IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 24, Issue 8 Ser. 1 (August. 2025), PP 01-07
www.iosrjournals.org

Comparative Evaluation: Validity And Reliability Of Artificial Intelligence As Public Sources Of Information On Composite Restoration.

Ritika Chellani, Dipti Barve, Pawan Patel, Pranav Dave, Vinay Sharma, Lakshmi Pandey

(Bds, Undergrad Student, Department Of Conservative Dentistry And Endodontics, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).
(Bds, Associate Professor, Department Of Conservative Dentistry And Endodontics, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).
(Bds, Professor, Department Of Conservative Dentistry And Endodontics, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).
(Bds, Professor, Department Of Pediatric Dentistry, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).
(Bds, Associate Professor, Department Of Conservative Dentistry And Endodontics, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).
(Bds, Senior Lecturer, Department Of Conservative Dentistry And Endodontics, Smbt Institute Of Dental Science And Research, Nashik, Maharashtra, India).

Abstract:
Aim:
The study aimed to evaluate and compare the validity and reliability of responses generated by Google Gemini to frequently asked questions related to composite restoration in dentistry.
Methodology:
Frequently asked questions were formulated by senior endodontists (n=3) and some from ChatGPT (n=25), These questions were posed to ChatGPT and GoogleGemini separately twice. These responses were separately evaluated by Three senior endodontists using a 5-point Likert scale and a Quality score. Shaprio-wilk test and ANOVA test were conducted for statistical analysis.
Result:
Both the Chatbots responded to all 25 questions. The validity of ChatGPT and Google Gemini had no significant difference between them. Both chatbots demonstrated a good level of reliability.
Conclusions*:*
ChatGPT and Google Gemini provided credible information on topics related to composite restoration.
Keywords: *ChatGPT; Google Gemini; Artificial Intelligence (AI); Composite Restoration*

Date of Submission: 23-07-2025 Date of Acceptance: 03-08-2025

I. Introduction:

Artificial Intelligence (AI) Chatbots have shown a new light to the digital world by providing information and a chance to ask tailored queries they have in their mind and the book may or may not answer them, Chatbots are learning new information daily due to their deep learning algorithm. This is provided by intimating as human brain. ChatGPT (OpenAI Inc.) and Google Gemini (Google LLC) are platforms that are leading in this evolution.

The generative pre trained transformer (ChatGPT) that is powered by openAI is based on based model transformation and is trained on diverse internet text. By analysing its patterns it employs unsupervised learning and improves its responses. (Tang Z)[1]. ChatGPT achieved 100 million users in January 2023, only 2 months after it was available in the market (Milmo, 2023)[2]. On the other side of the road, Google Gemini authorities the skills of Google in search and language.Google Gemini also uses transformers by bidirectional encoder representations (BERT) that have been specifically developed for understanding the context of words for better representation of language(Devlin et al, 2018][3];[Malik Sallam, 2025)[4]. These chatbots are actively interacting with their users. When Chatbots receive information from the user, This information is then processed in coding language by using a deep learning model. Then chatbots try to foresee the next possible response based on its

previously learned data patterns.

Chatbots are formulated as a source of information and facilitate data on comprehensive topics incorporating healthcare information. These can improvise patients understanding related to diagnosis, treatment , prognosis, outcomes as well as side effects and complications by providing to thiers queries . (Rasu ,2023)[5]. Valid and reliable content on the health care related information are critical for improving patient awareness of thier condition and avoid any harm . So they should be provided with accurate, evidence-based information from standardized sources, so they can make decisions about thier treatment. (Edwards et al ,2001][6].

This will allow patients to correctly understand thier disease, treatment, prognosis. These will allow them to provide consent , treatment option and what to expect as result (Austvoll-Dahlgren &Helseth, 2010)[7]. (Suárez et al.2023)[8], conducted a study regarding performance of ChatGPT related to endodontics questions based on Position Statements by European Society of Endodontology. Responses provided was 57.33% accurate. The performance of AI in the field of endodontics has not been assessed till now, so aim of this study was to compare & evaluate the validity and reliability of responses provided by two chatbots, ChatGPT and Google Gemini to frequently asked questions (FAQS) on composite restoration.

