African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 5, Issue 2, 2026
Page No: 217-225
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2025: 1.126 SJIFactor 2024: 6.752 معامل التأثير العربي: 1.62

An Empirical Comparative Analysis of ChatGPT and DeepSeek in NLP: Text Generation, Summarization, Translation, and User Feedback Evaluation

Aisha Hamed Bubaker [1 *], Nadia Mohammed Senussi [2]
[1,2] Computer Science Department, Faculty of Education - Qmens, University of Benghazi, Benghazi, Libya

دراسة تحليلية ومقارنة تجريبية بينChatGPT وDeepSeek في معالجة اللغة الطبيعية: توليد النصوص، الترجمة، وتقييم تجربة المستخدم

عائشة حمد بوبكر [1*]، نادية محمد السنوسي [2]
[2،1] قسم الحاسوب، كلية التربية - قمينس، جامعة بنغازي، بنغازي، ليبيا

[*]Corresponding author: aisha.hamed@uob.edu.ly

Received: March 14, 2026	Accepted: May 22, 2026	Published: June 04, 2026

Abstract

Understanding the capabilities of AI tools is incredibly important, especially as they become more integrated into our daily lives. This study examines two AI tools: ChatGPT and DeepSeek-V3. This research aims to conduct a comprehensive comparison between the two tools in terms of performance, flexibility. In this paper we evaluate the capabilities of each tool in natural language processing. Data was collected by performing the same activities with each tool, and the accuracy and quality of the outcomes were recorded. Additionally, each tool's adaptability to various criteria was tested. The findings demonstrated ChatGPT's strong performance in general tasks and natural language processing, as well as its user-friendliness and high degree of adaptability to a wide range of needs. On the other hand, DeepSeek-V3 demonstrated outstanding performance in specialized tasks, producing precise and effective results in particular domains such as translation.

Keywords: Generative AI, DeepSeek-V3, ChatGPT, Educational Applications, AI Tools in Education, (Natural Language Processing) NLP.

الملخص

فهم قدرات أدوات الذكاء الاصطناعي أمر بالغ الأهمية، خاصة مع تزايد استخدامها في حياتنا اليومية. يركز هذا البحث على أداتين للذكاء الاصطناعي ChatGPT وDeepSeek-V3, ويهدف إلى إجراء دراسة تحليلية لهاتين الأداتين وذلك من خلال القيام بمقارنة شاملة تعتمد على التجارب العملية والتحليل الكمي والنوعي. في هذه الورقة نقوم بتقييم قدرات كل أداة في مهام مختلفة، خصوصا في معالجة اللغة الطبيعية (NLP) من حيث تلخيص المستندات والترجمة الآلية وإنشاء النصوص. تم جمع البيانات من خلال القيام بنفس الأنشطة مع كل أداة، وتم تسجيل كفاءة وجودة النتائج وجمع أراء بعض المستخدمين. أظهرت النتائج أن ChatGPT يتمتع بقوة كبيرة في المهام العامة ومعالجة اللغة الطبيعية (NLP)، فضلاً عن سهولة استخدامه وقدرته العالية على التكيف مع مجموعة واسعة من الاحتياجات. وعلى العكس من ذلك، أظهر DeepSeek-V3 أداءً متميّزا في المهام المتخصصة، حيث أعطى نتائج دقيقة وفعالة في مجالات معينة مثل الترجمة.

الكلمات المفتاحية: الذكاء الاصطناعي التوليدي، ChatGPT، DeepSeek-V3, التطبيقات التعليمية، أدوات الذكاء الاصطناعي في التعليم، معالجة اللغة الطبيعية(NLP) .

