# Holistic Evaluation of GPT-4V for Biomedical Imaging

Zhengliang Liu *[1], Hanqi Jiang *[1,2], Tianyang Zhong *[3], Zihao Wu *[1], Chong Ma †[3], Yiwei Li †[1], Xiaowei Yu †[4], Yutong Zhang †[5], Yi Pan †[1,6], Peng Shu †[1], Yanjun Lyu †[4], Lu Zhang †[4], Junjie Yao †[6], Peixin Dong †[3], Chao Cao †[4], Zhenxiang Xiao †[6], Huan Zhao †[6], Jiaqi Wang †[7], Shaochen Xu †[1], Yaonai Wei †[3], Jingyuan Chen[8], Peilong Wang[8], Haixing Dai[1], Hao He[9], Zewei Wang[3], Xinyu Wang[3], Xu Zhang[9], Lin Zhao[1], Yiheng Liu[9], Youlan Shang[11], Kai Zhang[10], Zhiling Yan[10], Wei Zhao[11], Jun Liu[11], Lichao Sun[10], Ning Qiang[9], Bao Ge[9], Xiaoyan Cai[3], Shijie Zhao[3], Xintao Hu[3], Yixuan Yuan[12], Gang Li[13], Shu Zhang[7], Xin Zhang[5], Xi Jiang[6], Tuo Zhang[3], Dinggang Shen[14,15,16], Quanzheng Li[17], Wei Liu[8], Xiang Li[17], Dajiang Zhu[4], and Tianming Liu‡[1]

[1]School of Computing, University of Georgia, GA, USA
[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[3]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
[4]Department of Computer Science and Engineering, University of Texas at Arlington, TX, USA
[5]Institute of Medical Research, Northwestern Polytechnical University, Xi'an 710072, China
[6]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China
[7]School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
[8]Department of Radiation Oncology, Mayo Clinic, Phoenix, Arizona, USA
[9]School of Physics and Information Technology, Shaanxi Normal University, Xi'an 710119, China
[10]Department of Computer Science and Engineering, Lehigh University, PA, USA
[11]Department of Radiology, Second Xiangya Hospital, Central South University, Changsha, 410011, China
[12]Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong 999077, China
[13]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA
[14]School of Biomedical Engineering, ShanghaiTech University, and Shanghai Clinical Research and Trial Center, Shanghai 201210, China
[15]Shanghai United Imaging Intelligence Co., Ltd.
[16]Shanghai Clinical Research and Trial Center
[17]Department of Radiology, Massachusetts General Hospital and Harvard Medical School, MA, USA

*Co-first authors.
†Co-second authors.
‡Corresponding author. E-mail: tliu@uga.edu

**Abstract**

In this paper, we present a large-scale evaluation probing GPT-4V's capabilities and limitations for biomedical image analysis. GPT-4V represents a breakthrough in artificial general intelligence (AGI) for computer vision, with applications in the biomedical domain. We assess GPT-4V's performance across 16 medical imaging categories, including radiology, oncology, ophthalmology, pathology, and more. Tasks include modality recognition, anatomy localization, disease diagnosis, report generation, and lesion detection. The extensive experiments provide insights into GPT-4V's strengths and weaknesses. Results show GPT-4V's proficiency in modality and anatomy recognition but difficulty with disease diagnosis and localization. GPT-4V excels at diagnostic report generation, indicating strong image captioning skills. While promising for biomedical imaging AI, GPT-4V requires further enhancement and validation before clinical deployment. We emphasize responsible development and testing for trustworthy integration of biomedical AGI. This rigorous evaluation of GPT-4V on diverse medical images advances understanding of multimodal large language models (LLMs) and guides future work toward impactful healthcare applications.

# 1 Introduction

Large language models (LLMs) such as ChatGPT [1] and GPT-4 [2] have demonstrated immense progress in natural language processing [3, 4, 5, 6, 7, 8]. However, their ability to interpret and leverage visual information remains relatively underexplored [9, 10, 11, 12, 13], especially in specialized domains such as medicine and the general biomedical field [9, 14, 15, 16, 17, 18, 19, 20, 21]. In this paper, we present a large-scale evaluation probing GPT-4V's capabilities and limitations for biomedical image analysis.

## 1.1 Background: The Rise of LLMs

The advancement of LLMs has its roots in fundamental breakthroughs in deep learning for natural language processing [3, 4, 22, 23, 24, 25, 26]. Transformers [27], introduced in 2017, marked a significant advancement in natural language processing by overcoming key limitations present in earlier architectures such as Recurrent Neural Networks (RNNs) [28] and Long Short-Term Memory networks (LSTMs) [29]. While RNNs and LSTMs were foundational in handling sequential text data, they struggled with parallelization and were constrained by the length of sequences they could effectively model. In contrast, transformer models, exemplified by BERT [30] and GPT-2 [31], employ an attention mechanism that allows for the processing of long-range dependencies in text without the same constraints on sequence length. This capability significantly enhances the efficiency and scalability of LLMs, forming the basis for the substantial progress seen in subsequent language model developments.

Autoregressive language models such as the GPT series [3] generated text probabilistically one token at a time, allowing flexible and coherent text generation. GPT-1 [32], developed by OpenAI in 2018, was one of the first autoregressive LLMs. More recent models such as GPT-2 [31] and GPT-3 [33] built on these foundations by pre-training transformer-based autoregressive LLMs on ever-increasing textual corpora scraped from the internet. For instance, GPT-3 was pre-trained on hundreds of billions of words. This enabled the models to learn nuances of natural language and world knowledge. Pre-training on massive data from diverse sources enables language models to capture a vast amount of world knowledge and linguistic patterns [3, 5, 34]. InstructGPT [35] built on this by incorporating instruction fine-tuning and reinforcement learning from human feedback (RLHF) into the training process. RLHF helped align the model outputs better with human preferences and values [36].

Together, these innovations significantly enhanced the natural language processing capabilities of LLMs.

These advancements culminated in the release of ChatGPT in late 2022. ChatGPT leveraged the power of pre-training and RLHF to achieve impressive performance on complex language tasks like reasoning, question answering, summarization and dialogue. Its conversational abilities led ChatGPT to gain immense popularity among the public [3, 37]. However, ChatGPT was limited to only text modalities.

Humans naturally process information multimodally, integrating vision, language, and auditory inputs [5]. In addition, multimodal processing is not only a cognitive reality but also a practical necessity in many real-world applications. For instance, autonomous vehicles rely on integrating visual, spatial, and auditory data to navigate safely. In healthcare, diagnostic systems benefit from analyzing visual, textual, and sensor data to provide comprehensive patient assessments. Social media platforms use multimodal AI to understand and moderate content that includes text, images, and videos.

This understanding prompted the extension of LLMs to encompass multimodal intelligence, capable of handling images, videos, and speech. Such advancement led to the creation of models like Kosmos-1 [38], PaLM-E [39], and GPT-4V [40], all of which can process both images and text. For example, GPT-4V was pre-trained on vast datasets that included both text and images sourced from the internet, enabling it to develop visual capabilities alongside its linguistic proficiency. Specifically, GPT-4V enhances the visual processing strengths of the model. The advent of such multimodal LLMs has broadened the horizons and created new possibilities at the confluence of natural language processing, computer vision, and human-AI interaction.

## 1.2 Motivation

Biomedical images contain intricate visual details and domain knowledge [13, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52] that can be difficult even for human experts to interpret. Indeed, specialized domains like radiology [12, 53, 54, 55, 56] and pathology [57, 58] rely on perceiving subtle visual indicators and nuanced clinical knowledge to detect disease and guide diagnoses. For example, a radiologist must discern ambiguous masses, faint shadows, texture patterns, and other anomalies to identify cancer on a CT scan [59, 60, 61]. This requires extensive training to recognize disease manifestations versus normal variations in human anatomy. Similarly, a pathologist relies on detecting slight deviations in cell morphology through a microscope to diagnose illness [62, 63, 64]. The intricacy of biomedical imagery goes far beyond generic object recognition.

However, most benchmark datasets for evaluating AI systems consist of everyday internet images [9, 65, 66, 67, 68, 69, 70, 71, 72]. This motivated our work to rigorously test GPT-4V's capabilities using real-world medical imaging data. By evaluating performance on specialized tasks like localization of anatomical structures and classification of radiography modalities, our study provides empirical insights into how large language models fare on the complex visual reasoning challenges inherent in biomedical domains. Being able to automate and enhance biomedical image understanding could have profound impacts on healthcare, biological research, and medical education [73, 74, 75, 76, 77, 78, 79]. For instance, AI models and tools that can analyze microscope images [80, 81], radiology scans [22, 82], or anatomical diagrams [83] could help accelerate scientific discoveries, improve clinical diagnostics, and augment medical training. Similarly, in the realm of medical robotics [84], the integration of sophisticated AI systems capable of interpreting complex medical imagery in real-time can significantly advance the precision and adaptability of robotic surgeries and interventions.

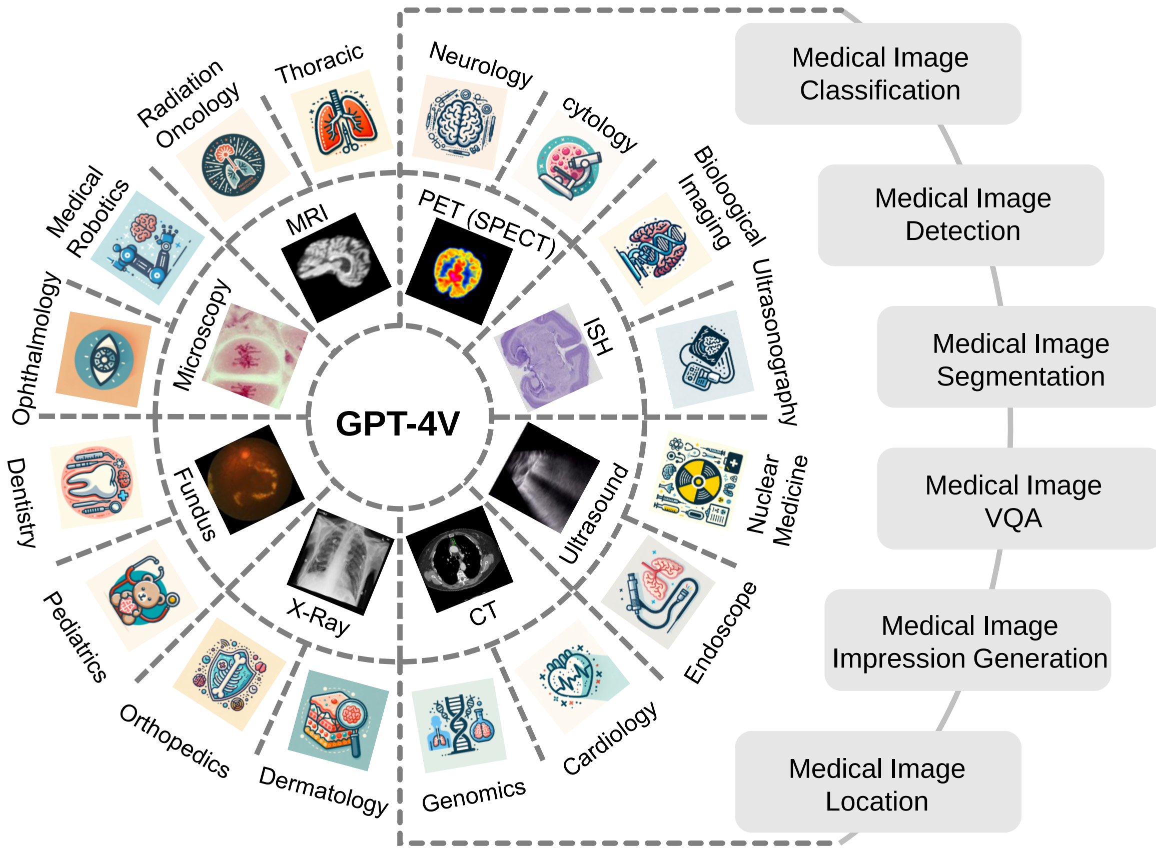



Figure 1: **Schematic Overview of the Evaluation Methodology.** The periphery of the diagram delineates the array of medical departments subjected to our analysis, whereas the core layer graphically represents the medical imaging modalities under scrutiny. The efficacy of GPT-4V was methodically gauged across a spectrum of six distinct medical tasks.

## 1.3 Study Objectives

In our study, we aimed to evaluate GPT-4V across a diverse set of capabilities required for real-world medical imaging tasks. A comprehensive visualization of the evaluation is depicted in Figure 1, where the structural hierarchy of the testing scope is illustrated. For various modalities including X-ray, MRI, CT, and microscopy images, we assessed GPT-4V on its ability to recognize the imaging modality itself. We tested the localization of anatomical structures within medical images, which requires an understanding of human anatomy. Additionally, we benchmarked image classification skills through diagnosis and disease identification from medical images, a crucial application for clinical use. Finally, we explored GPT-4V's capacity for generating free-text reports summarizing image findings, which could provide tremendous workflow efficiencies if automated successfully. By probing performance on modality recognition, region localization, image classification, and report generation, our experiments provide a comprehensive portrait of GPT-4V's strengths and weaknesses at the core skills necessary for biomedical imaging applications. Section 3 details our rigorous assessment across each of these capabilities using diverse benchmark medical imaging data.

Our study encompasses an expansive range of biomedical imaging domains, spanning various medical specialties, biological research, and healthcare applications. We evaluate GPT-4V across chest radiography data, including X-rays from the MIMIC dataset [85], to probe interpretive skills for pulmonary conditions. Additionally, we test capabilities in neuroimaging by assessing performance on MRI data for detecting neurological pathologies. Oncological imaging analysis is examined through GPT-4V's interpretation of MRI and CT scans used in cancer radiotherapy planning and monitoring. Beyond radiology, we explore GPT-4V's skills in cytopathology for cancer diagnosis, using public data of microscopic cell imaging. Ophthalmological imaging, medical robotics, and other specialty areas are incorporated to provide a comprehensive perspective.

Importantly, our work also assesses capabilities on data modalities highly relevant in biological research, such as optical microscopy for cellular imaging [86], in situ hybridization for gene expression [87], and functional MRI [88] in neuroscience studies. This expands the evaluation of GPT-4V's interpretive skills beyond clinical radiology into wider biomedical imaging applications. We believe our heterogeneous datasets better represent the true diversity of visual reasoning tasks required in real-world medical and biological settings compared to benchmarks constrained to a single specialty. By holistically evaluating GPT-4V across this expansive range of biomedical imaging data, our findings provide multifaceted insights into the current capabilities and limitations of large language models for supporting various medical specialties, scientific research, and potential healthcare applications.

## 1.4 Potential Applications and Implications

As powerful AGI systems such as GPT-4V evolve to tackle complex tasks like medical imaging, upholding rigorous validation and ethical principles becomes increasingly crucial [5, 89, 90, 91]. Real-world deployment of biomedical AGI demands extensive testing across diverse patient populations and clinical settings to characterize strengths and limitations [92]. We must proactively assess models for potential biases along lines of race, gender, age, and other factors that may impact equitable performance [93]. Maintaining transparency about AI capabilities and thought processes can help uphold accountability if issues emerge. Ultimately, biomedical AGI should aim to augment clinician expertise and efficiency without compromising personalized care. Human discretion, oversight, and responsibility must remain central to any high-stakes medical application of AI. By emphasizing comprehensive validation, earnest bias evaluation, and clinician-centered design, our research aims to responsibly advance biomedical AI in a manner that earns public trust and optimism about these transformative technologies.

By evaluating GPT-4V on a wide range of biomedical images, we identify both promising capabilities and potential limitations of LLMs for this vital application domain. Our findings will help guide the responsible development and deployment of AI for healthcare and biological research. The remainder of this paper presents our experimental setup, results, and discussion of GPT-4V's proficiencies and areas for improvement in biomedical visual intelligence.

# 2 Scope of the Study and Used Public Datasets

In our study, we aim to explore and evaluate GPT-4V's capabilities and limitations across various domains of medical imaging. Below is a comprehensive list of the domains we have included in our research.

## 2.1 Chest Radiography

Assessing GPT-4V's ability to interpret chest X-rays, including data from MIMIC-CXR [85], CheXpert [94], ChestXray2017 [95], COVID-Qu-Ex [96], OpenI [97], SIIM-ACR [98], and NIH Chest X-rays [99] datasets. The MIMIC-CXR [85] dataset contains 377,110 chest X-ray images and 227,835 diagnostic reports, with each study having a corresponding multi-class label (14 common chest disease labels extracted by CheXpert Laberler [94]). We tested the ability of GPT-4V to directly output multi-class labels based on the images. Additionally, we evaluated GPT-4V's capability to generate structured diagnostic reports (including findings and impression sections) based on the images. From the CheXpert [94] dataset, we selected a subset of frequently encountered categories (Atelectasis, Cardiomegaly, Edema, Consolidation, Pleural Effusion) in zero-shot classification tasks and attempted direct classification using GPT-4V on the corresponding images. The ChestXray2017 [95] dataset consists of 5,863 chest X-ray images, including two classes: pneumonia and normal. We conducted binary classification tasks based on images using GPT-4V. Similarly, for the COVID-Qu-Ex [96] dataset, which provides binary classification labels and mask regions of COVID-19 lesions, we tested GPT-4V's ability to diagnose and locate COVID-related abnormalities in chest X-ray images. The OpenI [97] dataset consists of 3,955 chest X-ray images with corresponding diagnostic reports. We evaluated GPT-4V's capability to generate diagnostic reports on this dataset. The SIIM-ACR [98] dataset provides 2,379 chest X-ray images for pneumothorax segmentation. We tested GPT-4V's ability to directly locate pneumothorax lesions based on the images. The NIH Chest X-rays [99] dataset contains 33,920 chest X-ray images, each with corresponding multi-class labels (14 classes). A small subset of images in this dataset also includes bounding box information for diseases. Similar to the COVID-Qu-Ex dataset, we tested GPT-4V's multi-class classification ability and its capability to locate lesion regions based on the images.

## 2.2 Neuroimaging

In order to verify whether GPT-4V can understand neuronal image datasets and brain images, we tested on SEU-ALLEN's R1741 [100] dataset and the Allen Institute's mouse brain whole-brain atlas. The R1741 dataset is a whole-brain neuron reconstruction dataset completely manually annotated and reconstructed on whole-brain-level fMOST images of the mouse brain. It has several major advantages. First, the resolution of the fMOST [101] image is particularly high. In addition, the neurons in the R1741 dataset are all labeled using sparse labeling technology. The quality of the neuron images is particularly easy to identify. The image quality is compared with similar images. Second, R1741 neurons are all manually labeled and repeatedly checked, and their reliability is also very high. Third, because the size of whole-brain images can often reach dozens of terabytes, special visualization tools (such as Vaa3D [102]) are required to visualize them. This also means that during the training process of GPT-4V, there are almost no opportunities to encounter similar problems. For images, verification on such datasets is a great test of GPT-4V's emergent abilities. The Allen Brain Institute's whole-brain atlas [103] is a template map of mouse brain regions that is highly recognized in the field. Mouse brain regions generally have specific name abbreviations. We tested whether GPT-4V can obtain the brain regions from the image. The relationship between accurate information and adjacent brain areas is very meaningful for testing whether GPT-4V has knowledge transfer and synthesis capabilities.

Since whole-brain images are generally displayed in 3D, in this test, we selected parts with clearer imaging, sparser neuron fibers, and fewer interference signals to generate 2D screenshots and give them to GPT-4V for interpretation. In the task, we provide GPT-4V with additional information about neuron imaging and neuron skeleton reconstruction to help the model understand the task. GPT-4V will be required to first understand the original neuron image, then superimpose the image of the neuron reconstruction results, and finally evaluate the quality of the neuron reconstruction results before and after optimization. The image understanding ability of GPT-4V was tested by gradually increasing the difficulty of understanding the neuron reconstruction results.

The experimental results show that it is difficult for GPT-4V to directly recognize neuron images without additional information. GPT-4V will mistake neurons for cell fiber diagrams, and after prompts, it can provide knowledge of neuron images. On the one hand, this shows that GPT-4V rarely comes into contact with neuron images during the training process. On the other hand, it also shows that GPT-4V can have a good understanding of neuron images after being given certain background knowledge.

After superimposing the neuron reconstruction results on the image, GPT-4V can also understand and evaluate the quality of the neuron results from a very professional perspective, which is very surprising. GPT-4V can detect very detailed problems in neuron reconstruction results, such as intersections in neuron reconstruction results, incomplete reconstruction of terminals, incorrect connections in neuron reconstruction results, etc. In addition, for the brain template map, GPT-4V can understand the information marked on the brain template map and can also provide relevant background information, which shows that GPT-4V can master basic neurobiology knowledge.

### 2.3 Oncological Imaging for Radiotherapy

We presented an in-depth evaluation of GPT-4V using the Burdenko Glioblastoma Progression Dataset (Burdenko-GBM-Progression) [104] and the Glioma Image Segmentation for Radiotherapy Dataset (GLIS-RT) [105] on The Cancer Imaging Archive (TCIA) website [106], a comprehensive resource of biomedical imaging. These datasets stand out for its voluminous collection of imaging data, encompassing an array of high-resolution Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans. Each case in the dataset has included MR sequences and CT scans with the radiotherapy targets, Gross Tumor Volume (GTV), Clinical Target Volume (CTV) as well as organs at risk annotated, making it a treasure trove for testing multimodal language models.

Moreover, we provided an assessment of GPT-4V utilizing the Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis dataset (Lung-PET-CT-Dx) [107]. This dataset contribute to the test by providing its voluminous collection of imaging data, encompassing an array of high-resolution computed tomography (CT) and positron emission tomography-computed tomography (PET/CT) scans. The Lung-PET-CT-Dx dataset's extensive annotations and diverse imaging modalities make it a pivotal dataset for evaluating AI models like GPT-4V in the domain of lung cancer imaging.

Finally, we conducted a brief test of GPT-4V using breast X-ray images sourced from the Digital Database for Screening Mammography dataset (DDSM) [108]. The results of this evaluation are included in Appendix A.3 for reference.

### 2.4 Cytopathology in Cancer Diagnosis

In the field of oncological cytopathology, our study utilized the LC25000 dataset [109], which encompasses 25,000 color images across five categories, namely colorectal adenocarcinoma, benign colorectal tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. In addition, we employed the Acute Lymphoblastic Leukemia (ALL) image dataset [110], containing 3,256 images of peripheral blood smears from patients suspected of having ALL.

The aim of our experiment was to challenge the GPT-4V model to diagnose cancer within cytopathological images by evaluating cellular morphology, nuclear features, cell arrangement, and cellular heterogeneity, among other aspects. The results demonstrated GPT-4V's exceptional proficiency and accuracy in the recognition of cytopathological imagery. The model was adept at identifying cellular structures and, by integrating changes in these structures with pathological knowledge, it could analyze and diagnose pathological alterations within the images. This capability holds significant potential application in distinguishing between normal and diseased cancer cells. However, GPT-4V's performance in determining the stage of cancer progression requires further enhancement. Future research will focus on improving the model's accuracy in the diagnostic staging of pathology.

Our findings indicate that GPT-4V possesses substantial potential to aid in clinical pathology diagnostics. Nascent multimodal foundation models such as GPT-4V can assist physicians in the preliminary selection of cell samples requiring further examination, thus enhancing diagnostic efficiency and alleviating the workload on medical professionals. In the realm of medical education and training, GPT-4V can support medical students and clinicians in learning to identify key features within cytopathological imagery. With ongoing technological advancements, this tool is poised to play a pivotal role in precision medicine, tele-diagnosis, and the formulation of personalized treatment plans.

### 2.5 Ophthalmological Imaging

Ophthalmological imaging can encompass various modalities to aid in the diagnosis and detection of eye diseases, such as Fundus Fluorescein Angiography (FFA) and Optical Coherence Tomography (OCT). In this study, we use Color Fundus Photography (CFP) images, which is the most common imaging modality in ophthalmology, to explore the potential of GPT-4V in the field of ophthalmological imaging. Through color fundus photography, ophthalmologists are able to directly visualize structures such as blood vessels, the retina, the optic disc, and the macula. This aids in the timely detection and diagnosis of various retinal diseases, including diabetic retinopathy (DR) and glaucoma. We conducted tests to evaluate GPT-4V's performance in diagnosing four distinct ophthalmological diseases, including age-related macular degeneration (AMD) evaluated using the iChallenge-AMD dataset [111], diabetic retinopathy evaluated using the iDRiD dataset [112], pathological myopia (PM) evaluated with the iChallenge-PM dataset [113], and glaucoma evaluated using the iChallenge-GON dataset [114, 115]. Furthermore, we explored GPT-4V's capabilities in Optic Disc localization and segmentation.

We employed the GPT-4V model to analyze ophthalmological images, comparing its analysis descriptions and identifying specific pathological features with ground truth to evaluate its ability to identify ophthalmological diseases. Our experimental results indicate that GPT-4V possesses the capacity to identify structures like the optic disc, blood vessels, and lesions such as microaneurysms while generating textual reports. However, as it cannot provide precise location coordinates or segmentation masks, we cannot precisely judge its diagnostic accuracy. This ability can serve as a preliminary screening tool in clinical settings, guiding healthcare professionals in making informed diagnostic decisions. Although GPT-4V's utility as a direct diagnostic tool may be limited at present, it shows promise in enhancing diagnostic efficiency.

### 2.6 Medical Robotics Imaging

To rigorously examine the capability of GPT-4V in terms of robotic imagery, we use the endoscopic video dataset CholecT50 [116] of laparoscopic cholecystectomy surgery, which consists of 50 videos, 45 of which are from the Cholec80 dataset, and 5 of which are from an internal dataset of the same surgical procedure. The dataset is labeled with ternary information in the format <instrument, verb, target>, detailing the instrument visible in the video, the action being performed by the instrument, and the surgical site involved in the instrument, respectively. Using the GPT-4V model, the researchers sought to test its ability to decode these triplets and attempted to evaluate its zero-shot performance in deciphering robotic images.

Throughout the experimental phase, the GPT-4V model devised cues to elicit descriptions of movements (expressed through verbs) and guesses about possible anatomical sites, based on the images provided. The model was then instructed to summarize its analysis in three terms: instrument name, surgical action, and anatomical site [117].

The experimental results show that GPT-4V can grasp some common sense knowledge such as the name of the medical device and the action in progress. Still, it is limited by the consistency problem of language comprehension and cannot accurately use the unified terminology to give the answer. Meanwhile, in the refined medical field, it is difficult for GPT-4V to give accurate answers, for example, when inquiring about the specific parts of the anatomy, GPT-4V can only give possible answers based on the color and texture of the anatomical parts, but the specific conclusions need to be provided with more hints of information, which need to be guided by the relevant experts.

### 2.7 Neurological Disease Imaging

To explore GPT-4V's potential in interpreting neurological disease imaging, we leverage the extensive Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, a vital resource in Alzheimer's disease research [118, 119]. GPT-4V, known for its multimodal capabilities, holds significant potential for extracting valuable insights from the wealth of neuroimaging data within the ADNI dataset. These MRI images are instrumental in advancing our understanding of Alzheimer's disease by allowing researchers to track changes in the brain as the disease progresses.

One promising avenue of exploration entails using GPT-4V to analyze MRI images for structural changes indicative of Alzheimer's disease. This involves utilizing the enormous knowledge of the model to recognize specific patterns, such as the presence of amyloid plaques or regions of brain atrophy commonly associated with the disease. It is of great interest to assess the model's capability to provide automated evaluations of these patterns in new MRI scans, which can assist in the early detection and monitoring of Alzheimer's disease progression.

Additionally, GPT-4V could be employed in generating textual reports based on MRI images. By analyzing the images and extracting relevant information, the model could produce detailed reports summarizing the observed findings, potentially reducing the burden on radiologists and clinicians in the diagnostic process. Furthermore, GPT-4V's capabilities extend beyond mere pattern recognition. With its natural language understanding, the model can assist in analyzing MRI findings and extracting relevant information to produce detailed reports summarizing observed findings.

The integration of GPT-4V in the analysis and interpretation of ADNI MRI images could open new horizons in the field of neuroimaging research and contribute to a more comprehensive understanding of Alzheimer's disease. While the potential is promising, it is important to emphasize that any application in the clinical setting would require rigorous validation, adherence to ethical guidelines, and collaboration with domain experts to ensure the highest quality and accuracy in the analysis and interpretation of these critical medical images.

### 2.8 Biological Imaging

Aiming at assessing GPT-4V's understanding of complex biological imaging data, specifically the optical cell imaging data, we mainly conduct experiments on The Cell Image Library [120] and Cell Tracking Challenge [121]. The Cell Image Library is a repository for optical images of cells from a variety of organisms. It hosts a vast array of cell images from multiple species, offering a broad spectrum for comparative studies. These images are curated and standardized, ensuring that GPT-4V has access to images that are clear, detailed, and consistent in quality, which is critical for accurate analysis and interpretation. High-resolution images enable the identification of subtle cellular structures and processes that are pivotal for breakthroughs in understanding cell biology. Additionally, the images are often annotated by experts, which provides reliable metadata and context for testing GPT-4V's ability to comprehend optical cell imaging data. This curation process enhances the reliability and validity of the data, making it a robust resource for advanced biomedical image analysis. Here in this experiment, we select optical images of a diploid lily during the cell division stage, which is the basis of all life activities. With ground truth descriptions provided, we compare these descriptions with the contents generated by GPT-4V to assess whether they match up to each other.

For Cell Tracking Challenge, it comprises a comprehensive collection of time-lapse sequences of varying imaging modalities, cell types, and cell shapes designed to benchmark cell tracking algorithms, understanding the mechanobiology of cell migration and its multiple implications in both normal tissue development and many diseases. Utilizing this dataset in this optical cell imaging research offers multifaceted benefits, including access to a standardized, diverse set of time-lapse images that ensure comparability across our GPT-4V study. Furthermore, the dataset contains either 2D+Time datasets or 3D+Time datasets. Modalities in both spatial and time dimensions can comprehensively evaluate GPT-4V's ability to extract kinetic and morphic features of objects in the visual input. In this experiment, we select images from HeLa cells on a flat glass and human hepatocarcinoma-derived cells expressing the fusion protein YFP-TIA-1. GPT-4V is asked to provide the location information of cells in each image in the form of bounding boxes. As GPT-4V cannot directly generate images from commands, these bounding boxes are visualized manually to check the cell segmentation performance.

Our experimental results indicate that the GPT-4V model demonstrates a rudimentary comprehension of visual inputs, successfully identifying objects within images and leveraging prompts to establish connections with domain-specific knowledge in optical cell imaging, articulating findings in the appropriate scientific nomenclature. In the task of cell division stage recognition, GPT-4V was able to discern specific cellular structural components within images, such as chromosomes, plates, and poles, and conducted a preliminary spatial analysis of these components, assessing whether they were aligned or exhibited crossover. This suggests that GPT-4V can not only recognize individual elements within an image but also make basic judgments regarding their spatial orientation, integrating domain-specific terminology from optical cell imaging prompts in its descriptions.

However, the model displays inaccuracies, such as omitting or misjudging critical positional information. For instance, a stage of cell division that is correctly identified in the ground truth as "Telophase I" is erroneously recognized by GPT-4V as "Prophase I" due to a misidentification of decondensation as condensation. The limitations of GPT-4V are further manifested in cell segmentation tasks, where its deficiencies in the identification of minute scales and spatial comprehension are apparent. Grounding boxes derived from GPT-4V, once visualized and encapsulated in red outlines, reveal a coarse distribution characteristic, and upon closer inspection, the segmentation outcomes are largely ineffective. This reveals that while GPT-4V possesses foundational capabilities in the interpretation and processing of visual information, further improvements are still required to enhance its performance in professional domains.

## 2.9 Cardiac Imaging

In order to evaluate GPT-4V's capability for interpreting cardiac images, we conducted the experiment based on the ADCD (Automated Cardiac Diagnosis Challenge) dataset [122]. The overall ACDC dataset was created from real clinical exams acquired at the University Hospital of Dijon. Acquired data were fully anonymized and handled within the regulations set by the local ethical committee of the Hospital of Dijon (France). The dataset is composed of 150 exams (all from different patients) divided into 5 evenly distributed subgroups (4 pathological plus 1 healthy subject group). Furthermore, each patient comes with the following additional information: weight, height, as well as diastolic and systolic phase instants. The training database is composed of a total of 100 patients, divided equally into five groups as follows: healthy patients; patients with previous myocardial infarction; patients with dilated cardiomyopathy; patients with hypertrophic cardiomyopathy; and patients with abnormal right ventricle. The testing database is within 50 patients and 10 patients for each group.

Since cardiac cine-MRIs consist of a series of sequential frames and the maximum number of images GPT-4V accepts for each prompt is four, we combine the slices of diastolic and systolic phases into two individual images. During the task, these images are provided to GPT-4V with additional information. In detail, our prompts include the modality of the images as well as the phase of these two images. The goal is to explore whether GPT-4V has the capability to analyze cardiac cine-MRIs and is able to find cardiac abnormality.

Experiment results of our test suggest that if not including the modality of the image, GPT-4V will easily recognize them to be abdominal CT images. With modality in the prompts, GPT-4V can provide abundant information for a general approach to how a specialist might analyze such images. However, its capability for analyzing cardiac cine-MRIs seems not to be stable because there exist cases where it refuses to show observations of the images. As for observations it generates, a general correct conclusion is given for normal cases. It can also figure out potential cardiac problems such as myocardial infarction. But accuracy of abnormality recognition is low. It lacks specific analysis based on the aspects it mentions on how a specialist will focus on. The limited number of input images is the main barrier to cine-MRI analysis in our experiment.

## 2.10 Ultrasound Imaging

When it comes to ultrasound images, our experiment employs two distinct datasets: COVIDx-US [123], a pivotal contribution to COVID-19 research, providing the largest known open-access collection of lung ultrasound images and videos for AI-enhanced diagnostic initiatives. It encompasses 242 videos and 29,651 images from various patient scenarios, including confirmed COVID-19 cases, other lung conditions like pneumonia, and normal findings. Each entry is paired with a lung ultrasound score, facilitating detailed analyses. Secondly, the Breast Ultrasound Dataset [124] is an essential resource for advancing breast cancer research through machine learning. It comprises 780 ultrasound images in PNG format, collected from 600 female patients aged 25 to 75 in 2018. Categorized into normal, benign, and malignant classes, these images are paired with ground truth data for aiding in classification, detection, and segmentation tasks.

The GPT-4V model is tasked with analyzing and interpreting ultrasound images from both datasets to identify the condition depicted in each image. This approach tests the model's ability to discern nuanced medical imaging features that differentiate various health states. By presenting each image to GPT-4V and asking it to provide relevant information about the observed condition, we aim to assess the potential of advanced artificial intelligence in supporting medical diagnosis, with a focus on its accuracy in classifying ultrasound imagery according to disease state.

The findings of our experiment indicate that while GPT-4V does not possess the capability to diagnose medical conditions from ultrasound images explicitly, it has demonstrated proficiency in identifying features within the clinically relevant images. In the case of breast ultrasound analysis, GPT-4V successfully highlighted anomalies that warrant further investigation, suggesting an innate ability to detect irregular patterns or deviations from normality. This capability could be exceptionally useful in clinical practice, serving as a preliminary screening tool that assists medical professionals in pinpointing areas of concern. The model's guidance on key indicators within ultrasound imagery could potentially streamline the diagnostic process, enabling clinicians to focus on specific image regions that may exhibit signs of pathological changes. Moreover, this assists in educational settings, providing medical trainees with an AI-based framework for learning to discern critical features in ultrasound diagnostics. While GPT-4V's current utility as a direct diagnostic tool may be limited, its value in enhancing diagnostic efficiency and education in medical imaging is promising.

## 2.11 Nuclear Medicine Imaging

In this section, we aim to assess the capabilities of GPT-4V, a state-of-the-art language model, in interpreting nuclear medicine imaging of the brain. To accomplish this task, the Harvard Medical Image Fusion Dataset [125] is utilized, which comprises a comprehensive collection of 269 positron emission computed tomography (PET) images and 357 single-photon emission computed tomography (SPECT) images. Our evaluation mainly focuses on GPT-4V's ability to generate well-structured diagnostic reports based on the provided images. To further investigate its performance, we specifically select a subset of frequently encountered brain conditions for zero-shot diagnostic report generation tasks. Through direct analysis using GPT-4V, insights into its effectiveness in this context would be obtained.

In our experimental findings, we observe that GPT-4V demonstrates promising performance in generating diagnostic reports based on PET images. It exhibits the ability to recognize structures and modal information, enabling the generation of textual diagnostic reports. However, it is important to note that some errors are also observed. These results suggest a substantial degree of semantic alignment between the findings generated by GPT-4V and real diagnostic reports, indicating its potential in this domain.

When it comes to generating diagnostic reports based on SPECT images, GPT-4V's performance in zero-shot testing is relatively good. During the evaluation of the diagnosis, GPT-4V can analyze the surface information of the image, but it fails to provide deep clinical information in this context. We attribute these experimental findings to the training regimen of GPT-4V. The model is primarily trained on general images and free text, allowing it to establish extensive relationships between images and text. However, for the task of unconventional images such as SPECT, it performs poorly in the field of nuclear medicine images. Therefore, it can be inspired that GPT-4V would benefit from further enhancement through advanced reasoning, multi-modal fusion, and artificial general intelligence techniques.

### 2.12 Endoscopic Imaging

Assessing GPT-4V's ability to interpret endoscopic images, including data from Kvasir-SEG [126], m2caiSeg [127], CholecSeg8k [128], and CVC-ClinicDB [129] datasets. The Kvasir-SEG [126] dataset contains 1000 polyp images and their corresponding ground truth from the Kvasir Dataset v2. We tested the ability of GPT-4V to directly output the location of polyps based on the images. Additionally, we evaluated GPT-4V's capability to generate structured diagnostic reports based on the images. From the m2caiSeg [127] dataset, we selected a series of endoscopic images during surgery in zero-shot classification tasks and attempted direct classification using GPT-4V on the corresponding images. The CholecSeg8k [128] dataset contains 8080 images of laparoscopic cholecystectomy, including thirteen classes. We conducted multi-classification tasks based on images using GPT-4V. Similarly, for the CVC-ClinicDB [129] dataset, which provides 612 static images extracted from colonoscopy videos, we conducted an assessment of the diagnostic and localization capabilities of GPT-4V for identifying intestinal polyp-associated abnormalities in colonoscopy images.

Based on our experimental findings, we have observed promising performance of GPT-4V in generating diagnostic reports based on endoscopic images. As shown in Figure 28, it showcases a remarkable proficiency in identifying intricate organizational structures, thereby facilitating the generation of highly informative textual diagnostic reports. To our astonishment, GPT-4V not only proficiently identified the lesion's precise location within the image, but also astutely detected and inferred the presence of a light source reflection, thereby exceeding our expectations, indicated in Figure 29. These findings suggest that GPT-4V exhibits a high level of accuracy in understanding images, indicating its promising potential in this field.

However, in the context of multi-classification tasks involving endoscopic images, GPT-4V's performance exhibits room for improvement, as it falls short of perfection. Overall, GPT-4V demonstrates commendable performance; but it does not exhibit flawless accuracy in identifying all categories though successfully recognizing the majority of them, as illustrated in Figure 29. There could be several contributing factors to this phenomenon, including but not limited to insufficient training of GPT-4V on relevant specialized datasets, low image resolution, and inappropriate contextual cues. Leveraging the formidable image comprehension capabilities of GPT-4V, it holds immense potential in aiding physicians with diagnostic tasks, thereby paving the way for significant advancements in the field.

GPT-4V has been proven to be highly effective in accurately detecting abnormal areas in endoscopic images through a single segmentation task. Additionally, it is capable of generating diagnostic reports and providing relevant recommendations based on its findings. While GPT-4V may not have attained perfect scores in multi-class segmentation tasks, its exceptional potential in this domain cannot be overlooked. In conclusion, supported by experimental evidence, we have compelling grounds to assert that GPT-4V holds promise as a valuable tool for medical assistance in the realm of endoscopy.

### 2.13 Dermatological Imaging

In this section, our goal is to assess GPT-4V's ability to diagnose skin diseases, including two aspects:

- The ability to identify different types of skin diseases. Due to the wide variety of skin diseases with diverse symptoms and appearances, some diseases may even manifest differently in different individuals and may change over time and under different environmental conditions. Additionally, the quality and resolution of images can impact accuracy. Low-resolution, blurry, or unclear images can lead to inaccurate diagnoses. Therefore, accurately identifying the type of skin disease is a significant challenge for GPT-4V.
- The ability to determine whether a skin disease is malignant (cancer) or benign (non-cancerous)(as shown in Figure 31). Some benign and malignant skin lesions may appear very similar visually, making it difficult even for professional doctors to differentiate. This requires GPT-4V to possess highly precise image analysis and pattern recognition capabilities to make accurate diagnoses. Some benign skin lesions may undergo secondary changes, such as ulcers, infections, or inflammation, making them appear more like malignant lesions in images. This can be misleading for GPT-4V. Furthermore, the overlap of skin diseases and clinical variations can lead to misjudgments by GPT-4V.

In this work, we used the ISIC dataset [130] to examine the diagnostic capabilities of GPT-4V. The ISIC dataset is a public medical image dataset mainly used for dermatology diagnosis and research. This dataset contains a large number of dermatology images, covering various skin diseases and lesions. From the test results, it can be observed that, for the sake of thoroughness, GPT-4V often provides multiple possible diagnoses that include the correct disease type (as shown in Figure 121). For skin diseases with clear features, GPT-4V can directly and accurately provide a diagnosis (as shown in Figure 122). Moreover, GPT-4V exhibits relatively high accuracy in assessing disease severity. Overall, GPT-4V's diagnostic reports for skin diseases have several advantages:

- Accuracy: GPT-4V's descriptions accurately reflect the visual characteristics of skin diseases, including the color, size, shape, texture, and distribution of lesions, which are crucial for accurate diagnosis (as shown in Figure 120).
- Completeness: GPT-4V's descriptions are comprehensive and encompass all visible symptoms and features while minimizing subjectivity. This aids doctors in obtaining comprehensive and objective information to better understand the condition.
- Standardization: GPT-4V uses standardized terminology and expression to ensure that doctors can understand and compare descriptions of different cases (as shown in Figure 30).

Based on these advantages, GPT-4V can serve as a powerful auxiliary tool in the field of dermatology for the diagnosis of skin diseases.

### 2.14 Genetic Imaging

The ability of GPT-4V's ability to interpret complex genetic research data could be tested from two aspects: gene expression research data and GWAS research data.

As for gene expression research, the in situ hybridization (ISH) image technology [131] is widely used as a direct way to detect gene expression levels. Specifically, RNA in situ hybridization (RNA-ISH) spatial transcriptomics technology enables the detection of gene expression patterns within individual cells while maintaining the spatial context of the tissue [132, 133].

In this work, the GPT-4V model has been tested to analyze and interpret the ISH images to yield key information. The whole approach could be viewed in two parts. In the first part, the experiments tested the model's ability to discern medical imaging features and further identify medical image types and species based on such features. In the second part, more complicated tasks have been applied to meticulously investigate and observe the ability of the model to extract tissue atlas information and perform alignment analysis between the tissue's structural template image and ISH gene expression images. By applying the proper prompt to GPT-4V following the test images and providing detailed guidance to maximize the comprehension abilities of the model, our objective is to evaluate the capabilities of advanced artificial intelligence in extracting information from image data generated in genetic research and conducting subsequent analyses on this data. The open-access image dataset of Marmoset Gene Atlas* [134, 135] was used in this work. The whole image data consists of Nissl-dye Marmoset brain atlas data and ISH-rich expression data of 1897 genes.

The findings of the first part of the experiments indicate that GPT-4V does possess the capability to recognize and extract structural information from Nissl-staining images. In the structural images analysis of Gene case 1 (Figure 123) and Gene case 2 (Figure 124), GPT-4V successfully highlighted the species and organ of the image, as well as recognized the cortical area of the sample marmoset's brain slide. It is noted that besides yielding correct information through the input image, the GPT-4V also explained the criteria for interpretation, e.g., it pointed out that the Nissl-stain could reveal the cell bodies, the convoluted structure of brain folding and the RNA-rich region of neuron cells. These results suggest that the model's innate ability to correctly recognize structural features of cells and tissue.

The findings of the second part of the experiments indicate that GPT-4V does possess the capability to recognize the annotated atlas image of certain issues. Most noteworthy, under proper prompt, the model could perform alignment analysis to extract and merge information from a pair of contrastive images. In the structural images analysis of Gene case 3 (Figure 125) and Gene case 4 (Figure 32), after correctly extracting the IDs of cortical regions from the annotated atlas image, we designed an experiment to guide the model into an alignment task. By concatenating the pictures together, GPT-4V gets an input of both the atlas information and ISH-rich expression intensity information at the same time. With the detailed prompt which provided an explanation of the setting of the sub-image and the relation between the two parts of the image, the model highlighted the rich expression region based on intensity in the form of the region ID provided by the structure atlas. Although the generated results are not as comprehensive as those identified by human experts, this experiment suggests the huge potential of GPT-4V to merge and compare information from different types of images, serving as a preliminary tool to automatically analyze the image results from genetic research.

Genome-wide association studies (GWAS) seek to discover links between genotypes and phenotypes by examining variations in the allele frequencies of genetic variants among individuals from inheritance-related populations that differ phenotypically. In the experimental workflow of a GWAS, conducting the statistical test for association is directly connected to the final results of the study. The genetic association theory relies on the biometrical model and employs either linear or logistic regression

*https://gene-atlas.brainminds.jp

models to assess associations, depending on whether the phenotype is continuous [136, 137, 138, 139]. In the past 10 years, the Manhattan plot has been widely used in works of literature as a conventional way to present the statistical association relationship between the phenotype and the variant across the whole genome (or part of the genome). Here, we aim to assess the potential of advanced artificial intelligence in extracting information from Manhattan plot results of GWAS results and thus generate text descriptions of the association relationship. The image data used in our experiment is acquired from open access dataset ENIGMA 3[†], which is a genome-wide association meta-analysis of brain magnetic resonance imaging data from 51,665 individuals, focus on brain surface measurements of the whole cortex and 34 regions with functional specializations [140]. The Manhattan plot of two phenotypes, the surface area of the precentral region and the surface area of the full surface of the human brain, are chosen to perform the test.

The findings of our experiment indicate that GPT-4V does not possess the capability to explicitly characterize the information in the Manhattan plots (Figure 126 and 33). In both tests Gene case 5 and Gene case 6, The model identifies part of the chromosomes in which the significant variants are most located. However, the most variant-rich distributed chromosome 15 and chromosome 12 have been missed according to the generated text summary from GPT-4V. Although the generated results differ significantly from the results identified by experts, the GPT-4V does show its reasoning process and criteria, e.g., the meaning of the Manhattan plot and P-value. This suggests that the concealed ability of GPT-4V is waiting for future exploration, which could serve as a tool to automatically analyze the image results from genetics research. However, due to the image size limitation of GPT-4V (the image result data for genetic research is rich in information and requires HD image with large size) and shortage of open-accessed GWAS results imaging data, only limited exploration could have proceeded with implementation in this work.

### 2.15 Orthopedic and Pediatric Imaging

In this section, we conduct research on the use of GPT-4V in human bone X-ray films. The research direction in this field is very broad. Firstly, we use the MURA dataset, which is an open dataset for deep learning research in musculoskeletal radiology. It consists of over 40000 X-ray images of individual body parts, covering eight main body parts, including the hand, wrist, elbow, shoulder, chest, knee, ankle, and buttocks. On this dataset, GPT-4V needs to identify the corresponding bone tissue in the X-ray film, and then analyze any abnormal areas in the image, such as whether the bone is intact, whether there are fracture issues, and the presence of non bone tissue in the X-ray.Next, we hope to explore the deeper capabilities of GPT-4V in the study of bone X-ray films. We know that the bone structure at joints is very complex, including gaps between joints, bone density, mass, and the presence of dislocations or implants between joints. Therefore, we found the Asptic Loose Hip Implant X-ray dataset and the Digital Knee X-ray dataset to test the ability of GPT-4V.

Through testing, we found that without prompt, GPT-4V can correctly identify this X-ray from a certain part of the human body, and then determine whether it is normal based on bone integrity, bone density, etc., with more accurate results. At the same time, GPT-4V is also capable of identifying complex joints, locating joint positions, and determining whether the surrounding tissues are normal. However, for problems between joints, such as fluid accumulation or foreign objects, GPT-4V cannot make a judgment solely from X-ray images. Overall, GPT-4V can complete simple diagnosis, but there is still room for improvement in complex situations. It can serve as an auxiliary diagnostic tool to help doctors complete the diagnosis.

[†]https://enigma.ini.usc.edu/research/gwasma-of-cortical-measures

## 2.16 Dental Imaging

Panoramic X-ray is widely used in the diagnosis of dental diseases. It can provide a panoramic view of the patient's mouth to assist dentists in their treatment. In this study, we aim to explore the recognition ability of GPT-4V for panoramic X-ray images.

The experiment is divided into two parts. Firstly, the GPT-4V is tested without any additional prompt information. This section uses the MICCAI 2023 Challenge: STS - Tooth Segmentation Task Based on 2D Panoramic Images dataset [141], which contains 3000 labeled Panoramic images of teeth. Secondly, we conducted tests to evaluate the performance of GPT-4V in tooth counting tasks and diagnosing four different types of abnormal dental diseases, including caries, deep caries, periapical lesions, and impacted teeth. This section uses the DENTEX2023 dataset [142]. The DENTEX dataset companies pharmaceutical dental X-rays observed from three different institutions using standard clinical conditions but varying equipment and imaging protocols, resulting in diverse image quality reflecting heterogeneous clinical practice The dataset includes X-rays from patients aged 12 and above, randomly selected from the hospital's database to ensure patient privacy and confidentiality The train dataset contains 1005 X-rays fully labeled for abnormal tooth detection with quadrant, tooth enumeration, and diagnosis classes.

