The European Research Journal 2026;12(7):730-738 Research Article

DOI: https://doi.org/10.18621/eurj.1839146 *Cardiovascular Surgery*

Comparative Analysis of Large Language Models in Hemodialysis Vascular Access: ChatGPT-5, Gemini-2.5, and DeepSeek-V3

Muhammet Hüseyin Erkan[1], Ömer Faruk Rahman[2], Abdullah Güner[3], Fevzi Ayyıldız[4], Emin Barbarus[1]

[1]Department of Cardiovascular Surgery, Balıkesir University, Faculty of Medicine, Balıkesir, Türkiye; [2]Department of Cardiovascular Surgery, İzmir Bakırçay University, Faculty of Medicine, İzmir, Türkiye; [3]Department of Cardiovascular Surgery, Konya City Hospital, Konya, Türkiye; [4]Department of Cardiovascular Surgery, Afyonkarahisar State Hospital, Afyonkarahisar, Türkiye

Abstract:

Objective: Vascular access is a key component of effective hemodialysis treatment. Patient education regarding selection, complications, and daily care remains challenging. Recently, large language models (LLMs) have been explored as supportive tools in this field. This study compared three widely used LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) in addressing patient-centered questions on hemodialysis vascular access.

Methods: Twenty-five frequently asked patient questions were compiled from literature, educational materials, and expert input. Each question was submitted to the three LLMs in standardized sessions, and answers were anonymized. Four cardiovascular surgeons independently evaluated responses in a blinded manner using 5-point Likert scales for accuracy, clarity, and scientificity. Statistical analyses compared model performances.

Results: DeepSeek-V3 provided significantly responses with a higher word count and achieved higher scientific depth scores compared to ChatGPT-5 and Gemini-2.5 ($P<0.01$). Accuracy scores showed ChatGPT-5 demonstrated significantly lower than both Gemini-2.5 and DeepSeek-V3 ($P<0.001$). No significant differences were observed among models regarding clarity. Overall, DeepSeek achieved the highest mean scores across all criteria.

Conclusion: DeepSeek-V3 demonstrated superior scientific depth and overall reliability, whereas ChatGPT-5 showed relative weaknesses in accuracy. The comparable clarity across models highlights LLMs' potential as supportive tools for patient education. Future studies should further validate these tools in real-world clinical settings and define appropriate safeguards for their use.

Keywords: Large Language Models, Artificial Intelligence, Patient Education, Hemodialysis, Vascular Access

Chronic kidney disease (CKD) is a global public health problem with increasing prevalence and incidence, significantly reducing quality of life and closely associated with mortality [1]. Hemodialysis is one of the most commonly used renal replacement therapies for patients with end-stage renal disease. Adequate and reliable vascular access is vital for the effective implementation of hemodialysis. Arteriovenous fistula (AVF), arteriovenous graft (AVG), and central venous catheter (CVC) are the

Submitted: December 9, 2025 **Accepted:** January 15, 2026 **Published Online:** January 20, 2026

How to cite this article: Erkan MH, Rahman ÖF, Güner A, Ayyıldız F, Barbarus E. Comparative Analysis of Large Language Models in Hemodialysis Vascular Access: ChatGPT-5, Gemini-2.5, and DeepSeek-V3. *Eur Res J.* 2026;12(7):730-738. doi: 10.18621/eurj.1839146

Corresponding author: Muhammet Hüseyin Erkan, MD., Assist. Prof., Phone: +90 266 612 14 61, E-mail: mhuseyinerkan@gmail.com

This is an **open-access** article distributed under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it.

Available Online at https://www.eurj.org.tr

main vascular access routes used today. International guidelines recommend AVF creation as the first option because it provides the highest possible patency, the lowest complication rate, and the best patient survival [2, 3].

However, patients may experience uncertainty and anxiety across a wide range of issues, from the choice of access type to the postoperative care process, from potential complications to the organization of daily life activities. Patients with insufficient knowledge may have difficulty recognizing clinical symptoms requiring medical intervention and, as a result, may delay seeking healthcare [4]. Furthermore, it has been reported that CKD patients with low health literacy have significantly lower quality of life parameters related to physical and mental health [5]. The literature shows that patients who can communicate effectively with healthcare professionals, evaluate reliable information sources, and understand and process health information have a higher prevalence of AVF use compared to catheters [6].