Introduction

The involvement of AI-powered technologies that assist students with a variety of activities, such as text generation, translation, and summarization, has significantly increased. Two well-known AI models with unique advantages in conversational AI and technical problem-solving are ChatGPT and DeepSeek (D R, & S,2025). ChatGPT, developed by OpenAI and released on November 30, 2022, is a robust text generating and conversational system (An, Ding, & Lin,2023). It is a natural language processing (NLP) model designed to produce human-like responses to user input. OpenAI is a company specializing in artificial intelligence (An, Ding, & Lin,2023). ChatGPT was trained using weighted algorithms developed by humans, based on a vast dataset of diverse texts, including books, research papers, scientific articles, Wikipedia, websites, and news media (Welsby & Cheung,2023). DeepSeek AI was launched in 2023 (Maes,2025). It is a specialized tool designed to help educators and learners develop various skills (Kotsis,2025). In 2024, DeepSeek introduced the DeepSeek Mixture-of-Experts (MoE) architecture, which utilizes a sparse activation method to shift the focus from computation-heavy tasks toward productivity-driven results (Neha & Bhati,2025). Subsequently, to streamline the software development lifecycle, a family of code-oriented models, DeepSeek Coder, was released, with parameter sizes ranging from 1B to 33B (Neha & Bhati,2025). DeepSeek published a series of smaller models in open-source code appropriate for application in low-resource contexts like edge computing devices or low-memory systems (Neha & Bhati,2025). These models allow cost-effectiveness and scalability, extending the potential of DeepSeek to provide high-end AI on diverse cases for use (Neha & Bhati,2025). The overall objective of this study is to examine the applicability, performance, and flexibility of the two AI technologies. This research paper's primary goal is to assess each tool's overall performance across various activities. Moreover, to evaluate the adaptability, flexibility to various needs and to investigate user experiments with the two tools.

Literature Review:

Recent studies show that students are extensively using artificial intelligence tools in higher education, mostly ChatGPT. One of these studies, titled Investigating Students Awareness Use and Perception of ChatGPT in Libyan Higher Education A Case Study at the University of Benghazi 2024. By Hamed and Senussi, addressed students Awareness of ChatGPT and its uses (Hamed & Senussi,2025). The results indicate that many students rely on ChatGPT to support their learning, assist with academic writing, and improve their overall academic performance (Hamed & Senussi,2025). As the authors pointed out students highly aware of ChatGPT's advantages., such as aiding research and providing personalized learning support. The study finding is" Several students recognize the benefits of ChatGPT, such as aiding research and providing personalized learning support. However, they also express concerns about over-reliance on AI, accuracy issues, reduced interactions with educators, and potential risks to academic integrity" (Hamed & Senussi2025). Therefore, recent research has shifted from simply examining students' awareness to comparing different AI models to determine which tools offer reasoning capabilities, accuracy, and efficiency. One example is the study conducted by Mondillo et al., titled *"Comparative Evaluation of Advanced AI Reasoning Models in Pediatric Clinical Decision Support: ChatGPT O1 vs. DeepSeek-R1."* The study aimed to evaluate the diagnostic accuracy and clinical utility of ChatGPT O1 and DeepSeek-R1 models in pediatric scenarios using the MedQA dataset (Mondillo, Colosimo, Perrotta, Frattolillo, & Masino,2025). The result of this study indicates that ChatGPT O1 achieved an accuracy of 92.8%, while DeepSeek-R1 achieved an accuracy of 87.0%. Cohen's Kappa (K=0.20) indicated low agreement between the models, reflecting their distinct reasoning strategies (Mondillo, Colosimo, Perrotta, Frattolillo, & Masino,2025). This study demonstrates ChatGPT O1's effectiveness in giving accurate and logical clinical reasoning, which makes it ideal for critical pediatric situations. Because of its adaptability and accessibility, DeepSeek-R1 is still a useful tool in environments with limited resources. Another study done by Albuhairy& Algaraady "DeepSeek vs. ChatGPT: Comparative Efficacy in Reasoning for Adults' Second Language Acquisition Analysis". This study compares the performance of DeepSeek and ChatGPT in assessing adult second language acquisition errors, particularly for South Asian Arabic learners, and evaluates their efficacy in detecting linguistic inaccuracies and diagnosing cases of L1 influence (Albuhairy, & Algaraady,2025). The objective of the study is to evaluate the performance of DeepSeek and ChatGPT in detecting linguistic inaccuracies (morphology, syntax, semantics) and diagnosing cases of L1 (first language) influence in adult L2 learners (Albuhairy & Algaraady,2025). The results of the study show that DeepSeek outperformed ChatGPT in context-driven error detection, while ChatGPT provided more relevant feedback (Albuhairy & Algaraady,2025). Study done by Gao and others about "A Comparison of DeepSeek and Other LLMs". The objectives of the study are to compare the performance of five LLMs on two classification problems, to propose a new dataset for evaluating academic research, and to investigate the similarity of predictions made by different LLMs (Gao, Jin, Ke, & Moryoussef,2025). Authorship classification and citation classification are the two classification settings used in the study. The study evaluates the performance of five LLMs: DeepSeek, Claude, Gemini, GPT, and Llama, for each scenario. CitaStat and MadStatAI are the two datasets used in the study (Gao, Jin, Ke, & Moryoussef,2025). The study finds that Claude consistently outperforms all other LLM approaches. DeepSeek-R1 underperforms Claude but outperforms Gemini, GPT, and Llama in most of the cases. GPT performs unsatisfactorily for AC1 and AC2, but it performs

