



Evaluating the Use of Large Language Models in Radiology and Histopathology Reporting: Expert-Based Assessment of Diagnostic Support and Patient-Oriented Simplification

Sümeyye Çelik^{1,a,*}, Alican Kuran^{1,b}, Oğuz Baysal^{1,c}, Umut Seki^{1,d}, Merve Soluk Tekkesin^{2,e}, Enver Alper Sinanoğlu^{1,f}

¹Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli University, Kocaeli, Türkiye.

²Department of Oral Pathology, Faculty of Dentistry, İstanbul University, İstanbul, Türkiye.

*Corresponding author

Research Article

History

Received: 19/11/2024

Accepted: 27/04/2025

ABSTRACT

Objectives: The aim of this study was to evaluate the effectiveness of two different versions of Chat-GPT, one of the large language models (LLMs), in the diagnosis and interpretation of cone beam computed tomography (CBCT) and histopathology reports.

Materials and Methods: In this study, Chat-GPT 3.5 and Chat-GPT 4 were tasked with generating preliminary diagnoses and differential diagnoses based on the findings from ten CBCT reports and ten histopathology reports. Additionally, both versions were asked to simplify these reports to a level understandable by patients. Dentomaxillofacial radiologists and pathologists, with varying levels of expertise, evaluated the responses of the LLMs and the performance of Chat-GPT 3.5 and Chat-GPT 4 in these tasks was subsequently compared based on these expert assessments.

Results: A comparison of diagnostic performance for radiology reports showed that Chat-GPT 4 was statistically superior to Chat-GPT 3.5 ($p < 0.001$), while no significant difference was observed between the two models in terms of report simplification scores ($P > 0.05$). In contrast, when evaluating histopathology reports, Chat-GPT 4 performed significantly better than Chat-GPT 3.5 in terms of both diagnostic accuracy and report simplification ($p < 0.05$).

Conclusions: The results demonstrated that Chat-GPT 4 achieved superior performance in the interpretation and evaluation of CBCT reports by LLMs. The strong performance of this latest version highlights the potential for LLMs to become valuable tools in the reporting processes of radiology and histopathology, as well as in numerous other fields, as advancements in technology continue to improve their capabilities.

Keywords: Large language models, Cone-Beam Computed Tomography, ChatGPT

Radyoloji ve Histopatoloji Raporlamasında Büyük Dil Modellerinin Performans Analizi: Tanı Desteğinden Hasta İletişimine

Araştırma Makalesi

Süreç

Geliş: 19/11/2024

Kabul: 27/04/2025

ÖZET

Amaç: Bu çalışmanın amacı, Chat-GPT'nin iki farklı sürümünün, büyük dil modellerinden biri olarak, konik ışınli bilgisayarlı tomografi (KİBT) ve histopatoloji raporlarının tanı ve yorumlanmasındaki etkinliğini değerlendirmektir.

Gereç ve Yöntemler: Bu çalışmada, Chat-GPT 3.5 ve Chat-GPT 4, on KİBT raporu ve on histopatoloji raporuna dayanarak ön tanı ve ayırıcı tanı üretmekle görevlendirilmiştir. Ek olarak, her iki sürümden de bu raporları hastaların anlayabileceği bir düzeye basitleştirmeleri istenmiştir. Farklı uzmanlık seviyelerine sahip ağız diş ve çene radyolojisi uzmanları ve oral patoloji uzmanları, Chat-GPT'nin yanıtlarını değerlendirmiş ve Chat-GPT 3.5 ile Chat-GPT 4'ün bu görevlerdeki performansı bu uzman değerlendirmelerine göre karşılaştırılmıştır.

Bulgular: Radyoloji raporlarına ilişkin tanısal performans karşılaştırması, Chat-GPT 4'ün Chat-GPT 3.5'e kıyasla istatistiksel olarak anlamlı şekilde üstün olduğunu göstermiştir ($p < 0,001$). Ancak, rapor basitleştirme puanları açısından iki model arasında anlamlı bir fark gözlemlenmemiştir ($P > 0,05$). Öte yandan, histopatoloji raporlarının değerlendirilmesinde, Chat-GPT 4 hem tanısal doğruluk hem de rapor basitleştirme açısından Chat-GPT 3.5'ten anlamlı derecede daha iyi performans göstermiştir ($p < 0,05$).

Sonuç: Sonuçlar, Chat-GPT 4'ün CBCT raporlarının yorumlanmasında ve değerlendirilmesinde eski versiyona kıyasla üstün performans gösterdiğini ortaya koymuştur. Bu son sürümün güçlü performansı, büyük dil modellerinin radyoloji ve histopatolojideki raporlama süreçlerinde ve teknolojik ilerlemelerle yeteneklerinin artmaya devam ettiği birçok başka alanda değerli araçlar haline gelebileceğini göstermektedir.

Anahtar Kelimeler: Büyük dil modelleri, Konik Işınli Bilgisayarlı Tomografi, ChatGPT.

Copyright



This work is licensed under
Creative Commons Attribution 4.0
International License

^a smyyelicik41@gmail.com

^c oguz_baysal@outlook.com

^e msoluk@istanbul.edu.tr

^b 0009-0003-0676-5098

^d 0009-0000-0360-0050

^f 0000-0002-7178-3335

^b alicankuran_15@hotmail.com

^d dtumut@outlook.com

^f alpersinanoglu@yahoo.com

^b 0000-0001-9677-8690

^d 0000-0002-0286-9792

^f 0000-0002-8349-3239

How to Cite: Çelik S, Kuran A, Baysal O, Seki U, Tekkesin MS, Sinanoğlu EA.(2025) Evaluating the Use of Large Language Models in Radiology and Histopathology Reporting: Expert-Based Assessment of Diagnostic Support and Patient-Oriented Simplification. Cumhuriyet Dental Journal, 28(2): 141-156.

Introduction

Artificial intelligence (AI) systems have the ability to interpret situations, events and objects and respond to them in a way that mimics human behaviour.¹ The recent developments in artificial intelligence technology have resulted in the introduction and extensive utilization of large language model-based (LLM) applications, which are readily accessible to the public. LLMs are artificial intelligence models that have been developed with the objective of understanding speech, responding accordingly and fulfilling the desired tasks. These models require large amounts of text data for training and can recognize the nuances of language using deep neural networks, through which they are also able to generate responses that mimic human interaction.²

In 2018, OpenAI (San Francisco, CA, USA) initiated the development of one of these LLM systems. The company released GPT, a generative pre-training transformer with a neural network comprising 1.5 billion parameters, trained on a dataset comprising 40 gigabytes of text. In 2022, an AI-based LLM, designated 'ChatGPT', was made available which had the ability to generate responses to text input comparable to human beings in terms of quality and sophistication. In 2023, a new version of the LLM, designated ChatGPT 4, was produced, which exhibited significantly enhanced capabilities in comparison to the previous version.³

Radiology and histopathology reports contribute significantly to the decision-making process of clinicians by providing a comprehensive evaluation of different diseases.⁴ However, the style and rules regarding the composition of reports may vary between institutions and according to the training received by the individual responsible for the report. This situation may cause confusion for patients who receive different explanations from multiple institutions, as well as limiting communication between physicians.⁵ Therefore, an effective and standardized approach to the formatting of radiological and histopathological reports is required.⁶⁻⁹

While artificial intelligence is currently employed primarily for image analysis in radiology and pathology studies, ChatGPT has yet to be fully investigated for its potential in reporting.^{10,11} In accordance with the radiologist's/pathologist's descriptions of findings in the report, ChatGPT has the capacity to perform preliminary diagnoses, differential diagnoses, and to elucidate the rationale behind its decisions. Furthermore, it has been proposed that ChatGPT could be employed to reduce the complexity of these reports, thereby facilitating patient comprehension.³ In recognition of this potential, the objective of the present study is to assess the capability of ChatGPT 3.5 and 4 to conduct preliminary and differential diagnoses based on data derived from radiology and histopathology reports, as well as their capacity to elucidate these reports to the patient.

Material and Methods

As this study was conducted in February 2024, the versions of Chat GPT that were active at that time were used. Approval for our study (project number 2024/112) was obtained from the Ethics Committee for Non-Interventional Clinical Research at Kocaeli University.