Traditionally, this information process has been carried out face-to-face by nephrologists, cardiovascular surgeons, and dialysis nurses. However, the intensity of healthcare systems, time constraints, and the large number of patients can make it difficult to allocate sufficient time to each patient. To fill this gap, patients are increasingly turning to the internet and digital resources. Revolutionary developments in artificial intelligence (AI) in recent years suggest that tools such as Large Language Models (LLMs) could play a potential role in patient education. Thanks to their natural language processing capabilities, these models have the potential to explain complex medical topics in language appropriate for patients. A study on the use of AI in nephrology reported that nephrologists who can integrate AI into their work can provide better care to their patients [7].

However, there are significant concerns regarding the quality, accuracy, and reliability of the content produced by these models. LLMs may reflect errors or biases in the datasets they were trained on, provide outdated information, or generate statements that sound plausible but are entirely unrealistic, known as “hallucinations” (8). When it comes to medical information, the risk that such errors could potentially harm patients necessitates a rigorous evaluation of these tools' performance. The current literature includes studies measuring the knowledge level of ChatGPT and other LLMs across various medical disciplines [9, 10]. However, no comprehensive study has been identified that provides a comparative evaluation of these models within a specific, multidisciplinary field such as hemodialysis vascular access, where patient education is crucial.

Therefore, access to accurate, understandable, and evidence-based information may play a pivotal role in patients’ decision-making regarding hemodialysis vascular access. Based on this premise, the aim of the present study is to evaluate the responses provided by three popular LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) to frequently asked questions by patients about hemodialysis vascular access; evaluating them using a blinded method by an expert panel based on the criteria of accuracy, clarity, and scientific depth.

METHODS

The methodology of this study was designed to compare the performance of three LLMs in answering patient questions about hemodialysis vascular access. The study was based on a panel of expert physicians evaluating the responses obtained from the models using a blinded method and comparing these responses in terms of objective metrics. The models evaluated were ChatGPT-5 (OpenAI, Microsoft Corporation, San Francisco, CA, USA, 2025), Gemini-2.5 (Google, Mountain View, CA, USA, 2025), and DeepSeek-V3 (DeepSeek, Beijing, CHINA, 2025) versions, which were accessed via their public interfaces on August 21, 2025.

The question set consists of 25 questions covering a variety of topics, including preparation for fistula surgery, fistula maturation, daily life and protection, fistula complications and their management, and catheter care. The identification of frequently asked questions was carried out through a systematic evaluation of widely used online platforms, including Google Trends, YouTube search recommendations, patient support forums, and official web portals of relevant healthcare organizations. Furthermore, questions recurrently reported by patients in daily clinical practice were included based on the investigators’ observations.

During the data collection process, each question was presented with a standardized prompt for each model. As an important methodological precaution, the browser cache and cookies were cleared for each new question, and a new chat session was always opened to prevent previous interactions from influencing the model output. All responses received were recorded in plain text, stripped of formatting, along with the relevant question and model ID.

To evaluate the collected responses, 25 responses from each model were randomly labeled with letters (Model X, Y, Z) and compiled into separate booklets. No information revealing the model ID was included anywhere in the booklets. The evaluation was conducted by a panel of four cardiovascular surgery specialists. A calibration meeting was held prior to the evaluation to ensure that the panelists interpreted the scales consistently. A minimum of 72 hours was allowed between the evaluation of one booklet and the distribution of the next to clear the evaluators' memory (washout). Each response was scored by the panelists on a 5-point Likert scale according to three main criteria: accuracy of medical content, clarity and comprehensibility of information for patients, and scientific depth and scope of information provided.

Statistical Analysis

All data were analyzed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA) software. The normality of data distribution was assessed using the Shapiro–Wilk test. Differences in response scores between the three models were examined using repeated measures analysis of variance for normally distributed data and the Friedman test for non-normally distributed data. Results for normally distributed data were reported as mean ± standard deviation, while results for non-normally distributed data were reported as median (minimum–maximum) values. A statistical significance level of $P<0.05$ was accepted for all analyses.

