



## **Evaluating the Potential of ChatGPT in CBCT Reporting: Are we Stepping into the Future?**

Dr. Vaishali Malik<sup>1\*</sup>, Dr. Gaurav Pratap Singh<sup>2</sup>, Dr. Trushna Rahangdale<sup>3</sup>, Dr. Amit Agarwal<sup>4</sup>, Dr. Purna Makhija<sup>5</sup>, Dr. Riya Dangi<sup>6</sup>

1. Assistant Professor, Department of Oral Medicine and radiology, Teerthanker Mahaveer Dental College and Research Center, Moradabad, Uttar Pradesh.
2. Professor and Head, Department of Oral Medicine and radiology, Index Institute of Dental Sciences, Indore, Madhya Pradesh.
3. Assistant Professor, Department of Oral Medicine and radiology, Index Institute of Dental Sciences, Indore, Madhya Pradesh.
4. Professor and Head, Department of Oral and Maxillofacial Surgery, Seema Dental College and Hospital, Rishikesh, Uttarakhand.
5. Consultant dental surgeon, Indore, Madhya Pradesh.
6. Consultant dental surgeon, Mumbai, Maharashtra.

**\*Correspondence to:** Dr. Vaishali Malik, Assistant Professor, Department of Oral Medicine and radiology, Teerthanker Mahaveer Dental College and Research Center, Moradabad, Uttar Pradesh.

### **Copyright**

© 2024 **Dr. Vaishali Malik**. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 04 December 2024

Published: 30 December 2024

DOI: <https://doi.org/10.5281/zenodo.14625593>

## ABSTRACT

**Introduction:** Radiology has undergone a significant transformation, becoming an essential component of modern medicine. Artificial Intelligence (AI), including tools like Chat Generative Pre-trained Transformer (ChatGPT), has demonstrated the ability to revolutionize workflows in dentistry and healthcare. ChatGPT, a large language model (LLM) trained on extensive datasets, can assist in diagnostic processes by generating CBCT reports.

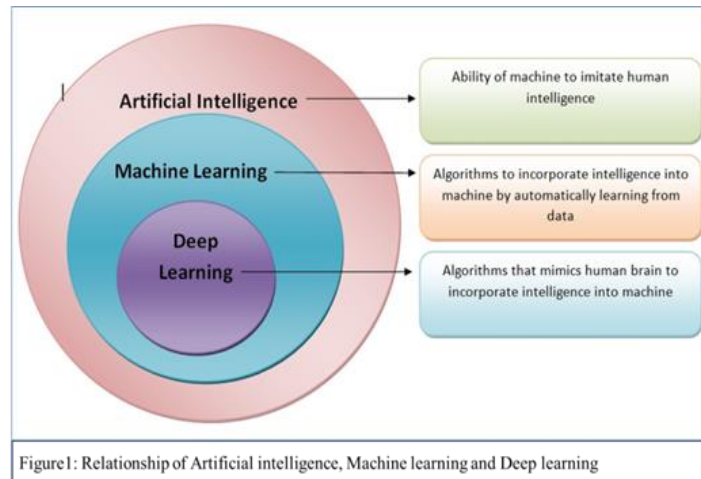
**Materials and Methods:** The study included 50 cases of intraosseous pathologies with CBCT and histopathology as primary investigations. Cases were evaluated by two oral radiologists to generate a primary radiographic diagnosis (PRD1) and differential diagnoses (DD). ChatGPT was then prompted to analyze the same cases, generating a second report (R2) with PRD2 and DD. The performance of ChatGPT was evaluated based on three criteria: comparison of PRD2 with PRD1, accuracy against the histopathologic diagnosis (gold standard), and quality of the generated reports as rated by radiologists on a Likert scale.

**Results:** ChatGPT-generated reports matched the primary radiographic diagnosis of radiologists in 64% of cases (Likert score: 4.24/5). The accuracy of PRD2 in matching histopathologic diagnosis was 62%, with a Likert score of 4.2/5. Regarding quality, 72% of ChatGPT reports were rated excellent (Likert score: 4.5/5). However, oral radiologists' accuracy in diagnosing pathologies was higher (88%) compared to ChatGPT (62%). ChatGPT performed better in cases where minimal clinical information was required but struggled in cases demanding detailed history or clinical context.