much better than random guessing for CC1 and CC2. Llama performs unsatisfactorily (Gao, Jin, Ke, & Moryoussef,2025). The results show that DeepSeek underperforms Claude but outperforms Gemini, GPT, and Llama in most cases. The percent of agreement between different LLMs is also reported (Gao, Jin, Ke, & Moryoussef,2025). Study done by Aydın about "Generative AI in Academic Writing: A Comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma". This study's primary goal is to assess new generation large language models' academic writing abilities by contrasting them with those of existing well-known AI systems. The study intends to evaluate Qwen 2.5 Max and DeepSeek v3's capacities to produce high quality academic content, taking into account their performance in academic writing procedures, content originality, semantic similarity, and readability (Aydın, Karaarslan, Erenay, & Bacanin,2025).The study's findings indicate that Qwen 2.5 Max and DeepSeek v3 can produce high-quality academic content, but questions remain regarding the content's originality and the models' capacity to avoid detection by AI detection tools readability (Aydın, Karaarslan, Erenay, & Bacanin,2025).

Material and methods

A quantitative and qualitative methodology was used in this research to evaluate the performance, flexibility of AI tools the DeepSeek-v3 and ChatGPT. A series of practical experiments were designed to collect data, focusing on comparing performance on various tasks, analyzing flexibility, and exploring user experience. A mixed methodology was adopted that combined quantitative and qualitative analysis. Quantitative analysis was used to evaluate performance based on specific metrics such as accuracy and speed, while qualitative analysis was used to understand flexibility and usability through user feedback and ratings. To assess the effectiveness and adaptability of the two tools, a number of practical tests were created. The experiments covered the natural language processing (NLP) in text summarization, translation and text generating. These assignments were selected to reflect various aspects and degrees of difficulty.

Natural Language Processing (NLP):

Natural language processing (NLP) is a subfield of artificial intelligence that seeks to make it possible for machines to comprehend, interpret, and produce human language (Chopra, Prashar, & Sain,2013). A number of tests were created for this study in order to assess how well DeepSeek and ChatGPT performed on different NLP tasks. The NLP tasks were selected to cover a variety of aspects, such as machine translation, text summarization, and text generating to reflect varying degrees of difficulty. These tests seek to determine each tool's advantages and disadvantages while comparing the precision, speed, and adaptability of each in handling natural language.

Summarizing the texts:

The first experiment was to ask ChatGPT and DeepSeek to summarize a research paper that called "ADOPTION AND USE OF AI TOOLS: A RESEARCH AGENDA GROUNDED IN UTAUT" Cited as: Venkatesh, V. "Adoption and Use of AI Tools: A Research Agenda Grounded in UTAUT," Annals of Operations Research, forthcoming. https://doi.org/10.1007/s10479-020-03918-9. The ChatGPT summary is consistent with the content of the document. The document attached in the link talks about AI tools in organizations, which the summary generated by ChatGPT covers in detail. The ChatGPT summary covered the main points and mentioned challenges such as "black box" AI models and potential errors as shown in Figure 1. It referred to the theoretical framework used (UTAUT) and how to improve AI adoption. It summarized the benefits (improved decision making) and obstacles (lack of transparency), However, the summary generated by DeepSeek talks about digital transformation in higher education institutions as shown in Figure 2, which is a completely different topic from the content of the document attached in the link. Therefore, the summary generated by DeepSeek is incorrect for this document

