

DOI: <https://dx.doi.org/10.18203/2320-1770.ijrcog20264008>

Short Communication

Assessing the suitability of ChatGPT and Google Gemini for patient education on the surgical management of gynaecological malignancies

Priyadarshini Jaishankar¹, Sharvari Kulkarni², Krishaa Nandini Suresh Babu^{3*},
Sreya Nimmagadda⁴

¹Department of Obstetrics and Gynaecology, Melmaruvathur Adhi Parasakthi Institute of Medical Science and Research, Melmaruvathur, Tamil Nadu, India

²Maharashtra Institute of Medical Education and Research Medical College, Pune, Maharashtra, India

³Department of Obstetrics and Gynaecology, Pilgrim Hospital, Boston, United Lincolnshire Teaching Hospital, United Kingdom

⁴NRI Medical College and General Hospital, Guntur, Andhra Pradesh, India

Received: 22 July 2026

Revised: 04 September 2026

Accepted: 05 September 2026

*Correspondence:

Dr. Krishaa Nandini Suresh Babu,
E-mail: sknandini2014@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Artificial intelligence (AI)-based conversational tools like ChatGPT and Google Gemini are increasingly being explored for their potential in patient education. In gynaecological oncology, clear and accessible communication about surgical options such as hysterectomy is vital. This study investigates the effectiveness of these tools in generating patient-friendly educational content. Educational content on total hysterectomy, radical hysterectomy, laparoscopic hysterectomy, abdominal hysterectomy, and supracervical hysterectomy was generated using ChatGPT and Google Gemini. Responses were analysed for word count, sentence structure, grade level, ease score, similarity percentage, and reliability. An independent sample t test was used for comparison. Shapiro-Wilk and Levene's tests were used to assess normality and variance equality, respectively. The analysis showed no statistically significant differences between ChatGPT and Google Gemini across all assessed metrics ($p > 0.05$). Both tools produced similar content in terms of word count, sentence length, readability (ease score and grade level), and reliability. The average reliability score was slightly higher for Gemini, while ChatGPT showed slightly better readability. Both ChatGPT and Google Gemini demonstrate comparable effectiveness in generating suitable educational content on surgical management of gynaecological malignancies. With appropriate oversight, they can serve as supplementary tools in enhancing patient understanding and shared decision-making in clinical settings.

Keywords: ChatGPT, Google Gemini, Gynaecological malignancies, Patient education, Hysterectomy, Artificial intelligence in healthcare, Total hysterectomy

INTRODUCTION

Gynaecology malignancies are cancer which involves women's reproductive organs. The five main organs involved in reproductive cancer are cervical, ovarian, uterine, vaginal and vulvar. Gynaecological cancers are heterogeneous diseases with varying risk factors, management paradigms and outcomes.¹ Due to the

increasing trend of obesity and overall public awareness in the general population.² These gynaecological malignancies have a broad range in the surgical and adjuvant treatments available across different countries.³ Therefore it is a priority to develop better methods for patient's education to understand the importance of these malignancies and its surgical management. The patient education helps in providing insight on identification of complications of surgical management, maintenance of

their corresponding adjuvant medication. Thereby ensuring early intervention to the said complications and preventing early mortality and morbidity.

As there is an increase in the number of individuals who turn to online platforms for health-related inquiries and truly embrace the responses given by such AI conversational tools.⁴ There is increased responsibility for these AI to ensure that right information has reached the patient which was accessible to them at their own comfort through these educational guides. Unfortunately, their accuracy still needs to be verified by the physicians, who can lose their personal connection with their patients. The patients can also develop over reliance with these tools and choose the options given by the AI over the carefully curated version given to them by the physician.

Chat-GPT is a generative AI chat box. It is based on GPT-4o large language which can generate human-like conversations toward a desired length, format, style, level of detail and language. It comes under a proprietary series of generative pre-trained transformers.^{5,6}

Google Gemini is a generative AI chatbot developed by google. It was based on a large language model which is a type of machine learning model designed for natural language processing tasks such as language generation.⁷ The aim of this study is to provide an assessment on the comparison between Chat- GPT and Google Gemini generated responses for writing patient education guides on surgical management of gynaecological malignancy. The AI tools were compared on the basis of comprehensibility and ease of understanding. Thereby showing the importance of the brochure.