**Conclusion:** Artificial intelligence, including ChatGPT, has shown promise as an assistant to oral radiologists in generating CBCT reports. While ChatGPT achieved satisfactory results in diagnostic accuracy and report quality, its performance in clinical scenarios requiring complex reasoning remains limited. This study highlights the potential of AI tools in radiology while emphasizing the need for further improvements to match the expertise of experienced radiologists.

## Introduction

Radiology has undergone a transformative journey since its inception, making a significant impact on modern medicine.<sup>1</sup>



Artificial Intelligence (AI) is an advancing field of computer science focused on developing machines capable of performing tasks that typically require human intelligence. AI encompasses various methodologies, including machine learning (ML), deep learning (DL), and natural language processing (NLP) (Figure 1). Large Language Models (LLMs) represent a subset of AI algorithms that utilize deep learning techniques and extensive datasets to understand, summarize, generate, and predict text-based content.<sup>2</sup> AI is used to reduce the working load on the humans with its precise working capabilities.<sup>3</sup>

Chat Generative Pre-trained Transformer (ChatGPT) is an AI-powered program designed to generate human-like responses to user prompts. It has been trained on vast amounts of data to enhance computational linguistics, communication capabilities, and responsiveness. ChatGPT, in its GPT-3.5 version, features an impressive 175 billion parameters, making it significantly more powerful than earlier versions.<sup>4</sup> The use of ChatGPT has elicited a debate about the advantages and disadvantages of AI technologies in routine clinical practice, including concerns about potential biases in the ChatGPT training datasets that could constrain its use<sup>5</sup>

Dento-maxillofacial imaging is an integral part of clinical dentistry and a main diagnostic aid to diagnose maxillofacial diseases<sup>6</sup>. The introduction of cone beam computed tomography (CBCT) devices, changed the way oral and maxillofacial radiology is practiced. CBCT was embraced into the dental settings very rapidly

due to its compact size, low cost, low ionizing radiation exposure when compared to medical computed tomography<sup>6</sup>

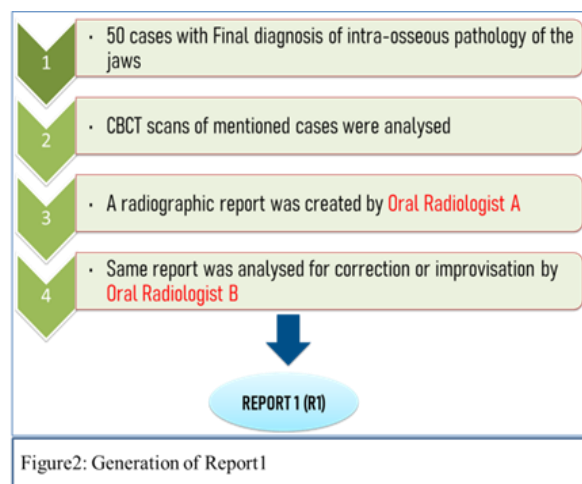
In dentistry and healthcare, ChatGPT offers diverse services for healthcare professionals, such as enhancing diagnostic accuracy, supporting decision-making, recording digital data, analyzing images, preventing and treating diseases, reducing treatment errors, and facilitating research. Specifically, in radiology, ChatGPT holds the potential to revolutionize workflows by reducing radiologists' workload.<sup>7</sup>

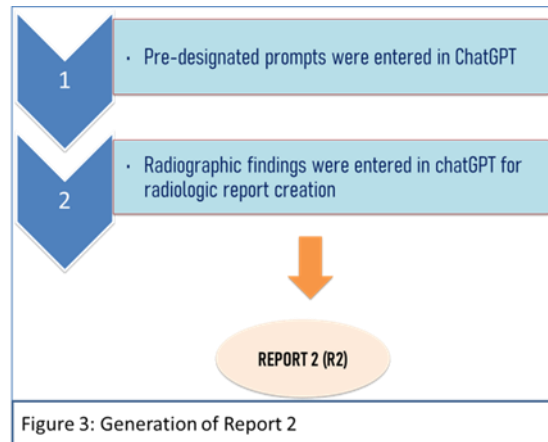
This study aims to assess the potential of chatGPT in creating CBCT reports. We anticipate that this research will provide valuable insights into the capabilities and limitations of AI language models in generating radiologic reports and decision making to diagnose the maxillofacial pathologies.

