Journal of Human-Information Interaction
Online ISSN: 2423-7418
https://hii.khu.ac.ir/

Comparing ChatGPT-4 and Gemini 2.5 Pro in Meeting the Information Needs of Doctoral Students at Kharazmi University

Mahdiyeh Khondabi[1], **Mohammad Zerehsaz**[2], **Ali Azimi**[3]✉

1. MSc in Knowledge and Information Science, Kharazmi University, Tehran, Iran. **Email**: mahdiehkhondabi@gmail.com
2. Department of Knowledge and Information Science, Kharazmi University, Tehran, Iran. **Email**: zerehsaz@khu.ac.ir
3. Corresponding author, Department of Knowledge and Information Science, Kharazmi University, Tehran, Iran. **Email**: azimia@khu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 27 June 2025
Revised in 4 September 2025
Accepted 16 September 2025
Published online 27 September 2025

Keywords:
ChatGPT-4, Gemini 2.5 Pro, Information needs, Information-seeking behavior, Persian, Doctoral students, Artificial intelligence literacy, Kharazmi University

ABSTRACT

Purpose: This study aimed to compare ChatGPT-4 and Gemini 2.5 Pro in meeting the information needs of doctoral students in Persian. The focus was to examine how accurately, relevantly, clearly, practically, diversely, and effectively each model supports academic, specialized, and Persian-language information-seeking needs.

Method: The study used a mixed-methods approach. The participants included 36 doctoral students from the Faculty of Psychology and Educational Sciences at Kharazmi University during the 2024–2025 academic year, selected through purposive sampling. An initial questionnaire and a comparative questionnaire were used, and the students evaluated the relevant indicators after direct interaction with both models. In the qualitative phase, think-aloud protocols, semi-structured interviews, and observation of the interaction process using Morae software were employed. Quantitative data were analyzed using the Wilcoxon signed-rank test, while qualitative data were analyzed through coding in MAXQDA.

Findings: The findings showed that the superiority of the two models does not follow a uniform pattern. ChatGPT-4 was rated significantly higher in overall performance, and users found it more effective in initiating searches, organizing interaction, formulating the initial problem, and enhancing information-seeking confidence. In contrast, Gemini 2.5 Pro showed significant superiority in the accuracy of scientific and specialized responses, the diversity of information, the coverage of different aspects of the topic, and the practical usefulness of responses. No significant difference was found between the two models in terms of comprehensibility. The qualitative findings also indicated that Gemini 2.5 Pro was more strongly associated with content depth, comprehensiveness, and support for research work, whereas ChatGPT-4 was mainly described in terms of conversational fluency, response organization, and user reassurance.

Conclusion: The results indicate that the two models are complementary tools. ChatGPT-4 is better suited to initiating searches and organizing users' thinking, while Gemini 2.5 Pro is more useful for specialized, research-oriented questions. However, neither model can replace academic databases, expert judgment, or the researcher's critical evaluation.

Cite this article: Khondabi, M., Zerehsaz, M., & Azimi, A. (2026). Comparing ChatGPT-4 and Gemini 2.5 Pro in Meeting the Information Needs of Doctoral Students at Kharazmi University. *Human-Information Interaction, 12*(3), 49-68.

 Publisher: University of Kharazmi.

Journal of Human-Information Interaction
kharazmi University
Online ISSN: 2423-7418
https://hii.khu.ac.ir/

Introduction
This study compares ChatGPT-4 and Gemini 2.5 Pro in meeting the information needs of doctoral students in Persian. The focus is not limited to fluency or length of generated answers; rather, the study examines how each system supports users when the information need is academic, discipline-related, and expressed in Persian. The comparison covers perceived overall performance, response accuracy, relevance, clarity, feedback, accessibility, information diversity, topic coverage, comprehensibility, applicability, interaction with complex and multi-stage questions, information-seeking confidence, perceived benefits, and user-reported challenges.

Methods and Materials
A mixed-methods design was used. The participants were 36 doctoral students from the Faculty of Psychology and Educational Sciences at Kharazmi University during the 2024–2025 academic year. Participants were selected through purposive sampling with attention to disciplinary diversity. Quantitative data were collected through an initial questionnaire and a comparative questionnaire administered after participants interacted directly with both models. The qualitative component used think-aloud protocols, semi-structured interviews, and screen-based observation with Morae software to capture users' reasoning, hesitation, prompt reformulation, satisfaction, and perceived barriers during interaction. Internal consistency was acceptable for the initial questionnaire ($\alpha = .703$), the ChatGPT-4 scale ($\alpha = .887$), and the Gemini 2.5 Pro scale ($\alpha = .923$). Because normality tests indicated departures from normality for most variables, paired comparisons were analyzed using the Wilcoxon signed-rank test. Qualitative data were coded through open, axial, and selective coding in MAXQDA.

Results and discussion
The findings showed a differentiated pattern rather than the absolute superiority of either model. For perceived overall performance, ChatGPT-4 was rated higher by most participants, and the difference was statistically significant ($z = -4.54$, $p < .001$). This result indicates that participants experienced ChatGPT-4 as smoother, more reassuring, and more supportive in beginning or organizing an information-seeking process. In contrast, Gemini 2.5 Pro received higher ratings for response accuracy on scientific and specialized questions ($z = -4.10$, $p < .001$) and also performed better in information diversity, topic coverage, and applicability. The difference in comprehensibility was not statistically significant ($z = -1.33$, $p = .185$), suggesting that both systems were understandable to doctoral students. However, applicability was significantly higher for Gemini 2.5 Pro ($z = -3.59$, $p < .001$), whereas information-seeking confidence was higher for ChatGPT-4 ($z = -3.14$, $p = .002$). The qualitative findings explained the quantitative pattern. Gemini 2.5 Pro was associated with greater content depth, more detailed responses, richer topic coverage, broader viewpoints, and stronger usefulness for research-oriented tasks. ChatGPT-4 was associated with more fluent interactions, clearer organization in some responses, appropriate brevity, example-based explanations, and user reassurance. Perceived benefits were grouped into three axes: interactional benefits, response quality, and applicability. Perceived challenges were grouped into performance-related, infrastructural, and user-centered challenges. These included overly general responses, limited critical analysis, instability in response generation, slow or unstable access, regional restrictions, cognitive load, dependence on generated answers, and difficulty judging when a generated answer was sufficiently accurate.

Conclusion
The study concludes that ChatGPT-4 and Gemini 2.5 Pro should be treated as complementary tools rather than interchangeable systems. ChatGPT-4 appears more suitable for initial exploration, problem formulation, accessible explanation, and increasing users' confidence. Gemini 2.5 Pro appears more suitable for scientific and specialized information needs that require accuracy, broader coverage, diversity of viewpoints, and practical academic usefulness. Neither model should replace scholarly databases, expert supervision, source verification, or critical judgment. For Persian-language academic

Journal of Human-Information Interaction
kharazmi University
Online ISSN: 2423-7418
https://hii.khu.ac.ir/

contexts, effective use of both systems requires artificial intelligence literacy, information literacy, prompt-writing skill, and explicit verification of generated content.

Ethical Considerations

Compliance with Research Ethics

The students' participation was informed and voluntary. Data confidentiality was maintained, and the participants' identifying information was not disclosed in the research report.

Authors' Contributions

First author: Conducting research, data collection, quantitative and qualitative analysis, and preparing the initial draft. **Second author**: Contributing to methodological design, providing scientific consultation, and reviewing the manuscript. **Third author**: Scientific supervision, quantitative and qualitative analysis, theoretical and methodological guidance, final review, and handling article correspondence as the corresponding author.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This article is derived from the master's thesis of Mahdieh Khondabi in Knowledge and Information Science at Kharazmi University, Tehran, Iran, conducted under the supervision of Dr. Ali Azimi and the advisory support of Dr. Mohammad Zarehsaz. The study was performed in the Kharazmi University Human Information Interaction Lab. Hereby, the authors would like to thank the students who cooperated in the data collection process.

نشریه تعامل انسان و اطلاعات
https://hii.khu.ac.ir/
شاپا الکترونیکی: ۲۴۲۳-۷۴۱۸
دانشگاه خوارزمی

مقایسه عملکرد چت‌جی‌پی‌تی ۴ و جمنای ۲,۵ پرو در پاسخ‌گویی به نیازهای اطلاعاتی دانشجویان دکتری دانشگاه خوارزمی

مهدیه خندابی[۱]، محمد زره‌ساز[۲]، علی عظیمی[۳]

۱. کارشناسی ارشد علم اطلاعات و دانش‌شناسی، دانشگاه خوارزمی، تهران، ایران. رایانامه: mahdiehkhondabi@gmail.com
۲. عضو هیات علمی گروه علم اطلاعات و دانش‌شناسی، دانشگاه خوارزمی، تهران، ایران. رایانامه: zerehsaz@khu.ac.ir
۳. نویسنده مسئول، گروه علم اطلاعات و دانش‌شناسی، دانشگاه خوارزمی، تهران، ایران. رایانامه: azimia@khu.ac.ir

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

تاریخ دریافت:
۰۳/۰۷/۱۴۰۴

تاریخ بازنگری:
۱۰/۰۹/۱۴۰۴

تاریخ پذیرش:
۱۴/۰۹/۱۴۰۴

تاریخ انتشار:
۳۰/۰۲/۱۴۰۵

کلیدواژه‌ها:
چت‌جی‌پی‌تی ۴، جمنای ۲,۵ پرو، نیاز اطلاعاتی، رفتار اطلاع‌یابی، زبان فارسی، دانشجوی دکتری، سواد هوش مصنوعی، دانشگاه خوارزمی

چکیده

هدف: پژوهش حاضر با هدف مقایسه چت‌جی‌پی‌تی ۴ و جمنای ۲,۵ پرو در رفع نیازهای اطلاعاتی دانشجویان دکتری به زبان فارسی انجام شد. تمرکز اصلی پژوهش بر این بود که هر مدل در پاسخ‌گویی به نیازهای اطلاعاتی دانشگاهی، تخصصی و فارسی‌زبان تا چه اندازه دقیق، مرتبط، قابل‌فهم، کاربردی، متنوع و پشتیبان فرایند جستجوی علمی عمل می‌کند.

روش: پژوهش با رویکرد ترکیبی انجام شد. مشارکت‌کنندگان شامل ۳۶ دانشجوی دکتری دانشکده روان شناسی و علوم تربیتی دانشگاه خوارزمی در سال تحصیلی ۱۴۰۳-۱۴۰۴ بودند که به شیوه هدفمند انتخاب شدند. در بخش کمی، از پرسشنامه اولیه و پرسشنامه مقایسه‌ای استفاده شد و دانشجویان پس از تعامل مستقیم با هر دو مدل، شاخص‌های مورد نظر را ارزیابی کردند. در بخش کیفی، پروتکل بلنداندیشی، مصاحبه نیمه‌ساختاریافته و مشاهده فرایند تعامل با کمک نرم‌افزار مورائه به کار رفت. داده‌های کمی با آزمون رتبه‌های علامت‌دار ویلکاکسون و داده‌های کیفی با کدگذاری در مکس‌کیودا تحلیل شدند.

