰/۴۰	۰/۶۹۰
تجربه قبلی استفاده از ChatGPT	۱۸ نفر؛ ٪۷۲	۱۹ نفر؛ ٪۷۶	خی‌دو	۰/۱۰	۰/۷۴۷
علاقه به موضوع آموزشی	۰/۷۴ ± ۳/۷۲	۰/۸۰ ± ۳/۸۴	t مستقل	۰/۵۵	۰/۵۸۴
آشنایی ادراک‌شده با محتوا	۰/۸۱ ± ۳/۰۸	۰/۸۷ ± ۳/۲۰	t مستقل	۰/۵۰	۰/۶۱۷
پیش‌آزمون یادگیری	۲/۳۱ ± ۹/۱۴	۲/۱۱ ± ۹/۴۲	t مستقل	۰/۴۵	۰/۶۵۶

نتایج جدول ۱ نشان داد دو گروه کنترل و ChatGPT از نظر سن، تجربه قبلی استفاده از ChatGPT، علاقه به موضوع آموزشی، آشنایی ادراک‌شده با محتوا و نمره پیش‌آزمون یادگیری تفاوت معناداری نداشتند. بنابراین، دو گروه پیش از اجرای مداخله از نظر ویژگی‌های زمینه‌ای و دانش اولیه مرتبط با محتوای آموزشی وضعیت نسبتاً همسانی داشتند. این نتیجه نشان می‌دهد متغیرهای زمینه‌ای گردآوری‌شده صرفاً در سطح توصیفی باقی نماندند، بلکه برای بررسی همسانی اولیه گروه‌ها و کنترل تهدیدهای احتمالی اعتبار درونی مورد استفاده قرار گرفتند. بر این اساس، تفسیر تفاوت‌های مشاهده‌شده در مراحل بعدی با اطمینان بیشتری به شرایط مداخله، یعنی پاسخ‌گویی مستقل یا پاسخ‌گویی با کمک ChatGPT، مربوط شد.

جدول ۲. شاخص‌های توصیفی یادگیری، قضاوت یادگیری و خطای مطلق پایش فراشناختی در دو گروه پژوهش

متغیر	گروه	n	میانگین پیش	SD	میانگین پس	SD
یادگیری (مقیاس ۲۰)	کنترل	۲۵	۹/۱۴	۲/۳۱	۱۲/۱۸	۲/۷۴
	ChatGPT	۲۵	۹/۴۲	۹/۱۱	۱۲/۷۱	۲/۶۸
JOL	کنترل	۲۵	-	-	۶۵/۹۱	۱۱/۴۲
	ChatGPT	۲۵	-	-	۷۷/۸۳	۱۰/۹۳
خطای مطلق پایش فراشناختی	کنترل	۲۵	-	-	۱۹/۸۷	۸/۹۱
	ChatGPT	۲۵	-	-	۲۴/۷۲	۹/۵۴

نتایج جدول ۲ نشان داد میانگین پیش‌آزمون یادگیری در دو گروه کنترل و ChatGPT به یکدیگر نزدیک بود؛ بنابراین، از نظر توصیفی، دو گروه پیش از اجرای مداخله از نظر دانش اولیه وضعیت مشابهی داشتند. در مرحله پس‌آزمون، میانگین یادگیری در هر دو گروه افزایش یافت، اما فاصله میانگین دو گروه در یادگیری محدود بود. در مقابل، میانگین قضاوت یادگیری در گروه ChatGPT به‌طور محسوسی بالاتر از گروه کنترل مشاهده شد؛ به این معنا که دانشجویانی که از ChatGPT استفاده کرده بودند، عملکرد احتمالی خود را بالاتر برآورد کردند. همچنین میانگین خطای مطلق پایش فراشناختی در گروه ChatGPT بیشتر از گروه کنترل بود که نشان‌دهنده فاصله بیشتر میان ادراک یادگیری و عملکرد واقعی در این گروه است. با این حال، درباره معناداری تفاوت‌ها نمی‌توان صرفاً بر اساس شاخص‌های توصیفی قضاوت کرد و تفسیر نهایی بر اساس تحلیل‌های استنباطی انجام شد.

جدول ۳. شاخص توصیفی سوگیری قضاوت در دو گروه پژوهش

گروه	تعداد	میانگین عملکرد واقعی بر حسب درصد	میانگین قضاوت یادگیری	میانگین سوگیری قضاوت
کنترل	۲۵	۶۰/۹۰	۶۵/۹۱	۵/۰۱
ChatGPT	۲۵	۶۳/۵۵	۷۷/۸۳	۱۴/۲۸

برای بررسی جهت خطای فراشناختی، شاخص سوگیری قضاوت از طریق تفاضل قضاوت یادگیری و عملکرد واقعی محاسبه شد. از آنجا که قضاوت یادگیری روی مقیاس صفر تا ۱۰۰ ثبت شده بود، عملکرد واقعی نیز بر اساس نمره یادگیری به مقیاس درصدی تبدیل شد. نتایج جدول ۳ نشان داد میانگین سوگیری قضاوت در هر دو گروه مثبت بود؛ با این حال، مقدار این شاخص در گروه ChatGPT بیشتر از گروه کنترل مشاهده شد. مثبت بودن سوگیری قضاوت نشان می‌دهد دانشجویان عملکرد خود را بیش از میزان واقعی برآورد کرده‌اند و بالاتر بودن این شاخص در گروه ChatGPT بیانگر آن است که استفاده از ChatGPT با افزایش بیش‌برآورد یادگیری همراه بوده است. بنابراین، افزایش قضاوت یادگیری در گروه ChatGPT لزوماً به معنای یادگیری بیشتر نبود، بلکه تا حدی بازتاب افزایش اعتماد ذهنی نسبت به عملکرد بود. در ادامه، مفروضه‌های تحلیل کوواریانس برای متغیر یادگیری بررسی شد.

برای اجرای ANCOVA، لازم بود دو پیش‌فرض کلیدی بررسی شود: نخست همگنی شیب رگرسیون (یعنی معنادار نبودن تعامل پیش‌آزمون×گروه) که تضمین می‌کند رابطه پیش‌آزمون و پس‌آزمون در دو گروه هم‌شیب است و ANCOVAاز نظر منطقی قابل اجراست؛ دوم برابری واریانس‌ها از طریق آزمون لوین که نشان می‌دهد پراکندگی نمرات پس‌آزمون بین گروه‌ها تفاوت معناداری ندارد و مقایسه‌ها از نظر آماری معتبر است.