II. Materials And Methods

Data collection

TwentyFive Frequently asked questions on composite restoration were formulated/selected, representing Patients concern. Sources of these questions are given below.

A. list of top 30 FAQs on composite restoration was Provided From ChatGPT

B. Some from suggested from endodontists that encounter Patient on daily bases .

25 questions are enumerate below in Table 1. Each question was posted twice to both the chatbotsto assess thier reliability.

Questions
1. What is composite restoration?
2. How does composite restoration differ from Amalgam filling?
3. What are the advantages of composite restorations?
4. How long does a composite restoration Typically last?
5. What is the process of getting a composite restoration?
6. Can composite restoration be used for front teeth ?
7. Are there Any limitations or considerations with composite Restoration?
8. Is composite restoration suitable for children ?
9. Can composite restoration be repaired if damaged?
10. How can I care for my teeth after getting a composite restoration?
11. Can Composite restoration be used to replace old amalgam filling?
12. Are there different types of composite materials used in restorations?
13. Is composite restoration more expensive than amalgam fillings ?
14. What should I do if my composite restoration feels rough or uncomfortable ?
15. Can composite restorations Stains over time ?
16. Are there Any risks associated with composite restorations?
17. Will composite filling match my natural tooth Color perfectly ?
18. Is the composite restoration process painful ?
19. What are alternative of composite restorations
20. Are composite filling more prone to decay compared to amalgam fillings?
21. Can composite filling be used for cavity In the back teeth?
22. Can composite restorations be used in people with braces ?
23. How Can I prevent my composite fillings from staining ?
24. What are smart Composite materials?
25. Can a composite restoration be placed over a root canal treated tooth ?

Scoring

The responses were independently evaluated by Three senior endodontists using a 5-point Likert scale. Global Quality Score (GQS) modified version was used to assign the responses (Bernard et al.,2007)[9].

Score 1 (Agree): The answer is Satisfactory

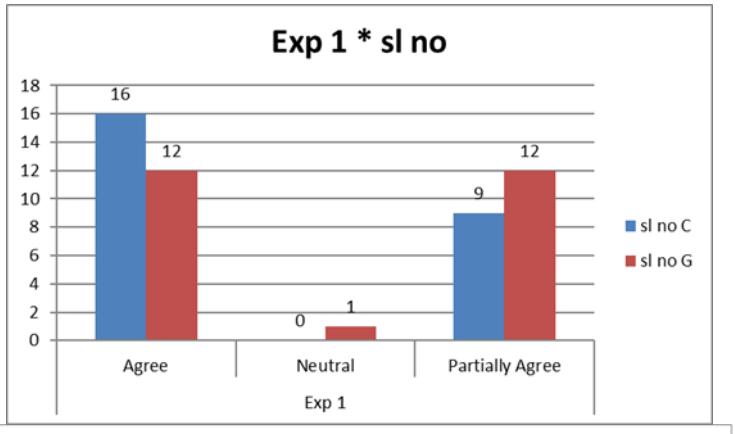
Score 2 (Partially Agree): Most of the content of the answer is correct.

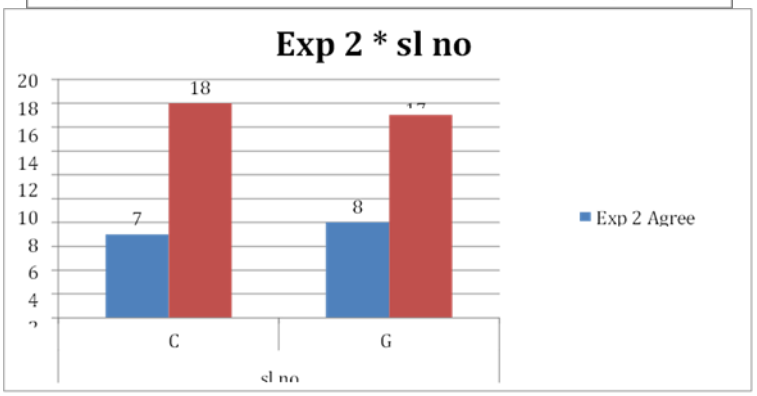
Score 3 (Neutral): The answer has some missing content in it .