یافته‌ها: این پژوهش نشان دادند که برتری دو مدل یکدست نیست. چت‌جی‌پی‌تی ۴ در عملکرد کلی به‌طور معناداری بالاتر ارزیابی شد و کاربران آن را در آغاز جستجو، نظم تعامل، صورت‌بندی اولیه مسئله و ایجاد اعتمادبه‌نفس اطلاع‌یابی مطلوب‌تر دانستند. در مقابل، جمنای ۲,۵ پرو در دقت ادراک شده کاربران در خصوص پاسخ‌های علمی و تخصصی، تنوع اطلاعات، پوشش ابعاد موضوع و کاربردی‌بودن پاسخ‌ها برتری معنادار داشت. تفاوت دو مدل از نظر قابل‌فهم بودن معنادار نبود. یافته‌های کیفی نیز نشان داد جمنای ۲,۵ پرو بیشتر با عمق محتوایی، جامعیت و پشتیبانی از کار پژوهشی پیوند دارد؛ در حالی که چت‌جی‌پی‌تی ۴ بیشتر با روانی گفت‌وگو، سازمان‌دهی پاسخ و اطمینان‌بخشی توصیف شد.

نتیجه‌گیری: نتایج نشان می‌دهد این دو مدل ابزارهایی مکمل‌اند. چت‌جی‌پی‌تی ۴ برای شروع جستجو و نظم‌بخشی ذهنی مناسب‌تر است و جمنای ۲,۵ پرو برای پرسش‌های تخصصی و پژوهشی کاربرد بیشتری دارد. بااین‌حال، هیچ‌یک جایگزین پایگاه‌های علمی، داوری متخصص و ارزیابی انتقادی پژوهشگر نیستند.

استناد: خندابی، مهدیه؛ زره‌ساز، محمد؛ و عظیمی، علی. مقایسه عملکرد چت‌جی‌پی‌تی ۴ و جمنای ۲,۵ پرو در پاسخ‌گویی به نیازهای اطلاعاتی دانشجویان دکتری دانشگاه خوارزمی. *تعامل انسان و اطلاعات*، ۱۲ (۳)، ۲۳-صص. ۴۹-۶۸.

ناشر: دانشگاه خوارزمی تهران.

مقدمه

در سال‌های اخیر، استفاده دانشگاهیان از ابزارهای هوش مصنوعی مولد[1] از یک رفتار حاشیه‌ای به بخشی از زیست‌بوم یادگیری، پژوهش و اطلاع‌یابی تبدیل شده است. شواهد پیمایشی در آموزش عالی نشان می‌دهد که دانشجویان از این ابزارها برای توضیح مفاهیم، خلاصه سازی متون، پیشنهاد ایده‌های پژوهشی، سازمان‌دهی نوشته‌ها و جستجوی اطلاعات استفاده می‌کنند و این روند با سرعتی چشمگیر در حال گسترش است (فریمن، ۲۰۲۵؛ شورای آموزش دیجیتال، ۲۰۲۴). بنابراین، مسئله اصلی دیگر صرفاً امکان استفاده از چت‌بات‌های هوشمند نیست، بلکه تشخیص کیفیت، حدود کاربرد، اعتمادپذیری و تناسب هر مدل با نوع نیاز اطلاعاتی کاربر است.

این تحول، رفتار اطلاع‌یابی دانشجویان را از جستجوی خطی و کلیدواژه‌محور به تعامل گفت‌وگومحور منتقل کرده است؛ تعاملی که در آن پاسخ به‌صورت منسجم، خلاصه‌شده و اغلب آماده استفاده عرضه می‌شود. در چارچوب علم اطلاعات و دانش‌شناسی، نیاز اطلاعاتی و رفتار اطلاع‌یابی همچنان به فاصله میان وضعیت دانشی موجود و مطلوب و کنش‌های کاربر برای تشخیص، جستجو، ارزیابی و استفاده از اطلاعات اشاره دارد (ویلسون، ۱۹۹۹؛ ویلسون، ۲۰۰۰؛ ناومر و فیشر، ۲۰۱۰؛ افزال، ۲۰۱۷)، اما در محیط‌های مبتنی بر هوش مصنوعی مولد، این رفتار با اعتماد به پاسخ، توان پرسش‌سازی، پیگیری مکالمه، اصلاح پرامپت، تشخیص خطا و راستی‌آزمایی خروجی گره می‌خورد.

مدل‌های زبانی بزرگ[2] را می‌توان جهشی مهم در آموزش و پژوهش دانست؛ زیرا علاوه بر تولید متن، می‌توانند نقش‌هایی مانند دستیار توضیح‌دهنده، ابزار خلاصه‌ساز، مولد ایده، سازمان‌دهنده محتوای پژوهشی، پشتیبان نگارش و همراه یادگیری شخصی‌سازی‌شده را ایفا کنند. مرورهای جدید نشان می‌دهد که این ابزارها ظرفیت بهبود بازخورد، دسترسی سریع به توضیح، خودتنظیمی یادگیری و پشتیبانی از فعالیت‌های پژوهشی را دارند، اما همزمان با خطرهایی مانند اتکای بیش از حد، خطای محتوایی، سوگیری، ابهام در منشأ اطلاعات، حریم خصوصی و کاهش داوری انتقادی همراه‌اند (گیزانی و همکاران، ۲۰۲۵؛ شی و همکاران، ۲۰۲۵؛ شهزاد و همکاران، ۲۰۲۴). از این رو، انتخاب مدل مناسب برای یک تکلیف علمی، تصمیمی روش‌شناختی و نه صرفاً ترجیحی فناورانه است.

از نظر فنی، چت‌بات‌هایی مانند چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو هر دو بر خانواده مدل‌های زبانی بزرگ و معماری‌های عمیق پردازش زبان طبیعی تکیه دارند؛ بااین‌حال، شیوه طراحی، آموزش، هم‌ترازی، پنجره زمینه، چندوجهی‌بودن و راهبرد پاسخ‌دهی آنها یکسان نیست. ادبیات فنی مدل‌های زبانی نشان می‌دهد که تفاوت در معماری، داده‌های آموزشی، تنظیم پس از آموزش، راهبردهای هم‌ترازی و قابلیت پردازش زمینه می‌تواند بر دقت، انسجام، طول پاسخ، توان استدلال، پوشش موضوعی و کاربردپذیری خروجی اثر بگذارد (کومار، ۲۰۲۴). گزارش فنی جی پی تی چهار این مدل را سامانه‌ای بزرگ‌مقیاس و چندوجهی معرفی می‌کند که می‌تواند ورودی متنی و تصویری را بپذیرد و خروجی متنی تولید کند؛ همچنین بر پیش‌آموزش مبتنی بر پیش‌بینی توکن بعدی و هم‌ترازی پس از آموزش برای بهبود واقع‌نمایی و پیروی از رفتار مطلوب تأکید دارد (اوپن‌ای‌آی، ۲۰۲۳).

در کاربرد آموزشی و اطلاع‌یابی، چت‌جی‌پی‌تی ۴ بیشتر در موقعیت‌هایی مزیت می‌یابد که کاربر به روشن‌سازی مسئله، ساده‌سازی مفهوم، نظم‌بخشی اولیه و کاهش ابهام شناختی نیاز دارد. از نظر «پشتیبانی شناختی»، این مدل برای شروع گفت‌وگو، بازنویسی پرسش، تولید طرح اولیه، ارائه مثال و تبدیل موضوع پیچیده به توضیح قابل فهم کارآمد است؛ در مقابل، جمنای ۲٫۵ پرو برای تحلیل پرسش‌های چندمرحله‌ای، مقایسه ابعاد مسئله و تولید پاسخ‌های گسترده‌تر ظرفیت بیشتری نشان می‌دهد (اوپن‌ای‌آی، ۲۰۲۳؛ اوپن‌ای‌آی، ۲۰۲۴؛ گوگل، ۲۰۲۵). البته این تمایز به معنای برتری مطلق یکی از دو مدل نیست، زیرا کیفیت خروجی به نوع پرامپت، زبان، موضوع، سطح تخصص کاربر و معیار ارزیابی وابسته است.

در مقابل، گزارش‌های رسمی گوگل درباره جمنای ۲٫۵ پرو بر «مدل اندیشنده[3]» بودن، استدلال پیش از پاسخ، قابلیت چندوجهی بومی، پردازش زمینه طولانی و توان کار با منابع متنوع مانند متن، تصویر، صوت، ویدئو و مخازن کد تأکید می‌کنند

[1] Generative Artificial Intelligence (GenAI)
[2] Large Language Models (LLMs).
[3] Thinking Model

(گوگل، ۲۰۲۵؛ تیم جمنای و همکاران، ۲۰۲۴). بنابراین، راهبرد پاسخ‌دهی جمنای ۲٫۵ پرو را می‌توان بیشتر با تحلیل چندمرحله‌ای، حفظ زمینه وسیع، گسترش ابعاد موضوع، مقایسه دیدگاه‌ها و ترکیب اطلاعات پراکنده توضیح داد؛ قابلیتی که در پرسش‌های تخصصی، چندبخشی و پژوهش‌محور اهمیت بیشتری می‌یابد.

مطالعات تطبیقی نیز از اصل زمینه‌مند بودن عملکرد مدل‌ها حمایت می‌کنند. در برخی پژوهش‌های آموزشی، جی پی تی ۴ از نظر ارزیابی ذهنی کاربران، انسجام یا کیفیت تولید محتوای آموزشی بهتر از جمنای گزارش شده است (لنگ و همکاران، ۲۰۲۴؛ بایتاک، ۲۰۲۴)؛ در حالی که برخی مطالعات در حوزه پزشکی و پژوهش‌های تخصصی، مزیت‌هایی برای جمنای در دقت، پوشش یا اعتبار ارجاع‌ها نشان داده‌اند و مرورهای مقایسه‌ای نیز نتایج را وابسته به نوع پرسش و حوزه کاربرد دانسته‌اند (عمر و همکاران، ۲۰۲۵؛ فتاح و همکاران، ۲۰۲۵). مطالعات جدیدتر نیز همین الگو را تأیید می‌کنند: آلپار (۲۰۲۵) تفاوت ابزارها را وابسته به مرحله نگارش و نوع بازخورد دانست؛ دومانیچ و بایجان (۲۰۲۵) نشان دادند قالب سؤال بر دقت و ثبات پاسخ اثر دارد؛ دیریچ و همکاران (۲۰۲۵) اعتماد کاربران را بیش از ویژگی‌های جمعیت‌شناختی به تجربه و احساس مهارت مرتبط دانستند؛ و کاوینکونلاسات (۲۰۲۵) نقش این ابزارها را در بازخورد، انگیزش و نگارش دانشگاهی همراه با چالش‌های اخلاقی و زیرساختی گزارش کرد.

در کنار این شواهد، یافته‌های ران و همکاران (۲۰۲۴ب)، بایتاک (۲۰۲۴)، ران و همکاران (۲۰۲۴الف) و اونو و موریتا (۲۰۲۴) نشان می‌دهد تفاوت چت‌جی‌پی‌تی و جمنای تابع نوع کارکرد است: چت‌جی‌پی‌تی معمولاً در روانی تعامل، بازخورد عملی، سازمان‌دهی پاسخ و شروع فرایند جستجو مزیت دارد، در حالی که جمنای در برخی موقعیت‌ها از نظر تنوع ایده، پوشش اطلاعاتی، تحلیل چندبعدی و پاسخ به نیازهای پژوهشی ظرفیت بیشتری نشان می‌دهد. در سطح زبان‌های غیرانگلیسی، مطالعات نیکلاس و بهاتیا (۲۰۲۳) و اختر و همکاران (۲۰۲۳) بر چالش‌هایی مانند افت دقت معنایی، حذف جزئیات، ضعف در انتقال ظرایف فرهنگی و محدودیت ترجمه تأکید کرده‌اند؛ ازاین‌رو، ارزیابی چت‌جی‌پی‌تی و جمنای در زبان فارسی باید مستقل و زمینه‌مند انجام شود. تلاش‌هایی مانند معرفی معیارهای ویژه برای ارزیابی مدل‌های زبانی فارسی نیز همین ضرورت را تأیید می‌کند (حسینبیگی و همکاران، ۲۰۲۵). بر همین اساس، پژوهش حاضر چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو را در رفع نیازهای اطلاعاتی فارسی‌زبان دانشجویان دکتری دانشگاه خوارزمی مقایسه می‌کند و بر شاخص‌هایی مانند دقت، ربط، شفافیت، تنوع اطلاعات، پوشش موضوعی، کاربردی‌بودن، قابل فهم بودن، تعامل با پرسش‌های پیچیده و تأثیر بر اعتمادبه‌نفس اطلاع‌یابی تمرکز دارد.