Study Design

In this retrospective study, radiology and histopathology reports written between 2020 and 2023, with the preliminary diagnosis sections deleted were uploaded to ChatGPT versions 3.5 and 4 in Turkish without translation into English and answers were requested in Turkish. The LLM was tasked with identifying a preliminary diagnosis and several differential diagnoses based on the findings presented in the reports. Subsequently, it was instructed to convey the report and diagnosis in a manner that would be readily comprehensible to the patient. The relative success of each version was evaluated by a group of experts in the field of dentomaxillofacial radiology and pathology, who were not aware of which version produced the answer. The workflow of the study is illustrated in Figure 1.

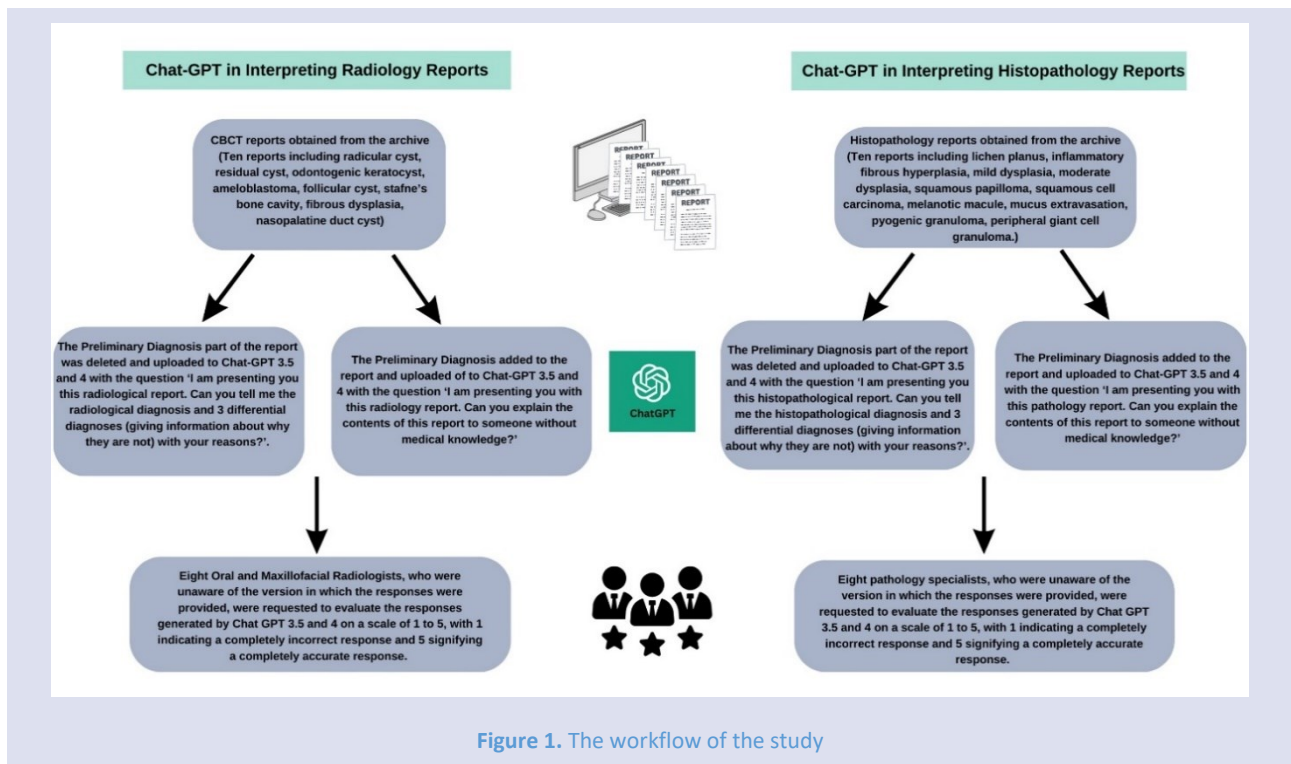


Figure 1. The workflow of the study

Radiographic Reports Data Set

Tomography reports written between 2020 and 2023 in the CBCT archive of the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli University, were retrospectively analysed. Reports were included only if they contained detailed information regarding the boundaries, shape, internal structure, and the lesion's relationship with adjacent anatomical structures. Reports that lacked any of these essential elements were excluded. To ensure diagnostic diversity and avoid overrepresentation, no more than two reports were included for each diagnosis. The reports included in the study consisted of 10 CBCT reports of radicular cyst, residual cyst, odontogenic keratocyst, ameloblastoma, follicular cyst, staphne bone cavity, fibrous dysplasia, nasopalatine duct cyst lesions.

Histopathology Reports Data Set

Histopathological reports written between 2020 and 2023 from İstanbul University, Faculty of Dentistry, Department of Oral Pathology were examined. Reports were selected based on the presence of comprehensive diagnostic information, including the type and structure of the affected tissue, cellular morphology, and the presence of inflammatory or vascular changes. Reports lacking

sufficient detail in these domains were excluded from the study. To maintain diagnostic variety and prevent bias, no more than two reports were included for each diagnosis. The histopathological reports included in the study consisted of 10 reports including lichen planus, inflammatory fibrous hyperplasia, mild dysplasia, moderate dysplasia, squamous papilloma, squamous cell carcinoma, melanotic macula, mucus extravasation, pyogenic granuloma, peripheral giant cell granuloma.

Uploading Reports to Chat-GPT

In the initial phase of report processing, the preliminary diagnosis sections, along with any personal information, were removed from the reports. Subsequently, Chat-GPT versions 3.5 and 4 were queried with the following question: "Could you please provide a preliminary diagnosis and differential diagnosis for the lesion described in this report, including your rationale?" The responses from both versions of the language model were documented for comparison (Table 1,2).

Table 1. Chat-GPT's responses to radiology reports

Radiological Preliminary Diagnosis	Radiology Report	Diagnosis		Report Simplification	
Radicular Cyst		ChatGPT 3.5 Preliminary Diagnosis:	ChatGPT 4 Preliminary Diagnosis:	ChatGPT 3.5	ChatGPT 4
	As a result of the evaluation, there is a cortical limited, unilocular, radiolucent lesion associated with the	Radicular Cyst Differential Diagnosis:	Radicular Cyst Differential Diagnosis:	The report contains the results of a detailed scan of your teeth and jaw. A significant issue has been found in one of your teeth, specifically tooth number 47,	This report explains some issues related to a lower right back tooth (tooth number 47). The top