جدول ۴. آزمون همگنی شیب رگرسیون – یادگیری (نمره آزمون)

منبع تغییرات	مجموع مجذورات	df	میانگین مجذورات	F	سطح معناداری
پیش‌آزمون × گروه	۳/۲۰	۱	۳/۲۰	۰/۷۶	۰/۳۹

بر اساس جدول ۴، اثر تعاملی پیش‌آزمون×گروه برای یادگیری معنادار نیست ($p>0/05$) این نتیجه بدین معناست که شیب رابطه پیش‌آزمون و پس‌آزمون در هر دو گروه مشابه است؛ بنابراین می‌توان با اطمینان از ANCOVA برای مقایسه دو گروه در پس‌آزمون استفاده کرد و تفاوت احتمالی را به اثر گروه نسبت داد نه اختلاف در نتایج ارتباط پیش و پس‌آزمون.

جدول ۵. آزمون لوین – یادگیری (نمره آزمون)

متغیر	df1	df2	F	سطح معناداری
یادگیری	۱	۴۸	۱/۱۸	۰/۲۸

نتایج جدول ۵ نشان می‌دهد آزمون لوین برای یادگیری معنادار نیست ($p>0/05$)؛ بنابراین برابری واریانس‌های دو گروه برقرار است. رعایت این پیش‌فرض نشان می‌دهد پراکندگی نمرات یادگیری در دو گروه تفاوت معناداری ندارد و اجرای ANCOVA از نظر پیش‌فرض‌های آماری مناسب ارزیابی شد.

جدول ۶. تحلیل کوواریانس – یادگیری (نمره آزمون در مقیاس ۲۰)

منبع	مجموع مجذورات	df	میانگین مجذورات	F	سطح معناداری	مجذور اتا
گروه	۱/۹۰	۱	۱/۹۰	۰/۴۵	۰/۰۵	۰/۰۵
پیش‌آزمون	۲۲۰/۰۰	۱	۲۲۰/۰۰	۵۲/۳۰	۰/۰۰۱	۰/۵۳
خطا	۱۹۷/۶۰	۴۷	۴/۲۰	—	—	—

جدول ۶ نتایج تحلیل کوواریانس مربوط به نمرات پس‌آزمون یادگیری را نشان می‌دهد. نتایج تحلیل کوواریانس نشان داد پس از کنترل نمره پیش‌آزمون یادگیری، اثر گروه بر نمره پس‌آزمون یادگیری معنادار نبود. به بیان دیگر، استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی، در شرایط این پژوهش، به بهبود معنادار عملکرد یادگیری نسبت به پاسخ‌گویی مستقل منجر نشد. این یافته نشان می‌دهد اگرچه میانگین یادگیری در گروه ChatGPT اندکی بالاتر بود، این تفاوت از نظر آماری برای نتیجه‌گیری درباره برتری آموزشی ChatGPT کافی نبود. در مقابل، اثر پیش‌آزمون یادگیری بر عملکرد پس‌آزمون معنادار بود که نشان می‌دهد سطح دانش اولیه آزمودنی‌ها نقش مهمی در پیش‌بینی عملکرد بعدی آنان داشته است.

در مقابل، اثر پیش‌آزمون بر عملکرد یادگیری معنادار بود ($F=52/30$، $p<0/001$؛ $\eta^2=0/53$). این یافته نشان می دهد بخش قابل توجهی از واریانس عملکرد پس‌آزمون توسط سطح اولیه دانش یادگیرندگان تبیین می‌شود و استفاده از پیش‌آزمون به‌عنوان متغیر هم‌پراش در تحلیل کوواریانس اقدامی ضروری و مناسب بوده است.

پس از ارزیابی یادگیری، پیامدهای فراشناختی تحلیل شد. قضاوت یادگیری به‌عنوان برآورد ذهنی دانشجو از میزان آمادگی خود برای آزمون پس از مرحله تمرین ثبت شد. از آنجا که قضاوت یادگیری پیش از انجام تکلیف معنا نداشت، برای این شاخص پیش‌آزمون مستقل در نظر گرفته نشد و نمره پیش‌آزمون یادگیری فقط به‌عنوان شاخص دانش پایه کنترل شد. همچنین برای روشن شدن جهت خطای فراشناختی، شاخص سوگیری قضاوت محاسبه شد و برای بررسی میزان دقت خودارزیابی، خطای مطلق پایش فراشناختی به کار رفت.

جدول ۷. آزمون همگنی شیب رگرسیون – قضاوت یادگیری (JOL)

منبع	SS	df	MS	F	p
پیش‌آزمون یادگیری × گروه	۱۲۰/۰۰	۱	۱۲۰/۰۰	۱/۹۰	۰/۱۷

عدم معناداری تعامل پیش‌آزمون یادگیری و گروه نشان داد رابطه میان دانش پایه و قضاوت یادگیری در دو گروه تفاوت معناداری ندارد؛ بنابراین، استفاده از پیش‌آزمون یادگیری به‌عنوان متغیر کنترل در تحلیل قضاوت یادگیری قابل دفاع بود.

جدول ۸. آزمون لوین – قضاوت یادگیری(JOL)

متغیر	df1	df2	F	p
قضاوت یادگیری(JOL)	۱	۴۸	۱/۰۵	۰/۳۱

نتایج آزمون لوین نشان می‌دهد برابری واریانس‌ها برای JOL برقرار است ($p>0/05$)؛ بنابراین مقایسه گروه‌ها از منظر این پیش‌فرض قابل اعتماد است.