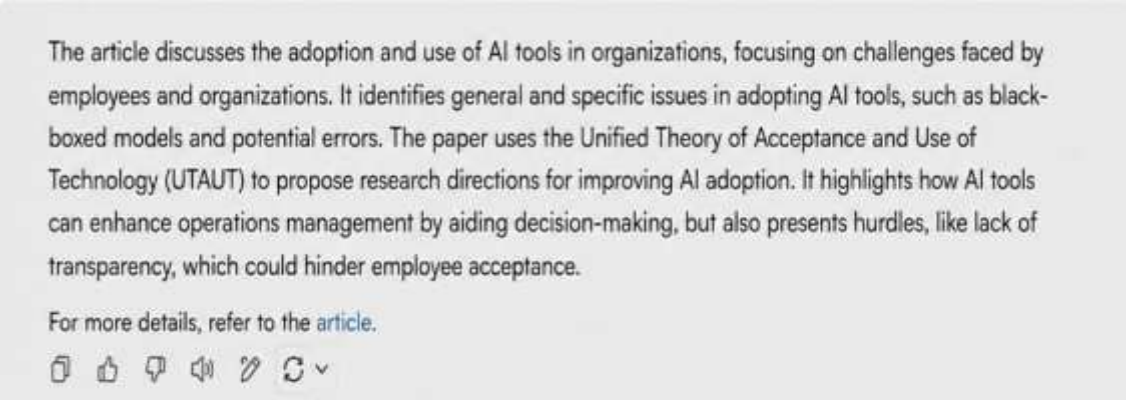

Figure1. shows text summarization of ChatGPT

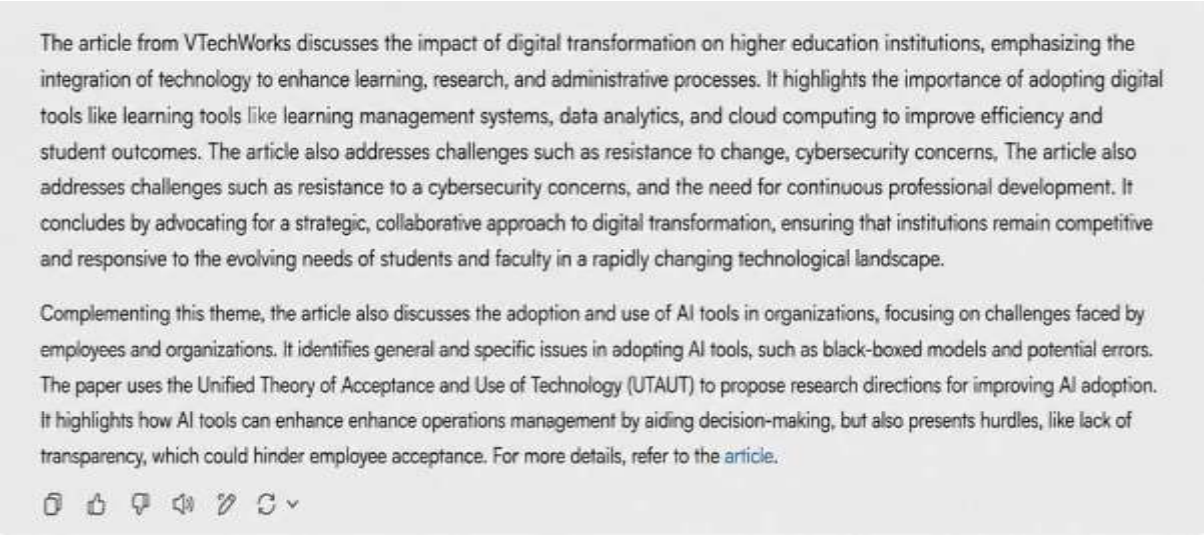

Figure2. shows text summarization of DeepSeek

Machine translation:

Machine translation is considered one of the most important practical applications of artificial intelligence technologies. There are several benefits and challenges to machine translation. One of its benefits is speed, as machine translation offers nearly instant translation, which makes it perfect for real-time applications (such live conversations or translations). In addition, it is inexpensive when compared to human translation. However, there are a number of difficulties and drawbacks with machine translation, such as its inability to accurately translate complicated materials because it may miss linguistic subtleties, particularly in technical, legal, or literary works. Additionally, informal language phrasing may be difficult for machine translation to comprehend, producing hilarious or incorrect translations. In addition to the previously stated, it could generate grammatical and linguistic faults in sentences. We asked Deep Seek to translate the abstract of the research paper that called "ADOPTION AND USE OF AI TOOLS: A RESEARCH AGENDA GROUNDED IN UTAUT" Cited as: Venkatesh, V. "Adoption and Use of AI Tools: A Research Agenda Grounded in UTAUT," Annals of Operations Research, forthcoming. https://doi.org/10.1007/s10479-020-03918-9. into Arabic language as shown in Figure3.