روش پژوهش

پژوهش حاضر از نظر هدف، کاربردی و از نظر رویکرد، ترکیبی بود. انتخاب طرح ترکیبی از آن جهت صورت گرفت که مقایسه چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در رفع نیازهای اطلاعاتی دانشجویان، صرفاً با سنجش کمی قابل توضیح نبود و لازم بود تجربه واقعی کاربران، برداشت آنان از کیفیت پاسخ‌ها، شیوه اصلاح پرسش، تردیدها، مزایا و چالش‌های ادراک‌شده نیز بررسی شود. بنابراین، در بخش کمی، عملکرد دو مدل بر اساس ارزیابی زوجی کاربران مقایسه شد و در بخش کیفی، تجربه تعامل کاربران با دو مدل از طریق پروتکل بلنداندیشی، مشاهده و مصاحبه تحلیل گردید. پروتکل بلنداندیشی به پژوهشگر امکان می‌دهد فرایندهای شناختی کاربر هنگام تعامل با سامانه، از جمله تردید، مقایسه، اصلاح پرسش و تشخیص کفایت پاسخ، در زمان انجام فعالیت ثبت شود (ون سومرن و همکاران، ۱۹۹۴).

جامعه پژوهش شامل دانشجویان دکتری دانشکده روان‌شناسی و علوم تربیتی دانشگاه خوارزمی بود. نمونه پژوهش را ۳۶ نفر از دانشجویان تشکیل دادند که به شیوه هدفمند و با توجه به تنوع رشته‌ای انتخاب شدند. معیار اصلی ورود به پژوهش، اشتغال به تحصیل در مقطع دکتری و توانایی طرح نیازهای اطلاعاتی علمی، عمومی و تخصصی به زبان فارسی بود. هر مشارکت‌کننده در یک جلسه کنترل‌شده با هر دو مدل تعامل کرد و پس از انجام وظایف تعیین‌شده، کیفیت پاسخ‌های دریافت‌شده را بر اساس شاخص‌های پژوهش ارزیابی نمود.

برای اجرای پژوهش، ابتدا شرکت‌کنندگان با هدف پژوهش، نحوه انجام جلسه و شیوه کار با ابزارها آشنا شدند. سپس در آزمایشگاه تعامل انسان و اطلاعات دانشگاه خوارزمی، امکانات لازم برای تعامل با دو مدل در اختیار آنان قرار گرفت. اجرای

برنامه کار با مدل ها در مارس تا می ۲۰۲۵ انجام شد، هر برنامه تعامل حدود نیم ساعت با نسخه چت جی پی تی ۴، نسخه پلاس و گوگل جمنای ۲,۵ پرو انجام شد. در زمان اجرای پژوهش، هر دو مدل در محیط کاربری مورد استفاده پژوهشگران با قابلیت دسترسی به اطلاعات آنلاین فعال بودند؛ بااین‌حال، پاسخ‌ها بر اساس تجربه کاربران ارزیابی شد و پژوهش حاضر داوری مستقلی درباره صحت منابع بازیابی شده انجام نداد. همچنین پرسش های طراحی شده برای هر دو مدل کاملا یکسان بود. فرایند تعامل با کمک نرم‌افزار مورائه ثبت شد. این نرم‌افزار امکان ضبط فعالیت‌های کاربر روی صفحه‌نمایش، ثبت کنش‌ها و واکنش‌ها، و مشاهده فرایند تعامل را فراهم کرد. شرکت‌کنندگان ابتدا به پرسش‌های جمعیت‌شناختی و پرسشنامه اولیه پاسخ دادند. سپس وظایف طراحی شده در نرم‌افزار مورائه به‌ترتیب برای آنان نمایش داده شد. بخشی از وظایف شامل پرسش‌های عمومی برای سنجش پاسخ‌گویی مدل‌ها به نیازهای ساده‌تر و بخشی دیگر شامل پرسش‌های تخصصی برگرفته از زمینه پژوهشی یا پروپوزال شرکت‌کنندگان برای بررسی پاسخ‌گویی به نیازهای پیچیده‌تر بود. در حین انجام وظایف، از مشارکت‌کنندگان خواسته شد افکار، تردیدها، دلایل ترجیح، برداشت از کیفیت پاسخ و مشکلات احتمالی خود را با صدای بلند بیان کنند.

ابزارهای گردآوری داده‌ها شامل پرسشنامه اولیه، پرسشنامه مقایسه‌ای دو مدل، پروتکل بلنداندیشی، مصاحبه نیمه ساختاریافته و مشاهده فرایند تعامل بود. پرسشنامه اولیه برای سنجش آمادگی و تجربه اطلاعاتی کاربران و پرسشنامه مقایسه‌ای برای ارزیابی شاخص‌هایی مانند سودمندی، دقت ادراک‌شده، تنوع اطلاعات، کیفیت پاسخ، قابل‌فهم بودن، کاربردی‌بودن و تأثیر بر اعتمادبه‌نفس اطلاع‌یابی به کار رفت. در این مقاله، منظور از دقت، «دقت ادراک شده از سوی کاربران» است؛ زیرا پاسخ‌ها بر اساس ارزیابی مشارکت‌کنندگان و تجربه آنان از تعامل با مدل‌ها سنجیده شد، نه بر اساس داوری مستقل متخصصان درباره صحت علمی تک‌تک پاسخ‌ها.

روایی ابزارها از طریق روایی صوری و محتوایی بررسی شد. پرسشنامه‌ها ابتدا در اختیار استادان راهنما و مشاور و سپس تعدادی از اعضای گروه علم اطلاعات و دانش‌شناسی قرار گرفت و بر اساس نظر آنان برخی گویه‌ها اصلاح، حذف یا اضافه شد. برای اطمینان از فهم‌پذیری ابزار، نسخه اصلاح‌شده پرسشنامه در اختیار تعدادی از دانشجویان قرار گرفت و بازخورد آنان درباره وضوح و قابل‌فهم بودن گویه‌ها در اصلاح نهایی لحاظ شد. پایایی پرسشنامه‌ها با ضریب آلفای کرونباخ سنجیده شد. مقدار آلفا برای پرسشنامه اولیه ۰،۷۰۳، برای بخش مربوط به چت‌جی‌پی‌تی ۴ برابر ۰،۸۸۷ و برای بخش مربوط به جمنای ۲,۵ پرو برابر ۰،۹۲۳ بود که نشان‌دهنده کفایت همسانی درونی ابزارهاست.

برای سنجش پایایی داده‌های کیفی، کدگذاری بخشی از داده‌های حاصل از پروتکل بلنداندیشی توسط دو کدگذار، شامل پژوهشگر و استاد راهنما، به‌صورت مستقل انجام شد. سپس میزان توافق میان کدگذاران با ضریب کاپا محاسبه گردید. مقدار کاپا برای وظایف عمومی ۰،۹۳ و برای وظایف تخصصی ۰،۹۰ به دست آمد که نشان‌دهنده توافق بالا و پایایی قابل قبول فرایند کدگذاری بود.

داده‌های کمی ابتدا از نظر نرمال بودن بررسی شدند. نتایج آزمون‌های کولموگوروف- اسمیرنوف و شاپیرو - ویلک نشان داد که توزیع بیشتر متغیرها نرمال نیست؛ بنابراین، برای مقایسه زوجی دو مدل از آزمون رتبه‌های علامت‌دار ویلکاکسون استفاده شد. تحلیل‌های آماری با نرم‌افزار اس‌پی‌اس‌اس انجام گرفت. داده‌های کیفی حاصل از بلنداندیشی، مشاهده و مصاحبه نیز با استفاده از نرم‌افزار مکس‌کیودا و از طریق کدگذاری باز، محوری و انتخابی تحلیل شدند و مقوله‌های اصلی مربوط به مزایا و چالش‌های ادراک‌شده کاربران از هر دو مدل استخراج گردید.

پرسش‌های پژوهش

۱. چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو از نظر شاخص‌های کمی عملکرد در پاسخ‌گویی به نیازهای اطلاعاتی دانشجویان دکتری چه تفاوتی دارند؟

۲. دانشجویان دکتری چه مزایایی را در استفاده از چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو برای پاسخ‌گویی به نیازهای اطلاعاتی فارسی‌زبان ادراک می‌کنند؟

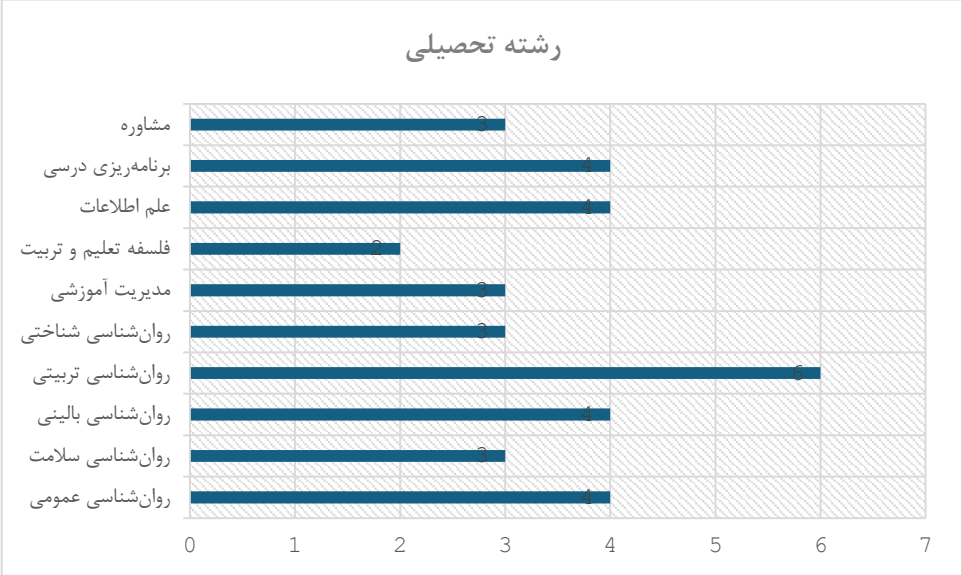
۳. بر اساس یافته‌های کمی و کیفی، هر یک از دو مدل برای کدام نوع نیاز اطلاعاتی دانشجویان دکتری مناسب‌تر ارزیابی می‌شود؟

۴. دانشجویان دکتری در استفاده از چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو به زبان فارسی با چه چالش‌های عملکردی، زیرساختی و کاربرمحور روبه‌رو هستند؟

یافته‌های پژوهش

ویژگی‌های جمعیت‌شناختی مشارکت‌کنندگان

از ۳۶ مشارکت‌کننده، ۲۵ نفر زن و ۱۱ نفر مرد بودند. بیشترین گروه سنی مربوط به ۲۵ تا ۳۴ سال بود (جدول ۱) و رشته‌های تحصیلی مشارکت‌کنندگان شامل روان‌شناسی عمومی، روان‌شناسی سلامت، روان‌شناسی بالینی، روان‌شناسی تربیتی، روان‌شناسی شناختی، مدیریت آموزشی، فلسفه تعلیم و تربیت، علم اطلاعات، برنامه‌ریزی درسی و مشاوره بود. شکل ۱ نشان می‌دهد رشته روان‌شناسی تربیتی بیشترین فراوانی را دارد و توزیع سایر رشته‌ها نسبتاً متوازن است.