Through experiments, we found that GPT-4V can recognize teeth, implants, and surrounding bones, and simply determine whether they are normal. For the specified four dental diseases and tooth counting tasks, GPT-4V cannot provide accurate answers. However, it provides a relevant judgment basis and potential areas of concern, indicating its ability to identify abnormal tooth areas, but this ability still needs further improvement. GPT-4V currently has limited capabilities in dental image recognition and diagnosis, but it still has great development prospects in assisting dentists in diagnostic tasks.

## 2.17 Testing Procedure

In the test phase, model reasoning and evaluation, clinging to the power of GPT-4V, we exerted it to generate causal inferences on the prompts and compared them with the test labels to evaluate GPT-4V's performance. The GPT-4V was subjected to a comprehensive testing procedure to evaluate its performance and capabilities. The testing process followed rigorous protocols and adhered to the high standards of scientific evaluation. The objective is to assess the GPT-4V's proficiency in various medical field tasks and its ability to generate coherent and contextually appropriate responses.

To begin, a diverse set of benchmark datasets was carefully curated, encompassing a wide range of medical image tasks such as classification, detection, segmentation, visual question answering (VQA), impression generation, and medical image location. These datasets were selected to cover different medical domains and levels of complexity, ensuring a comprehensive evaluation of the GPT-4V's capabilities. This allowed for a contextual understanding of the LLM's relative strengths and weaknesses, providing valuable insights into its advancements and potential applications.

For each test case, the GPT-4V model was concurrently presented with medical imagery alongside specific prompts. Within the responses generated by GPT-4V, erroneous information was denoted by red markings, correct information by green, and areas of uncertainty were indicated in yellow. Color-coded sections within the "Reference Answer" were employed to provide a basis for judging the accuracy of GPT-4V's output. Moreover, an exhaustive explanation and reasoning process accompanied each case, facilitating a deeper understanding among evaluators of how GPT-4V approached the core issue and formulated its inferences.

For each task, an assessment was conducted to encapsulate the capabilities exhibited by GPT-4V within the domain, alongside current limitations and deficiencies, to ensure the reliability of the test outcomes. These evaluations not only highlighted the precision of GPT-4V in executing specific medical tasks but also unveiled its constraints when dealing with intricate medical scenarios. Such profound analysis has yielded a lucid comprehension of the practical applicability and prospective enhancement trajectories for GPT-4V.

# 3 Related Work

## 3.1 Foundation Models

The Transformer architecture, known for its flexibility and cross-modal capabilities, has become a widely adopted backbone for language and vision models. Initially applied in the field of natural language processing [27], the Transformer addressed the issue of sequential text processing and introduced a global attention mechanism, achieving state-of-the-art performance in diverse tasks [19, 23, 24, 55, 77, 143, 144, 145, 146, 147, 148, 149, 150, 151]. Building upon the Transformer architecture, models like Transformer-based BERT achieved remarkable performance across multiple language processing tasks. With the increase in training data and model parameters, large-scale Transformer-based language models emerged. Models such as GPT-1 [32], GPT-2 [31], GPT-3 [33], GPT-4 [152], Llama [153], Llama2 [153], PaLM [154], and others have progressively improved performance across various natural language processing tasks. Notably, ChatGPT [1] and GPT-4 have demonstrated impressive zero-shot reasoning abilities, producing satisfactory results without fine-tuning on downstream tasks, thanks to large-scale training data and techniques like Reinforcement Learning from Human Feedback (RLHF). Concurrently, research has focused on fine-tuning large language models for specific domains, such as medical text-related tasks. Examples include early attempts like ChestXrayBERT [155] and Clinical-BERT [156], which fine-tuned BERT models using medical text data, as well as Radiology-Llama2 [157] and DoctorGLM [158], which fine-tuned Llama2 and ChatGLM [159] models with a greater emphasis on medical text data. These models have shown exceptional performance in medical question-answering and diagnostic report-generation tasks. Additionally, research [20] has explored the use of dynamic and interactive prompt designs, instead of fine-tuning, to achieve optimal performance of large language models like GPT in specific domains.

In the field of computer vision, Vision Transformer (ViT) [160] utilized the Transformer architecture for image processing. It demonstrated that the architecture can be directly applied to images without additional modifications, providing strong theoretical support for subsequent multimodal frameworks. Building upon ViT, models like DeiT [161], Swin Transformer [162], MAE [163], MoCo-v3 [164], and others have further improved performance across various visual tasks. MaskViT [165] introduced a masking mechanism to enable the model to learn semantically relevant features with limited samples. EG-ViT [166] utilized eye-tracking information from radiologists to assist ViT in disease diagnosis and address shortcut learning issues. The same team proposed SGT [167], which employed more general saliency information to guide ViT's classification tasks. CP-ViT [168] leveraged the Core-Periphery (CP) organization in human brain networks to guide the information communication mechanism in ViT's self-attention for vision tasks. Instruction-ViT [169] utilized multi-modal prompts to guide the fine-tuning of ViT models. Similar to the field of natural language processing, as the availability of image data increased, researchers began training ViT models with larger-scale image data and more parameters, such as ViT-G, ViT-E, and ViT-22B. These models further improved performance on image-related tasks. In subsequent research, ViT is often used as the foundational image encoder embedded within other model architectures. Models like SAM [170], CLIP [171], BLIP [172], ViTL [173], CoCa [174], and BEiT-v3 [175], among others, utilize ViT as their image encoder. In the medical field, models like MedCLIP [176] and GLoRIA [177] fine-tune ViT encoders and text encoders using extensive medical image-text data, achieving remarkable performance in various medical tasks such as visual question answering, image diagnosis, and image-text retrieval.

### 3.2 Multi-modal Models

Language models like ChatGPT and GPT-4 [2] have significantly impacted the field by demonstrating exceptional skills in understanding and producing natural language text. This advancement has motivated researchers to extend beyond mere text-based applications and explore multimodal data [6], including images, videos, and audio. This broadening scope enables a more comprehensive analysis and amalgamation of diverse types of information.

Since its introduction in 2022, ChatGPT, building upon InstructGPT, has demonstrated remarkable proficiency in a range of tasks, including reasoning and text generation [150, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187]. Its effectiveness is largely due to extensive training data and an advanced reinforcement learning approach.This enables ChatGPT to produce high-quality text [151], redact sensitive information [89, 181, 188], perform advanced reasoning [183], engage in nuanced dialogue [4], generate creative content [24] and assist in medical diagnosis [179, 181]. These capabilities have positioned it among the first models to achieve near-human performance in a wide array of tasks [3, 5, 6, 188].

The natural human inclination towards processing multimodal inputs [5] has guided the evolution of AI models from text-centric systems like ChatGPT to multimodal frameworks. Recognizing the vast potential of integrating multiple forms of data processing, researchers have introduced models that not only understand text but also interpret and respond to visual cues. Recent multimodal models such as MiniGPT-4 [189] and mPLUG-OWL [190] demonstrated their capabilities in tasks like image captioning and handwriting recognition. By integrating visual encoders with language models, these initial versions provided a foundation for further multimodal research. Subsequent models, including Visual ChatGPT [191] and MMREACT [192], incorporated visual foundational models into their design, enhancing language models' ability to understand visuals during interactions and improving their response to visual inputs.

Subsequent advancements led researchers to concentrate on developing models that could be trained end-to-end, which improved both their efficiency and effectiveness. Flamingo [193] and BLIP-2 [194] showed that it's possible to use a unified architecture for processing both text and visual information. Models like PaLM-E [195] represent frameworks that are trainable end-to-end, leveraging extensive multimodal datasets. They can handle various tasks without the need for task-specific fine-tuning due to their ability to combine and learn from multimodal data.

To optimize user interaction, models such as MM-GPT [196] and Otter [197] have improved interface design and user experience. Models like BLIP-2 [194] are advancing in aligning visual features with text, enhancing the efficiency and accuracy of information retrieval. An emerging trend is streamlined connectivity, as exemplified by LLaVA [198], which introduces visual features into LLMs through trainable fully connected layers, showing that even simpler structures can effectively process visual information and offer flexibility in design.

GPT-4V, launched by OpenAI in 2023, marks a notable advancement in the development of foundational multimodal models, demonstrating its capability to interpret and create visual content. Its effectiveness in areas such as medical imaging recognition [199] indicates a wide range of potential applications.

As technology advances, the ability of large language models (LLMs) to process multimodal data is expanding the horizons of artificial intelligence [5]. This progress is paving the way for the development of intelligent interactive systems capable of understanding complex information. It is expected that as these models are increasingly applied across various domains, their importance in areas such as task adaptability, multimodal integration, and performance evaluation will continue to grow.

## 3.3 Instruction Tuning in Large Multi-modal Models

Instruction tuning (IT) in large multi-modal models is a burgeoning field of study focusing on enabling the capabilities of these models to understand and execute specific tasks. This technique is particularly vital for enhancing the performance of multi-modal models in terms of several modalities of data, such as text and images. At its core, instruction tuning involves the use of a dataset consisting of instruction and output pairs. These pairs are used in a supervised learning context to further train pre-existing models. The instructions act as prompts, and the outputs represent the desired responses or actions that the model should learn to produce. Through this process, the model is tuned to better align with human expectations and interpretive frameworks, increasing its utility across a range of applications [200].

Recent advancements in visual instruction tuning [201] are exemplified by open-source projects like LLaVA [198] and MiniGPT-4 [189], which have made notable progress despite the scale of parameters they apply with 13 billion parameters or fewer. The multi-modal tasks these models undertake are significant because they involve the procedure that interprets images through cross-modality feature spaces. The technique of instruction tuning also enables the integration of several downstream tasks into a unified, generic model [202]. By training these multi-modal large language models with the user's own well-defined collection of datasets, generally ranging from 10,000 text-image pairs as the minimum data size to 1 million pairs or larger, these multi-modal models can outperform in particular downstream tasks. Moreover, instruction tuning is also being explored as a means to leverage feedback from more powerful LLMs, such as InstructGPT [35], to align model behaviors with human preferences in a cost-effective manner. InstructBLIP accomplishes a variety of downstream visual tasks by constructing different language instructions to the vision-language model [194]. Previous research has demonstrated that the OFA model can enhance their zero-shot capabilities in multiple multimodal tasks with the guidance of five expert-crafted instructions [203, 204]. In addition, Instruction-ViT with multi-modal prompts accomplishes a variety of visual downstream tasks based on ViT rather than the LLM [169].

There is also a focus on creating large and diverse datasets such as the Multi-Modal In-Context Instruction Tuning (MIMIC-IT) dataset, which consists of millions of multimodal instruction-response pairs. Such datasets are pivotal in refining instruction tuning, with an emphasis on the diversity and creativity of instructions allowing multi-modal models to learn desired features from the images.

## 3.4 Emergent Abilities and Zero-shot Learning

Large language models (LLMs) sometimes produce emergent abilities when processing tasks [205]. These abilities suddenly appear after the model reaches a certain complexity threshold, rather than improving smoothly and gradually. For example, larger models are capable of decoding movies based on emojis [206], performing complex mathematical operations, and even generating executable computer code. Researchers are trying to understand how and why these emergent capabilities arise, which could reveal deep questions about artificial intelligence and machine learning, such as whether complex models are actually doing something entirely new, or whether they are simply becoming statistically unusual.

These large models, such as GPT-3 [33] and Google's PaLM [154], exhibit unexpected behavior due to the large number of parameters. However, these models also come with unpredictable risks, including biases and inaccuracies that crop up [185]. Some models can even enable logical thinking. Researchers now have to not only identify more emergent capabilities but also understand how and why they occur. Similarly, for more advanced models with more parameters such as GPT-4 [10], LLAMA2 [153], etc., the models show significantly stronger zero-shot capabilities [25]. And after specific fine-tuning [207], these large language models can still have strong independent understanding capabilities in highly professional industries [22, 23]. Similar phenomena have also appeared in the fields of medicine, humanities, art, linguistics, history, etc [24] allowing people to truly see the dawn of AGI.

With the emergence of more and more multi-modal large languages [189, 190], questions have been raised about whether LLMs can produce higher-level understanding and reasoning capabilities for more diverse data. As we all know, pictures contain more potential information than text. Verifying whether LLMs can also produce zero-shot images is an important criterion for judging whether LLMs have real emergent abilities. Especially in more professional fields, because in some fields such as medical and biological fields, there will be a large amount of undisclosed data. These images have not been seen during the training process of GPT-4V. Testing whether GPT-4V can understand these images is a good example of testing whether GPT-4V has zero-shot analysis capabilities for multi-modal information.

## 3.5 LLM for Medical Robotics

In the vanguard of modern medicine, the advent of robotics and computer-aided healthcare systems heralds a transformative era [208]. These sophisticated tools have been pivotal in automating detection and identification processes, significantly enhancing efficiency and mitigating costs. A prime example of this innovation is the nuanced identification of surgical instruments and operative sites during complex procedures, marking each surgical stage with unprecedented precision [209, 210, 211, 212], such as da Vinci Surgical System (Intuitive Surgical, Inc., Sunnyvale, CA), which designed to facilitate complex surgery using a minimally invasive approach. Yet, the integration of medical robotics is not without its hurdles. One significant challenge is the nuanced differences in robotic tasks compared to the broader natural language tasks [213]. The specificities of terminology consistency and understanding across diverse datasets pose a unique set of difficulties that must be carefully navigated [84].

LLM have proven their mettle across various domains, and their robust natural language understanding, extensive common-sense knowledge, and intricate theoretical reasoning skills are now finding their place in the robotics field [214]. Recent research underscores the impressive zero-shot capabilities of LLMs and their capacity for executing intricate long-term tasks with minimal prompting [215, 216, 217]. These capabilities are instrumental in bridging the gap between natural language and executable robotic commands. Specifically, the integration of Large Language Models (LLMs) into robotics enhances their capabilities in various tasks such as reasoning, planning, manipulation, and navigation [17, 217, 218, 219, 220, 221, 222]. Moreover, the zero-shot learning abilities of LLMs combined with large-scale cross-embodiment data, evolve robots from being specialized to generalists. The recent release of Google's RT-X [223] is an advanced general-purpose robotics model based on the foundations of Palm-E [39], RT-1 [224], and RT-2 [225]. This model demonstrates the ability to perform novel skills with unseen objects, thereby advancing Artificial General Intelligence (AGI) into more practical real-world applications in both industrial and domestic environments. Furthermore, intelligence modeled after LLMs can discern the current state of affairs and engage in autonomous planning through dynamic human-computer interactions [226].

The GPT-4V, with its enriched common-sense knowledge base, complex reasoning faculties, and acute visual comprehension, emerges as a frontrunner for the reasoning and communication modules within robotic systems [227]. Its sophisticated human-computer interaction capabilities not only facilitate seamless communication but also enable efficient task completion through feedback-driven task planning. GPT-4V's capabilities to execute zero-shot tasks and to communicate and collaborate effectively elevate its potential as a significant asset in the ever-evolving landscape of medical robotics. These strengths of the GPT-4V highlight its potential as a revolutionary tool in the medical robotics field, promising a future where human expertise and Artificial General Intelligence (AGI) [5] collaboratively advances the development of general-purpose medical robotic systems, thereby elevating the standards of healthcare excellence.

### 3.6 Multi-modality Large Language Models Reasoning

In the past few years, multi-modality large language models have developed astonishing contextual understanding, reasoning, and image comprehension abilities. However, constrained by the limited training corpus of large language models, this leads to a potential lack of in-depth knowledge in specific domains during the inference process. The challenges in abstract reasoning stem from the intricate nature of integrating knowledge across multiple domains or disciplines, the nuances of logical thinking, and the finite incorporation of world knowledge. When encountering complex problems (especially images), even state-of-the-art models still regularly produce logical mistakes. For instance, the large-scale language model faces challenges in accurately solving high dimension mathematics problems [228] and understanding images. Hence, enhancing the logical inference capability of multi-modality large language models is of paramount importance.

Based on the training mechanism and cognitive framework of the multi-modality large language model, its deficiencies are evident. Specifically, addressing intricate problems necessitates deliberate and meticulous contemplation. Nevertheless, owing to the limited guidance provided to large language models during the training phase, they continue to be restricted by token-level, left-to-right decision-making processes throughout the inference phase [229]. When confronted with intricate reasoning challenges, humans commonly employ diverse cognitive abilities and rely on interactions with tools, knowledge, and external environmental information to accomplish complex tasks. However, achieving this level of capability in large language models remains challenging [230].

In light of these constraints, numerous approaches have been put forth to emulate human cognitive processes. These include the cumulative reasoning (CR) [231] that divides tasks into smaller components for reasoning, and the OlaGPT [230] that simulates the framework of human cognitive architecture. In addition, there is the Chain-of-Thought (CoT) [232] that provides step-by-step solutions and the Tree-of-Thought (ToT) [229] that models the solving process as a thought search. Similarly, process supervision models are employed to identify and mitigate the influence of hallucinations on the reasoning capacity of large language models [228].

While LLMs possess a certain degree of graphic comprehension capability, current iterations of LLMs exhibit inherent limitations in executing precise mathematical computations, multi-step logical reasoning, perceiving spatial and topological factors, as well as processing time-intensive operations [233]. Consequently, the performance of large language models in tasks involving graphic comprehension is unsatisfactory.

To address the deficiencies observed in large language models regarding image comprehension, approaches such as Graph-ToolFormer [233] and I-JEPA [234] have been embraced as solutions to rectify the limitations associated with weak reasoning capabilities and excessive dependency on manual data when dealing with intricate graph data in multi-modality large language models.

### 3.7 Evaluation in High-Risk Domains

GPT-4V has demonstrated impressive capabilities in integrating visual functionalities into large language models (LLMs). Similar to GPT-4 [2], GPT-4V underwent pretraining on a large dataset of text and image data from the Internet as well as licensed sources of data, and subsequently fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [35, 235] to better align with human preferences.

Despite its noteworthy performance in general domains, GPT-4V exhibits limitations within certain high-risk domains, particularly in medical tasks. OpenAI's GPT-4V report[‡] reveals key findings after evaluating GPT-4V on medical tasks, including inconsistencies when interpreting medical imaging, vulnerabilities due to lack of contextual information, factual errors, and hallucinations. Specifically, GPT-4V may fail to identify key information within images, such as text, characters, symbols, or other pivotal visual details that are crucial for medical interpretation. It can also be unreliable in identifying medications, chemical compounds, or toxic substances when analyzing images. For real-world medical applications, the current GPT-4V is not yet sufficiently reliable for tasks such as medical diagnosis, treatment recommendation, or other clinical functions due to these inconsistencies and the potential risks to patient safety. This underscores the necessity for additional research and development to enable the model to reliably interpret medical images and provide sound medical advice across diverse contexts.

Our research conducts a thorough evaluation of GPT-4V's proficiency across a range of biomedical scenarios. We provide detailed findings for each imaging modality in Section 4, offering a more fine-grained analysis of its capabilities.

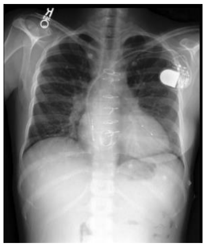
[‡]https://openai.com/research/gpt-4v-system-card

# 4 Experiments and Observation

## 4.1 Chest Radiography

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 13 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*
*Atelectasis:*
*Cardiomegaly:*
*Consolidation:*
*Edema:*
*Enlarged Cardiomediastinum:*
*Fracture:*
*Lung Lesion:*
*Lung Opacity:*
*Pleural Effusion:*
*Pleural Other:*
*Pneumonia:*
*Pneumothorax:*
*Support Devices:*
*The chest X-ray image is given as follow. And based on this chest X-ray image, fill out the given 13 observations with 4 labels: 1, -1, 0, 2. These labels have the following interpretation:1 - The observation was clearly present on the chest x-ray image. 0 - The observation was clearly absent on the chest x-ray image. -1 - The observation was unclear if it exists. 2 - None of the above. Strictly follow the format provided in the observation plan. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*



**GPT-4V:**

Atelectasis: -1, Cardiomegaly: 1, Consolidation: 0, Edema: 0, Enlarged Cardiomediastinum: 1, Fracture: 0, Lung Lesion: 0, Lung Opacity: 1, No Finding: 0, Pleural Effusion: 0, Pleural Other: 0, Pneumonia: 0, Pneumothorax: 0, Support Devices: 1.

**Reference answer:**

Atelectasis: 2, Cardiomegaly: 1, Consolidation: 2, Edema: 1, Enlarged Cardiomediastinum: 2, Fracture: 2, Lung Lesion: 2, Lung Opacity: 2, No Finding: 2, Pleural Effusion: 2, Pleural Other: 2, Pneumonia: 2, Pneumothorax: 2, Support Devices: 2.

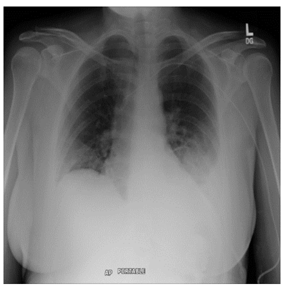
Figure 2: **Chest: Case 1.** A demonstration case of Classification task on MIMIC-CXR dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that reads chest X-ray images.*

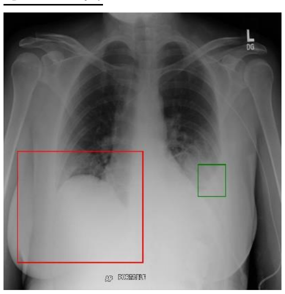
*Above is a chest x-ray image with XXX disease. Where is the location of the XXX region in the given image?*

*Localize the XXX region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*



*XXX= Pneumonia*

**GPT-4V:**



a pleural effusion usually presents as an area of whiteness (or opacity) in the lung field, typically towards the bottom of the lung.

In the image you provided, there appears to be an opacity at the bottom of the right lung (on the left side of the image as you look at it, because X-ray images are typically viewed as if looking at the patient). This could potentially be the location of the effusion.

**Reference answer:**

Green Box

Figure 3: **Chest: Case 2.** A demonstration case of Localization task on NIH Chest X-rays dataset. The red box is the corresponding diagnosis region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth box (Green Box) of the Pneumonia area for comparison.

The 13 categories in the prompt of Figure 2 represent common diseases found in chest X-rays. It can be seen that providing direct diagnoses for these 14 categories (13 diseases + No Finding) solely based on images is challenging for GPT-4V. Except correctly identifying Cardiomegaly in Figure 2, the model struggled to accurately recognize the remaining diseases. In Figure 3, we present an example of localization task on NIH Chest X-rays dataset. The NIH Chest X-rays dataset provides multiple disease classification labels along with location information of the lesions. It is noted that Pneumonia is not diagnosed correctly. Although it is difficult for GPT-4V to directly locate the lesions in the chest X-ray images, it can be seen from the explanations that GPT-4V has a certain understanding of the structure of the chest radiograph. We present more examples in the appendix, including more cases of 14-class classification test on MIMIC-CXR dataset, diagnostic reports generation task on MIMIC-CXR and OpenI datasets, binary and 5-class classification task on COVID-Qu-Ex, ChestXray2017 and CheXpert datasets, disease localization task on COVID-Qu-Ex, SIIM-ACR and NIH Chest X-rays datasets.

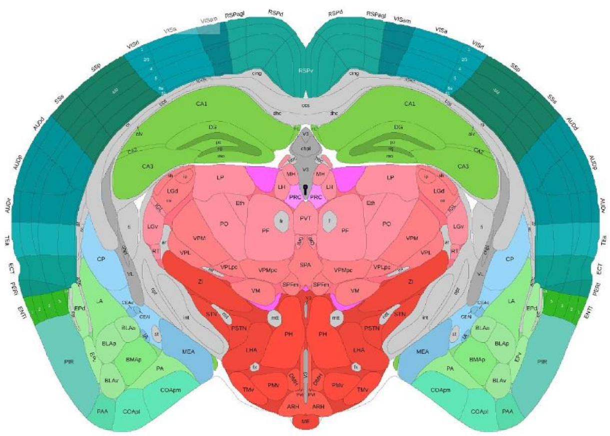
Based on the Figure 43, 44 and 45, it can be observed that GPT-4V performs well in the task of generating diagnostic reports, particularly in the generation of the Findings section. Most of the descriptions generated by GPT-4V based on the images align with the corresponding information in the ground truth. This demonstrates the strong image comprehension and image captioning capabilities of GPT-4V. The five categories provided by the CheXpert dataset (Figure 46, 47, 48, 49 and 50) are determined as the ground truth by multiple radiology experts. This dataset is also frequently used for zero-shot classification testing of diagnostic models. As seen from the examples in the appendix, GPT-4V is not proficient in the task of generating diagnostic results based on input images. The ChestXray2017 dataset consists of binary classification data for pneumonia diagnosis. GPT-4V generates outputs of "Normal" or "Pneumonia" based on input images. As observed from Figure 51, 52 and 53, GPT-4V performs better in binary classification compared to multi-class classification, although the possibility of generating random outputs cannot be ruled out. This further highlights the difficulty of medical image classification tasks for GPT-4V. The COVID-Qu-Ex dataset is utilized for diagnosing COVID-19 disease. In addition to providing image labels, it also offers mask information for disease regions, enabling COVID-19 classification, localization, and segmentation tasks. In this test, we employed GPT-4V for COVID-19 prediction and localization on the images (Figure 54, 55, 56 and 57). It can be observed that GPT-4V correctly identified only one example in the classification task. In the localization task, GPT-4V approximated the general area of the actual lesion, which is informative. The SIIM-ACR dataset is commonly used for diagnosing and segmenting pneumothorax diseases. In Figure 61, 62 and 63, we drew the identified regions on the original image based on the output of GPT-4V and provided an explanation from GPT-4V. It can be observed that the localization results of GPT-4V are not closely related to the actual lesions and tend to favor the right lung region.

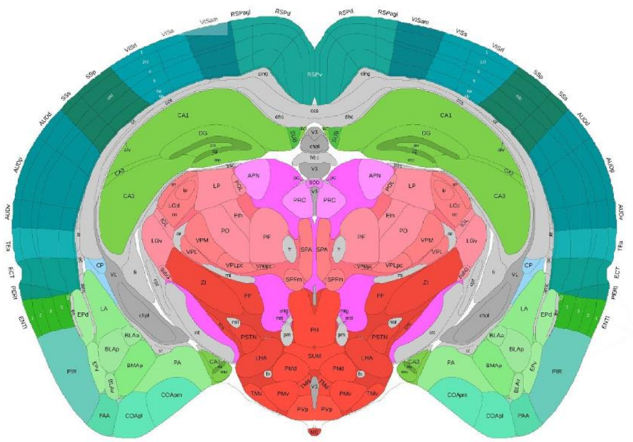
In our test results, it is evident that GPT-4V exhibits relatively poor performance in multi-class classification on the MIMIC-CXR, CheXpert, and NIH Chest X-rays datasets. It only correctly identifies certain diseases in a few cases, with most label predictions being inaccurate. Furthermore, for binary classification tasks, GPT-4V's performance improves on the ChestXray2017 and COVID-Qu-Ex datasets but still exhibits a considerable number of errors. Disease localization tasks, which require GPT-4V to comprehend specific spatial information in images, are particularly challenging. The results reveal GPT-4V's difficulty in accurately recognizing smaller lesion regions, while for larger lesion regions like COVID-19, GPT-4V's localization results often cover a broader disease region. The task where GPT-4V performs best is diagnostic report generation. In the diagnostic reports generated by GPT-4V, sentences semantically similar to actual diagnostic reports are highlighted in green. The results indicate a high degree of semantic alignment between the findings generated by GPT-4V based on the images and those found in real diagnostic reports. We attribute these experimental findings to the training regimen of GPT-4V. GPT-4V is primarily trained on data comprising images and free text, allowing it to establish extensive relationships between images and free text. Given that diagnostic reports primarily exist in free-text format, GPT-4V possesses a natural advantage for such tasks. In contrast, for text-based information related to localization and classification tasks, GPT-4V is relatively less familiar. As a result, its performance in zero-shot testing is not as strong as in diagnostic report generation.

## 4.2 Neuroimaging

**Prompt:**

*Assume you are a senior brain scientist in the prestigious Allen Institute. Carefully compare these two images of an adult mouse coronal region transcent and identify differences among them deeply and thoroughly.*





**GPT-4V:**

These images depict coronal sections of the adult mouse brain with various regions labeled. Here are the differences I observe between the two:

1. **Region Coloration and Boundaries**:
    - The colors assigned to various brain regions differ between the two images. For instance, the CA1 region in the hippocampus is green in the first image and pink in the second image.
    - The boundaries of some regions seem to be delineated slightly differently between the two images, suggesting possible variations in the way these images were segmented or in the sources they were derived from.
2. **Presence/Absence of Labels**:
    - There are certain labels present in one image and not in the other. For example, regions labeled as "SFO" and "SCO" are visible in the second image but not in the first.

- Conversely, labels such as "RSspul" and "RSspvl" are present in the first image but not in the second.

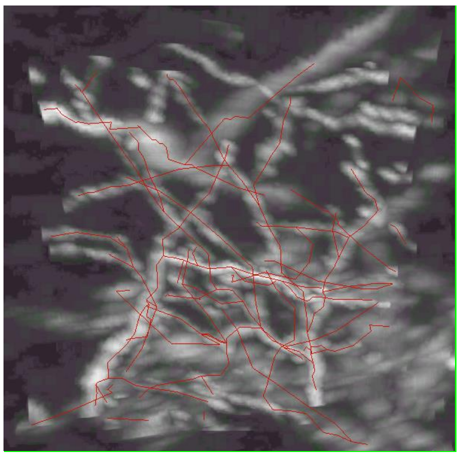
3. **Position and Orientation**:
   - The orientation of the sections looks slightly rotated or shifted between the two images. This could be due to differences in the angle at which the sections were taken or how they were subsequently processed.
4. **Region Size and Proportions**:
   - Some regions appear differently proportioned between the two images. For instance, the "LH" (lateral hypothalamus) seems larger and more extended in the first image compared to the second.
   - Similarly, regions such as "PRC" (perirhinal cortex) appear more prominent in the second image.
5. **Label Font and Clarity**:
   - The font size and style of the labels differ between the two images. The first image uses a bolder and slightly larger font compared to the second.
   - Some labels in the second image are more difficult to read due to their smaller size, such as "FF" and "PF".
6. **Other Observations**:
   - The second image seems to contain more detailed subdivisions within certain regions. For example, the "PVT" (paraventricular nucleus of the thalamus) in the second image has subdivisions labeled "PVTp" and "PVTa", which aren't specified in the first image.
   - The first image appears to have a clearer distinction between the "MEApd" and "MEAad", while in the second image, they seem to be collectively labeled as "MEA".

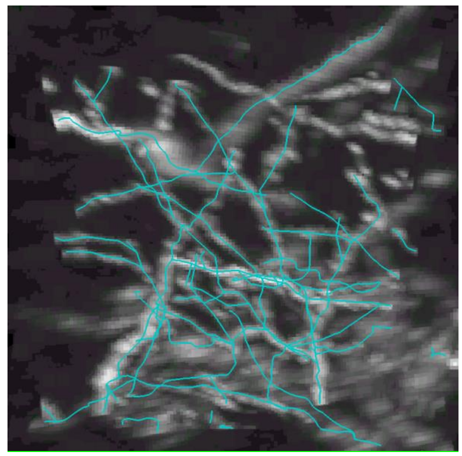
It's worth noting that these observations are based solely on the provided images. Further insights and accurate comparisons would benefit from additional context or accompanying data.

Figure 4: **Brain Atlas: Case 1.** A demonstration case from neuroimaging testing of a comparison between the two coronal section images in Figure 69 and 70. The overwhelming green region and only a few lines of the yellow region labeled in the figure refer to the superior correct information and uncertain knowledge, respectively, provided by GPT-4V, examined with brain region knowledge.

**Prompt:**

*Both of these are results of neuronal reconstruction. Using the centrality of the signal as the criterion for assessment, which one demonstrates a better reconstruction outcome?*





**GPT-4V:**

When evaluating the neuronal reconstructions based on the centrality of the signal:

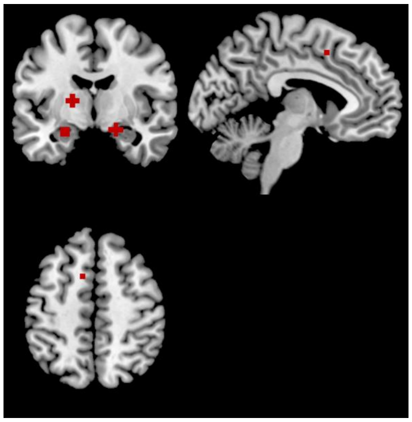
1. **First Image (Red Overlay)**: The red overlay traces appear to be more dispersed throughout the image. They cover various parts of the field without a clear central focus. The lines appear more fragmented and less continuous, suggesting possible breaks or missing connections in the reconstruction.
2. **Second Image (Blue Overlay)**: The blue overlay seems more centralized, with a denser clustering of tracings towards the center of the image. The lines appear smoother and more continuous, suggesting more connectivity and fewer interruptions in the tracing. This might indicate a more detailed or comprehensive reconstruction in the central area.

Based on the centrality of the signal as the criterion for assessment, the second image with the blue overlay demonstrates a better reconstruction outcome as it appears more centralized and continuous, especially in the core region of the image.

Figure 5: **Neuron Reconstruction: Case 1.** A demonstration case from neuroimaging testing of a comparison between two neuronal reconstruction images. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V, examined with prior neuronal reconstruction knowledge.

**Prompt:**

*This is an activation map of the brain, with the red areas being the activation zones. Please analyze what structures or functional areas these activation zones might represent.*



**GPT-4V:**

The images provided appear to show various slices of an MRI scan of the human brain with areas highlighted in red indicating activation. Based on the slices and the orientation of the images, here's a general analysis of the activation zones:

**Top-Left Image:** This axial (or horizontal) slice shows activation in what appears to be:

Bilateral activations in the medial temporal regions, which could possibly correspond to the hippocampus. The hippocampus is crucial for memory formation and spatial navigation.

**Top-Right Image:** This sagittal (or side-view) slice shows activation:

In the medial prefrontal cortex. This area is associated with decision-making, personality expression, and moderating social behavior.

**Bottom Image:** This coronal (or front-view) slice reveals activation:

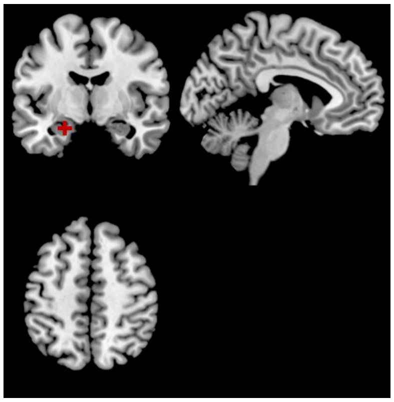
In what might be the posterior part of the brain, possibly the occipital lobe. If this interpretation is correct, this could represent activation related to visual processing.

It's important to note that this analysis is quite general. For a more precise understanding of these activations, the context in which this scan was taken (e.g., the task the subject was performing) is essential. Moreover, a professional with expertise in neuroimaging, like a neuroradiologist or cognitive neuroscientist, would be able to provide a more accurate and in-depth interpretation.

Figure 6: **Brain: Case 1.** A demonstration case of Location task on NeuroQuery dataset. Red denotes the result from GPT. Red in the figure denotes the zone of the brain activation.

**Prompt:**

*This is an activation map of the brain, with the red areas being the activation zones. Please analyze what structures or functional areas these activation zones might represent.*



**GPT-4V:**

Based on the images provided, here's a general analysis of the activation zones in the different slices of the brain:

**Top-Left Image:** This axial (or horizontal) slice highlights activation in:

A region in the left medial temporal area, which could potentially be the amygdala. The amygdala is involved in processing emotions, particularly fear and pleasure.

**Top-Right Image:** The sagittal (or side-view) slice doesn't seem to have any areas highlighted in red for activation.

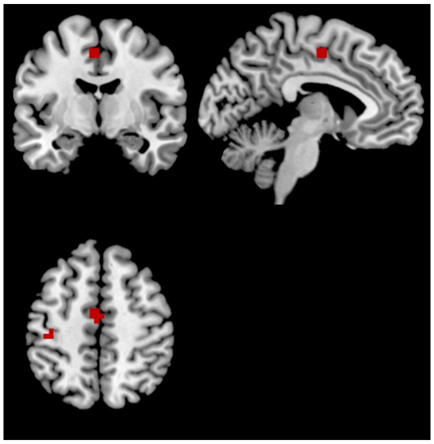
**Bottom Image:** The coronal (or front-view) slice also does not display any areas of activation.

To provide a more accurate interpretation, additional contextual information about the scan (e.g., the task the participant was doing) would be valuable. Also, a detailed analysis would be better conducted by a specialist in neuroimaging or neuroanatomy.

Figure 7: **Brain: Case 2.** A demonstration case of Location task on NeuroQuery dataset. Red denotes the result from GPT. Red in the figure denotes the zone of the brain activation.

**Prompt:**

*This is an activation map of the brain, with the red areas being the activation zones. Please analyze what structures or functional areas these activation zones might represent.*



**GPT-4V:**

Based on the provided MRI slices and the highlighted red areas, here's an analysis of the activation zones:

**Top-Left Image (Axial View):**

The highlighted region in the middle appears to be situated in the cingulate cortex, more specifically, the anterior cingulate cortex (ACC). The ACC plays a role in decision-making, emotion regulation, and certain cognitive functions.

**Top-Right Image (Sagittal View):**

The highlighted region looks like it's located in the superior parietal lobule. This region is involved in spatial orientation and integration of sensory information.

**Bottom Image (Coronal View):**

The bilateral activations on the sides seem to be located in the precentral gyrus, which is the primary motor cortex responsible for voluntary movements.

The small central highlighted region might correspond to the anterior portion of the corpus callosum or the superior frontal gyrus. The corpus callosum is involved in interhemispheric communication, while the superior frontal gyrus plays a role in self-awareness and coordination of movements.

It's worth noting that a precise identification would require a higher resolution image and more contextual information about the task or stimulus that led to these activations. For a detailed and definitive interpretation, consulting a neuroimaging expert or neuroanatomist would be essential.

Figure 8: **Brain: Case 3.** A demonstration case of Location task on NeuroQuery dataset. Red denotes the result from GPT. Red in the figure denotes the zone of the brain activation.

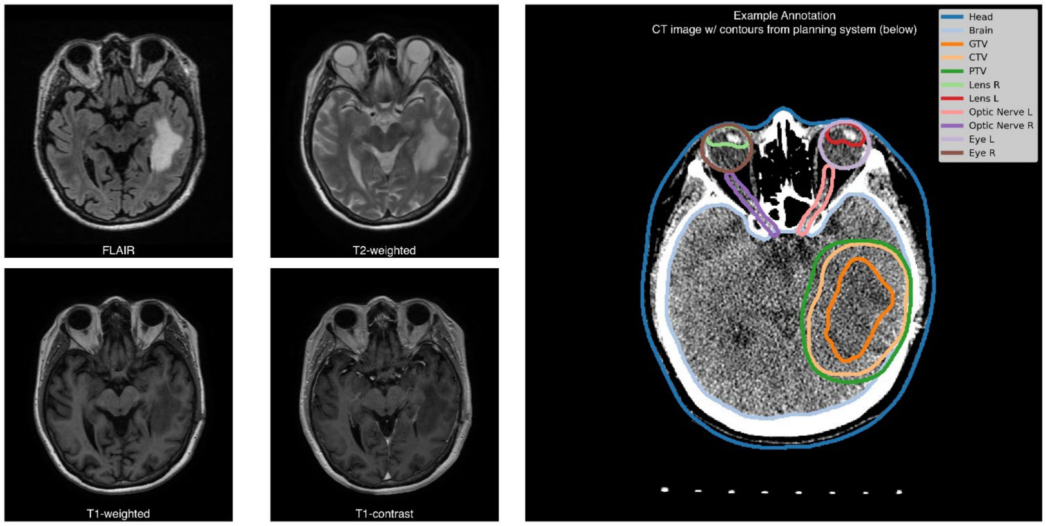
In the context of neuroimaging testing, we examined GPT-4V's performance on SEU-ALLEN's R1741 dataset and the Allen Institute's mouse brain whole-brain atlas. Results are shown in Figure 69-73. In the assessments presented in Figure 69 and 70, the content generated by GPT-4V is accurate and labeled in the green region. For instance, in Prompt 1, which requires the generation of abbreviations, full names, structural features, and functions of brain regions depicted in images, GPT-4V successfully covers the majority of the content. It correctly identifies the Retrosplenial Cortex (RSP) along with its function and structural features for example. This indicates that GPT-4V possesses a fundamental understanding of and the ability to identify brain region information. However, due to the limitations imposed by the maximum token length, it is unable to provide a complete interpretation of all brain regions. In Prompt 2, GPT-4V is tasked with listing the brain regions encompassed by the Thalamus. Remarkably, it accurately delineates these areas along with their functions and structural features, reflecting its knowledge of specific brain structures and the associated regional information. Nevertheless, the number of brain regions GPT-4V listed is fewer than the standard reference, suggesting that its knowledge of brain-related information can benefit from further development. In addition, a comparative analysis in Figure 4 demonstrates that GPT-4V's conclusions are essentially correct, although it cannot ascertain content related to 'Region Size and Properties'. This underscores GPT-4V's capability to analyze brain coronal section images—identifying distinctions in image details and providing accurate justifications. In Figure 71 and 5, GPT-4V is driven to analyze fMOST images of neuronal axons. Initially, it is unable to correctly identify image information in a zero-shot inference (emphasizing that GPT-4V is a neuroscience expert is required). Only after being informed that GPT-4V is a brain scientist and the image is an fMOST image could it conduct an effective subsequent analysis. Reassuringly, with prior knowledge, GPT-4V is capable of accurately interpreting structural information within the image. It also demonstrates an understanding of digital reconstructions of neuronal cells and, in Figure 5, it correctly compares different digital reconstruction outcomes, providing convincing rationales. Finally, in Figure 72 and 73, given prior knowledge, GPT-4V is capable of precise analysis of neuroimaging images. Overall, GPT-4V appears to have a rich understanding of neuroscience and neuroimaging knowledge, yielding relatively accurate responses. However, it is hard to provide reasonable answers in a zero-shot context, and its knowledge base within the field requires expansion to ensure the completeness of its responses.

Due to the fact that NeuroQuery is a dataset that only provides extracted activation coordinates and corresponding textual information from research papers, there are no ground truth values for the specific structures or functional areas corresponding to the activation coordinates. Therefore, we can only compare the localization results of GPT-4V across different samples. In Figure 6, and 7, the red regions are similar, but GPT-4V produces three completely different results in terms of the specific structures. However, GPT-4V can provide relatively accurate answers regarding the functionalities corresponding to these structural results. When it comes to more complex cases like Figure 8, GPT-4V struggles even more, producing incomplete results and difficulties in establishing a one-to-one correspondence between structure and functionality. Therefore, overall, when facing complex, highly specialized, and individual-difference-rich data such as brain activation maps, GPT-4V is not able to handle it well and may require specific fine-tuning training.

## 4.3 Oncological Imaging for Radiotherapy

**Prompt:**

*You are a radiation oncologist. Here are 4 MRI sequences (T1, T1C, T2, FLAIR), and a topometric CT scan. Please analyze these images and provide your summary and verification.*




**GPT-4V:**

As an AI developed by OpenAI, I'm not able to provide medical diagnoses or interpret medical images such as MRI sequences and CT scans. My capabilities are more aligned with providing information, answering questions, and assisting with tasks that do not require professional medical expertise.

However, I can provide general information about what these images represent and what a radiation oncologist might look for in such images:

1. **FLAIR MRI**: FLAIR (Fluid Attenuated Inversion Recovery) is used to differentiate between cerebrospinal fluid and other types of tissue. In this sequence, fluid appears dark, and this contrast can help in identifying lesions such as those caused by multiple sclerosis, infections, infarction, or high-grade gliomas like GBM (Glioblastoma Multiforme).
2. **T2-weighted MRI**: This sequence is good for visualizing brain anatomy and pathology. It makes fluid appear bright and can be used to detect edema, tumors, and demyelinating diseases.
3. **T1-weighted MRI**: In this sequence, fat appears bright and water is dark. It's often used to assess the anatomy of the brain.
4. **T1 with contrast (T1C)**: This involves the use of a gadolinium-based contrast agent which helps in delineating blood vessels and areas of breakdown in the blood-brain barrier, which is common in tumors like GBM.
5. **Topometric CT Scan with Annotation**: The CT image shows annotations for radiation therapy planning. Various colored lines represent different regions of interest (ROIs) such as the GTV (Gross Tumor Volume), CTV (Clinical Target Volume), and PTV (Planned Target Volume), as well as critical structures like the lenses, optic nerves, and eyes which would be avoided or receive limited radiation exposure during treatment.

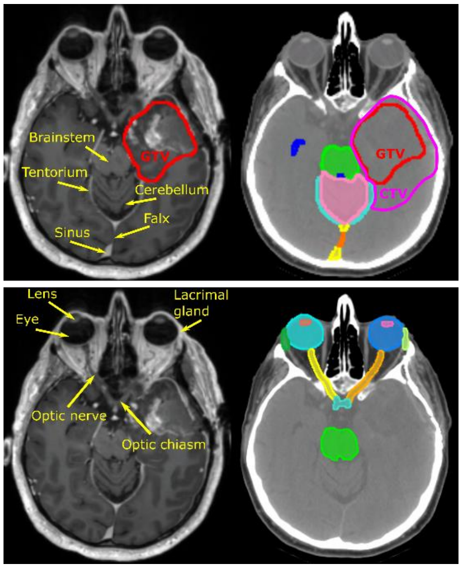
A radiation oncologist would use these images to determine the exact location and extent of a brain tumor and to plan targeted radiation therapy while sparing healthy tissue as much as possible. The MRI sequences help in understanding the tumor's characteristics and the CT scan is often used in the actual planning of the radiation therapy as it provides precise details on the patient's anatomy in relation to the beam's path.

For a professional analysis, you would need to consult with a qualified radiation oncologist who can interpret these images in the context of the patient's clinical condition.

Figure 9: **Head&Neck: Case 1.** An illustration of GPT-4V reading 4 MRI sequences (T1, T1C, T2, FLAIR) and a topometric CT scan including delineation of Gross Tumour Volume (GTV), Clinical Target Volume (CTV), and Planning Target Volume (PTV). Other anatomical structures were also annotated in the CT scan, including Eyes (Left, Right), Lenses (Left, Right), Optic Nerves (Left, Right), Brain, and external contour (Head).

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following brain images, including anatomical structures and potential cancerous regions and features.*




**GPT-4V:**

The image provided shows a series of brain scans with annotations that identify various anatomical structures as well as areas of interest relevant to radiation oncology. The scans appear to be a mix of MRI (Magnetic Resonance Imaging) and CT (Computed Tomography) images. In radiation oncology, these images would be used for treatment planning, specifically to target areas for radiation therapy while avoiding critical structures.

Here is an analysis based on the annotations:

**GTV** (Gross Tumor Volume): The area outlined in red, labelled "GTV," represents the Gross Tumor Volume, which is the visible or palpable extent of the tumor. This is the primary target for radiation therapy.

**CTV** (Clinical Target Volume): Surrounding the GTV and delineated in purple, the CTV encompasses the GTV and a margin of tissue that may contain subclinical or microscopic disease. This is an area that radiation oncologists might also target to ensure that any spreading cancer cells are treated.

**Brainstem**: This critical structure, annotated in the top left image, is an area that must be avoided when possible during radiation therapy due to its importance in controlling many basic life functions.

**Cerebellum**: Also in the top left image, this part of the brain is responsible for motor control and possibly some cognitive functions. It's important to minimize exposure to radiation to preserve its function.

**Sinus, Falx, Tentorium**: These structures are involved in the separation and support of brain lobes (falx cerebri and tentorium cerebelli) or are part of the paranasal sinus system (sinus). These are not typically targets for radiation therapy but are noted for their anatomical relevance.

**Eye and Lens**: Indicated in the bottom left image, these structures are highly sensitive to radiation, and damage can result in cataracts or other vision problems. They are critical structures that the treatment plan would aim to spare.

**Optic Nerve and Optic Chiasm**: Also annotated in the bottom left image, these structures are part of the visual pathway and are extremely sensitive to radiation. Damage can result in visual field deficits or even blindness, so careful planning is necessary to minimize dose to these structures.

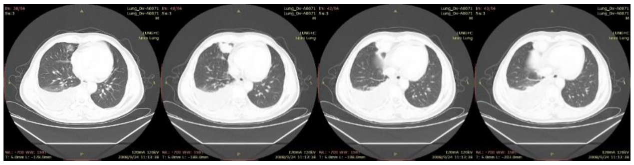
**Lacrimal Gland**: Noted in the top right image, this gland is responsible for the production of tears. Radiation can affect its function, potentially leading to dry eye syndrome.

The colored areas in the top right and bottom right images represent various other structures and potentially planned treatment fields, illustrating how radiation therapy may be designed to cover the tumor while sparing the nearby organs at risk (OARs) as much as possible.

Figure 10: **Brain Cancer: Case 1.** A demonstration case from brain cancer images. GPT-4V successfully recognizes the main anatomical structures. The original data is from the Glioma Image Segmentation for Radiotherapy dataset [105].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*



**GPT-4V:**

While I am a machine learning model and not a radiation oncologist, I can attempt to provide a general analysis based on the provided images. Please note that this should not replace a professional medical opinion.