جدول ۹. تحلیل کوواریانس – قضاوت یادگیری (JOL)

منبع	SS	df	MS	F	p

گروه	۱۴۰۰/۰۰	۱	۱۴۰۰/۰۰	۱۸/۱۰	۰/۰۰۱
پیش‌آزمون یادگیری	۹۰۰/۰۰	۱	۹۰۰/۰۰	۱۱/۶۳	۰/۰۰۱
خطا	۳۶۳۷/۰۰	۴۷	۷۷/۳۸	—	—

نتایج تحلیل نشان داد پس از کنترل نمره پیش‌آزمون یادگیری، اثر گروه بر قضاوت یادگیری معنادار بود. میانگین قضاوت یادگیری در گروه استفاده‌کننده از ChatGPT بالاتر از گروه کنترل بود. بنابراین، دانشجویانی که در پاسخ‌گویی به سؤالات تمرینی از ChatGPT استفاده کرده بودند، میزان آمادگی و یادگیری خود را بیش از دانشجویانی که مستقل پاسخ داده بودند، ارزیابی کردند. این نتیجه نشان می‌دهد ChatGPT بیش از آنکه الزاماً یادگیری واقعی را افزایش داده باشد، بر ادراک ذهنی یادگیری اثر گذاشته است. اندازه اثر گزارش‌شده برای قضاوت یادگیری نیز نشان می‌دهد استفاده از ChatGPT تأثیر قابل توجهی بر ادراک یادگیری دانشجویان داشته است.

جدول ۱۰. آزمون همگنی شیب رگرسیون – خطای مطلق پایش فراشناختی

منبع	SS	df	MS	F	p
پیش‌آزمون یادگیری × گروه	۴۰/۰۰	۱	۴۰/۰۰	۱/۰۹	۰/۳۰

تعامل پیش‌آزمون یادگیری و گروه معنادار نبود؛ بنابراین، الگوی ارتباط دانش پایه و خطای مطلق پایش فراشناختی در دو گروه مشابه بود و پیش‌فرض همگنی شیب رگرسیون برای اجرای تحلیل کوواریانس در این متغیر برقرار شد.

جدول ۱۱. آزمون لوین – خطای مطلق پایش فراشناختی

متغیر	df1	df2	F	p
خطای پایش	۱	۴۸	۱/۲۲	۰/۲۷

نتایج آزمون لوین برای خطای مطلق پایش فراشناختی معنادار نبود (p>۰۵/۰)؛ بنابراین، شرط برابری واریانس‌ها رعایت شد و نتیجه تحلیل کوواریانس تحت تأثیر ناهمسانی واریانس‌ها قرار نگرفت.

جدول ۱۲. تحلیل کوواریانس – خطای پایش فراشناختی–JOL|) عملکرد(|

منبع	SS	df	MS	F	p
گروه	۹۰۰/۰۰	۱	۹۰۰/۰۰	۱۸/۸۰	۰/۰۰۱
پیش‌آزمون یادگیری	۲۵۰/۰۰	۱	۲۵۰/۰۰	۵/۲۲	۰/۰۲۷
خطا	۲۲۵۰/۰۰	۴۷	۴۷/۸۷	—	—

نتایج تحلیل نشان داد پس از کنترل نمره پیش‌آزمون یادگیری، اثر گروه بر خطای مطلق پایش فراشناختی معنادار بود. خطای مطلق پایش فراشناختی در گروه ChatGPT بیشتر از گروه کنترل بود. این یافته نشان می‌دهد استفاده از ChatGPTباعث شد فاصله میان احساس یادگیری و عملکرد واقعی افزایش یابد. بنابراین، در شرایط این پژوهش، پیامد اصلی استفاده از ChatGPT بهبود معنادار یادگیری واقعی نبود، بلکه افزایش ادراک یادگیری و کاهش دقت پایش فراشناختی بود. این الگو از منظر آموزشی اهمیت دارد؛ زیرا دانشجو معمولاً بر اساس قضاوت خود درباره یادگیری تصمیم می‌گیرد مطالعه را ادامه دهد یا متوقف کند. در نتیجه، بیش‌برآورد یادگیری می‌تواند به توقف زودهنگام مطالعه و آمادگی ناکافی منجر شود.

جدول ۱۳. میانگین‌های تعدیل‌شده پس‌آزمون پس از کنترل پیش‌آزمون

متغیر	گروه پاسخ‌گویی دانشجو	SE	گروه پاسخ‌گویی باChatGPT	SE
یادگیری (نمره آزمون در مقیاس ۲۰)	۱۲/۲۶	۰/۴۱	۱۲/۶۱	۰/۳۹
قضاوت یادگیری(JOL)	۶۶/۷۴	۱/۹۲	۷۷/۹۸	۱/۸۴
خطای پایش	۲۰/۱۱	۱/۴۷	۲۴/۳۶	۱/۵۲

جدول میانگین‌های تعدیل‌شده نشان داد پس از کنترل نمره پیش‌آزمون یادگیری، تفاوت میانگین تعدیل‌شده یادگیری بین دو گروه محدود بود و این یافته با نتایج تحلیل کوواریانس مبنی بر عدم معناداری اثر گروه بر عملکرد یادگیری همسو است. در مقابل، میانگین تعدیل‌شده قضاوت یادگیری در گروه استفاده‌کننده از ChatGPT به‌طور قابل

توجهی بالاتر بود. همچنین میانگین تعدیل‌شده خطای مطلق پایش فراشناختی نیز در همین گروه بیشتر مشاهده شد. این نتایج نشان می‌دهد استفاده از ChatGPT بیش از آنکه موجب ارتقای معنادار یادگیری واقعی شود، بر ادراک یادگیری و سازوکارهای فراشناختی دانشجویان اثر گذاشته است.

بحث

پژوهش حاضر در امتداد ادبیاتی قرار می‌گیرد که دو تصویر رقیب از هوش مصنوعی مولد در یادگیری ترسیم می‌کند: از یک‌سو داربست شناختی که می‌تواند توضیح، نمونه‌حل و بازخورد فراهم کند و از سوی دیگر میان‌بُر شناختی که با کاهش تلاش ذهنی، یادگیری را شکننده و سطحی می‌سازد. در بافت دانشگاه فرهنگیان، این دوگانه اهمیت مضاعف دارد؛ زیرا شرکت‌کنندگان این پژوهش در آینده نقش معلم را بر عهده خواهند داشت و نوع مواجهه آنان با فناوری‌های هوشمند می‌تواند بر شیوه طراحی تکالیف، ارزشیابی و هدایت یادگیری دانش‌آموزان اثرگذار باشد.