	<p>roots of tooth 47 with excessive crown destruction. The lesion caused minimal expansion of the cortical bones but caused perforation of the cortical bone on the lingual side. No resorption was observed in the roots of the related tooth. In addition, the lesion is in close relationship with the right mandibular canal from the lingual side.</p>	<p>Dentigerous Cyst Odontogenic Keratocyst Ameloblastoma</p>	<p>Ameloblastoma Odontogenic Keratocyst Odontogenic Myxoma</p>	<p>with a type of cystic lesion associated with the roots. This cyst has caused the formation of hard bone surrounding the lesion within the bone structure. Additionally, the cyst has damaged the bones inside the mandible, but there has been no expansion beyond the bone. There is no damage to the roots of your tooth. Your doctor believes this cyst may be a type called a radicular cyst. Your doctor will provide you with appropriate treatment options for this condition. If you require further information or explanation, feel free to consult your doctor.</p>	<p>part of your tooth is severely damaged, and a small cyst (cavity) has formed at the tip of the roots in the jawbone. This cyst has caused a hole in the jawbone, but the bone itself has not expanded. The cyst may have originated from an infection starting in the tooth's root, and it is located near the main nerve canal in your jawbone. For a definitive diagnosis, a sample from the cyst will need to be examined in more detail.</p>
Radicular Cyst	<p>The evaluation revealed a 12x15x11 mm unilocular, cortically defined, circular radiolucent lesion associated with the roots of teeth numbered 11-12, which have undergone root canal treatment. The lesion has not caused any resorption in the affected teeth. However, it has perforated the buccal cortex and is near the incisive canal.</p>	<p>Preliminary Diagnosis: Periapical Radicular Cyst Differential Diagnosis: Periapical Granuloma Ameloblastoma Periapical Abscess</p>	<p>Preliminary Diagnosis: Radicular Cyst Differential Diagnosis: Dentigerous Cyst Odontogenic Keratocyst Ameloblastoma</p>	<p>A lesion related to the roots was detected after a previous root canal treatment on a tooth. This lesion has created a hole in the bone and is associated with the nerve canal of the front teeth. Doctors believe that this lesion is a type of cyst. The cyst is a fluid-filled sac that forms in the tissues around the tooth roots. The size and characteristics of the lesion support the likelihood of it being a cyst.</p>	<p>A cyst measuring approximately 12x15x11 mm has been detected next to the roots of two teeth in the upper jaw. These teeth have previously undergone root canal treatment. The cyst has penetrated one of the jaw bones and is in contact with a nerve canal. This type of cyst has been identified as a 'radicular cyst,' which typically results from infections in the tooth root. This is a condition concerning dental or jaw health, and it is necessary to consult a dentist for appropriate treatment.</p>
Stafne Bone Cavity	<p>As a result of the evaluation, it has been determined that the bone cavity observed in the left mandibular angle region is located inferior to the mandibular canal. The bone cavity has a concave shape, starting from the mandibular lingula. The boundaries of the bone cavity are surrounded by</p>	<p>Preliminary Diagnosis: Mandibular Osteomyelitis Differential Diagnoses: Mandibular Osteoma, Ameloblastoma, Cystic Lymphoma</p>	<p>Preliminary Diagnosis: Stafne Bone Cavity Differential Diagnoses: Ameloblastoma Odontogenic Cysts Existential Variations of the Mandibular Canal</p>	<p>According to the radiological report, a bone cavity has been detected in the left corner of the jaw. This cavity is located on the inner part of the jaw, beneath the teeth, and is surrounded by bone. The condition is referred to as a Stafne bone cavity. It is generally harmless and does not require treatment. However, it is important to follow your doctor's recommendations.</p>	<p>Of course, this report indicates that there is a small, stable cavity on the left side of your jaw, within the jawbone, which is present without causing harm. This cavity likely points to a Stafne Bone Cavity, a natural condition that is usually discovered incidentally and generally does not require treatment. It</p>

	cortical bone, and the integrity of the bone has been preserved				is most often a benign condition and has no negative impact on your health
Odontogenic Keratocyst	The evaluation revealed a unilocular, radiolucent lesion with sclerotic borders, measuring 14x26x22 mm, associated with the impacted tooth number 38, extending from the left mandibular molar region toward the ramus. The lesion has perforated the mandibular base and the medial surface of the ramus. The mandibular canal is within the lesion. The lesion has caused expansion in the buccal and lingual cortices, with sclerotic bone formation present on the buccal side.	Preliminary Diagnosis: Ameloblastoma Differential Diagnosis: Odontogenic Keratocyst Dentigerous Cyst Bone Cyst	Preliminary Diagnosis: Odontogenic Keratocyst Differential Diagnosis: Ameloblastoma Dentigerous Cyst Odontogenic myxoma	This report indicates that a lesion called an odontogenic keratocyst has been detected in the lower left jaw area, at the back of the lower jaw. This lesion has affected the surrounding tissues, causing expansion and hardening of the jawbones. Treatment will likely require surgical intervention.	A cyst about the size of a hazelnut has been found under your jawbone, next to an unerupted tooth. This cyst has caused some changes in the jawbone and is a condition that requires treatment. Doctors believe it is a specific type of cyst related to the tooth's development.
Fibrous Dysplasia	The evaluation revealed a radiopaque lesion, starting from the region of tooth number 48, involving the entire ramus, and extending towards the right condyle, giving a ground-glass appearance in the trabecular bone. The lesion has caused expansion in the condyle and ramus, resulting in asymmetry. Additionally, there is severe resorption on the superior surface of the left condyle.	Preliminary Diagnosis: Temporomandibular Joint (TMJ) tumor or lesion Differential Diagnoses: Ameloblastoma Fibrous Dysplasia Radiolucent Lesions (cystic lesions)	Preliminary Diagnosis: Fibrous Dysplasia Differential Diagnoses: Osteosarcoma, Paget's Disease, Osteomyelitis	According to this report, a ground-glass appearance lesion has been detected in the right jaw joint (condyle) region, starting from the area of tooth number 48. This lesion has caused growth in the jaw joint and jawbone, leading to facial asymmetry. Additionally, significant bone loss (resorption) has been identified on the left jaw joint. Based on the evaluation, the lesion is suspected to be a condition called fibrous dysplasia. Fibrous dysplasia is a condition that causes abnormal growth of bone tissue and often requires surgical treatment.	The report indicates that there is abnormal tissue growth (lesion) on the right side of the jaw. This growth has caused asymmetry (unevenness) and deformation in the jaw joint and jawbone. Additionally, bone loss has been detected on the left jaw joint. Doctors believe that this abnormal growth and changes are due to a condition called 'fibrous dysplasia.' Fibrous dysplasia is a rare disease where the normal structure of bones is disrupted, potentially affecting their function.
Ameloblastoma	The evaluation revealed a radiolucent lesion measuring 24x38x26 mm, extending from the left mandibular premolar region to	Preliminary Diagnosis: Ameloblastoma Differential Diagnoses: Ameloblastoma, Odontogenic Cyst, Osteosarcoma	Preliminary Diagnosis: Ameloblastoma Differential Diagnoses: Odontogenic Keratocyst,	Of course, the report mentions a lesion in the left lower jaw area that is growing and contains air spaces and flat septa. This lesion has caused expansion in the jawbone and affected the surrounding tissues. Additionally, it has damaged the roots of a	Doctors have found a large mass in your lower jaw, starting from the area where some of your teeth are located and extending toward the back of

the right mandibular angle. The lesion contains areas of air spaces and thin, flat septa. The lesion has caused severe expansion in the buccal and lingual cortices and has destroyed the apex of the alveolar ridge. There is a radiopaque band around the lesion, indicating a chronic process. The lesion has caused root resorption in tooth number 45. It has displaced the mandibular canal cortex buccally and superiorly. Within the lesion, tooth number 48 is located in an inverted position toward the mandibular ramus. The obtained images indicate that the observed lesion is of a chronic nature, but its severe expansion should be taken into account. Magnetic resonance imaging and histopathological examination are recommended for assessing soft tissue involvement and the contents of the lesion

Odontogenic Myxoma,
Central Giant Cell Granuloma

tooth and displaced the jaw canal. Imaging and biopsy are recommended for better evaluation of the lesion's contents. Doctors believe that this lesion could be a tumor called ameloblastoma.

the jaw. This mass has disrupted the normal structure of your jawbone and damaged some of your teeth. Additionally, this mass has caused swelling by expanding your jawbone. This condition is a type of tumor called 'ameloblastoma,' which is usually benign. This slow-growing tumor occupies space in the jawbone and can damage surrounding tissues if left untreated. Doctors have recommended additional tests to gather more information and determine the appropriate treatment. This is an important issue for your jaw health, and it is crucial to follow your doctor's advice and pursue treatment

Infected follicular cyst

The evaluation revealed that tooth number 38 is impacted in an oblique position and that there is a unilocular, radiolucent lesion around its crown. The lesion measures 25x38x29 mm. The lesion, which is in close proximity to tooth number 37, has not caused resorption in the roots. The mandibular canal passes through the

Preliminary Diagnosis: Dentigerous Cyst
Differential Diagnoses: Follicular Cyst, Ameloblastoma, Osteomyelitis

Preliminary Diagnosis: Ameloblastoma
Differential Diagnoses: Odontogenic Keratocyst, Dentigerous Cyst, Central Giant Cell Granuloma

The evaluation revealed that tooth number 38 is impacted in an oblique position and that there is a unilocular, radiolucent lesion around its crown. The lesion measures 25x38x29 mm. The lesion, which is in close proximity to tooth number 37, has not caused resorption in the roots. The mandibular canal passes through the lesion. The lesion has caused expansion in the mandible and has perforated the buccal and lingual cortices. Sclerotic structures have been identified in the bone surrounding the lesion. Based on the obtained information, it is thought that the relevant lesion is an infected follicular cyst

The report indicates that there is a misaligned tooth trapped in the jawbone beneath your mouth, which should normally have erupted, and a large cyst has formed around this tooth. This cyst has damaged the jawbone by causing expansion and has perforated both the outer surfaces of the jawbone and is close to an important nerve that runs through the