RESULTS

In terms of response length, defined as the total word count of each generated answer, statistically significant differences were observed among ChatGPT-5, Gemini-2.5, and DeepSeek-V3. DeepSeek-V3 produced the longest responses, with a mean word count of 439.12, followed by Gemini-2.5 (257.08) and ChatGPT-5 (149.60). These differences were statistically significant ($P<0.001$). Pairwise comparisons demonstrated that DeepSeek-V3 responses contained a significantly higher number of words than those of both Gemini-2.5 and ChatGPT-5 ($P<0.01$ for both comparisons), and that Gemini-2.5 responses were also significantly longer than those of ChatGPT-5 ($P<0.001$). Detailed data on response length are presented in Table 1.

For Rater 1, a statistically significant difference was found between ChatGPT-5, Gemini-2.5, and DeepSeek-V3 in the scientificity category (Figure 1). In pairwise comparisons, this difference was observed between ChatGPT-5 and DeepSeek-V3 ($P=0.022$). DeepSeek's median value was found to be significantly higher than ChatGPT-5's (5 and 4, respectively; $P=0.04$). No statistically significant differences were found in the other categories evaluated for R1 ($P=0.294$).

No statistically significant differences were found among the three models in the accuracy category for Rater 2 ($P=0.312$). Statistically significant differences were found between ChatGPT-5, Gemini-2.5, and DeepSeek-V3 in the scientificity category for R2 ($P<0.001$). In the pairwise comparisons, DeepSeek scored significantly higher than ChatGPT-5 ($P<0.001$) and was also significantly superior to Gemini-2.5 ($P<0.001$). No significant difference was found among the three models in the clarity category ($P=0.113$).

For Rater 3, statistically significant differences were found among the three models in the accuracy category ($P<0.001$). In the pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 and DeepSeek-V3 ($P<0.001$ for both comparisons). In contrast, no significant difference was found between Gemini-2.5 and DeepSeek-V3 ($P=0.832$). For R3, a statistically significant difference was found among the three models in the scientificity category ($P<0.001$). In pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 ($P=0.007$) and DeepSeek-V3 ($P<0.001$). In contrast, no significant difference was found between Gemini-2.5 and DeepSeek-V3 ($P=0.437$). No statistically significant difference was found among the models in the clarity category for R3 ($P=1.000$).

TABLE 1. Comparison of Response Length (Measured as Word Count) Among ChatGPT-5, Gemini-2.5, and DeepSeek-V3

Questions	ChatGPT-5	Gemini-2.5	DeepSeek-V3
1. Why is it necessary to establish a vascular access for hemodialysis?	171	321	433
2. What is the difference between an arteriovenous fistula and a catheter? Which one is safer?	180	366	524
3. Which vascular access option is more suitable for me for hemodialysis?	243	438	570
4. What tests are performed before fistula surgery?	171	285	574
5. Is fistula surgery difficult? How long does it take?	116	143	482
6. Do I need to do any special preparation before fistula surgery?	195	214	410
7. How long after fistula surgery will it be ready for use?	102	185	333
8. What does it mean for a fistula to "mature"? How long does it take?	160	242	419
9. What should I do after surgery to protect my fistula?	194	283	489
10. How can I tell if my fistula is working?	143	250	449
11. Can I use my fistula arm? Can I lift heavy objects?	123	215	374
12. Can blood pressure measurements or blood draws be performed on my fistula arm?	73	181	242
13. What should I be careful about in daily life to avoid damaging my fistula?	177	237	408
14. What should I do if my fistula becomes swollen, painful, or bruised?	128	243	489
15. What happens if my fistula becomes blocked? Can it be reopened?	103	302	347
16. How can I tell if a clot has formed in my fistula?	120	201	445
17. Is it dangerous if a balloon (aneurysm) develops in my fistula?	155	250	463
18. How will I undergo dialysis if my fistula does not function?	136	247	435
19. Why is a catheter inserted, and why is it preferred over a fistula?	158	355	548
20. How can I tell if my dialysis catheter is infected?	125	317	287
21. Can I bathe with my dialysis catheter, and what should I pay attention to in daily care?	173	224	404
22. What should I do if my dialysis catheter becomes blocked? Does it need to be replaced?	118	292	393
23. What is a fistula graft, and in which cases is it preferred?	206	276	521
24. If my fistula fails, can I have another surgery?	114	190	441
25. If my vascular access is no longer viable, what is the last resort for dialysis?	156	170	498
Mean±SD	149.6±38.68	257.08±68.27	439.12±82.58
P-value		**<0.001**	
F		517.55	

SD, standard deviation; F, One-way Anova test statistic. Statistically significant P-value is shown in bold.