Score 4 (Partially Disagree): Most of the answer is incorrect.

Score 5 (Disagree): The answer is incorrect.

Some criteria were applied to evaluate the response generated from Chatbots ,considering both the accuracy of the information and authenticity of answer obtained.





Once all the responses were obtained, the score sheet was excharged between three reviewers, responses were reviewed and discussed. Disagreement was resolved through evidence-based discussions . For statistical analysis a single score sheet as prepared for all 50 responses.

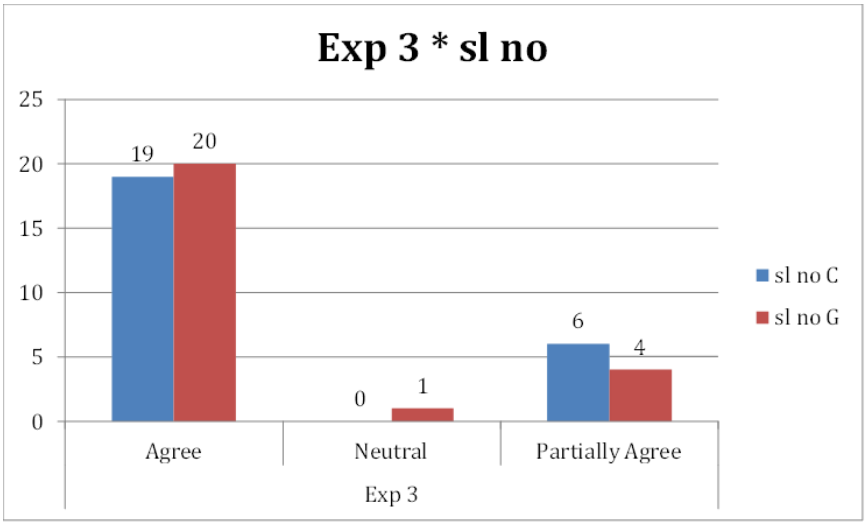
Statistical analysis

Analysis of Validity

For analysis the validity of responses (ie. To check the degrees of accuracy of intended answer), two categories were made as valid and invalid. As well as two tests were involved: low threshold validity test and high threshold validity test.

As acceptance level of low threshold was 4 . If responses ≥4 , responses was valid . If responses scored <4, then it was considered invalid.

For high-threshold acceptance level was 5 , if responses scored 5 then it was valid , and if <5 then it was invalid. For this Shaprio-wilk test was conducted to compare the validity of answers between both the chatbots . The level was set at<0.5.

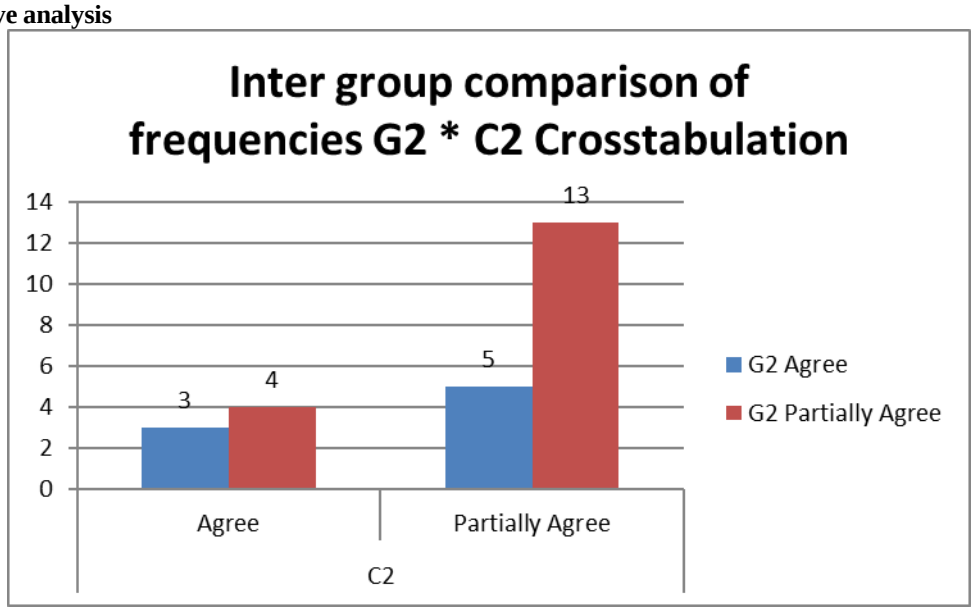


Analysis of reliability

To define reliability (producting of similar answer when repeatedly asked). Was also analyzed by 5 points Likert scale score which was obtained by asking same question twice. To evaluate the consistency of responses, anova test was used for all 25 questions responses.