شکل ۱. توزیع رشته تحصیلی مشارکت‌کنندگان

در شکل ۱، بیشترین فراوانی مربوط به رشته روان‌شناسی تربیتی است و سایر رشته‌ها توزیعی نسبتاً متوازن دارند؛ بنابراین یافته‌های مقاله باید به‌عنوان تجربه گروهی از دانشجویان دکتری در رشته‌های نزدیک به روان‌شناسی، علوم تربیتی و علم اطلاعات تفسیر شود.

جدول ۱. توزیع سنی مشارکت‌کنندگان

طبقه سن	فراوانی	درصد
۲۵ تا ۳۴ سال	۲۳	۶۳,۹
۳۵ تا ۴۴ سال	۱۲	۳۳,۵
بالای ۴۵ سال	۱	۲,۸

توزیع نمونه پژوهش از نظر جنسیت ۲۵ زن (۶۹,۴٪) و ۱۱ مرد (۳۰,۶٪) و توزیع سن (جدول ۱) همگن نیست و بیشتر مشارکت‌کنندگان در گروه سنی ۲۵ تا ۳۴ سال قرار دارند. این ویژگی برای تفسیر یافته‌ها مهم است، زیرا تجربه استفاده از ابزارهای هوش مصنوعی مولد ممکن است با سن، رشته، میزان تجربه پژوهشی و آشنایی قبلی با فناوری تغییر کند.

۱- مقایسه کمی عملکرد چت‌جی‌پی‌تی ۴ و جمنای ۲,۵ پرو

یافته‌های کمی در جدول ۲ نشان می‌دهد تفاوت دو مدل در بیشتر شاخص‌های ارزیابی از نظر آماری معنادار است. در شاخص عملکرد کلی، اکثریت مشارکت‌کنندگان چت‌جی‌پی‌تی ۴ را بالاتر ارزیابی کردند؛ اما در شاخص‌های دقت، تنوع اطلاعات، پو شش مو ضوعی، کاربردی‌بودن و تعامل با پر سش‌های چندمرحله‌ای، جمنای ۲,۵ پرو مزیت آ شکارتری دا شت. در شاخص قابل‌فهم بودن پاسخ‌ها، تفاوت معناداری میان دو مدل مشاهده نشد. بنابراین، یافته‌های کمی از یک الگوی دوگانه حمایت می‌کند: چت‌جی‌پی‌تی ۴ در تجربه کلی و اعتماد کاربر، و جمنای ۲,۵ پرو در کیفیت محتوایی و پوشش اطلاعاتی قوی‌تر ظاهر شده است.

جدول ۲ مهم‌ترین شواهد کمی مقاله را نشان می‌دهد. جهت رتبه‌ها بیانگر آن است که ترجیح کاربران در شاخص‌های مختلف تغییر می‌کند؛ یعنی چت‌جی‌پی‌تی ۴ در تجربه کلی و اعتمادبه‌نفس اطلاع‌یابی مزیت دارد، اما جمنای ۲,۵ پرو در دقت علمی، تنوع اطلاعات و پوشش موضوعی قوی‌تر ارزیابی شده است.

جدول ۲. خلاصه نتایج کمی مقایسه دو مدل بر اساس آزمون ویلکاکسون

شاخص مقایسه	شواهد آماری اصلی	آماره Z	معناداری	نتیجه تفسیری
عملکرد کلی	۳۱ رتبه به نفع چت‌جی‌پی‌تی ۴، ۳ رتبه به نفع جمنای، ۲ تساوی	-۴,۵۴۳	۰,۰۰۱>	برتری کلی چت‌جی‌پی‌تی ۴
دقت ادراک‌شده پاسخ‌ها از دیدگاه کاربران	۲۷رتبه به نفع جمنای، ۱ رتبه به نفع چت‌جی‌پی‌تی ۴، ۸ تساوی	-۴,۰۹۷	۰,۰۰۱>	برتری جمنای ۲,۵ پرو
رفع نیازهای اطلاعاتی	تفاوت معنادار میان دو مدل در شاخص‌های عملیاتی	-۴,۰۹۷	۰,۰۰۱>	تفاوت معنادار در ارزیابی کاربران از کارآمدی دو مدل
تعامل با پرسش‌های چندمرحله‌ای	۳۰رتبه به نفع جمنای، ۵ رتبه به نفع چت‌جی‌پی‌تی ۴، ۱ تساوی	-۴,۰۹۷	۰,۰۰۱	برتری جمنای در نیازهای پیچیده
دقت در حوزه‌های علمی، عمومی و تخصصی	۲۷رتبه به نفع جمنای، ۱ رتبه به نفع چت‌جی‌پی‌تی ۴، ۸ تساوی	-۴,۱۹۹	۰,۰۰۱>	برتری جمنای در دقت حوزه‌ای
تنوع اطلاعات	۳۱رتبه به نفع جمنای، ۳ رتبه به نفع چت‌جی‌پی‌تی ۴، ۲ تساوی	-۴,۷۲۲	۰,۰۰۱>	برتری جمنای در تنوع دیدگاه‌ها
پوشش ابعاد موضوعی	۳۱رتبه به نفع جمنای، ۳ رتبه به نفع چت‌جی‌پی‌تی ۴، ۲ تساوی	-۴,۷۲۲	۰,۰۰۱>	برتری جمنای در پوشش موضوع
قابل‌فهم بودن پاسخ‌ها	۹رتبه منفی، ۱۳ رتبه مثبت، ۱۴ تساوی	-۱,۳۲۶	۰,۱۸۵	تفاوت معنادار نیست
کاربردی‌بودن پاسخ‌ها	۲۵رتبه به نفع جمنای، ۴ رتبه به نفع چت‌جی‌پی‌تی ۴، ۷ تساوی	-۳,۵۹۴	۰,۰۰۱>	برتری جمنای در کاربردی‌بودن
افزایش اعتمادبه‌نفس	تفاوت معنادار طبق تفسیر پایان‌نامه	-۳,۱۳۵	۰,۰۰۲	برتری ادراک‌شده چت‌جی‌پی‌تی ۴

مقایسه داده های جدول ۲ با مطالعات جدید نشان می‌دهد نتیجه پژوهش حاضر با اصل زمینه‌مندی عملکرد مدل‌های زبانی[4]سازگار است. در پژوهش لنگ و همکاران (۲۰۲۴)، جی پی تی ۴ در تولید موارد آموزشی از نظر ارزیابی ذهنی بهتر از جمنای گزارش شده است، درحالی‌که عمر و همکاران (۲۰۲۵) در زمینه تولید پژوهش پزشکی، مزیت جمنای را در دقت و اعتبار ارجاعات نشان داده‌اند. همچنین مرور فتاح و همکاران (۲۰۲۵) نشان می‌دهد در برخی مطالعات پزشکی، جی پی تی ۴ از نظر دقت و کوتاهی پاسخ بهتر ظاهر شده است. بنابراین، یافته‌های جدول ۲ را نباید به‌عنوان برتری مطلق یکی از دو مدل تفسیر کرد، بلکه باید آن را نشانه تفاوت کارکردی مدل‌ها بر اساس نوع نیاز اطلاعاتی دانست.

مفهوم «توانایی تعامل در درک پرسش‌های پیوسته و پیچیده» به‌صورت عملیاتی تعریف و تحلیل شد. در این پژوهش، تعامل به توانایی مدل در حفظ زمینه گفت‌وگو، درک وابستگی میان پیام‌های متوالی و ارائه پاسخ منسجم در شرایط چندمرحله‌ای اطلاق شده است. برای سنجش این مفهوم، شاخص‌های حفظ زمینه، انسجام در پاسخ‌های چندمرحله‌ای، درک وابستگی بین پیام‌ها، توانایی بازسازی درخواست‌های چندبخشی و نیاز به تکرار یا اصلاح پرسش در نظر گرفته شد.

برای بررسی توانایی دو مدل در تعامل با پرسش‌های چندمرحله‌ای و چندبخشی، شاخص‌هایی مانند حفظ زمینه گفت‌وگو، انسجام پاسخ در مراحل متوالی، درک وابستگی میان پیام‌ها، بازسازی درخواست‌های چندبخشی و کاهش نیاز به تکرار یا اصلاح پرسش در نظر گرفته شد. نتایج آزمون رتبه‌های علامت‌دار ویلکاکسون نشان داد که بین چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در این شاخص تفاوت معنادار وجود دارد. مقدار آماره z آزمون برابر با ۴٫۰۹۷- و سطح معناداری برابر با ۰٫۰۰۱ گزارش شد. همچنین، مقایسه رتبه‌ها نشان داد که ۳۰ رتبه به نفع جمنای ۲٫۵ پرو، ۵ رتبه به نفع چت‌جی‌پی‌تی ۴ و یک مورد تساوی بوده است. بنابراین، جمنای ۲٫۵ پرو از دیدگاه مشارکت‌کنندگان در حفظ زمینه، پاسخ‌گویی منسجم به پرسش‌های پیوسته و مدیریت درخواست‌های چندبخشی عملکرد مطلوب‌تری داشته است.

برای بررسی تفاوت دو مدل از نظر دقت پاسخ‌ها در حوزه‌های علمی، عمومی و تخصصی، میانگین ارزیابی کاربران از پاسخ‌های هر حوزه با استفاده از آزمون ویلکاکسون مقایسه شد. نتایج نشان داد که دقت ادراک‌شده پاسخ‌های جمنای ۲٫۵ پرو به‌طور معناداری بالاتر از چت‌جی‌پی‌تی ۴ ارزیابی شده است. این یافته نشان می‌دهد که از دیدگاه مشارکت‌کنندگان، جمنای ۲٫۵ پرو در ارائه پاسخ‌های دقیق‌تر در موقعیت‌های اطلاعاتی گوناگون، به‌ویژه در پرسش‌های علمی و تخصصی، عملکرد مطلوب‌تری داشته است. با این حال، این نتیجه باید در چارچوب ارزیابی کاربران تفسیر شود و به معنای داوری عینی درباره صحت علمی همه پاسخ‌ها نیست. نتایج آزمون ویلکاکسون نشان داد که بین دو مدل از نظر تنوع اطلاعات ارائه‌شده تفاوت معناداری وجود دارد. بر اساس ارزیابی کاربران، جمنای ۲٫۵ پرو در بیشتر موارد اطلاعات متنوع‌تر، ابعاد بیشتر و دیدگاه‌های گسترده‌تری نسبت به چت‌جی‌پی‌تی ۴ ارائه کرده است. این یافته نشان می‌دهد که جمنای ۲٫۵ پرو برای موقعیت‌هایی که کاربر به دریافت چند زاویه تحلیلی، اطلاعات مکمل و گسترش دامنه موضوع نیاز دارد، از نظر کاربران کارآمدتر ارزیابی شده است.