METHODS

Study design

A cross-sectional comparative study was conducted over a period of three months to evaluate and compare the suitability of patient information guides generated by ChatGPT and Google Gemini for surgical management of gynaecological malignancies. Data collection commenced on 12th January 2025 and was completed within the three-month study period.

As this study involved the evaluation of text generated by publicly available AI language models and did not involve human participants, patient data, or biological specimens, institutional ethics committee approval was not required.

Data collection

The process of data collection was equally distributed among the participating researchers. Each of the fellow researchers were responsible for collecting data pertaining to different surgical procedures in the field of gynaecology. Six common surgical procedures in the field of obstetrics and gynaecology were selected. The surgical

management includes-Total hysterectomy, radical hysterectomy, laparoscopic hysterectomy, abdominal hysterectomy, supracervical hysterectomy and robotic hysterectomy.

The two AI tools selected, namely ChatGPT version 4.0 and Google Gemini, were used for generating brochures for patient education. Prompts were given to both AI tools: write a patient guide for total hysterectomy, write a patient guide radical hysterectomy, write a patient guide laparoscopic hysterectomy, write a patient guide abdominal hysterectomy, write a patient guide supracervical hysterectomy, write a patient guide robotic hysterectomy.

Responses generated were then collected in a Microsoft word document and graded by the help of three standardised materials which includes:

Flesch- Kincaid calculator

This calculator is responsible for providing the readability of the patient guide provided by the respective AI tools.⁸ In our research it was used for word count, sentence count ease of understanding which helped in assessing the readability of the patient guide by the common men.⁹

Quillbolt plagiarism tool

This tool was used for finding the similarities of the information provided by the AI tools. This helped in identifying the plagiarism.¹⁰ Between the corresponding information provided by ChatGPT and Gemini.¹¹

Modified DISCERN score

It was used for assessing the reliability of the material provided by the AI tools. It is based on a 5 point scoring system which evaluates the reliability of the medical information with higher scores indicates greater reliability.¹²

Statistical analysis

The data was then exported to Microsoft Excel sheet. After which statistical analysis was done using R version 4.3.2. R core team (2023). R: A Language and Environment for Statistical Computing. R foundation for statistical computing, Vienna, Austria. Independent sample t- test has been used to compare the means between ChatGPT and Google Gemini. Normality of the variables was assessed using Shapiro-Wilk Test and equality of variances was tested using Levene's Test. Based on the p values obtained, there is no statically significant difference between any of the responses generated by the 2 AI tools.

RESULTS

ChatGPT and Google Gemini was used to generate brochures on patient education for total hysterectomy,

radical hysterectomy, laparoscopic hysterectomy, abdominal hysterectomy and the supracervical hysterectomy.

Table 1 compares the characteristics of the patient guides generated by the two AI platforms. There were no statistically significant differences between ChatGPT and Google Gemini in terms of total word count (792.8 ± 465.2 vs. 793.2 ± 451.7 words, $p=0.999$), number of sentences (93.8 ± 43.5 vs. 76.2 ± 45.1 , $p=0.547$), average syllables per word (1.96 ± 0.09 vs. 1.98 ± 0.08 , $p=0.724$), Flesch-Kincaid grade level (10.8 ± 1.17 vs. 11.9 ± 0.87 , $p=0.130$), Flesch Reading Ease score (32.5 ± 7.6 vs. 28.6 ± 6.6 , $p=0.415$), similarity percentage ($50.0 \pm 13.9\%$ vs. $52.4 \pm 20.9\%$, $p=0.839$), or modified DISCERN reliability score (3.0 ± 0.0 vs. 3.4 ± 0.55 , $p=0.178$).

The average number of words per sentence was lower for ChatGPT than for Google Gemini (8.38 ± 1.76 vs. 10.52 ± 1.10), approaching statistical significance ($t=$

2.305 , $p=0.050$), suggesting that ChatGPT tended to produce shorter and potentially more readable sentences. However, this difference did not reach conventional statistical significance when adopting a threshold of $p < 0.05$. Overall, both AI platforms generated patient information guides with comparable readability and reliability metrics.

Figure 1 compares the readability, similarity, and reliability of patient education guides generated by ChatGPT and Google Gemini. Both AI models produced educational materials with comparable readability and reliability across all hysterectomy procedures. ChatGPT generally generated shorter sentences and achieved slightly higher Flesch Reading Ease scores, whereas Gemini demonstrated marginally higher modified DISCERN scores for radical and abdominal hysterectomy. Similarity analysis demonstrated moderate overlap between the two AI-generated texts, ranging from approximately 25% to 70% across procedures.