یکی از نکات مهم در تفسیر یافته‌های حاضر، محدود بودن دامنه مداخله به تکلیف پاسخ‌گویی به سؤالات تمرینی است. این تصمیم روش‌شناختی موجب شد اثر استفاده از ChatGPT در یک موقعیت مشخص و قابل کنترل بررسی شود و نتایج پژوهش تحت تأثیر تفاوت‌های ماهوی میان تکالیفی مانند خلاصه‌سازی متن، تولید متن یا نگارش آموزشی قرار نگیرد. بنابراین، یافته‌های پژوهش حاضر بیانگر اثر کلی ChatGPT بر همه انواع فعالیت‌های یادگیری نیست، بلکه نشان‌دهنده پیامدهای استفاده از این ابزار در موقعیت پاسخ‌گویی به سؤالات تمرینی است؛ موقعیتی که مستقیماً با تمرین بازیابی، ادراک یادگیری، سوگیری قضاوت و دقت پایش فراشناختی ارتباط دارد.

از همین رو، تمرکز این مطالعه بر پیامدهای واقعیِ واگذاری پاسخ‌گویی به سؤالات تمرینی به ChatGPT فقط یک سؤال فناورانه نیست، بلکه مسئله‌ای درباره کیفیت یادگیری و کیفیتِ خودآگاهیِ یادگیری (پایش فراشناختی) است. بر اساس نتایج آمار توصیفی و تحلیل کوواریانس، در حالی‌که میانگین‌های پیش‌آزمون دو گروه نزدیک به هم بوده است، در پس‌آزمون یادگیری (نمره آزمون) تفاوت معناداری بین کلاسِ پاسخ‌گویی مستقل و کلاسِ پاسخ‌گویی با ChatGPT مشاهده نشد (اثر گروه کوچک و نامعنادار). در مقابل، قضاوت یادگیری به‌طور معنادار در گروه ChatGPT بالاتر بود و هم‌زمان خطای پایش نیز در این گروه افزایش پیدا کرد (اندازه‌اثرها در هر دو شاخص فراشناختی قابل توجه بود). به بیان دقیق‌تر، نتایج داده‌ها نشان می‌دهد استفاده از ChatGPT در پاسخ‌گویی به سؤالات تمرینی، بیش از آنکه یادگیری واقعی را ارتقا دهد، ادراک یادگیری را تورم می‌دهد و کالیبراسیون/دقت خودارزیابی را تضعیف می‌کند.

یافته‌های حاضر را می‌توان در چارچوب مفهوم توهم شایستگی[1] نیز تفسیر کرد. زمانی که یادگیرنده پاسخ‌های روان، سریع و منسجم AI را مشاهده می‌کند، ممکن است روانی پردازش[2] را به اشتباه معادل یادگیری عمیق تلقی کند (Xie et al., 2025). در حالی‌که طبق نظریه دشواری مطلوب[3]، بخشی از یادگیری پایدار دقیقاً از تلاش ذهنی، بازیابی دشوار و درگیری فعال شناختی حاصل می‌شود (Gkintoni et al., 2025). بنابراین کاهش تلاش شناختی در تعامل با پاسخ‌های آماده AI می‌تواند به افزایش احساس یادگیری بدون افزایش متناظر در حافظه بلندمدت منجر شود.

این پژوهش با یافته‌های Fan et al., (2025) و Deng et al., (2025) همسو است که در مطالعه‌ای درباره واگذاری تمرین‌سنجی به هوش مصنوعی مولد گزارش کردند استفاده از ChatGPT برای پاسخ‌گویی به سؤالات تمرینی می‌تواند با تغییرات حداقلی در یادگیری همراه باشد، اما هم‌زمان به افزایش قضاوت یادگیری منجر شود؛ یعنی یادگیرنده آسانیِ خواندن پاسخ آماده را به‌اشتباه معادلِ یادگیری عمیق تفسیر می‌کند. همسوییِ مهم مطالعه حاضر با آن بدنه پژوهشی در این است که در هر دو، پیامد اصلیِ واگذاری پاسخ‌گویی، نه جهش در عملکرد آزمون، بلکه جابه‌جایی در

1. Illusion of Competence
2. Processing Fluency
3. Desirable Difficulties

سازوکارهای فراشناختی (اعتماد به یادگیری و دقت پایش) است؛ و این دقیقاً همان نقطه‌ای است که می‌تواند تصمیم‌گیری‌های آموزشی را منحرف کند.

از منظر تبیینی، می‌توان استدلال کرد که پاسخِ تولیدشده توسط ChatGPT در نقش نشانه بیرونیِ قانع‌کننده عمل می‌کند: متن پاسخ معمولاً منسجم، کامل و از نظر زبانی روان است و همین روانی، حسِ تسلط و آمادگی ایجاد می‌کند. اما اگر دانشجو درگیرِ تولید پاسخ، بازیابی از حافظه، خطایابی و بازسازی استدلال نشود، احتمال دارد پردازش شناختی در سطح دریافتِ منفعلانه باقی بماند و بنابراین سود یادگیریِ ناشی از تمرین کامل محقق نشود. این تبیین با رویکردهایی هم‌خوان است که اثر هوش مصنوعی را وابسته به چگونگی استفاده می‌دانند AI :وقتی داربست یادگیری است که یادگیرنده را به توضیح‌دهی، مقایسه، نقد و بازنگری وادار کند؛ و وقتی میان‌بُر است که جای پاسخ‌گویی و استدلال را بگیرد.

یافته‌های فراشناختی مطالعه (افزایش JOL و افزایش خطای پایش) از نظر آموزشی حساس‌تر از نتیجه یادگیری‌اند؛ زیرا دانشجو بر مبنای همین قضاوت‌ها تصمیم می‌گیرد مطالعه را ادامه دهد یا متوقف کند، کدام بخش را مرور کند، و در کجا به تمرین بیشتر نیاز دارد. در ادبیات تربیت معلم نیز نشان داده شده است که آگاهی فراشناختی و تشویق فرایندهای فراشناختی با کیفیت یادگیری و عملکرد تحصیلی پیوند دارد (Agrawal et al., 2025)؛ بنابراین هر عامل بیرونی که دقت پایش را کاهش دهد، می‌تواند به‌طور غیرمستقیم به تصمیم‌های ضعیف‌تر در خودتنظیمی مطالعه منجر شود (Xue & Khalid, 2026). از این زاویه، یافته حاضر هشدار می‌دهد که اعتماد به نفس کاذب ناشی از پاسخ آماده، ممکن است به توقف زودهنگام مطالعه و آمادگی ناکافی در موقعیت‌های واقعی بینجامد.