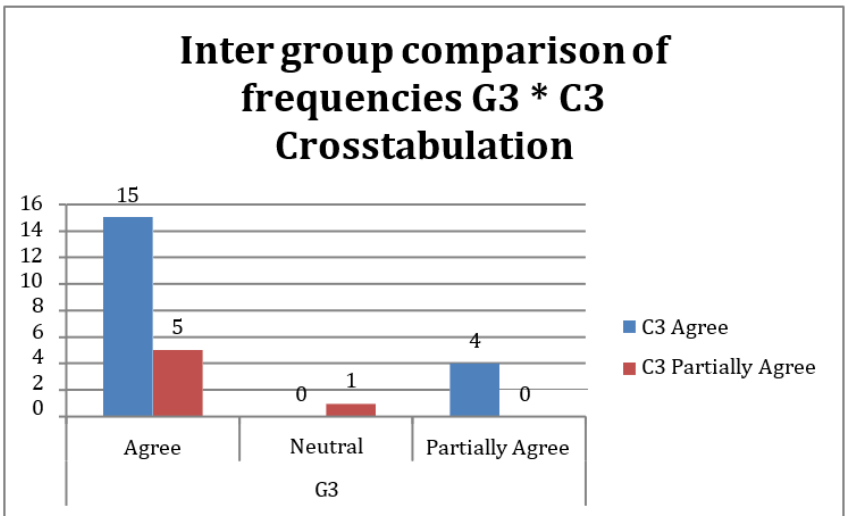
III. Result

Descriptive analysis



The two chatbot responses to 25 questions a total of 50 answers were obtained.
The answers from ChatGPT and Google Gemini were long and responses contained satisfactory details.
Both the chatbot reminded their users to see a dentist or an endodontist for further queries to the same questions.

Statistical analysis:



IV. Discussion:

AI Chatbots are new emerging and easy source of information with ability to change the mindset of the a person to receive and process data , including healthcare (Gilson A,2023)[10] . For the potential of changing the mindset of the individuals , it is necessary for researchers to evaluate the response provided by chatbots and spread awareness to society about there benefits and threats of data .

I t is impractical to eliminate misinformation that are being spread through chatbots. So, researchers and scientists should take this upon their own self to verify the content provided by them. Chatbots usage has increased in healthcare in today's era to decrease healthcare cost and increase patient satisfaction.

(Campbell K,2020)[11]. In field of dentistry, particularly in endodontics, the implementation of chatbots

for asking frequently asked questions by patients serves as significant resources for individuals seeking information regarding denial condition, prossible care and treatment results.

Questions that were used , were aim to reflect public's concern related to composite restoration in dentistry. Twenty Five questions that were used were formulated by senior endodontists that are daliy in contact with patients. Other options was to personally ask patients about there queries regarding composite restoration. But this method is challenging to collect queries from from different socioeconomic/ place/ mindset of people worldwide. Some questions were formulated from ChatGPT as most frequently asked questions in field of endodontics regarding composite restoration. By querying ChatGPT to formulated its most asked questions in endodontics, a representative of selection of text Utilized for ChatGPT training was acquired, covering a wide range of online materials and scholarly publications.(Kung HT , 2022][12]; [Lievin V 2022)[13].

The drawback of this study was that , the study was conducted on a small scale, to assess the responses only three senior Endodontists were involved. It would have been more engaging to have larger panel for validation of responses . So it is advisable for endodontics associations should take the task to publish regarding evaluation of qulity of endodontics information provided by chatbots.