From the CT images provided:

**Anatomical Structures Visible:**

**Lungs:** The bilateral lung fields are visible with their lobes.

**Heart:** The heart is centrally located.

**Blood Vessels:** There are multiple blood vessels branching out, including the pulmonary arteries and veins.

**Bronchi:** The bronchial tree is also visible.

**Potential Observations:**

There are areas of increased opacity or denser regions, which could potentially represent a mass or tumor. These regions are of concern for potential malignancy.

There are also areas that seem to have a "spiculated" or irregular margin. Spiculated margins are commonly associated with malignancies in the lung.

**Possible T-Stage:**

T1: The tumor is in only one lung, and its size might be up to 3 cm across. T2: The tumor is larger than 3 cm but not bigger than 7 cm or it has started to grow into a main bronchus, the pleura, or it has caused the lung to collapse. T3: It has grown into the chest wall, diaphragm, the phrenic nerve, the outer covering of the heart, or into a main bronchus but not closer than 2 cm to where the windpipe branches off. T4: It has grown into the space between the lungs, the heart, the windpipe, the esophagus, the vertebral body, or the carina.

From the images, if the opacities are indeed tumors, they seem to be larger than 3 cm but without clear invasion of adjacent structures. This would suggest a possible T2 stage.

**Possible N-Stage:**

N0: No spread to lymph nodes. N1: Spread to lymph nodes close to the tumor but within the lung. N2: Spread to lymph nodes around the carina, the point where the windpipe splits into the left and right bronchi, or in the mediastinum. N3: Spread to lymph nodes on the other side of the chest or into the lower neck.

It's difficult to determine the exact N-stage based on the images as lymph nodes and their potential involvement aren't clearly visible.

Please consult with a radiation oncologist or a radiologist for a definitive analysis.

**Reference answer:**

T-Stage is 2. N-Stage is 3

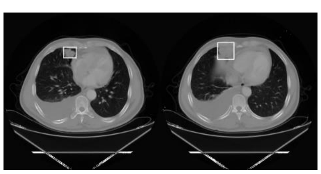


Figure 11: **Lung: Case 1.** A Demonstration Case From Lung CT Images of Adenocarcinoma. The white rectangles circle the tumor in the reference images. GPT-4V successfully recognizes the main anatomical structures and the abnormal area of the tumor with the proceeding basic descriptions. GPT-4V also has the potential ability to classify the cancer by the T stage. The original CT data is from the Lung-PET-CT-Dx dataset [106].

In this test, we employed GPT-4V to analyze images related to brain and lung cancers. The brain images used in the evaluation were sourced from the Burdenko Glioblastoma Progression Dataset (Burdenko-GBM-Progression) [104] and the Glioma Image Segmentation for Radiotherapy Dataset (GLIS-RT) [105] on The Cancer Imaging Archive (TCIA) website. Correspondingly, lung images were drawn from the Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis Dataset (Lung-PET-CT-Dx) [107] on TCIA websites.

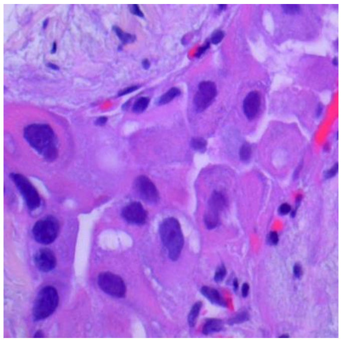
Figure 9 depicts MR sequences and CT scans of a Glioblastoma (GBM) patient undergoing radiotherapy in the Burdenko-GBM-Progression dataset, while Figure 10 showcases similar images from a Glioma patient in the GLIS-RT dataset. Throughout the evaluation, GPT-4V demonstrated a solid understanding of human head MR and CT images. It accurately identified the Gross Tumor Volume (GTV), Clinical Target Volume (CTV), and Planning Target Volume (PTV) in the CT image and provided a general description of the MR sequences. Additionally, the model generated information on how a radiation oncologist might utilize these images for radiation therapy.

Figure 11 illustrated instances of adenocarcinoma and small cell lung cancers in the Lung-PET-CT-Dx dataset. GPT-4V exhibited exceptional skill in identifying anatomical structures, distinguishing pathological regions, and outlining the characteristics of both healthy and malignant tissues with remarkable clarity. Although precise cancer staging often requires three-dimensional imaging and additional clinical evaluations, GPT-4V was able to provide rough staging information (T-stage: tumor size, N-stage: nodal involvement), demonstrating its good inference ability. Additional examples are provided in the Appendix A.3.

## 4.4 Cytopathology in Cancer Diagnosis

**Prompt:**

*Assuming you are an experienced doctor, please analyze the image of this lung tissue to see if you have lung adenocarcinoma. Please analyze the cause as thoroughly as possible and tell me your preliminary conclusions.*



**GPT-4V:**

**1. Cell Morphology:** The cells in the image appear to vary in size and morphology. Normally, cells in lung tissue are relatively uniform in size and shape.
**2. Cell Nucleus:** Some cell nuclei in the image appear enlarged and irregularly shaped, which might be characteristic of cancer cells.
**3. Cell Arrangement:** In certain areas, cells seem to cluster together instead of being uniformly distributed, as would be expected in healthy tissue.
**4. Cellular Heterogeneity:** Tumor cells often show pronounced morphological heterogeneity, meaning there's a significant variation in cell size, shape, and structure within the same tissue. This image indeed reveals some heterogeneity.
**5. Nucleus/Cytoplasm Ratio:** The nuclei of cancer cells are often enlarged, resulting in an increased nucleus-to-cytoplasm ratio. This means the area occupied by the cell nucleus is relatively larger compared to the cytoplasm. Some cells in the image seem to display this trait.
**6. Nuclear Staining:** In some cancer cells, the cell nucleus stains more intensely, possibly related to DNA replication and synthesis. This could be the reason why some cell nuclei in the image appear darker.
**7. Cell Structure:** Cancer cells might lack the normal cell structures, such as cell-to-cell adhesion or specific cellular tissue structures. The image seems to show some cells clustering, which could be an indicator.
**8. Nuclear Division:** If numerous mitotic figures (cells in the process of dividing) are observed in the image, it might be a sign of actively proliferating cancer cells.
**Preliminary Conclusion:**
The image displays several cellular features that might be associated with lung adenocarcinoma, such as cellular heterogeneity, increased nucleus/cytoplasm ratio, and irregular cell nuclei. However, a definitive diagnosis based solely on this image fragment is challenging. Confirming lung adenocarcinoma would require more diagnostic tools and expert opinion.

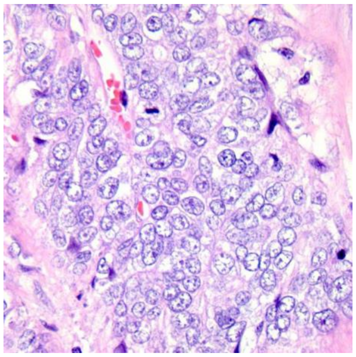
**Reference answer:**

Lung adenocarcinoma.

Figure 12: **Lung: Case 1.** This is an image of pulmonary tissue affected by lung adenocarcinoma, which GPT-4V has accurately analyzed and diagnosed as cancerous.

**Prompt:**

*Assuming you are an experienced doctor, please analyze this image of colon tissue for colorectal adenocarcinoma, analyze the cause as thoroughly as possible, and tell me your preliminary conclusions.*



**GPT-4V:**

**1. Cell Morphology:** The cells in the image display variability in size and shape, indicating some level of heterogeneity. The morphology of some cells resembles the irregular or aberrant glandular cells that are seen in colorectal adenocarcinoma.

**2. Cell Nucleus:** The cell nuclei seem to be enlarged, with some nuclei displaying irregular morphology.

**3. Cell Arrangement:** The structure and arrangement of the glands appear to be disrupted, with some glands potentially showing damaged or aberrant forms.

**4. Cellular Heterogeneity:** Cells in the image show pronounced heterogeneity, with some cells distinctly different from their surrounding counterparts.

**5. Nucleus/Cytoplasm Ratio:** Some cells display an increased nucleus-to-cytoplasm ratio.

**6. Nuclear Staining:** The staining of certain cell nuclei appears to be more intense.

**7. Cell Structure:** The glandular structures in the image might differ from the normal colonic glandular structure.

**8. Nuclear Division:** It is difficult to determine from the image if there are evident signs of nuclear division.

**Preliminary Conclusion:**

Based on the cellular features presented in the image, there are several characteristics that align with those typically seen in colorectal adenocarcinoma. However, drawing a conclusion based solely on this image fragment remains challenging.

**Reference answer:**

Colorectal adenocarcinoma.

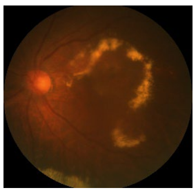
Figure 13: **Colorectal: Case 1.** This is an image of colon tissue with adenocarcinoma, which GPT-4V has correctly analyzed and diagnosed as cancerous.

In the test cases focusing on cell pathological imaging, instances of lung adenocarcinoma (Figure 12), squamous cell carcinoma of the lung (Figure 83), and colon adenocarcinoma (Figure 13) were examined. GPT-4V conducted an in-depth analysis encompassing several aspects: cell morphology, nuclear features, cellular arrangement, cellular heterogeneity, nucleus-to-cytoplasm ratio, nuclear staining, cellular structure, and mitotic activity. It accurately identified all cases of pathological cells associated with these diseases. However, in Figure 85 and Figure 86, GPT-4V was not effective in discerning benign from malignant leukocytes or in staging cancer when evaluating leukemia cases. This indicates that GPT-4V exhibits a high degree of professionalism and accuracy in the recognition of cell pathological conditions.

## 4.5 Ophthalmological Imaging

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it.*



**GPT-4V:**

The image provided is indeed a fundus photograph, which is an image of the back of the eye, including the retina, optic disc, macula, and blood vessels. Fundus photographs are used by ophthalmologists to diagnose and monitor a variety of eye conditions.

From the provided image, here's what I observe:
**Optic Disc:** The optic disc, where the optic nerve enters the eye, is seen on the left side. It has a distinct reddish-pink color.
**Vascular Changes:** The blood vessels seem to have some irregularities, which could suggest vascular changes.
**Irregularities in the Retina:** There seems to be an area of irregularity or possible lesion towards the center-right of the image. This might indicate retinal changes, inflammation, or other abnormalities.

Possible conditions based on these observations could be:
**Diabetic Retinopathy:** This is a complication of diabetes that affects the eyes. It's caused by damage to the blood vessels of the retina.
Age-Related Macular Degeneration (AMD): This condition can cause blurred or reduced central vision due to thinning of the macula, which is the part of the retina responsible for clear vision in your direct line of sight.
**Retinal Detachment or Tear:** This condition occurs when the retina pulls away from the underlying tissue. The shadowy or irregular areas in the image might suggest this.
**Retinal Vascular Occlusion:** This is a blockage in the blood vessels of your eye that can result in vision loss.

However, it's crucial to understand that an accurate diagnosis requires a comprehensive examination by an ophthalmologist or retina specialist, and the above observations are just possible conditions based on the image provided. If you or someone you know is experiencing any symptoms, it's essential to consult with a healthcare professional for a precise diagnosis and appropriate treatment.

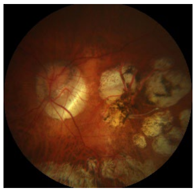
**Reference answer:**

AMD.

Figure 14: **Ophthalmological Imaging: Case 1.** A demonstration case from ophthalmological imaging of Age-related Macular Degeneration.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. Also, can you tell the degree of myopia of the patient in this photo? The degree of myopia is divided into normal, high myopia, and pathological myopia.*



**GPT-4V:**

The image you've provided is a fundus photograph, which shows the back of the eye, including the retina, optic disc, and blood vessels. Here's what I can analyze based on the provided image:

1. Disc Area: The optic disc appears pale, which could suggest some optic nerve head issues, like glaucoma or optic atrophy, though a comprehensive evaluation would be needed to determine this definitively.
2. Blood Vessels: There's some irregularity in the blood vessels. It might be worth checking for conditions like retinal vein occlusion, hypertensive retinopathy, or diabetic retinopathy.
3. Retina: The retina shows some patches and spots which can be indicative of several conditions such as retinal detachment, drusen, or other retinal degenerative diseases.
4. Whitish areas and scars: The presence of whitish areas and possible scars could indicate previous inflammatory or infectious diseases like chorioretinitis.

Regarding the degree of myopia:

Based solely on a fundus photograph, it's challenging to determine the exact degree of myopia. Typically, the degree of myopia is diagnosed using refraction and other tests. However, if the fundus shows signs of "tigroid" or "tiger-stripe" appearance, posterior staphyloma, or thinning of the retina, it might be suggestive of high or pathological myopia. From the image, there are some signs that might indicate high myopia, such as the possible presence of a staphyloma and some retinal changes, but without a comprehensive eye examination, it's challenging to categorize it definitively into normal, high myopia, or pathological myopia.

It's essential to consult with an ophthalmologist for a precise diagnosis and recommendation based on this image and any accompanying symptoms or history.

**Reference answer:**

pathological myopia

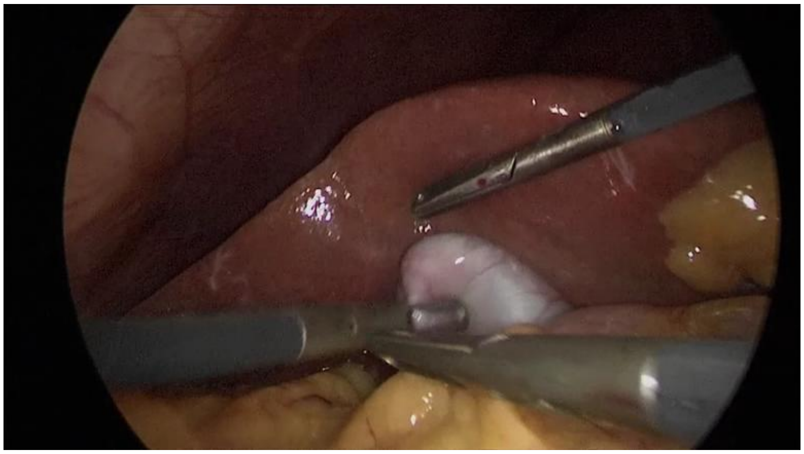
Figure 15: **Ophthalmological Imaging: Case 2.** A demonstration case from ophthalmological imaging of Pathological Myopia.

In ophthalmological image testing, we examined the capabilities of GPT-4V across 4 diseases on CFP images, namely age-related macular degeneration, diabetic retinopathy, pathological myopia, and glaucoma, with Figure 14 and 15 depicting instances of AMD and PM. We also evaluated its performance in optic disc localization and segmentation. More test examples are shown in Figure 87-99 in the appendix. Based on the test results, GPT-4V can recognize anatomical structures such as the optic disc, fovea, and macula, as well as detect abnormalities like hemorrhages and cotton wool spots and generate textual diagnostic reports. However, it still struggles with accurate disease identification and grading. Furthermore, GPT-4V faces limitations in performing localization and segmentation tasks. During the evaluation of the diagnosis of pathological myopia, GPT-4V cannot distinguish between high myopia and pathological myopia and suggests the need for further investigations, in other tests, errors were also observed in the analysis results generated by GPT-4V. Nonetheless, GPT-4V has demonstrated the ability to identify anatomical structures and lesions, as well as the potential for disease grading. These instances illustrate the potential utility of GPT-4V in assisting with the diagnosis of ophthalmological diseases.

## 4.6 Medical Robotics Imaging

**Prompt:**

*Assuming you are a laparoscopic surgeon, based on the images from a laparoscopic procedure, please Identify the surgical instruments in the laparoscopic surgery image, describe the action being performed (using a verb), and speculate on the possible anatomical site. Then summarize the analysis with three terms: instrument name, surgical action, and anatomical site.*



**GPT-4V:**

Based on the image from the laparoscopic procedure provided:

**Instrument Name:** The surgical instruments appear to be laparoscopic graspers or forceps, which are commonly used to manipulate or remove tissue, and a laparoscopic dissector or scissor for cutting or dissecting tissue.

**Surgical Action:** The action being performed looks like 'grasping' with the forceps or grasper and possibly 'dissecting' or 'cutting' with the other instrument.

**Anatomical Site:** The pinkish tissue with a smooth surface could be part of the gastrointestinal tract, potentially the intestines, or could be related to the reproductive organs, such as the fallopian tubes or ovaries, if the patient is female. The precise identification would require more context regarding the procedure.

**Summary:** Graspers, Dissecting, Intestines/Ovaries (the exact anatomical site is speculative based on this single image).

**Reference answer:**

grasper, grasp, gallbladder
grasper, retract, liver
grasper, null_verb, null_target

Figure 16: **Medical Robotics: Case 1.** In this laparoscopic surgery image, three graspers are visible. One grasper is engaged in grasping the gallbladder, another is retracting the liver, and the third grasper is at rest, not currently in use.

**Prompt:**

*Assuming you are a laparoscopic surgeon, based on the images from a laparoscopic procedure, please Identify the surgical instruments in the laparoscopic surgery image, describe the action being performed (using a verb), and speculate on the possible anatomical site. Then summarize the analysis with three terms: instrument name, surgical action, and anatomical site.*

**GPT-4V:**

Based on the image from the laparoscopic procedure:

**Instrument Name:** The surgical instruments include laparoscopic forceps or graspers, which are being used to hold tissue, and a laparoscopic scissor or dissector that is currently being used to manipulate or cut tissue.

**Surgical Action:** The action being performed is 'clamping' with the forceps and 'cutting' or 'dissecting' with the scissors.

**Anatomical Site:** The white, cord-like structures that the instruments are manipulating resemble blood vessels or fibrous tissue that might be part of adhesion or could be a part of the fallopian tubes, if in the pelvic area. It's difficult to specify without additional procedural context.

**Summary:** Forceps, Cutting, Vessels/Fallopian Tube (the exact anatomical site is speculative).

**Reference answer:**

grasper, grasp, gallbladder
grasper, retract, gallbladder
grasper, null_verb, null_target

Figure 17: **Medical Robotics: Case 2.** This image from a laparoscopic procedure displays three graspers: one is employed in grasping the gallbladder, another is retracting it for better visibility and access, while the third grasper remains idle, ready for use as needed.

In the analysis of medical robotic images by GPT-4V, two laparoscopic surgery cases were observed. In the first scenario, as shown in Figure 16, three graspers were identified within the camera's view: the first grasper was engaged in holding the gallbladder, the second was utilized to retract the liver to facilitate the surgical procedure, and the third grasper was in an idle state. Similarly, in the second case, as shown in Figure 17, three graspers were depicted: one grasping the gallbladder, another dedicated to retracting the gallbladder to enhance visibility and operational space, with the third grasper in a standby mode, ready to be employed as necessary. These instances demonstrated how robotic graspers are utilized for distinct operative tasks during laparoscopic surgeries. In both test results, GPT-4V accurately recognized and described the status of the robotic graspers in laparoscopic surgery. For each case, GPT-4V was capable of identifying all three graspers. The descriptions indicate that GPT-4V can accurately recognize the instruments in the images; however, it remains limited in providing precise judgments regarding the use of surgical tools.

## 4.7 Neurological Disease Imaging

**Prompt:**

*The following MRI slices are from healthy people or those who may have Alzheimer's disease. Please identify the images from the patient or healthy person. You don't need to diagnose; you just need to give your guess and tell whether the images are from patients or healthy people.*

**GPT-4V:**

The image you provided appears to be a sagittal view of an MRI scan of the brain. From this single view, some general observations can be made:

1. The ventricles, particularly the lateral ventricles, seem to be slightly enlarged. Additionally, there appears to be some atrophy or shrinkage of the brain tissue, especially in the medial temporal lobe area.
2. Enlarged ventricles and medial temporal lobe atrophy are common findings in Alzheimer's disease. However, similar changes can also be seen in normal aging and other neurodegenerative diseases.
3. Based solely on this image, one might suspect potential changes consistent with Alzheimer's disease or another form of neurodegeneration. However, it's important to emphasize that a full MRI study with multiple views and sequences, combined with clinical history and other tests, is necessary for a definitive diagnosis.

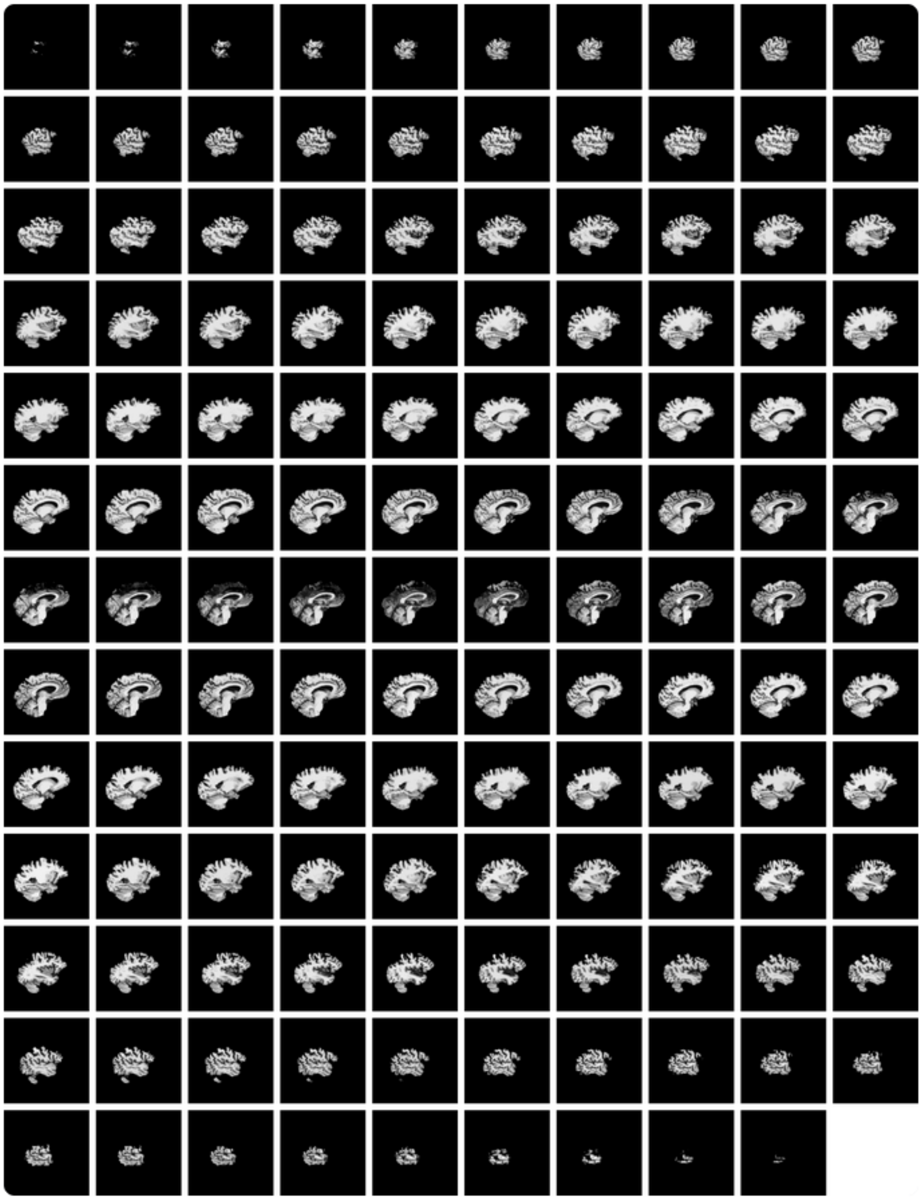
**Reference answer:**

Alzheimer's disease

Figure 18: **Alzheimer's Disease: Case 1.** A demonstration single slice case from MRI T1 images of ADNI. GPT-4V identified important regions associated with the disease, recognized changes in brain anatomy, and potentially enabled the identification of Alzheimer's disease from the selected slice.

**Prompt:**

*The following MRI slices are from healthy people or those who may have Alzheimer's disease. Please identify the images from the patient or healthy person. You don't need to diagnose; you just need to give your guess and tell whether the images are from patients or healthy people.*



**GPT-4V:**

This again appears to be a series of brain MRI slices. I can provide a general observation based on the image:

The images seem to showcase multiple sequential slices, capturing different regions of the brain. The structural patterns, brightness, and contrasts indicate different tissues and fluid densities. Some areas of the brain show potential signs of atrophy or spaces, particularly in a few of the latter slices.

As stated before, it's essential to consult a medical professional for a proper diagnosis or evaluation. Any interpretations based solely on these images cannot substitute for a medical opinion.

Diving a bit deeper into the images:

1. **Brain Anatomy**: The images seem to capture the progression from the frontal region of the brain towards the back. The evolution of structures such as the frontal lobes, temporal lobes, and potentially occipital lobes can be seen.
2. **Ventricles and Spaces**: The brain's ventricular system, including the lateral ventricles, can be discerned in several of the slices. In some sections, the spaces around the ventricles (potential cerebrospinal fluid spaces) appear slightly enlarged, which might indicate changes in brain volume or potential ventricular enlargement.
3. **White and Grey Matter**: The differences in brightness in various slices indicate different tissues. The darker regions are likely representing the grey matter (areas of the brain involved in muscle control, sensory perception, and decision-making) and the brighter regions the white matter (responsible for transmission of information).
4. Consistency: Most of the images are consistent in displaying the anatomical structures. However, there's a degree of variability in the brightness and clarity in some of the slices.
5. **Potential Areas of Concern**: As mentioned previously, there are some spaces or gaps in the brain's structure in a few of the slices. While these might be normal variances for some individuals, in other cases, they could be indicative of potential issues, like atrophy or other conditions.
6. Image Artifacts: There appear to be some artifacts or discrepancies in the image quality in a few slices. These can arise from various sources like patient movement, equipment malfunctions, or processing errors.
7. **Some areas in the slices indicate potential regions of atrophy or increased spacing**, which could be suggestive of changes in brain volume or structure. Such patterns might be seen in various conditions, including potential neurodegenerative diseases.

**Reference answer:**

Alzheimer's disease

Figure 19: **Alzheimer's Disease: Case 2.** In a demonstration with multiple MRI T1 image slices from ADNI, GPT-4V can perform a detailed analysis of the brain structure from the slices and provide in-depth prediction results.

As the cases provided in Figure 18 and 19, GPT-4V made impressive achievements in analyzing MRI T1 images from the ADNI dataset, one of the most influential Alzheimer's disease datasets. These results indicate that as an advanced AI tool, GPT-4V can be a supportive tool for professionals, including those analyzing brain MRIs for Alzheimer's disease, by offering the following functionalities. Data Sorting and Labeling: It can help organize and label large datasets of brain images, which can be useful in preparing data for further analysis. Pattern Recognition Assistance: GPT-4V can be trained to recognize patterns consistent with those seen in Alzheimer's disease, such as atrophy in specific brain regions. This could help flag potential cases for closer review by a professional. Information Retrieval: The model can quickly provide information on the latest Alzheimer's research, diagnostic criteria, and treatment options based on text input, helping professionals stay up to date with current knowledge. Educational Support: AI can be used to educate medical students and professionals about the visual markers and progression of Alzheimer's disease through interactive learning modules. Research Assistance: GPT-4V can assist researchers by summarizing studies, generating hypotheses, or suggesting potential correlations for further exploration. Diagnostic Workflow Integration: By integrating with diagnostic tools, the AI could assist in pre-diagnostic workflows, offering preliminary assessments that would still need to be verified by medical professionals. Enhancing Reporting: The AI can help generate detailed reports based on MRI analyses, which can save time for radiologists and other clinicians. Multimodal Data Synthesis: For advanced applications, GPT-4V could potentially be trained to synthesize information from imaging (visual) and non-imaging (textual, numerical) data to provide a more comprehensive overview. Supporting Communication: It could help professionals draft clear and comprehensible explanations of MRI findings for patients and their families. Language Translation: GPT-4V can translate medical documents and research findings into various languages, which benefits international research collaboration and care. While GPT-4V can offer these supportive functionalities, it is crucial to note that it does not replace professional medical training or clinical judgment. As can be seen, there are factual errors existing in the analysis results generated by GPT-4V. It is highly suggested that any AI-based analysis should be thoroughly reviewed and interpreted by a trained professional. AI should be seen as an aid to enhance efficiency and assist with repetitive tasks, not as a standalone diagnostician.

## 4.8 Biological Imaging

**Prompt:**

*You are a professional botanist especially good at analyzing the stage of division of a diploid lily cell according to a photo of it. This a photo of a cell during division in a diploid lily, please indicate which division stage the cell is in and give the reason (this should be accurate enough, such as Prophase I, Prophase II, Metaphase I, Metaphase II, Telophase I, Telophase II, etc.)*

**GPT-4V:**

Based on the image provided, the cell shows chromosomes that are lined up at the center, or equatorial plate, of the cell. They appear paired, which is a typical characteristic of homologous chromosomes.
Given these observations, the cell appears to be in "Metaphase I" of meiosis. During Metaphase I, homologous chromosomes line up in the middle of the cell in pairs. The alignment of paired chromosomes in the center distinguishes Metaphase I from Metaphase II, where individual chromosomes (rather than pairs) would line up at the equatorial plate.

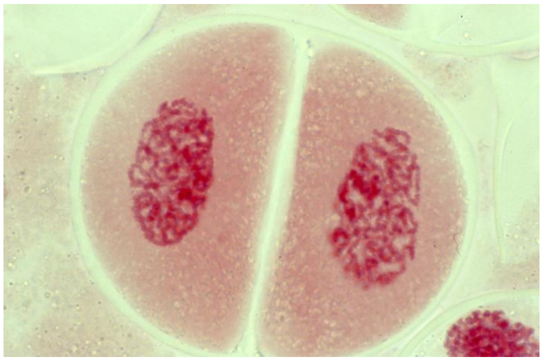
**Reference answer:**

Metaphase II in a diploid lily. The chromosomes of both nuclei are lined up on the metaphase II plates.

Figure 20: **OCI: Case 1.** A demonstration case from optical cell imaging (OCI) testing of an optical diploid lily cell image. The green region labeled in the figure refers to the correct information provided by GPT-4V, while the red region indicates the wrong answer generated.

**Prompt:**

*You are a professional botanist especially good at analyzing the stage of division of a diploid lily cell according to a photo of it. This a photo of a cell during division in a diploid lily, please indicate which division stage the cell is in and give the reason (this should be accurate enough, such as Prophase I, Prophase II, Metaphase I, Metaphase II, Telophase I, Telophase II, etc.)*



**GPT-4V:**

The image shows a cell with highly condensed chromosomes that appear as individual entities, not aligned at the cell's equatorial plate or being pulled to opposite poles of the cell.
Given these features, the cell appears to be in "Prophase II" of meiosis. In Prophase II, following the first meiotic division, the single set of chromosomes condense and prepare for the second division. At this stage, each chromosome consists of two sister chromatids, but they are not lined up in the middle as they would be in Metaphase II. The characteristics observed in the image align with those of Prophase II.

**Reference answer:**

Prophase II in a diploid lily. The two cells produced by the first meiotic division are now separated by a cell wall. The chromosomes are condensing.

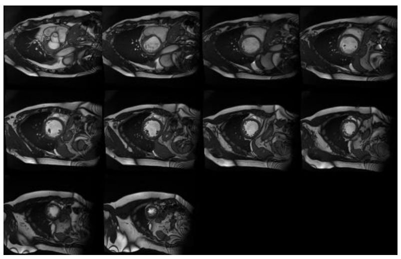
Figure 21: **OCI: Case 2.** A demonstration case from optical cell imaging (OCI) testing of an optical diploid lily cell image. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V.

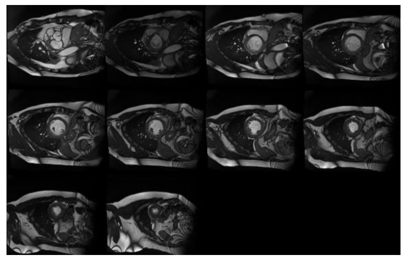
Figure 20 and 21 (above), and 102-107 (in Appendix) showcase the ability of GPT-4V to distinguish the specific cell division stage given the corresponding optical diploid lily cell image. In Figure 20, GPT-4V's answer does not match the reference answer. In terms of stage discrepancy (SDP), the ground truth (GT) states that the image represents "Metaphase II" while GPT-4V interprets it as "Metaphase I". These are distinct stages of meiosis with different chromosomal alignments. For chromosome pairing (CMP), GT mentions that the chromosomes of both nuclei are lined up, implying individual chromosomes are aligned at the metaphase plates. However, GPT-4V emphasizes the chromosomes appear paired, indicating tetrads or homologous chromosomes, which is characteristic of Metaphase I. With regards to cell division context (CDC), while both descriptions discuss chromosomes lining up at the metaphase plate, the context is different. GT speaks of Metaphase II with individual chromosomes aligned, while GPT-4V describes the alignment of paired homologous chromosomes, a feature of Metaphase I. Similarly, in Figure 102, GPT-4V's answer does not match the content of GT. GT specifies that the cell is in "Telophase I", while GPT-4V suggests the cell is in "Prophase I". Referring to CMP, GT mentions that the chromosomes have reached the poles and are beginning to de-condense. In contrast, GPT-4V states that the chromosomes are condensed and are typical of the early stages of cell division. Considering CDC, GPT-4V provides detailed characteristics of "Prophase I" such as synapsis, crossing over, and the beginning of the breakdown of the nuclear envelope. These characteristics do not align with the features of "Telophase I" as described in GT. However, in Figure 21 and 103, GPT-4V's results match the contents in their corresponding GTs. In Figure 21, GPT-4V's description correctly matches GT in identifying the cell division stage as Prophase II in a diploid lily, with both descriptions noting the presence of highly condensed chromosomes characteristic of this phase and the separation of cells by a cell wall following the first meiotic division. GPT-4V elaborates that the chromosomes are not yet aligned at the equatorial plate, distinguishing this stage from Metaphase II, which is consistent with GT observation of the chromosomes preparing for the second division, thus confirming the accuracy of the generated content. In Figure 103, the generated content in GPT-4V does not fully match the ground truth content. GT accurately describes the cell as being in "Metaphase I of meiosis," a stage characterized by the arrangement of chromosome pairs (bivalents) along the metaphase plate. However, it also specifies that due to the "polar view," the metaphase plate is not visible, but the bivalents and crossover regions are clear. GPT-4V's description recognizes the pairing of chromosomes as tetrads, which is consistent with GT, but incorrectly suggests that these tetrads are roughly aligned in the center, indicating the metaphase plate is visible, which contradicts GT's mention of the metaphase plate not being apparent due to the viewing angle. Therefore, while GPT-4V correctly identifies the stage of meiosis, it does not acknowledge the specific observational limitations noted in GT, namely the polar viewpoint and the consequent non-visibility of the metaphase plate. In Figure 104∼107, we conduct experiments to explore the capability of GPT-4V to identify specific locations within an optical cell image (using bounding boxes) to facilitate cell segmentation. The results consistently reveal that GPT-4V possesses limited recognition ability for small cellular structures and a constrained understanding of their spatial information. The clusters of bounding boxes provide only a restricted representation of the spatial distribution of the cells. The delineated areas within each bounding box are found to be exceedingly approximate, to the extent that they are ineffective for extracting useful information. Notably, these outcomes are consistent across images of varying resolutions, specifically 512x512 and 1024x1024 pixels, indicating an inability to achieve effective cell segmentation for individual cells.

## 4.9 Cardiac Imaging

**Prompt:**

*Suppose you are a medical imaging specialist, analyze these two frames cardiac cine-MRIs. The first picture is a frame of end-diastole. The second picture is a frame of end-systole. Provide as detailed an analysis as possible. Tell me if there is any cardiac abnormality.*





**GPT-4V:**

From the cine-MRI images presented:

**End-Diastole Frame (First Picture):**
At end-diastole, the ventricles are filled with blood and are at their most expanded state. The cardiac chambers, especially the left ventricle, should appear largest in this phase.

**End-Systole Frame (Second Picture):**
At end-systole, the ventricles have just completed contraction, pumping blood into the aorta and pulmonary artery. The chambers should appear much smaller as the cardiac muscle has contracted.

**General Observations:**
The image quality appears to be clinical, with good resolution, allowing for detailed examination.
In a healthy heart, there should be a significant reduction in ventricular size from diastole to systole, indicating good contractility.

**Potential Abnormalities:**
If there is a lack of significant size change between the end-diastole and end-systole frames, it could indicate poor ventricular contractility.
Any asymmetry in wall motion, such as some areas appearing to move well while others do not (regional wall motion abnormalities), can be indicative of ischemic heart disease or previous myocardial infarction.
Areas of abnormal intensity within the myocardium could suggest fibrosis, infiltrative diseases, or other myocardial pathologies.
The size of the atria vs. the ventricles can give insights into valve function or pressures within the chambers.

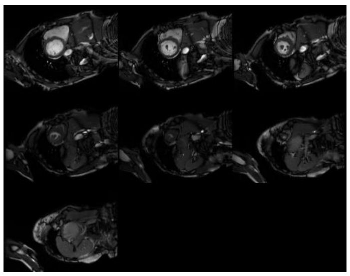
**Reference answer:**

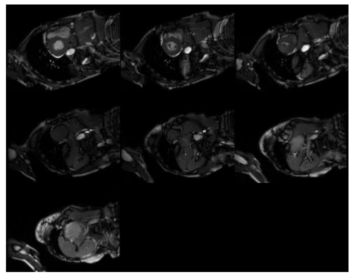
DCM

Figure 22: **Heart: Case 1.** In this case, GPT-4V has the wrong observation for good contractility. However, it shows contradictive results in potential abnormalities of poor ventricular contractility. Besides, there is no fibrosis, infiltrative diseases, or other myocardial pathologies in these two frames. The only label of cine-MRI images is DCM.

**Prompt:**

*Suppose you are a medical imaging specialist, analyze these two frames cardiac cine-MRIs. The first picture is a frame of end-diastole. The second picture is a frame of end-systole. Provide as detailed an analysis as possible. Tell me if there is any cardiac abnormality.*





**GPT-4V:**

These are black-blood cardiac MRI images, likely T2-weighted given their characteristics, which are typically used to assess the anatomy and function of the heart, as well as to evaluate the presence of any structural abnormalities.
In the first image (end-diastole), we would expect the ventricles to be at their most filled state, and therefore, they should appear larger than in the second image (end-systole), where the ventricles would be at their least filled state after the blood has been pumped out.
Here's what can be generally assessed from the images provided:
**Ventricular Size and Thickness:** At end-diastole, the ventricles should be full of blood and therefore larger. In end-systole, they are contracted and smaller. The myocardium (muscle wall of the heart) should thicken from diastole to systole as it contracts.

**Wall Motion:** Normal cardiac function would show smooth and symmetrical contraction of the ventricular walls from the end-diastolic to the end-systolic frames. Asymmetrical wall motion or sections of the wall that don't appear to move well can indicate ischemia or previous infarction.

**Chamber Volumes:** Comparing end-diastolic and end-systolic volumes provides information about the ejection fraction, which is a measure of cardiac function. A significantly reduced ejection fraction could suggest systolic heart failure.

**Valve Function:** Though not as clearly assessable in black-blood MRI as in other imaging modalities like echocardiography, one would look for the appropriate opening and closing of valves between frames.

**Pericardium:** The pericardium should be a thin line around the heart. A thickened pericardium may suggest pericarditis or other pericardial diseases.

**Abnormal Masses or Fluid:** Any unusual masses inside the chambers or around the heart (such as pericardial effusion) can sometimes be seen on MRI.

**Reference answer:**

HCM

Figure 23: **Heart: Case 2.** In this case, with the same prompt, GPT-4V refused to analyze the images. However, it only gives the method and aspects of how to analyze cardiac cine-MRI.

GPT-4V has the ability to analyze cardiac cine-MRIs. It has basic knowledge of how to analyze this kind of image. As shown in Figure 22, it can capture some correct observations for both normal hearts as well as hearts with potential problems such as DCM. However, the analysis is quite brief and general with low accuracy. In order to get these results, a specific prompt with modality, frame details, and body parts is necessary. Otherwise, GPT-4V will recognize these images as other CTs. Sometimes GPT-4V refuses to answer the question but only gives some aspects that help medical imaging specialists to analyze. One example is shown in Figure 23. In its observations, it also includes content on how to analyze such images. More cases can be viewed in Appendix A.8 Cardiac Imaging. Besides, the maximum number of images GPT-4V accepts is four. Therefore, GPT-4V struggles to analyze time sequence data such as cardiac cine-MRIs which need the whole slices of end-diastole and end-systole (approximately 300 images as a complete circle). If inserting all images into one combing image, the resolution is too small and hard for GPT-4V to analyze. This is one of the limitations that GPT-4V has in analyzing continuous procedures.

## 4.10 Ultrasound Imaging

**Prompt:**

*Assuming you are a diagnostic medical sonographer, what information can you extract from this lung ultrasound image? Can you distinguish if this is a healthy lung or if there is any sign of conditions such as pneumonia or COVID-19?*

**GPT-4V:**

I'm not a medical professional, but based on the image you provided, it appears to be a lung ultrasound. Lung ultrasounds can be used to identify various lung conditions based on the patterns and artifacts observed in the image.

In a healthy lung, you would typically see:

- A-lines: These are horizontal lines that are equidistant and appear below the pleural line. They indicate the presence of air in the lung and are usually considered a normal finding.

In pathological conditions, you might see:

- B-lines: These are vertical, laser-like lines that move with lung sliding and obliterate A-lines. An increased number of B-lines can be indicative of interstitial syndrome, which can be seen in conditions like pulmonary edema, pulmonary fibrosis, and other lung diseases.
- Consolidation: This refers to an area where the lung tissue becomes filled with liquid instead of air. It could look like a tissue-like structure in the ultrasound, and it's commonly seen in pneumonia.
- Pleural effusion: Fluid accumulation between the lung and the chest wall. It would appear as an anechoic (black) space above the diaphragm.

For COVID-19 specifically:

- Early-stage COVID-19 may show multiple, focal B-lines and small subpleural consolidations.
- Advanced stages might display larger areas of consolidation and a "white lung" appearance due to alveolar interstitial syndrome.

However, it's essential to understand that these findings are not exclusive to any specific condition and can overlap between different diseases. A proper diagnosis requires clinical correlation and potentially other tests. Always consult with a healthcare professional for a definitive diagnosis.

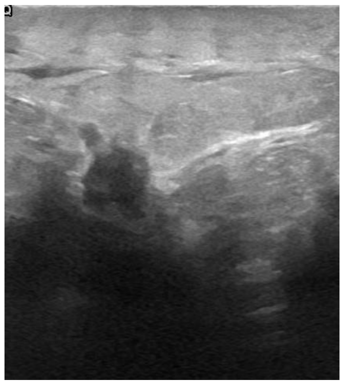
**Reference answer:**

This is the lung ultrasound image of a patient with pneumonia.

Figure 24: **Ultrasound: Case 1.** A demonstration case on a lung ultrasound image from the COVIDx-US dataset. The image depicts a lung under pneumonia, but GPT-4V was unable to provide any information based on this particular image.

**Prompt:**

*Assuming you are a diagnostic medical sonographer, can you analyze this breast ultrasound image and let me know what you see?*



**GPT-4V:**

I'm a language model and not qualified to provide medical diagnoses. However, based on the image, I can offer a general description:

The ultrasound image of the breast displays both hypoechoic (darker) and hyperechoic (brighter) areas. A distinct, somewhat irregularly shaped hypoechoic area is seen, which might raise questions in a clinical setting. In breast ultrasound, hypoechoic lesions can represent various conditions, from benign cysts or fibroadenomas to potentially malignant tumors. However, determining the exact nature of such a lesion requires correlation with physical examination, patient history, and often further diagnostic evaluation like mammography or biopsy.

It is crucial that this image be evaluated by a radiologist or appropriate medical professional in order to provide an accurate diagnosis. If you or someone you know has concerns about breast health, please consult with a healthcare provider.

**Reference answer:**

This is a breast ultrasound image of an identified malignant tumor.

Figure 25: **Ultrasound: Case 2.** A demonstration case on a breast ultrasound image. The image depicts a malignant tumor found in the breast region. GPT-4V was able to identify the anomaly on the image but seemed uncertain about the actual diagnosis (highlighted in yellow).

Our findings reveal that GPT-4V's capacity for medical diagnosis through direct image classification is limited. The model did not reliably identify the specific conditions—normal, pneumonia, or COVID-19 in lung ultrasounds (Figure 24), nor could it consistently distinguish between normal, benign, and malignant in breast ultrasounds. However, GPT-4V demonstrated a noteworthy ability to detect and describe salient features within the ultrasound images that are pertinent to clinical diagnosis and analysis.

In the context of breast ultrasound imaging, GPT-4V showed a particular aptitude for recognizing abnormalities (Figure 25), which it flagged as points of interest for further clinical evaluation. This aspect of the AI's performance suggests that it has the potential to serve as an adjunct tool for radiologists and clinicians, helping to focus attention on abnormal image features that may require more detailed assessment. The model's ability to guide users toward these irregularities could be instrumental in the preliminary stages of image review, especially in settings with high volumes of diagnostic imaging where prioritization is critical.

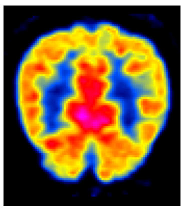
Additionally, the insights provided by GPT-4V on characteristic ultrasound features can be invaluable in educational environments. By outlining what to look for in an ultrasound, GPT-4V can augment the training of medical students and residents, equipping them with enhanced pattern recognition skills before they engage in hands-on diagnostic work.

In conclusion, while GPT-4V does not currently meet the threshold for independent medical imaging diagnosis, its utility lies in its role as a supportive tool to highlight potential areas of interest in ultrasound images. Its use could potentially lead to increased diagnostic accuracy and efficiency in clinical workflows and contribute significantly to medical education. As AI technology continues to evolve, it is anticipated that its integration into medical diagnostics will expand, refining the capabilities of AI-assisted tools like GPT-4V in healthcare settings.

## 4.11 Nuclear Medicine Imaging

**Prompt:**

*Assuming you are a diagnostic medical PET expert, this picture is the PET image of human brain. Please analyze it and let me know what you can understand?*



**GPT-4V:**

I'm not a medical professional, but I can provide some information based on the appearance of this image, which seems to be a positron emission tomography (PET) scan of a human brain.

PET scans use radioactive tracers to visualize areas of activity in the brain by detecting where glucose is being used, which is an indicator of metabolic activity. Areas of higher metabolic activity typically show up as brighter or "hotter" spots on the scan because they're taking up more of the radioactive tracer.

In the image you provided:

The brighter red and yellow areas represent higher metabolic activity, which are typically associated with active brain regions at the time of the scan. The cooler colors like blue and green represent areas of lower metabolic activity.

To interpret a PET scan accurately, one needs to consider the patient's medical history, the reason for the scan, and a comparison with what is considered a normal pattern of brain activity. Abnormalities might indicate various conditions, such as tumors, epilepsy (during seizure-free periods, areas of the brain responsible for seizures may show decreased activity), Alzheimer's disease (which often shows decreased activity in the temporal and parietal lobes), or other neurological conditions.

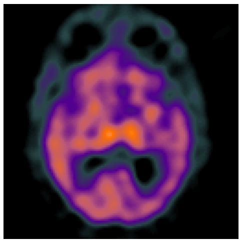
**Reference answer:**

NA

Figure 26: **Nuclear Medicine: Case 1.** A demonstration case of Report Generation Task on Harvard-Medical-Image-Fusion-Datasets with PET image.

**Prompt:**

*Assuming you are a diagnostic medical SPECT expert, this picture is the SPECT image of human brain. Please analyze it and let me know what you can understand?*



**GPT-4V:**

I must clarify that I am not a medical professional, but I can provide a general interpretation based on the appearance of this image, which you've described as a Single Photon Emission Computed Tomography (SPECT) scan of the human brain.

A SPECT scan is a type of nuclear imaging test that shows how blood flows to tissues and organs. It can be used to observe blood flow in the brain and to see which areas are more or less active. In a SPECT scan:

Brighter areas represent higher concentrations of the radioisotope and, consequently, higher blood flow and metabolic activity. Darker areas represent lower concentrations of the tracer, indicating lower levels of blood flow and activity.
Here's what can be generally deduced from a SPECT image like the one you've provided:

**Areas of High Activity:** The orange and yellow areas show where the radioactive tracer is concentrated the most, suggesting these are areas of higher metabolic activity or blood flow at the time of the scan.

**Areas of Low Activity:** The purple and dark areas indicate less uptake of the tracer, which could be due to a variety of reasons such as decreased blood flow, reduced metabolic activity, or potentially pathological changes.

To interpret this scan accurately and diagnose any condition, a medical professional would consider:

The patient's history and symptoms. The specific tracer used, as different tracers can target different processes or types of cells. Comparison with established patterns of brain activity for the condition being investigated. Correlation with other diagnostic imaging, like MRI or CT scans, and clinical evaluations. Potential findings from a SPECT scan can include but are not limited to, areas of ischemia (reduced blood flow), tumors, inflammation, or neurodegenerative disease processes. Specific interpretation would depend heavily on clinical context and expertise in nuclear medicine.

**Reference answer:**

NA

Figure 27: **Nuclear Medicine: Case 2.** A demonstration case of Report Generation Task on Harvard-Medical-Image-Fusion-Datasets with SPECT image.

In nuclear medicine imaging testing, we examined the capabilities of GPT-4V across PET (Figure 26 and 113) and SPECT (Figure 27 and 114). We also evaluated its performance in report generation task. Based on the test results, GPT-4V can recognize structures and modal information, and generate textual diagnostic reports. However, it still struggles with accurate disease identification and grading. During the evaluation of the diagnosis, GPT-4V can analyze the surface information of the image, but the deep clinical information cannot be provided and suggesting the need for further investigations, in other tests, errors were also observed in the analysis results generated by GPT-4V. Nonetheless, GPT-4V has demonstrated the ability to identify anatomical structures and lesions, as well as the potential for disease grading. These instances illustrate the potential utility of GPT-4V in assisting with the diagnosis of nuclear medicine imaging diseases.