در مقایسه با پژوهش‌هایی که نشان داده‌اند فعالیت‌هایی نظیر طراحی و پاسخ‌گویی به سؤالات، بحث گروهی و دریافت بازخورد همتایان می‌تواند به بهبود یادگیری منجر شود، یافته حاضر نشان داد که استفاده از ChatGPT لزوماً به افزایش عملکرد یادگیری منتهی نشده است. این تفاوت را می‌توان با ماهیت متفاوت فرایندهای یادگیری در دو موقعیت توضیح داد. در فعالیت‌های مشارکتی، یادگیرنده معمولاً در تولید پاسخ، استدلال درباره پاسخ و ارزیابی دیدگاه‌های دیگران نقش فعالی دارد؛ فرایندهایی که مستلزم بازیابی اطلاعات، پردازش عمیق و درگیری شناختی هستند. در مقابل، هنگام استفاده از هوش مصنوعی، بخشی از این فعالیت‌های شناختی ممکن است به ابزار واگذار شود و دانشجو بدون مشارکت فعال در فرایند تولید پاسخ، مستقیماً به نتیجه نهایی دست یابد. در چنین شرایطی، اگرچه پاسخ ارائه‌شده توسط AI می‌تواند از نظر محتوایی صحیح باشد، اما کاهش درگیری شناختی ممکن است مانع از شکل‌گیری یادگیری عمیق شود. افزون بر این، روانی، انسجام و اطمینان ظاهری پاسخ‌های تولیدشده توسط AI می‌تواند احساس تسلط بیشتری در یادگیرنده ایجاد کند؛ احساسی که الزاماً با عملکرد واقعی او در آزمون همخوان نیست و در نتیجه به افزایش خطای پایش فراشناختی منجر می‌شود.

با توجه به اینکه مطالعه حاضر در بافت واقعیِ کلاس (دو کلاسِ موجود) انجام شده، از نظر کاربردپذیری یک مزیت جدی دارد؛ اما از نظر تبیین‌گری، پیام روش‌شناختیِ نتایج این است که قضاوت‌های فراشناختی باید در کنار نمره آزمون گزارش و تفسیر شوند. حتی اگر نمره آزمون تغییر چشمگیری نکند، تغییر در JOL و خطای پایش می‌تواند پیامد آموزشی بزرگ‌تری داشته باشد، چون مسیر تصمیم‌گیریِ یادگیرنده را تغییر می‌دهد.

از نظر دلالت‌های عملی، یافته‌ها پیشنهاد می‌کند که سیاست مجاز/غیرمجاز بودن ChatGPT به‌تنهایی راهگشا نیست؛ بلکه باید قواعد استفاده یادگیرنده‌محور تعریف شود تا AI به جای جانشینِ پاسخ‌گویی، به محرکِ پردازش عمیق تبدیل گردد. دو راهکار کم‌هزینه و قابل‌اجرا در کلاس‌های دانشگاه فرهنگیان عبارت‌اند از: (۱) الزام دانشجو به پاسخ اولیه بدون AI و سپس مقایسه با پاسخ AI همراه با نوشتن توجیه/نقد کوتاه (خودتوضیح‌دهی)، و (۲) استفاده از پرامپت‌های راهنما و سرنخ به جای پاسخ نهایی (مثلاً درخواست از AI برای طرح سؤالِ هدایتگر یا ارائه معیار ارزیابی پاسخ). چنین

طراحی‌هایی با منطق پژوهش‌های پروتکلیِ جدید درباره اثر هوش مصنوعی بر کوشش شناختی و عملکرد همسو است و می‌تواند هم‌زمان از تورم قضاوت یادگیری بکاهد و مشارکت شناختی را بالا ببرد.

در عین حال، چند محدودیت می‌تواند توضیح دهد چرا اثر یادگیری در این مطالعه معنادار نشده است و چگونه می‌توان آن را در پژوهش‌های بعدی دقیق‌تر آزمود: نخست، کلاس‌محور بودن با دو خوشه، احتمال اثرات کلاسی/مدرس و هم‌بستگی درون‌کلاسی را بالا می‌برد و این مسئله در مطالعات آینده با افزایش تعداد کلاس‌ها یا استفاده از مدل‌های چندسطحی بهتر مدیریت می‌شود. دوم، اگرچه استفاده از AI استانداردسازی شده، اما تفاوت در میزان درگیرشدن با پاسخ (خواندن سطحی در برابر بازسازی استدلال) می‌تواند ناهمگنیِ اثر ایجاد کند؛ بنابراین افزودن شاخص‌های فرایندی (زمان، لاگ تعامل، کیفیت توجیه دانشجو) می‌تواند سازوکار اثر را روشن‌تر کند.

نتیجه‌گیری

پژوهش حاضر با هدف بررسی تأثیر استفاده از هوش مصنوعی مولد (ChatGPT) در پاسخ‌گویی به سؤالات تمرینی بر یادگیری و پایش فراشناختی دانشجومعلمان انجام شد. نتایج نشان داد که استفاده از ChatGPT اگرچه موجب افزایش قضاوت یادگیری دانشجویان شد، اما به بهبود معنادار عملکرد یادگیری منجر نگردید و هم‌زمان با افزایش خطای پایش فراشناختی همراه بود. این یافته بیانگر آن است که بهره‌گیری از هوش مصنوعی مولد لزوماً به یادگیری عمیق‌تر منجر نمی‌شود و در برخی شرایط ممکن است احساس یادگیری را بیش از یادگیری واقعی افزایش دهد.