بررسی رتبه‌ها و نتایج آزمون ویلکاکسون نشان داد که جمنای ۲٫۵ پرو در پوشش گسترده‌تر ابعاد موضوعی نسبت به چت‌جی‌پی‌تی ۴ برتری معنادار داشته است. به بیان دیگر، مشارکت‌کنندگان پاسخ‌های جمنای را از نظر جامعیت، توجه به جنبه‌های مختلف مسئله و پوشش موضوعی گسترده‌تر ارزیابی کردند. بنابراین، در نیازهای اطلاعاتی‌ای که مستلزم بررسی چندبعدی موضوع، مقایسه دیدگاه‌ها و دریافت پاسخ گسترده‌تر است، جمنای ۲٫۵ پرو عملکرد مناسب‌تری نشان داده است.

بر اساس داده های جدول ۲، نتایج آزمون ویلکاکسون نشان می‌دهد که جمنای ۲٫۵ پرو از نظر درک و پاسخ‌گویی به پرسش‌های پیچیده و چندبخشی عملکرد مطلوب‌تری داشته است. این یافته نشان می‌دهد که در برخی موقعیت‌های تعاملی، به‌ویژه زمانی که پرسش نیازمند پیگیری منطقی، تحلیل چندبخشی و حفظ انسجام پاسخ است، جمنای ۵.۲ پرو از دیدگاه کاربران توانسته است تجربه روان‌تر و منسجم‌تری فراهم کند.

نتایج نشان داد که از نظر قابل‌فهم بودن پاسخ‌ها، تفاوت معناداری میان چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو مشاهده نشد. بنابراین، از دیدگاه دانشجویان دکتری، هر دو مدل توانسته‌اند پاسخ‌هایی نسبتاً قابل‌فهم و قابل استفاده در زبان فارسی ارائه کنند. با این حال، در شاخص کاربردی‌بودن، تفاوت معنادار به نفع جمنای ۲٫۵ پرو گزارش شد. این یافته نشان می‌دهد که اگرچه

[4] The context-dependent nature of large language model performance

پاسخ‌های دو مدل از نظر فهم‌پذیری نزدیک به هم ارزیابی شده‌اند، پاسخ‌های جمنای از نظر قابلیت استفاده عملی در فرایند جستجو، پژوهش و حل مسئله اطلاعاتی مفیدتر تلقی شده‌اند. نتایج آزمون ویلکاکسون نشان داد که تفاوت دو مدل از نظر تأثیر بر اعتمادبه‌نفس کاربران در فرایند جستجوی اطلاعات معنادار است. بر اساس یافته‌ها، چت‌جی‌پی‌تی ۴ بیش از جمنای ۲٫۵ پرو در تقویت احساس اطمینان، کنترل و اعتمادبه‌نفس دانشجویان هنگام جستجوی اطلاعات مؤثر ارزیابی شده است. این نتیجه نشان می‌دهد که برتری یک مدل صرفاً به دقت یا جامعیت پاسخ محدود نمی‌شود؛ بلکه سبک گفت‌وگو، روانی تعامل، سرعت دریافت پاسخ و احساس تسلط کاربر نیز در تجربه اطلاع‌یابی نقش مهمی دارد.

۲- مزایای ادراک‌شده کاربران

یافته‌های کیفی، الگوی کمی را تکمیل می‌کند. مطابق جدول ۳، کاربران جمنای ۲٫۵ پرو را بیشتر با عمق محتوایی، جامعیت، جزئیات بیشتر، تنوع دیدگاه، سازمان‌دهی مناسب محتوا و پشتیبانی از فرایند تحقیق پیوند داده‌اند. در مقابل، چت‌جی‌پی‌تی ۴ بیشتر با سهولت تعامل، سرعت پاسخ، مدیریت گفت‌وگو، اختصار متناسب، مثال‌محوری و سازمان‌دهی پاسخ توصیف شده است. این تفاوت نشان می‌دهد ارزیابی کاربران فقط تابع صحت پاسخ نیست، بلکه سبک تعامل و احساس کنترل کاربر بر فرایند جستجو نیز بر پذیرش مدل اثر می‌گذارد.

جدول ۳. مقایسه مزایای ادراک‌شده کاربران از دو مدل

مدل	محور اصلی	کد محوری	کدهای باز منتخب
جمنای ۲٫۵ پرو	تعاملی	سهولت و روانی تعامل	دسترسی آسان، سرعت بالا، سازمان‌دهی مناسب محتوا، وضوح پاسخ‌ها
جمنای ۲٫۵ پرو	کیفیت پاسخ‌دهی	عمق و غنای محتوایی	دقت بالا، جامعیت، جزئیات بیشتر، پایگاه دانشی گسترده
جمنای ۲٫۵ پرو	کاربردپذیری	پشتیبانی از فرایند تحقیق	ارائه راهکار عملی، چندبعدی بودن پاسخ، تنوع دیدگاه
چت‌جی‌پی‌تی ۴	تعاملی	سهولت تعامل کاربر	سرعت بالا، مدیریت گفت‌وگو، اختصار متناسب پاسخ
چت‌جی‌پی‌تی ۴	کیفیت پاسخ‌دهی	دقت و انسجام پاسخ	وضوح پاسخ، مثال‌محوری، نمایش گرافیکی و جذابیت بصری
چت‌جی‌پی‌تی ۴	کاربردپذیری	سازمان‌دهی و پشتیبانی علمی	سازمان‌دهی محتوا، نتیجه‌گیری مدل‌محور، ارائه چند دیدگاه

جدول ۳ نشان می‌دهد که مزیت‌های ادراک‌شده فقط به کیفیت محتوای پاسخ محدود نیستند. کاربران هنگام قضاوت درباره مدل، همزمان روانی تعامل، سرعت، نظم پا سخ، عمق محتوا و قابلیت ا ستفاده در کار پژوه شی را در نظر گرفته‌اند. این نکته برای آموزش سواد هوش م صنوعی اهمیت دارد، زیرا تجربه خو شایند گفت‌وگو الزاماً به معنای دقت علمی بی شتر نی ست. اظهارات م شارکت‌کنندگان ن شان داد که مزیت‌های دو مدل از نظر کاربران، صرفاً به کیفیت محتوای پا سخ محدود نبود، بلکه سهولت د سترسی، روانی تعامل، سرعت، دقت ادراک شده، جامعیت و تناسب با کار علمی نیز در قضاوت آنان نقش داشت. یکی از م شارکت‌کنندگان درباره جمنای ۲٫۵ پرو اظهار کرد: «از نظر د ستر سی، کار با جمنای ۲٫۵ پرو خیلی آ سان بود و ما خیلی راحت می‌توانستیم با سرعت بالا از آن استفاده کنیم»(مشارکت کننده ۱۲)؛ در حالی که درباره چت‌جی‌پی‌تی ۴ افزود: «چت‌جی‌پی‌تی ۴ تعاملی‌تر بود و مدیریت گفت‌وگو در آن راحت‌تر انجام می‌شد»(مشارکت کننده ۸). این نقل‌قول نشان می‌دهد که کاربران، جمنای را بی شتر از منظر د ستر سی و سرعت، و چت‌جی‌پی‌تی را بی شتر از منظر مدیریت گفت‌وگو و تعامل تجربه کرده‌اند.

در مجموع، نقل‌قول‌های کاربران نشان می‌دهد که تجربه استفاده از دو مدل، یک تجربه یکدست و تک‌بعدی نبوده است. برخی کاربران جمنای ۲٫۵ پرو را به دلیل گستردگی، دقت ادراک‌شده، پشتیبانی از زبان فارسی، توضیح اصطلاحات تخصصی و ارائه راهنمایی‌های کاربردی ترجیح داده‌اند؛ در حالی که چت‌جی‌پی‌تی ۴ بیشتر با چارچوب‌مندی، روانی تعامل، مدیریت بهتر گفت‌وگو و پاسخ‌های منظم پیوند خورده است. یکی از مشارکت‌کنندگان این تفاوت را چنین جمع‌بندی کرد: «به نظر من جمنای از نظر گسترده‌تر عمل کردن بهتر بود، اما چت‌جی‌پی‌تی هم کیفیت پاسخ‌گویی خوبی داشت. در کل به رفتار و عادت فرد بستگی دارد و هرکدام از این مدل‌ها می‌توانند مبنا قرار بگیرند و نیازهای پژوهشی را تا حد مورد انتظار تأمین کنند» (مشارکت کننده ۳). این برداشت، این ایده را تقویت می‌کند که دو مدل نه جایگزین کامل یکدیگر، بلکه ابزارهایی مکمل برای موقعیت‌های متفاوت اطلاع‌یابی هستند.

جدول ۳ نشـــان می‌دهد که ادراک مثبت کاربران از جمنای ۲٫۵ پرو بیشـــتر از مســـیر عمق پاســـخ، جزئیات، چندبعدی‌بودن و پ شتیبانی از فرایند تحقیق شکل گرفته ا ست. برتری جمنای در دقت و پو شش مو ضوعی ا ست. کدهای باز ا ستخراج شده حول محورهای تعاملی، کیفیت پا سخ‌دهی و کاربردپذیری سازمان یافته‌اند و ن شان می‌دهند کاربران مزیت‌ها را همزمان از منظر محتوایی و تجربه کاربری ارزیابی کرده‌اند. اظهارات مشـارکت‌کنندگان همچنین نشـان داد که جمنای ۲٫۵ پرو در تجربه برخی کاربران با دقت، جامعیت و تناسب با کار علمی پیوند خورده است. یکی از مشارکت‌کنندگان بیان کرد: «جمنای ۲٫۵ پرو هم دقت بالایی داشت و هم مناسب کار علمی بود؛ پاسخ‌هایش خوب بود و من از آن استفاده می‌کردم». مشارکت‌کننده دیگری نیز تأکید کرد: «جمنای کیفیت پاســخ‌گویی خیلی خوبی داشــت؛ هم جامع بود، هم جزئیات را نگه می‌داشــت و زبان فارســی را خیلی خوب پشــتیبانی می‌کرد، به‌ویژه در برابر اصــطلاحات تخصــصــی علمی». این اظهارات با یافته‌های کمی درباره برتری جمنای در دقت ادراک‌شده، تنوع اطلاعات، پوشش موضوعی و کاربردی‌بودن پاسخ‌ها همسو است.

مزیت اصلی چت‌جی‌پی‌تی ۴ در تجربه ارتباطی و سازمان‌دهی پاسخ نمایان شده است. تراکم کدهای مربوط به سهولت تعامل، مدیریت گفت‌وگو، مثال‌محوری و اختصار مناسب نشان می‌دهد این مدل برای مرحله اکتشاف، روشن‌سازی مسئله و کاهش ابهام اولیه برای دانشجو کارکرد برجسته‌تری دارد. کدهای باز استخراج‌شده حول محورهای تعاملی، کیفیت پاسخ‌دهی و کاربردپذیری سازمان یافته‌اند و نشان می‌دهند کاربران مزیت‌ها را همزمان از منظر محتوایی و تجربه کاربری ارزیابی کرده‌اند. بخشی از کاربران چت‌جی‌پی‌تی ۴ را به دلیل ساختار پاسخ، چارچوب‌مندی و مدیریت بهتر گفت‌وگو مطلوب ارزیابی کردند. یکی از مشارکت‌کنندگان اظهار کرد: «چت‌جی‌پی‌تی هم کیفیت پاسخ‌گویی خوبی داشت و چارچوب آن برای من جالب بود؛ به نظر من خیلی چارچوب‌مند بود و سعی می‌کرد دستوراتی را که به آن داده می‌شد با فرمت مناسبی رعایت کند»(مشارکت کننده ۲۱). این برداشت نشان می‌دهد که مزیت چت‌جی‌پی‌تی ۴ در تجربه کاربران بیشتر به سازمان‌دهی پاسخ، نظم تعاملی و ایجاد حس کنترل در فرایند گفت‌وگو مربوط بوده است.