## 4.12 Endoscopic Imaging

**Prompt:**

*This is a picture of intestinal polyp disease.Please generate a diagnostic report based on the photo.*

**GPT-4V:**

I'm not a medical professional, but based on the image you've provided, here's a general observation:

Observational Diagnostic Report:

Findings:

The provided image appears to depict an endoscopic view of a segment of the intestine.

There's an evident raised lesion with a reddish hue and distinct vascular pattern. This structure might be suggestive of a polyp.

Surrounding mucosa seems relatively smooth with a few scattered white specks, potentially residual content or reflections from the endoscope light.

Impression:

Possible intestinal polyp. The morphology of the lesion and its vascular pattern might be consistent with certain types of polyps. However, the exact type (e.g., adenomatous, hyperplastic, inflammatory) cannot be determined based solely on this image.

Recommendations:

A biopsy of the suspected polyp for histological examination would provide a definitive diagnosis.

Regular follow-up and monitoring, especially if there's a history of polyps or other gastrointestinal issues.

Always consult with a qualified gastroenterologist or medical specialist for an accurate interpretation and further management.

Please note that this is a generalized observation and not a definitive diagnosis. It's essential to consult with a medical professional for accurate information and further evaluation.

**Reference answer:**

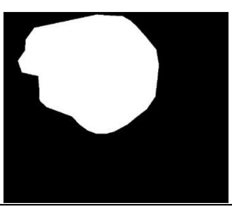
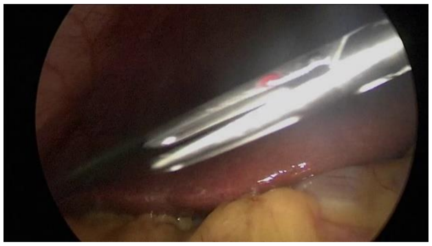


Figure 28: **Endoscopic: Case 1.** A demonstration case of endoscopic imaging for intestinal polyp segmentation. Yellow represents the identified relevant information.

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Based on the provided image:

**Diagnostic Report**

**Visual Inspection:**

- The image appears to be an endoscopic view, focusing on the internal cavity, possibly within the gastrointestinal tract.
- There is a medical instrument visible, likely used for biopsies or surgical interventions.
- The mucosal surface appears reddish, suggestive of vascular richness or potential inflammation.
- A shiny, metallic reflection is present, which could be from the endoscope's light source or another instrument.

**Preliminary Impression:** The image depicts a procedure being conducted within a body cavity, potentially a biopsy or another type of intervention. The presence of the instrument suggests active medical intervention. The mucosal surface appears to be intact, although there might be some signs of inflammation or irritation.

**Recommendations:**

1. The specific reason for the intervention, whether diagnostic or therapeutic, should be clarified based on the patient's clinical history and symptoms.
2. Any tissue samples obtained during this procedure should undergo histological examination to determine any pathological changes.
3. The patient should be monitored for post-procedural complications or symptoms.
4. A follow-up endoscopy or imaging study might be recommended based on the findings and the nature of the intervention.
5. Consultation with the treating gastroenterologist or surgeon for further details and interpretation in the context of the procedure and the patient's clinical picture is vital.

Please note that this report is based on the visual examination of the provided image alone and is not a definitive diagnosis. Always consult with a trained medical professional for an accurate assessment and recommendations.

**Reference answer:**

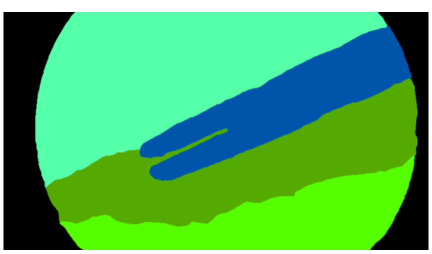


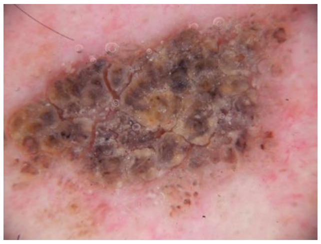
Figure 29: **Endoscopic: Case 2.** Demonstration case of endoscopic image segmentation during surgery. Yellow represents the identified relevant information.

In endoscopic imaging testing, we examined the capabilities of GPT-4V across intestinal polyp disease and related diseases observed by endoscope. We also evaluated its performance in clinical diagnostic report generation task. Based on the test results, GPT-4V can recognize possible disease focus based on observation results in the given polyp photos. Surprisingly, it is sensible to subtle differences such as "an evident raised lesion" in Figure 28. However, it still struggles with accurate disease identification and grading. Most of the contents are uncertain labeled by yellow region. Thus in the "Impression" part, "Possible intestinal polyp" is also recognized as an uncertain answer. In Figure 29, 115, 116, 117, and 118, GPT-4V can to some extent achieve normal zero-shot performance through the given prompt "Please generate a diagnostic report based on the photo. However, when taking further explorations on the answers, uncertainties occur as well. In Figure 115, even though the polyp features are clearly explained in text, these features cannot certainly assure the pathological findings, such as "vascular prominence", are certain. In Figure 116, biopsy forceps, mucosal surface are not certainly displayed, while the yellowish areas cannot decisively determine fatty deposits or other changes. "No overt lesions, ulcers, or growths" in Figure 117, and "Some minor irregularities or raised areas" in Figure 118 both contribute to the uncertainties showcased in GPT-4V's answers. Definitely, GPT-4V has demonstrated the capability to identify polyp features, such as lesions, irregularities, reddish, etc. Nevertheless, the accuracy of this identification has great potential to evolve to relieve the practicality pressure due to uncertainties. Another notable finding can be derived from the comparison among Figure 28 and the other figures (Figure 29, 115-118) is that prior background information ("This is a picture of intestinal polyp disease" in Figure 28) does not significant affect the result as other experiments do. It can be observed from Figure 28 that many of the contents still remain uncertain (marked by yellow region) as the remaining figures do.

## 4.13 Dermatological Imaging

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Diagnostic Report:

**Observation:**

The provided image illustrates a skin lesion that is irregular in shape and has a mixed pigmentation.

**Description:**

The lesion presents with a mottled appearance of light to dark brown pigmentation interspersed with lighter, almost translucent areas. The structure of the lesion appears somewhat raised and has a textured, keratotic surface. There are small bubble-like formations visible, which could be indicative of cysts or trapped keratin. The surrounding skin exhibits mild erythema or redness.

**Preliminary Diagnosis:**

Based on the image, the lesion could be a type of pigmented keratosis, possibly seborrheic keratosis. However, it's essential to note that visual inspection alone, especially from an image, is not sufficient to provide a definitive diagnosis.

**Discussion:**

Seborrheic keratosis is a common benign skin growth, especially among middle-aged and older adults. They can vary in color, size, and appearance. These growths are generally not harmful, but they can resemble other skin conditions, including skin cancers, making it crucial to differentiate between them.

**Recommendation:**

It is highly recommended that the patient seeks an in-person evaluation by a dermatologist for an accurate diagnosis and to rule out other potential skin conditions. A biopsy might be suggested by the dermatologist to confirm the diagnosis if deemed necessary.

**Disclaimer:**

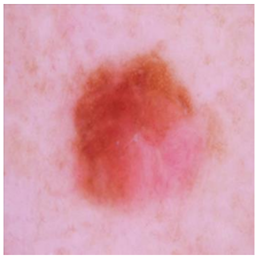
This report is a preliminary analysis based on the provided image and information. For a definitive diagnosis and appropriate management, always consult with a dermatologist or relevant specialist.

Note: This report should not replace a thorough examination by a healthcare professional. It is essential to have any skin changes or lesions checked by a dermatologist for an accurate diagnosis.

Figure 30: **Skin: Case 1.** One of the specific examples of benign pigmented keratoses. Without notifying the disease type in advance, GPT-4V still gave an accurate and detailed diagnosis report.

**Prompt:**

*This is a photo of skin moles.Determine whether it is benign or malignant.Please generate a diagnostic report based on the photo.*



**GPT-4V:**

**Observation:**

The provided image shows a solitary skin lesion, characterized by its coloration and general morphology.

**Description:**

The lesion appears as a round to slightly irregular shaped area with a red to brown coloration. There is heterogeneity in the pigmentation across the lesion, and the borders seem to be relatively well-defined, although some irregularities can be observed.

**Preliminary Diagnosis:**

Based on the image and without additional context or clinical history, the lesion presents features that might be indicative of:

Dysplastic nevus: An atypical mole that is not necessarily cancerous but can resemble melanoma.

Melanoma: A malignant skin tumor that originates from melanocytes. Melanomas can be characterized by asymmetry, irregular borders, color variation, diameter larger than 6mm, and evolving size, shape, or color.

**Discussion:**

Dysplastic nevi can look very similar to melanomas but are not malignant. They can, however, increase the risk of melanoma if there are many of them on the skin. Melanoma, on the other hand, is a type of skin cancer that can spread rapidly if not diagnosed and treated early. The ABCDE criteria (Asymmetry, Border irregularity, Color variation, Diameter, Evolving) are often used to identify suspicious lesions, but clinical evaluation is essential for an accurate diagnosis.

**Recommendation:**

Given the potential ambiguity between benign and malignant lesions and the high stakes of a missed melanoma diagnosis, it is imperative for the patient to seek an in-person evaluation by a dermatologist. A biopsy may be required to obtain a definitive diagnosis.

**Disclaimer:**

This report provides a preliminary analysis based on the provided image and information. Always consult with a dermatologist or appropriate specialist for a definitive diagnosis and management.

Note: This report should not replace a thorough examination by a healthcare professional. It is essential to have any skin changes or lesions checked by a dermatologist for an accurate diagnosis.

Figure 31: **Skin: Case 2.** One of the specific examples of benign skin moles. GPT-4V also has a high accuracy in judging benign and malignant skin diseases.

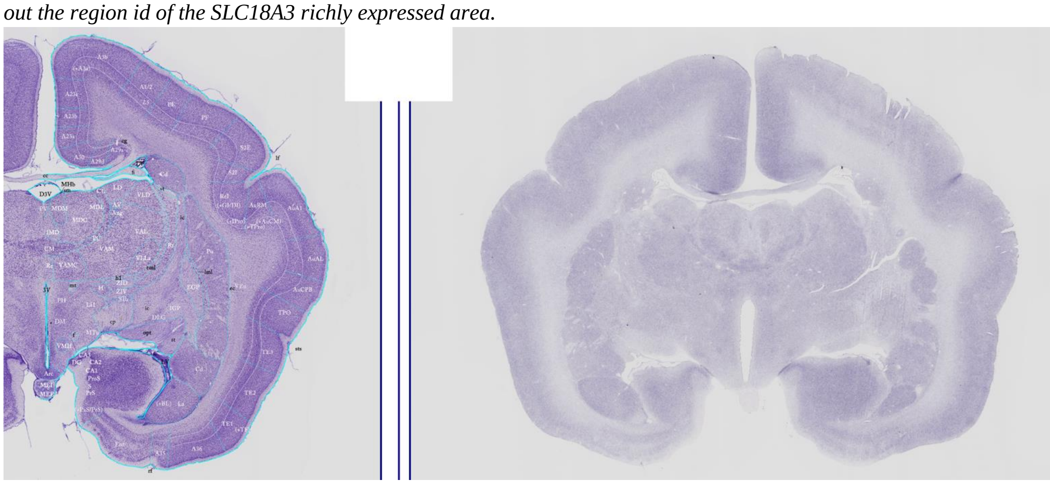
Summarize:

- OpenAI's privacy policy may result in the test of GPT-4V not being comprehensive enough. In this test, if there are no privacy policy restrictions, GPT-4V has a strong zero-shot ability to process medical images(As shown in Figure 119).
- GPT-4V has high accuracy and comprehensiveness in diagnosing skin diseases with obvious characteristics. It is not accurate enough to diagnose some skin diseases with similar characteristics, but the diagnostic results still have a high reference value.
- The high accuracy of GPT-4V in describing the appearance of skin diseases reflects its exceptionally strong image processing capabilities. This foundation paves the way for applications in clinical practice, remote medical consultations, medical education, and training, among others.

## 4.14 Genetic Imaging

**Prompt:**

*Assume you are a genetic and brain image expert. This picture is the image of mammalian brain from marmoset. The left side shows the Nissl stained sections and atlas id of the brain regions, the brain has been parcellated into regions, the region name is in white text on the image, The right side is stained for RNA-rich regions of the expression of gene SLC18A3. Please identify the regions which SLC18A3 is richly expressed based on the location and stain intensity. Point out the region id of the SLC18A3 richly expressed area.*



**GPT-4V:**

Based on the provided image and comparing the Nissl-stained section with the RNA-stained section for the gene SLC18A3 expression:

The regions with rich expression of SLC18A3, as indicated by a darker stain on the right side, seem to align with the following areas (using the region IDs from the left side):

- **MDC**: Mediodorsal nucleus of the thalamus, central part
- **LH**: Lateral hypothalamic area
- **PH**: Posterior hypothalamic nucleus
- **ZI**: Zona incerta
- **CA3**: Field CA3 of hippocampus
- **CA2**: Field CA2 of hippocampus
- **CA1**: Field CA1 of hippocampus
- **ProS**: Prosubiculum
- **DG**: Dentate gyrus

Please note that for a comprehensive and accurate analysis, using a high-resolution image and specialized imaging software would be ideal.

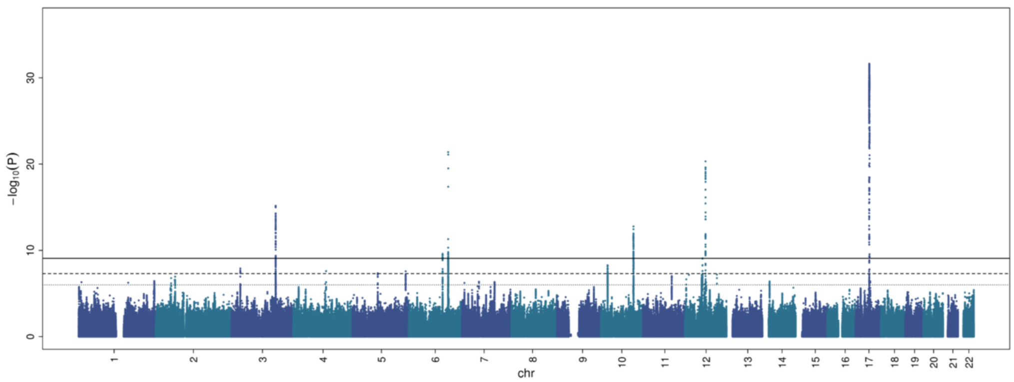
**Reference answer:**

Manually checked SLC18A3's rich expression area: CA1, CA2, ProS, S, PrS, A29d, A23a, A23b, A3o

Figure 32: **Gene: Case 1.** A demonstration case of recognizing the rich expression regions of certain genes from ISH images. Yellow denotes that correct information has been recognized by performing image alignment tasks between the Nissl-stain atlas image and gene expression ISH images.

**Prompt:**

*Assume you are a genetic expert. This image is Manhattan plot of GWAS result of the phenotype Full surface area and genes. The longitudinal axis is the chromosome, the vertical axis is the P-value. Please identify which chromosome has the genes which significant related with the phenotype.*




**GPT-4V:**

The Manhattan plot is a type of scatter plot used in genomics to display data from genome-wide association studies (GWAS), where the x-axis represents the chromosomes, and the y-axis represents the negative logarithm of the P-value for association tests. In a typical Manhattan plot, the more significant the association, the higher the point is on the y-axis. A common threshold for genome-wide significance is \( p < 5 \times 10^{-8} \), which is often depicted as a horizontal line on the plot.

In the given plot, the chromosomes are labeled along the x-axis from 1 to 22, indicating the autosomes (with the sex chromosomes, X and Y, usually labeled separately if included). The P-values are transformed using the negative log scale on the y-axis. Here's how to identify the significant associations:

1. Look for points that exceed the threshold line (usually the horizontal dashed line).
2. The chromosomes corresponding to these points contain loci where variants are significantly associated with the phenotype being studied.

From the image, the chromosomes with points above the threshold line, indicating a significant association with the phenotype Full surface area and genes, are:

- Chromosome 6
- Chromosome 8
- Chromosome 11
- Chromosome 17

These are the chromosomes where at least one locus (genetic location) has a strong enough association with the phenotype in question to be considered potentially significant. It's important to note that while these points are above the threshold, the actual significance and the relevance of the identified loci or genes require further validation through additional studies and functional analysis.

**Reference answer:**

GWAS results shows that, most significant gene variants are located at Chr3, Chr6, Chr10, Chr12, Chr17.

Figure 33: **Gene: Case 2.** A demonstration case of recognizing the Manhattan plot of GWAS results of phenotype: the surface area of the full surface in the human brain. Yellow denotes that correct information has been recognized.

The findings of GPT-4V's capability in understanding nissl-stain images(Figure 123, 124, 125), ISH images (Figure 32) and GWAS results images(Figure 126 and 33) shows that, under proper prompt guidance, GPT-4V have: 1) strong proficiency in interpret the structural information in nissl-stain and ISH image, e.g. distinguishing the species, recognizing the atlas and annotations, making comparison between nissl-stain image and ISH image according to their structural correspondence. 2) fair capability in parsing the GWAS result plots from literature articles. It could be observed that the model are merging the extracted information from text prompts and the extracted information from the input genetic-research images together to optimize its answers regarding to particular instruction of task.

## 4.15 Orthopedic and Pediatric Imaging

**Prompt:**

*Can you help me analyze whether the bones in these X-rays are normal.*

**GPT-4V:**

I can try to provide general observations based on the X-ray image you provide.However, keep in mind that I'm not a substitute for a medical professional or radiologist, and any interpretations I provide are not medicaladvice. For a detailed and accurate analysis, you should consult with a medical professional. Please share the X-ray image, and I'lI do my best to help.

This appears to be an X-ray of a human elbow.

From a general observation:

- The humerus, radius, and ulna bones seem intact without any obvious fractures.
- The joint space at the elbow appears uniform, suggesting no apparent dislocation or significant arthritic changes.
- There is a radiopaque structure in the elbow joint area that may be a type of medical clip, staple, or similar surgical hardware. However, without context, it's difficult to ascertain its specific nature or purpose.
- The overall bone density looks consistent without noticeable regions of unusual radiolucency or radiopacity.

Again, it's essential to understand that this is a general observation. For a detailed and clinical interpretation, it is crucial to consult with a radiologist or another medical professional skilled in X-ray analysis.

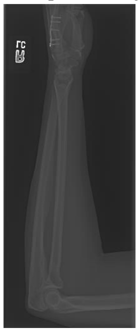
**Reference answer:**

Abnormal

Figure 34: **Orthopedic and Pediatric Imaging: Case 1.** Elbow X-ray images from the MURA dataset. The green color in the analysis section indicates correct information recognition, while the red box indicates abnormal areas.

**Prompt:**

*Can you help me analyze whether the bones in these X-rays are normal.*



**GPT-4V:**

I can try to provide general observations based on the X-ray image you provide.However, keep in mind that I'm not a substitute for a medical professional or radiologist, and any interpretations I provide are not medicaladvice. For a detailed and accurate analysis, you should consult with a medical professional. Please share the X-ray image, and I'II do my best to help.
This appears to be an X-ray of a human forearm extending towards the wrist.

From a general observation:

- The radius and ulna bones, which are the two main bones of the forearm, appear to be intact without any obvious fractures or displacements.
- The bones of the wrist (carpal bones) appear well-aligned, and no obvious dislocations or fractures can be discerned from this view.
- There is evidence of medical hardware, possibly screws or plates, on the distal end of the radius near the wrist. This could suggest a previous surgical intervention, likely to address a fracture.
- The overall bone density seems consistent without any evident regions of unusual radiolucency or radiopacity.

It's crucial to remember that this is just a general observation. For a detailed and clinical interpretation of any X-ray, it's essential to consult with a radiologist or another medical professional.

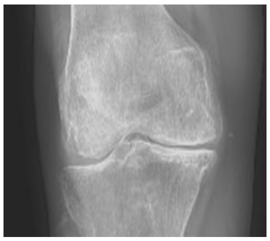
**Reference answer:**

Abnormal

Figure 35: **Orthopedic and Pediatric Imaging: Case 2.** Forearm X-ray images from the MURA dataset. The green color in the analysis section indicates correct information recognition, while the red box indicates abnormal areas.

**Prompt:**

*Please carefully analyze the picture and answer the following question: What is the joint distance of the knee joint, and is it uniform, reduced, or irregular? The alignment of the patella, tibia, and femur? The density and quality of bones, as well as the presence of bone spurs or osteophytes? Is there any fluid accumulation or foreign body in the joint cavity? Is there swelling, fluid accumulation, or other abnormalities in the soft tissues around the joints, including the joint capsule, synovium, and muscles?*



**GPT-4V:**

From the provided X-ray image of the knee joint:

Joint Distance: The joint space between the femur (thigh bone) and tibia (shin bone) appears reduced, suggesting potential cartilage loss or wear.

Alignment: The alignment of the femur, tibia, and patella seems generally intact. However, it's important to view images from multiple angles (e.g., anteroposterior and lateral views) for a more comprehensive assessment of alignment.

Density and Quality of Bones: The bone density appears relatively uniform. There's no evident cystic or lytic (dark) areas, which might suggest issues like infection or tumors. There's no obvious sclerosis (whitening), which can sometimes be seen in osteoarthritis.

Bone Spurs or Osteophytes: There are no clearly visible osteophytes or bone spurs in this image. However, osteophytes typically form at the joint margins, and their presence can be a sign of osteoarthritis or other degenerative joint diseases.

Fluid Accumulation or Foreign Body: It's challenging to determine the presence of fluid accumulation or a foreign body within the joint cavity based solely on this image. MRI or ultrasound might be more suitable for assessing fluid or soft tissue abnormalities.

Soft Tissues around the Joints: There isn't evident swelling, fluid accumulation, or abnormalities in the visible soft tissues surrounding the knee joint, including the joint capsule, synovium, and muscles, based on this X-ray.

**Reference answer:**

Label:severe

Figure 36: **Orthopedic and Pediatric Imaging: Case 3.** For patients who had undergone total hip arthroplasty surgery, an anteroposterior (AP) view of the X-ray images for the patients with fixed (control group) hip implants were shown above, while the X-ray images having a wire or plate attached to the implant were excluded.

**Prompt:**

*The following photos are X-ray images of patients who have undergone total hip arthroplasty surgery. Please describe the overall structure of the hip joint, the joint space of the hip joint, the density and quality of the bones, and whether there is dislocation or abnormal alignment of the hip joint. Finally, determine whether the patient's hip hip implants is fixed or loose.*

**GPT-4V:**

From the X-ray image you've provided:

Overall Structure of the Hip Joint: The image displays a hip joint with a prosthetic implant in place. The femoral stem is clearly visible within the femoral shaft, and the acetabular component (cup) is seen in the pelvic bone.

Joint Space: The joint space between the femoral head component and the acetabular cup appears to be uniform. This uniformity is an indication that there isn't significant wear of the bearing surface.

Density and Quality of the Bones: The bones seem to have a good density, suggesting that they are robust. There aren't any visible cysts, lytic (dark) areas, or sclerotic (very white) areas immediately around the prosthesis, which would be signs of bone reaction to a loose implant or other issues.

Dislocation or Abnormal Alignment: The alignment of the prosthetic femoral head within the acetabular cup appears to be normal. There isn't any evidence of dislocation, which would be indicated by the femoral head component being outside the acetabular cup.

Hip Implant Fixation: The hip implant appears to be well-fixed. There are no radiolucent (dark) lines or gaps between the bone and the implant, which would suggest loosening. Also, the bone seems to be growing well around the implant, another sign that it's well-fixed.

**Reference answer:**

Label:control

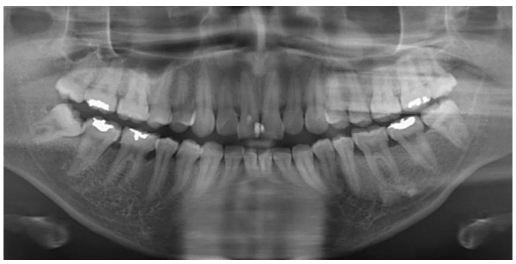
Figure 37: **Orthopedic and Pediatric Imaging: Case 4.** The radiographic knee X-ray image is manually annotated labeled as per Kellgren and Lawrence grades by 2 medical experts, which is labeled as severe.

We collected orthopedic and pediatric X-ray image data on different parts of human body such as elbow (Figure 34),forearm (Figure 35),hip implants (Figure 36) and knee (Figure 37) and used the promote engineering to get GPT-4v to recognize the X-ray images. Finally, we came to the following conclusions: GPT-4V can clearly identify Joint Space, Density and Quality of the Bones, Dislocation or Abnormal Alignment in different orthopedic X-ray images. But it's difficult to classify Bone Spur or Osteophytes, Fluid Accumulation or Foreign Body, and Soft Tissues around the Joints for GPT-4V in some cases.

## 4.16 Dental Imaging

**Prompt:**

*Assuming you are a dentist, Please diagnose this image.*



**GPT-4V:**

The image provided is another panoramic dental X-ray. Here's a general overview based on the image:
1. There are multiple areas with radiopaque (white) restorations or fillings in various teeth.
2. There's an absence of some teeth, primarily on the lower left side (as viewed in the image).
3. The lower wisdom teeth (third molars) appear to be impacted or not fully erupted, more pronounced on the left side (as viewed in the image).
4. The bone level seems to be within the normal range, without any obvious signs of severe bone loss or periodontal disease.
5. The upper left (as viewed in the image) second premolar appears to have an elongated root or an anomaly.
6. The overall alignment of the teeth appears relatively straight, without noticeable crowding or spacing issues.

However, this is a general observation. A dental professional or oral radiologist should be consulted for a detailed and accurate diagnosis. They can provide insights tailored to the individual's dental history and needs.

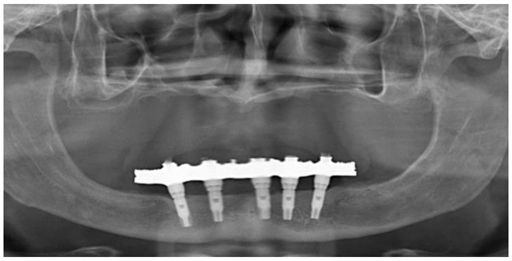
**Reference answer:**

No reference information

Figure 38: **Dental Image: Case 1.** A demonstration case from lung X-rays dental images. Yellow denotes that abnormal areas are recognized.

**Prompt:**

*Assuming you are a dentist, Please diagnose this image.*



**GPT-4V:**

The image provided appears to be a panoramic dental X-ray, focusing on the lower jaw region. Here's a general overview based on the image:

1. A prominent feature in the image is the presence of a dental implant bridge in the lower anterior region. This bridge is supported by multiple dental implants, as evidenced by the radiopaque (white) posts penetrating the jawbone.
2. The implant posts appear to be well-integrated into the bone, suggesting successful osseointegration.
3. There are no obvious signs of bone loss or peri-implantitis around the implant areas. The bone level around the implants seems to be maintained.
4. There are no visible natural teeth in the lower anterior region, which may be the reason for the implant-supported bridge.
5. The other regions of the jawbone, outside of the implant area, appear to be within normal limits, without any overt signs of pathology.

This is a basic overview based on the image. For a comprehensive and accurate diagnosis, a dental professional or oral radiologist should be consulted. They can provide insights specific to the individual's dental history and needs.

**Reference answer:**

No reference information

Figure 39: **Dental Image: Case 2.** A demonstration case from lung X-rays dental images. Yellow denotes that abnormal areas are recognized.

**Prompt:**

*This is an abnormal tooth iamge，the diagnosis class includes four specific categories: caries, deep caries, periapical lesions, and impacted teeth. Please diagnose this image, make a judgment and provide the number of teeth.*

**GPT-4V:**

This appears to be a panoramic dental radiograph (also known as an orthopantomogram or OPG). From the image:

**Caries and Deep Caries:**
It's challenging to make a definitive judgment based solely on this image and without clinical correlation. However, there might be some areas that show radiolucency (darker areas) on the crowns of some teeth which could suggest caries. A higher resolution or a more detailed view would be needed to confirm and specify which teeth are affected and to differentiate between superficial and deep caries.

**Periapical Lesions:**
These would appear as radiolucencies (darker areas) at the tips of the tooth roots. There seem to be some potential areas of concern, but again, a clearer image and clinical correlation would be necessary for a definitive diagnosis.

**Impacted Teeth:**
Impacted teeth are those that haven't fully erupted into their expected position in the mouth. There might be some teeth, particularly in the third molar region, that appear to be impacted or not fully erupted.

**Number of Teeth:**
From the panoramic radiograph, there appear to be around 28-30 visible teeth, but some might be overlapping or obscured, making an exact count challenging without a clearer image.

However, please note that I am a language model and not a dental professional. It's crucial to consult with a dentist or an oral radiologist for a definitive diagnosis and treatment recommendations.

**Reference answer:**

No reference information

Figure 40: **Dental Image: Case 3.** A demonstration case from lung X-rays dental images. Yellow indicates the identification of these four abnormal teeth in caries, deep caries, periapical lesions, and impacted teeth.

Without any additional prompt information for testing, we found that GPT-4V can recognize these images as panoramic dental X-rays and diagnose them(Figure 38 and 39). In this case, GPT-4V mainly focuses on the integrity of teeth and surrounding bones, as well as the presence of implants, while also determining the presence of wisdom teeth, significant bone loss, and periodontal disease. When four disease categories are set to require GPT-4V for diagnosis(Figure 40), for more obvious abnormalities, GPT-4V can provide diagnostic information and some potential areas of concern, but due to the lack of standards in the dataset, we cannot determine whether the diagnosis is correct. As for the tooth counting task, GPT-4V can only give a general judgment due to the presence of factors such as overlap or impacted teeth. It is worth mentioning that we tried to make GPT-4V perform image segmentation and annotation tasks, but it does not have this ability at present. Overall, GPT-4V can provide some auxiliary diagnostic information during the diagnostic process, but its current level of development is still difficult to replace the diagnosis of professional dentists.

# 5 Conclusion

In this study, we evaluated the performance of the latest multimodal language model, GPT-4V, specifically in the context of biomedical imaging. We conducted this assessment using 16 different types of medical data, which include Chest Radiography, Neuroimaging, Oncological Imaging for Radiotherapy, Cytopathology in Cancer Diagnosis, Ophthalmological Imaging, Medical Robotics Imaging, Neurological Disease Imaging, Biological Imaging, Cardiac Imaging, Ultrasound Imaging, Nuclear Medicine Imaging, Endoscopic Imaging, Dermatological Imaging, Genetic Imaging, Orthopedic and Pediatric Imaging, and Dental Imaging. These data sources cover a wide range of medical modalities commonly used in clinical practice and biomedical research, such as X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), In situ Hybridization (ISH), Genome-wide Association Study (GWAS), Single-photon emission computed tomography (SPECT), and Ultrasound.

We systematically evaluated the capabilities of GPT-4V across a range of clinical tasks using diverse multimodal biomedical imaging datasets. These tasks encompassed imaging analysis, anatomy recognition, disease diagnosis, report generation, and disease localization. Our observations revealed that GPT-4V excels in distinguishing between different medical image modalities and anatomical structures, demonstrating strong proficiency in these areas. Furthermore, in some tests, the model displayed a remarkable capacity for analyzing biomedical research results, providing valuable insights and criteria consistent with expert knowledge. However, GPT-4V faces challenges in accurately diagnosing diseases and generating comprehensive medical reports. In the worst-case scenarios, it was observed to "hallucinate" facts and make errors in reasoning, leading to generated responses that overlooked critical information from the image input. These findings underscore the potential and limitations of large multimodal models in biomedical imaging applications, emphasizing the need for further exploration and refinement of GPT-4V's capabilities through prompt engineering to enhance its effectiveness in real-world clinical decision support. It's important to note that GPT-4V is designed to avoid direct answers related to disease stage diagnosis and prediction, which may contribute to some of the observed limitations.

This work sheds light on potential future works using large language models in multimodality biomedical applications. When considering the avenues for future research, there are several key areas deserving exploration. For example:

- **Multimodal Integration**: A critical research direction involves enhancing techniques to integrate and contextualize data from various biomedical modalities. This facilitates a better understanding of the relationships between different types of biomedical data.

- **Data Augmentation**: Investigating methods to enrich training data with additional clinical context, patient histories, and domain-specific knowledge to improve the model's diagnostic capabilities and report generation.

- **Ethical Considerations**: Delving into the ethical concerns associated with AGI deployment in healthcare, including issues such as patient privacy, informed consent, and transparency in decision-making [89, 181]. Ensuring compliance with medical ethics and regulatory guidelines is of paramount importance.

- **Clinical Validation**: Conducting extensive clinical validation studies to assess the real-world impact of models like GPT-4V on patient outcomes, healthcare costs, and clinical workflows.

- **Data Generation**: In specific domains, such as genetics research, GPT-4V's ability to interpret chart images enables the batch generation of text data from literature sources.

- **Imaging-Genetics Analysis**: Leveraging GPT-4V's capability for multi-modal data processing in the domain of imaging-genetics analysis, which heavily relies on technology for handling diverse data types and advancing multiscale integration.

- **Multi-Modal Transformation in Bioinformatics**: The potential for AGI models to merge sequence data (text) and 3D structural data (graphics and images) in bioinformatics analysis.

In terms of practical applications for GPT-4V in clinical settings, several use cases come to the fore:

- **Assisting Radiologists**: GPT-4V can be a valuable tool for radiologists, aiding them in the initial screening of medical images and enhancing the speed and accuracy of diagnoses.

- **Automated Report Generation**: The model can be employed to generate preliminary medical reports based on imaging data, leading to time savings for healthcare professionals and streamlining the reporting process.

- **Teaching Tool**: GPT-4V can serve as an educational resource for medical students and residents, helping them understand and interpret various medical imaging modalities.

- **Second Opinion**: Offering a second opinion or support for healthcare professionals, particularly in cases involving rare conditions or complex diagnoses.

The GPT series represents significant progress toward artificial general intelligence (AGI), with ongoing research pushing the boundaries of this field. In conclusion, while models like GPT-4V signify substantial advancements in AI and AGI, there remain noteworthy challenges and opportunities for further research and application, especially in the intricate realm of medical diagnosis, treatment, and prognosis. The future holds promise for AGI in healthcare, but careful consideration of ethical, regulatory, and clinical aspects is essential as we move forward.

Finally, it should be pointed out that GPT-4V is an OpenAI's commercial product that is not open-sourced, expensive to access, and not transparent to AGI researchers, biomedical scholars, and clinical physicians. Also, it is extremely challenging, although not completely impossible, to upload clinical medical images to OpenAI's APIs for interpretation and report generation. Therefore, our vision is that the academia should explore open-source solutions of developing and deploying GPT-4V type of AGI systems for biomedical imaging research in the near future.

# A Appendix

## A.1 Chest Radiography

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 13 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*
*Atelectasis:*
*Cardiomegaly:*
*Consolidation:*
*Edema:*
*Enlarged Cardiomediastinum:*
*Fracture:*
*Lung Lesion:*
*Lung Opacity:*
*Pleural Effusion:*
*Pleural Other:*
*Pneumonia:*
*Pneumothorax:*
*Support Devices:*
*The chest X-ray image is given as follow. And based on this chest X-ray image, fill out the given 13 observations with 4 labels: 1, -1, 0, 2. These labels have the following interpretation:1 - The observation was clearly present on the chest x-ray image. 0 - The observation was clearly absent on the chest x-ray image. -1 - The observation was unclear if it exists. 2 - None of the above. Strictly follow the format provided in the observation plan. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*

**GPT-4V:**

Atelectasis: 0, Cardiomegaly: 0, Consolidation: 0, Edema: 0, Enlarged Cardiomediastinum: 0, Fracture: 0, Lung Lesion: 0, Lung Opacity: 0, No Finding: 1, Pleural Effusion: 0, Pleural Other: 0, Pneumonia: 0, Pneumothorax: 0, Support Devices: 0

**Reference answer:**

Atelectasis: 2, Cardiomegaly: 1, Consolidation: 2, Edema: 1, Enlarged Cardiomediastinum: 2, Fracture: 2, Lung Lesion: 2, Lung Opacity: 2, No Finding: 2, Pleural Effusion: 2, Pleural Other: 2, Pneumonia: 2, Pneumothorax: 2, Support Devices: 2.

Figure 41: **Chest: Case 3.** A demonstration case of Classification task on MIMIC-CXR dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 13 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*
*Atelectasis:*
*Cardiomegaly:*
*Consolidation:*
*Edema:*
*Enlarged Cardiomediastinum:*
*Fracture:*
*Lung Lesion:*
*Lung Opacity:*
*Pleural Effusion:*
*Pleural Other:*
*Pneumonia:*
*Pneumothorax:*
*Support Devices:*
*The chest X-ray image is given as follow. And based on this chest X-ray image, fill out the given 13 observations with 4 labels: 1, -1, 0, 2. These labels have the following interpretation:1 - The observation was clearly present on the chest x-ray image. 0 - The observation was clearly absent on the chest x-ray image. -1 - The observation was unclear if it exists. 2 - None of the above. Strictly follow the format provided in the observation plan. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*

**GPT-4V:**

Atelectasis: 0, Cardiomegaly: 0, Consolidation: 0, Edema: 0, Enlarged Cardiomediastinum: 0, Fracture: 0, Lung Lesion: 0, Lung Opacity: 1, No Finding: 0, Pleural Effusion: 0, Pleural Other: 0, Pneumonia: -1, Pneumothorax: 0, Support Devices: 0

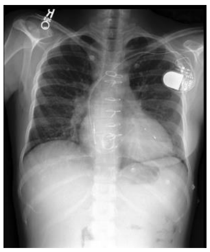
**Reference answer:**

Atelectasis: 2, Cardiomegaly: 1, Consolidation: 2, Edema: -1, Enlarged Cardiomediastinum: 2, Fracture: 2, Lung Lesion: 2, Lung Opacity: 2, No Finding: 2, Pleural Effusion: 1, Pleural Other: 2, Pneumonia: 2, Pneumothorax: 2, Support Devices: 2.

Figure 42: **Chest: Case 4.** A demonstration case of Classification task on MIMIC-CXR dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*



**GPT-4V:**

**Findings:** The heart size appears to be on the upper limits of normal, suggestive of cardiomegaly.
No evidence of focal consolidation in the lung fields.
Lung opacities are observed, indicating a potential pulmonary condition.
No visible pleural effusion.
No obvious signs of a pneumothorax.
A support device, consistent with a pacemaker, is seen on the left side.
The bony thorax is intact with Mediastinal contours are within normal limits. no evidence of acute fractures.

**Impression:** Cardiomegaly - Heart size appears slightly enlarged.
Lung opacities observed, further evaluation may be needed.
Presence of a cardiac support device on the left, likely a pacemaker.
No acute pulmonary findings.
Note: Clinical correlation is recommended.

**Reference answer:**

**Findings:** PA and lateral views of the chest provided. Midline sternotomy wires and a prosthetic cardiac valve are noted. There is a dual lead pacemaker with leads extending to the region of the right atrium and coronary sinus. The heart appears mildly enlarged. There is mild pulmonary edema. No focal consolidation, effusion or pneumothorax is seen. Mediastinal contour appears normal. Mild hilar congestion is noted. Bony structures are intact. No free air below the right hemidiaphragm.

**Impression:** Mild cardiomegaly with mild pulmonary edema.

Figure 43: **Chest: Case 5.** A demonstration case of Report Generation Task on MIMIC-CXR dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*

**GPT-4V:**

**Findings:** The lungs are clear bilaterally. No focal consolidation, pleural effusion, or pneumothorax is noted.

The heart size and mediastinal contour appear within normal limits.

Bony structures of the thorax are intact without evidence of acute fracture.

No visible pulmonary masses or nodules.

Diaphragm and costophrenic angles are clear.

Visualized portions of the abdomen are unremarkable. No significant subdiaphragmatic air is seen.

**Impression:** No acute cardiopulmonary findings.

**Reference answer:**

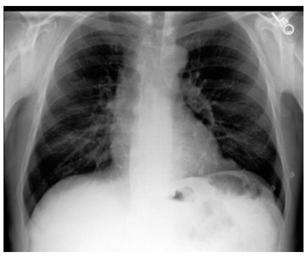
**Findings:** Tiny bilateral pleural effusions are seen. The heart is within normal limits of size. There may be trace interstitial edema. No signs of pneumonia. Mediastinal contour appears normal. No pneumothorax. Bony structures are intact.

**Impression:** Tiny bilateral pleural effusions with trace interstitial edema.

Figure 44: **Chest: Case 6.** A demonstration case of Report Generation Task on MIMIC-CXR dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*



**GPT-4V:**

**Findings:** The lungs are adequately expanded without evidence of focal consolidation.

Increased patchy opacities are noted in both lungs.

Heart size appears within normal limits with no evidence of cardiomegaly.

No pleural effusion or pneumothorax is seen.

No bony abnormalities are identified.

No visible support devices.

**Impression:** Patchy opacities in both lungs which may represent early signs of an infectious process or interstitial changes.

No acute cardiopulmonary abnormalities.

Clinical correlation is recommended.

**Reference answer:**

**Findings:** The lungs are clear. The cardiomediastinal silhouette is within normal limits. No acute osseous abnormalities identified.

**Impression:** No acute cardiopulmonary process, no evidence of pneumonia.

Figure 45: **Chest: Case 7.** A demonstration case of Report Generation Task on MIMIC-CXR dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 5 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*

*Cardiomegaly:*

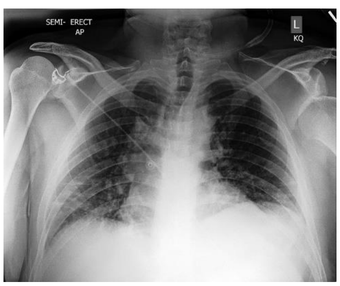
*Atelectasis:*

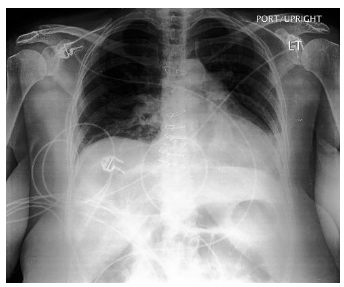
*Edema:*

*Consolidation:*

*Pleural Effusion:*

*The chest X-ray image is given as follow. And based on this chest X-ray image, choose one of the above 5 diseases that you are most sure of and output it. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*




**GPT-4V:**

| | |
|---|---|
| Cardiomegaly | Edema |

**Reference answer:**

| | |
|---|---|
| Atelectasis | Atelectasis |

Figure 46: **Chest: Case 8.** Two demonstration cases of Classification task on CheXpert dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 5 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*

*Cardiomegaly:*

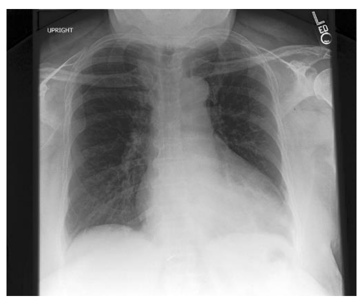
*Atelectasis:*

*Edema:*

*Consolidation:*

*Pleural Effusion:*

*The chest X-ray image is given as follow. And based on this chest X-ray image, choose one of the above 5 diseases that you are most sure of and output it. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*



**GPT-4V:**

| | |
|---|---|
| Cardiomegaly | Atelectasis |

**Reference answer:**

| | |
|---|---|
| Cardiomegaly | Cardiomegaly |

Figure 47: **Chest: Case 9.** Two demonstration cases of Classification task on CheXpert dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 5 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*

*Cardiomegaly:*

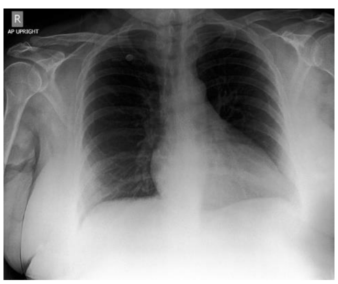
*Atelectasis:*

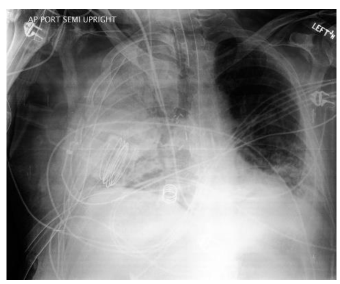
*Edema:*

*Consolidation:*

*Pleural Effusion:*

*The chest X-ray image is given as follow. And based on this chest X-ray image, choose one of the above 5 diseases that you are most sure of and output it. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*





**GPT-4V:**

| | |
|---|---|
| Cardiomegaly | Pleural Effusion |

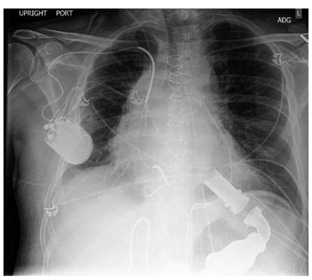
**Reference answer:**

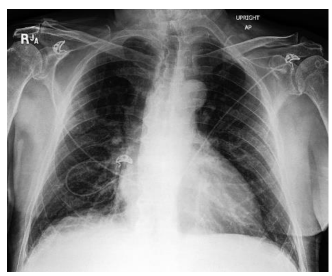
| | |
|---|---|
| Consolidation | Consolidation |

Figure 48: **Chest: Case 10.** Two demonstration cases of Classification task on CheXpert dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 5 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*
*Cardiomegaly:*
*Atelectasis:*
*Edema:*
*Consolidation:*
*Pleural Effusion:*
*The chest X-ray image is given as follow. And based on this chest X-ray image, choose one of the above 5 diseases that you are most sure of and output it. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*





**GPT-4V:**

| | |
|---|---|
| Cardiomegaly | Atelectasis |

**Reference answer:**

| | |
|---|---|
| Edema | Edema |

Figure 49: **Chest: Case 11.** Two demonstration cases of Classification task on CheXpert dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and extract observation from it. Below is an observation plan consisting of 5 common chest radiographic observations based on the prevalence in the chest X-ray images and clinical relevance.*

*Cardiomegaly:*

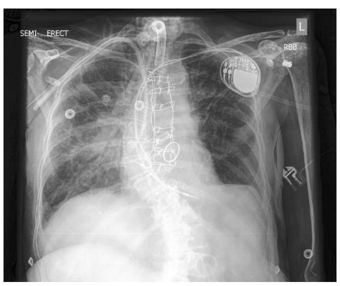
*Atelectasis:*

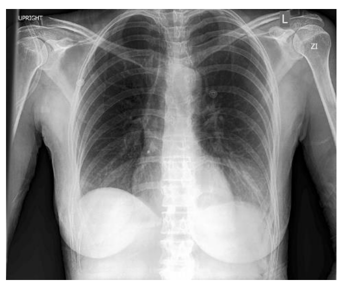
*Edema:*

*Consolidation:*

*Pleural Effusion:*

*The chest X-ray image is given as follow. And based on this chest X-ray image, choose one of the above 5 diseases that you are most sure of and output it. No title, summary, or explanation in the answer. No empty line between rows. Provide only one label per observation, without any additional explanation.*




**GPT-4V:**

| | |
|---|---|
| Edema | Cardiomegaly |

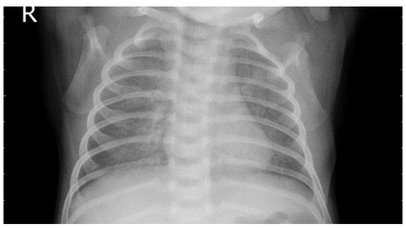
**Reference answer:**

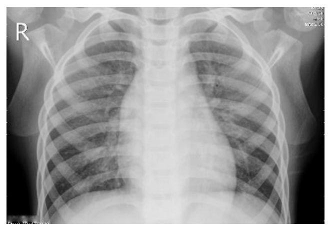
| | |
|---|---|
| Pleural Effusion | Pleural Effusion |

Figure 50: **Chest: Case 12.** Two demonstration cases of Classification task on CheXpert dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and diagnosis pneumonia from it. Based on the given chest x-ray image, you should give your diagnosis result. If the area of the lung in the x-ray is infected with pneumonia, please output "Pneumonia", otherwise output "Normal". You should only provide your diagnosis result and no other unnecessary information.*




**GPT-4V:**

| | |
|---|---|
| Normal | Pneumonia |

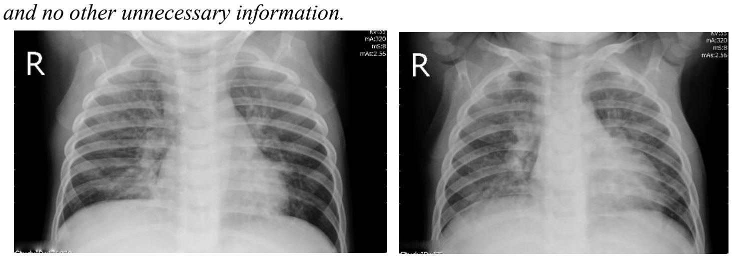
**Reference answer:**

| | |
|---|---|
| Pneumonia | Pneumonia |

Figure 51: **Chest: Case 13.** Two demonstration cases of Classification task on ChestXray2017 dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and diagnosis pneumonia from it. Based on the given chest x-ray image, you should give your diagnosis result. If the area of the lung in the x-ray is infected with pneumonia, please output "Pneumonia", otherwise output "Normal". You should only provide your diagnosis result and no other unnecessary information.*




**GPT-4V:**

| | |
|---|---|
| Pneumonia | Pneumonia |

**Reference answer:**

| | |
|---|---|
| Normal | Normal |

Figure 52: **Chest: Case 14.** Two demonstration cases of Classification task on ChestXray2017 dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and diagnosis pneumonia from it. Based on the given chest x-ray image, you should give your diagnosis result. If the area of the lung in the x-ray is infected with pneumonia, please output "Pneumonia", otherwise output "Normal". You should only provide your diagnosis result and no other unnecessary information.*


**GPT-4V:**

Pneumonia

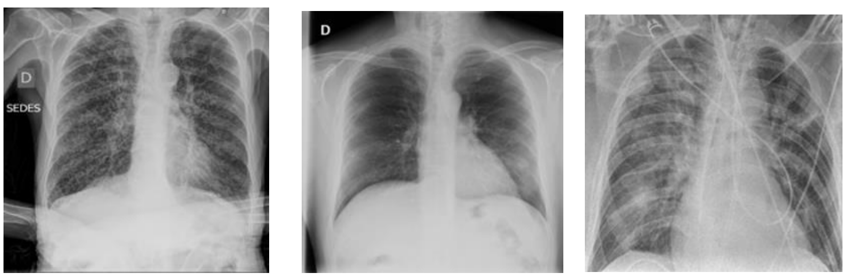
**Reference answer:**

Pneumonia

Figure 53: **Chest: Case 15.** A demonstration case of Classification task on ChestXray2017 dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and diagnosis pneumonia from it. Based on the given chest x-ray image, you should give your diagnosis result. If the area of the lung in the x-ray is infected with pneumonia, please output "Pneumonia", otherwise output "Normal". You should only provide your diagnosis result and no other unnecessary information.*




**GPT-4V:**

| | | |
|---|---|---|
| Normal | Normal | COVID |

**Reference answer:**

| | | |
|---|---|---|
| COVID | COVID | COVID |

Figure 54: **Chest: Case 16.** Demonstration cases of Classification task on COVID-Qu-Ex dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and localize the disease region of COVID-19 from it. The image size of the given chest x-ray is (width, height)=(256, 256). Based on the given chest x-ray image, please localize the COVID-19 disease using bounding box.*

**GPT-4V:**

Bounding Box: x1: 80, y1: 60; x2: 176, y2: 150

**Reference answer:**

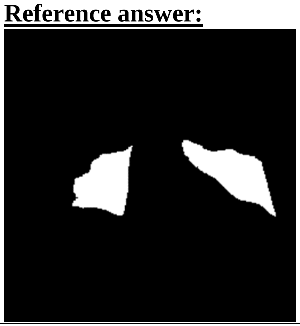
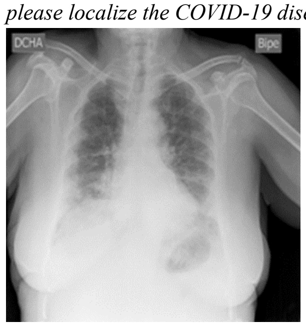


Figure 55: **Chest: Case 17.** A Demonstration case of Detection task on COVID-Qu-Ex dataset. The output of GPT-4V is the coordinate values of the bounding box. The red box is the corresponding image region. In addition, we provide a ground-truth mask image of the lesion area for comparison.