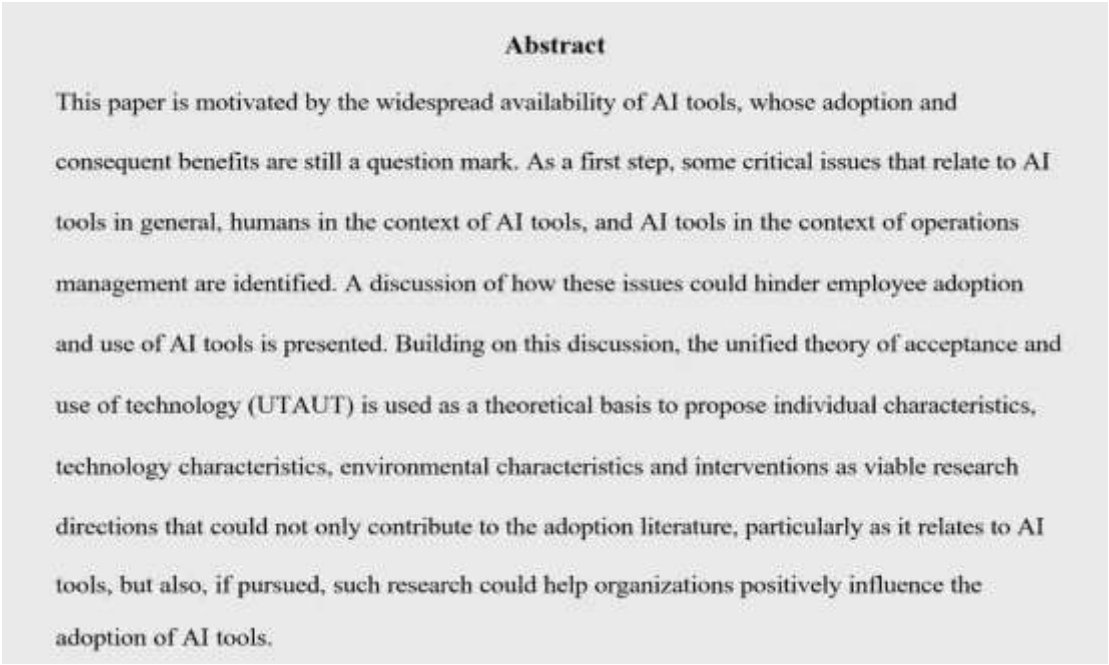

Figure 3. Shows the abstract that translated by ChatGPT and DeepSeek

When we compare the translations produced by the two tools, we find that DeepSeek translation outperforms ChatGPT translation in terms of covering the complete document, translating it accurately, and choosing the right words. As seen in Figure 4, DeepSeek translation is excellent and faithfully captures the information in the research paper's original abstract. While translation by ChatGPT is generally accurate, it lacks some important details as shown in Figure 5.

Figure 4. Shows the translation generated by DeepSeek

Figure 5. Shows the translation generated by ChatGPT

Text Generation:
In Figure 6 and Figure 7, we asked the tools to generate text based on a specific command. The task was: “Write a 100-word paragraph about the benefits and challenges of artificial intelligence for students.”