۳- چالش‌های ادراک‌شده کاربران

بر اساس داده های جدول ۴، چالش‌های کاربران در سه سطح عملکردی، زیرساختی و کاربرمحور قرار می‌گیرد. در سطح عملکردی، محدودیت در کیفیت پاسخ، کلی‌گویی، ضعف در تنوع اطلاعات، اختصار بیش از حد، نقدگرایی محدود و سوگیری الگوریتمی ادراک‌شده مطرح شده است. در سطح زیرساختی، کندی یا قطعی اینترنت، خطاهای سیستمی، محدودیت IP ایران و نیاز به فیلترشکن از مهم‌ترین موانع بوده‌اند. در سطح کاربرمحور نیز فشار شناختی، سردرگمی در تشخیص خروجی، نیاز به تمرکز بالا و وابستگی شناختی به داده‌های تولیدشده مشاهده شد. این یافته‌ها با مرورهای جدید درباره چالش‌های مدل‌های زبانی در آموزش عالی، از جمله اتکای بیش از حد، مسائل فنی و نگرانی درباره اعتبار پاسخ‌ها، همسو است (شهزاد و همکاران، ۲۰۲۴؛ شی و همکاران، ۲۰۲۵).

جدول ۴. مقایسه چالش‌های ادراک‌شده کاربران در استفاده از دو مدل

مدل	محور اصلی	کد محوری چالش	کدهای باز منتخب
جمنای ۲٫۵ پرو	عملکردی	محدودیت در کیفیت پاسخ	اختصار بیش از حد، کلی‌بودن پاسخ‌ها، ضعف در تنوع اطلاعات در برخی موارد
جمنای ۲٫۵ پرو	زیرساختی	مشکلات اتصال و دسترسی	کندی اینترنت، قطعی اینترنت، خطای سیستمی، محدودیت IP ایران
جمنای ۲٫۵ پرو	کاربرمحور	فشار شناختی و سردرگمی	نیاز به تمرکز بالا، حجم زیاد پاسخ، ابهام در تشخیص خروجی
چت‌جی‌پی‌تی ۴	عملکردی	محدودیت محتوایی و تحلیلی	کلی‌گویی، ضعف در تنوع اطلاعات، نقدگرایی محدود، سوگیری الگوریتمی ادراک‌شده
چت‌جی‌پی‌تی ۴	زیرساختی	محدودیت دسترسی	کندی یا قطعی اینترنت، ضرورت استفاده از فیلترشکن
چت‌جی‌پی‌تی ۴	کاربرمحور	پیچیدگی تجربه و وابستگی شناختی	گیج‌کنندگی برخی ابزارها، حدس کاربر در تفسیر پاسخ، وابستگی به داده‌ها

تحلیل کیفی داده‌ها نشان داد که کاربران در استفاده از چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو به زبان فارسی با سه دسته چالش روبه‌رو بوده‌اند: چالش‌های عملکردی، چالش‌های زیرساختی و چالش‌های کاربرمحور (جدول ۵). در سطح عملکردی، مواردی مانند کلی‌گویی، اختصار بیش از حد، ضعف در تنوع اطلاعات، محدودیت در نقد تحلیلی، ناپایداری پاسخ و دشواری در دریافت خروجی قطعی گزارش شد. در سطح زیرساختی، کندی یا قطعی اینترنت، خطاهای سیستمی، محدودیت دسترسی، نیاز به فیلترشکن و محدودیت‌های منطقه‌ای از مهم‌ترین موانع تجربه کاربران بود. در سطح کاربرمحور نیز فشار شناختی، نیاز به تمرکز بالا برای فهم پاسخ‌ها، سردرگمی در تشخیص کفایت خروجی و وابستگی ذهنی به پاسخ تولیدشده مشاهده شد. این یافته‌ها نشان می‌دهد که چالش استفاده از مدل‌های زبانی در بافت فارسی فقط به کیفیت فنی مدل مربوط نیست، بلکه از تعامل میان توان مدل، کیفیت زیرساخت، محدودیت‌های دسترسی، مهارت پرامپت‌نویسی و توان کاربر در ارزیابی خروجی شکل می‌گیرد.

در کنار مزیت‌های محتوایی جمنای، حجم زیاد پاسخ و نیاز به تمرکز برای تشخیص کفایت خروجی می‌تواند فشار شناختی ایجاد کند. این یافته بیانگر آن است که پاسخ جامع، در نبود مهارت ارزیابی و گزینش اطلاعات، ممکن است به دشواری تصمیم‌گیری علمی منجر شود. بنابراین محدودیت‌های ادراک‌شده فقط ناشی از مدل نیست، بلکه حاصل تعامل میان کیفیت پاسخ، زیرساخت دسترسی و مهارت کاربر در تفسیر خروجی است.

بر اساس داده های جدول شماره ۴، چالش اصلی چت‌جی‌پی‌تی ۴ در تجربه کاربران بیشتر به کلی‌گویی، محدودیت نقد تحلیلی و وابستگی کاربر به پاسخ آماده مربوط است. روانی پاسخ می‌تواند اعتماد سریع ایجاد کند، اما همین اعتماد اگر با راستی‌آزمایی همراه نشود، خطر پذیرش غیرانتقادی خروجی را افزایش می‌دهد. این شکل نشان می‌دهد که محدودیت‌های ادراک‌شده فقط ناشی از مدل نیست، بلکه حاصل تعامل میان کیفیت پاسخ، زیرساخت دسترسی و مهارت کاربر در تفسیر خروجی است.

بحث و نتیجه‌گیری

نتایج این پژوهش نشان داد که چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در رفع نیازهای اطلاعاتی دانشجویان دکتری فارسی‌زبان دو کارکرد متفاوت و مکمل دارند. چت‌جی‌پی‌تی ۴ در عملکرد کلی ادراک‌شده، روانی تعامل، نظم‌دهی پاسخ، شروع جستجو و افزایش اعتمادبه‌نفس اطلاع‌یابی مزیت داشت؛ در مقابل، جمنای ۲٫۵ پرو در دقت ادارک شده در خصوص پاسخ‌های علمی، تنوع اطلاعات، پوشش ابعاد موضوع، کاربردی‌بودن و پاسخ به نیازهای پیچیده و چندمرحله‌ای قوی‌تر ارزیابی شد. بنابراین، نتیجه اصلی مقاله این نیست که یکی از دو مدل به‌طور مطلق برتر است، بلکه این است که انتخاب مدل باید تابع نوع نیاز اطلاعاتی، مرحله جستجو، سطح تخصص پرسش و میزان آمادگی کاربر برای ارزیابی انتقادی خروجی باشد (لنگ و همکاران، ۲۰۲۴؛ عمر و همکاران، ۲۰۲۵؛ فتاح و همکاران، ۲۰۲۵).

از منظر رفتار اطلاع‌یابی، چت‌جی‌پی‌تی ۴ را می‌توان ابزار مناسب مرحله اکتشاف اولیه دانست؛ یعنی زمانی که دانشجو هنوز مسئله خود را دقیق صورت‌بندی نکرده، به توضیح روان، مثال، ساختار اولیه و کاهش ابهام نیاز دارد. برتری این مدل در اعتمادبه‌نفس اطلاع‌یابی نیز نشان می‌دهد که کیفیت گفت‌وگو و احساس کنترل کاربر بر تعامل، بخشی از موفقیت جستجوی دانشگاهی است (شهزاد و همکاران، ۲۰۲۴؛ اوپن‌ای‌آی، ۲۰۲۴). بااین‌حال، همین ویژگی می‌تواند خطر اتکای بیش از حد و پذیرش سریع پاسخ را افزایش دهد؛ ازاین‌رو، استفاده از آن باید با آموزش مهارت ارزیابی و راستی‌آزمایی همراه باشد.

از منظر نیازهای تخصصی و پژوهشی، به نظر می رسد که جمنای ۲٫۵ پرو برای پرسش‌هایی مناسب‌تر است که به پوشش گسترده‌تر، مقایسه دیدگاه‌ها، حفظ زمینه در پرسش‌های چندبخشی و پاسخ کاربردی‌تر نیاز دارند. این تفسیر با گزارش‌های فنی درباره توانایی مدل‌های جمنای در پردازش زمینه‌های طولانی و استدلال پیچیده همخوان است (تیم جمنای و همکاران، ۲۰۲۴؛ گوگل، ۲۰۲۵). در عین حال، جامعیت بیشتر پاسخ همواره به معنای کفایت علمی نیست و ممکن است فشار شناختی کاربر را افزایش دهد؛ بنابراین، جمنای نیز باید به‌عنوان دستیار پژوهشی اولیه، نه مرجع نهایی، به کار رود.

یافته مربوط به پرسش‌های چندمرحله‌ای و چندبخشی نشان می‌دهد که جمنای ۲٫۵ پرو در موقعیت‌هایی که نیازمند حفظ زمینه، پیگیری وابستگی میان پیام‌ها و پاسخ‌گویی به درخواست‌های چندلایه است، از دیدگاه کاربران عملکرد بهتری داشته

است. این نتیجه با الگوی کلی یافته‌های مقاله هماهنگ است؛ زیرا جمنای ۲٫۵ پرو در شاخص‌هایی مانند دقت ادراک‌شده، تنوع اطلاعات، پوشش ابعاد موضوع و کاربردی‌بودن پاسخ‌ها نیز برتری نشان داده است. بنابراین، می‌توان نتیجه گرفت که این مدل برای نیازهای اطلاعاتی پیچیده‌تر، پژوهشی‌تر و چندمرحله‌ای مناسب‌تر ارزیابی می‌شود؛ در حالی که چت‌جی‌پی‌تی ۴ همچنان در آغاز جستجو، نظم‌بخشی اولیه، روانی تعامل و افزایش اعتمادبه‌نفس اطلاع‌یابی نقش مکمل دارد.

یافته‌های پژوهش نشان داد که جمنای ۲٫۵ پرو در دقت ادراک‌شده، تنوع اطلاعات، پوشش ابعاد موضوعی و کاربردی‌بودن پاسخ‌ها نسبت به چت‌جی‌پی‌تی ۴ برتری داشته است. این نتیجه بیانگر آن است که جمنای در موقعیت‌هایی که دانشجو به پاسخ‌های گسترده‌تر، چندبعدی‌تر و قابل استفاده‌تر در کار پژوهشی نیاز دارد، می‌تواند ابزار مناسب‌تری باشد. این یافته تا حدی با پژوهش‌هایی همسو است که بر توانایی جمنای در ارائه اطلاعات متنوع، پوشش گسترده‌تر و عملکرد مناسب در برخی حوزه‌های تخصصی تأکید کرده‌اند؛ از جمله پژوهش ساب، ونگ، استوتز و همکاران (۲۰۲۴)

درباره قابلیت‌های جمنای در کاربردهای پزشکی و پژوهش فتاح، صالح و همکاران (۲۰۲۵) که نشان داد عملکرد مدل‌ها به نوع پرسش و حوزه کاربرد وابسته است. با وجود این، نتایج پژوهش حاضر در برخی موارد با مطالعات دیگر همسو نیست. برای نمونه، سلمان و همکاران (۲۰۲۵) در حوزه داروشناسی قلبی ـ عروقی، چت‌جی‌پی‌تی ۴ را از نظر دقت پاسخ‌گویی به پرسش‌های چندگزینه‌ای و کوتاه‌پاسخ برتر گزارش کرده‌اند. همچنین استرژالکوفسکی و همکاران (۲۰۲۴) در موضوع جداشدگی شبکیه نشان داده‌اند که چت‌جی‌پی‌تی ۴ در برخی پرسش‌های دشوارتر از نظر درستی، انسجام و کیفیت کلی عملکرد بهتری از جمنای داشته است. این تفاوت‌ها نشان می‌دهد که عملکرد مدل‌های زبانی بزرگ ماهیتی زمینه‌مند دارد و به زبان، حوزه موضوعی، نوع پرسش، سطح تخصص کاربر و معیار ارزیابی وابسته است.