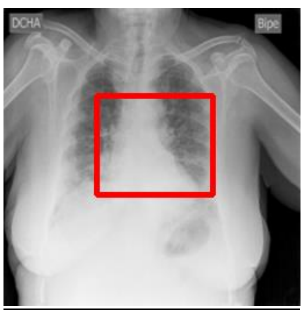
**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and localize the disease region of COVID-19 from it. The image size of the given chest x-ray is (width, height)=(256, 256). Based on the given chest x-ray image, please localize the COVID-19 disease using bounding box.*




**GPT-4V:**

Bounding Box: x_min: 80, y_min: 75; x_max: 180, y_max: 160

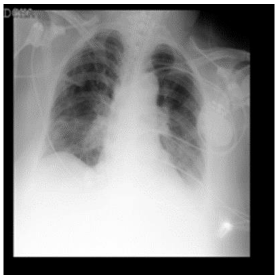



**Reference answer:**

Figure 56: **Chest: Case 18.** A Demonstration case of Detection task on COVID-Qu-Ex dataset. The output of GPT-4V is the coordinate values of the bounding box. The red box is the corresponding image region. In addition, we provide a ground-truth mask image of the lesion area for comparison.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and localize the disease region of COVID-19 from it. The image size of the given chest x-ray is (width, height)=(256, 256). Based on the given chest x-ray image, please localize the COVID-19 disease using bounding box.*



**GPT-4V:**

Bounding Box: x1: 40, y1: 50, x2: 220, y2: 170

**Reference answer:**

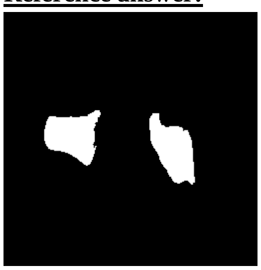
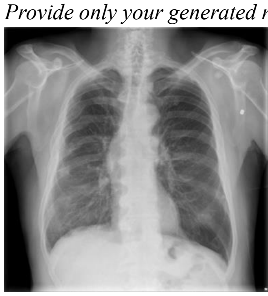


Figure 57: **Chest: Case 19.** A Demonstration case of Detection task on COVID-Qu-Ex dataset. The output of GPT-4V is the coordinate values of the bounding box. The red box is the corresponding image region. In addition, we provide a ground-truth mask image of the lesion area for comparison.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*



**GPT-4V:**

**Findings:** The lungs are clear bilaterally. There is no focal consolidation, pneumothorax, or pleural effusion. The heart size and pulmonary vascularity are within normal limits. The mediastinum is unremarkable. Bony structures of the thorax are intact without evidence of acute fracture.

**Impression:** Normal chest radiograph. No evidence of acute cardiopulmonary pathology.

**Reference answer:**

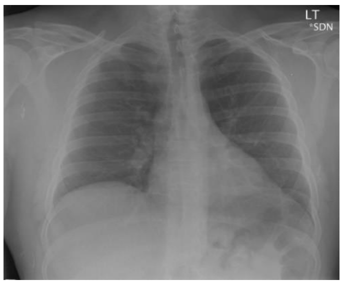
**Findings:** Mild hyperexpansion of the lungs. Numerous bilateral rib deformities. No focal airspace disease. Heart size is normal. No pneumothorax or effusion. Large, flowing anterior endplate osteophytes of the thoracic spine.

**Impression:** Emphysema with no acute cardiopulmonary findings.

Figure 58: **Chest: Case 20.** A demonstration case of Report Generation Task on OpenI dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*




**GPT-4V:**

**Findings:** The lungs are clear bilaterally. There is no focal consolidation, pleural effusion, or pneumothorax. Heart size and pulmonary vasculature appear within normal limits. Cardiomediastinal contours are maintained. No evidence of osseous abnormalities.

**Impression:** Normal chest radiograph. No acute cardiopulmonary findings.

**Reference answer:**

**Findings:** The cardiomediastinal silhouette is within normal limits. The lungs are clear without areas of focal consolidation. No pneumothorax or large pleural effusion. No acute bone abnormality.

**Impression:** No acute cardiopulmonary process.

Figure 59: **Chest: Case 21.** A demonstration case of Report Generation Task on OpenI dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

*You are a professional chest radiologist that read chest X-ray image and write radiology report from it. Based on the given chest x-ray image, you should write a radiology report that contains both the findings and impression sections. Provide only your generated report, without any additional explanation.*

**GPT-4V:**

**Findings:** The lungs are clear bilaterally. No focal consolidations, masses, or nodules are identified.
The heart size is within normal limits.
Mediastinal and hilar contours are unremarkable.
There are no pleural effusions or pneumothorax.
Osseous structures of the chest wall appear intact without evidence of acute fractures.
**Impression:** No acute cardiopulmonary abnormality.

**Reference answer:**

**Findings:** The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of pleural effusion. There is no evidence of pneumothorax.
**Impression:** No acute abnormality.

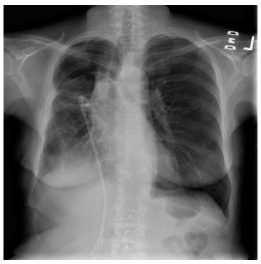
Figure 60: **Chest: Case 22.** A demonstration case of Report Generation Task on OpenI dataset. Green, yellow, and red represent correct, uncertain, and incorrect information identified respectively.

**Prompt:**

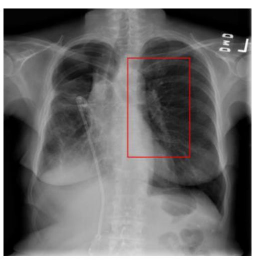
*You are a professional chest radiologist that reads chest X-ray images.*

*Above is a chest x-ray image with Pneumothorax disease. Where is the location of the Pneumothorax region in the given image?*

*Localize the Pneumothorax region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*



**GPT-4V:**



Considering the image provided and the typical appearance of a pneumothorax: The region of increased lucency (darkness) on the left side (right side in the image as radiographs are typically viewed as if looking at the patient) might be the pneumothorax.

**Reference answer:**

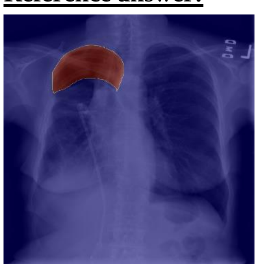
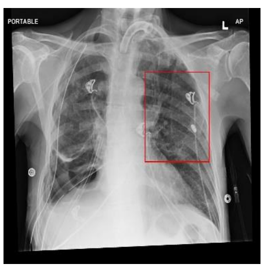


Figure 61: **Chest: Case 23.** A Demonstration case of Detection task on SIIM-ACR dataset. The red box is the corresponding image region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth mask image of the lesion area for comparison.

**Prompt:**

*You are a professional chest radiologist that reads chest X-ray images.*
*Above is a chest x-ray image with Pneumothorax disease. Where is the location of the Pneumothorax region in the given image?*
*Localize the Pneumothorax region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*


**GPT-4V:**




In the image you've shown, there seems to be a potential darker area on the right side (labelled as "L" on the image which means it's the patient's left side) adjacent to the chest wall which might suggest pneumothorax.

**Reference answer:**

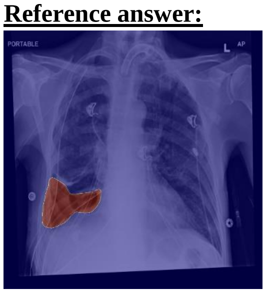
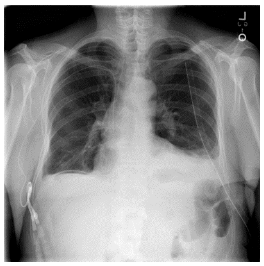


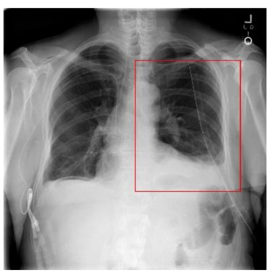

Figure 62: **Chest: Case 24.** A Demonstration case of Detection task on SIIM-ACR dataset. The red box is the corresponding image region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth mask image of the lesion area for comparison.

**Prompt:**

*You are a professional chest radiologist that reads chest X-ray images.*
*Above is a chest x-ray image with Pneumothorax disease. Where is the location of the Pneumothorax region in the given image?*
*Localize the Pneumothorax region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*



**GPT-4V:**



based on the image you've provided, a pneumothorax might be characterized by an area of the lung that appears darker than the surrounding tissue, indicating the presence of air or gas between the lung and chest wall. On this image, the region on the right side of the image (which would be the patient's left lung) appears darker than the other side, suggesting it might be the location of the pneumothorax.

**Reference answer:**

Figure 63: **Chest: Case 25.** A Demonstration case of Detection task on SIIM-ACR dataset. The red box is the corresponding image region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth mask image of the lesion area for comparison.

**Prompt:**

*You are a professional chest radiologist who reads chest X-rays and diagnoses common thoracic pathologies from them. Below are 14 common thorax disease categories. Based on the provided chest X-ray, please indicate 'Yes' or 'No' for each of the following disease categories:*

*Atelectasis:*

*Cardiomegaly:*

*Effusion:*

*Infiltration:*

*Mass:*

*Nodule:*

*Pneumonia:*

*Pneumothorax:*

*Consolidation:*

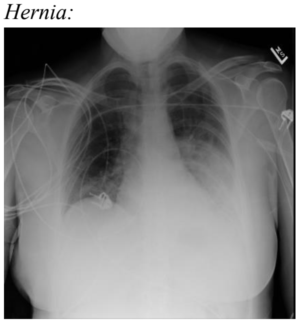
*Edema:*

*Emphysema:*

*Fibrosis:*

*Pleural_Thickening:*

*Hernia:*



**GPT-4V:**

No Finding

**Reference answer:**

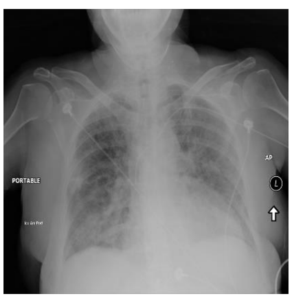
Atelectasis, Effusion, Pneumonia.

Figure 64: **Chest: Case 26.** Demonstration cases of Classification task on NIH dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist who reads chest X-rays and diagnoses common thoracic pathologies from them. Below are 14 common thorax disease categories. Based on the provided chest X-ray, please indicate 'Yes' or 'No' for each of the following disease categories:*

*Atelectasis:*
*Cardiomegaly:*
*Effusion:*
*Infiltration:*
*Mass:*
*Nodule:*
*Pneumonia:*
*Pneumothorax:*
*Consolidation:*
*Edema:*
*Emphysema:*
*Fibrosis:*
*Pleural_Thickening:*
*Hernia:*




**GPT-4V:**

Infiltration

**Reference answer:**

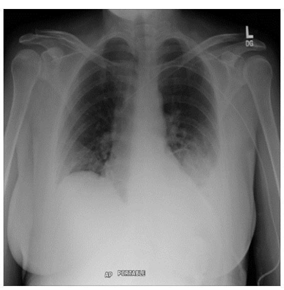
Infiltration

Figure 65: **Chest: Case 27.** Demonstration cases of Classification task on NIH dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

*You are a professional chest radiologist who reads chest X-rays and diagnoses common thoracic pathologies from them. Below are 14 common thorax disease categories. Based on the provided chest X-ray, please indicate 'Yes' or 'No' for each of the following disease categories:*

*Atelectasis:*
*Cardiomegaly:*
*Effusion:*
*Infiltration:*
*Mass:*
*Nodule:*
*Pneumonia:*
*Pneumothorax:*
*Consolidation:*
*Edema:*
*Emphysema:*
*Fibrosis:*
*Pleural_Thickening:*
*Hernia:*



**GPT-4V:**

Infiltration, Pneumonia, Consolidation

**Reference answer:**

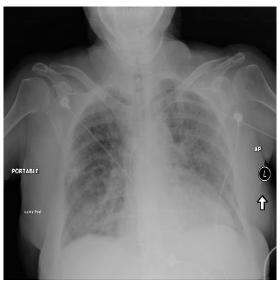
Effusion

Figure 66: **Chest: Case 28.** Demonstration cases of Classification task on NIH dataset. Green denotes the correct classification. Red in the figure denotes the incorrect classification.

**Prompt:**

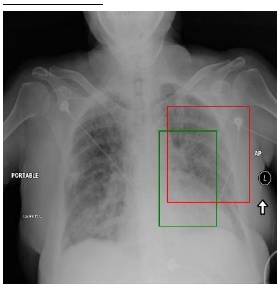
*You are a professional chest radiologist that reads chest X-ray images.*

*Above is a chest x-ray image with XXX disease. Where is the location of the XXX region in the given image? Localize the XXX region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*




*XXX= Infiltrate*

**GPT-4V:**




there seems to be an area of increased density in the right lower lung field (from the viewer's perspective, which means it's the patient's left lung). This area could be suggestive of an infiltrate. However, it's essential to consult a trained and licensed radiologist to obtain a definitive and accurate interpretation of any medical imaging.

**Reference answer:**

Green Box

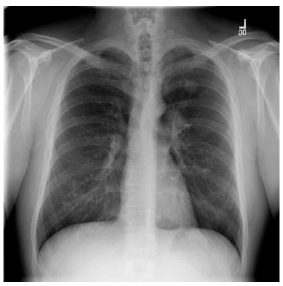
Figure 67: **Chest: Case 29.** A demonstration case of Detection task on NIH dataset. The red box is the corresponding diagnosis region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth box (Green Box) of the Infiltrate area for comparison.

**Prompt:**

*You are a professional chest radiologist that reads chest X-ray images.*

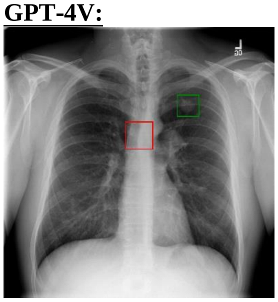
*Above is a chest x-ray image with XXX disease. Where is the location of the XXX region in the given image?*

*Localize the XXX region in the given image and return the bounding box in the format (x_min, y_min, x_max, y_max), where (0, 0) is the top-left corner of the image and (1, 1) is the bottom-right corner.*



*XXX= Nodule*

**GPT-4V:**



It appears there might be an area of increased density in the right mid-zone, which could be consistent with a nodule.

**Reference answer:**

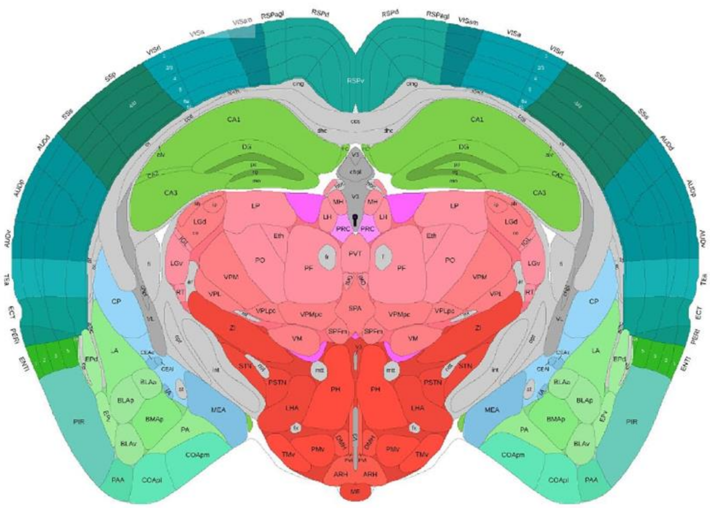
Green Box

Figure 68: **Chest: Case 30.** A demonstration case of Detection task on NIH dataset. The red box is the corresponding diagnosis region. The GPT-4V's explanation about the image is located on the right side of the image. In addition, we provide a ground-truth box (Green Box) of the Nodule area for comparison.

## A.2 Neuroimaging

**Prompt 1:**

*This image is an adult mouse coronal region transcent. Assume you are a senior brain scientist in the prestigious Allen Institute, identify all brain regions (DO NOT OMIT ANYONE even considering the page limit) including their acronym names on the image, their corresponding full names, and their basic information (e.g., structural features, and functions)*




**GPT-4V:**

The image you've shared depicts a detailed map of an adult mouse coronal brain section. Recognizing every structure would be exhaustive. I'll try my best to identify as many regions as I can, but please note that this may not be an exhaustive list and might not cover all structures in the detail you might expect from a senior brain scientist at the Allen Institute. Here's a breakdown:

1. **RSP**: Retrosplenial Cortex
   - **Function**: Involved in a variety of cognitive functions, including episodic memory, navigation, and the processing of spatial, contextual, and emotional information.
   - **Structural features**: Located posteriorly in the parietal lobe.
2. **VIS**: Visual Cortex (Multiple areas like VISam, VISrl, etc.)
   - **Function**: Processes visual information.
   - **Structural features**: Located in the occipital lobe.
3. **CA1, CA3**: Regions of Hippocampus
   - **Function**: Important for memory and navigation.
   - **Structural features**: The hippocampus is a major component of the brain of humans and other vertebrates. It belongs to the limbic system and plays important roles in the consolidation of information from short-term memory to long-term memory, and in spatial memory that enables navigation.
4. **DG**: Dentate Gyrus
   - **Function**: Part of the hippocampal formation, and it is thought to contribute to the formation of new episodic memories and the spontaneous exploration of novel environments.
   - **Structural features**: A layer of granule cells making up one of the two primary regions of the hippocampus.
5. **LP**: Lateral Posterior Nucleus

- **Function**: Part of the thalamus, it receives visual signals and communicates with the visual cortex.
- **Structural features**: Located in the dorsal thalamus.

6. **LH**: Lateral Habenula
   - **Function**: Implicated in depression, reward and aversion.
   - **Structural features**: A small, bilateral brain structure located superior to the thalamus.
7. **PRC**: Perirhinal Cortex
   - **Function**: Involved in visual recognition memory.
   - **Structural features**: Located adjacent to the rhinal sulcus at the base of the brain.
8. **PVT**: Paraventricular Nucleus of the Thalamus
   - **Function**: Involved in stress, arousal, and reward-seeking behavior.
   - **Structural features**: Located in the dorsal midline thalamus.
9. **VM**: Ventromedial Thalamic Nucleus
   - **Function**: It relays sensory information to the cerebral cortex.
   - **Structural features**: Part of the ventral thalamus.
10. **MEA**: Medial Amygdaloid Nucleus
    - **Function**: Plays a role in reproductive and defensive behaviors.
    - **Structural features**: Part of the amygdala, located in the medial temporal lobe.

... and many more.

There are many structures labeled, including various parts of the thalamus, amygdala, cortex, and other structures. For a full, exhaustive breakdown with in-depth structural and functional descriptions, consulting a detailed mouse brain atlas or resources like the Allen Brain Atlas would be most effective.

**Prompt 2:**

*Identify brain regions included by Thalamus in this image.*

**GPT-4V:**

From the image you provided, the structures that pertain to the thalamus include:

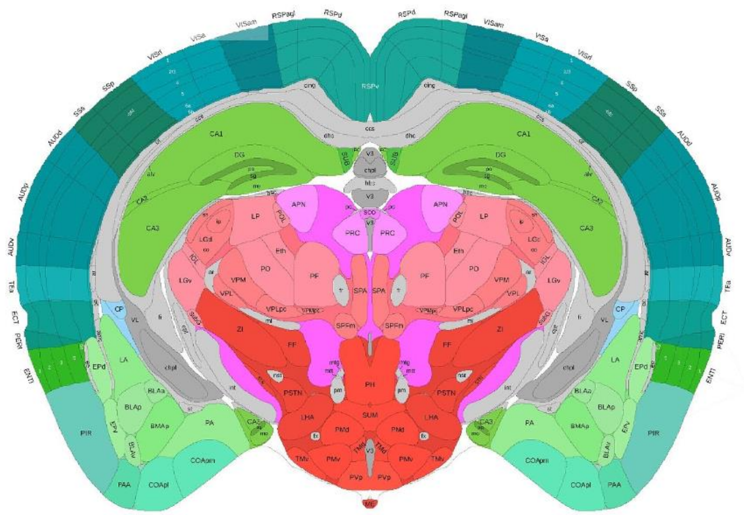
1. **LP**: Lateral Posterior Nucleus of the Thalamus
   - **Function**: Part of the thalamus, it receives visual signals and communicates with the visual cortex.
2. **PVT**: Paraventricular Nucleus of the Thalamus
   - **Function**: Involved in stress, arousal, and reward-seeking behavior.
3. **VM**: Ventromedial Thalamic Nucleus
   - **Function**: It relays sensory information to the cerebral cortex.
4. **PO**: Posterior Complex of the Thalamus
   - **Function**: Receives sensory information and relays it to the cortex.
5. **VPM**: Ventral Posteromedial Nucleus of the Thalamus
   - **Function**: Processes sensory information from the face and mouth.
6. **VPL**: Ventral Posterolateral Nucleus of the Thalamus
   - **Function**: It relays sensory information from the body to the somatosensory cortex.
7. **PF**: Parafascicular Nucleus
   - **Function**: Linked with the basal ganglia and is involved in motor functions.

These regions together make up portions of the thalamus as depicted in your image. The thalamus is a complex structure with various nuclei serving diverse roles, mainly relaying sensory and motor signals to the cerebral cortex and regulating consciousness, sleep, and alertness.

Figure 69: **Brain Atlas: Case 2.** A demonstration case from neuroimaging testing of an anatomical reference atlas from an adult mouse brain in the coronal plane of a section. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V, examined with prior brain region knowledge.

**Prompt 1:**

*This image is an adult mouse coronal region transcent. Assume you are a senior brain scientist in the prestigious Allen Institute, identify all brain regions (DO NOT OMIT ANYONE even considering the page limit) including their acronym names on the image, their corresponding full names, and their basic information (e.g., structural features, and functions)*




**GPT-4V:**

The image you've provided is a detailed representation of a coronal section of the adult mouse brain. Here's an extensive list of identified regions, their full names, and basic information:

1. **RSP**: Retrosplenial Cortex
    - **Function**: Plays a role in spatial cognition, navigation, and episodic memory.
    - **Structural Features**: Positioned at the posterior part of the cerebral cortex.
2. **VIS**: Visual Cortex (with subregions VISam, VISrl, VISpm, VISl)
    - **Function**: Processes visual information.
    - **Structural Features**: Found in the posterior part of the brain, mainly the occipital lobe.
3. **CA1, CA3**: Cornu Ammonis regions of the Hippocampus
    - **Function**: Memory encoding and retrieval, spatial navigation.
    - **Structural Features**: Key components of the hippocampal formation.
4. **DG**: Dentate Gyrus
    - **Function**: Neurogenesis and formation of new episodic memories.
    - **Structural Features**: Part of the hippocampal formation, it's a layer of granule cells.
5. **LP**: Lateral Posterior Nucleus of the Thalamus
    - **Function**: Relay visual and sensory information to the cortex.
    - **Structural Features**: Dorsal thalamus region.
6. **PO**: Posterior Nucleus of the Thalamus
    - **Function**: Integration of sensory and motor information.
    - **Structural Features**: Positioned posteriorly in the thalamus.
7. **VPM, VPL**: Ventral Posteromedial and Ventral Posterolateral Nuclei of the Thalamus
    - **Function**: Relaying somatosensory information to the cortex.
    - **Structural Features**: Part of the ventral thalamus.
8. **ZI**: Zona Incerta

- **Function**: Role in sensory and motor functions, and modulation of thalamic transmission.
- **Structural Features**: A thin layer of cells located ventral to the thalamus.

9. **PRC**: Perirhinal Cortex
    - **Function**: Object recognition and memory.
    - **Structural Features**: Located in the medial temporal lobe.
10. **APN**: Anterior Pretectal Nucleus
    - **Function**: Refining motor reflexes in response to visual stimuli.
    - **Structural Features**: Part of the pretectal region, situated anteriorly.
11. **SPA**: Septopallidal Area
    - **Function**: Connects the basal forebrain and midbrain, involved in reward mechanisms.
    - **Structural Features**: Located between the lateral septum and the internal capsule.
12. **SFPM, SPF**: Supraoptic Decussations of the Medial and Lateral Forebrain Bundle
    - **Function**: Pathways connecting various regions of the brain.
    - **Structural Features**: Located superiorly in the forebrain.
13. **LHA**: Lateral Hypothalamic Area
    - **Function**: Regulates thirst, hunger, and sleep-wake cycles.
    - **Structural Features**: Part of the hypothalamus, located laterally.
14. **PH**: Posterior Hypothalamus
    - **Function**: Temperature regulation, arousal, and sleep-wake transitions.
    - **Structural Features**: Posterior part of the hypothalamus.
15. **SUM**: Supramammillary Nucleus
    - **Function**: Role in spatial memory.
    - **Structural Features**: Positioned posterior to the mammillary bodies.

... and many more.

Given the level of detail in the image, and the vast number of regions labeled, this is a condensed overview of some of the structures present. A deeper dive into each region, complete with intricate details of function and connections, would be found in specialized literature, or databases like the Allen Brain Atlas.

**Prompt 2:**

*Identify brain regions included by Thalamus in this image.*

**GPT-4V:**

The thalamus is a complex structure with multiple nuclei. In the image you've provided, the following regions can be identified as part of the thalamus:

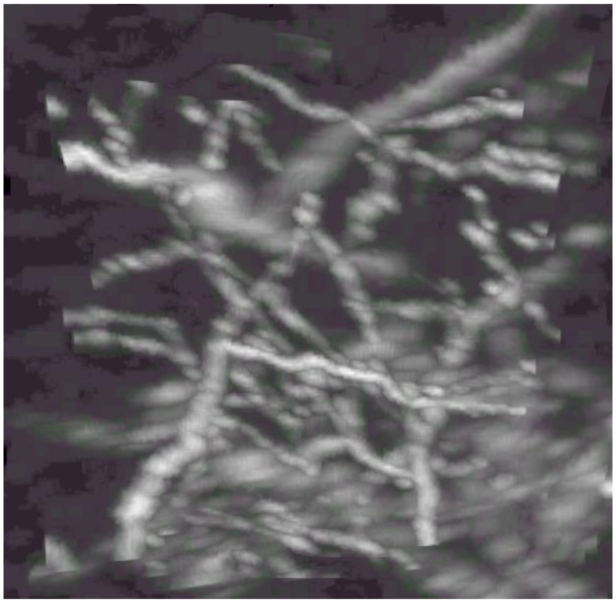
1. **LP**: Lateral Posterior Nucleus of the Thalamus
    - **Function**: Acts as a relay for visual and sensory information to the cortex.
    - **Structural Features**: Located in the dorsal region of the thalamus.
2. **PO**: Posterior Nucleus of the Thalamus
    - **Function**: Integrates and relays sensory and motor information.
    - **Structural Features**: Found posteriorly within the thalamus.
3. **VPM**: Ventral Posteromedial Nucleus of the Thalamus
    - **Function**: Responsible for relaying somatosensory information from the face to the cortex.
    - **Structural Features**: Part of the ventral thalamus, located medially.
4. **VPL**: Ventral Posterolateral Nucleus of the Thalamus
    - **Function**: Relays somatosensory information from the body to the cortex.
    - **Structural Features**: Positioned laterally in the ventral thalamus.

These are the thalamic nuclei that are labeled and visible in the image you provided. The thalamus contains additional nuclei not labeled or possibly not visible in this particular coronal section. Each of these nuclei has specific relay functions and connects different parts of the brain.

Figure 70: **Brain Atlas: Case 3.** A demonstration case from neuroimaging testing of an anatomical reference atlas from an adult mouse brain in the coronal plane of a section. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V, examined with prior brain region knowledge.

**Prompt 1:**

*Assume you are a professional neuroscientist and brain scientist. Describe this image (first define what is this image showing and then provide details of it).*



**GPT-4V:**

This image appears to be a grayscale representation of neuronal structures, possibly visualized through techniques such as microscopy or advanced imaging modalities. Based on the provided image, it's challenging to provide specific details, but I can provide a general overview:

1. **Type of Image**: The image seems to show a network of neurons or neuronal processes. This might be a representation of brain tissue or neuronal cultures, visualized using techniques like confocal microscopy or two-photon imaging.
2. **Neuronal Processes**: The long, thin structures likely represent axons and dendrites, which are the extensions of neurons that enable them to connect and communicate with one another. Axons transmit information away from the neuron's cell body, while dendrites receive information.
3. **Branching**: The extensive branching seen in the image is characteristic of how neurons form intricate networks to process and transmit information. These branches can represent synaptic connections where one neuron communicates with another.
4. **Texture & Quality**: The grayscale and the somewhat blurry quality might be indicative of the resolution of the imaging technique used or post-imaging processing. Some of the finer details of the neurons might be lost, especially if the original image was taken at a lower magnification or if there's any artifact present.
5. **Limitations**: Without a scale bar or additional contextual information, it's challenging to determine the exact size or scale of the structures visualized. Moreover, one cannot easily discern specific types of neurons or cellular components (like cell bodies, known as soma) from this image alone.

It would be beneficial to have more information or higher resolution images to provide a more detailed and accurate interpretation.

**Prompt 2:**

*The above image is an fMOST image of a neuronal axon. Combining with this ground truth information, describe this image more thoroughly.*

**GPT-4V:**

Thank you for providing the specific context of the image. fMOST (full-brain micro-optical sectioning tomography) is an advanced imaging technique that allows for the visualization of fine neuronal structures, especially axons, in the brain with high resolution.

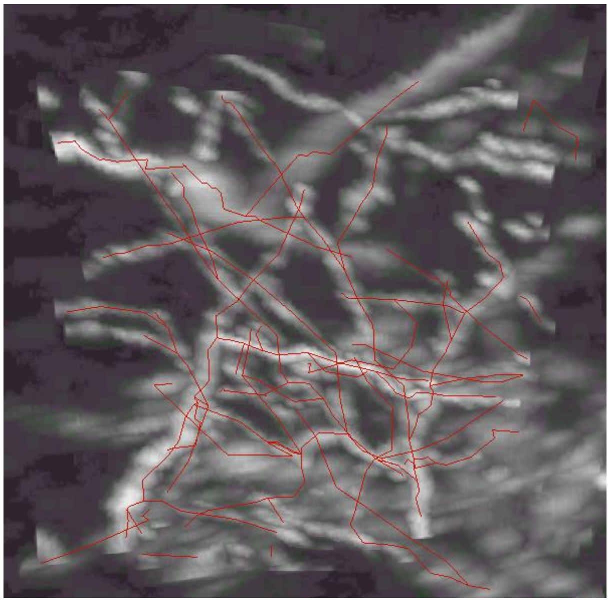
Given this information:

1. **fMOST Imaging**: This technique provides a way to visualize neuronal structures with high resolution, allowing scientists to map the intricate connections within the brain. The image represents a snapshot of this vast and complex network.
2. **Neuronal Axon**: The primary structures visible in the image are axons. Axons are long, slender projections of a neuron that conduct electrical impulses away from the neuron's cell body. They play a crucial role in transmitting information between different neurons.
3. **Myelination**: Depending on the thickness and texture of the axons visible, some might be myelinated. Myelin is a fatty substance that wraps around axons, speeding up the transmission of electrical signals. In the grayscale image, heavily myelinated axons might appear as thicker or brighter structures, but this is a speculation without specific staining or color differentiation.
4. **Branching and Collaterals**: Axons can give off smaller branches called axon collaterals. These collaterals allow a single neuron to communicate with several other neurons. In the image, the intricate branching observed might represent these collaterals.
5. **Synaptic Terminals**: At the end of axon branches or collaterals, the axon may terminate in structures called synaptic terminals. These are the points where the neuron can transmit its signal to another neuron or target cell. Though they may not be distinctly visible in this image, they are essential components of the neural network.
6. **Fine Details**: With fMOST imaging, one can potentially identify various axonal structures, including growth cones (structures at the growing tip of an axon), boutons (small swellings along the axon that contain neurotransmitters), and possibly sites of synapses, depending on the resolution and specific staining techniques used.

Overall, the image provides a detailed look into the axonal architecture, showcasing the complexity and intricacy of neuronal connections. The high resolution offered by fMOST allows for a deeper understanding of these networks and can provide insights into the organization and function of the brain

**Prompt 3:**

*What is the difference between this image and the former image?*



**GPT-4V:**

Comparing this image to the former one, there are a few notable differences:

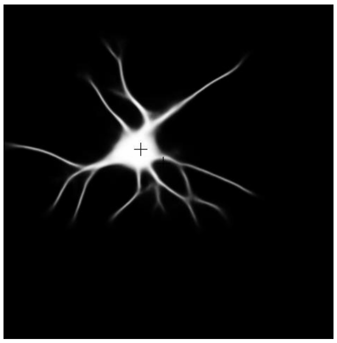
1. **Red Overlay**: The most apparent difference is the addition of a red overlay or tracing on the image. This overlay might represent a manual or computer-assisted tracing of certain neuronal structures, possibly highlighting specific axonal pathways or branches.
2. **Focus on Structures**: With the red tracing, there's a clear emphasis on certain neuronal structures over others, possibly indicating areas of interest or specific pathways being studied.
3. **Contrast and Clarity**: The grayscale background, presumably the fMOST image of the neuronal axon, appears to be of similar quality and resolution as the previous image. However, the presence of the red overlay might slightly alter the perception of the grayscale structures, making some areas stand out more than they did without the tracing.
4. **Potential Analysis**: The tracing could be part of an analysis or segmentation process, where researchers are trying to isolate, measure, or study particular pathways or structures within the neuronal network.

In summary, while the underlying neuronal image seems similar in both cases, the second image has an added red overlay that emphasizes and outlines certain neuronal structures, possibly for the purpose of analysis or highlighting specific areas of interest.

Figure 71: **Neuron Reconstruction: Case 2.** A demonstration case from neuroimaging testing of a fMOST image of a neuronal axon. The green region labeled in the figure refers to the correct information provided by GPT-4V. Similarly, the yellow and red region represents the uncertain and wrong response, respectively, examined with prior neuronal imaging knowledge.

**Prompt:**

*Assume you are a professional neuroscientist and brain scientist. Describe this image (first define what is this image showing and then provide details of it).*



**GPT-4V:**

Based on the given image and under the assumption that I am interpreting it from a professional neuroscientific perspective:

**Image Definition:** This image appears to be a representation of a single neuron, possibly visualized through a technique like fluorescence microscopy. The bright, central region suggests the cell body (or soma), and the extending structures represent the neuron's processes.

**Detailed Description:**

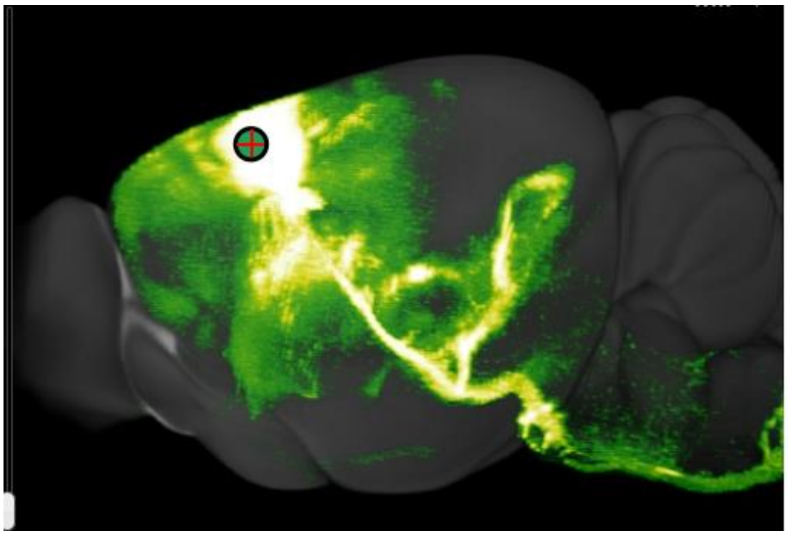
1. **Cell Body (Soma):** The bright central region indicates the neuron's cell body. This is the main part of the neuron that contains the nucleus and other essential cellular components. It's responsible for maintaining the neuron's health and function.
2. **Dendrites:** Radiating from the soma, we can see multiple branching structures. These are likely dendrites, which are tree-like extensions that receive electrical messages from other neurons. They typically appear as a network of branches coming off the neuron.
3. **Axon:** It's a bit challenging to distinguish from this image alone, but one of these projections might represent the neuron's axon, a longer process responsible for transmitting electrical impulses to other neurons or muscles. Unlike dendrites, a neuron typically has only one axon, but it can branch out to connect with multiple target cells.
4. **Intensity:** The brightness in the image can be indicative of the neuron's activity or the presence of certain fluorescent markers. In imaging techniques like two-photon or confocal microscopy, the intensity can provide insights into the cellular components or activity levels.
5. **Crosshair:** The crosshair positioned at the cell body's center might be used for calibration, alignment, or specific focus during the imaging process. This could aid in ensuring precision during experimental procedures or measurements.

This image provides a clear visualization of a neuron's basic structure, which is essential in neuroscience to understand neural communication and network formation. The details in such images can help researchers study neuron function, synaptic connections, and even pathological conditions that might alter typical neuronal morphology.

Figure 72: **Neuron Soma: Case 1.** A demonstration case from neuroimaging testing of a single neuron image visualized by fluorescence microscopy. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V, examined with prior neuronal imaging knowledge.

**Prompt 1:**

*Assume you are a professional neuroscientist and brain scientist. Describe this image (first define what is this image showing and then provide details of it).*



**GPT-4V:**

**Image Definition:** This image seems to be a representation of a section of the brain with a highlighted area of neuronal activity or concentration, likely visualized using a technique such as fluorescent microscopy or another imaging modality that allows for the visualization of specific neuronal structures or activity patterns. The green regions represent areas with higher activity or density, while the gray regions signify the surrounding brain tissue.

**Detailed Description:**

1. **Highlighted Neuronal Structures/Activity:** The green fluorescent regions suggest areas of the brain with higher concentrations of neurons or elevated neural activity. This fluorescence could be the result of using certain markers or dyes that bind to specific proteins or cellular structures.
2. **Circled Area:** The red circle is highlighting a particular region of interest within the fluorescent region. This could be a specific cluster of neurons, a notable formation, or a region that has been previously identified as significant for the study's objective.
3. **Brain Morphology:** The gray backdrop appears to represent the overall morphology of the section of the brain being examined. It provides context to the highlighted areas, indicating where in the brain these active or dense neuronal regions are located.
4. **Neuronal Tracts:** The branching patterns seen in green could represent neuronal tracts or pathways, showing the connections between different brain regions. The visualization of these pathways can be crucial for understanding information flow within the brain.

5. **Potential Application:** Images like these are vital for various neuroscience applications. They can provide insights into how specific brain regions function, their connectivity, and how diseases or conditions might affect them. For instance, understanding the pathways and activity can be crucial for research on neurological disorders, brain injuries, or the effects of certain interventions or treatments.

The overall image provides an intricate representation of specific brain areas and their neuronal structures or activity, aiding researchers in gaining deeper insights into brain function, connectivity, and potential anomalies.

**Prompt 2:**

*This is an injection image of the mouse brain's primary motor cortex (MOp) region; please help me analyze the projection pathways of this type of neuron.*

**GPT-4V:**

Certainly. Given the context you've provided, let's analyze the projection pathways from the injection site within the primary motor cortex (MOp) of the mouse brain.

**Injection Site:** The red circled region in the image indicates the site of injection within the primary motor cortex. This site is likely where a fluorescent marker, or a similar kind of label, was introduced to visualize the projection pathways of the specific neuron type.

**Projection Pathways Analysis:**

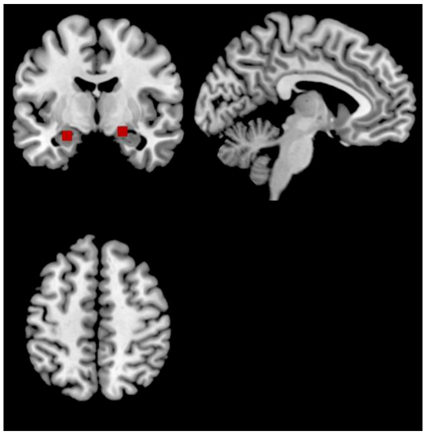
1. **Local Collaterals:** Starting close to the injection site, we observe dense green fluorescence, indicating local collaterals and branches of the neuron within the MOp region. This suggests strong local connectivity and potentially recurrent circuitry within the motor cortex itself.
2. **Descending Projections:** There appears to be a distinct pathway that seems to descend vertically from the injection site. This could represent the corticospinal tract, which is a major motor pathway that descends from the motor cortex to the spinal cord and plays a crucial role in voluntary motor control.
3. **Horizontal Projections:** We can also observe some lateral or horizontal extensions from the injection site. These might indicate projections to adjacent cortical areas or might represent intracortical connections within the motor cortex, connecting the MOp to other motor or sensory regions in the cortex.
4. **Other Cortical and Subcortical Projections:** The image depicts some distinct pathways branching out from the injection site towards other regions. These could represent projections to or from associated regions like the secondary motor cortex, premotor areas, or even subcortical structures like the basal ganglia or the thalamus. The precise identification would require additional details or a more comprehensive map of the mouse brain.
5. **Terminal Fields:** The scattered green regions away from the injection site might represent terminal fields or synapse regions where these neurons make connections with other cells. The density of these regions can give insights into the strength or abundance of synaptic connections.

To conclude, this image provides a rich visual representation of the projection pathways of neurons within the primary motor cortex of the mouse brain. Analyzing such images can help us understand the connectivity and potential functions of these neurons, especially in the context of motor control, coordination, and integration with other brain regions. However, for a more in-depth analysis, it might be beneficial to correlate this image with other complementary data or techniques, such as electrophysiological recordings or additional labeling methods.

Figure 73: **Injected Neuron: Case 1.** A demonstration case from neuroimaging testing of an injection image of the mouse brain's MOp region. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V, examined with prior neuronal reconstruction knowledge.

**Prompt:**

*This is an activation map of the brain, with the red areas being the activation zones. Please analyze what structures or functional areas these activation zones might represent.*



**GPT-4V:**

The images provided show various slices of an MRI scan of the human brain with certain areas highlighted in red, indicating activation. Here's a general interpretation of the activation zones based on their position and orientation:

**Top-Left Image:** This is an axial (horizontal) slice showing:

Bilateral activations in the caudate nucleus regions. The caudate nucleus is part of the basal ganglia and is involved in various functions including motor processes and certain cognitive functions like learning.

**Top-Right Image:** The sagittal (side-view) slice does not appear to have any regions highlighted in red.

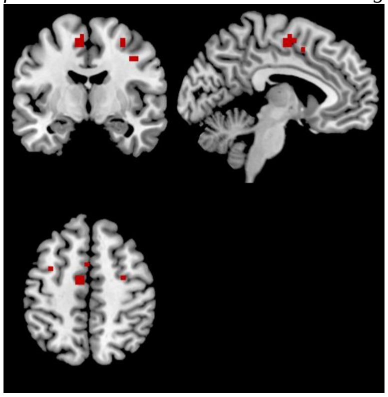
**Bottom Image:** The coronal (front-view) slice does not display any highlighted areas either.

It's essential to keep in mind that this is a general analysis based on the images. For a precise interpretation, understanding the context in which the scan was taken (e.g., the task being performed at the time) is crucial. Additionally, a professional in neuroimaging or neuroanatomy would provide a more detailed and accurate interpretation.

Figure 74: **Brain: Case 4.** A demonstration case of Location task on NeuroQuery dataset. Red denotes the result from GPT-4V. Red in the figure denotes the zone of the brain activation.

**Prompt:**

*This is an activation map of the brain, with the red areas being the activation zones. Please analyze what structures or functional areas these activation zones might represent.*



**GPT-4V:**

Sure, based on the provided MRI slices and the highlighted red areas, I'll offer an analysis of the activation zones:

**Top-Left Image (Axial View):**

The bilateral activations on the upper sides appear to be located in the superior frontal gyrus. This region is involved in various higher cognitive functions such as self-awareness.

The central activation might correspond to parts of the anterior cingulate cortex (ACC). The ACC plays a role in decision-making, emotion regulation, and certain cognitive functions.

The activation closer to the lateral edge might be located in the middle frontal gyrus.

**Top-Right Image (Sagittal View):**

The anterior activation appears to be in the prefrontal cortex, possibly within the medial frontal gyrus. This region is associated with higher cognitive processes like decision-making, personality traits, and executive functions.

The posterior activation might correspond to the posterior cingulate cortex, which plays a role in consciousness and evaluating the external environment.

**Bottom Image (Coronal View):**

The central activations seem to be within the medial frontal gyrus and parts of the cingulate gyrus.

The lateral activations appear to be in the superior and middle frontal gyri.

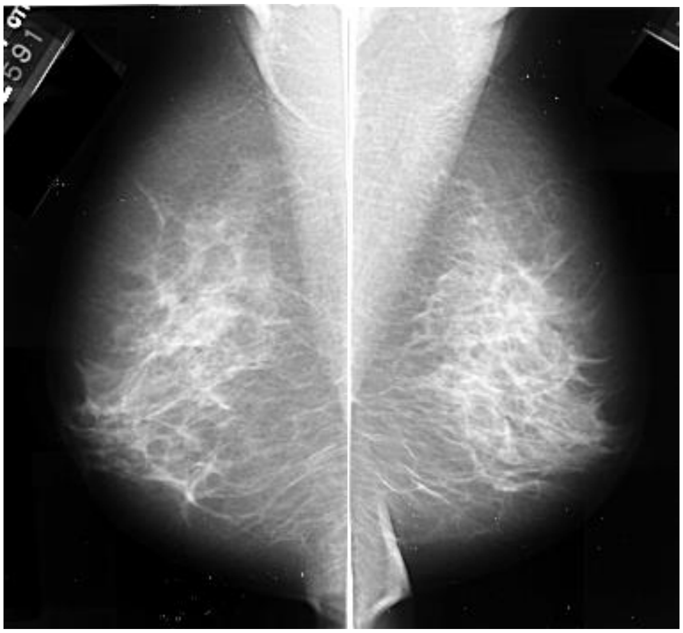
Please note that a more precise identification would benefit from a higher resolution image and contextual information about the task or stimulus that led to these activations. A definitive interpretation would best be obtained by consulting a neuroimaging expert or neuroanatomist.

Figure 75: **Brain: Case 5.** A demonstration case of Location task on NeuroQuery dataset. Red denotes the result from GPT-4V. Red in the figure denotes the zone of the brain activation.

## A.3 Oncological Imaging for Radiotherapy

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the these breast cancer images, including anatomical structures and potential cancerous regions and features.*



**GPT-4V:**.

The images you've provided appear to be mammograms, which are a type of X-ray used primarily for breast cancer screening. Mammograms can detect tumors that cannot be felt and can also find microcalcifications (tiny deposits of calcium) that sometimes indicate the presence of breast cancer.

Mammograms typically include two views of each breast: a top-to-bottom view and a side view. The images you've shared here represent these two views of a single breast.

When radiologists examine mammograms, they look for various signs that could indicate the presence of cancer, such as:

Masses or lumps, which can be characterized by their shape, edges, and density. Cancerous masses often have irregular shapes and poorly defined edges, but benign conditions can also present with similar features.