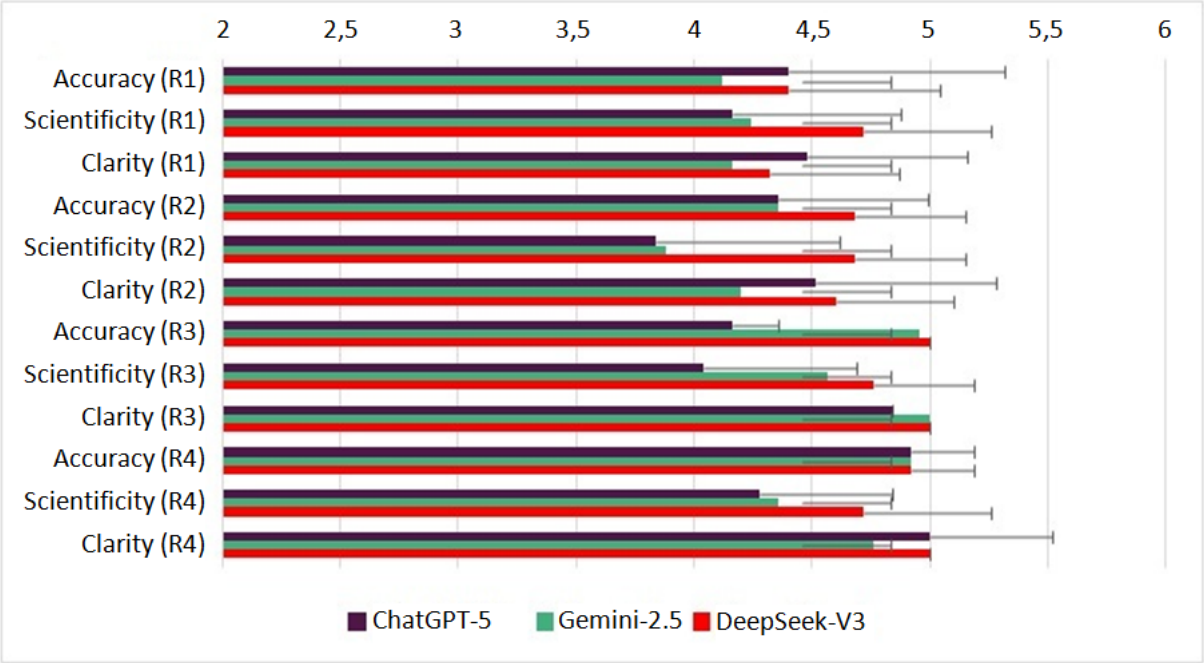


FIGURE 1. Comparison of ChatGPT, Gemini, and DeepSeek's scores in the accuracy, scientificity, and clarity categories using a bar chart.

TABLE 2. Comparison of ChatGPT-5, Gemini-2.5, and DeepSeek-V3's Accuracy, Scientificity, and Clarity Categories and Overall Average Scores