از سوی دیگر، چت‌جی‌پی‌تی ۴ در پژوهش حاضر از نظر افزایش اعتمادبه‌نفس اطلاع‌یابی کاربران عملکرد مطلوب‌تری داشته است. این یافته نشان می‌دهد که کیفیت تجربه کاربر تنها به محتوای پاسخ وابسته نیست، بلکه روانی تعامل، سرعت، سازمان‌دهی پاسخ و ایجاد حس کنترل در کاربر نیز بر پذیرش و استفاده از مدل اثر دارد. این نتیجه با پژوهش توریبیو (۲۰۲۳) همسو است؛ زیرا وی نشان داد استفاده از چت‌جی‌پی‌تی و سایر ابزارهای هوش مصنوعی، از طریق ارائه راهنمایی مرحله‌به‌مرحله، توضیح‌های فوری و حمایت از یادگیری خودتنظیم، می‌تواند اضطراب تحصیلی را کاهش دهد و احساس خودکارآمدی و اعتمادبه‌نفس یادگیرندگان را تقویت کند.

چالش‌های شناسایی‌شده در استفاده از دو مدل به زبان فارسی نشان می‌دهد که تجربه کاربران فارسی‌زبان صرفاً تابع توانایی زبانی یا فنی مدل نیست. حتی زمانی که پاسخ‌ها از نظر محتوایی مفید ارزیابی می‌شوند، عواملی مانند محدودیت دسترسی، ناپایداری اینترنت، نیاز به فیلترشکن، ابهام در تشخیص صحت پاسخ، فشار شناختی و ضعف در راستی‌آزمایی منابع می‌تواند کیفیت تجربه کاربر را کاهش دهد. از این نظر، استفاده دانشگاهی از چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در زبان فارسی باید همراه با سواد اطلاعاتی، سواد هوش مصنوعی، مهارت طراحی پرامپت و توان ارزیابی انتقادی خروجی باشد.

این پژوهش با یافته‌های پژوهش، یاری و احمدی (۱۳۹۳) هم راستا است چرا که چالش‌هایی نظیر محدودیت‌های زیرساختی، فیلترینگ برخی منابع اطلاعاتی، و عدم آگاهی از تکنیک‌های جستجوی پیشرفته، موانع مهمی در رفتار جستجوی اطلاعات کاربران محسوب می‌شوند اهمیت این نکته در بافت فارسی بیشتر است، زیرا ارزیابی مدل‌های زبانی در زبان فارسی باید با معیارهای مستقل و زمینه‌مند انجام شود و نمی‌توان نتایج ارزیابی‌های انگلیسی‌محور را بی‌واسطه به کاربران فارسی‌زبان تعمیم داد (حسین‌بیگی و همکاران، ۲۰۲۵).

بر اساس یافته های این پژوهش (جدول ۵) مشخص شد که چت‌جی‌پی‌تی ۴ برای آغاز جستجو، تولید ایده، روشن‌سازی مسئله و دریافت توضیح قابل‌فهم؛ و جمنای ۲٫۵ پرو برای گسترش ابعاد موضوع، مقایسه دیدگاه‌ها، پرسش‌های تخصصی و سنجش کاربردپذیری پاسخ بهتر عمل می کند. با این حال خروجی مدل باید با پایگاه‌های علمی، منابع معتبر، داوری متخصص و قضاوت پژوهشگر کنترل شود. بنابراین، سواد هوش مصنوعی، سواد اطلاعاتی، توان طراحی پرامپت و ارزیابی انتقادی خروجی باید بخشی از آموزش پژوهش در تحصیلات تکمیلی باشد (شی و همکاران، ۲۰۲۵؛ عمران و المشرف، ۲۰۲۴). در بخش کیفی، اظهارات مشارکت‌کنندگان نشان داد که مزیت‌های ادراک‌شده دو مدل ماهیتی متفاوت دارد.

جدول ۵. جمع‌بندی کاربردی انتخاب مدل بر اساس نوع نیاز اطلاعاتی

نوع نیاز اطلاعاتی/کاربرد	مدل مناسب‌تر بر اساس یافته‌ها	دلیل اصلی
شروع جستجو و روشن‌سازی مسئله	چت‌جی‌پی‌تی ۴	روانی تعامل، سازمان‌دهی پاسخ و افزایش اعتمادبه‌نفس
پرسش تخصصی و دقیق	جمنای ۲٫۵ پرو	دقتادارک شده بالاتر و پوشش موضوعی بیشتر در داده‌های پژوهش
نیاز چندمرحله‌ای و چندبخشی	جمنای ۲٫۵ پرو	توان بهتر در حفظ زمینه و پاسخ به پرسش پیچیده
توضیح سریع و قابل‌فهم	هر دو مدل	تفاوت معنادار در قابل‌فهم بودن مشاهده نشد
استفاده پژوهشی و مقایسه دیدگاه‌ها	جمنای ۲٫۵ پرو، همراه با راستی‌آزمایی	تنوع اطلاعات و جامعیت بیشتر
افزایش اطمینان کاربر در جستجو	چت‌جی‌پی‌تی ۴	تأثیر مثبت‌تر بر اعتمادبه‌نفس ادراک‌شده

درباره جمنای ۲٫۵ پرو، کاربران بیشتر به سهولت دسترسی، سرعت، دقت ادراک‌شده، جامعیت، پوشش جزئیات، پشتیبانی از زبان فارسی و تناسب با کار علمی اشاره کردند؛ چنان‌که یکی از مشارکت‌کنندگان بیان کرد: «جمنای کیفیت پاسخ‌گویی خیلی خوبی داشت؛ هم جامع بود، هم جزئیات را نگه می‌داشت و زبان فارسی را خیلی خوب پشتیبانی می‌کرد»(مشارکت کننده ۳). در مقابل، چت‌جی‌پی‌تی ۴ بیشتر با تعامل‌پذیری، مدیریت بهتر گفت‌وگو، چارچوب‌مندی پاسخ و نظم ارائه محتوا توصیف شد. یکی از کاربران اظهار کرد: «چت‌جی‌پی‌تی ۴ تعاملی‌تر بود و مدیریت گفت‌وگو در آن راحت‌تر انجام می‌شد» (مشارکت کننده ۱۲). بنابراین، داده‌های کیفی نشان می‌دهد که جمنای بیشتر در عمق، گستردگی و کاربرد پژوهشی، و چت‌جی‌پی‌تی ۴ بیشتر در روانی تعامل و سازمان‌دهی پاسخ برجسته شده است.

پیشنهادهای کاربردی

۱. برای دانشجویان تحصیلات تکمیلی، استفاده مرحله‌ای از دو مدل پیشنهاد می‌شود: ابتدا چت‌جی‌پی‌تی ۴ برای صورت‌بندی مسئله، تولید پرسش‌های دقیق‌تر و دریافت چارچوب اولیه؛ سپس جمنای ۲٫۵ پرو برای گسترش ابعاد موضوع، تنوع دیدگاه‌ها و دریافت پاسخ‌های پژوهشی‌تر.

۲. کتابخانه‌های دانشگاهی می‌توانند کارگاه‌های سواد هوش مصنوعی و سواد اطلاعاتی برگزار کنند که در آن دانشجویان یاد بگیرند پاسخ‌های دو مدل را مقایسه، خطاها را شناسایی، منابع را بررسی و پرامپت‌های دقیق‌تر طراحی کنند.

۳. استادان راهنما می‌توانند از دانشجویان بخواهند در کارهای پژوهشی، استفاده از مدل‌های زبانی را شفاف گزارش کنند؛ برای مثال نوع مدل، هدف استفاده، متن پرامپت‌های اصلی، نحوه راستی‌آزمایی و منابع نهایی بررسی‌شده را در پیوست یا یادداشت روش‌شناختی ثبت کنند.

۴. در درس‌های روش تحقیق و نگارش علمی، تمرین‌هایی طراحی شود که دانشجویان یک پرسش علمی را به هر دو مدل بدهند، پاسخ‌ها را از نظر دقت، ربط، پوشش، کاربردی‌بودن و منابع مقایسه کنند و در نهایت با پایگاه‌های علمی معتبر پاسخ نهایی را اصلاح کنند.

محدودیت‌ها و پیشنهاد برای پژوهش‌های آینده

۱. نخستین محدودیت پژوهش، حجم نمونه ۳۶ نفری و تمرکز آن بر دانشجویان دکتری یک دانشکده در دانشگاه خوارزمی است؛ بنابراین تعمیم نتایج به همه دانشجویان تحصیلات تکمیلی یا رشته‌های دیگر باید با احتیاط انجام شود.

۲. دومین محدودیت، وابستگی نتایج به نسخه‌های مورد استفاده از چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در زمان اجرای پژوهش است. مدل‌های زبانی به‌سرعت به‌روزرسانی می‌شوند و ممکن است در نسخه‌های بعدی، دقت، سرعت، دسترسی یا نحوه تعامل آن‌ها تغییر کند.

۳. سومین محدودیت، نقش مهارت پرامپت‌نویسی کاربران است. کیفیت پاسخ مدل‌ها تا حد زیادی به وضوح پرسش، زمینه ارائه‌شده، پیگیری کاربر و توان او در اصلاح پرسش بستگی دارد. بنابراین بخشی از تفاوت‌های مشاهده‌شده ممکن است ناشی از شیوه تعامل کاربران با مدل‌ها باشد.

۴. چهارمین محدودیت، اتکای بخشی از داده‌ها به ارزیابی ادراکی کاربران است. اگرچه استفاده از مورائه و پروتکل بلنداندیشی کیفیت داده‌ها را افزایش داد، برای سنجش دقیق‌تر دقت پاسخ‌ها، پژوهش‌های آینده باید از داوری تخصصی مستقل، سناریوهای استاندارد و معیارهای عینی مانند نرخ خطا، کیفیت ارجاعات و زمان پاسخ نیز استفاده کنند.

۵. از آنجا که ترتیب استفاده از دو مدل برای همه مشارکت‌کنندگان یکسان بود، احتمال اثر ترتیب، یادگیری یا خستگی بر ارزیابی کاربران وجود دارد؛ بنابراین، در پژوهش‌های آینده پیشنهاد می شود ترتیب ارائه مدل‌ها به‌صورت چرخشی یا تصادفی کنترل شود.

ملاحظات اخلاقی

در اجرای پژوهش، مشارکت دانشجویان آگاهانه و داوطلبانه بود. محرمانگی داده‌ها رعایت شد و اطلاعات هویتی مشارکت‌کنندگان در گزارش مقاله افشا نشده است.

مشارکت نویسندگان

نویسنده اول: گردآوری داده‌ها، اجرای جلسات، انجام تحلیل‌های کمی و کیفی و تهیه پیش‌نویس اولیه مقاله، **نویسنده دوم**: استاد مشاور، مشارکت در طراحی پژوهش، مشاوره علمی و بازبینی مقاله. **نویسنده سوم**: استاد راهنمای پایان‌نامه، انجام تحلیل‌های کمی و کیفی، نظارت بر مراحل انجام پژوهش، بازبینی علمی نتایج و اصلاح مقاله.