بر این اساس، به نظر می‌رسد اثربخشی آموزشی هوش مصنوعی بیش از آنکه به خودِ فناوری وابسته باشد، به شیوه استفاده از آن بستگی دارد. استفاده از AI زمانی می‌تواند به بهبود یادگیری کمک کند که در قالب فعالیت‌هایی طراحی شود که یادگیرنده را به بازیابی اطلاعات، استدلال، خودتوضیح‌دهی و ارزیابی انتقادی پاسخ‌ها وادار کند. در غیر این صورت، احتمال دارد بخشی از فرایندهای شناختی یادگیرنده به ابزار واگذار شود و فاصله میان ادراک یادگیری و عملکرد واقعی افزایش یابد.

پژوهش حاضر با چند محدودیت همراه بود. نخست، مطالعه در دو کلاس موجود و با تخصیص تصادفی در سطح کلاس انجام شد؛ بنابراین، اگرچه پیش‌آزمون و متغیرهای زمینه‌ای برای کنترل تفاوت‌های اولیه به کار رفتند، امکان تصادفی‌سازی کامل فردی وجود نداشت. دوم، دامنه مداخله به پاسخ‌گویی به سؤالات تمرینی محدود بود و نتایج آن به سایر کاربردهای ChatGPT مانند خلاصه‌سازی متن، تولید محتوای نوشتاری یا انجام تکالیف ترکیبی قابل تعمیم مستقیم نیست. سوم، قضاوت یادگیری ماهیتی خودگزارشی داشت و ممکن است تحت تأثیر اعتمادبه‌نفس، نگرش به فناوری یا تجربه قبلی استفاده از هوش مصنوعی قرار گرفته باشد. چهارم، حجم نمونه محدود و متعلق به یک واحد دانشگاهی بود؛ ازاین‌رو، تعمیم نتایج به سایر رشته‌ها، دانشگاه‌ها و گروه‌های سنی باید با احتیاط انجام شود. پنجم، میزان درگیری شناختی واقعی دانشجویان هنگام استفاده از ChatGPT به‌صورت مستقیم اندازه‌گیری نشد و این موضوع می‌تواند در پژوهش‌های آینده با شاخص‌های فرایندی دقیق‌تر بررسی شود.

با توجه به یافته‌های پژوهش، پیشنهاد می‌شود در طراحی تکالیف مبتنی بر هوش مصنوعی، نقش AI به‌عنوان داربست یادگیری مورد توجه قرار گیرد نه جایگزین فرایند یادگیری. همچنین توصیه می‌شود در برنامه‌های تربیت معلم، آموزش سواد هوش مصنوعی و مهارت‌های فراشناختی مرتبط با ارزیابی انتقادی خروجی‌های AI مورد تأکید قرار گیرد. در نهایت، انجام پژوهش‌های آتی با حجم نمونه بیشتر، تعداد کلاس‌های بالاتر، شاخص‌های فرایندی دقیق‌تر و بررسی ماندگاری یادگیری در بازه‌های زمانی طولانی‌تر می‌تواند به تبیین بهتر سازوکارهای اثرگذاری هوش مصنوعی مولد در محیط‌های آموزشی کمک کند.

پیروی از اصول اخلاق پژوهش

نویســـندگان اصـــول اخلاقی را در انجام و انتشـــار این پژوهش علمی رعایت نموده‌اند و این موضـــوع موردتأیید همه آن‌هاست.

مشارکت نویسندگان

سهم و مشارکت نویسندگان در نگارش مقاله به میزان برابر و مساوی است.

تشکر و قدردانی

نویسندگان از تمامی افراد شرکت کننده در پژوهش تقدیر و تشکر می نمایند.

تعارض منافع

نویسندگان هیچگونه تعارض منافعی ندارند.

ORCID

Amin Hosein Pour https://orcid.org/0009-0008-3554-6923
Reza Alam https://orcid.org/0009-0000-8776-6946
Seyed Ali Mosallemi https://orcid.org/0009-0005-0375-4064

منابع

ابراهیمی، فاطمه و ضرابیان، فروزان. (۱۳۹۷). نقش فناوری‌های جدید آموزشی در تعلیم و تربیت: مقاله مروری. *کنفرانس ملی دستاوردهای نوین جهان در تعلیم و تربیت، روانشناسی، حقوق و مطالعات فرهنگی اجتماعی*. بازیابی از https://sid.ir/paper/898013/fa

اسماعیلی، لیدا، سهرابی، نادره، مهریار، امیرهوشنگ، و خیر، محمد. (۱۳۹۹). مدل علی راهبردهای یادگیری (شناختی و فراشناختی) با نقش واسطه‌گری امید تحصیلی برای خودکارآمدی تحصیلی در دانش‌آموزان متوسطه. *روش‌ها و مدل‌های روان‌شناختی، ۱۱(۴۱)*، ۳۴-۱۷. https://dorl.net/dor/20.1001.1.22285516.1399.11.41.2.6

امیدی، مصطفی، و رستمی بیرق، مهسا. (۱۴۰۳). ادراک و نگرش معلمان نسبت به استفاده از هوش مصنوعی در فرآیند تدریس و ارزشیابی (مطالعه موردی: معلمان شاغل در دوره متوسطه کلانشهر تهران و حومه). *مجله شناخت، رفتار، یادگیری، ۱(۴)*، ۹۴-۷۷. https://doi.org/10.61838/jcbl.1.4.7

جدیدی، حمید، و افشاری، ابوالفضل. (۱۴۰۴). نقش هوش مصنوعی در تغییر مکانیسم‌های ارزیابی و بازخورد در آموزش و پرورش. *یازدهمین همایش علمی‌پژوهشی علوم تربیتی و روانشناسی، آسیب‌های اجتماعی و فرهنگی ایران، تهران*، ایران. https://civilica.com/doc/2471483

جعفری، رضا، و رحمتی، حسینعلی. (۱۴۰۴). کاربرد هوش مصنوعی در مشاورهٔ اخلاقی؛ از ظرفیت‌های تحول‌آفرین تا چالش‌های اخلاقی. *فصلنامه اخلاق‌پژوهی، ۸(۲)*، ۷۲-۴۹. https://doi.org/10.22034/ethics.2025.51991.1790