	ChatGPT-5	Gemini-2.5	DeepSeek-V3	stat	P-value
Accuracy (R1)	5 (3-5)	4 (2-5)	4 (3-5)	$\chi2$=2.452	0.294
Scientificity (R1)	4 (3-5)	4 (3-5)	5 (3-5)	$\chi2$=7.600	**0.022**
Clarity (R1)	5 (3-5)	4 (3-5)	4 (3-5)	$\chi2$=2.233	0.327
Accuracy (R2)	4 (4-5)	4 (3-5)	5 (4-5)	$\chi2$=6.500	0.312
Scientificity (R2)	4 (3-5)	4 (3-5)	5 (4-5)	$\chi2$=21.816	**<0.001**
Clarity (R2)	5 (3-5)	4 (3-5)	5 (4-5)	$\chi2$=4.353	0.113
Accuracy (R3)	4 (4-5)	5 (4-5)	5 (5-5)	$\chi2$=38.273	**<0.001**
Scientificity (R3)	4 (3-5)	5 (3-5)	5 (4-5)	$\chi2$=18.620	**<0.001**
Clarity (R3)	5 (4-5)	5 (5-5)	5 (5-5)	$\chi2$=8.000	1.0
Accuracy (R4)	5 (4-5)	5 (4-5)	5 (4-5)	$\chi2$=0	1.0
Scientificity (R4)	4 (3-5)	4 (3-5)	5 (3-5)	$\chi2$=9.864	**0.007**
Clarity (R4)	5 (5-5)	5 (3-5)	5 (5-5)	$\chi2$=10.000	0.051
Accuracy (average)	4.5 (4-5)	4.75 (3.5-5)	4.75 (4.25-5)	$\chi2$=18.074	**<0.001**
Scientificity (average)	4 (3.5-4.75)	4.25 (3.25-5)	4.75 (4-5)	$\chi2$=22.422	**<0.001**
Clarity (average)	4.75 (4.25-5)	4.5 (3.75-5)	4.75 (4.5-5)	$\chi2$=6.000	0.05
Three categories (average)	4.41±0.21	4.46±0.34	4.73±0.17	F=34.811	**<0.01**

$\chi2$, Friedman test statistic; F, one-way Anova test statistic; SD, standard deviation; Min, minimum; Max, maximum; R, rater (cardiovascular surgery specialist).

Statistically significant P-values are shown in bold.

No statistically significant difference was found among the models in the accuracy category for Rater 4 (P=1.000). A statistically significant difference was found between the models in the scientificity category for R4 (P=0.007). In the pairwise comparisons, this difference was observed between DeepSeek-V3 and ChatGPT-5 (P=0.02), and DeepSeek's scores were found to be significantly higher. No statistically significant difference was found between the models in the clarity category for R4 (P=0.051).

When the models were compared in terms of average accuracy scores, a statistically significant difference was found (P<0.001). In pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 (P=0.034) and DeepSeek-V3 (P<0.001). In contrast, no significant difference was found between Gemini-2.5and DeepSeek-V3 (Table 2).

When comparing the models in terms of average scientificity scores, a statistically significant difference was found (P<0.01). In pairwise comparisons, DeepSeek-V3 scored significantly higher than both Gemini-2.5 (P=0.002) and ChatGPT-5 (P<0.001). No statistically significant difference was found between the models in terms of average clarity scores (P=0.051).

When the average scores for the three categories were compared (Figure 2), a statistically significant difference was found between ChatGPT-5, Gemini, and DeepSeek-V3 (P<0.001). DeepSeek-V3's average scores were found to be significantly higher than both Gemini-2.5 and ChatGPT-5 (4.73 compared to 4.46 and 4.41; P<0.01). In contrast, no significant difference was found between Gemini-2.5 and ChatGPT-5 in terms of average scores (P=1.000).

DISCUSSION

This is among the first studies to the performance of three LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) in a specific medical field where patient education is critically important, such as hemodialysis vascular access, against expert evaluation. The findings reveal that, overall, the DeepSeek-V3 model demonstrated a statistically significant superiority over ChatGPT-5 and Gemini-2.5 in terms of accuracy, scientificity, and overall average scores. However, no significant difference was found among the models in the clarity category. This suggests that DeepSeek-V3's detailed responses do not reduce comprehensibility.