Table 1: Characteristics of responses generated by ChatGPT and Google Gemini.

Variables	ChatGPT		Google Gemini		T (8)*	P value
	Mean	SD	Mean	SD		
Words	792.80	465.19	793.20	451.65	-0.001	0.999
Sentences	93.80	43.45	76.20	45.07	0.629	0.547
Average words per sentence	8.38	1.76	10.52	1.10	-2.305	0.050
Average syllables per word	1.96	0.09	1.98	0.08	-0.365	0.724
Grade level	10.80	1.17	11.90	0.87	-1.689	0.130
Ease score	32.50	7.56	28.64	6.60	0.860	0.415
Similarity %	50.02	13.93	52.38	20.88	-0.210	0.839
Reliability score	3.00	0.00	3.40	0.55	-1.633	0.178

*T (8) is the t statistic at 8 degrees of freedom. P-values < 0.05 are considered statistically significant.

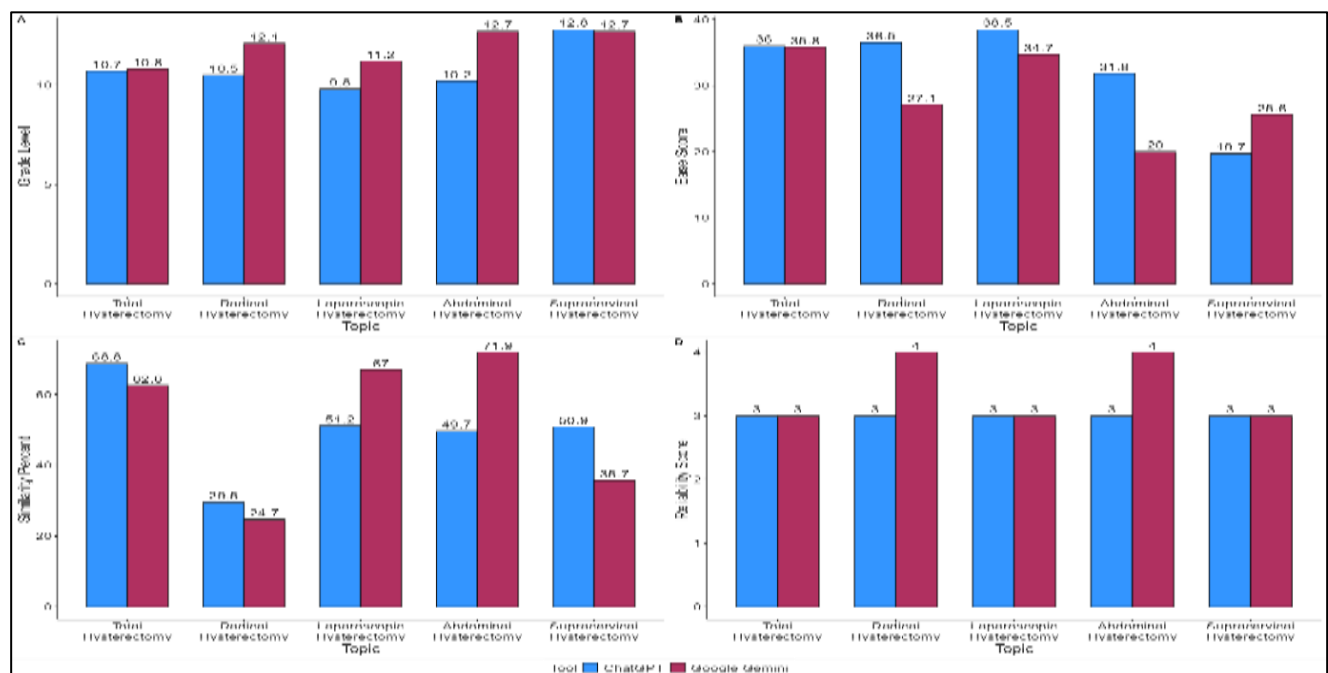


Figure 1 (A-D): Graphical representation of comparison between grade level, ease score, similarity percent and reliability score, for the patient education guide generated by ChatGPT and Google Gemini.