	<p>lesion. The lesion has caused expansion in the mandible and has perforated the buccal and lingual cortices. Sclerotic structures have been identified in the bone surrounding the lesion.</p>				<p>jawbone. The bone surrounding the cyst has hardened slightly. This indicates a problem in the jawbone that requires treatment.</p>
Nasopalatine Canal Cyst	<p>The evaluation revealed a unilocular, radiolucent lesion measuring 9x9x12 mm with cortical borders in the anterior maxillary region along the midline. The lesion is not associated with teeth numbers 11 and 21. The lesion shows continuity with the incisive canal and has caused expansion in the incisive canal. Additionally, the lesion has caused perforation in the palatal cortex</p>	<p>Preliminary Diagnosis: Odontogenic Keratocyst Differential Diagnoses: Dentigerous Cyst, Periapical Cyst, Ameloblastoma</p>	<p>Preliminary Diagnosis: Nasopalatine Canal Cyst Differential Diagnoses: Radicular Cyst, Odontogenic Keratocyst, Globulomaxillary Cyst</p>	<p>A small fluid-filled sac has been detected in the anterior upper jaw, just above the midline. This sac is not related to the teeth and is located near a structure called the nasopalatine canal. The sac has caused expansion within the canal and created a perforation in the inner part of the upper jawbone. Based on these findings, it is thought to be a nasopalatine canal cyst.</p>	<p>There is an approximately 1-centimeter hollow structure (cyst) in the upper jaw, not related to the teeth. This cyst has enlarged a canal in the front part of the jaw and has caused a perforation in the inner part of the jawbone. This condition is usually not serious but may sometimes require treatment</p>
Residual Cyst	<p>The evaluation revealed a round, unilocular, radiolucent lesion measuring 18x19x18 mm in the left maxillary molar region where a tooth extraction was previously performed. The lesion has extended into the left maxillary sinus but has not perforated the sinus. Additionally, the lesion has a cortical boundary. It has caused severe expansion with perforation in the buccal cortex</p>	<p>Preliminary Diagnosis: Fungal Sinusitis Differential Diagnoses: Dentigerous Cyst, Ameloblastoma, Odontogenic Keratocyst</p>	<p>Preliminary Diagnosis: Residual Cyst Differential Diagnoses: Dentigerous Cyst, Ameloblastoma, Odontogenic Keratocyst</p>	<p>According to the radiological report, a cyst measuring 18x19x18 mm was detected in the area of the extracted tooth in the left upper back tooth region of the jawbone. The cyst has grown towards the maxillary sinus but has not perforated its membrane. It has a cortical boundary and has caused significant expansion in the buccal cortex. It is thought to be a residual cyst, likely originating from the extracted tooth.</p>	<p>A walnut-sized cyst (cavity) has been found in the upper part of your mouth, on the left rear side, where a tooth was previously extracted. This cyst has grown towards an air-filled space inside your mouth called the sinus but has not damaged it. The cyst has pushed against the bone on the cheek side, causing expansion in the area. This type of cyst is called a 'residual cyst,' and it usually forms where teeth have been extracted. Your doctor will likely recommend surgery to remove the cyst</p>

In the second phase, the complete, unredacted versions of the reports, including the diagnosis sections, were provided to Chat-GPT versions 3.5 and 4. These versions were then asked the question: "I am presenting this radiology/pathology report to you. Can you explain its

contents in a way that someone without medical knowledge would understand?" The responses from both versions were similarly recorded for analysis (Table 1,2).

Table 2. Chat-GPT's responses to histopathology reports

Pathology result	Pathology Report	Diagnosis		Report Simplification	
Lichen Planus	Sample Area: Buccal Mucosa In the examined sections, the surface is covered by stratified squamous epithelium showing acanthosis and hyperorthokeratosis, with a distinct granular layer. Beneath the epithelium, there is band-like lymphocytic infiltration in the connective tissue. These lymphocytes have migrated into the epithelium in some areas	ChatGPT 3.5 Preliminary Diagnosis: Aphthous Mucositis Differential Diagnoses: Lupus Erythematosis, Pemphigus Vulgaris, Chronic Aphthous Stomatitis	ChatGPT 4 Preliminary Diagnosis: Lichen Planus Differential Diagnoses: Oral Lichenoid Reaction, Psoriasis, Squamous Cell Carcinoma	ChatGPT 3.5 The report states that a sample taken from the buccal tissue inside the mouth has been examined. It notes thickening in the tissue, accumulation of immune cells in the connective tissue, and migration of some cells into the tissue. As a result, a diagnosis of "Lichen Planus," an inflammatory condition, has been made.	ChatGPT 4 This report concerns the examination of a tissue sample taken from the inside of your mouth. Changes have been found in the tissue indicating thickening of the upper layer of the skin and the accumulation of dead cells. Additionally, white blood cells, which the body uses to fight infections, have been observed within this tissue and sometimes entering skin cells. Based on these findings, a diagnosis of "Lichen Planus" has been made, which is a chronic skin condition that often causes uncomfortable sores in the mouth. This condition is associated with the body's immune system mistakenly attacking its own tissues, and treatment is aimed at alleviating symptoms.
Inflammatory Fibrous Hyperplasia	Sample Area: Alveolar Mucosa The examined sections show a stratified squamous epithelium covered with parakeratosis and irregular acanthosis. Beneath this, there is dense connective tissue consisting of hyalinized collagen fibers, containing a patchy chronic inflammatory infiltrate. Inflammatory cells have occasionally migrated into the epithelium.	Histopathological Diagnosis: Chronic Ulcerative Mucositis Differential Diagnoses: Ulcerative Colitis, Hypersensitivity Reaction, Chronic Leukoplakia	Histopathological Diagnosis: Oral mucosal lesion consistent with chronic inflammation Differential Diagnoses: Lichen Planus, Squamous Cell Carcinoma, Aphthous Ulcer	The report indicates that there is chronic inflammation beneath the examined mucosa, resulting in thickening and densification of the tissue. This condition has been diagnosed as "inflammatory fibrous hyperplasia."	The report explains the result of examining a tissue from the oral cavity under a microscope and summarizes the following: The oral tissue (mucosa) is inflamed and thickened. There are hardened fibers and inflammatory cells within the tissue. This condition indicates that a fibrous (fibrous) growth has occurred in response to prolonged irritation or inflammation in the tissue. As a result, a condition known as "Inflammatory Fibrous Hyperplasia" has been identified, meaning there is excessive and abnormal growth associated with inflammation in the tissue. This signifies an issue in the oral cavity, suggesting that a specialist should be consulted for further evaluation and treatment.
Mild Dysplasia	Sample Area: Lateral Tongue Mucosa The examined serial sections show that the surface of the specimens is covered by multilayered squamous epithelium exhibiting parakeratosis, irregular acanthosis,	Histopathological Diagnosis: Mucosal Hyperplasia Differential Diagnoses: Lichen Planus, Leukoplakia, HPV Reaction	Histopathological Diagnosis: Oral Leukoplakia or Oral Epithelial Dysplasia Differential Diagnoses: Oral Candidiasis, Lichen Planus, Squamous Cell Carcinoma	The sample taken from the tongue mucosa was examined. The results showed mild cellular changes (dysplasia) on the mucosa. These changes were detected alongside some signs of	This report describes the results of a detailed examination conducted on a sample taken from your tongue. The examination revealed some abnormal changes in the tissue of your tongue. These changes indicate that the cells in certain areas of the tissue are growing irregularly and