Previous studies evaluating LLMs in patient education across different medical domains have reported generally high performance in terms of clarity and patient-oriented explanations, with more variable

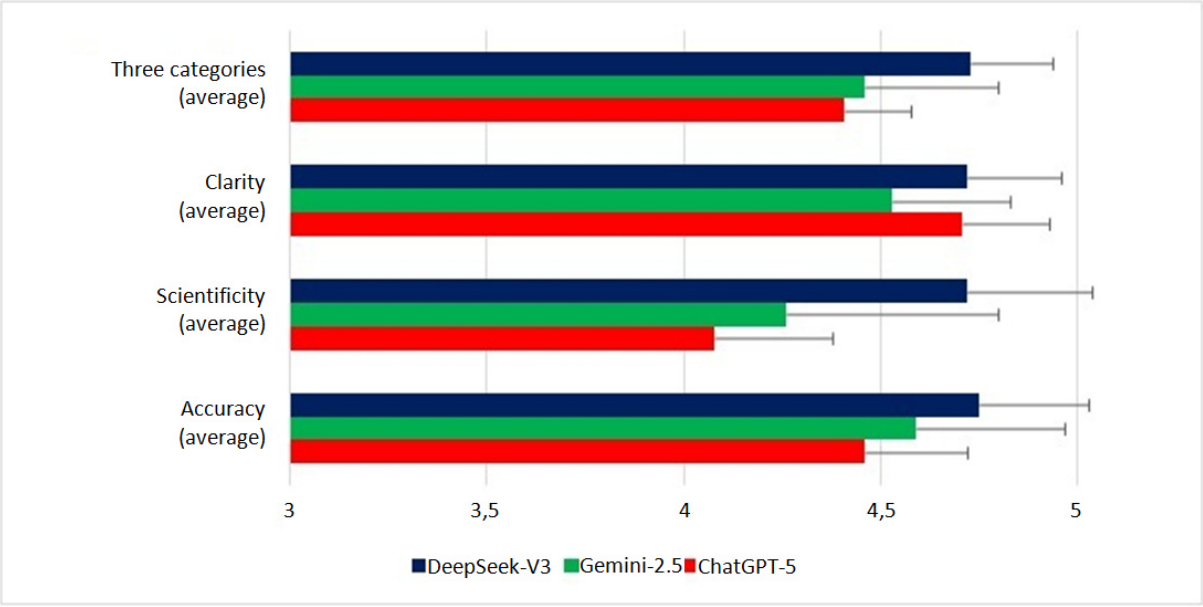


FIGURE 2. Comparison of ChatGPT, Gemini, and DeepSeek's accuracy, scientificity, clarity, and overall average scores using a bar chart.

results regarding scientific depth and accuracy. The literature has shown that LLMs can effectively communicate medical information, but performance can vary significantly depending on the clinical context and evaluation criteria [9, 10]. Consistent with these findings, our results indicate that although all evaluated models performed similarly in terms of clarity, notable differences were observed in scientific depth and accuracy, particularly in a multidisciplinary and procedure-specific field such as hemodialysis vascular access.

The most striking finding of our study is that DeepSeek-V3 scored significantly higher than the other two models on the scientificity criterion. This may be related to the scope, recency, and quality of the medical literature in the model's dataset. In particular, the evaluations by Rater 2 and Rater 3 showed that DeepSeek-V3 tends to present medical content in a more in-depth and evidence-based manner. This finding is extremely important when considering the role of LLMs in providing medical information, as it is essential to maintain scientific accuracy as well as comprehensibility for patients [11]. In this study, DeepSeek-V3's ability to generate responses with a higher word count may have contributed to its high scientific score. However, long responses containing unnecessary details may cause confusion in patient education. Therefore, when evaluating the scientific depth of responses, attention was paid not only to the number of words but also to content density, evidence-based information presentation, and literature references. However, it should be noted that the relationship between response length and quality is not linear; unnecessarily lengthy and off-topic responses can have negative consequences in patient education [12].

In the accuracy category, it is concerning that ChatGPT-5 has a significantly lower average score compared to both Gemini-2.5 and DeepSeek-V3. ChatGPT-5's most noteworthy response was providing clear information that a fistula would mature and be ready for dialysis within four weeks in response to a question about fistula maturation (Question 8: What does it mean for a fistula to “mature”? How long does it take?). It did not provide a time frame and did not mention the valuable “Rules of 6” found in the guidelines [2]. This may indicate that ChatGPT-5 has a higher risk of producing hallucinations in this specific domain. Rater 3's evaluations were where this difference was most pronounced. One of the major criticisms of LLM use in medicine - the risk of hallucination - —has once again been confirmed by the results of our study [13, 14]. The statistically similar performance of Gemini-2.5 and DeepSeek-V3 in terms of accuracy suggests that these models may be more consistent in providing reliable information.