حسینی تلی، بتول، رودینی، طاهره، جمشیدزهی، عبدالرزاق، و بمپوری، ادهم. (۱۴۰۴). آموزش هوشمند: چگونه AI مسیر یادگیری را برای هر دانش‌آموز تغییر می‌دهد؟ *پنجمین کنفرانس بین‌المللی پژوهش‌های مدیریت، تعلیم و تربیت در آموزش و پرورش*، تهران، ایران. بازیابی از https://civilica.com/doc/2330823

حیدرزاده، آبتین، و پنجه‌علی‌بیک، سمانه. (۱۴۰۱). دوازده نکته‌ی کلیدی در منشور اخلاقی و رفتار حرفه‌ای استفاده از هوش مصنوعی در آموزش علوم پزشکی: یک مرور نظام‌مند. *طب و تزکیه، ۳۱(۱)*، ۳۶-۲۱. بازیابی از https://sid.ir/paper/1072863/fa

خانی هالانی، مجتبی، و محمودی، مهدی. (۱۴۰۴). مروری بر روش‌های نوین ارزشیابی با بهره‌گیری از هوش مصنوعی در آموزش دانشگاهی. *اولین همایش ملی پژوهش‌های نوین در تاب‌آوری، امید به زندگی و نوآوری‌های آموزشی*، سبزوار، ایران. بازیابی از https://civilica.com/doc/2304806

رجبیان ده زیره، مریم. (۱۴۰۳). شناسایی چالش‌ها و قابلیت‌های هوش مصنوعی در آموزش و یادگیری با ارائه راهکارها. *فناوری آموزش، ۱۸(۴)*، ۹۵۰-۹۲۱. https://doi.org/10.22061/tej.2024.10777.3058

شمالی احمدآبادی، مهدی، و برخورداری احمدآبادی، عاطفه. (۱۴۰۴). نقش نگرش به هوش مصنوعی، هدفمندی در زندگی و احساس ارزشمندی در پیش‌بینی بی‌تفاوتی اخلاقی نوجوانان در فضای مجازی. *کاوش‌های نوین در علوم محاسباتی و مدیریت رفتاری*، e221432. https://doi.org/10.22034/necsbm.2025.512396.1111

عنابستانی، علی‌اکبر، توکلی‌نیا، جمیله، و نیکنامی، نسیم. (۱۴۰۳). تبیین اثرگذاری هوش مصنوعی بر بهبود کیفیت زندگی شهروندان با رویکرد آینده‌پژوهی (مورد مطالعه: کلان‌شهر مشهد). *اقتصاد و برنامه‌ریزی شهری، ۵(۱)*، ۲۱-۶. https://doi.org/10.22034/uep.2024.442663.1461

Agrawal, P. K., Gore, R., Kumar, M., Kushwaha, V., Goenka, S., & Agrawal, S. (2025). Metacognitive Awareness and Academic Performance: Implications from a Cognitive Neuroscience Perspective in Pre-service Teacher Education. *Annals of neurosciences, 09727531251361976*. *Advance online*

publication. https://doi.org/10.1177/09727531251361976

Ahmed, A., Kerr, E., & O'Malley, A. (2025). Quality assurance and validity of AI-generated single best answer questions. *BMC medical education, 25(1*), 300. https://doi.org/10.1186/s12909-025-06881-w

Benatar, S., & Daneman, D. (2020). Disconnections between medical education and medical practice: A neglected dilemma. *Global public health, 15(9),* 1292–1307. https://doi.org/10.1080/17441692.2020.1756376

Bishop, D. V. M., Thompson, J., & Parker, A. J. (2022). Can we shift belief in the 'Law of Small Numbers'?. *Royal Society open science, 9(3),* 211028. https://doi.org/10.1098/rsos.211028

Carpenter, S. K., Northern, P. E., Tauber, S. U., & Toftness, A. R. (2020). Effects of lecture fluency and instructor experience on students' judgments of learning, test scores, and evaluations of instructors. Journal of experimental psychology. *Applied, 26(1),* 26–39. https://doi.org/10.1037/xap0000234

Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R. F., Fan, Y., Li, Y., Lu, B., Yuan, M., Zhang, J., Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: study protocol for a randomized controlled experiment among college students. *Trials, 26(1),* 244. https://doi.org/10.1186/s13063-025-08950-3

Chien, C. Y., Tsai, S. L., Huang, C. H., Wang, M. F., Lin, C. C., Chen, C. B., Tsai, L. H., Tseng, H. J., Huang, Y. B., & Ng, C. J. (2024). Effectiveness of Blended Versus Traditional Refresher Training for Cardiopulmonary Resuscitation: Prospective Observational Study. *JMIR medical education, 10,* e52230. https://doi.org/10.2196/52230

Daher, W., & Hashash, I. (2022). Mathematics Teachers' Encouragement of Their Students' Metacognitive Processes. *European journal of investigation in health, psychology and education, 12(9),* 1272–1284. https://doi.org/10.3390/ejihpe12090088

Dell, K. A., & Wantuch, G. A. (2017). How-to-guide for writing multiple choice questions for the pharmacy instructor. *Currents in pharmacy teaching & learning, 9(1),* 137–144. https://doi.org/10.1016/j.cptl.2016.08.036

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56(2),* 489-530.

Fisher, M., & Oppenheimer, D. M. (2021). The illusion of knowledge and explanatory depth. *Cognitive Science, 45(2),* e12919.

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy. *Brain sciences, 15(2),* 203. https://doi.org/10.3390/brainsci15020203

Guilding, C., Pye, R. E., Butler, S., Atkinson, M., & Field, E. (2021). Answering questions in a co-created formative exam question bank improves summative exam performance, while students perceive benefits from answering, authoring, and peer discussion: A mixed methods analysis of PeerWise. *Pharmacology research & perspectives, 9(4),* e00833. https://doi.org/10.1002/prp2.833

Jaeger, B., & Wilks, M. (2023). The Relative Importance of Target and Judge Characteristics in Shaping the Moral Circle. *Cognitive science, 47(10),* e13362. https://doi.org/10.1111/cogs.13362

Logren, A., Ruusuvuori, J., & Laitinen, J. (2017). Group members' questions shape participation in health counselling and health education. *Patient education and counseling, 100(10),* 1828–1841. https://doi.org/10.1016/j.pec.2017.05.003