	and areas of papillomatosis. In the basal layer of the epithelium, there is a localized increase in cell numbers and loss of polarity. The rete ridges display a thickened appearance resembling "fil" (foot) in certain areas. Below these structures, particularly in the superficial sections, chronic inflammatory cell infiltration is observed. In the deeper part of the larger specimen, a salivary gland lobe with dilated ducts can be seen			inflammation. Although the current risk of cancer is low, regular monitoring is recommended.	are inflamed. However, this condition is currently mild and does not indicate cancer. Nevertheless, such changes could potentially lead to more serious issues in the future, so your doctor will closely monitor the situation and may recommend treatment if necessary.
Moderate dysplasia + Proliferative verrucous leukoplakia	<p>Sample Area: Mucosa of the Floor of the Mouth</p> <p>In all sections, the covering stratified squamous epithelium shows advanced hyperkeratosis (hypergranulosis), acanthosis, and papillomatosis. In the areas identified as lesions, prominent finger-like projections extending towards the surface are observed. The rete ridges extend into the connective tissue and thicken as they merge. These changes gradually diminish toward the surgical margins, where the epithelium appears hyperkeratotic and smooth, with occasional areas of edema. Below these layers, there is connective tissue with superficial lymphocytic infiltration, followed by well-developed mature adipose tissue and deeper striated muscle tissue</p>	<p>Histopathological Diagnosis: Epithelial Dysplasia</p> <p>Differential Diagnoses: Aphthous Ulceration, Leukoplakia, Fibroma</p>	<p>Histopathological Diagnosis: Oral Lichen Planus or Oral Leukoplakia</p> <p>Differential Diagnoses: Oral Candidiasis, Squamous Cell Carcinoma, Erythroplakia</p>	<p>The report indicates that there are some abnormal changes in the tissue at the floor of the mouth. These changes are characterized by thickening, protrusions, and cellular abnormalities in the surface of the tissue. These findings could be precancerous and may require appropriate treatment and monitoring</p>	<p>The report indicates that there are some abnormal changes in the tissue lining the inside of your mouth. These changes include: The top layer of tissue is thicker and harder than normal. There are irregular protrusions on the surface of the tissue. Signs of inflammation are present, with healthy fat and muscle tissue found beneath the affected area. The diagnosis suggests abnormal cell growth, and a condition known as Proliferative Verrucous Leukoplakia (PVL), which carries a risk of turning into cancer, has been identified. This indicates unhealthy changes in some areas of the tissue inside your mouth. It's important to discuss with your doctor what this means and what actions should be taken.</p>
Squamous Cell Carcinoma	<p>Sample Area: Alveolar Mucosa</p> <p>In the examined sections, tumor islands are observed,</p>	<p>Histopathological Diagnosis: Squamous Cell Carcinoma</p>	<p>Histopathological Diagnosis: Squamous Cell Carcinoma</p>	<p>Of course, I will try to understand. Samples taken from an area called the alveolar mucosa were</p>	<p>The report diagnoses a type of skin cancer called squamous cell carcinoma. This indicates the presence of abnormal (different in</p>

	some of which are associated with the stratified squamous epithelium covering the fragments, while others are spread as separate masses within the connective tissue. The tumor islands consist of atypical squamous layer cells. Most of them show keratin plugs in their centers. Some tumor cells are distributed either individually or in smaller groups. A mild inflammatory reaction is noted surrounding the tumor.	Differential Diagnoses: Ameloblastoma, Papilloma, Adenocarcinoma	Differential Diagnoses: Verrucous Carcinoma, Basal Cell Carcinoma, Papilloma	examined. In some areas, there was normal tissue known as stratified squamous epithelium. In other areas, however, tumors with abnormal growths were found. These tumors contain cells that are normally not present, and some of them showed abnormalities. There is a mild inflammatory response around the tumors. All these findings indicate the presence of a type of cancer called "squamous cell carcinoma." Appropriate steps can be taken for treatment and monitoring.	structure and appearance from normal) and uncontrolled growth of cells grouped together in the skin or mucous membranes. Most of these cells show an accumulation of a protein called keratin, and a mild inflammatory response has been observed around them. In summary, cancerous cells have been found in the tissue samples, and they exhibit specific characteristics.
Mucous Extravasation	Sample Area: Lip Mucosa In the examined sections, the surface is covered by stratified squamous epithelium showing parakeratosis. Below this, expanded salivary gland ducts and mucus lakes are observed, spread over a wide area. Within and around these lakes, there is dense infiltration of lymphocytes, plasma cells, occasional neutrophils, and histiocytic cells.	Histopathological Diagnosis: Chronic lip mucosa inflammation Differential Diagnoses: Leukoplakia, Lichen Planus, Lip Cancer	Histopathological Diagnosis: Chronic Mucositis Differential Diagnoses: Sjögren's Syndrome, Acute Mucositis, Lichen Planus	Sure, the report examines a section of the inner surface of the lip. There is a layer called parakeratosis on top, beneath which are enlarged salivary gland ducts and mucus lakes. It is observed that immune cells have accumulated around these lakes, which may indicate a possible infection. As a result, increased secretion from the mucous glands and accumulation of immune cells have been detected.	This report describes some changes in the inner surface of your lips. There is a thickening in the cells that cover the inside of the lips, as well as an enlargement of some of the salivary glands in your lips. Additionally, mucus accumulations have been observed, which contain various immune cells in high density as a sign of inflammation. The diagnosis indicated is mucus extravasation, which shows that mucus is leaking into the tissues in an abnormal way. This is usually not a serious condition, but your doctor may provide recommendations for treatment or follow-up.
Pyogenic Granuloma	Sample Area: Gingival Mucosa The examined sections are covered by stratified squamous epithelium showing parakeratosis. In some areas, the epithelium has been lost, with exudate and debris present in these regions. Underneath, numerous lumens lined with young endothelial cells are observed. Some of	Histopathological Diagnosis: Chronic inflammatory changes Differential Diagnosis: Gingivitis, Periodontitis, Aphthous stomatitis	Histopathological Diagnosis: Periodontitis Differential Diagnosis: Lichen planus, Squamous cell carcinoma, Aphthous ulcer	The report indicates that a condition known as "pyogenic granuloma" has been identified in the gingival mucosa. This suggests a loss of the protective layer in the upper part of the mucosa, along with signs of infection and tissue damage. Surrounding it, there are various white blood cells involved in fighting the infection.	The report indicates the presence of a "pyogenic granuloma," which is an inflamed and rapidly growing benign tissue growth in the gums. The gum surface shows damage and inflammation in some areas. Surrounding it, there is a concentration of new blood vessels and white blood cells, which indicates inflammation. This condition usually arises as a result of oral irritation or injury and may require treatment.

Giant Cell Granuloma	<p>these lumens are unopened, while others are filled with erythrocytes. Surrounding them, infiltration composed of numerous polymorph neutrophils, lymphocytes, and plasma cells can be seen.</p> <p>Sample Area: Gingival Mucosa In the examined sections, the epithelium covering the areas has largely disappeared. Here, there is exudate and debris present. Underneath, a vascular-rich focal lesion is observed, consisting of mononuclear young mesenchymal cells and multinucleated osteoclastic giant cells. Additionally, there is a significant accumulation of hemosiderin and small areas of new bone formation scattered throughout.</p>	<p>Histopathological Diagnosis: Chronic Periodontitis</p> <p>Differential Diagnosis: Leukoplakia, Periodontal Abscess, Gingival Hyperplasia</p>	<p>Histopathological Diagnosis: Chronic Osteomyelitis</p> <p>Differential Diagnoses: Acute Osteomyelitis, Periapical Abscess, Paget's Disease</p>	<p>According to the report, there is a loss of epithelium in large areas of the gingival mucosa, with inflammation and debris present in these regions. Beneath this, a vascular-rich focal lesion has been observed, containing young mesenchymal cells and multinucleated giant cells. Additionally, there are deposits of hemosiderin and areas of new bone formation. The diagnosis has been made as "giant cell granuloma." This is typically a type of healing disorder that arises due to trauma or infection around the gums.</p>	<p>This report is the result of an examination conducted on a sample taken from the gums. The report indicates that the surface layer of the gums has been damaged and has disappeared in some areas. In these damaged regions, there is inflammation and remnants of necrotic tissue. Additionally, there are signs that the body is attempting to repair the damage: young cells coming in for repair and large cells facilitating bone destruction and remodeling are observed. The report also mentions damaged areas filled with blood vessels, iron deposits from old bleeding, and areas of new bone formation. In conclusion, this condition is defined as a "giant cell granuloma," which is typically characterized as a benign type of tumor. This indicates an inflammatory condition and the body's repair process. It is advisable to consult a specialist for treatment.</p> <p>Of course! The "oral melanotic macule" mentioned in the report is a harmless spot that forms due to the accumulation of melanin pigment, which determines skin color, in the lip tissue inside your mouth. An increase in melanin and its transport to certain cells (melanocytes) in the cells covering the inner surface of your lips has been observed. This is generally a change in color seen on the skin and may not indicate a serious condition. However, if you notice any changes, it is important to</p>
Melanotic Macule	<p>Sample Area: Lip Mucosa The examined sections reveal a covering of parakeratotic stratified squamous epithelium. In the basal layer of the epithelium, there are areas with an increase in the number of melanocytes and basal cells containing melanin pigment. Additionally, melanin incontinence is observed in the superficial layers of the connective tissue.</p>	<p>Histopathological Diagnosis: Simple Lentigo</p> <p>Differential Diagnoses: Lentigo, Melanosis, Solar Lentigo</p>	<p>Histopathological Diagnosis: Melanotic macular pigmentation</p> <p>Differential Diagnoses: Melanoma, Nevus, Actinic Keratosis</p>	<p>The report examines a sample taken from the lip mucosa. There is an abnormal condition in the epithelial layer covering the inner surface of the lips. Some cells show an increase in melanin pigment, and there is an increase in the number of melanocytes. It is observed that the distribution of melanin pigment in the connective tissue is abnormal. This condition is referred to as "oral melanotic macule." It is generally benign and</p>	