In contrast, Gemini-2.5 demonstrated a more balanced and moderate performance profile across the evaluated domains. While it did not exhibit pronounced strengths comparable to ChatGPT-5 in clarity or to DeepSeek-V3 in scientific depth, it also did not show marked weaknesses. This intermediate performance may suggest a more conservative response generation approach, prioritizing consistency over specialization, which could be advantageous in certain patient education contexts.

Interestingly, all three models scored similarly high in the clarity category, with no statistically significant difference observed between them. This finding demonstrates that current LLMs have generally reached a mature level of ability to express complex medical concepts in understandable language appropriate for the patient level. This is a promising result that strengthens their potential for use as a supportive tool, particularly for patient groups with low health literacy [15, 16].

The scoring variance among evaluators (e.g., Rater 1 not observing a significant difference in scientific accuracy, while other raters did) highlights the role of subjective judgment in evaluating medical information. Different clinical experiences and emphases can lead to different interpretations of the quality of a response. This highlights the value of study designs using multi-rater, blinded methods in evaluating LLM outputs and points to the limitations of assessments based on a single expert opinion.

Strengths and Limitations

This study’s main strength lies in its direct, blinded comparison of three LLMs in the context of hemodialysis vascular access, a field where patient education is critically important. The use of clinically relevant, patient-centered questions and independent evaluation by multiple expert raters enhances the reliability of the findings. Moreover, the separate assessment of accuracy, clarity, and scientific depth

allows for a clear delineation of each model's strengths and limitations in patient education. Nevertheless, this study has some limitations. First, the models' performance was evaluated at a specific point in time (August 2025). Since LLMs are constantly updated and model parameters (e.g., temperature settings) can affect the randomness of outputs, their performance may vary over time. Additionally, differences in model architecture, training datasets, and the rapid update cycles of LLMs may limit direct quantitative comparisons with previously published studies evaluating earlier model versions. Second, although the 25-question set used in the study is comprehensive, it may not cover all possible scenarios and complications related to hemodialysis access. Thirdly, the evaluator panel consisted solely of cardiovascular surgeons, and the exclusion of nephrologists and dialysis nurses is a factor that limits the generalizability of the study. Another limitation of this study is that intraclass correlation coefficient (ICC) analyses did not yield interpretable results due to the inherently subjective nature of expert-based evaluations of LLM-generated responses, which may have contributed to variability in the findings. Finally, the real-world impact of the responses on patient groups (comprehension, compliance, satisfaction) has not been measured; this is an important area for future research.

CONCLUSION

In conclusion, this study demonstrates that DeepSeek-V3 may be a more reliable source in terms of scientific depth and overall quality regarding hemodialysis vascular access; while ChatGPT-5's relative weakness in accuracy reinforces the fact that LLMs should be critically verified rather than automatically accepted for medical use. The success of all models in terms of clarity, however, shows that this technology holds promise as a support tool in patient education. However, ethical concerns must also be considered when using LLMs in patient education. Patients may place a high level of trust in AI-based information, which could create potential risks if incorrect or incomplete information is provided. Therefore, clinicians should review LLM-based educational materials and ensure the accuracy and clinical appropriateness of the information provided.

Ethics Approval and Consent to Participate

As this study was based solely on the evaluation of AI-generated responses and did not involve any patient data, ethical committee approval was not required.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: MHE; Study Design: MHE, ÖFR; Supervision: MHE, ÖFR; Funding: N/A; Materials: N/A; Data Collection and/or Processing: ÖFR, AG, FA, EB; Statistical Analysis and/or Data Interpretation: ÖFR; Literature Review: AG, FA, EB; Manuscript Preparation: MHE; and Critical Review: ÖFR, AG, FA, EB.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The large language model-based chatbots ChatGPT-5 (OpenAI, Microsoft Corporation, San Francisco, CA, USA, 2025), Gemini-2.5 (Google, Mountain View, CA, USA, 2025) and DeepSeek-V3 (DeepSeek, Beijing, CHINA, 2025) were used solely to respond to patient questions about hemodialysis vascular access. The final content was critically reviewed and approved by human authors, who take full responsibility for the integrity and accuracy of the work. The language model was not involved in authorship or editorial decision-making.