## Materials and Methods

Cases of intra-osseous pathologies, that have CBCT and histopathology as primary investigations, were accessed from the database of Oral radiology department. 50 cases with no or insignificant artifacts were selected to include in the study.

Every case was analyzed by 2 radiologists; by first one (Radiologist A) for initial report writing and by second one (Radiologist B) for improvisation, if required. Final report that was prepared by oral radiologists, that was noted as Report 1 (R1), had one primary radiographic diagnosis (PRD1) and at least 5 Radiographic differential diagnoses (DD). (Figure 2)





Same cases were analyzed by chatGPT to create a CBCT report by first using following prompt (Figure 8):

“\*You are an experienced Dento-maxillo-facial Radiologist.

\*Writing down chain of thoughts in every thinking step.

\*Generate a CBCT radiological report only containing the following sections:

a) Radiographic Analysis

b) Primary Radiographic diagnosis

c) List of atleast 5 differential diagnosis (without including primary radiographic diagnosis) (in order of probability of occurrence)

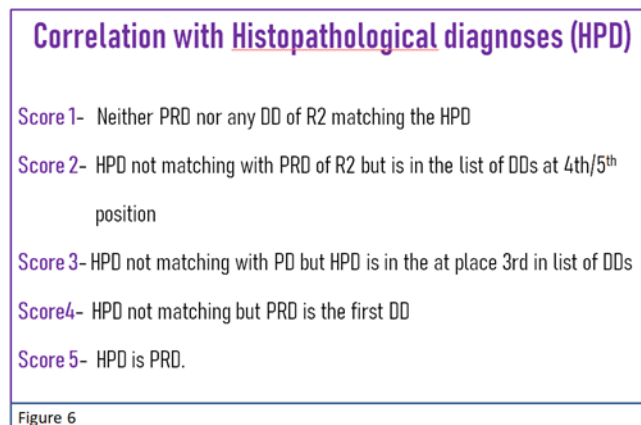
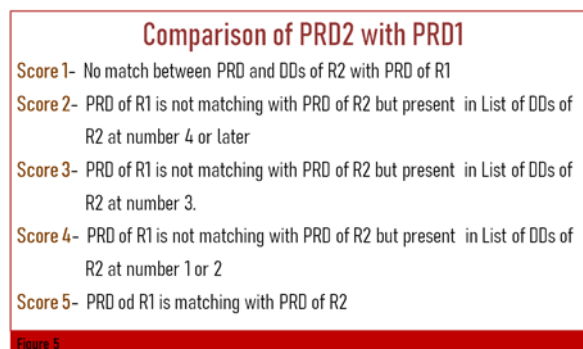
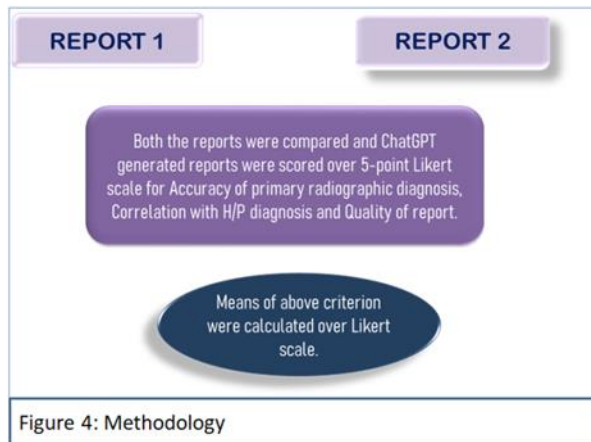
d) Suggestion about further diagnostic and therapeutic intervention.

\*Only say YES if you understand my requirement.