DISCUSSION

A cross-sectional study conducted to compare responses generated by two AI tools ChatGPT and Google Gemini for brochures on patient education for total hysterectomy, radical hysterectomy, laparoscopic hysterectomy, abdominal hysterectomy and supracervical hysterectomy, revealed that there was no statistical difference between the responses of both AI tools. AI is being used by almost everyone to gather information. There have been a lot of studies about how AI is generating responses and if they are readable and accurate.¹³

For easy-to-read brochures, the level of understanding (ease score) should be such that even a high school person can understand. The recommended level is that a 6-8th grade person must understand the resources available to people for better understanding.¹⁴ This would help for better patient education.

AI is available for free and would be an important tool that can be utilised to increase patient education. Reading material available online can be converted into text that was easier to read by AI in one study.¹⁴

In this study, Chat GPT and Google Gemini was used to create patient education guide on types of hysterectomy. The mean ease score of ChatGPT was 32.5 and that of Google Gemini was 28.6-which, according to Flesch Kincaid scale means that only a college graduate can understand the responses generated.⁹

This means that the AI tools generated education guides were not catering to high school graduated and would not be helpful in educating people, universally. Giving prompts to create education guide with instructions that high school person can understand would improve the outcomes from AI. In one particular study, information on orthopaedic diseases was converted to a text understood by a fifth-grader by AI.¹⁵ This shows that the AI is a great help to improve patient education, but needs to be screened by professionals.

Chatbots provide answers to personalised queries unlike medial pamphlets and enlighten the patients and reduce their anxiety.¹⁶ Another study used 5 different AI tools to study patient education material for *H. pylori*. They found that the information was reliable but not complete. AI also could not translate these English texts in other language properly.¹⁷

The readability of the patient education material was high, when AI was used to give the information on an ostomy guide.¹⁸

AI is pre-trained to use existing literature, which can lead to plagiarism. There was around 50% plagiarism found in both the responses of AI. Plagiarism is considered a misconduct in medical literature. Scientific writing should be accurate and plagiarism leads to lack of honesty and

creativity.¹⁹ As AI can access large amount of data available on the internet, it can lead to similarities in the responses to the scientific literature-leading to ethical issues.²⁰

Discern score is a standardized score which assesses the quality of healthcare information available online.²¹ The average discern score of google Gemini was slightly higher (3.40) as compared to ChatGPT which was an average of 3. This signifies that Gemini Gives slightly more correct responses than ChatGPT. The findings of reliability using discern score is like other studies done for studying the AI responses.²²

One similar study compared responses of both AI for generating patient education guides for epilepsy. In that case also the discern score from both AI was similar.²³

Limitations

There were some limitations to this study. There are lot of Artificial Intelligence tools available on the internet. In this study, we have only used two of them. Other tools should also be used to see which gives better information for patient education. This study only included the types of hysterectomy. There are innumerable diseases, out of which the most common ones should also be studied. ChatGPT used is 4.0 version, which is a older version. A newer version might provide better information.

CONCLUSION

The findings from this study suggest that both AI tools produced content with comparable readability, linguistic quality, and reliability, with no statistically significant differences across all assessed parameters. Although Google Gemini demonstrated slightly higher reliability scores and ChatGPT showed marginally better readability, both tools failed to generate text at a level suitable for universal patient comprehension, as the average ease scores indicated a college-level reading requirement. This limits their standalone utility in broad-based patient education.

Nevertheless, with appropriate clinician oversight, tailored prompting, and content simplification, AI-generated materials can serve as supplementary educational resources to enhance patient understanding, support informed consent, and reduce preoperative anxiety. Their ability to rapidly generate consistent, structured, and relevant information highlights their potential role in modern clinical practice.