Squamous Papilloma	Sample Area: Buccal Mucosa The examined sections reveal a hyperparakeratotic stratified squamous epithelium covered by papillomatous and irregular acanthosis. Numerous small and large vascular sections are present between the papillaries. Koilocytic cells are particularly observed in the superficial regions of the epithelium.	Histopathological Diagnosis: Papillomatosis Differential Diagnoses: Candidiasis, Leukoplakia, Lichen Planus	Histopathological Diagnosis: Papillomatosis Differential Diagnoses: Verruca Vulgaris, Lichen Planus, Squamous Cell Carcinoma	usually does not require treatment. However, expert evaluation is important. This report summarizes the examination of samples taken from the oral mucosa. The results indicate that the surface of the mucosa is raised, the epithelial layer is abnormally thickened, and there are numerous blood vessel cross-sections present. Additionally, "koilocytic cells" have been detected in the superficial regions of the epithelial layer. These findings suggest a benign condition known as squamous papilloma. This is a common type of benign lesion in the oral cavity. It is important to consult your doctor for treatment and management.	The report indicates that there are some abnormal changes in the tissue covering the inner surface of the mouth (buccal mucosa). There is thickening of the tissue, excessive keratin production, and small protrusions present. Additionally, certain specific cells have been detected that may be indicative of an HPV infection. As a result, these findings suggest the presence of a "squamous papilloma," which is generally a benign and harmless tumor. While it may require treatment, it is usually not a serious condition.
---------------------------	---	--	---	---	---

Experts' Evaluation

Eight expert oral and maxillofacial radiologists and eight expert pathologists, each with at least five years of clinical and academic experience in their respective fields, evaluated the responses generated by ChatGPT 3.5 and 4 using a questionnaire created on Google Forms. The evaluators were not informed of the version to which each response belonged. The questionnaire comprised 10 responses in which the LLMs interpreted the diagnosis from the radiology/pathology reports and 10 responses in which they simplified the radiology/pathology reports for the patient. In this evaluation phase, the experts were requested to rate the answers on a scale of 1 to 5, with 1 indicating a response that was 'completely wrong', 2 representing a response that was 'mostly wrong', 3 signifying an 'average' response, 4 indicating a response that was 'mostly correct', and 5 denoting a response that was 'completely correct'.

Statistical Analysis

Descriptive statistics were calculated to summarize the experts' evaluating scores. The Mann-Whitney U test was used to compare the differences in scores between Chat-GPT 3.5 and Chat-GPT 4 for both radiology and histopathology tasks. To assess inter-rater reliability among experts, the Intraclass Correlation Coefficient (ICC)

was computed. A p-value of <0.05 was considered statistically significant.

All statistical analyses were performed using IBM SPSS Statistics for Windows, version 20.0 (IBM Corp., Armonk, NY, USA)

Results

The mean number of years' experience of the eight dentomaxillofacial radiologists who evaluated the responses provided by the Chat-GPT versions was 6.87 years (minimum = 1, maximum = 17). The results of the ICC analysis, which was conducted to assess the compatibility of the scores provided by the radiology specialists, indicated a moderate statistically significant agreement between the experts regarding the scores assigned to the responses generated by Chat-GPT 3.5 ($r = 0.661$, $p < 0.001$). Similarly, for the responses generated by Chat-GPT 4, the experts demonstrated a moderately significant level of agreement ($r = 0.602$, $p < 0.001$).

The mean number of years' experience of the pathology experts who evaluated the responses of Chat-GPT on pathology reports was 13.87 (minimum = 6, maximum = 29). The ICC analysis, conducted to assess the inter-rater reliability of the pathology experts' evaluations, revealed a moderate statistically significant agreement between the experts in assessing the

responses generated by Chat-GPT 3.5 ($r = 0.545$, $p < 0.001$). Similarly, the evaluation of the responses provided by Chat-GPT 4 revealed a moderate statistically significant agreement between the experts ($r = 0.729$, $p < 0.001$).

The overall performance of Chat-GPT 3.5 and Chat-GPT 4 in diagnosing lesions based on CBCT reports and simplifying radiological findings for patient understanding was evaluated by dentomaxillofacial radiologists on a total of 50 points (10 questions with 5 points each). Each evaluator scored 10 distinct model responses for each task (diagnosis and simplification) using a 5-point Likert scale (1 = completely wrong, 5 = completely correct), resulting in a maximum possible score of 50 per category. Chat-GPT 3.5 scored an average of 23.87 ± 4.61

(minimum= 15 maximum= 30) for diagnostic accuracy, while Chat-GPT 4 achieved a significantly higher average score of 37.00 ± 3.33 (minimum= 32 maximum= 42). Comparison of diagnostic performance using the Mann-Whitney U test revealed that Chat-GPT 4 was statistically superior to Chat-GPT 3.5 ($U = 3.000$, $p < 0.001$). Regarding the ability to simplify radiological reports for patients, Chat-GPT 3.5 scored a mean of 35.12 ± 4.94 (minimum= 27 maximum= 44), while Chat-GPT 4 scored slightly higher with a mean of 37.75 ± 3.33 (minimum= 32 maximum= 42).

However, the comparison between the two versions for report simplification, as analyzed by the Mann-Whitney U test, did not show a statistically significant difference ($U = 30.500$, $p = 0.143$) (Table 3).

Table 3. Comparison of Chat-GPT 3.5 and 4 scores in diagnosis and report simplification

Task	LLM Version	Expert (n)	Mean (\pm SD)	Mean Rank	p
Lesion diagnosis from the radiology report	Chat-GPT 3.5	8	23.87 ± 4.61	5.80	$p < 0.001$
	Chat-GPT 4	8	37.00 ± 3.33	15.20	
Simplifying the radiology report for the patient	Chat-GPT 3.5	8	35.12 ± 4.94	8.15	$p > 0.05$
	Chat-GPT 4	8	37.75 ± 3.33	12.45	
Lesion diagnosis from the histopathology report	Chat-GPT 3.5	8	21.25 ± 3.91	7.65	$p < 0.05$
	Chat-GPT 4	8	30.62 ± 5.37	13.35	
Simplifying the histopathology report for the patient	Chat-GPT 3.5	8	30.87 ± 4.61	6.75	$p < 0.05$
	Chat-GPT 4	8	38.37 ± 7.15	14.25	

The overall performance of Chat-GPT 3.5 and Chat-GPT 4 in diagnosing lesions from pathology reports and simplifying pathology reports for patient understanding was evaluated by pathology experts, with scores given out of a total of 50 points. Chat-GPT 3.5 achieved a mean score of 21.25 ± 3.91 (minimum = 18, maximum = 29) for diagnostic accuracy, while Chat-GPT 4 achieved a higher mean score of 30.62 ± 5.37 (minimum = 19, maximum = 37). A comparison of the diagnostic performance of the two LLMs using the Mann-Whitney U test revealed that Chat-GPT 4 was statistically superior to Chat-GPT 3.5 ($U = 21.500$, $p = 0.029$). In terms of simplifying pathology reports for patient comprehension, Chat-GPT 3.5 received a mean score of 30.87 ± 4.61 (minimum = 23, maximum = 39), while Chat-GPT 4 scored higher with a mean of 38.37 ± 7.15 (minimum = 26, maximum = 47). The Mann-Whitney U test showed that Chat-GPT 4 was significantly more successful in simplifying reports compared to Chat-GPT 3.5 ($U = 12.500$, $p = 0.003$) (Table 3).