Statistical analysis showed that there was no significant difference between the frequencies of group (p>0.05). In some severe condition, endodontic diseases are life threatening associated with significant health complications and fatalities at both dental and general level. In case of apical abscess , if it is left untreated or improperly treated it can led to death. (Rampa et al ,2019)[14]. The publications of misinformation with limited access to credible information can lead to a mass scale distribution in healthcare. Recently public reluctant to seek dental treatment during Coronavirus Disease 2019 (COVID- 19) due to fear of contracting the SARS-CoV-2 virus(Jessica Langella ,2020][15];[Yu et al ,2020)[16], as well as misinformation about relation between infertility and COVID-19 vaccine was spread that lead to hesitation in taking the vaccine . (Abbasi, 2022)[17].

A through and meticulous assessment should be done about the information that is generated by chatbots on endodontics before they are entitled as trustworthy sources to public. The responses provided by chatbots can be diverse due to philosophies of different companies. Different algorithm they have employed, training datasets and objectives (Bhardwaz & Kumar, 2023][18].

Consistency is a integral part of assessing performance of chatbots. Both the chatbot have an justifiable level of reliability. Chatbots are designed on deep algorithm learning that have some level of inherent. So, the responses generated are not necessitation. (Lu et al, 2023)[19]. Both the chatbot have prospective to generate content with more extensive understanding of topic.

Inter group comparison of frequencies of G1 * C1 Crosstabulation

		C1						
		Agree	Partially Agree	Total	Chi-Square value	P value of Chi-Square test	Kappa Value	P value of Kappa Value
G1	Agree	7	5	12	.767	.682#	.	.
	Neutral	1	0	1				
	Partially Agree	8	4	12				
	Total	16	9	25				

There was a statistically non significant difference seen for the frequencies between the groups (p>0.05)

Inter group comparison of frequencies of G2 * C2 Crosstabulation

		C2						
		Agree	Partially Agree	Total	Chi-Square value	P value of Chi-Square test	Kappa Value	P value of Kappa Value
G2	Agree	3	5	8	.527	.468#	.144	.206#
	Partially Agree	4	13	17				
	Total	7	18	25				

There was a statistically non significant difference seen for the frequencies between the groups (p>0.05)

Inter group comparison of frequencies of G3 * C3 Crosstabulation

		C3						
		Agree	Partially Agree	Total	Chi-Square value	P value of Chi-Square test	Kappa Value	P value of Kappa Value
G3	Agree	15	5	20	4.441	.109#	.	.
	Neutral	0	1	1				

	Partially Agree	4	0	4				
	Total	19	6	25				

There was a statistically non significant difference seen for the frequencies between the groups ($p>0.05$)

Both the Chatbots have high valid scores, but there were some error in there content. Many responses could have prospective to mislead public on some specific topics.

Information provided by chatbots often lack depth of understanding, that shows that they have limited knowledge of complex topics. A significant drawback of chatbots which was observed related to healthcare questions was that they are not able to provide references or citations of information. Without reference, they claim there responses to be authentic , that can be verified. This is issue can be problematic as the information can be from outdated, factually incorrect, incomplete or biased.

Additionally, this information can be from sources that are not public health organizations or high-impact scientific journals. This absence of transparency regarding the references of data used by chatbots raised doubts about reliability of them for healthcare information. There are pressing need to enhance the capabilities of chatbots to reliable data sources and it establish responses to health care quires.

V. Conclusion:

Study though light on the use of AI Chatbots as public sources of information on composite restoration . Both the chatbot had optimistic results concerning the validity of responses. There was no significant different between their validity and reliability. As chatbot are becoming an integral part , it is crucial for researchers to evaluate their achievements and publicize their shortcomings to keep the public informed of valid information and possible deception provided by chatbots .

Author contributions

Funding Information
No Funding was provided

Conflict Of Interest Statement
There was no Conflict of interest.

Data Availability Statement
Data availability statement are mentioned in supplementary materials.

Ethics Statement
There was no involvement of human so no need for ethical approval.