Future research should explore other surgical procedures, involve multilingual assessment, and test newer AI versions to assess progress in reliability and accessibility. While AI cannot replace clinician communication, it can be a valuable adjunct in improving patient engagement and health literacy.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

1. Suneja G, Viswanathan A. Gynecologic malignancies. *Hematol/Oncol Clin.* 2020;34(1):71-89.
2. Blair AR, Casas CM. Gynecologic cancers. *Primary Care Clin Office Practice.* 2009;36(1):115-30.
3. Fotopoulou C, Kraetschell R, Dowdy S, Fujiwara K, Yaegashi N, Larusso D, et al. Surgical and systemic management of endometrial cancer: an international survey. *Arch Gynecol Obstet.* 2015;291(4):897-905.
4. Hernandez CA, Gonzalez AE, Polianovskaia A, Sanchez RA, Arce VM, Mustafa A, et al. The future of patient education: AI-driven guide for type 2 diabetes. *Cureus.* 2023;15(11):e48919.
5. OpenAI. ChatGPT-Release notes. San Francisco (CA): OpenAI. Available at: <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>. Accessed on 08 August 2026.
6. Roumeliotis KI, Tselikas ND. Chatgpt and open-ai models: A preliminary review. *Future Internet.* 2023;15(6):192.
7. Google DeepMind. Gemini 2.5: Our most intelligent AI model. 2025. Available at: <https://deepmind.google/technologies/gemini/>. Accessed on 08 August 2026.
8. Williamson JM, Martin AG. Analysis of patient information leaflets provided by a district general hospital by the Flesch and Flesch-Kincaid method. *Int J Clin Pract.* 2010;64(13):1824-31.
9. Good Calculators. Flesch Kincaid Calculator. Good Calculators; 2019. Available at: <https://goodcalculators.com/flesch-kincaid-calculator>. Accessed on 08 August 2026.
10. QuillBot. QuillBot AI Writing and Paraphrasing Tool. Chicago (IL): QuillBot. Available at: <https://quillbot.com>. Accessed on 08 August 2026.
11. Karnan N, Francis J, Vijayvargiya I, Tan CR, Rubino C. Analyzing the effectiveness of AI-generated patient education materials: A comparative study of ChatGPT and Google Gemini. *Cureus.* 2024;16(11):1.
12. Adithya S, Aggarwal S, Sridhar J, Kavya VS, John VP, Singh C. An observational study to evaluate readability and reliability of AI-generated brochures for emergency medical conditions. *Cureus.* 2024;16(8):10.
13. Paluszek O, Loeb S. Artificial intelligence and patient education. *Curr Opinion Urol.* 2025;35(3):219-23.
14. Shah YB, Ghosh A, Hochberg A, Mark JR, Lallas CD, Shah MS. Artificial intelligence improves urologic oncology patient education and counseling. *Can J Urol.* 2024;31(5):12013-8.
15. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? *Clin Orthopaed Related Res.* 2023;481(11):2260-7.
16. Didier AJ, Moss G, Sutton JM. Applications of artificial intelligence for cancer patient education. *J Cancer Educat.* 2025;40(3):313-4.
17. Zeng S, Kong Q, Wu X, Ma T, Wang L, Xu L, et al. Artificial Intelligence-Generated Patient Education Materials for Helicobacter pylori Infection: A Comparative Analysis. *Helicobacter.* 2024;29(4):e13115.
18. Yücel Kaçmaz H, Kahraman H, Akutay S, Dağdelen D. Development and validation of an artificial intelligence-assisted patient education material for ostomy patients: a methodological study. *J Adv Nursing.* 2025;81(7):3859-67.
19. Min SK. Plagiarism in medical scientific research: can continuing education and alarming prevent this misconduct? *Vascular Specialist Int.* 2020;36(2):53.
20. Karnan N, Francis J, Vijayvargiya I, Tan CR, Rubino C. Analyzing the effectiveness of AI-generated patient education materials: A comparative study of ChatGPT and Google Gemini. *Cureus.* 2024;16(11):1.
21. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Commun Health.* 2019;53(2):105-11.
22. Varsaale N, George CR, Gunturi LM, Saravanan H, Menon GA. A cross-sectional study comparing patient education guides created by ChatGPT and Google Gemini for common cardiovascular-related conditions. *Cureus.* 2025;17(1):1.
23. Arora S, Ramesh M, Moe AT, Giri T, Parrikk K, Challa HV. Evaluating artificial intelligence (AI)-generated patient education guides on epilepsy: a cross-sectional study of ChatGPT and Google Gemini. *Cureus.* 2024;16(11):e73212.

Cite this article as: Jaishankar P, Kulkarni S, Suresh Babu KN, Nimmagadda S. Assessing the suitability of ChatGPT and Google Gemini for patient education on the surgical management of gynaecological malignancies. *Int J Reprod Contracept Obstet Gynecol* 2026;15:4563-7.