The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. Kidney Int. 2019;96(5):1048-1050. doi: 10.1016/j.kint.2019.07.012.
2. Lok CE, Huber TS, Lee T, et al.; National Kidney Foundation. KDOQI Clinical Practice Guideline for Vascular Access: 2019 Update. Am J Kidney Dis. 2020 Apr;75(4 Suppl 2):S1-S164. doi: 10.1053/j.ajkd.2019.12.001.
3. Schmidli J, Widmer MK, Basile C, et al., ESVS Guidelines Committee, EEVS Guidelines Reviewers. Editor's Choice - Vascular Access: 2018 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS). Eur J Vasc Endovasc Surg. 2018;55(6):757-818. doi: 10.1016/j.ejvs.2018.02.001.
4. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav. 2007;31 Suppl 1:19-26. doi: 10.5555/ajhb.2007.31.supp.S19.
5. Skoumalova I, Madarasova Geckova A, Rosenberger J, et al. Health Literacy and Change in Health-Related Quality of Life in Dialysed Patients. Int J Environ Res Public Health. 2022;19(2):620. doi: 10.3390/ijerph19020620.
6. Zavacka M, Skoumalova I, Geckova AM, et al. Does Health Literacy of Hemodialyzed Patients Predict the Type of Their Vascular Access? A Cross-Sectional Study on Slovak Hemodialyzed Population. Int J Environ Res Public Health. 2020;17(2):675. doi: 10.3390/ijerph17020675.
7. Singh P, Goyal L, Mallick DC, et al. Artificial Intelligence in Nephrology: Clinical Applications and Challenges. Kidney Med. 2024;7(1):100927. doi: 10.1016/j.xkme.2024.100927.
8. Ji Z, Lee N, Frieske R, et al. Survey of Hallucination in Natural Language Generation. ACM Comput Surv. 2023;55(12):248. doi: 10.1145/3571730.
9. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29(3):721-732. doi: 10.3350/cmh.2023.0089.
10. Akçay O, Öztürk Ö, Acar T, Gürsoy S. Accuracy and Reliability of ChatGPT in Answering Patient Questions About Lung Cancer and Its Surgery: An Expert Panel Evaluation by Thoracic Surgeons. J Cancer Educ. 2025 Jul 4. doi: 10.1007/s13187-025-02682-3.
11. Almagazzachi A, Mustafa A, Eighaei Sedeh A, et al. Generative Artificial Intelligence in Patient Education: ChatGPT Takes on Hypertension Questions. Cureus. 2024;16(2):e53441. doi: 10.7759/cureus.53441.
12. Sivaramakrishnan G, Almuqahwi M, Ansari S, Lubbad M, Alagamawy E, Sridharan K. Assessing the power of AI: a comparative evaluation of large language models in generating patient education materials in dentistry. BDJ Open. 2025;11(1):59. doi: 10.1038/s41405-025-00349-1.
13. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 2023;11(6):887. doi: 10.3390/healthcare11060887.
14. Huang Y, Gomaa A, Semrau S, et al. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology. Front Oncol. 2023;13:1265024. doi: 10.3389/fonc.2023.1265024.
15. Swisher AR, Wu AW, Liu GC, Lee MK, Carle TR, Tang DM. Enhancing Health Literacy: Evaluating the Readability of Patient Handouts Revised by ChatGPT's Large Language Model. Otolaryngol Head Neck Surg. 2024;171(6):1751-1757. doi: 10.1002/ohn.927.
16. Eid K, Eid A, Wang D, Raiker RS, Chen S, Nguyen J. Optimizing Ophthalmology Patient Education via ChatBot-Generated Materials: Readability Analysis of AI-Generated Patient Education Materials and The American Society of Ophthalmic Plastic and Reconstructive Surgery Patient Brochures. Ophthalmic Plast Reconstr Surg. 2024;40(2):212-216. doi: 10.1097/IOP.0000000000002549.