After having a response from chatGPT, radiological findings of the CBCT scans were entered in the chatGPT(Figure 9). Following which a CBCT report was generated by chatGPT that was noted as Report 2 (R2). R2 also has one Primary radiographic diagnosis (PRD2) and at least 5 radiographic differential diagnoses. (Figure 10)

Following this R2 were assessed on following criteria:

1. Comparison of PRD2 with PRD1
2. Accuracy of PRD2 taking histopathologic diagnosis(HPD) as gold standard.
3. Quality of chatGPT generated report (R2)



Quality of report	
Score 1-	Very poor reporting
Score 2-	Poor reporting
Score 3-	Average reporting
Score 4-	Good reporting
Score 5-	Excellent reporting

Figure 7

ChatGPT ▾

You are an experienced Dento-maxillo-facial Radiologists Writing down chain of thoughts in every thinking steps Generate a CBCT radiological report only containing the following sections: a) Radiographic Analysis b) Primary Radiographic diagnosis c) List of atleast 5 differential diagnosis (without including primary radiographic diagnosis) (in order of probability of occurrence) d) suggestion about further diagnostic and therapeutic intervention. 4. Only say YES if you understand my requirement.

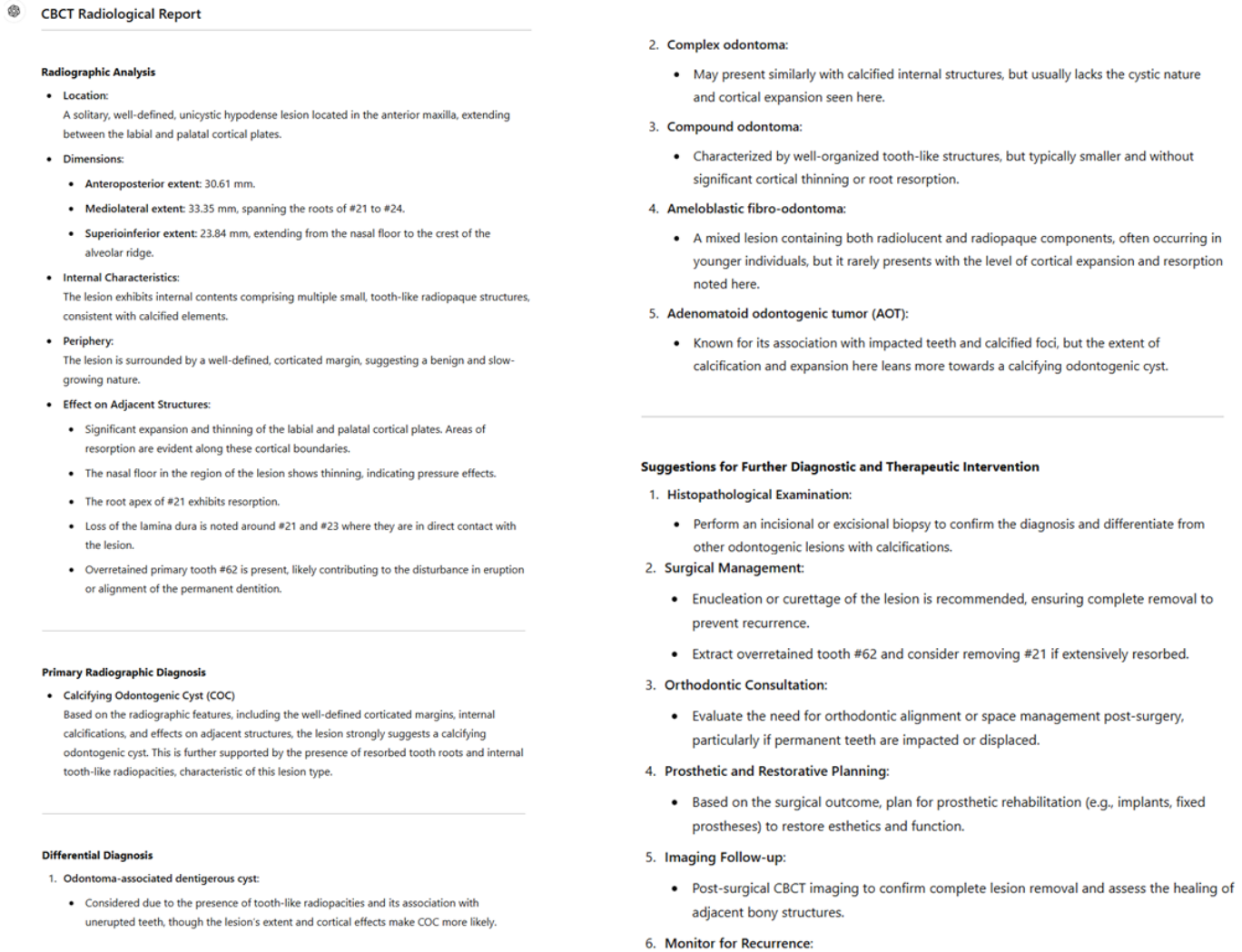
 YES

**Figure 8: First prompt entered in ChatGPT**

 YES

A 17 year old male patient solitary unicystic hypodense area present in anterior maxilla extending anteroposterior 30.61mm between labial and palatal cortical plates, mediolaterally 33.35mm between roots of #21 to #24 superioinferioly 23.84mm between nasal floor and crest of alveolar ridge. Internally lesion contains multiple small tooth like radiopacities. periphery of lesion are well defined and corticated. Labial and palatal cortical plates and expanded and resorbed. Nasal floor is also thinned out in the region of Ision, lesion, root apex of #21 is resorbed while there is loss of lamina dura irt #21 #23 at point of contact with the lesion. There is also overretained #62