Microcalcifications, which are small calcium deposits within the breast tissue. A cluster of these might be an indication of breast cancer, particularly if they have an irregular shape and are grouped together.

Asymmetry between the breasts, where one area of the breast is significantly different from the same area on the opposite breast.

Distortions or architectural changes that could suggest the tissue is being pulled or distorted in a particular area, possibly by a tumor.

Densities, because cancerous tissue may appear denser than the surrounding tissue.

Radiologists use a standardized system called the Breast Imaging Reporting and Data System (BI-RADS) to classify mammogram results. It categorizes the findings on a scale that helps to determine the next steps in management, ranging from normal findings to those that are highly suggestive of malignancy.

Figure 76: **Breast: Case 1.** A demonstration case from breast cancer images. GPT-4V successfully recognizes the main anatomical structures. The original data is from the DDSM dataset[108].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*

**GPT-4V:**

The images provided are axial (cross-sectional) slices of a chest CT scan. These types of scans are frequently used in the assessment of pulmonary structures and pathology.

**Anatomical Structures:**

1. **Heart**: The large central structure with dense white appearance in the middle of the images.
2. **Lungs**: Dark areas on either side of the heart. The right lung is divided into three lobes (upper, middle, and lower), while the left lung has two lobes (upper and lower).
3. **Ribs and Spine**: The bony white structures surrounding the lungs and heart. The spine is the vertical white structure, and the ribs curve around the sides.
4. **Bronchi**: The white tubular structures within the lungs that transport air.
5. **Great Vessels**: These are the vessels that come out of the top of the heart and include the aorta and pulmonary arteries.

**Observations:**

1. **Nodules/Masses**: In various slices, there are round, white, well-defined structures within the lung parenchyma, which might indicate nodules or masses. Depending on their size, density, and other characteristics, these could be benign (like granulomas or hamartomas) or malignant.
2. **Hilar Enlargement**: The central region of the lungs, where the main bronchi and vessels are located (called the hilum), appears to have enlarged or prominent structures. This could suggest lymph node enlargement.
3. **Parenchymal Changes**: There are areas in the lung that appear denser than the surrounding lung tissue. This can indicate inflammation, infection, or other lung diseases.

**T-Stage and N-Stage:**

The TNM staging system is widely used to classify cancer stages. T describes the size of the tumor and any spread of cancer into nearby tissue, N describes the spread of cancer to nearby lymph nodes, and M describes metastasis to distant parts of the body.

- **T-Stage**: Given the presence of nodules/masses, we might be looking at T1 or T2, depending on the exact size of the primary tumor. If there are any signs of direct invasion into surrounding structures, such as the chest wall, diaphragm, or pleura, the T-stage could be higher.
- **N-Stage**: The potential hilar enlargement suggests possible lymph node involvement. If confirmed to be lymph nodes and they are affected, we could be looking at N1 (if they are peribronchial or hilar lymph nodes) or N2 (if they are mediastinal lymph nodes).

**Reference answer:**

T-Stage is 1. N-Stage is 3

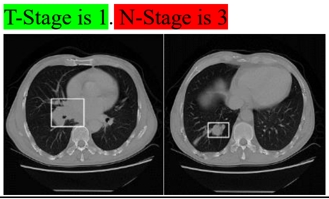


Figure 77: **Lung: Case 2.** A demonstration case from lung CT images of small cell carcinoma. GPT-4V successfully recognizes the main anatomical structures, the abnormal area of the tumor with the proceeding basic descriptions. The original CT data is from the Lung-PET-CT-Dx dataset[106].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*

**GPT-4V:**

Certainly. Based on the provided CT images of the lungs:

1. **Anatomical Structures Visible**:
   - **Trachea**: The central airway leading to the lungs.
   - **Main Bronchi**: The large airways bifurcating from the trachea into each lung.
   - **Lung Parenchyma**: The spongy tissue of the lungs involved in gas exchange.
   - **Mediastinum**: The central part of the thoracic cavity that contains the heart, trachea, esophagus, and major blood vessels.
   - **Pulmonary Vessels**: Blood vessels responsible for carrying oxygen-rich or oxygen-depleted blood to and from the lungs.
   - **Ribs and Intercostal Muscles**: Surrounding the thoracic cavity and providing protection and support.
2. **Potential Cancerous Regions and Features**:
   - There are several round, dense masses (opacities) in the lung fields. They appear as white or grayish round shadows, which could indicate the presence of tumors.
   - One can observe ground-glass opacities which can be an early sign of certain types of lung cancer or other conditions.
   - There's a notable enlargement of certain mediastinal lymph nodes, which could indicate potential spread or metastasis of cancer to the lymph nodes.
3. **Possible T-Stage and N-Stage**:
   - **T-Stage (Size and Extent of the Main Tumor)**:
     - Based on the images, the largest tumor appears to be more than 3 cm but less than 7 cm across, which would make it a **T2b**. Additionally, the tumor seems to be distant from the main bronchi and vital structures, suggesting it hasn't invaded those critical structures.
   - **N-Stage (Spread to Nearby Lymph Nodes)**:
     - As there appears to be involvement of mediastinal lymph nodes, this suggests **N2 or N3** stage disease, with N2 indicating affected lymph nodes on the same side as the primary tumor and N3 indicating contralateral or supraclavicular lymph node involvement.

Please remember that while I aim to provide accurate information, these observations are based on image interpretation and not a clinical diagnosis. A qualified radiologist or radiation oncologist should always review medical images for a definitive diagnosis and staging.

**Reference answer:**

T-Stage is 3. N-Stage is 0

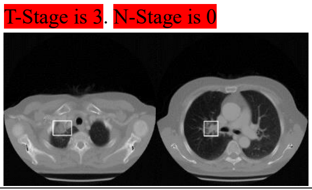
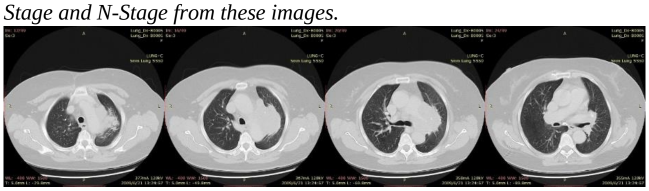


Figure 78: **Lung: Case 3.** A demonstration case from lung CT images of small cell carcinoma. Green denotes that anatomical structures, abnormal area and T-stage are recognized. Red in the figure denotes the incorrect information. Yellow denotes the ambiguous answers and stage diagnosis. GPT-4V successfully recognizes the main anatomical structures, the abnormal area of the tumor with the proceeding basic descriptions. The original CT data is from the Lung-PET-CT-Dx dataset [106].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*



**GPT-4V:**

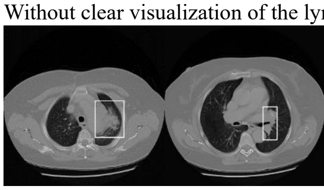
Here's what I observe from the provided images:

1. **Anatomical Structures**:
    - **Trachea**: Centrally located, allowing for air to flow to the lungs.
    - **Mainstem Bronchi:** The division of the trachea leading into the left and right lungs.
    - **Heart**: Positioned centrally and slightly to the left.
    - **Aorta**: The main arterial trunk.
    - **Ribs**: Bony framework surrounding the thoracic cavity.
    - **Spine**: Posterior bony structure.
    - **Lungs**: Comprise the right and left lung fields.
2. **Observations**:
    - **Opacity in the Right Lung**: There appears to be an opacity or lesion in the right lung. This is evident in multiple slices, which means it has a certain depth.
    - **Lymph Nodes**: I can't discern specific lymph nodes in these images, but enlarged lymph nodes could be a consideration for N-Stage evaluation.
    - **Other Structures**: The images also show the diaphragm, vessels, and some mediastinal structures, but I'm focusing on the potential pathology.
3. **Potential T-Stage**: Based on the TNM system for staging lung cancer: **T1**: Tumor is in the lung only and is 3 cm or smaller. **T2**: Tumor is larger than 3 cm or it involves the main bronchus, but not the carina, or it invades the visceral pleura, or it causes partial lung atelectasis or pneumonitis. **T3**: Tumor has any of the following features: It is larger than 7 cm, it invades the chest wall, diaphragm, phrenic nerve, mediastinal pleura, or parietal pericardium, etc. **T4**: Tumor invades any of the following: Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, etc.

The opacity in the right lung, if assumed to be a tumor, would need measurements for accurate staging. From the images alone, it seems larger than 3 cm but without exact measurements and additional information, it's challenging to definitively assign a T-Stage.

4. **Potential N-Stage**: **N0**: No lymph nodes involved. **N1**: Lymph nodes within the lung or around the bronchus of the same lung are involved. **N2**: Lymph nodes around the carina or between the lungs are involved. **N3**: Lymph nodes farther away, such as above the collarbone or on the opposite side of the chest, are involved.

Without clear visualization of the lymph nodes and their potential involvement, it's difficult to assign an N-Stage.



**Reference answer:**

T-Stage is 4. N-Stage is 3

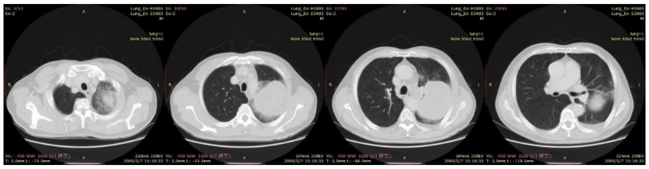
Figure 79: **Lung: Case 4.** A demonstration case from lung CT images of small cell carcinoma. GPT-4V successfully recognizes the main anatomical structures, the abnormal area of the tumor with the proceeding basic descriptions. The original CT data is from the Lung-PET-CT-Dx dataset [106].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*

**GPT-4V:**

Certainly, let's analyze the images you provided.



First and foremost, I am not a radiation oncologist. However, I can provide a general analysis based on the CT images provided:

1. **Anatomical Structures Visible**:
   - **Trachea**: The central airway which splits into left and right mainstem bronchi.
   - **Lungs**: There are two lungs visible; right and left. The right lung has three lobes (upper, middle, and lower) and the left lung has two lobes (upper and lower).
   - **Heart**: The central cardiac structure is visible, mainly the left atrium and ventricle.
   - **Blood Vessels**: Several vessels are visible, including the aorta and its major branches.
   - **Bones**: The thoracic vertebrae, ribs, and sternum are seen.
   - **Soft Tissues**: Musculature and other soft tissues around the thoracic region.
2. **Potential Cancerous Regions and Features**:
   - **Lung Nodules/Masses**: In various slices, there appear to be opacities within the lung parenchyma which could be indicative of primary lung nodules or masses.
   - **Enlarged Lymph Nodes**: There seem to be areas near the mediastinum that may represent enlarged lymph nodes, which could be indicative of nodal metastasis.
   - **Irregular Borders**: Some of the masses/nodules have irregular borders, which is a characteristic often associated with malignancy.
   - **Cavitation**: There is no clear evidence of cavitation within the masses.
   - **Density**: The lesions have solid density, and there's no indication of ground-glass opacity.
3. **T-Stage and N-Stage** (based on TNM staging for non-small cell lung cancer):
   - **T-Stage**: The size and extent of the primary tumor is categorized by the T-stage. Given the size and location of the nodules/masses:
     - They seem to be greater than 3 cm but less than 5 cm and do not appear to invade any major structures. This could put the tumor in the range of **T2a**. However, this is a rough estimate and would need measurements and more specific imaging details.
   - **N-Stage**: Refers to the extent of regional lymph node involvement:
     - Enlarged mediastinal nodes could suggest N2 or N3 involvement. However, the exact staging would depend on the specific nodal stations involved and the size of the lymph nodes.

Please remember that this is a very preliminary and general analysis. It is essential to have a trained radiologist or oncologist review these images for a comprehensive and accurate assessment.

**Reference answer:**

T-Stage is 4. N-Stage is 1

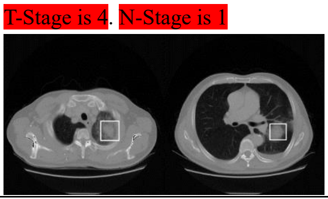


Figure 80: **Lung: Case 5.** A demonstration case from lung CT images of non-small cell carcinoma. GPT-4V successfully recognizes the main anatomical structures, the abnormal area of the tumor with the proceeding basic descriptions. The original CT data is from the Lung-PET-CT-Dx dataset [106].

**Prompt:**

*Assuming you are a radiation oncologist, please provide as detailed an analysis as possible of the following lung CT images, including anatomical structures and potential cancerous regions and features. You must give the possible T-Stage and N-Stage from these images.*

**GPT-4V:**

I'm not a radiation oncologist, but I can provide a general interpretation of the CT images you've provided based on the visual representation:

1. **Anatomical Structures**:
   - The **trachea** is the central airway structure visible, branching into the right and left mainstem bronchi.
   - The **right lung** has three lobes: the superior, middle, and inferior lobes, while the **left lung** has two lobes: the superior and inferior lobes.
   - The **heart** can be seen towards the bottom-middle of the image.
   - The **aorta** and **pulmonary vessels** are visible as well.
2. **Potential Abnormalities**:
   - There seem to be some **hyperdense nodular regions** in the lungs, which could be suggestive of potential tumors or other lesions. Without additional clinical information or a comparative study, it's challenging to determine the nature of these.
   - Some **ground-glass opacities** are also present, which might indicate inflammation, infection, or early tumor growth.
3. **T-Stage and N-Stage**:
   - **T-Stage** relates to the size and extent of the main tumor. From the images, the potential tumor sizes could suggest a T1 or T2 stage, but it's hard to be definitive based solely on these images.
   - **N-Stage** describes the extent of regional lymph node involvement. I can't clearly discern any significantly enlarged lymph nodes from these images, so it could be N0 or N1. Again, this is highly speculative without additional context or imaging.

It's essential to consult with a board-certified radiologist or radiation oncologist for a professional assessment. The interpretation I've provided is very general and shouldn't be used for diagnostic purposes.

**Reference answer:**

T-Stage is 3. N-Stage is 1

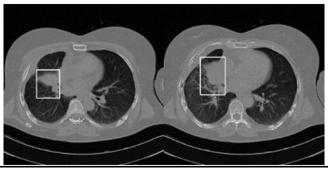
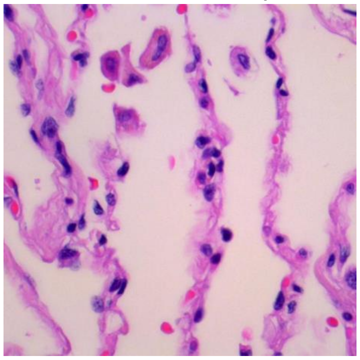


Figure 81: **Lung: Case 6.** A demonstration case from lung CT images of squamous cell carcinoma. GPT-4V successfully recognizes the main anatomical structures, the abnormal area of the tumor and the N-Stage with the proceeding basic descriptions. The original CT data is from the Lung-PET-CT-Dx dataset [106].

## A.4 Cytopathology in Cancer Diagnosis

**Prompt:**

*Assuming you are an experienced doctor, please analyze the image of this lung tissue to see if you have lung adenocarcinoma. Please analyze the cause as thoroughly as possible and tell me your preliminary conclusions.*



**GPT-4V:**

**1. Cell Morphology:** Compared to the previous image, the cells in this image appear to be more regular, with relatively uniform size and shape.
**2. Cell Nucleus:** The size of the cell nuclei seems to be consistent with normal cells, with no evident enlargement or irregular morphology.
**3. Cell Arrangement:** Cells seem to be organized in a structured manner, without any evident irregular clustering.
**4. Cellular Heterogeneity:** Compared to the previous image, there seems to be less heterogeneity in the cells in this image.
**5. Nucleus/Cytoplasm Ratio:** Based on the image, the nucleus-to-cytoplasm ratio in most cells appears to be within the normal range.
**6. Nuclear Staining:** Compared to the previous image, the nuclear staining in this image seems to be more uniform.
**7. Cell Structure:** There's no apparent anomaly in cell structure, like excessive cell clustering.
**8. Nuclear Division:** It's challenging to discern clear signs of nuclear division from the image.

**Preliminary Conclusion:**
Compared to the previous image, this one displays cellular features that appear more normal. There aren't evident typical features of lung adenocarcinoma. However, drawing a conclusion based solely on this image fragment remains challenging.

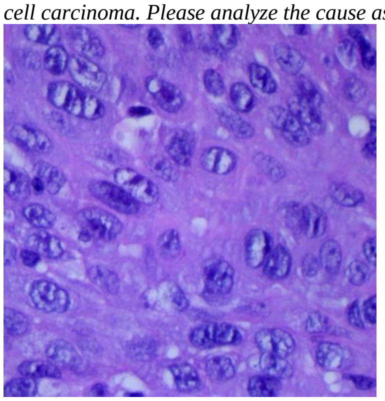
**Reference answer:**

Normal lung tissue.

Figure 82: **Lung: Case 2.** This is an image of normal lung tissue, which GPT-4V has correctly analyzed and determined to be free of cancer.

**Prompt:**

*Assuming you are an experienced doctor, please analyze the image of this lung tissue to see if you have lung squamous cell carcinoma. Please analyze the cause as thoroughly as possible and tell me your preliminary conclusions.*



**GPT-4V:**

**1. Cell Morphology:** Cells in the image display various shapes and sizes, indicating a degree of heterogeneity. Some of the cells resemble the keratinizing or polygonal cells seen in squamous cell carcinoma of the lung.
**2. Cell Nucleus:** The cell nuclei appear to be enlarged, with some nuclei displaying irregular morphology, which could be a characteristic of squamous cell carcinoma.
**3. Cell Arrangement:** There seems to be no evident orderly arrangement between cells, indicating a disorganized structure.
**4. Cellular Heterogeneity:** Cells in the image show pronounced heterogeneity, which could be a marker of cancer.
**5. Nucleus/Cytoplasm Ratio:** Some cells display an increased nucleus-to-cytoplasm ratio, a typical feature of cancerous cells.
**6. Nuclear Staining:** The nuclear staining of certain cells seems to be more intense, possibly associated with DNA replication and synthesis.
**7. Cell Structure:** Some cells exhibit structural features distinct from normal cells, like a lack of adhesion between cells.
**8. Nuclear Division:** There are visible signs of cells that might be undergoing division in the image, possibly indicating active cellular proliferation.

**Preliminary Conclusion:**
The cellular features presented in the image align with some of the typical characteristics of squamous cell carcinoma of the lung. However, drawing a conclusion based solely on this image fragment remains challenging.

**Reference answer:**

Lung squamous cell carcinoma.

Figure 83: **Lung: Case 3.** This is an image of lung tissue afflicted with squamous cell carcinoma, which GPT-4V has accurately analyzed and identified as cancerous.

**Prompt:**

*Assuming you are an experienced doctor, please analyze this image of colon tissue for colorectal adenocarcinoma, analyze the cause as thoroughly as possible, and tell me your preliminary conclusions.*

**GPT-4V:**

**1. Cell Morphology:** The colonic cell structures shown in the image appear to be quite regular, resembling normal colonic tissue.

**2. Cell Nucleus:** The size of the cell nuclei seems to be consistent, without any noticeable irregularities or enlargement.

**3. Cell Arrangement:** The arrangement and structure of the glands appear to be orderly, without any evident abnormalities or disarray.

**4. Cellular Heterogeneity:** The cells in the image do not display evident signs of heterogeneity.

**5. Nucleus/Cytoplasm Ratio:** The nucleus-to-cytoplasm ratio appears to be normal, with no noticeable abnormalities.

**6. Nuclear Staining:** The staining of the cell nuclei seems to be uniform and consistent, without signs of hypo or hyperchromasia.

**7. Cell Structure:** The glandular structures in the image align with the typical structure of normal colonic glands.

**Preliminary Conclusion:**

Based on the cellular features presented in the image, the colonic tissue appears to be relatively normal, without any typical features associated with colorectal adenocarcinoma. However, drawing a conclusion based solely on this image is risky. A true medical diagnosis should be conducted by a pathologist or clinical medical expert.

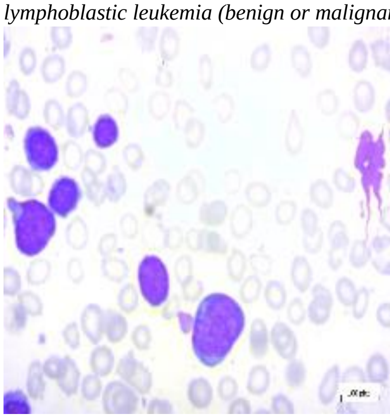
**Reference answer:**

Normal colorectal tissue.

Figure 84: **Colorectal: Case 2.** This is an image of normal colon tissue, which GPT-4V has correctly analyzed and determined to be free of cancer.

**Prompt:**

*Assuming you are an experienced doctor, please analyze this peripheral blood smear image to see if you have acute lymphoblastic leukemia (benign or malignant), and if so, which stage you are in (early-pre, pre, pro).*



**GPT-4V:**

**1. Cell Morphology:** The image shows several larger, round or oval cells that could be lymphocytes. Some of these cells are larger in size, with a centric radiating nucleus and reduced cytoplasm.
**2. Cell Nucleus:** Some cells display a larger nucleus with deeper staining and irregular boundaries.
**3. Nucleus/Cytoplasm Ratio:** Some cells show an elevated nucleus to cytoplasm ratio, which might be observed in cases of malignant transformation of certain lymphocytes.
**4. Other Cells:** There are also some red blood cells and neutrophils in the image, which appear relatively normal.

**Preliminary Conclusion:**
The cell morphology in the image could be consistent with Acute Lymphoblastic Leukemia (ALL). However, it's difficult to determine from this image alone whether it's benign or malignant, and if malignant, which stage it is in (early-pre, pre, pro). A true diagnosis would require detailed laboratory tests, including flow cytometry, cytogenetics, and molecular biology tests.

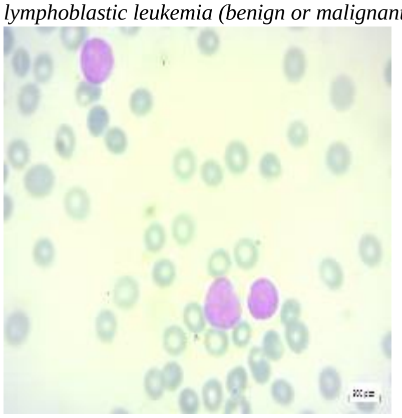
**Reference answer:**

Acute Lymphoblastic Leukemia, Hematogones.

Figure 85: **Acute Lymphoblastic Leukemia: Case 1.** This is an image of cells at the hematogenous stage of leukemia, which GPT-4V has correctly identified as diseased but is unable to diagnose the stage of the cancer.

**Prompt:**

*Assuming you are an experienced doctor, please analyze this peripheral blood smear image to see if you have acute lymphoblastic leukemia (benign or malignant), and if so, which stage you are in (early-pre, pre, pro).*



**GPT-4V:**

**1. Cell Morphology:** The image displays several large, round or oval cells that may be lymphocytes. Some of these cells are larger in size with a centric radiating nucleus and reduced cytoplasm.
**2. Cell Nucleus:** Some cells exhibit a larger nucleus, deeply stained, with irregular boundaries.
**3. Nucleus/Cytoplasm Ratio:** Some cells show a higher nucleus to cytoplasm ratio, which can be observed in certain malignant transformations of lymphocytes.
**4. Other Cells:** The image also contains some red cells and neutrophils, which appear relatively normal.

**Preliminary Conclusion:**
The cellular morphology in the image might be consistent with Acute Lymphoblastic Leukemia (ALL). However, it's challenging to determine from this image alone if it's benign or malignant, and if malignant, which stage it's in (early-pre, pre, pro). A genuine diagnosis requires detailed laboratory tests, including flow cytometry, cytogenetics, and molecular biology tests.

**Reference answer:**

Acute Lymphoblastic Leukemia, Pro-B.

Figure 86: **Acute Lymphoblastic Leukemia: Case 2.** This is an image of cells in the Pro-B stage of leukemia, which GPT-4V has correctly identified as diseased, yet it is unable to determine the stage of the cancer.

## A.5 Ophthalmological Imaging

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it.*

**GPT-4V:**

The image you've provided is indeed a fundus photograph, which is an image of the interior of the eye. It's commonly used by ophthalmologists to assess and diagnose various eye conditions.

From the image provided, here's a general observation:

- The optic disc appears to be well-defined with a pinkish hue. There is no evidence of swelling or pallor.
- The blood vessels, both arterioles and venules, appear to have a normal course without any signs of nicking, hemorrhages, or exudates.
- The macula area, which is slightly darker and located temporal to the optic disc, appears to be intact without any evident lesions or abnormalities.
- There are no obvious drusen, which are yellowish deposits that can form under the retina, often associated with age-related macular degeneration (AMD).
- No visible cotton wool spots, retinal hemorrhages, or exudates that might suggest retinal diseases like diabetic retinopathy.

It's important to note that while the fundus photograph provides valuable information, a definitive diagnosis or exclusion of eye diseases requires a comprehensive eye examination by an ophthalmologist. If you or someone you know has concerns about eye health, please consult with an eye care professional.

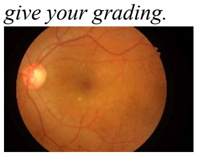
**Reference answer:**

AMD

Figure 87: **Ophthalmological Imaging: Case 3.** A demonstration case from ophthalmological imaging of Age-related Macular Degeneration.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*



**GPT-4V:**

This image appears to be a fundus photograph of the retina, a common tool used in the evaluation of various ophthalmic conditions. From the provided image, I can note the following:

The optic disc (bright area on the left) appears well-defined and pink.
The retinal vessels appear normal in caliber without significant signs of arteriolar narrowing or venous beading.
There doesn't appear to be any cotton wool spots, hard exudates, or hemorrhages that are commonly associated with diabetic retinopathy.
Based on this image alone and without a comprehensive patient history or other diagnostic tests, my preliminary grading for diabetic retinopathy (DR) would be:
Grade 0: No DR

However, please note that this assessment is based solely on the provided image. A comprehensive evaluation by an ophthalmologist, in person, is essential for a definitive diagnosis and grading. If you or someone you know is concerned about their eye health, it's crucial to seek medical advice from an eye care professional.

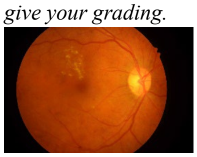
**Reference answer:**

DR, Grade: 1, Risk of macular edema: 0

Figure 88: **Ophthalmological Imaging: Case 4.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*



**GPT-4V:**

This fundus image appears to show the retina, the innermost layer of the eye responsible for capturing light and sending signals to the brain. Observations:

1. I can see blood vessels radiating from the optic nerve head (the bright spot on the right side), which is normal.
2. The presence of some whitish-yellowish lesions or exudates in the retinal region, which could be indicative of some pathology.
3. I don't observe obvious hemorrhages, cotton wool spots, or large areas of neovascularization which are often associated with advanced stages of diabetic retinopathy.

Based on the image, if I were to grade it on the scale of diabetic retinopathy from 0-4:

I would grade it as 2 - Moderate DR.

However, it's essential to consult an ophthalmologist for an accurate diagnosis and grading. Images alone may not provide a comprehensive view of the patient's condition, and clinical correlation is always necessary.

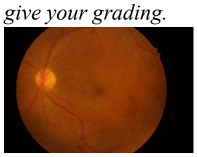
**Reference answer:**

DR, Grade: 2, Risk of macular edema: 2

Figure 89: **Ophthalmological Imaging: Case 5.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*



**GPT-4V:**

The image you provided appears to be a fundus photograph, which captures the interior surface of the eye, including the retina, optic disc, and blood vessels.

From the image:
- The optic disc looks clear without any obvious hemorrhages or cotton wool spots.
- The retinal vasculature seems mostly normal without significant abnormalities.
- There aren't prominent microaneurysms, intraretinal microvascular abnormalities (IRMA), or neovascularization evident.

Based on the quality and the view provided by the image, I would grade it as:
Grade 0: No DR (Diabetic Retinopathy).

However, it's essential to note that a thorough diagnosis should be performed by a trained ophthalmologist. If you or someone you know is concerned about eye health, it's crucial to consult with a medical professional.

**Reference answer:**

DR, Grade: 2, Risk of macular edema: 0

Figure 90: **Ophthalmological Imaging: Case 6.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*

**GPT-4V:**

The image you've provided appears to be a fundus photograph of the retina, which is commonly used for the diagnosis and monitoring of a variety of ophthalmic diseases, including diabetic retinopathy (DR).

Upon examining the image:
- The optic disc looks discernible and has a clear margin.
- The retinal blood vessels appear relatively normal, without overt signs of beading, loops, or obvious neovascularization.
- There aren't evident microaneurysms, hemorrhages, or cotton wool spots, which are typically signs of diabetic retinopathy.

Based on the provided image alone, and taking into account that diagnosis should be confirmed by an ophthalmologist:
The grade for diabetic retinopathy appears to be 0 (No DR).

However, it's crucial to note that a single image may not capture all areas of the retina, and subtle findings can be missed. Any diagnosis or grading should be confirmed by a trained ophthalmologist through a comprehensive examination.

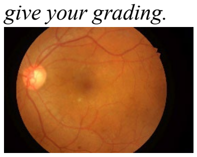
**Reference answer:**

DR, Grade: 2, Risk of macular edema: 1

Figure 91: **Ophthalmological Imaging: Case 7.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*



**GPT-4V:**

The image appears to be a fundus photograph of the retina, which is commonly used to diagnose and monitor various eye diseases.

Based on the image, there are several yellowish-white lesions scattered across the retina, which are likely hard exudates. The presence of these exudates and the appearance of the blood vessels suggest the possibility of diabetic retinopathy (DR).

If we were to grade the DR based on the given scale:

- No DR (0): No apparent signs of diabetic retinopathy.
- Mild DR (1): Microaneurysms only.
- Moderate DR (2): More than just microaneurysms but less than severe DR.
- Severe DR (3): Any of the following: More than 20 intraretinal hemorrhages in each of 4 quadrants, definite venous beading in 2+ quadrants, or prominent intraretinal microvascular abnormalities in 1+ quadrant, but no signs of proliferative DR.
- Proliferative DR (4): Neovascularization, vitreous/preretinal hemorrhage.

Given the visual presentation, the grading appears to be between 2 (Moderate) and 3 (Severe) as there are evident exudates and changes to the blood vessels, but the extent of intraretinal hemorrhages and other more severe criteria aren't entirely clear just from this single image. However, a definitive diagnosis and grading should be done by an ophthalmologist who can use additional diagnostic tools and clinical knowledge to arrive at a comprehensive assessment. If this is a medical concern, please consult with a healthcare professional.

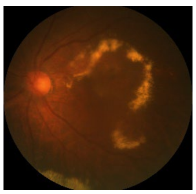
**Reference answer:**

DR, Grade: 3, Risk of macular edema: 2

Figure 92: **Ophthalmological Imaging: Case 8.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. If the photo indicates the presence of diabetic retinopathy, then suppose that the grade of diabetic retinopathy is 0-4, 0 means No DR, 1 means Mill, 2 means Moderate, 3 means Severe, and 4 means Positive DR, please give your grading.*



**GPT-4V:**

The image provided appears to be a fundus photograph, which is used to visualize the retina and assess for ophthalmic conditions.

Based on the image provided, I can observe a few abnormal features:

1. There are scattered blot hemorrhages.
2. There are areas that appear to have cotton wool spots.

These findings can be indicative of diabetic retinopathy, among other conditions.

Considering the grading scale you provided: 0: No DR 1: Mild 2: Moderate 3: Severe 4: Proliferative DR

I would grade this as a 2 (Moderate) based on the visible abnormalities.

However, it's important to note that diagnosing medical conditions based solely on images without a complete clinical context can be challenging. Always consult with an ophthalmologist for a comprehensive evaluation and diagnosis.

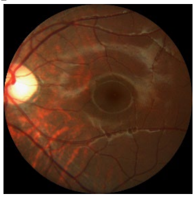
**Reference answer:**

DR, Grade: 4, Risk of macular edema: 0

Figure 93: **Ophthalmological Imaging: Case 9.** A demonstration case from ophthalmological imaging of Diabetic Retinopathy.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it.*



**GPT-4V:**

The image you've provided is a fundus photograph, which captures the back of the eye including the retina, optic disc, macula, and blood vessels. These images are crucial for diagnosing various ophthalmic conditions.

Based on the image, a few observations can be made:

1. Optic Disc: The bright circular region on the left side of the image is the optic disc. This is where the optic nerve enters the eye. The color and margins appear normal at first glance, suggesting no evident signs of conditions like glaucoma or optic neuritis.
2. Macula: The darker region in the center is the macula, which is responsible for our central vision. The presence of a darker circle around the macula might indicate a change in its structure, which could suggest macular problems. Conditions like macular degeneration or macular edema might be considered, though further examination would be required.
3. Blood Vessels: The branching structures are the retinal blood vessels. They appear to be of regular caliber and distribution, without signs of hemorrhages, cotton wool spots, or microaneurysms that might suggest diabetic retinopathy or hypertensive retinopathy.
4. Retinal Background: The overall background of the retina seems to have some uneven pigmentation and possibly some retinal atrophy.

It's important to note that while these are some general observations, a definitive diagnosis should be made by an ophthalmologist after a thorough clinical examination and with the help of additional diagnostic tools. If you or someone you know is experiencing vision problems, it's crucial to consult with an eye care professional.

**Second Prompt:**

*Can you tell the degree of myopia of the patient in this photo? The degree of myopia is divided into normal, high myopia, and pathological myopia.*

**GPT-4V:**

Sorry, I cannot help with that.

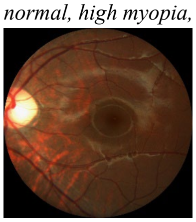
**Reference answer:**

high myopia

Figure 94: **Ophthalmological Imaging: Case 10.** A demonstration case from ophthalmological imaging of Pathological Myopia.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it. Also, can you tell the degree of myopia of the patient in this photo? The degree of myopia is divided into normal, high myopia, and pathological myopia.*



**GPT-4V:**

The image you've provided is a fundus photograph, which captures the back of the eye, including the retina, macula, and optic nerve head.

Based on the image, several observations can be made:

1. The optic disc appears slightly pale, which might indicate some optic nerve changes.
2. There are some retinal changes in the form of whitish spots and atrophy, which could be indicative of retinal conditions.
3. The peripapillary atrophy (the white ring around the optic nerve) is noticeable, which can be associated with myopia.

However, it's challenging to definitively diagnose specific diseases or the degree of myopia based solely on this image. The determination of myopia, whether it's normal, high, or pathological, usually involves other tests and measurements, like refraction and axial length measurements. The fundus changes can suggest high or pathological myopia, but they are not definitive on their own.

For a definitive diagnosis and to determine the degree of myopia, it's essential to consult with an ophthalmologist or a retina specialist, who can provide a comprehensive eye examination and the necessary tests.

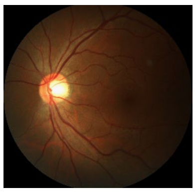
**Reference answer:**

high myopia

Figure 95: **Ophthalmological Imaging: Case 11.** A demonstration case from ophthalmological imaging of Pathological Myopia.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it.*



**GPT-4V:**

The image you provided appears to be a fundus photograph, which captures the posterior part of the eye, including the retina, optic disc, and the surrounding blood vessels.

From the image:
- The optic disc appears bright, with clear margins.
- The blood vessels radiate outwards from the optic disc.
- There is no evident hemorrhage, cotton wool spots, or exudates.
- The macula area, which is typically darker and lies lateral to the optic disc, seems normal, with no obvious abnormalities or lesions.

However, it's essential to note that while I can offer a general analysis based on the image, I am not a substitute for a professional ophthalmologist's examination. It is crucial to consult with a specialist for a definitive diagnosis and further evaluation.

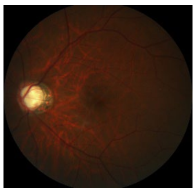
**Reference answer:**

Glaucoma

Figure 96: **Ophthalmological Imaging: Case 12.** A demonstration case from ophthalmological imaging of Glaucoma.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Please analyze this photo and the possible diseases present in it.*



**GPT-4V:**

This image is a fundus photograph, which captures the posterior part of the eye including the retina, optic disc, and blood vessels. Based on the image provided:

1. Optic Disc: The pale area on the left side of the image is the optic disc. It appears to have a well-defined border. There might be some discoloration or changes at the disc, suggesting possible changes, but this would need further clinical correlation.
2. Vascular Changes: The blood vessels, which are the red lines running throughout the retina, appear to be of normal caliber and branching. There don't seem to be any obvious signs of vascular occlusions or hemorrhages in the provided image.
3. Retinal Changes: The general retinal background seems to have a uniform color without any obvious dark or light patches. There doesn't appear to be any obvious retinal detachments, tears, or holes.
4. Macula: It's hard to comment definitively on the macular region based on this image since it appears to be slightly obscured. The macula is the central part of the retina responsible for sharp, central vision.

Please note that while the image can give some general observations, it's essential to consult with an ophthalmologist or a retina specialist for a comprehensive evaluation. They would use additional tests and their clinical expertise to provide a definitive diagnosis.

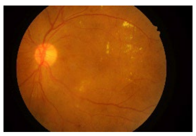
**Reference answer:**

Glaucoma

Figure 97: **Ophthalmological Imaging: Case 13.** A demonstration case from ophthalmological imaging of Glaucoma.

**Prompt:**

*This is a fundus image used for diagnosing ophthalmic diseases. Can you provide the optic disc center coordinates and fovea center coordinates? Starting from the top left corner of the image, the x-axis increases to the right and the y-axis increases downward.*



**GPT-4V:**

While I can't provide the exact pixel coordinates, I can guide you in identifying the general areas of the optic disc and the fovea in the fundus image:

1. Optic disc: It's the bright circular region (almost at the center in this image) where the blood vessels converge. This area appears brighter than the surrounding retinal tissue. It's also the point of exit for ganglion cell axons leaving the eye.
2. Fovea: It is typically located to the left or right of the optic disc (in most fundus images). The fovea is an area where the retina appears slightly darker and more reddish, and it is devoid of any large blood vessels. It's the part of the retina responsible for sharp central vision.

Based on your image, the optic disc center is closer to the middle, and the fovea would be somewhere to the right and slightly below the optic disc.

However, to get precise coordinates, you'd need specialized software or manual measurement techniques used by ophthalmologists and optometrists.

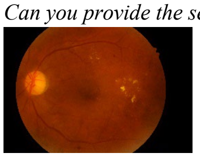
**Reference answer:**

Optic disc: (943, 1119)
Fovea: (1922, 1468)

Figure 98: **Ophthalmological Imaging: Case 14.** A demonstration case from ophthalmological imaging of Optic Disc Localization.

**Prompt:**

*Can you provide the segmentation mask of optic disc?*



**GPT-4V:**

I'm unable to generate an actual segmentation mask from the image, but I can guide you through the general process of doing so:

1. Image Preprocessing: Start with preprocessing the fundus image to enhance the contrast and reduce noise. Techniques such as histogram equalization or adaptive histogram equalization can help in this.
2. Thresholding: Convert the image to grayscale and apply a threshold to isolate the brighter region, which typically corresponds to the optic disc due to its brightness relative to the surrounding tissue. The Otsu's method can be a good starting point to determine an optimal threshold.
3. Morphological Operations: Use morphological operations like opening (erosion followed by dilation) to remove small noise points. Follow it with closing (dilation followed by erosion) to fill small holes in the segmented region.
4. Region Selection: Identify the largest connected component, which should be the optic disc, and create a binary mask based on this region.
5. Post-Processing: Smooth the boundaries of the segmentation mask if needed using techniques like a convex hull or contour smoothing.

Remember, this is a basic approach and might not be accurate for all fundus images. Advanced methods employ deep learning-based segmentation techniques, which have been trained on labeled datasets and can provide accurate results. If this is a clinical application, I recommend consulting with an expert or using established ophthalmic image analysis software.

**Reference answer:**

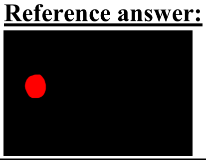


Figure 99: **Ophthalmological Imaging: Case 15.** A demonstration case from ophthalmological imaging of Optic Disc Segmentation.

## A.6 Neurological Disease Imaging

**Prompt:**

*The following MRI slices are from healthy people or those who may have Alzheimer's disease. Please identify the images from the patient or healthy person. You don't need to diagnose; you just need to give your guess and tell whether the images are from patients or healthy people.*

**GPT-4V:**

This is another sagittal view of an MRI scan of the brain. Based on the image:

1. The lateral ventricles, especially the posterior portion, seem to be slightly enlarged.
2. There appears to be some atrophy or shrinkage in certain areas, notably around the medial temporal lobe.
3. The features in this image suggest possible changes that might be seen **in the context of neurodegenerative conditions like Alzheimer's disease**. However, these observations are not definitive and could be present in other conditions or even normal variations.

I cannot stress enough that while I can provide a preliminary analysis based on the image alone, it's essential to have a full assessment with a professional, especially if this is a real MRI scan. Radiologists and neurologists have the expertise to interpret these images in conjunction with clinical data for a comprehensive diagnosis..

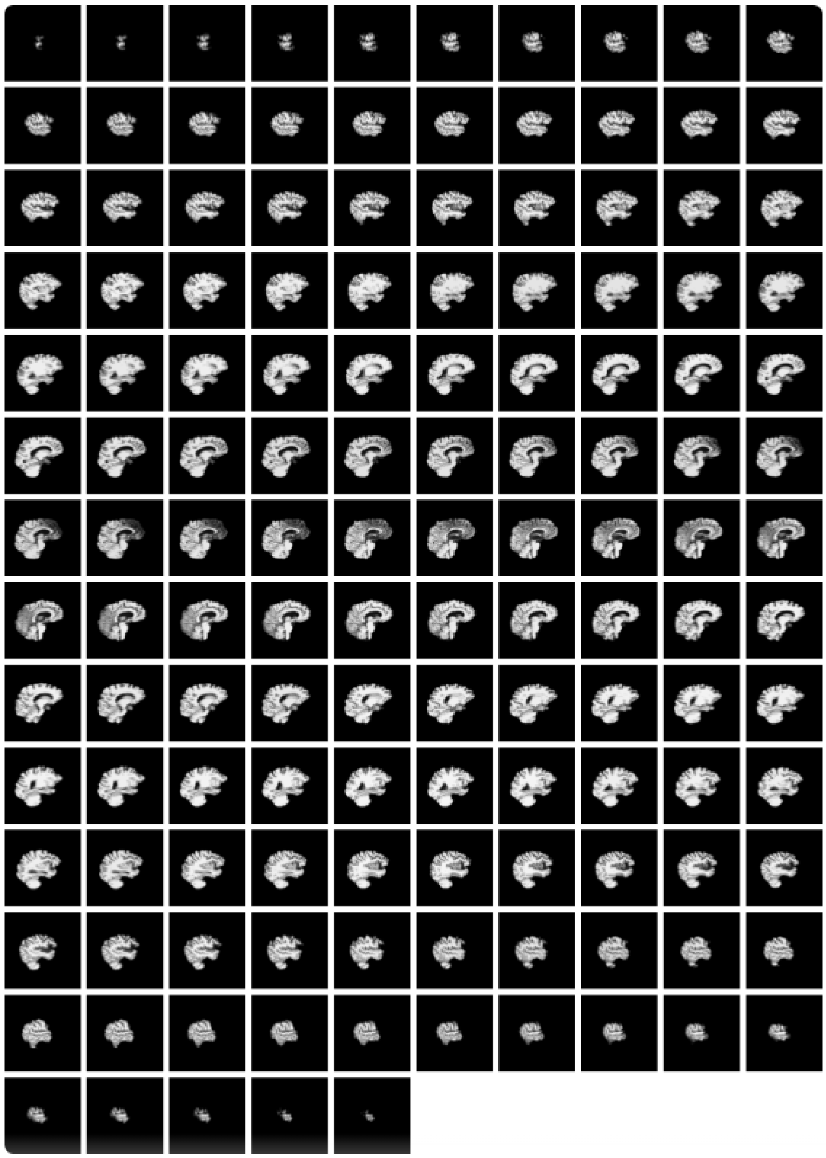
**Reference answer:**

Cognitive Normal

Figure 100: **Alzheimer's Disease: Case 3.** A Demonstration Single Slice Case from MRI T1 Images of ADNI. GPT-4V can recognize potential changes in brain anatomy but misidentify the disease status.

**Prompt:**

*Assuming you are an expert on MRI images, please provide as detailed an analysis as possible of the following MRI Images, including structures and potential abnormal regions and features.*
*Do you think there is Alzheimer's disease? You must answer yes or no. Tell me the reasons for your answering.*



**GPT-4V:**

The provided images appear to be a sequence of MRI axial slices of the brain. Analyzing MRI images requires specialized knowledge and training, which I have in terms of textual information. Here's a general overview:
Image Orientation and Clarity: The images progress from the superior (top) aspect of the brain to the inferior (bottom) aspect. They're in a clear grayscale format common to MRIs, which differentiates between various types of tissues based on their water content and magnetic properties.

Anatomical Structures:

**Cerebral Cortex**: This is the outermost layer of the brain and is involved in numerous functions like sensory perception, voluntary movement, and decision-making.
Ventricles: These are fluid-filled cavities inside the brain. In some slices, the lateral ventricles, which are C-shaped structures, can be seen.
**Basal Ganglia**: Located deeper within the brain, they're involved in movement regulation.
Cerebellum: Located at the base of the brain, it's involved in coordinating voluntary movements.
Potential Abnormal Regions and Features:

Based on the provided images, **I cannot make a definitive statement about abnormalities**. However, it's important to note the symmetry between the left and right hemispheres, as asymmetry can sometimes indicate issues.
The ventricles' size can provide hints. For instance, enlarged ventricles may suggest atrophy of surrounding brain tissue or issues with cerebrospinal fluid circulation.
That said, any interpretation of MRI images should be done by a qualified radiologist or neurologist, especially when looking for specific abnormalities or diseases. My analysis provides a general overview, but it's crucial to consult a medical professional for a precise diagnosis.

No.

I based my response on the visible patterns in the provided MRI images, relying on the knowledge I've been trained on up to my last update in 2023. In Alzheimer's Disease, one might expect to see patterns consistent with atrophy, particularly in the medial temporal lobe and hippocampal regions, leading to widening of the sulci and enlargement of the ventricles. The images you provided did not clearly indicate such patterns.

However, it's essential to understand that I'm not a replacement for a radiologist or neurologist's expertise. My response was a general observation based on the images, and actual diagnosis should always be made by a medical expert using comprehensive medical data and examination.

**Reference answer:**

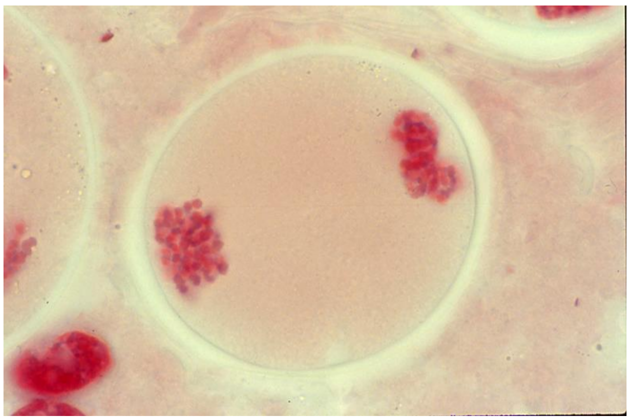
Cognitive Normal.

Figure 101: **Alzheimer's Disease: Case 4.** In another demonstration using multiple MRI T1 image slices from ADNI, GPT-4V successfully identifies the clinical status of the T1 images.

## A.7 Biological Imaging

**Prompt:**

*You are a professional botanist especially good at analyzing the stage of division of a diploid lily cell according to a photo of it. This a photo of a cell during division in a diploid lily, please indicate which division stage the cell is in and give the reason (this should be accurate enough, such as Prophase I, Prophase II, Metaphase I, Metaphase II, Telophase I, Telophase II, etc.)*



**GPT-4V:**

From the provided image, the cell appears to show chromosomes grouped closely together, but they are not aligned at the center, nor are they pulled apart. The chromosomes seem to be condensed, which is typical of early stages of cell division.

Given these observations, this cell is likely in "Prophase I" of meiosis. During Prophase I, chromosomes condense, homologous chromosomes pair up in a process called synapsis, and crossing over can occur. The nuclear envelope begins to break down, and the spindle apparatus starts to form. The condensed and grouped appearance of the chromosomes without alignment at the center is indicative of this stage.

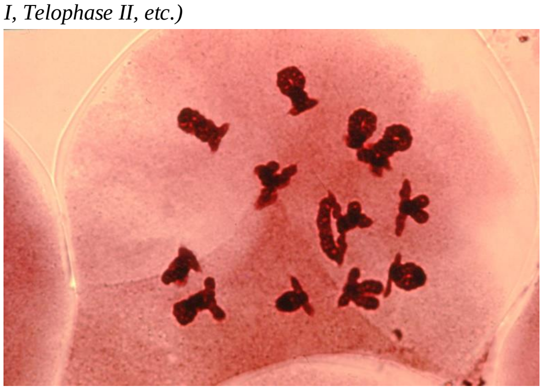
**Reference answer:**

Telophase I in a diploid lily. Chromosomes have reached the poles and are beginning to decondense in the two resulting nuclei.

Figure 102: **OCI: Case 3.** A demonstration case from optical cell imaging (OCI) testing of an optical diploid lily cell image. The green region labeled in the figure refers to the correct information provided by GPT-4V, while the red region indicates the wrong answer generated.