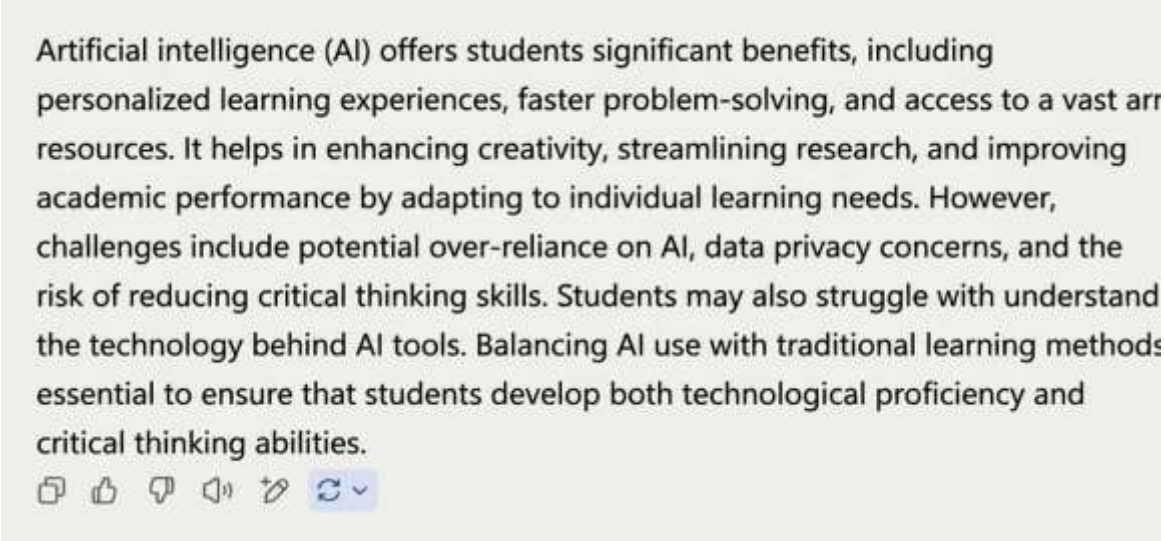

Figure 6. Shows the text generated by ChatGPT

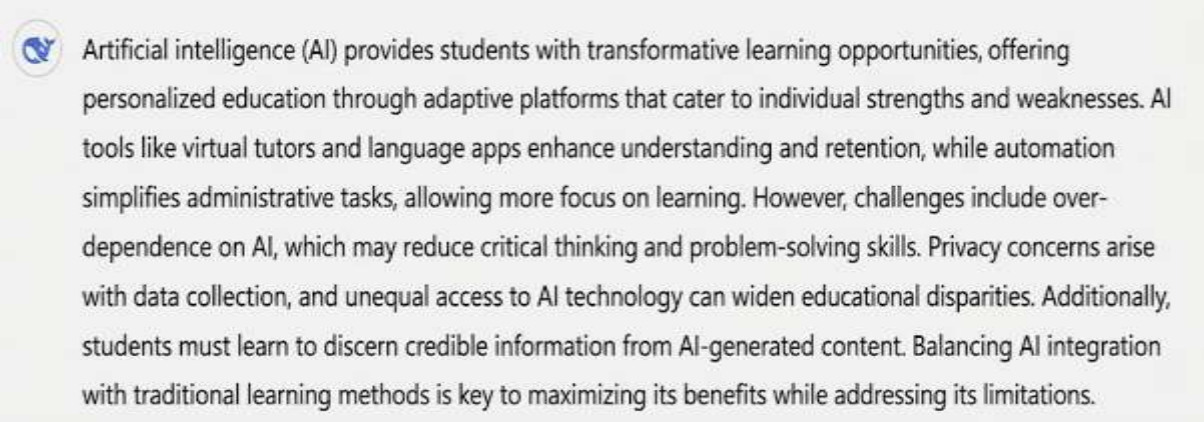

Figure 7. Shows the text generated by DeepSeek

The text generated by DeepSeek is better at covering the desired topic as shown in Figure 7, which is the benefits and challenges of AI for students. This is because of the following reasons: Benefit Comprehensiveness: The text that was generated by DeepSeek mentions more detailed benefits, such as adaptive learning platforms, virtual resources (such as virtual tutors and language apps), and the automation of administrative tasks, giving a broader picture of how students benefit from AI. Challenges Coverage: The DeepSeek generated text addresses challenges in more depth, such as overreliance on AI, privacy concerns, the digital divide, and the need to distinguish reliable information from AI-generated content. These points cover important aspects that may affect students. Balance: The DeepSeek generated text offers a good balance of benefits and challenges, emphasizing the importance of integrating AI with traditional educational methods. In comparison, the text generated by ChatGPT covers benefits and challenges in general without sufficient detail, making it less comprehensive than the text that was generated by DeepSeek as shown in Figure 6. Qualitative analysis is based on students' opinions (user feedback) about the usability, flexibility and accuracy. In this study, one hundred students from the University of Benghazi, representing different academic majors and wide range of familiarity with AI technologies, were selected at random and asked about their thoughts on the two tools. Their responses were as follows:

Ease of use:

As illustrated in Figure 8, 73.3% of users perceive ChatGPT to be easier to use compared to DeepSeek, while 15.8% of users report no significant difference between the two tools

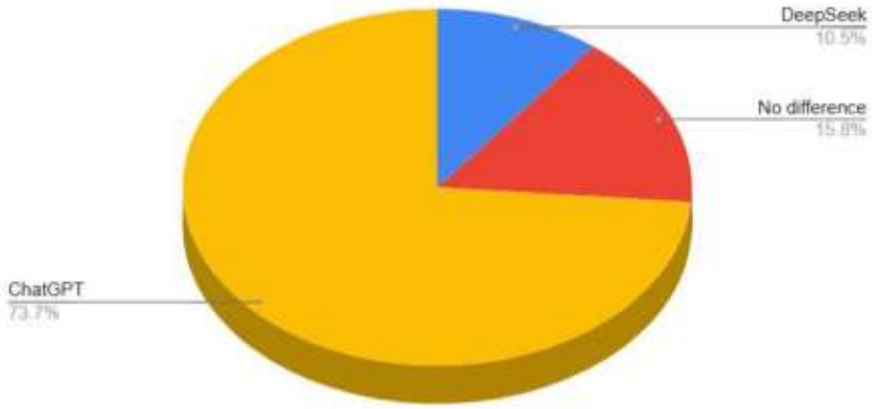

Figure 8. user opinion of usability

Accuracy:

When students were asked about the accuracy of the results provided by both tools, 60.5% of respondents expressed that they believe DeepSeek offers more accurate results compared to ChatGPT as illustrated in Figure 9. This suggests a general preference for DeepSeek in terms of precision and reliability in the information it provides. It is crucial to take into account the remaining percentage, as they either believed ChatGPT to be more accurate or did not observe a significant difference. This could suggest that user experiences vary based on the specific tasks or queries performed.

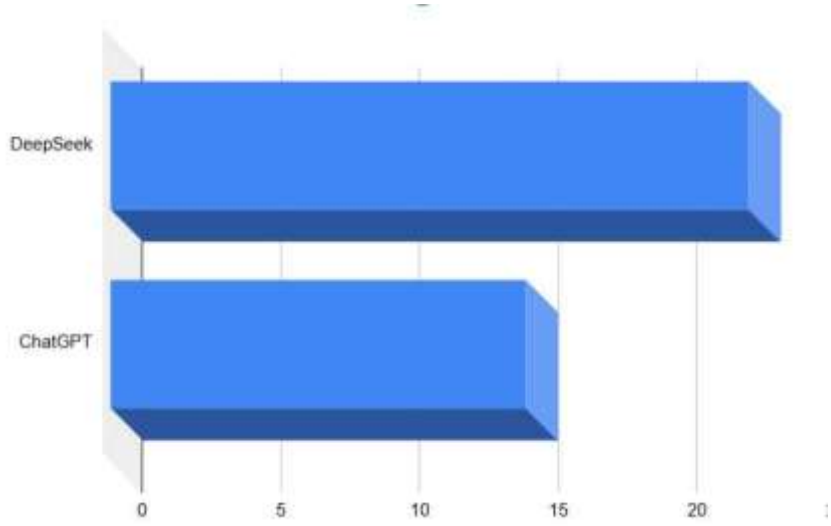

Figure 9. user opinion of accuracy

Discussion

The performance of the ChatGPT and DeepSeek tools was compared based on data collected from experiments and user opinions according to their performance in specific task focusing on natural language processing (translation and text generation). The results showed clear differences in performance between the two tools, which can be explained based on the strengths and weaknesses of each. A general comparison between ChatGPT and DeepSeek was conducted based on user feedback regarding accuracy, flexibility in handling tasks, and performance speed. Additionally, we evaluated the strengths and weaknesses of both technologies, taking into account user experiences and previous studies, as outlined in Table 1. This comparison aims to provide a comprehensive understanding of how each tool performs under different conditions, highlighting areas where they excel and where improvements might be needed.

Table 1. shows difference between ChatGPT and DeepSeek in general

Criteria	ChatGPT	DeepSeek
Accuracy	Good accuracy in general tasks	High accuracy in specialized tasks.
Speed	Faster on simple tasks.	Faster on complex tasks
Flexibility	Good flexibility in general tasks.	High flexibility in adapting to specific requirements
Strengths	-Ease of use -Strong performance in general tasks.	-High accuracy in specialized tasks -Flexibility in adapting to requirements
Weaknesses	-Limited in specialized tasks -May generate inaccurate results in complex tasks.	-Requires technical knowledge -Less sophisticated user interface

In typical natural language processing and text generation tasks, Deep Seek performed extremely well since it demonstrated an effective understanding of context and the capacity to generate cohesive messages. It also performed exceptionally well in translation, covering the complete document with high accuracy, but ChatGPT performed very well in document summarization. For instance, when it came to summarizing a lengthy text, ChatGPT delivered a clear and accurate summary in contrast to Deep Seek, which included inaccurate material not present in the research that was to be summarized. ChatGPT is simple to use and adaptable for basic jobs, while DeepSeek excels at specific tasks with excellent accuracy and speed. However, although ChatGPT may generate errors in specialized or sophisticated jobs, DeepSeek necessitates significant technical knowledge and may struggle to adapt to unexpected tasks. The quality of training data, the user interface's ease of use, and specialization versus generality are some of the elements that can account for these performance disparities. While DeepSeek is more suited for specialized activities, ChatGPT is more versatile but less accurate in specialized tasks because it can handle a wide range of general tasks.