To evaluate whether Chat-GPT 3.5's performance in diagnosing lesions based on report findings and simplifying reports for patient understanding was more effective in radiology reports compared to pathology reports, the scores assigned by experts were analyzed using the Mann-Whitney U test. The results indicated that there was no statistically significant difference in Chat-GPT 3.5's ability to make a diagnosis from radiology reports versus pathology reports ($U = 18.500$, $p = 0.161$), nor in its ability to simplify the report for patient comprehension ($U = 16.000$, $p = 0.105$). In contrast, when Chat-GPT 4's performance was analyzed, the scores awarded by dentomaxillofacial radiologists for interpreting radiology reports were significantly higher than those given by pathology experts for interpreting pathology reports ($U = 8.500$, $p = 0.010$). However, no statistically significant difference was observed between radiology and pathology reports in Chat-GPT 4's ability to simplify reports for patients ($U = 29.000$, $p = 0.798$) (Table 4).

Table 4. Comparison of Chat-GPT's performance in radiology reports versus histopathology reports

LLM Version	Task	Expert (n)	Mean (\pm SD)	Mean Rank	p
Chat-GPT 3.5	Lesion diagnosis from the radiology report	8	23.87 ± 4.61	10.19	$p > 0.05$
	Lesion diagnosis from the histopathology report	8	21.25 ± 3.91	6.81	
	Simplifying the radiology report for the patient	8	35.12 ± 4.94	10.50	$p > 0.05$
	Simplifying the histopathology report for the patient	8	30.87 ± 4.61	6.50	
Chat-GPT 4	Lesion diagnosis from the radiology report	8	37.00 ± 3.33	11.44	$p = 0.010$

Lesion diagnosis from the histopathology report	8	30.62 ± 5.37	5.56	p>0.05
Simplifying the radiology report for the patient	8	37.75 ± 3.33	8.13	
Simplifying the histopathology report for the patient	8	38.37 ± 7.15	8.88	

Discussion

Chat-GPT, consisting of its basic chat-based interface, provides a user-friendly experience that is easily accessible even to individuals with no prior experience in AI technologies. Regarding its simple design and capacity to provide precise and contextually relevant replies, Chat-GPT has been increasingly adopted across a wide range of research disciplines.¹² Two areas of research that have garnered significant attention are the potential of this tool to enhance both research and clinical workflows in radiology and pathology.^{12,13} In both domains, studies have demonstrated that Chat-GPT and similar LLMs can be effectively utilized for various purposes, including the training of specialists, report generation, patient education, and aiding in the diagnosis and follow-up of diseases.^{2,14-19} The use of the complete capabilities of domain-specific LLMs and integrating these tools into academic and clinical settings could lead to substantial enhancements in efficiency, precision, and the advancement of scientific knowledge, which will help all those involved in the healthcare sector, from clinicians to patients.¹³ Considering this information, the present study investigated the diagnostic capabilities of various versions of Chat-GPT based on findings from radiology and histopathology reports, as well as their ability to present these reports in a manner comprehensible to patients.

The integration of ChatGPT and similar artificial intelligence models into radiology raises important ethical questions. First of all, clear rules on patient privacy and data security should be determined, and the use of artificial intelligence should be clearly communicated to patients and referring physicians. The contribution of AI to diagnostic processes may affect patients' trust in the quality of care and the healthcare system. Therefore, it should be emphasised that AI is only a supportive tool and the final decision should be made by expert radiologists. Furthermore, responsibility for diagnostic errors caused by AI should be clearly defined and accountability mechanisms should be established both legally and professionally. Potential biases in AI algorithms should be minimised by using diverse and inclusive training data, thus ensuring fair and reliable results. Finally, obtaining informed consent from patients regarding the use of AI will play a critical role in ensuring compliance with ethical standards by increasing transparency in the integration of technology into healthcare.²⁰

Many studies have been conducted to evaluate the accuracy of ChatGPT in reporting accurate radiology and general information about pathophysiology, and mostly positive results have been obtained. Bhayana et al. used 150 multiple-choice questions with similar style, content and difficulty level to the Royal College of Canada and

American Board of Radiology examinations between 25 February and 3 March 2023. The model performed well on low-level cognitive thinking (84%) and clinical management questions (89%), but poorly on higher-level thinking (60%), identification of imaging findings (61%), calculation and classification (25%). Physics questions (40%) were found to be less successful than clinical questions (73%). ChatGPT always used confident language, even when giving incorrect answers. The results suggest that the model may be helpful in radiology education but should be used with caution in clinical decision-making.²¹

Expert evaluations of Chat-GPT 4's performance in diagnosing lesions from radiology reports revealed a score of 37.00 ± 3.33 out of 50. This suggests that Chat-GPT 4 has the potential to assist clinicians during the preliminary diagnosis phase by interpreting findings from expert-written CBCT reports. However, this level of performance was not observed in Chat-GPT 3.5, which received a significantly lower score of 23.87 ± 4.61 out of 50 for its ability to make a preliminary diagnosis from CBCT reports. The performance disparity between Chat-GPT 3.5 and Chat-GPT 4, indicating that Chat-GPT 3.5 is substantially less reliable than Chat-GPT 4 for this specific task. Expert evaluations of Chat-GPT 4's performance in simplifying radiology reports for patient comprehension resulted in a score of 37.75 ± 3.33 out of 50. These findings indicate that Chat-GPT 4 is capable of effectively translating CBCT reports into a patient-friendly format, which could improve physician-patient communication and support patient education. While Chat-GPT 3.5 received a slightly lower score of 35.12 ± 4.94, the performance difference between the two versions was not statistically significant. Therefore, Chat-GPT 3.5 also shows potential in making radiology reports more accessible to patients, though with somewhat lower effectiveness compared to Chat-GPT 4. This difference may be attributed to the enhanced language comprehension and instruction-following capabilities of Chat-GPT 4, which allow it to simplify technical medical content more effectively than Chat-GPT 3.5. In contrast, Chat-GPT 3.5 may struggle with nuanced expression or resort to terminology that is still overly technical for laypersons.

In the context of tasks involving histopathology reports, Chat-GPT 4 demonstrated significantly better performance in both areas evaluated. Specifically, Chat-GPT 4 received scores of 30.62 ± 5.37 and 38.37 ± 7.15 out of 50 for these tasks, indicating that it shows promise in simplifying histopathology reports for patient comprehension. However, the score of 30.62 ± 5.37 for making a preliminary diagnosis from histopathology reports is notably lower than the 37.00 ± 3.33 score

achieved for preliminary diagnosis from radiology reports. This suggests that while Chat-GPT 4 is effective in processing both types of reports, it is significantly more proficient in deriving preliminary diagnoses from radiology findings compared to histopathology findings. It is stated that ChatGPT 4 can solve complex problems with higher accuracy thanks to its improved reasoning ability and wider knowledge base.²² Chat-GPT 4 has a more advanced architecture and a more comprehensive training data. In this way, it can analyse medical terms, anatomical structures and findings in reports more accurately. Chat-GPT 3.5, on the other hand, is insufficient in synthesising information, especially in the clinical context, as it has a more limited understanding capacity.

This discrepancy may be explained by the nature of the content in histopathology reports, which often involve more complex terminology, abstract concepts, and context-dependent findings at the microscopic level. In contrast, radiology reports typically include more structured and standardized language describing macroscopic anatomical features. These differences may affect how well the model interprets and synthesizes information. Furthermore, it is possible that the training data used for Chat-GPT 4 included relatively fewer examples of histopathological language compared to radiological terminology, contributing to the observed performance gap.