Mehta, N., Benjamin, J., Agrawal, A., Valanci, S., Masters, K., & MacNeill, H. (2025). Addressing educational overload with generative AI through dual coding and cognitive load theories. *Medical teacher, 1–3. Advance online publication*. https://doi.org/10.1080/0142159X.2025.2543548

Niźnikowski, T., Arnista, P., Sadowski, J., Mastalerz, A., Romero-Ramos, O., Fernández-Rodríguez, E.,

Łuba-Arnista, W., Biegajło, M., Różański, P., Niźnikowska, E., Karaś, A., Kuśmierczyk, P., & Nogal, M. (2025). Enhancing the learning of sports skills through verbal feedback. *Frontiers in sports and active living, 7*, 1519365. https://doi.org/10.3389/fspor.2025.1519365

Otto, A. R., Braem, S., Silvetti, M., & Vassena, E. (2022). Is the juice worth the squeeze? Learning the marginal value of mental effort over time. *Journal of experimental psychology. General, 151(10)*, 2324–2341. https://doi.org/10.1037/xge0001208

Remington, R. W., Yuen, H. W., & Pashler, H. (2016). With practice, keyboard shortcuts become faster than menu selection: A crossover interaction. *Journal of experimental psychology. Applied, 22(1)*, 95–106. https://doi.org/10.1037/xap0000069

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20(9)*, 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Sun, Q., Chen, F., & Yin, S. (2023). The role and features of peer assessment feedback in college English writing. *Frontiers in psychology, 13*, 1070618. https://doi.org/10.3389/fpsyg.2022.1070618

Tao, W., Zhang, M., & Liu, Y. (2024). Mastering delegation to artificial intelligence creative tools: The concept, dimensions, and development of a scale to measure cognitive outsourcing. *Social Behavior and Personality: an international journal*, *52*(12), 1-15. https://doi.org/10.2224/sbp.13907

Wilson, J. T., & Bauer, P. J. (2024). Generative and active engagement in learning neuroscience: A comparison of self-derivation and rephrase. *Cognition, 245*, 105709. https://doi.org/10.1016/j.cognition.2023.105709

Xie, X., Zhang, L. J., & Wilson, A. J. (2025). Comparing ChatGPT Feedback and Peer Feedback in Shaping Students' Evaluative Judgement of Statistical Analysis: A Case Study. *Behavioral sciences (Basel, Switzerland), 15(7)*, 884. https://doi.org/10.3390/bs15070884

Xue, Y., & Khalid, F. (2026). Investigating the impact of metacognitive regulation on students' academic performance: a quantitative study of Chinese distance learners. *Frontiers in psychology, 17*, 1726956. https://doi.org/10.3389/fpsyg.2026.1726956

Sample: References in Persian

Anabestani, A. A., Tavakolinia, J., & Niknami, N. (2024). The role of artificial intelligence in improving citizens' quality of life using a futures studies approach (Case study: Mashhad Metropolis). *Urban Economics and Planning, 5*(1), 6–21. https://doi.org/10.22034/uep.2024.442663.1461 [In Persian].

Ebrahimi, F., & Zarabian, F. (2018). The role of new educational technologies in education: A review article. In *Proceedings of the National Conference on New World Achievements in Education, Psychology, Law, and Socio-Cultural Studies*. https://sid.ir/paper/898013/fa [In Persian].

Esmaeili, L., Sohrabi, N., Mehryar, A. H., & Khayyer, M. (2020). A causal model of learning strategies (cognitive and metacognitive) with the mediating role of academic hope for academic self-efficacy in high school students. *Journal of Psychological Methods and Models, 11*(41), 17–34. https://dorl.net/dor/20.1001.1.22285516.1399.11.41.2.6 [In Persian].

Heidarzadeh, A., & Panjeh Ali Beik, S. (2022). Twelve key points in codes of ethics and professional conduct of using artificial intelligence in medical sciences education: A systematic review. *Medicine and Spiritual Cultivation, 31*(1), 21–36. https://sid.ir/paper/1072863/en [In Persian].

Hosseini Telli, B., Roudini, T., Jamshidzahi, A., & Bampouri, A. (2025). Smart education: How artificial intelligence changes the learning path for every student. In *Proceedings of the 5th International Conference on Management, Teaching, and Education Research in the Education System*. https://civilica.com/doc/2330823 [In Persian].

Jadidi, H., & Afshari, A. (2025). The role of artificial intelligence in transforming assessment and feedback mechanisms in education. In *Proceedings of the 11th Scientific Research Conference on Educational Sciences and Psychology, and Social and Cultural Harms in Iran*. https://civilica.com/doc/2471483 [In Persian].

Jafari, R., & Rahmati, H. (2025). The application of artificial intelligence in ethical counseling: From transformative potentials to ethical challenges. *Journal of Moral Studies, 8*(2), 49–72. https://doi.org/10.22034/ethics.2025.51991.1790 [In Persian].

Khani Halani, M., & Mahmoudi, M. (2025). A review of modern assessment methods using artificial intelligence in higher education. In *Proceedings of the First National Conference on Emerging Research in Resilience, Hope for Life, and Educational Innovations*. https://civilica.com/doc/2304806 [In Persian].

Omidi, M., & Rostami Beyragh, M. (2025). Teachers' perceptions and attitudes toward the use of artificial intelligence in teaching and evaluation processes (Case study: Secondary school teachers in Greater Tehran and surrounding areas). *Cognition, Behavior, Learning, 1*(4), 77–94. https://doi.org/10.61838/jcbl.1.4.7 [In Persian].

Rajabiyan Dehzireh, M. (2024). Identifying the challenges and capabilities of artificial intelligence in teaching and learning by providing solutions. *Technology of Education Journal, 18*(4), 921–950. https://doi.org/10.22061/tej.2024.10777.3058 [In Persian].

Shomali Ahmadabadi, M., & Barkhordari Ahmadabadi, A. (2026). The role of attitudes toward artificial intelligence, purpose in life, and self-esteem in predicting moral disengagement among adolescents in the digital space. *Novel Explorations in Computational Science and Behavioral Management, 3*(2), 134–148. https://doi.org/10.22034/necsbm.2025.512396.1111 [In Persian].