**Figure 9: Radiological findings entered in ChatGPT**



**Figure 10: ChatGPT generated report**

Above 3 criteria were rated over Likert scale from 1 to 5 on the basis of inference given in Figure 5, 6 and 7. For criteria 3, R2 of all 50 cases were provided to 3 radiologists randomly to grade their quality from 0 to 5 (Figure 4). Accuracy of PRD1 and PRD2 was also compared.

## Results

A Total of 50 cases were taken in to study without categorization of age and gender. Distribution of pathologies among sample is shown in Table 1.



S/N	Final Diagnosis	Number
1.	Radicular Cyst	13
2.	OKC	6
3.	Dentigerous cyst	4
4.	Ameloblastoma	6
5.	COC	1
6.	Metastatic Carcinoma	1
7.	CGCG	2
8.	Odontome	6
9.	Lateral Periodontal Cyst	1
10.	Ossifying fibroma	8
11.	Nasopalatine cyst	1
12.	Residual cyst	1
	Total	50

Table 1: Distribution of pathologies in the study

When R1 and R2 were compared, it was found that in 64% (n=32) cases primary radiographic diagnosis of R1 and R2 were same and they were rated 5/5 over Likert scale while in 4% cases (n=2), Primary Radiographic diagnosis provided by Oral radiologist was not mentioned by chatGPT even in list of DDs and they were scored 1/5 over Likert scale. Overall, for comparison of PRD2 with PRD1, ChatGPT generated report was rated 4.24/5 on Likert scale.

When accuracy of chatGPT generated report was analyzed it was found that in 62% (n =31) chatGPT has provided primary radiographic diagnosis that was similar to histopathologic diagnosis that was considered as gold standard for final diagnosis and such cases were rated 5/5 on Likert scale. 6% (n=3) cases were rated 1/5 as histopathologic diagnosis of those cases were not even present the list of DDs provided by chatGPT. Overall accuracy of chatGPT generated reports overated 4.2/5 over Likert scale.

For assestment of quality of chatGPT generated reports all the 50 reports were distributed randomly among 5 oral radiologists of more or less similar clinical experience and oral radiologists were asked to rate quality of text provided in the report over Likert scale from 1 to 5. 72% (n=36) chatGPT generated reports were graded as excellent (5/5) by oral radiologists while only 2% (n=1) report was graded as very poor (1/5). Overall rating of chatGPT generated reports were found to be 4.5/5 over Likert scale.