In their systematic review on the application of Chat-GPT in radiology, Keshavarz et al. reported that across eleven studies in the field, researchers compared the performance of Chat-GPT 4 with that of Chat-GPT 3.5. In 90% of these studies, Chat-GPT 4 was found to outperform Chat-GPT 3.5, particularly in solving complex problems and demonstrating advanced reasoning skills.¹² In their study, Rao et al. evaluated the performance of ChatGPT-3.5 and GPT-4 in clinical decision support for radiology, focusing on two scenarios: breast cancer screening and breast pain. Using both open-ended and select-all-that-apply (SATA) prompts based on ACR Appropriateness Criteria, they found that GPT-4 significantly outperformed GPT-3.5. Specifically, GPT-4 achieved 98.4% accuracy in SATA prompts for breast cancer screening compared to 88.9% for GPT-3.5, and 77.7% versus 58.3% for breast pain scenarios. These results demonstrate GPT-4's improved contextual understanding and reasoning capabilities in complex clinical tasks.²³ Consistent with these findings, our study showed that Chat-GPT 4 was superior to Chat-GPT 3.5 in all tasks, except for simplifying radiology reports for patients. Similarly, Doshi et al. reported no significant difference between Chat-GPT 3.5 and Chat-GPT 4 in simplifying radiology reports, paralleling the results observed in our study. This suggests that software updates do not always lead to improved performance, particularly in tasks related to readability.²⁴

In a separate study, Hirosowa et al. compared the diagnostic and differential diagnosis capabilities of Chat-GPT 3 with those of physicians across various clinical scenarios. Although their findings indicated that

physicians' responses were significantly more accurate, they demonstrated that general AI chatbots, like Chat-GPT 3, were also capable of generating well-differentiated diagnostic lists for common chief complaints.²⁵ In the present study, while the diagnostic outputs of Chat-GPT were not directly compared with those of medical experts, the responses generated by the language models were evaluated by experts, revealing that Chat-GPT, particularly its latest version, can achieve high accuracy in deriving differential diagnoses and preliminary diagnoses from radiology and histopathology reports. Nevertheless, it is important to recognize that the generated diagnoses are not flawless, and large language models such as Chat-GPT are not intended to replace physicians, but rather to function as supportive tools within clinical workflows.

A key limitation of this study is the exclusive use of Chat-GPT among LLMs. Future research should adopt a more comprehensive approach, evaluating the performance of other LLMs in executing these specific tasks and conducting comparative analyses between them. Such an approach would offer more robust and reliable insights into the performance capabilities of various chatbots. Additionally, since the reports in this study were written in Turkish, both the input prompts and the generated responses were in Turkish, which may have influenced Chat-GPT's performance. Further studies could explore the impact of language on the accuracy and efficiency of LLMs.

Conclusions

Chat-GPT has the potential to serve as a valuable tool in supporting specialists by aiding in the preliminary diagnosis of radiology and histopathology reports, generating differential diagnoses, and providing explanations for these diagnostic possibilities. Additionally, it may prove beneficial in simplifying complex reports for patients, making them more accessible to those without specialized knowledge. However, it is crucial that the use of Chat-GPT or other LLMs for these purposes remains under the supervision of experts, with the final decisions and interpretations being made by qualified professionals.

Acknowledgements

This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

During the preparation of this work the author(s) used Chat-GPT in order to translate the histopathological/radiological reports and the answers given by Chat-GPT in the study into English for publication since they were in Turkish. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

The radiological data of this study was presented at the Genç ORAD Congress held in Adana between March

5–9, 2024, and was awarded first prize among the oral presentations.

Conflicts of Interest Statement

The authors declare no conflicts of interest. They have no financial, personal, or institutional relationships that could have influenced the work reported in this study.

References

- Buldur B, Teke F, Kurt MA, Sağtaş K. Perceptions of Dentists Towards Artificial Intelligence: Validation of a New Scale. *Cumhuriyet Dental Journal* 2024;27(2): 109–117.
- Elkasssem AA, Smith AD. Potential Use Cases for ChatGPT in Radiology Reporting. *AJR Am J Roentgenol* 2023;221(3):373-376.
- Sacoransky E, Kwan BYM, Soboleski D. ChatGPT and assistive AI in structured radiology reporting: A systematic review. *Curr Probl Diagn Radiol* 2024;53(6):728-737.
- Sun Z, Ong H, Kennedy P, et al. Evaluating GPT4 on Impressions Generation in Radiology Reports. *Radiology* 2023;307(5):e231259.
- Zhong T, Zhao W, Zhang Y, et al. ChatRadio-Valuer: A Chat Large Language Model for Generalizable Radiology Report Generation Based on Multi-institution and Multi-system Data. Published online October 9, 2023. doi:10.48550/arXiv.2310.05242
- Persigehl T, Baumhauer M, Baeßler B, et al. Structured Reporting of Solid and Cystic Pancreatic Lesions in CT and MRI: Consensus-Based Structured Report Templates of the German Society of Radiology (DRG). *ROFO Fortschr Geb Rontgenstr Nuklearmed* 2020;192(7):641-656.
- Brook OR, Brook A, Vollmer CM, Kent TS, Sanchez N, Pedrosa I. Structured reporting of multiphasic CT for pancreatic cancer: potential effect on staging and surgical planning. *Radiology* 2015;274(2):464-472.
- Kabadi SJ, Krishnaraj A. Strategies for Improving the Value of the Radiology Report: A Retrospective Analysis of Errors in Formally Over-read Studies. *J Am Coll Radiol* 2017;14(4):459-466.
- Nobel JM, van Geel K, Robben SGF. Structured reporting in radiology: a systematic review to explore its potential. *Eur Radiol* 2022;32(4):2837-2854.
- Kaka H, Zhang E, Khan N. Artificial Intelligence and Deep Learning in Neuroradiology: Exploring the New Frontier. *Can Assoc Radiol J* 2021;72(1):35-44.
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18(8):500-510.
- Keshavarz P, Bagherieh S, Nabipoorashrafi SA, et al. ChatGPT in radiology: A systematic review of performance, pitfalls, and future perspectives. *Diagn Interv Imaging* 2024;105(7-8):251-265.
- Omar M, Ullanat V, Loda M, Marchionni L, Umeton R. ChatGPT for digital pathology research. *Lancet Digit Health* 2024;6(8):e595-e600.
- Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a Radiology Board-style Examination: Insights into Current Strengths and Limitations. *Radiology* 2023;307(5):e230582.
- Jj C, Dh K, Tt G, et al. Accuracy of Information Provided by ChatGPT Regarding Liver Cancer Surveillance and Diagnosis. *AJR Am J Roentgenol* 2023;221(4).
- Lyu Q, Tan J, Zapadka ME, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Vis Comput Ind Biomed Art* 2023;6:9.
- Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of Breast Cancer Prevention and Screening Recommendations Provided by ChatGPT. *Radiology* 2023;307(4):e230424.
- Darzidehkalani E. ChatGPT in Medical Publications. *Radiology* 2023;307(5):e231188.
- Gunn AJ. Commentary: The Emerging Role of Artificial Intelligence for Patient Education. *J Vasc Interv Radiol* 2023;34(10):1769-1770
- Ethan Sacoransky, Benjamin Y.M. Kwan, Donald Soboleski, ChatGPT and assistive AI in structured radiology reporting: A systematic review, *Current Problems in Diagnostic Radiology* 2024;53(6): 728-737,
- R. Bhayana, S. Krishna, R.R. Bleakney Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. *Radiology* 2023; 307 (5):e230582.
- Massey, Patrick A. MD, MBA; Montgomery, Carver MD; Zhang, Andrew S MD. Comparison of ChatGPT–3.5, ChatGPT–4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. *Journal of the American Academy of Orthopaedic Surgeons* 2023;31(23):1173-1179.
- A. Rao, J. Kim, M. Kamineni, M. Pang, W. Lie, K.J. Dreyer, M.D. Succi Evaluating GPT as an adjunct for radiologic decision making: GPT-4 versus GPT-3.5 in a breast imaging pilot. *J Am Coll Radiol* 2023;20(10):990-997.
- Doshi R, Amin K, Khosla P, Bajaj S, Chheang S, Forman HP. Utilizing Large Language Models to Simplify Radiology Reports: a comparative analysis of ChatGPT3.5, ChatGPT4.0, Google Bard, and Microsoft Bing. Published online June 5, 2023:2023.06.04.23290786.
- Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. *Int J Environ Res Public Health* 2023;20(4):3378.