References

[1] Tang Z, Kejriwal M. A Pilot Evaluation Of Chatgpt And DALL-E 2 On Decision Making And Spatial Reasoning. Cornell University Research Journal 2023 .
[2] Milmo, D. (2023) Chatgpt Reaches 100 Million Users Two Months After Launch. The Guardian 2023.
[3] Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pretraining Of Deep Bidirectional Transformers For Language Understanding Cornell University 2018 .
[4] Malik Sallam , Kholoud-Al-Mahzoum ,Mohammed Sallam, Maad M Mijwil, Deepseek: Is It End Of Generative AI Monopoly Or The Mark Of The Impending Doomsday? Mesopotamian Journal Of Big Data 2025.
[5] Rafia Rasu, Walter Agbor Bawa, Richard Suminski , Kathleen Snell, Bradley Warady Health Literacy Impact On National Healthcare Utilization And Expenditure. International Journal Of Health Policy And Management 2015.
[6] Edwards, A., Elwyn, G., Smith, C., Williams, S. & Thornton, H. Consumers' Views Of Quality In The Consultation And Their Relevance To 'Shared Decision-Making' Approaches. Health Expectations 2001.
[7] Austvoll-Dahlgren, A. & Helseth, S. What Informs Parents' Decision-Making About Childhood Vaccinations? Journal Of Advanced Nursing 2010.
[8] Suárez, A., Díaz-Flores García, V., Algar, J., Gómez Sánchez, M., Llorente De Pedro, M. & Freire, Y. Unveiling The Chatgpt Phenomenon: Evaluating The Consistency And Accuracy Of Endodontic Question Answers. International Endodontic Journal. 2010.
[9] Bernard, A., Langille, M., Hughes, S., Rose, C., Leddin, D. & Van Zanten, S.V. A Systematic Review Of Patient Inflammatory Bowel Disease Information Resources On The World Wide Am J Gastroenterol. 2007.
[10] Gilson A, Safranek C, Huang T, Socrates V, Chi L, Taylor RA, Et Al. How Does Chatgpt Perform On The United States Medical Licensing Examination? The Implications Of Large Language Models For Medical Education And Knowledge Assessment. JMIR Med Educ. Feb 08, 2023.
[11] Campbell K, Louie P, Levine B, Gililland J. Using Patient Engagement Platforms In The Postoperative Management Of Patients. Curr Rev Musculoskelet Med 2020 .
[12] Kung HT, Cheatham M, Chatgpt, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V , Performance Of Chatgpt On USMLE: Potential Or AI-Assisted Medical Education Using Large Language Models. PLOS Digital Health 2022.
[13] Liévin V, Egeberg Hother C, Winther O , Can Large Language Models Reason About Medical Questions? Cornell University, 2022.

[14] Rampa, S., Veeratrishul, A., Raimondo, M., Connolly, C., Allareddy, V. & Nalliah, R.P. Hospital-Based Emergency Department Visits With Periapical Abscess: Updated Estimates From 7 Years. Journal Of Endodontics, 2019 .
[15] Jessica Langella, Britta Magnuson, Matthew D Finkelman, Robert Amato Clinical Response To COVID-19 And Utilization Of An Emergency Dental Clinic In An Academic Institution Journal Of Endodontics 2020.
[16] Jingjing Yu Et Al , Characteristics Of Endodontic Emergencies During Coronavirus Disease 2019 Outbreak In Wuhan. Journal Of Endodontics 2020.
[17] Abbasi, J. Widespread Misinformation About Infertility Continues To Create COVID-19 Vaccine Hesitancy. Journal Of The American Medical Association, 2022.
[18] Bhardwaz, S. & Kumar, J. An Extensive Comparative Analysis Of Chatbot Technologies Chatgpt, Google BARD And Microsoft Bing. In: 2023 2nd International Conference On Applied Artificial Intelligence And Computing (ICAAIC). 2023.
[19] Lu, Q., Qiu, B., Ding, L., Xie, L. & Tao, D. Error Analysis Prompting Enables Human-Like Translation Evaluation In Large Language Models, Cornell University, 2023.
[20] Mohammad-Rahimi, H., Ourang, S.A., Pourhoseingholi, M.A., Dianat, O., Dummer, P.M.H. & Nosrat, A. Validity And Reliability Of Artificial Intelligence Chatbots As Public Sources Of Information On Endodontics. International Endodontic Journal, 2024.