Likert Scale/Criteria	Comparison of PRD2 with PRD1	Accuracy of PRD2 with Histopathologic diagnosis	Quality of ChatGPT generated report
Score 5	32	31	36
Score 4	7	7	7
Score 3	4	6	3
Score 2	5	3	3
Score 1	2	3	1
Average score	4.24	4.20	4.50

**Table 2: Rating of ChatGPT generated reports over Likert scale**

Data related to above 3 criteria have been summarized in table 2.

Overall accuracy of oral radiologists generated CBCT reports were found to be 88% while it was 62% for chatGPT generated reports.

## Discussion

In the recent years, artificially intelligence has come up with an adjunct to maxillofacial radiologist as an assistant and as a decision maker as well. In our study we found that chatGPT can actually think like an oral radiologists while preparing CBCT reports and also can be an average decision maker when comes to diagnosing the correct condition.

When oral radiologists generated reports and chatGPT generated reports were compared, it was found that both of them had produced similar results, as far as primary radiographic is concerned, in 64% of the cases. A study conducted by Yanni Hu et.al.<sup>7</sup> concluded that in 48.1% of the cases first diagnosis (PRD1 in our study) was in alignment in pathologic diagnosis (PRD2 in our study)

In our study 72% of chatGPT generated reports were considered as excellent on the basis of tax policy and presentation. In the study conducted by Yanni Hu et.al.<sup>7</sup> , 88.7% of chatGPT generated reports were considered as error free.

ChatGPT generated reports had accurate primary radiographic diagnosis (PRD2) matching with the histopathologic diagnosis in 62% of the cases while large number (38%) were not in alignment with histopathologic diagnosis. High accuracy of chatGPT to provide diagnosis in accordance with histopathology was observed in cases which usually need no or minimum history/clinical information for diagnosing intraosseous pathologies but it failed to provide accurate PRD in cases where history/clinical information was an important component to reach accurate diagnosis such as in case of Residual cyst.

Comparing accuracy of oral radiologists generated reports and chatGPT generated reports, oral radiologists had higher accuracy (88%) in diagnosing pathology correctly than that of chatGPT (62%).

## Conclusion

In the changing world, artificial intelligence has become an important and somewhat unavoidable part of it, including its role in health care. In multiple studies including our study chatGPT has shown promising results to be an excellent assistant to maxillofacial radiologist in generating radiologic/CBCT reports. Our study concluded the chatGPT can be a good decision maker in diagnosing pathologies radiographically but has high scope for improvement to actually think like an oral and maxillofacial radiologist.

## References

1. Najjar R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics*. 2023; 13(17):2760.
2. Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 23, 689 (2023).
3. Dhiman S, Sharma S, Mehta R. (2024) Artificial Intelligence in Oral Medicine and Radiology. *J Oral Med and Dent Res*. 5(1):1-6.
4. Alhaidry HM, Fatani B, Alrayes JO, Almanan AM, Alkhaed NK. ChatGPT in Dentistry: A Comprehensive Review. *Cureus*. 2023;15(4):e38317. Published 2023 Apr 30. doi:10.7759/cureus.38317

- 
5. Pedram Keshavarz et al; ChatGPT in radiology: A systematic review of performance, pitfalls, and future perspectives. *Diagnostic and Interventional Imaging*, Volume 105, Issues 7–8, Pages 251-265.
  6. Karpagavalli S, Malik V, Singh GP, Akhtar N, Mehta R; Patient Perception Of Dental Radiographs Its Hazards And Safety Protection: A Questionnaire Based Cross -Sectional Study June 2023 *European Chemical Bulletin* 12(Special issue):3108-3120
  7. Venkatesh E, Elluru SV. Cone beam computed tomography: basics and applications in dentistry. *J Istanbul Univ Fac Dent* 2017;51(3 Suppl 1):S102-S121.
  8. Hu, Y., Hu, Z., Liu, W. et al. Exploring the potential of ChatGPT as an adjunct for generating diagnosis based on chief complaint and cone beam CT radiologic findings. *BMC Med Inform Decis Mak* 24, 55 (2024).