The evaluation criteria for creating a summary of long texts are accuracy, length, clarity, and consistency. In terms of accuracy, ChatGPT is more accurate in summarizing the document and covered the main points, while Deep Seek has some points that are not mentioned in the main research. Length: Both tools gave a summary close to the required length and comprehensive. In terms of clarity and consistency, both summaries generated by ChatGPT and Deep Seek were clear and consistent enough to understand the research topic. The evaluation criteria for machine translation are accuracy, completeness, sequence, clarity, consistency, and efficiency. The translation by Deep Seek was very accurate and covered the entire summary to be translated. The translation was very sequential, clear, consistent, and very fast, while the translation by ChatGPT was good, but very brief, not sequential, and did not cover all the points in the summary to be translated. As for the evaluation criteria that were focused on in generating texts, they are accuracy, appropriate length, consistency, style, and creativity. DeepSeek's response is better since it is more structured, more understandable, and well-balanced.

Conclusion

ChatGPT and DeepSeek are two of the most powerful artificial intelligence tools, each with its own unique strengths and weaknesses, as highlighted in this research. DeepSeek stands out as the ideal choice for tasks that demand high precision and domain-specific expertise. Its ability to handle complex, niche requirements make it particularly valuable for professionals in fields like engineering, medicine, and scientific research, where accuracy and depth of information are critical. On the other hand, for general jobs that call for conciseness and clarity, ChatGPT is the greatest options. Applications like abstract creation and general tasks that call for conciseness and clarity can benefit from the use of ChatGPT. Although each tool has its capabilities, their benefits are different based on the situation. ChatGPT's adaptability and user-friendliness make it more appropriate for daily applications and general knowledge jobs. However, DeepSeek is more accurate which makes it a better choice for technical and complicated tasks. Users can choose the best outcomes catered to their individual requirements by utilizing the advantages of both technologies**.**

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

Albuhairy, M. M., & Algaraady, J. (2025). DeepSeek vs. ChatGPT: Comparative efficacy in reasoning for adults' second language acquisition analysis. *Humanities and Educational Sciences Journal, 44*, 864–883. https://doi.org/10.55074/hesj.vi44.1313

An, J., Ding, W., & Lin, C. (2023). ChatGPT: Tackle the growing carbon footprint of generative AI. *Nature, 615*(7953), 586. https://doi.org/10.1038/d41586-023-00843-2

Aydın, Ö., Karaarslan, E., Erenay, F. S., & Bacanin, N. (2025). Generative AI in academic writing: A comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma. *arXiv preprint arXiv:2503.04765*. https://doi.org/10.48550/arXiv.2503.04765

Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. *International Journal of Technology Enhancements and Emerging Engineering Research, 1*(4), 131–134.

D R, A., & I. S., S. (2025). Advancements in AI-powered NLP models: A critical analysis of Manus AI, Gemini, Grok AI, DeepSeek, and ChatGPT. *SSRN*. https://doi.org/10.2139/ssrn.5185131

Gao, T., Jin, J., Ke, Z. T., & Moryoussef, G. (2025). A comparison of DeepSeek and other LLMs. *arXiv preprint arXiv:2502.03688*. https://doi.org/10.48550/arXiv.2502.03688

Hamed, A., & Senussi, N. (2025). Investigating students' awareness, usage, and perceptions of ChatGPT in Libyan higher education: A case study at the University of Benghazi in 2024. *AlQalam Journal of Medical and Applied Sciences, 8*(2), 1166–1172. https://doi.org/10.54361/ajmas.258296

Kotsis, K. T. (2025). ChatGPT and DeepSeek evaluate one another for science education. *EIKI Journal of Effective Teaching Methods, 3*(1).

Maes, S. H. (2025). The circle of life for LLMs: Was the reaction to DeepSeek justified? *Zenodo*. https://doi.org/10.5281/zenodo.14838733

Mondillo, G., Colosimo, S., Perrotta, A., Frattolillo, V., & Masino, M. (2025). Comparative evaluation of advanced AI reasoning models in pediatric clinical decision support: ChatGPT O1 vs. DeepSeek-R1. *medRxiv*. https://doi.org/10.1101/2025.01.27.25321169

Neha, F., & Bhati, D. (2025). A survey of DeepSeek models. *Authorea Preprints*. https://doi.org/10.22541/aupreprint/244531

Welsby, P., & Cheung, B. M. Y. (2023). ChatGPT. *Postgraduate Medical Journal, 99*(1176), 1047–1048. https://doi.org/10.1093/postmj/qgad056