تعارض منافع

نویسندگان اعلام می‌کنند که این مقاله تعارض منافع ندارد.

سپاسگزاری

این مقاله مستخرج از پایان‌نامه کارشناسی ارشد مهدیه خندابی در رشته علم اطلاعات و دانش‌شناسی دانشگاه خوارزمی با عنوان «مقایسه چت‌جی‌پی‌تی ۴ و جمنای ۲٫۵ پرو در رفع نیازهای اطلاعاتی دانشجویان تحصیلات تکمیلی در زبان فارسی» است که با راهنمایی نویسنده سوم و مشاوره نویسنده دوم انجام شده است. همچنین این مطالعه در «آزمایشگاه تعامل انسان و اطلاعات» دانشکده روانشناسی و علوم تربیتی دانشگاه خوارزمی انجام شده و بدین وسیله از همه دانشجویانی که در فرایند گردآوری داده‌ها همکاری کردند سپاسگزاری می‌شود.

منابع

اکبری، علی؛ ریگی، طاهره؛ و فتاحی، سید رحمت‌الله (۱۳۹۸). از رفتار اطلاع‌یابی تا رفتار دانش‌یابی: واکاوی سیر تحول مفهومی و نظری. *پردازش و مدیریت اطلاعات*، ۳۴(۴)، ۱۹۳۹–۱۹۶۰.

اسفندیاری، شهرام؛ و قمری، پرهام (۱۴۰۳). مروری نظام‌مند بر تأثیر مدل‌های هوش مصنوعی مولد ChatGPT و Gemini در آموزش زبان انگلیسی: فرصت‌ها و چالش‌ها. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*، ۱۴(۴)، ۶۱۱–۶۴۱. https://doi.org/10.22059/jflr.2025.386926.1173

پرهام‌نیا، فرشاد (۱۴۰۰). شناسایی عوامل مؤثر روان‌شناختی بر رفتار اطلاع‌یابی کاربران اطلاعاتی: مطالعه مرور نظام‌مند. *مطالعات کتابداری و علم اطلاعات*، ۱۳(۳)، ۸۲–۱۰۵. https://doi.org/10.22055/slis.2021.32105.1682

ستوده، داود؛ و امیری‌تهرانی‌زاده، امین (۱۴۰۲). مدل زبانی مبتنی بر BERT جهت تحلیل محتوای ورزشی در زبان فارسی. *کارافن*، ۲۰(۱)، ۳۴۱–۳۶۲. https://doi.org/10.48301/kssa.2023.357227.2251

نصیری، محمدرضا؛ و عبدالله عموقین، جعفر (۱۴۰۳). بررسی نقش هوش مصنوعی در آموزش با استفاده از رویکرد رهنگاشت فناوری. *مطالعات کاربردی علم‌سنجی*، ۱(۲)، ۱۱۳–۱۲۹.

یاری، شیوا؛ و احمدی، حمید (۱۳۹۳). مروری بر متون رفتار اطلاع‌یابی در ایران. پژوهش‌نامه پردازش و مدیریت اطلاعات، ۳۰(۱)، ۱۷۳–۱۹۷. https://doi.org/10.35050/JIPM010.2014.006

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Afzal, W. (2017). Conceptualization and measurement of information needs: A literature review. *Journal of the Australian Library and Information Association, 66*(2), 116–138. https://doi.org/10.1080/24750158.2017.1306165

Akter, S. N., Yu, Z., Muhamed, A., Ou, T., Bäuerle, A., Cabrera, Á. A., ... & Neubig, G. (2023). An in-depth look at Gemini's language abilities. *arXiv preprint arXiv:2312.11444*.

Alpar, Ö. (2025). Evaluating generative AI tools for improving English writing skills: A preliminary comparison of ChatGPT-4, Google Gemini, and Microsoft Copilot. *European Journal of Educational Research, 14*(4), 1291–1308. https://doi.org/10.12973/eu-jer.14.4.1291

Baytak, A. (2024). The content analysis of the lesson plans created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology, 9*(1), 329–350. https://doi.org/10.46303/ressat.2024.19

Đerić, E., Frank, D., & Milković, M. (2025). Trust in generative AI tools: A comparative study of higher education students, teachers, and researchers. *Information, 16*(7), Article 622. https://doi.org/10.3390/info16070622

Digital Education Council. (2024, August 2). *Digital Education Council global AI student survey 2024: Insights from over 3,800 students across 16 countries*. Retrieved from https://www.digitaleducationcouncil.com/resource-library-items/digital-education-council-global-ai-student-survey-2024

Fattah, F. H., Salih, A. M., Salih, A. M., Asaad, S. K., Ghafour, A. K., Bapir, R., ... Kakamad, F. H. (2025). Comparative analysis of ChatGPT and Gemini (Bard) in medical inquiry: A scoping review. *Frontiers in Digital Health, 7*, Article 1482712. https://doi.org/10.3389/fdgth.2025.1482712

Freeman, J. (2025). *Student generative AI survey 2025* (HEPI Policy Note 61). Oxford, England: Higher Education Policy Institute. Retrieved from https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/

Gemini Team, Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... Blanco, L. (2023). Gemini: A family of highly capable multimodal models. arXiv preprint arXiv:2312.11805. Retrieved from https://arxiv.org/abs/2312.11805

Gemini Team, Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., ... Batsaikhan, B. O. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv preprint arXiv:2403.05530. Retrieved from https://arxiv.org/abs/2403.05530

Google AI for Developers. (2025, September 25). *Gemini API release notes*. Retrieved from https://ai.google.dev/gemini-api/docs/changelog

Google. (2025, March 25). *Gemini 2.5: Our most intelligent AI model*. The Keyword. Retrieved from https://blog.google/innovation-and-ai/models-and-research/google-deepmind/gemini-model-thinking-updates-march-2025/

Goto, T., Ono, K., & Morita, A. (2024). A comparative analysis of large language models to evaluate robustness and reliability in adversarial conditions. TechRxiv. https://doi.org/10.36227/techrxiv.171173447.70655950/v1

Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., & Hamam, H. (2025). A systematic literature review to implement large language model in higher education: Issues and solutions. *Discover Education, 4*, Article 35. https://doi.org/10.1007/s44217-025-00424-7

Hosseinbeigi, S. B., Rohani, B., Masoudi, M., Shamsfard, M., Saaberi, Z., Karimi Manesh, M., & Abbasi, M. A. (2025). Advancing Persian LLM evaluation. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 2711–2727). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-naacl.147

Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments, 11*, Article 22. https://doi.org/10.1186/s40561-024-00310-z

Kawinkoonlasate, P. (2025). A comparative study of Google Gemini and ChatGPT in enhancing English language learning for EFL learners: A case study of the English research writing course. *Pedagogical Research, 10*(4), Article em0251. https://doi.org/10.29333/pr/17670

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371. https://doi.org/10.1002/(SICI)1097-4571(199106)42:5

Kumar, P. (2024). Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artificial Intelligence Review, 57*, Article 260. https://doi.org/10.1007/s10462-024-10888-y

Lang, G., Triantoro, T., & Sharp, J. H. (2024). Large language models as AI-powered educational assistants: Comparing GPT-4 and Gemini for writing teaching cases. *Journal of Information Systems Education, 35*(3), 390–407. https://doi.org/10.62273/YCIJ6454

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511626388

Naumer, C. M., & Fisher, K. E. (2010). Information needs. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2452–2458). Boca Raton, FL: CRC Press. https://doi.org/10.1081/E-ELIS3-120043243

Nicholas, G., & Bhatia, A. (2023). Lost in translation: Large language models in non-English content analysis. arXiv preprint arXiv:2306.07377. Retrieved from https://arxiv.org/abs/2306.07377

Omar, M., Nassar, S., Hijazi, K., Glicksberg, B. S., Nadkarni, G. N., & Klang, E. (2025). Generating credible referenced medical research: A comparative study of OpenAI's GPT-4 and Google's Gemini. *Computers in Biology and Medicine, 185*, Article 109545. https://doi.org/10.1016/j.compbiomed.2024.109545

Ono, K., & Morita, A. (2024). Evaluating large language models: ChatGPT-4, Mistral 8x7B, and Google Gemini benchmarked against MMLU. TechRxiv. https://doi.org/10.36227/techrxiv.170956672.21573677/v1

OpenAI. (2024). *GPT-4o system card*. Retrieved from https://openai.com/index/gpt-4o-system-card

Popović Šević, N., Šević, A., Slijepčević, M., & Krstić, J. (2025). AI adoption in higher education: Exploring attitudes and perceived benefits between users and non-users. *Online Journal of Communication and Media Technologies, 15*(4), Article e202528. https://doi.org/10.30935/ojcmt/17246

Rane, N., Choudhary, S., & Rane, J. (2024a). Gemini or ChatGPT? Efficiency, performance, and adaptability of cutting-edge generative artificial intelligence (AI) in finance and accounting. SSRN. https://doi.org/10.2139/ssrn.4731283

Rane, N., Choudhary, S., & Rane, J. (2024b). Gemini versus ChatGPT: Applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence, 5*(1), 69–93. https://doi.org/10.48185/jaai.v5i1.1052

Saab, K., Tu, T., Weng, W. H., Tanno, R., Stutz, D., Wulczyn, E.,... & Natarajan, V. (2024). Capabilities of gemini models in medicine. arXiv preprint arXiv:2404.18416.

Salman, I. M., Ameer, O. Z., Khanfar, M. A., & Hsieh, Y. H. (2025). Artificial intelligence in healthcare education: evaluating the accuracy of ChatGPT, Copilot, and Google Gemini in cardiovascular pharmacology. *Frontiers in Medicine*, 12, 1495378. https://doi.org/10.3389/fmed.2025.1495378

Shahghasemi, E. (2025). AI: A human future. *Journal of Cyberspace Studies, 9*(1), 145–173. https://doi.org/10.22059/jcss.2025.389027.1123

Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. *International Journal of Educational Technology in Higher Education, 21*, Article 46. https://doi.org/10.1186/s41239-024-00478-x

Shi, Y., Yu, K., Dong, Y., & Chen, F. (2025). Large language models in education: A systematic review of empirical applications, benefits, and challenges. *Computers and Education: Artificial Intelligence, 10*, Article 100529. https://doi.org/10.1016/j.caeai.2025.100529

Strzalkowski, P., Strzalkowska, A., Chhablani, J., Pfau, K., Errera, M. H., Roth, M.,... & Guthoff, R. (2024). Evaluation of the accuracy and readability of ChatGPT-4 and Google Gemini in providing information on retinal detachment: a multicenter expert comparative study. *International Journal of Retina and Vitreous,* 10(1), 61.

Toribio, N. F. (2023). Analysis of ChatGPT and other AI's ability to reduce anxiety of science-oriented learners in academic engagements. *Journal of Namibian Studies: History, Politics, Culture*, 33, 5320–5337. https://doi.org/10.59670/jns.v33i.3132

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modeling cognitive processes*. London, England: Academic Press.

Wilson, T. D. (1999). Models in information behavior research. *Journal of Documentation, 55*(3), 249–270. https://doi.org/10.1108/EUM0000000007145

Wilson, T. D. (2000). Human information behavior. *Informing Science, 3*(2), 49–56. doi:10.28945/576

Yildiz Domanic, K., & Baycan, S. (2025). Evaluating and comparing student responses in examinations from the perspectives of human and artificial intelligence (GPT-4 and Gemini). *BMC Medical Education, 25*, Article 1282. https://doi.org/10.1186/s12909-025-07835-y