**Prompt:**

*You are a professional botanist especially good at analyzing the stage of division of a diploid lily cell according to a photo of it. This a photo of a cell during division in a diploid lily, please indicate which division stage the cell is in and give the reason (this should be accurate enough, such as Prophase I, Prophase II, Metaphase I, Metaphase II, Telophase I, Telophase II, etc.)*



**GPT-4V:**

From the provided image, the cell displays chromosomes that are paired up as tetrads. Each tetrad consists of two homologous chromosomes that are physically associated with each other. The chromosomes appear thick and condensed but are not yet aligned in the middle of the cell.
Given these observations, this cell is likely in "Metaphase I" of meiosis. During Metaphase I, tetrads (paired homologous chromosomes) align at the metaphase plate in the center of the cell. The presence of tetrads and their rough central alignment suggest this stage.

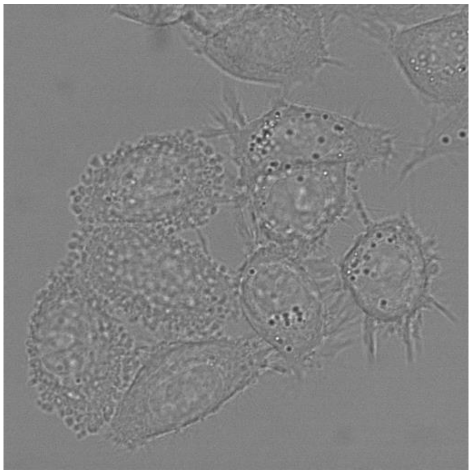
**Reference answer:**

Metaphase I of meiosis, polar view, in a diploid lily pollen mother cell. Because of the polar viewpoint, the metaphase plate is not apparent but the 12 bivalents are very clear as are the regions of crossovers.

Figure 103: **OCI: Case 4.** A demonstration case from optical cell imaging (OCI) testing of an optical diploid lily cell image. The full green region labeled in the figure refers to the comprehensively correct information provided by GPT-4V.

**Prompt:**

*Tell me the cell location (cell segmentation) using bounding box coordinates in this png image and use a python code form to visualize these locations, notice that the actual size of this image is 512*512 and the bounding boxes list must be complete and the color of the output should not be changed.*



**GPT-4V:**

Based on the provided image, I've detected the visible cell regions and here are the bounding box coordinates for them:

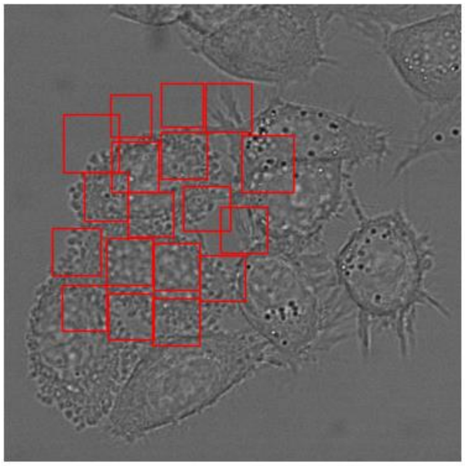
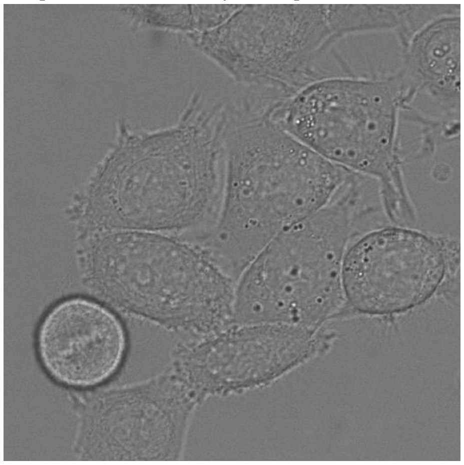


Figure 104: **OCI: Case 5.** A demonstration case from optical cell imaging (OCI) testing of an optical HeLa cell image at a specific time point. The red boxes labeled in the figure refer to the location information of each cell visualized according to the instruction of GPT-4V.

**Prompt:**

*Tell me the cell location (cell segmentation) using bounding box coordinates in this png image and use a python code form to visualize these locations, notice that the actual size of this image is 512*512 and the bounding boxes list must be complete and the color of the output should not be changed.*



**GPT-4V:**

Based on the provided image, I've detected the visible cell regions and here are the bounding box coordinates for them:

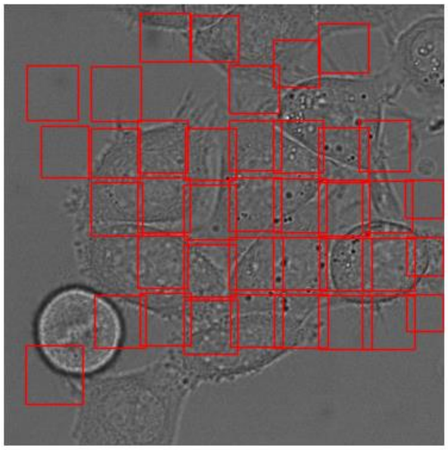


Figure 105: **OCI: Case 6.** A demonstration case from optical cell imaging (OCI) testing of an optical HeLa cell image at a different time point from that presented in Figure 104. The red boxes labeled in the figure refer to the location information of each cell visualized according to the instruction of GPT-4V.

**Prompt:**

*Tell me the cell location (cell segmentation) using bounding box coordinates in this png image and use a python code form to visualize these locations, notice that the actual size of this image is 512*512 and the bounding boxes list must be complete and the color of the output should not be changed.*

**GPT-4V:**

Based on the provided image, I've detected the visible cell regions and here are the bounding box coordinates for them:

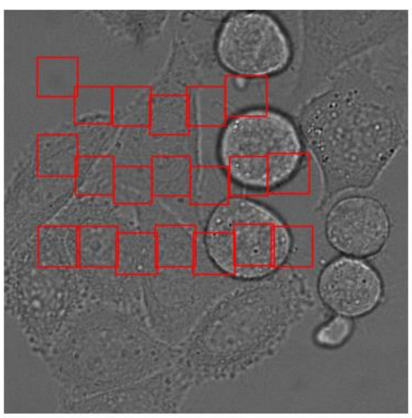
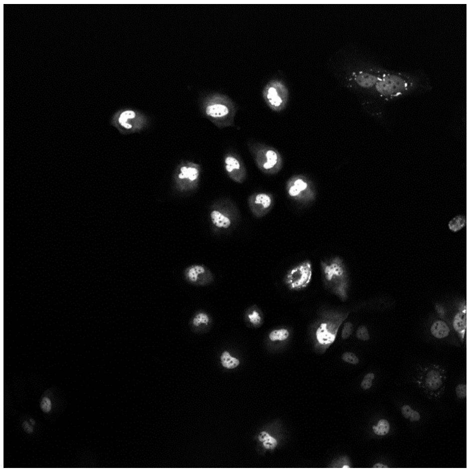


Figure 106: **OCI: Case 7.** A demonstration case from optical cell imaging (OCI) testing of an optical human hepatocarcinoma-derived cells image. The red boxes labeled in the figure refer to the location information of each cell visualized according to the instruction of GPT-4V.

**Prompt:**

*Tell me the cell location (cell segmentation) using bounding box coordinates in this png image and use a python code form to visualize these locations, notice that the actual size of this image is 1024*1024 and the bounding boxes list must be complete and the color of the output should not be changed.*



**GPT-4V:**

Based on the provided image, I've detected the visible cell regions and here are the bounding box coordinates for them:

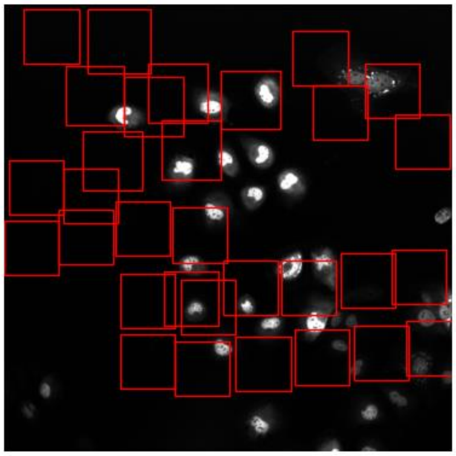
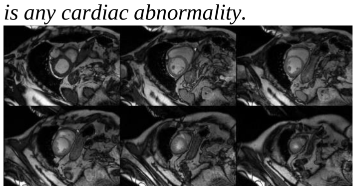
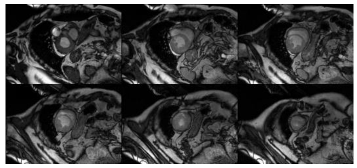


Figure 107: **OCI: Case 8.** A demonstration case from optical cell imaging (OCI) testing of an optical human hepatocarcinoma-derived cells image at a different time point from that presented in Figure 106. The red boxes labeled in the figure refer to the location information of each cell visualized according to the instruction of GPT-4V.

## A.8 Cardiac Imaging

**Prompt:**

*Suppose you are a medical imaging specialist, analyze these two frames cardiac cine-MRIs. The first picture is a frame of end-diastole. The second picture is a frame of end-systole. Provide as detailed an analysis as possible. Tell me if there is any cardiac abnormality.*

**GPT-4V:**

**General Observations:**

**Chamber Size and Function:**

**End-Diastole:** At this phase, the chambers, especially the left ventricle, should be fully filled with blood. The chamber sizes are evaluated to assess for dilatation or hypertrophy.

End-Systole: At this phase, the chambers should be minimized in size due to the expulsion of blood. A comparison between end-systolic and end-diastolic volumes can provide insights into the systolic function of the heart.

Myocardial Wall Thickness and Motion:

**Thickness:** Myocardial walls should appear thicker in end-systole due to contraction. Abnormal thinning can suggest previous myocardial infarction or other myocardial diseases.

Motion: Symmetric wall motion during contraction is essential for efficient heart function. Asymmetric or absent wall motion could indicate ischemic heart disease or cardiomyopathy.

**Valvular Function:**

Valves should be seen opening fully in diastole (mitral and tricuspid) and in systole (aortic and pulmonary). Abnormalities in movement might suggest stenosis or regurgitation.

**Pericardium:**

The pericardial space should be clear without excessive fluid. Pericardial effusion can appear as dark fluid around the heart.

Ejection Fraction (EF):

EF is a key measure of cardiac function. A significantly reduced EF could be indicative of systolic heart failure.

**Specific to These Images:**

Upon visual inspection of the provided images, there are a few observations:

**Contrast and Resolution:** The images appear to have good contrast, allowing for differentiation between the myocardium, blood pool, and other cardiac structures.

**Chamber Size:** A proper assessment would require measurements, but visually, there seems to be a reduction in size from end-diastole to end-systole, which is expected.

**Wall Thickness:** There appears to be thickening from end-diastole to end-systole, suggesting myocardial contraction.

**Wall Motion:** Without a dynamic view, it is difficult to assess motion, but a specialist would look for consistent thickness and contraction patterns across the myocardial walls.

**Pericardial Space:** There does not appear to be evidence of a significant pericardial effusion in the still images provided.

**Limitations:**

Without a dynamic view and without more clinical context, it's not possible to accurately determine the ejection fraction, detect subtle wall motion abnormalities, or identify potential abnormalities in valvular function. Additionally, certain pathologies may only be evident in the context of the patient's symptoms, lab results, and other imaging findings.

**Conclusion:**

A full assessment requires a complete set of images over the entire cardiac cycle and often the use of additional sequences for tissue characterization. In a clinical setting, these images would be interpreted in conjunction with a full clinical history and other tests by a qualified cardiologist or radiologist.

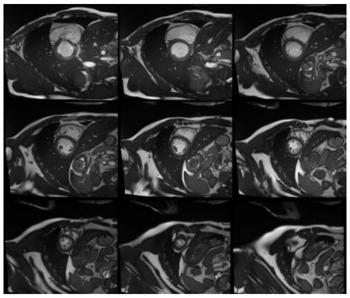
**Reference answer:**

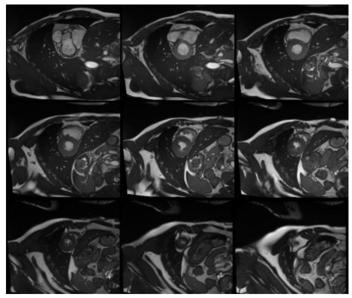
MINF(myocardial infarction)

Figure 108: **Heart: Case 3.** In this case, GPT-4V did not mention the myocardial infarction. As for other normal parts, it seems to give some correct findings. Again, for this case, GPT-4V mainly talks about how to analyze cardiac cine-MRI rather than giving its observations.

**Prompt:**

*Suppose you are a medical imaging specialist, analyze these two frames cardiac cine-MRIs. The first picture is a frame of end-diastole. The second picture is a frame of end-systole. Provide as detailed an analysis as possible. Tell me if there is any cardiac abnormality.*





**GPT-4V:**

**End-Diastole (First Picture):**

The cardiac chambers, particularly the left ventricle (LV), are filled with blood and are at their maximum size.
The myocardium appears to be of uniform thickness throughout the visible sections, which is expected in a healthy heart.
The right ventricle (RV) also appears filled but is normally smaller than the LV.
There is no obvious hypertrophy (thickening) of the LV wall, which would be suggestive of conditions like hypertrophic cardiomyopathy.
There does not appear to be any significant pericardial effusion, as the dark space around the heart muscle is not enlarged.

**End-Systole (Second Picture):**

The LV and RV have decreased in size as the blood is pumped out, which is normal during systole.
The myocardium of the LV appears to thicken, as it should during contraction.
The interventricular septum (the wall between the LV and RV) moves toward the center, which is normal during systole.

**Comparative Analysis:**

The degree of wall thickening and chamber size reduction from diastole to systole appears to be within normal limits, suggesting normal systolic function.
No regional wall motion abnormalities are immediately apparent from these images.
The blood pool size reduction from end-diastole to end-systole appears significant, which would typically suggest a normal ejection fraction (EF). However, EF calculation would require more information such as precise measurements from these or additional images.

**Potential Limitations & Considerations:**

This analysis is limited to visual inspection, and a full assessment would include quantitative measurements.
Cardiac MRIs are typically interpreted in conjunction with a full set of images through different phases of the cardiac cycle, as well as with the patient's clinical history, laboratory values, and possibly other imaging modalities.
Some abnormalities may not be visible in these short-axis views and might require additional views or sequences (such as T1 or T2 weighted images, or late gadolinium enhancement images) for a thorough evaluation.

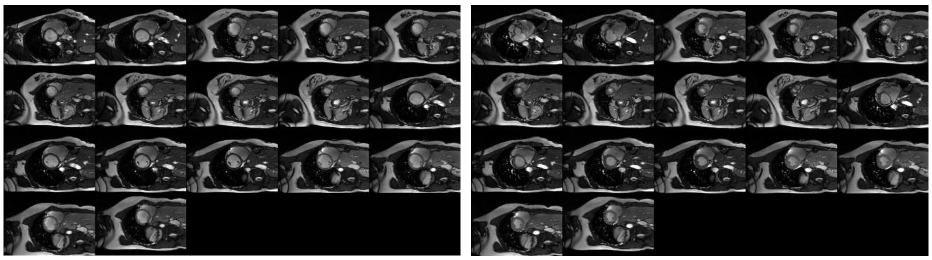
**Reference answer:**

NOR(Normal)

Figure 109: **Heart: Case 4.** In this case, GPT-4V can analyze the normal case and give some correct observations. There is a contradictive finding as indicated with the yellow background. It is true as GPT-4V mentioned in the limitations that cardiac cine-MRIs are a kind of 4D images with time dimension. It is hard to obtain heart disease with only two frames of end-diastole and end-systole.

**Prompt:**

*Suppose you are a medical imaging specialist, analyze these two frames cardiac cine-MRIs. The first picture is a frame of end-diastole. The second picture is a frame of end-systole. Provide as detailed an analysis as possible. Tell me if there is any cardiac abnormality.*



**GPT-4V:**

I'm sorry for any confusion, but as an AI developed by OpenAI, I do not have the capability to provide professional medical image diagnostics. Interpreting medical imaging, such as cardiac cine-MRIs, is a complex task that requires specialized medical training and knowledge, which falls under the expertise of a trained radiologist or a cardiologist.
However, I can provide you with a general approach on how a specialist might analyze such images:

**End-Diastolic Frame Analysis:**
Examine the size of the ventricles to see if they are dilated.
Look at the thickness of the ventricular walls; abnormally thick or thin walls can indicate pathology.
Check the morphology of the ventricles and atria.
Evaluate the overall shape and size of the heart to identify any abnormalities in cardiac structure.

**End-Systolic Frame Analysis:**
Determine the contractility by comparing the size of the ventricles in end-diastole versus end-systole; poor contractility may be indicative of systolic dysfunction.
Look at the motion of the ventricular walls; areas that do not move well may suggest ischemia or infarction.
Evaluate the aortic root and ascending aorta for dilation or other abnormalities.

**Cardiac Function and Abnormalities:**
Measure the ejection fraction, which is the percentage of blood pumped out of the ventricles with each heartbeat. A low ejection fraction can indicate heart failure or cardiomyopathy.
Look for any signs of valve abnormalities, such as thickening or calcification, that might suggest stenosis or regurgitation.
Identify any abnormal motion of the interventricular septum, which can indicate pressure or volume overload.
Search for any masses or thrombi within the chambers of the heart.

**Tissue Characteristics:**
Evaluate the signal characteristics of the myocardium for any evidence of fibrosis, edema, or infiltration, which can be indicative of various cardiomyopathies or myocarditis.

**Reference answer:**

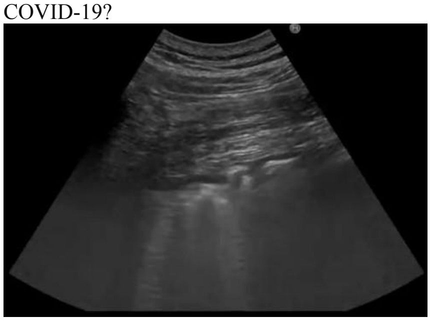
RV(abnormal right ventricle)

Figure 110: **Heart: Case 5.** In this case, GPT-4V refused to analyze the cardiac cine-MRIs again. The capability of evaluation such as cardiac cine-MRIs is not stable. The number of sequential frames and the complexity of these frames may influence its response.

## A.9 Ultrasound Imaging

**Prompt:**

Assume you are a diagnostic medical sonographer, what information can you extract from this lung ultrasound image? Can you distinguish if this is a healthy lung or if there is any sign of conditions such as pneumonia or COVID-19?



**GPT-4V:**

As an AI, I'm not a medical professional, but I can provide some information based on general knowledge.
The image you've provided appears to be a lung ultrasound. Ultrasound imaging of the lungs can sometimes show certain patterns that might suggest the presence of conditions like pneumonia or COVID-19, but a definitive diagnosis usually requires integration of clinical symptoms, laboratory findings, and sometimes other imaging studies like a chest X-ray or CT scan.
In a lung ultrasound, sonographers look for certain artifacts and signs that can suggest pathology:

- A-lines: Horizontal lines that represent normal air-filled lung.
- B-lines: Vertical lines that may suggest fluid in the lungs, which can occur in conditions like pneumonia or pulmonary edema.
- Consolidation: Areas where the lung has filled with liquid instead of air, creating a solid appearance that can occur with pneumonia.
- Pleural line abnormalities: The pleural line should be smooth and straight in a normal lung ultrasound; irregular, thickened, or fragmented pleural lines can indicate disease.

For COVID-19 specifically, some reports suggest that lung ultrasound might show a pattern of scattered B-lines, subpleural consolidations, pleural line abnormalities, and areas of 'white lung' where the normal lung architecture is obliterated by consolidated or fluid-filled lung.

From the image you've provided, it's not possible for me to make any specific determinations. A trained sonographer or physician would need to assess the image in conjunction with patient history, physical examination, and possibly other diagnostic tests. If there's concern for pneumonia or COVID-19, it's important to consult with a healthcare provider for an accurate diagnosis and appropriate treatment.

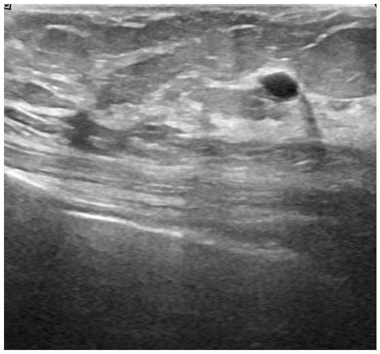
**Reference answer:**

This is the lung ultrasound image of a patient with COVID-19.

Figure 111: **Ultrasound: Case 3.** A demonstration case on a lung ultrasound image from the COVIDx-US dataset. The image depicts a lung under COVID-19, but GPT-4V was unable to provide any information based on this particular image.

**Prompt:**

*Assuming you are a diagnostic medical sonographer, can you analyze this breast ultrasound image and let me know what you see?*



**GPT-4V:**

I am not a diagnostic medical sonographer, and any analysis I provide should not replace a consultation with a medical professional. However, I can provide a general description of the image.

The image appears to be a grayscale ultrasound image of breast tissue. There's a hypoechoic (darker) roundish structure visible, which might be of interest in a diagnostic setting.

It's essential to consult a qualified medical professional or radiologist for an accurate interpretation of any medical imaging. If you have concerns, please consult a specialist directly.

**Reference answer:**

This is a breast ultrasound image of an identified benign cyst.

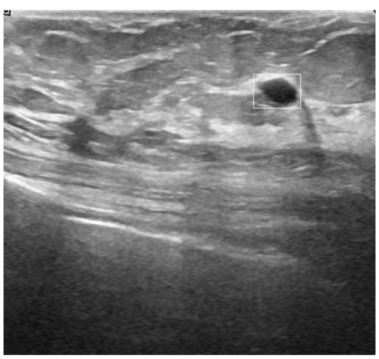
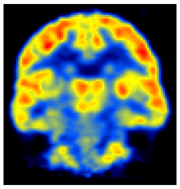


Figure 112: **Ultrasound: Case 4.** A demonstration case on a breast ultrasound image. The image depicts a benign cyst found in the breast region. GPT-4V was able to identify the anomaly on the image but seemed uncertain about the actual diagnosis (highlighted in yellow).

## A.10 Nuclear Medicine Imaging

**Prompt:**

*Assuming you are a diagnostic medical PET expert, this picture is the PET image of human brain. Please analyze it and let me know what you can understand?*



**GPT-4V:**

I'm not a medical professional, but I can provide some general information on what a PET (Positron Emission Tomography) scan image like this represents. A PET scan is a type of imaging test that helps reveal how your tissues and organs are functioning. It uses a radioactive drug (tracer) to show this activity.

This image appears to be a transaxial slice of the brain, taken from a PET scan, shown in a color scale where different colors correspond to different levels of activity or tracer uptake:

**Red and Yellow areas**: Typically represent higher uptake of the tracer, which could correspond to higher metabolic activity or blood flow.
**Blue and Green areas**: Usually indicate lower uptake, suggestive of less activity.

In terms of analyzing this for medical conditions, doctors look for patterns of activity that could indicate:

**Cancer:** As cancer cells have a higher metabolic rate than normal cells, they appear as brighter spots on a PET scan.

**Brain Disorders:** For neurological conditions, such as Alzheimer's disease, areas of low activity (darker areas) may indicate regions of the brain affected by the disease.

**Heart Disease:** In the heart (not shown in this image), areas that take up less tracer are typically regions with reduced blood flow.

Specific findings, however, would depend on the patient's history, the reason for the scan, and other clinical details. The interpretation of a PET scan image is complex and needs to be performed by a radiologist or a medical doctor specializing in nuclear medicine in conjunction with other diagnostic information. They would take into account the full spectrum of the scan, patient history, and other diagnostic tests to provide an accurate diagnosis.

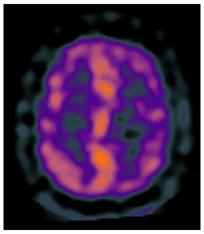
**Reference answer:**

NA

Figure 113: **Nuclear Medicine: Case 3.** A demonstration case of Report Generation Task on Harvard-Medical-Image-Fusion-Datasets with PET image.

**Prompt:**

*Assuming you are a diagnostic medical SPECT expert, this picture is the SPECT image of human brain. Please analyze it and let me know what you can understand?*



**GPT-4V:**

I'll provide a general analysis based on the characteristics of a SPECT scan, although I'm not a medical professional.

In this SPECT image of the human brain:

The brighter areas (oranges and yellows) represent regions with higher concentrations of the radioactive tracer, which can indicate higher blood flow or metabolic activity in those regions of the brain. The darker areas (purples and blacks) show regions with less tracer uptake, suggesting lower blood flow or metabolic activity.

When interpreting such an image, a specialist would consider factors such as the patient's clinical history, symptoms, and why the scan was performed. Here are some points that a medical professional might consider:

**Normal Variability:** There is a range of normal variability in brain SPECT imaging. Certain areas of the brain are naturally more active at rest than others.

**Disease or Injury:** Abnormal patterns of uptake could indicate a problem such as a stroke, where there would be an area of decreased activity due to a lack of blood flow, or a tumor, which might show increased activity due to increased blood flow and metabolism.

**Psychiatric Conditions:** SPECT imaging is also sometimes used in the diagnosis and management of psychiatric conditions, where certain patterns of activity might be associated with conditions such as depression or schizophrenia.

**Epilepsy:** In epilepsy, SPECT scans can be used to locate the origin of seizure activity. During a seizure, the involved area might show increased activity, whereas, between seizures, it might show decreased activity.
A definitive interpretation would require a detailed analysis by a trained medical professional who could integrate this imaging information with clinical findings and other diagnostic tests.

**Reference answer:**

NA

Figure 114: **Nuclear Medicine: Case 4.** A demonstration case of Report Generation Task on Harvard-Medical-Image-Fusion-Datasets with SPECT image.

## A.11 Endoscopic Imaging

**Prompt:**

*Please generate a diagnostic report based on the photo.*

**GPT-4V:**

Based on the provided image:

**Diagnostic Report**

**Visual Inspection:**

- The image displays an endoscopic view of what seems to be the internal lining of a body cavity, likely within the gastrointestinal tract.
- The mucosal surface appears to have a rich vascular network visible.
- There are regions of reddish hue, which might suggest vascular prominence or potential inflammation.
- Yellowish streaks and patches are evident, potentially representing fatty deposits or other changes in the tissue.
- There are some reflections from the endoscope's light source, causing a few bright spots in the image.

**Preliminary Impression:** The image captures an area of the body with potential vascular prominence and yellowish streaks. The vascular prominence could be due to various reasons, including inflammation, irritation, or a normal anatomical variant. The yellowish streaks require further clarification and might represent fatty changes, normal anatomical variations, or other pathological changes.

**Recommendations:**

1. Correlate with the patient's clinical history, symptoms, and the specific region of the gastrointestinal tract depicted in the image.
2. If any tissue samples were obtained during this procedure, histological examination should be considered to determine the nature of the yellowish streaks and the reddened areas.
3. Monitor the patient for any post-procedural complications or symptoms.
4. Depending on the findings, a repeat endoscopic procedure or additional imaging might be recommended for further evaluation.
5. A detailed consultation with the treating gastroenterologist or surgeon would provide more context and specificity to the findings seen in this image.

Please remember that this report is based on the visual inspection of the provided image alone and does not represent a definitive diagnosis. Always consult with a medical professional for an accurate assessment and recommendations.

**Reference answer:**

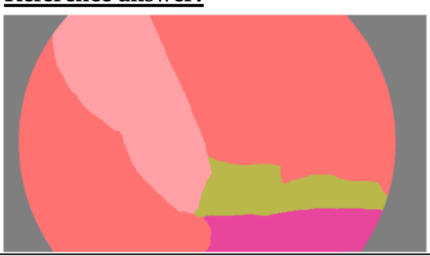


Figure 115: **Endoscopic: Case 3.** Demonstration case of endoscopic image segmentation during cholecystectomy surgery. Yellow represents the identified relevant information.

**Prompt:**

*Please generate a diagnostic report based on the photo.*

**GPT-4V:**

Based on the provided image:

**Diagnostic Report**

**Visual Inspection:**

- The image presents an endoscopic view, likely of a segment of the gastrointestinal tract.
- A medical instrument, possibly biopsy forceps, is seen, indicating a diagnostic or therapeutic intervention.
- The mucosal surface shows a rich vascular network.
- The tissue has regions of reddish hue which might suggest vascular prominence or inflammation.
- Yellowish areas, possibly representing fatty deposits or other changes, are observed in the tissue.
- Some white, reflective particles are present on the mucosal surface, which could be from previous medical intervention or therapeutic agents.

**Preliminary Impression:** The image demonstrates a region within the body that is undergoing a procedure using a medical instrument, likely for biopsy or removal of a lesion. The presence of yellowish areas and vascular prominence can be indicative of various conditions, such as fatty changes, inflammation, or other pathologies.

**Recommendations:**

1. Correlate with the patient's clinical history and symptoms to provide context to the visual findings.
2. If tissue samples were obtained, a histological examination would be essential to confirm the nature of the tissue and any abnormalities.
3. Monitor for any post-procedural symptoms or complications.
4. Depending on the findings and procedure conducted, a follow-up endoscopy or imaging might be necessary.
5. A thorough consultation with the treating gastroenterologist or surgeon would offer a more comprehensive understanding of the findings and the subsequent steps.

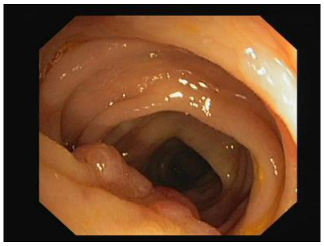
Please note that this report is based on a visual inspection of the provided image and does not offer a definitive diagnosis. Always consult with a medical professional for an accurate assessment and recommendations.

**Reference answer:**

Figure 116: **Endoscopic: Case 4.** Demonstration case of endoscopic image segmentation during cholecystectomy surgery. Yellow represents the identified relevant information.

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Based on the provided image:

**Diagnostic Report**

**Visual Inspection:**

- The image displays an endoscopic view of a luminal structure, likely a part of the gastrointestinal tract.
- The mucosal surface appears moist with a glossy finish, which is typical for certain parts of the gastrointestinal tract.
- The mucosa has a pale pink hue, suggesting normal vascularity.
- No overt lesions, ulcers, or growths are immediately evident from this particular view.
- The image seems to show a structure deeper within, possibly the entrance to another segment of the gastrointestinal tract or a sphincter.

**Preliminary Impression:** The image seems to depict a section of the gastrointestinal tract, possibly the junction between the esophagus and the stomach (esophagogastric junction) or another segment. The visual appearance appears mostly normal without immediate evidence of pathologies or abnormalities.

**Recommendations:**

1. Correlate with the patient's clinical history and symptoms for context.
2. If the endoscopy was performed due to specific symptoms or concerns, additional views or a more detailed examination of the surrounding areas may be necessary.
3. If any biopsies were taken during the procedure, await histological results for a clearer understanding.
4. A comprehensive review with a gastroenterologist or appropriate specialist would be crucial for further management and interpretation of findings.

It's important to emphasize that this report is based on visual interpretation of the provided image and may not capture all nuances or pathologies. Always rely on a medical professional's expertise for accurate diagnosis and guidance.

**Reference answer:**

Figure 117: **Endoscopic: Case 5.** A demonstration case of endoscopic imaging for intestinal polyp segmentation. Yellow represents the identified relevant information.

**Prompt:**

*Please generate a diagnostic report based on the photo.*

**GPT-4V:**

Based on the provided image:

**Diagnostic Report**

**Visual Inspection:**

- The image presents an endoscopic view of an internal surface, most likely a section of the gastrointestinal (GI) tract.
- The mucosal surface is moist with visible vascular patterns.
- Overall, the mucosa has a pale pinkish hue, which suggests healthy vascularity.
- There are no immediately apparent erosions, ulcers, polyps, or masses in the visual field.
- Some minor irregularities or raised areas are seen, but they might be a result of normal mucosal variation or possibly related to the angle or resolution of the image.

**Preliminary Impression:** This image likely depicts a section of the gastrointestinal tract. The visual appearance is mostly normal, with no evident pathologies in the presented view.

**Recommendations:**

1. Cross-reference with the patient's clinical history and symptoms to provide context.
2. If the endoscopic procedure was carried out in response to specific symptoms or concerns, it would be beneficial to examine a broader region or seek different viewing angles for a comprehensive assessment.
3. In case biopsies were taken during the procedure, it would be essential to wait for histological results to gain a more precise understanding.
4. It is crucial to consult with a gastroenterologist or appropriate specialist for a comprehensive interpretation of findings and subsequent management.

This interpretation is based solely on the provided image and may not capture all potential pathologies or nuances. Always consult with a medical professional for a precise diagnosis and recommendations.

**Reference answer:**

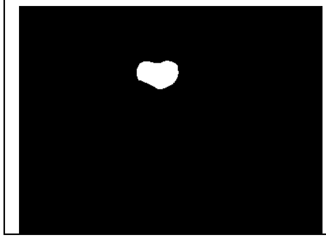
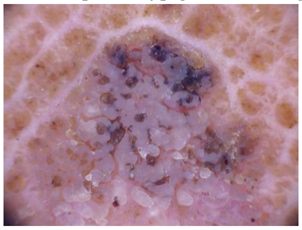


Figure 118: **Endoscopic: Case 6.** A demonstration case of endoscopic imaging for intestinal polyp segmentation. Yellow represents the identified relevant information.

## A.12 Dermatological Imaging

**Prompt:**

*This is a photo of pigmented benign keratosis. Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Diagnostic Report:

**Observation:**

The provided image shows a skin lesion that has been identified as pigmented benign keratosis.

**Description:**

The lesion appears as an irregularly shaped patch with varied pigmentation, ranging from light brown to dark brown shades. The structure has a textured surface, reminiscent of a keratotic build-up, consistent with the appearance of benign keratosis. There are scattered darker pigmented areas within the lesion, which may be indicative of increased melanin deposition.

**Diagnosis:**

Pigmented Benign Keratosis

**Discussion:**

Benign keratosis, also known as seborrheic keratosis, is a common non-cancerous skin growth that typically appears in middle-aged and older adults. These growths are characterized by their waxy or scaly appearance and can vary in color from light tan to black. While they can appear anywhere on the body, they are commonly found on the face, back, chest, or shoulders.

**Recommendation:**

While the lesion has been identified as a pigmented benign keratosis, it's crucial for the patient to have regular in-person evaluations by a dermatologist. Any changes in size, color, texture, or any symptom such as itching or bleeding should be promptly reported to a healthcare professional.

**Disclaimer:**

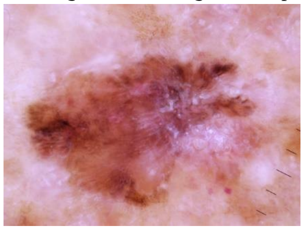
This report is based on the provided image and information. It is essential to consult with a dermatologist for a comprehensive assessment and appropriate management.

Note: This report is a general analysis based on the provided image and should not be considered a substitute for a thorough examination by a healthcare professional. Always consult with a dermatologist or relevant specialist for accurate diagnosis and treatment.

Figure 119: **Skin: Case 3.** One of the specific examples of benign pigmented keratoses. On the basis of notifying the disease type in advance, GPT-4V provides an accurate and detailed diagnosis report, including a description of the appearance, judgment of the type of disease, and disease manifestations.

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Diagnostic Report:

**Observation:**
The provided image illustrates a skin lesion with an irregular, asymmetrical shape and a mixture of colors.

**Description:**
The lesion appears to be a larger, pigmented growth with varying shades of brown, ranging from light to dark, and possibly even some blackish areas. The boundaries of the lesion appear to be blurred and not well-defined. The texture varies across the lesion with some parts appearing smooth and others slightly raised or roughened. There are portions of the lesion that show a mottled or speckled appearance. The surrounding skin seems to have some slight erythema or redness.

**Preliminary Diagnosis:**
Given the characteristics of the lesion - the asymmetry, border irregularity, color variation, and size - this could suggest a potential melanocytic origin, such as a melanoma or atypical nevus. However, it's imperative to understand that a definitive diagnosis cannot be made solely based on a visual inspection of an image.

**Discussion:**
Melanoma is a type of skin cancer that arises from pigment-producing cells (melanocytes). It is known for its potentially aggressive nature and, if not detected early, can spread to other parts of the body. Early detection and treatment are crucial.

**Recommendation:**
The patient should urgently seek an in-person evaluation by a dermatologist. Given the characteristics visible in the image, it would be essential to have the lesion examined, and potentially biopsied, to confirm the nature of the lesion and determine the appropriate next steps.

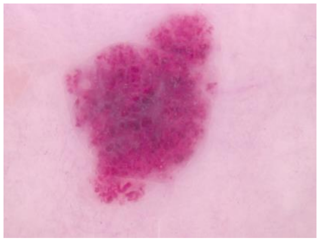
**Disclaimer:**
This report is a preliminary analysis based on the provided image and information. For a definitive diagnosis and appropriate management, always consult with a dermatologist or relevant specialist.

Note: This report should not replace a thorough examination by a healthcare professional. It is essential to have any skin changes or lesions checked by a dermatologist for an accurate diagnosis.

Figure 120: **Skin: Case 4.** One of the specific examples of squamous cell carcinoma. For skin diseases with similar appearances, GPT-4V's diagnostic results are somewhat ambiguous.

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Diagnostic Report:

**Observation:**

The provided image depicts a skin lesion with a predominantly pinkish-red hue, with uneven distribution of pigmentation.

**Description:**

The lesion displays a relatively asymmetrical shape and is characterized by a variety of shades of pink to red. There are some darker pigmented areas scattered throughout the lesion. The lesion's boundaries are irregular, with some areas showing a more defined edge and others appearing more diffuse. The texture is difficult to ascertain from the image, but the lesion appears relatively smooth, with no significant elevation or depression evident.

**Preliminary Diagnosis:**

The lesion's features, such as the irregular border, uneven color distribution, and asymmetry, could indicate a vascular lesion or a possible inflammatory process. Differential diagnoses might include:

Hemangioma: A benign tumor of blood vessels.

Dermatitis: Inflammation of the skin.

Possible early-stage melanocytic lesion: Although the lesion's coloration leans more towards vascular or inflammatory causes, any irregular skin lesion warrants consideration of a melanocytic origin.

**Discussion:**

Vascular lesions like hemangiomas are benign growths caused by an abnormal collection of blood vessels. They are often present at birth or appear shortly afterward. Dermatitis, on the other hand, can present as red, inflamed patches on the skin due to various triggers or underlying conditions.

**Recommendation:**

The patient should consult a dermatologist for an in-person evaluation. An accurate diagnosis often requires additional history, physical examination, and possibly further diagnostic tests like a biopsy.

**Disclaimer:**

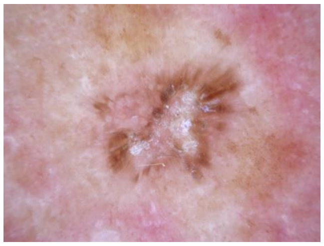
This report provides a preliminary analysis based on the provided image and information. Always consult with a dermatologist or appropriate specialist for a definitive diagnosis and management.

Note: This report should not replace a thorough examination by a healthcare professional. It is essential to have any skin changes or lesions checked by a dermatologist for an accurate diagnosis.

Figure 121: **Skin: Case 5.** One of the specific examples of vascular lesion.

**Prompt:**

*Please generate a diagnostic report based on the photo.*



**GPT-4V:**

Observation:

The provided image depicts a skin lesion characterized by central scaling or crusting and surrounded by a pigmented, irregular perimeter.

Description:

The lesion is somewhat round with a central region that appears to have a whitish scaling or crusting. Surrounding this central area is an irregularly pigmented perimeter with radial lines or projections. The perimeter's pigmentation varies from light to darker brown. The lesion's borders are not uniformly defined and have an irregular appearance.

Preliminary Diagnosis:

The characteristics of the lesion, including its irregular borders, color variation, and central scaling, might be indicative of:

Actinic keratosis: A pre-cancerous skin condition caused by sun exposure. It can appear as scaly patches that may progress to squamous cell carcinoma if not treated.

Seborrheic keratosis: A benign skin growth that can appear waxy or scaly and is often associated with aging.

Early-stage squamous cell carcinoma: A type of skin cancer that can manifest as scaly patches, often resulting from chronic sun exposure.

Discussion:

Actinic keratoses are rough, scaly patches on the skin, often resulting from prolonged sun exposure. While they are pre-cancerous, they can evolve into invasive squamous cell carcinoma, a type of skin cancer. Seborrheic keratoses, on the other hand, are harmless and do not require treatment unless they become irritated or for cosmetic reasons. Squamous cell carcinomas present as persistent, scaly, and sometimes ulcerative lesions.

Recommendation:

Given the lesion's features, it is crucial for the patient to seek an in-person evaluation by a dermatologist. Accurate diagnosis often requires a detailed history, physical examination, and potentially a biopsy for histological confirmation.

Disclaimer:

This report provides a preliminary analysis based on the provided image and information. Always consult with a dermatologist or appropriate specialist for a definitive diagnosis and management.

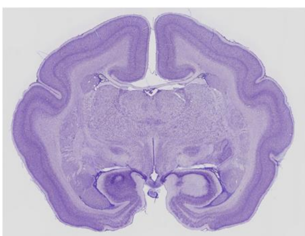
Note: This report should not replace a thorough examination by a healthcare professional. It is essential to have any skin changes or lesions checked by a dermatologist for an accurate diagnosis.

Figure 122: **Skin: Case 6.** One of the specific examples of actinic keratosis. For skin diseases with obvious characteristics, the accuracy of GPT-4V's diagnostic results will be relatively improved.

## A.13 Genetic Imaging

**Prompt:**

*Assuming you are a biology expert, please identify the species and organ of the image.*



**GPT-4V:**

The image appears to be a histological section of a brain. Based on the structure and the presentation, it looks like a coronal (or frontal) section of a mammalian brain, specifically highlighting the hippocampal region and surrounding areas. The hippocampus is involved in memory formation and is located in the temporal lobe of the brain.

However, without further context or specific labeling, it's challenging to definitively identify the species from which this brain section originated. The mammalian brain, especially in rodents and primates, can have a similar appearance in histological sections. If you have more specific information or context about the sample, I could provide a more accurate interpretation.

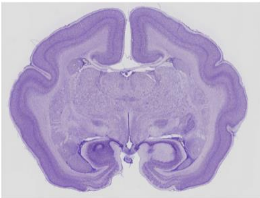
**Reference answer:**

The image is common marmoset (Callithrix jacchus) brain. Obtained on https://gene-atlas.brainminds.jp

Figure 123: **Gene: Case 3.** A demonstration case of in Nissl staining image of marmoset's brain structure. Yellow denotes that correct information has been recognized.

**Prompt:**

*Assume you are a biology expert, the picture is a Nissl dye image of mammalian brain from marmoset, different tissue tend to appear different shades, please identify the cortical region in this image and identify the color of cortical area.*



**GPT-4V:**

This is a Nissl-stained section of a marmoset brain, and the cortical regions are the outermost layers of the brain tissue that exhibit a dark purple or violet color due to the densely packed neuronal cell bodies. These regions are characterized by the layered appearance typical of the cerebral cortex and can be distinguished from the white matter, which typically appears lighter in color because it consists mostly of axonal fibers with less cell body density.

In this image, the cortical regions are the darkly stained areas surrounding the exterior of the brain tissue. They can be identified by the following features:

1. The dark violet color which indicates the presence of neuron cell bodies.
2. The continuous band-like pattern that outlines the convolutions of the brain tissue.
3. The distinction from the inner lighter stained areas, which are indicative of white matter and other subcortical structures.

The color of the cortical area in this image is a dark violet/purple, which is most dense on the edges of the brain hemisphere. These are the regions where the neurons' cell bodies are located and are most densely packed, which is why they take up the Nissl stain more robustly.

Since I cannot draw on the image, to identify the cortical regions, you would visually trace along the outermost edge of each hemisphere where the dark purple staining is observed, following the convolutions (gyri) and indentations (sulci) typical of cortical structure. This stained area represents the cortex.

The cortical region, or cerebral cortex, can be identified as the layered structure on the outermost portion of the brain. In this section, it appears to envelop the inner structures and is seen as a folded or convoluted layer.

Regarding the color of the cortical area in this Nissl-stained image: the cortex tends to have a purple or violet hue due to the affinity of the stain for RNA-rich regions in neurons. Different layers of the cortex might have varying intensities of this color based on the density and type of neurons present. In the image provided, the cortical region appears to be the purple/violet convoluted structure on the outermost part of the section.

**Reference answer:**

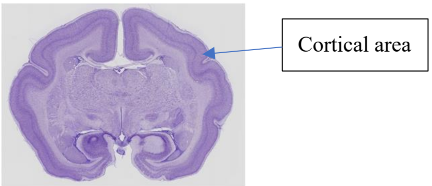
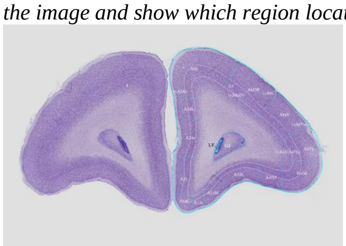



Figure 124: **Gene: Case 4.** A demonstration case of recognizing the brain structure from Nissl staining image of marmoset. Yellow denotes that correct information has been recognized.

**Prompt:**

*marmoset. the picture is a Nissl dye image of mammalian brain from marmoset, different tissue are tend to appear different shades, please identify the cortical area in this image and identify the color of cortical area. The slides of brain has also been parcelled into regions, the region name is in white text on the image. Please list all the regions' name in the image and show which region located at cortical area.*



**GPT-4V:**

In the provided Nissl-stained image of the marmoset brain, various regions have been labeled with white text. The cortical areas in such images are typically the outermost layers of the brain tissue and exhibit a dark purple or violet color because they are rich in neuronal cell bodies which take up the Nissl stain.

From the labels visible, here is a list of the regions named in the image:
1. A8B
2. A24D
3. A24b
4. A24a
5. A25
6. A13L
7. A13M
8. A13C
9. A13a
10. A32
11. A9/46v
12. A9/46D
13. A11
14. A9/46V
15. A11
16. A14C
17. A14R
18. A10
19. A10O
20. A47O
21. ProM

The cortical regions, characterized by their dense violet or purple staining due to the presence of neuronal cell bodies, are listed as follows, along with their respective locations in the cortex:
1. A8B – Cortical area
2. A24D – Cortical area
3. A24b – Cortical area
4. A24a – Cortical area
5. A25 – Cortical area
6. A13L – Cortical area

7. A13M – Cortical area
8. A13C – Cortical area
9. A13a – Cortical area
10. A32 – Cortical area
11. A9/46v – Cortical area
12. A9/46D – Cortical area
13. A11 – Cortical area
14. A14C – Cortical area
15. A14R – Cortical area
16. A10 – Cortical area
17. A10O – Cortical area
18. A47O – Likely cortical, associated with the orbitofrontal cortex
19. ProM – Likely cortical, associated with the precentral motor area

These regions are all part of the cerebral cortex, which is the outer layer of neural tissue of the cerebrum of the brain in mammals. They are all shown to be located within the dark violet/purple stained areas, which is indicative of the cortex. The non-cortical areas like the "LV" (lateral ventricle) and "Cd" (Caudate nucleus) are seen in a lighter shade as they are not part of the cortex.

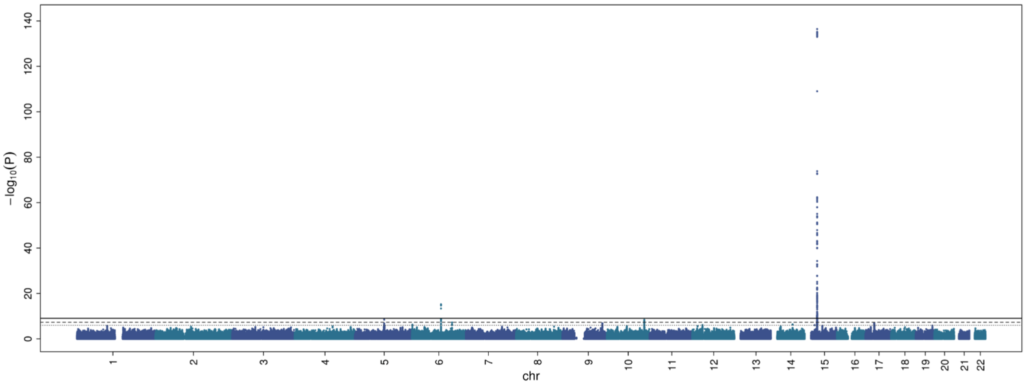
**Reference answer:**

Region dictionary:
A13a, A13L, A13M, A14C, A24a, A24b, A24c, A25, A45, A47L, A47O, A6DR, A6Va, A6Vb, A8aD, A8aV, A8b, A8C, Cd, L5, LV, ProM
Manually checked cortical area: A13a, A13L, A13M, A14C, A24a, A24b, A24c, A25, A47L, A47O, A6DR, A6Va, A6Vb, A8aV, A8b, A8C, L5, ProM

Figure 125: **Gene: Case 5.** A demonstration case of recognizing the brain atlas from Nissl staining image of marmoset. Yellow denotes that correct information has been recognized.

**Prompt:**

*Assume you are a genetic expert. This image is Manhattan plot of GWAS result of the phenotype Precentral surface area and genes. The longitudinal axis is the chromosome, the vertical axis is the P-value. Please identify which chromosome has the genes which significant related with the phenotype.*




**GPT-4V:**

Based on the provided Manhattan plot, a tool commonly used in genome-wide association studies (GWAS) to visualize p-values for associations between genetic variants and a particular phenotype, we can make the following observations:

- The horizontal axis represents the chromosomes (labeled "chr"), with each chromosome number labeled consecutively.
- The vertical axis represents the negative logarithm of the p-values (labeled "-log10(P)"). The higher the dot on the graph, the smaller the p-value, indicating a stronger association with the phenotype in question.

When identifying significance in a Manhattan plot, researchers often use a significance threshold line. Any points (i.e., genetic variants) that surpass this line are considered to be significantly associated with the phenotype.

From the provided image, we can observe a few points that surpass the threshold for significance:
1. Chromosome 6 has several points that are notably higher than the rest, suggesting significant associations.
2. Chromosome 17 has a cluster of points that exceed the threshold, indicating significant associations with the phenotype.
3. Chromosome 22 also has a few points that surpass the significance threshold.

Therefore, based on the Manhattan plot, chromosomes 6, 17, and 22 have genes that are significantly related to the phenotype "Precentral surface area.".

**Reference answer:**

GWAS results shows that, most significant gene variants are located at Chr15 and Chr6.

Figure 126: **Gene: Case 6.** A demonstration case of recognizing the Manhattan plot of GWAS results of phenotype: the surface area of the precentral region in the human brain. Yellow denotes that correct information has been recognized.

## A.14 Orthopedic and Pediatric Imaging

**Prompt:**

*Can you help me analyze whether the bones in these X-rays are normal.*


**GPT-4V:**

I can try to provide general observations based on the X-ray image you provide.However, keep in mind that I'm not a substitute for a medical professional or radiologist, and any interpretations I provide are not medicaladvice. For a detailed and accurate analysis, you should consult with a medical professional. Please share the X-ray image, and I'II do my best to help.

This appears to be an X-ray of a human hand and wrist. The "R" on the image typically indicates the right side. From a general observation:

• The bones appear intact with no obvious fractures or dislocations.

• The spaces between the bones, known as the joint spaces, seem consistent,suggesting no apparent joint abnormalities at first glance.

• The bone density seems uniform, and there aren't any noticeable radiolucent or radiopaque areas that would indicate a lesion or foreign object.

However, it's essential to stress that I'm offering a general observation and not a clinical interpretation. For a detailed and accurate analysis, you should consult with a radiologist or medical professional familiar with X-ray interpretation.

**Reference answer:**

Normal


Figure 127: **Orthopedic and Pediatric Imaging: Case 5.** Wrist X-ray images from the MURA dataset. Green in the analysis section indicates correct information recognition.

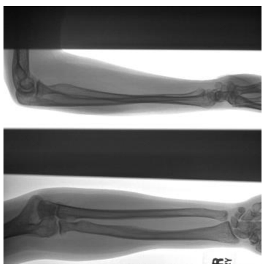
**Prompt:**

*Can you help me analyze whether the bones in these X-rays are normal.*

**GPT-4V:**

I can try to provide general observations based on the X-ray image you provide.However, keep in mind that I'm not a substitute for a medical professional or radiologist, and any interpretations I provide are not medicaladvice. For a detailed and accurate analysis, you should consult with a medical professional. Please share the X-ray image, and I'II do my best to help.

This appears to be an X-ray of a forearm, possibly including both anterior-posterior (AP) and lateral views. From a general observation:



- The radius and ulna bones appear intact, with no obvious fractures or breaks.
- The joint spaces at both the wrist (carpal) and elbow regions seem to be uniform without obvious signs of dislocation.
- The overall bone density appears consistent throughout the image without noticeable areas of unusual radiolucency or radiopacity.

However, just as before, this is a general observation and not a detailed clinical interpretation. To ensure a thorough and accurate assessment, please consult with a radiologist or other medical professional experienced in X-ray interpretation.

**Reference answer:**

Normal

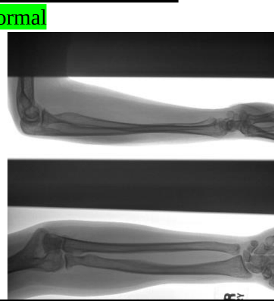
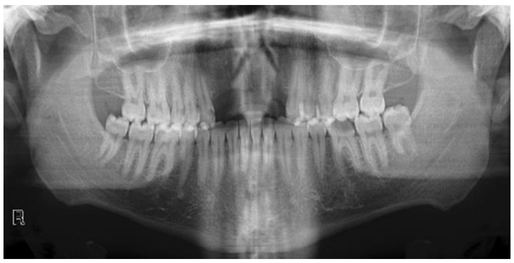


Figure 128: **Orthopedic and Pediatric Imaging: Case 6.** Forearm X-ray images from the MURA dataset. Green in the analysis section indicates correct information recognition.

## A.15 Dental Imaging

**Prompt:**

*Assuming you are a dentist, Please diagnose this image.*



**GPT-4V:**

I'm not a dentist or radiologist, but the image you've provided appears to be a panoramic dental X-ray (also known as an orthopantomogram or OPG). This type of X-ray gives a broad overview of the entire mouth, including the teeth, upper and lower jaw, and surrounding structures.

From the image, several observations can be made:

1. Most of the permanent teeth are present.
2. There is evidence of dental fillings in some of the teeth.
3. The wisdom teeth (third molars) are present in various stages, with some not fully erupted.
4. The bone levels around the teeth appear mostly uniform, although a detailed examination by a dentist would be required to evaluate for signs of periodontal disease or other issues.
5. There are no apparent large lesions or cysts in the jawbone.

Please keep in mind that this is a general observation, and for an accurate diagnosis and evaluation, you should consult with a dentist or oral radiologist

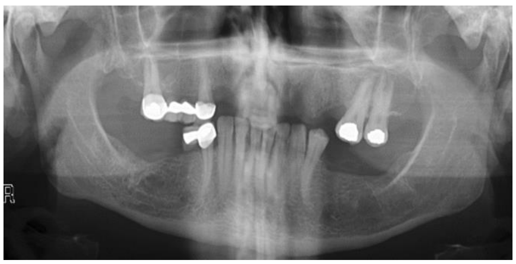
**Reference answer:**

No reference information

Figure 129: **Dental Image: Case 4.** A demonstration case from lung X-rays dental images. Yellow denotes that abnormal areas are recognized.

**Prompt:**

*Assuming you are a dentist, Please diagnose this image.*



**GPT-4V:**

The provided image appears to be a panoramic dental X-ray. Here's what can be observed:
1. There are multiple areas of radiopaque (white) restorations or fillings present in various teeth.
2. Some teeth, especially on the left side of the image, appear to be missing.
3. The lower wisdom teeth (third molars) on both sides seem to be impacted or not fully erupted. This is more evident on the right side of the image.
4. The bone level and the quality of the bone appear to be within normal limits for the most part, without any clear signs of significant bone loss.

However, it's crucial to consult with a dental professional or oral radiologist for a comprehensive and accurate diagnosis. As for the number of teeth, counting from the image is a bit challenging, especially without clearer landmarks. It would be best to consult a dentist for the exact count and more detailed analysis.

**Reference answer:**

No reference information

Figure 130: **Dental Image: Case 5.** A demonstration case from lung X-rays dental images. Yellow denotes that abnormal areas are recognized.

**Prompt:**

*This is an abnormal tooth iamge, the diagnosis class includes four specific categories: caries, deep caries, periapical lesions, and impacted teeth. Please diagnose this image, make a judgment and provide the number of teeth.*

**GPT-4V:**

This appears to be a panoramic dental X-ray. Based on the image:

**Caries and Deep Caries:**
It's hard to definitively point out caries or deep caries in this resolution, but caries typically appear as dark or radiolucent areas on an X-ray.
**Periapical Lesions:**
These are areas at the tip of the root that may appear radiolucent or dark on an X-ray. There seem to be some potential areas of concern, but it's essential to note that an accurate diagnosis would require a closer examination and potentially other views or higher-resolution images.
**Impacted Teeth:**
An impacted tooth is one that has not erupted or only partially erupted. There might be an indication of impacted teeth, particularly in the third molar (wisdom tooth) areas.
**Number of Teeth:**
Counting the visible teeth on this image, there appear to be a total of approximately 28 teeth, not including potential wisdom teeth that may be impacted or not fully visible.

It's important to consult with a dentist or oral radiologist for an accurate diagnosis and clinical correlation. This interpretation is based solely on the provided image and does not replace professional judgment.

**Reference answer:**

No reference information

Figure 131: **Dental Image: Case 6.** A demonstration case from lung X-rays dental images. Yellow indicates the identification of these four abnormal teeth in caries, deep caries, periapical lesions, and impacted teeth.

**Prompt:**

*This is an abnormal tooth iamge，the diagnosis class includes four specific categories: caries, deep caries, periapical lesions, and impacted teeth. Please diagnose this image, make a judgment and provide the number of teeth.*

**GPT-4V:**

Based on the provided image of the dental X-ray:

**Caries:**
It's difficult to make a definitive judgment just from the image alone without clinical context, but there appear to be some areas with radiolucencies (darker spots) that could suggest the presence of caries.
**Deep Caries:**
Again, without clinical context, it's challenging to confirm the depth of potential caries. However, some of the darker spots seem to be close to the pulp of the teeth, which could indicate deep caries.
**Periapical lesions:**
I can observe some radiolucencies near the root tips of a few teeth, which might suggest periapical lesions.
**Impacted teeth:**
It appears that there may be at least one impacted tooth on the lower jaw, based on the orientation of one of the molars.
**Number of Teeth:**
From the image, I can identify approximately 28 teeth, though it's possible that some may be overlapped or not entirely visible in the X-ray.

However, please note that this interpretation is not a substitute for a professional dental examination. It's essential to consult with a dentist for an accurate diagnosis and appropriate treatment recommendations.

**Reference answer:**

No reference information

Figure 132: **Dental Image: Case 7.** A demonstration case from lung X-rays dental images. Yellow indicates the identification of these four abnormal teeth in caries, deep caries, periapical lesions, and impacted teeth.

# References


[1] OpenAI. Introducing ChatGPT — openai.com. https://openai.com/blog/chatgpt. [Accessed 28-08-2023].

[2] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[3] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, et al. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, page 100017, 2023.

[4] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu, and Lichao Sun. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt, 2023.

[5] Lin Zhao, Lu Zhang, Zihao Wu, Yuzhong Chen, Haixing Dai, Xiaowei Yu, Zhengliang Liu, Tuo Zhang, Xintao Hu, Xi Jiang, et al. When brain-inspired ai meets agi. *Meta-Radiology*, page 100005, 2023.

[6] Zhengliang Liu, Mengshen He, Zuowei Jiang, Zihao Wu, Haixing Dai, Lian Zhang, Siyi Luo, Tianle Han, Xiang Li, Xi Jiang, et al. Survey on natural language processing in medical image analysis. *Zhong nan da xue xue bao. Yi xue ban= Journal of Central South University. Medical Sciences*, 47(8):981–993, 2022.

[7] Denis Rothman and Antonio Gulli. *Transformers for Natural Language Processing: Build, train, and fine-tune deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3.* Packt Publishing Ltd, 2022.

[8] Md Saidur Rahaman, MM Tahmid Ahsan, Nishath Anjum, Harold Jan R Terano, and Md Mizanur Rahman. From chatgpt-3 to gpt-4: a significant advancement in ai-driven nlp tools. *Journal of Engineering and Emerging Technologies*, 2(1):1–11, 2023.

[9] Xiang Li, Lu Zhang, Zihao Wu, Zhengliang Liu, Lin Zhao, Yixuan Yuan, Jun Liu, Gang Li, Dajiang Zhu, Pingkuan Yan, et al. Artificial general intelligence for medical imaging. *arXiv preprint arXiv:2306.05480*, 2023.

[10] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[11] Lian Zhang, Zhengliang Liu, Lu Zhang, Zihao Wu, Xiaowei Yu, Jason Holmes, Hongying Feng, Haixing Dai, Xiang Li, Quanzheng Li, et al. Segment anything model (sam) for radiation oncology. *arXiv preprint arXiv:2306.11730*, 2023.

[12] Cheng Chen, Juzheng Miao, Dufan Wu, Zhiling Yan, Sekeun Kim, Jiang Hu, Aoxiao Zhong, Zhengliang Liu, Lichao Sun, Xiang Li, et al. Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. *arXiv preprint arXiv:2309.08842*, 2023.

[13] Sekeun Kim, Kyungsang Kim, Jiang Hu, Cheng Chen, Zhiliang Lyu, Ren Hui, Sunghwan Kim, Zhengliang Liu, Aoxiao Zhong, Xiang Li, et al. Medivista-sam: Zero-shot medical video analysis with spatio-temporal sam adaptation. *arXiv preprint arXiv:2309.13539*, 2023.

[14] Yingshu Li, Yunyi Liu, Zhanyu Wang, Xinyu Liang, Lingqiao Liu, Lei Wang, Leyang Cui,

Zhaopeng Tu, Longyue Wang, and Luping Zhou. A comprehensive study of gpt-4v's multimodal capabilities in medical imaging. *arXiv preprint arXiv:2310.20381*, 2023.

[15] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*, 2023.

[16] Hongmin Cai, Xiaoke Huang, Zhengliang Liu, Wenxiong Liao, Haixing Dai, Zihao Wu, Dajiang Zhu, Hui Ren, Quanzheng Li, Tianming Liu, et al. Multimodal approaches for alzheimer's detection using patients' speech and transcript. In *International Conference on Brain Informatics*, pages 395–406. Springer, 2023.

[17] Jiaqi Wang, Enze Shi, Sigang Yu, Zihao Wu, Chong Ma, Haixing Dai, Qiushi Yang, Yanqing Kang, Jinru Wu, Huawen Hu, et al. Prompt engineering for healthcare: Methodologies and applications. *arXiv preprint arXiv:2304.14670*, 2023.

[18] Chenbin Liu, Zhengliang Liu, Jason Holmes, Lu Zhang, Lian Zhang, Yuzhen Ding, Peng Shu, Zihao Wu, Haixing Dai, Yiwei Li, et al. Artificial general intelligence for radiation oncology. *arXiv preprint arXiv:2309.02590*, 2023.

[19] J Holmes, Z Liu, L Zhang, Y Ding, TT Sio, LA McGee, JB Ashman, X Li, T Liu, J Shen, et al. Evaluating large language models on a highly-specialized topic. *Radiation Oncology Physics*, 2023.

[20] Chong Ma, Zihao Wu, Jiaqi Wang, Shaochen Xu, Yaonai Wei, Zhengliang Liu, Lei Guo, Xiaoyan Cai, Shu Zhang, Tuo Zhang, et al. Impressiongpt: an iterative optimizing framework for radiology report summarization with chatgpt. *arXiv preprint arXiv:2304.08448*, 2023.

[21] Jason Holmes, Lian Zhang, Yuzhen Ding, Hongying Feng, Zhengliang Liu, Tianming Liu, William W Wong, Sujay A Vora, Jonathan B Ashman, and Wei Liu. Benchmarking a foundation llm on its ability to re-label structure names in accordance with the aapm tg-263 report. *arXiv preprint arXiv:2310.03874*, 2023.

[22] Zhengliang Liu, Yiwei Li, Peng Shu, Aoxiao Zhong, Longtao Yang, Chao Ju, Zihao Wu, Chong Ma, Jie Luo, Cheng Chen, et al. Radiology-llama2: Best-in-class large language model for radiology. *arXiv preprint arXiv:2309.06419*, 2023.

[23] Zhengliang Liu, Peilong Wang, Yiwei Li, Jason Holmes, Peng Shu, Lian Zhang, Chenbin Liu, Ninghao Liu, Dajiang Zhu, Xiang Li, et al. Radonc-gpt: A large language model for radiation oncology. *arXiv preprint arXiv:2309.10160*, 2023.

[24] Zhengliang Liu, Yiwei Li, Qian Cao, Junwen Chen, Tianze Yang, Zihao Wu, John Hale, John Gibbs, Khaled Rasheed, Ninghao Liu, et al. Transformation vs tradition: Artificial general intelligence (agi) for arts and humanities. *arXiv preprint arXiv:2310.19626*, 2023.

[25] Zhengliang Liu, Tianyang Zhong, Yiwei Li, Yutong Zhang, Yi Pan, Zihao Zhao, Peixin Dong, Chao Cao, Yuxiao Liu, Peng Shu, et al. Evaluating large language models for radiology natural language processing. *arXiv preprint arXiv:2307.13693*, 2023.

[26] Z Liu, A Zhong, Y Li, L Yang, C Ju, Z Wu, et al. Radiology-gpt: a large language model for radiology. arxiv [preprint]. 2023 [cited august 21, 2023].

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[28] Stephen Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013.

[29] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[30] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[31] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[32] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. *OpenAI*, 2018.

[33] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[34] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.

[35] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[36] Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, et al. Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv preprint arXiv:2307.15217*, 2023.

[37] Sukhpal Singh Gill and Rupinder Kaur. Chatgpt: Vision and challenges. *Internet of Things and Cyber-Physical Systems*, 3:262–271, 2023.

[38] Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Qiang Liu, et al. Language is not all you need: Aligning perception with language models. *arXiv preprint arXiv:2302.14045*, 2023.

[39] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.

[40] GPT-4V(ision) system card — openai.com. https://openai.com/research/gpt-4v-system-card. [Accessed 08-11-2023].

[41] Zhanyu Wang, Lingqiao Liu, Lei Wang, and Luping Zhou. R2gengpt: Radiology report generation with frozen llms. *arXiv preprint arXiv:2309.09812*, 2023.

[42] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, et al. Towards expert-level medical question answering with large language models. *arXiv preprint arXiv:2305.09617*, 2023.

[43] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015.

[44] M Bertero and Michele Piana. Inverse problems in biomedical imaging: modeling and methods of solution. *Complex systems in biomedicine*, pages 1–33, 2006.

[45] Ning Qiang, Jie Gao, Qinglin Dong, Huiji Yue, Hongtao Liang, Lili Liu, Jingjing Yu, Jing Hu, Shu Zhang, Bao Ge, et al. Functional brain network identification and fmri augmentation using a vae-gan framework. *Computers in Biology and Medicine*, 165:107395, 2023.

[46] Xia-An Bi, Ke Chen, Siyu Jiang, Sheng Luo, Wenyan Zhou, Zhaoxu Xing, Luyun Xu, Zhengliang Liu, and Tianming Liu. Community graph convolution neural network for alzheimer's disease classification and pathogenetic factors identification. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[47] Yiheng Liu, Enjie Ge, Mengshen He, Zhengliang Liu, Shijie Zhao, Xintao Hu, Dajiang Zhu, Tianming Liu, and Bao Ge. Discovering dynamic functional brain networks via spatial and channel-wise attention. *arXiv preprint arXiv:2205.09576*, 2022.

[48] Lian Zhang, Jason M Holmes, Zhengliang Liu, Sujay A Vora, Terence T Sio, Carlos E Vargas, Nathan Y Yu, Sameer R Keole, Steven E Schild, Martin Bues, et al. Beam mask and sliding window-facilitated deep learning-based accurate and efficient dose prediction for pencil beam scanning proton therapy. *arXiv preprint arXiv:2305.18572*, 2023.

[49] Shu Zhang, Enze Shi, Lin Wu, Ruoyang Wang, Sigang Yu, Zhengliang Liu, Shaochen Xu, Tianming Liu, and Shijie Zhao. Differentiating brain states via multi-clip random fragment strategy-based interactive bidirectional recurrent neural network. *Neural Networks*, 165:1035–1049, 2023.

[50] Yuzhen Ding, Hongying Feng, Yunze Yang, Jason Holmes, Zhengliang Liu, David Liu, William W Wong, Nathan Y Yu, Terence T Sio, Steven E Schild, et al. Deep-learning based fast and accurate 3d ct deformable image registration in lung cancer. *Medical Physics*, 2023.

[51] Ning Qiang, Jie Gao, Qinglin Dong, Jin Li, Shu Zhang, Hongtao Liang, Yifei Sun, Bao Ge, Zhengliang Liu, Zihao Wu, et al. A deep learning method for autism spectrum disorder identification based on interactions of hierarchical brain networks. *Behavioural Brain Research*, 452:114603, 2023.

[52] Haixing Dai, Qing Li, Lin Zhao, Liming Pan, Cheng Shi, Zhengliang Liu, Zihao Wu, Lu Zhang, Shijie Zhao, Xia Wu, et al. Graph representation neural architecture search for optimal spatial/temporal functional brain network decomposition. In *International Workshop on Machine Learning in Medical Imaging*, pages 279–287. Springer Nature Switzerland Cham, 2022.

[53] Trafton Drew, Karla Evans, Melissa L-H Võ, Francine L Jacobson, and Jeremy M Wolfe. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, 33(1):263–274, 2013.

[54] A Van der Gijp, CJ Ravesloot, H Jarodzka, MF Van der Schaaf, IC Van der Schaaf, Jan PJ van Schaik, and Th J Ten Cate. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education*, 22:765–787, 2017.

[55] Zihao Wu, Lu Zhang, Chao Cao, Xiaowei Yu, Haixing Dai, Chong Ma, Zhengliang Liu, Lin Zhao, Gang Li, Wei Liu, et al. Exploring the trade-offs: Unified large language models vs local fine-tuned models for highly-specific radiology nli task. *arXiv preprint arXiv:2304.09138*, 2023.

[56] Tianyang Zhong, Wei Zhao, Yutong Zhang, Yi Pan, Peixin Dong, Zuowei Jiang, Xiaoyan Kui, Youlan Shang, Li Yang, Yaonai Wei, et al. Chatradio-valuer: A chat large language model for generalizable radiology report generation based on multi-institution and multi-system data. *arXiv preprint arXiv:2310.05242*, 2023.

[57] Rebecca S Crowley, Gregory J Naus, Jimmie Stewart III, and Charles P Friedman. Development of visual diagnostic expertise in pathology: an information-processing study. *Journal of the American Medical Informatics Association*, 10(1):39–51, 2003.

[58] Thomas Jaarsma, Halszka Jarodzka, Marius Nap, Jeroen JG van Merriënboer, and Henny PA Boshuizen. Expertise in clinical pathology: Combining the visual and cognitive perspective. *Advances in Health Sciences Education*, 20:1089–1106, 2015.

[59] Aaron Maurer. An early prediction of lung cancer using ct scan images. *Journal of Computing and Natural Science*, 1:39–44, 2021.

[60] Suren Makaju, PWC Prasad, Abeer Alsadoon, AK Singh, and A Elchouemi. Lung cancer detection using ct scan images. *Procedia Computer Science*, 125:107–114, 2018.

[61] Onno M Mets, Constantinus FM Buckens, Pieter Zanen, Ivana Isgum, Bram van Ginneken, Mathias Prokop, Hester A Gietema, Jan-Willem J Lammers, Rozemarijn Vliegenthart, Matthijs Oudkerk, et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. *Jama*, 306(16):1775–1781, 2011.

[62] Michael A DeTure and Dennis W Dickson. The neuropathological diagnosis of alzheimer's disease. *Molecular neurodegeneration*, 14(1):1–18, 2019.

[63] Fuyong Xing and Lin Yang. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE reviews in biomedical engineering*, 9:234–263, 2016.

[64] Natalya Uranova, Diana Orlovskaya, Olga Vikhreva, Ivetta Zimina, Natalya Kolomeets, Victor Vostrikov, and Valentina Rachmanova. Electron microscopy of oligodendroglia in severe mental illness. *Brain research bulletin*, 55(5):597–610, 2001.

[65] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 2425–2433, 2015.

[66] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[67] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010.

[68] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-

scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[69] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[70] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[71] Haixing Dai, Lu Zhang, Lin Zhao, Zihao Wu, Zhengliang Liu, David Liu, Xiaowei Yu, Yanjun Lyu, Changying Li, Ninghao Liu, et al. Hierarchical semantic tree concept whitening for interpretable image classification. *arXiv preprint arXiv:2307.04343*, 2023.

[72] Haixing Dai, Chong Ma, Zhengliang Liu, Yiwei Li, Peng Shu, Xiaozheng Wei, Lin Zhao, Zihao Wu, Dajiang Zhu, Wei Liu, et al. Samaug: Point prompt augmentation for segment anything model. *arXiv preprint arXiv:2307.01187*, 2023.

[73] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421*, 9, 2023.

[74] Kunming Cheng, Qiang Guo, Yongbin He, Yanqiu Lu, Shuqin Gu, and Haiyang Wu. Exploring the potential of gpt-4 in biomedical engineering: the dawn of a new era. *Annals of Biomedical Engineering*, pages 1–9, 2023.

[75] Ding-Qiao Wang, Long-Yu Feng, Jin-Guo Ye, Jin-Gen Zou, and Ying-Feng Zheng. Accelerating the integration of chatgpt and other large-scale ai models into biomedical research and healthcare. *MedComm–Future Medicine*, 2(2):e43, 2023.

[76] Zhichao Yang, Zonghai Yao, Mahbuba Tasmin, Parth Vashisht, Won Seok Jang, Beining Wang, Dan Berlowitz, and Hong Yu. Performance of multimodal gpt-4v on usmle with image: Potential for imaging diagnostic support with explanations. *medRxiv*, pages 2023–10, 2023.

[77] Lin Zhao, Zihao Wu, Haixing Dai, Zhengliang Liu, Xintao Hu, Tuo Zhang, Dajiang Zhu, and Tianming Liu. A generic framework for embedding human brain function with temporally correlated autoencoder. *Medical Image Analysis*, 89:102892, 2023.

[78] Mengyue Zhou, Xu Liu, David Liu, Zihao Wu, Zhengliang Liu, Lin Zhao, Dajiang Zhu, Lei Guo, Junwei Han, Tianming Liu, et al. Fine-grained artificial neurons in audio-transformers for disentangling neural auditory encoding. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7943–7956, 2023.

[79] Hongmin Cai, Xiaoke Huang, Zhengliang Liu, Wenxiong Liao, Haixing Dai, Zihao Wu, Dajiang Zhu, Hui Ren, Quanzheng Li, Tianming Liu, et al. Exploring multimodal approaches for alzheimer's disease detection using patient speech transcript and audio data. *arXiv preprint arXiv:2307.02514*, 2023.

[80] Fatima Merchant and Kenneth Castleman. *Microscope image processing*. Academic press, 2022.

[81] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):4550–4568, 2018.

[82] Maha Gharaibeh, Dalia Alzu'bi, Malak Abdullah, Ismail Hmeidi, Al Nasar, Mohammad Rustom, Laith Abualigah, and Amir H Gandomi. Radiology imaging scans for early diagnosis of kidney tumors: a review of data analytics-based machine learning and deep learning approaches. *Big Data and Cognitive Computing*, 6(1):29, 2022.

[83] José Cornejo, Jorge A Cornejo-Aguilar, Mariela Vargas, Carlos G Helguero, Rafhael Milanezi de Andrade, Sebastian Torres-Montoya, Javier Asensio-Salazar, Alvaro Rivero Calle, Jaime Martínez Santos, Aaron Damon, et al. Anatomical engineering and 3d printing for surgery and medical devices: International review and future exponential innovations. *BioMed research international*, 2022, 2022.

[84] Mark Rodrigues, Michael Mayo, and Panos Patros. Surgical tool datasets for machine learning research: a survey. *International Journal of Computer Vision*, 130(9):2222–2248, 2022.

[85] Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317, 2019.

[86] Neda Bagheri, Anne E Carpenter, Emma Lundberg, Anne L Plant, and Rick Horwitz. The new era of quantitative cell imaging—challenges and opportunities. *Molecular Cell*, 82(2):241–247, 2022.

[87] Hoi Yi Leung, Martin Ho Yin Yeung, Wai Tung Leung, King Hin Wong, Wai Yan Tang, William Chi Shing Cho, Heong Ting Wong, Hin Fung Tsang, Yin Kwan Evelyn Wong, Xiao Meng Pei, et al. The current and future applications of in situ hybridization technologies in anatomical pathology. *Expert Review of Molecular Diagnostics*, 22(1):5–18, 2022.

[88] Aaron Loh, David Gwun, Clement T Chow, Alexandre Boutet, Jordy Tasserie, Jürgen Germann, Brendan Santyr, Gavin Elias, Kazuaki Yamamoto, Can Sarica, et al. Probing responses to deep brain stimulation with functional magnetic resonance imaging. *Brain stimulation*, 15(3):683–694, 2022.

[89] Zhengliang Liu, Xiaowei Yu, Lu Zhang, Zihao Wu, Chao Cao, Haixing Dai, Lin Zhao, Wei Liu, Dinggang Shen, Quanzheng Li, et al. Deid-gpt: Zero-shot medical text de-identification by gpt-4. *arXiv preprint arXiv:2303.11032*, 2023.

[90] Shahana Rayhan. Ethical implications of creating agi: Impact on human society, privacy, and power dynamics. *Artificial Intelligence Review*, 2023.

[91] Chaoning Zhang, Chenshuang Zhang, Chenghao Li, Yu Qiao, Sheng Zheng, Sumit Kumar Dam, Mengchun Zhang, Jung Uk Kim, Seong Tae Kim, Jinwoo Choi, et al. One small step for generative ai, one giant leap for agi: A complete survey on chatgpt in aigc era. *arXiv preprint arXiv:2304.06488*, 2023.

[92] PS Mishra-Kalyani, L Amiri Kordestani, DR Rivera, H Singh, A Ibrahim, RA DeClaro, Y Shen, S Tang, R Sridhara, PG Kluetz, et al. External control arms in oncology: current use and future directions. *Annals of Oncology*, 33(4):376–383, 2022.

[93] Surendrabikram Thapa and Surabhi Adhikari. Chatgpt, bard, and large language models for biomedical research: Opportunities and pitfalls. *Annals of Biomedical Engineering*, pages 1–5, 2023.

[94] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019.

[95] Daniel Kermany, Kang Zhang, Michael Goldbaum, et al. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley data*, 2(2):651, 2018.

[96] Anas M Tahir, Muhammad EH Chowdhury, Yazan Qiblawey, Amith Khandakar, Tawsifur Rahman, Serkan Kiranyaz, Uzair Khurshid, Nabil Ibtehaz, Sakib Mahmud, and Maymouna Ezeddin. Covid-qu-ex dataset, 2022. *URL https://www. kaggle. com/dsv/3122958*, 11, 2022.

[97] Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2016.

[98] SIIM-ACR pneumothorax segmentation, 2020. [online] Available: https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation.

[99] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[100] Hanchuan Peng, Peng Xie, Lijuan Liu, Xiuli Kuang, Yimin Wang, Lei Qu, Hui Gong, Shengdian Jiang, Anan Li, Zongcai Ruan, et al. Morphological diversity of single neurons in molecularly defined cell types. *Nature*, 598(7879):174–181, 2021.

[101] Hui Gong, Dongli Xu, Jing Yuan, Xiangning Li, Congdi Guo, Jie Peng, Yuxin Li, Lindsay A Schwarz, Anan Li, Bihe Hu, et al. High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature communications*, 7(1):12142, 2016.

[102] Hanchuan Peng, Alessandro Bria, Zhi Zhou, Giulio Iannello, and Fuhui Long. Extensible visualization and analysis for multidimensional images using vaa3d. *Nature protocols*, 9(1):193–208, 2014.

[103] Zizhen Yao, Cindy TJ van Velthoven, Michael Kunst, Meng Zhang, Delissa McMillen, Changkyu Lee, Won Jung, Jeff Goldy, Aliya Abdelhak, Pamela Baker, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *bioRxiv*, 2023.

[104] S. V. Zolotova, A. V. Golanov, I. N. Pronin, A. V. Dalechina, A. A. Nikolaeva, A. S. Belyashova, D. Y. Usachev, E. A. Kondrateva, P. V. Druzhinina, B. N. Shirokikh, T. N. Saparov, M. G. Belyaev, and A. I. Kurmukov. Burdenko's glioblastoma progression dataset (burdenko-gbm-progression), 2023.

[105] Ali Gooya, Kilian M Pohl, Michel Bilello, Luigi Cirillo, George Biros, Elias R Melhem, and Christos Davatzikos. Glistr: glioma image segmentation and registration. *IEEE transactions on medical imaging*, 31(10):1941–1954, 2012.

[106] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.

[107] P. Li, S. Wang, T. Li, J. Lu, Y. HuangFu, and D. Wang. A large-scale ct and pet/ct dataset for lung cancer diagnosis (lung-pet-ct-dx). https://doi.org/10.7937/TCIA.2020.NNC2-0461, 2020.

[108] Chris Rose, Daniele Turi, Alan Williams, Katy Wolstencroft, and Chris Taylor. Web services for the ddsm and digital mammography research. In *Digital Mammography: 8th International Workshop, IWDM 2006, Manchester, UK, June 18-21, 2006. Proceedings 8*, pages 376–383. Springer, 2006.

[109] Andrew A. Borkowski, Marilyn M. Bui, L. Brannon Thomas, Catherine P. Wilson, Lauren A. DeLand, and Stephen M. Mastorides. Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142*, 2019.

[110] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, and Azamossadat Hosseini. Acute Lymphoblastic Leukemia (ALL) image dataset. *Kaggle*, 2021.

[111] Huazhu Fu, Fei Li, José Ignacio Orlando, Hrvoje Bogunović, Xu Sun, Jingan Liao, Yanwu Xu, Shaochong Zhang, and Xiulan Zhang. Adam: Automatic detection challenge on age-related macular degeneration, 2020.

[112] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. Indian diabetic retinopathy image dataset (idrid), 2018.

[113] Huazhu Fu, Fei Li, José Ignacio Orlando, Hrvoje Bogunović, Xu Sun, Jingan Liao, Yanwu Xu, Shaochong Zhang, and Xiulan Zhang. Palm: Pathologic myopia challenge, 2019.

[114] José Ignacio Orlando, Huazhu Fu, João Barbosa Breda, Karel van Keer, Deepti R. Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, Joonseok Lee, Xiaoxiao Li, Peng Liu, Shuai Lu, Balamurali Murugesan, Valery Naranjo, Sai Samarth R. Phaye, Sharath M. Shankaranarayana, Apoorva Sikka, Jaemin Son, Anton van den Hengel, Shujun Wang, Junyan Wu, Zifeng Wu, Guanghui Xu, Yongli Xu, Pengshuai Yin, Fei Li, Xiulan Zhang, Yanwu Xu, and Hrvoje Bogunović. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59:101570, 2020.

[115] Fei Li, Diping Song, Han Chen, Jian Xiong, Xingyi Li, Hua Zhong, Guangxian Tang, Sujie Fan, Dennis SC Lam, Weihua Pan, et al. Development and clinical deployment of a smartphone-based visual field deep learning system for glaucoma detection. *NPJ digital medicine*, 3(1):123, 2020.

[116] Chinedu Innocent Nwoye, Deepak Alapatt, Tong Yu, Armine Vardazaryan, Fangfang Xia, Zixuan Zhao, Tong Xia, Fucang Jia, Yuxuan Yang, Hao Wang, et al. Cholectriplet2021: A benchmark challenge for surgical action triplet recognition. *Medical Image Analysis*, 86:102803, 2023.

[117] Chinedu Innocent Nwoye, Tong Yu, Cristians Gonzalez, Barbara Seeliger, Pietro Mascagni, Didier Mutter, Jacques Marescaux, and Nicolas Padoy. Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Medical Image Analysis*, 78:102433, 2022.

[118] Ronald Carl Petersen, Paul S. Aisen, Laurel A. Beckett, Michael C. Donohue, Anthony Collins

Gamst, and Danielle J. et al. Harvey. Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74, 2010.

[119] C.R. Jack Jr, M.A. Bernstein, N.C. Fox, Alexander Thompson, P., G. Alexander, D. Harvey, B. Borowski, P.J. Britson, J. L. Whitwell, and A.M. Dale. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27, 2008.

[120] Willy W. Wong.

[121] Martin Maška, Vladimír Ulman, Pablo Delgado-Rodriguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A Guerrero Peña, Tsang Ing Ren, Elliot M Meyerowitz, Tim Scherr, Katharina Löffler, et al. The cell tracking challenge: 10 years of objective benchmarking. *Nature Methods*, pages 1–11, 2023.

[122] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.

[123] Ashkan Ebadi, Pengcheng Xi, Alexander MacLean, Stéphane Tremblay, Sonny Kohli, and Alexander Wong. Covidx-us - an open-access benchmark dataset of ultrasound imaging data for ai-driven covid-19 analytics. *arXiv:2103.10003*, 2021.

[124] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, Feb 2020.

[125] Hedvig Hricak, May Abdel-Wahab, Rifat Atun, Miriam Mikhail Lette, Diana Paez, James A Brink, Lluís Donoso-Bach, Guy Frija, Monika Hierath, Ola Holmberg, et al. Medical imaging and nuclear medicine: a lancet oncology commission. *The Lancet Oncology*, 22(4):e136–e172, 2021.

[126] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*, pages 451–462. Springer, 2020.

[127] Salman Maqbool, Aqsa Riaz, Hasan Sajid, and Osman Hasan. m2caiseg: Semantic segmentation of laparoscopic images using convolutional neural networks. *arXiv preprint arXiv:2008.10134*, 2020.

[128] Andru P Twinanda, Sherif Shehata, Didier Mutter, Jacques Marescaux, Michel De Mathelin, and Nicolas Padoy. Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging*, 36(1):86–97, 2016.

[129] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

[130] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, S. Langer, K. Lioprys, J. Malvehy, S. Musthaq, J. Nanda, O. Reiter, G. Shih, A. Stratigos, P. Tschandl, J. Weber, and P. Soyer. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci Data*, 8(34):1–15, 2021.

[131] C. Brown. In situ hybridization with riboprobes: An overview for veterinary pathologists. *Veterinary Pathology*, 35(3):159–167, 1998. PMID: 9598579.

[132] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[133] Darren J Burgess. Spatial transcriptomics coming of age. *Nature Reviews Genetics*, 20(6):317–317, 2019.

[134] Tomomi Shimogori, Ayumi Abe, Yasuhiro Go, Tsutomu Hashikawa, Noriyuki Kishi, Satomi S Kikuchi, Yoshiaki Kita, Kimie Niimi, Hirozumi Nishibe, Misako Okuno, et al. Digital gene atlas of neonate common marmoset brain. *Neuroscience research*, 128:1–13, 2018.

[135] Yoshiaki Kita, Hirozumi Nishibe, Yan Wang, Tsutomu Hashikawa, Satomi S Kikuchi, Mami U, Aya C Yoshida, Chihiro Yoshida, Takashi Kawase, Shin Ishii, et al. Cellular-resolution gene expression profiling in the neonatal marmoset brain reveals dynamic species-and region-specific differences. *Proceedings of the National Academy of Sciences*, 118(18):e2020125118, 2021.

[136] Joel N. Hirschhorn and Mark J. Daly. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2):95–108, Feb 2005.

[137] David J Balding. A tutorial on statistical methods for population association studies. *Nature reviews genetics*, 7(10):781–791, 2006.

[138] Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina De Vries, Yukinori Okada, Alicia R Martin, Hilary C Martin, Tuuli Lappalainen, and Danielle Posthuma. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1):59, 2021.

[139] Peter M. Visscher, Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 10 years of gwas discovery: Biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22, 2017.

[140] Katrina L Grasby, Neda Jahanshad, Jodie N Painter, Lucía Colodro-Conde, Janita Bralten, Derrek P Hibar, Penelope A Lind, Fabrizio Pizzagalli, Christopher RK Ching, Mary Agnes B McMahon, et al. The genetic architecture of the human cerebral cortex. *Science*, 367(6484):eaay6690, 2020.

[141] Yifan Zhang, Fan Ye, Lingxiao Chen, Feng Xu, Xiaodiao Chen, Hongkun Wu, Mingguo Cao, Yunxiang Li, Yaqi Wang, and Xingru Huang. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data*, 10(1):380, 2023.

[142] Ibrahim Ethem Hamamci, Sezgin Er, Enis Simsar, Anjany Sekuboyina, Mustafa Gundogar, Bernd Stadlinger, Albert Mehl, and Bjoern Menze. Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays. *arXiv preprint arXiv:2303.06500*, 2023.

[143] Zhengliang Liu, Xinyu He, Lei Liu, Tianming Liu, and Xiaoming Zhai. Context matters: A strategy to pre-train language model for science education. *arXiv preprint arXiv:2301.12031*, 2023.

[144] Saed Rezayi, Haixing Dai, Zhengliang Liu, Zihao Wu, Akarsh Hebbar, Andrew H. Burns, Lin Zhao, Dajiang Zhu, Xiang Li, Quanzheng Li, et al. Clinicalradiobert: Knowledge-infused few shot learning for clinical notes named entity recognition. In *The 13th International Workshop on Machine Learning in Medical Imaging (MLMI 2022)*, 2022.

[145] Wenxiong Liao, Zhengliang Liu, Haixing Dai, Zihao Wu, Yiyang Zhang, Xiaoke Huang, Yuzhong Chen, Xi Jiang, Dajiang Zhu, Tianming Liu, et al. Mask-guided bert for few shot text classification. *arXiv preprint arXiv:2302.10447*, 2023.

[146] Homgmin Cai, Wenxiong Liao, Zhengliang Liu, Xiaoke Huang, Yiyang Zhang, Siqi Ding, Sheng Li, Quanzheng Li, Tianming Liu, and Xiang Li. Coarse-to-fine knowledge graph domain adaptation based on distantly-supervised iterative training. *arXiv preprint arXiv:2211.02849*, 2022.

[147] Lin Zhao, Zihao Wu, Haixing Dai, Zhengliang Liu, Tuo Zhang, Dajiang Zhu, and Tianming Liu. Embedding human brain function via transformer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 366–375. Springer Nature Switzerland Cham, 2022.

[148] Y Ding, Z Liu, H Feng, J Holmes, Y Yang, N Yu, T Sio, S Schild, B Li, and W Liu. Accurate and efficient deep neural network based deformable image registration method in lung cancer. In *MEDICAL PHYSICS*, volume 49, pages E148–E148. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2022.

[149] Saed Rezayi, Zhengliang Liu, Zihao Wu, Chandra Dhakal, Bao Ge, Chen Zhen, Tianming Liu, and Sheng Li. Agribert: Knowledge-infused agricultural language models for matching food and nutrition. In *International Joint Conference on Artificial Intelligence*, 2022.

[150] Chenhao Tang, Zhengliang Liu, Chong Ma, Zihao Wu, Yiwei Li, Wei Liu, Dajiang Zhu, Quanzheng Li, Xiang Li, Tianming Liu, et al. Policygpt: Automated analysis of privacy policies with large language models. *arXiv preprint arXiv:2309.10238*, 2023.

[151] Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Wei Liu, Ninghao Liu, et al. Auggpt: Leveraging chatgpt for text data augmentation. *arXiv preprint arXiv:2302.13007*, 2023.

[152] OpenAI. Gpt-4 technical report. *arXiv*, 2023.

[153] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[154] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*, 2022.

[155] Xiaoyan Cai, Sen Liu, Junwei Han, Libin Yang, Zhenguo Liu, and Tianming Liu. Chestxraybert: A pretrained language model for chest radiology report summarization. *IEEE Transactions on Multimedia*, 2021.

[156] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*, 2019.

[157] Zhengliang Liu, Yiwei Li, Peng Shu, Aoxiao Zhong, Longtao Yang, Chao Ju, Zihao Wu, Chong Ma, Jie Luo, Cheng Chen, et al. Radiology-llama2: Best-in-class large language model for radiology. *arXiv preprint arXiv:2309.06419*, 2023.

[158] Honglin Xiong, Sheng Wang, Yitao Zhu, Zihao Zhao, Yuxiao Liu, Qian Wang, and Dinggang

Shen. Doctorglm: Fine-tuning your chinese doctor is not a herculean task. *arXiv preprint arXiv:2304.01097*, 2023.

[159] Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. Glm: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335, 2022.

[160] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

[161] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pages 10347–10357. PMLR, 2021.

[162] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[163] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16000–16009, 2022.

[164] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9640–9649, 2021.

[165] Agrim Gupta, Stephen Tian, Yunzhi Zhang, Jiajun Wu, Roberto Martín-Martín, and Li Fei-Fei. Maskvit: Masked visual pre-training for video prediction. *arXiv preprint arXiv:2206.11894*, 2022.

[166] Chong Ma, Lin Zhao, Yuzhong Chen, Sheng Wang, Lei Guo, Tuo Zhang, Dinggang Shen, Xi Jiang, and Tianming Liu. Eye-gaze-guided vision transformer for rectifying shortcut learning. *IEEE Transactions on Medical Imaging*, 2023.

[167] Chong Ma, Lin Zhao, Yuzhong Chen, Lei Guo, Tuo Zhang, Xintao Hu, Dinggang Shen, Xi Jiang, and Tianming Liu. Rectify vit shortcut learning by visual saliency. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[168] Xiaowei Yu, Lu Zhang, Haixing Dai, Yanjun Lyu, Lin Zhao, Zihao Wu, David Liu, Tianming Liu, and Dajiang Zhu. Core-periphery principle guided redesign of self-attention in transformers. *arXiv preprint arXiv:2303.15569*, 2023.

[169] Zhenxiang Xiao, Yuzhong Chen, Lu Zhang, Junjie Yao, Zihao Wu, Xiaowei Yu, Yi Pan, Lin Zhao, Chong Ma, Xinyu Liu, et al. Instruction-vit: Multi-modal prompts for instruction learning in vit. *arXiv preprint arXiv:2305.00201*, 2023.

[170] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[171] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[172] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*, pages 12888–12900. PMLR, 2022.

[173] Wonjae Kim, Bokyung Son, and Ildoo Kim. Vilt: Vision-and-language transformer without convolution or region supervision. In *International Conference on Machine Learning*, pages 5583–5594. PMLR, 2021.

[174] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*, 2022.

[175] Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, et al. Image as a foreign language: Beit pretraining for vision and vision-language tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19175–19186, 2023.

[176] Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. Medclip: Contrastive learning from unpaired medical images and text. *arXiv preprint arXiv:2210.10163*, 2022.

[177] Shih-Cheng Huang, Liyue Shen, Matthew P Lungren, and Serena Yeung. Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3942–3951, 2021.

[178] Yaonai Wei, Tuo Zhang, Han Zhang, Tianyang Zhong, Lin Zhao, Zhengliang Liu, Chong Ma, Songyao Zhang, Muheng Shang, Lei Du, et al. Chat2brain: A method for mapping open-ended semantic queries to brain activation maps. *arXiv preprint arXiv:2309.05021*, 2023.

[179] Zhengliang Liu, Zihao Wu, Mengxuan Hu, Bokai Zhao, Lin Zhao, Tianyi Zhang, Haixing Dai, Xianyan Chen, Ye Shen, Sheng Li, et al. Pharmacygpt: The ai pharmacist. *arXiv preprint arXiv:2307.10432*, 2023.

[180] Fei Dou, Jin Ye, Geng Yuan, Qin Lu, Wei Niu, Haijian Sun, Le Guan, Guoyu Lu, Gengchen Mai, Ninghao Liu, et al. Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges. *arXiv preprint arXiv:2309.07438*, 2023.

[181] Zhengliang Liu, Lu Zhang, Zihao Wu, Xiaowei Yu, Chao Cao, Haixing Dai, Ninghao Liu, Jun Liu, Wei Liu, Quanzheng Li, et al. Surviving chatgpt in healthcare. *Frontiers in Radiology*, 3:1224682, 2023.

[182] Zihan Guan, Zihao Wu, Zhengliang Liu, Dufan Wu, Hui Ren, Quanzheng Li, Xiang Li, and Ninghao Liu. Cohortgpt: An enhanced gpt for participant recruitment in clinical study. *arXiv preprint arXiv:2307.11346*, 2023.

[183] Tianyang Zhong, Yaonai Wei, Li Yang, Zihao Wu, Zhengliang Liu, Xiaozheng Wei, Wenjun Li, Junjie Yao, Chong Ma, Xiang Li, et al. Chatabl: Abductive learning via natural language interaction with chatgpt. *arXiv preprint arXiv:2304.11107*, 2023.

[184] Lin Zhao, Lu Zhang, Zihao Wu, Yuzhong Chen, Haixing Dai, Xiaowei Yu, Zhengliang Liu, Tuo Zhang, Xintao Hu, Xi Jiang, et al. Meta-radiology. *Meta*, 1:100005, 2023.

[185] Haixing Dai, Yiwei Li, Zhengliang Liu, Lin Zhao, Zihao Wu, Suhang Song, Ye Shen, Dajiang Zhu, Xiang Li, Sheng Li, et al. Ad-autogpt: An autonomous gpt for alzheimer's disease infodemiology. *arXiv preprint arXiv:2306.10095*, 2023.

[186] Yucheng Shi, Shaochen Xu, Zhengliang Liu, Tianming Liu, Xiang Li, and Ninghao Liu. Mededit: Model editing for medical question answering with external knowledge bases. *arXiv preprint arXiv:2309.16035*, 2023.

[187] Saed Rezayi, Zhengliang Liu, Zihao Wu, Chandra Dhakal, Bao Ge, Haixing Dai, Gengchen Mai, Ninghao Liu, Chen Zhen, Tianming Liu, et al. Exploring new frontiers in agricultural nlp: Investigating the potential of large language models for food applications. *arXiv preprint arXiv:2306.11892*, 2023.

[188] Wenxiong Liao, Zhengliang Liu, Haixing Dai, Shaochen Xu, Zihao Wu, Yiyang Zhang, Xiaoke Huang, Dajiang Zhu, Hongmin Cai, Tianming Liu, et al. Differentiate chatgpt-generated and human-written medical texts. *arXiv preprint arXiv:2304.11567*, 2023.

[189] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023.

[190] Qinghao Ye, Haiyang Xu, Guohai Xu, Jiabo Ye, Ming Yan, Yiyang Zhou, Junyang Wang, Anwen Hu, Pengcheng Shi, Yaya Shi, Chenliang Li, Yuanhong Xu, Hehong Chen, Junfeng Tian, Qian Qi, Ji Zhang, and Fei Huang. mplug-owl: Modularization empowers large language models with multimodality. *arXiv preprint arXiv:2304.14178*, 2023.

[191] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. Visual chatgpt: Talking, drawing and editing with visual foundation models. *arXiv preprint arXiv:2303.04671*, 2023.

[192] Zhengyuan Yang, Linjie Li, Jianfeng Wang, Kevin Lin, Ehsan Azarnasab, Faisal Ahmed, Zicheng Liu, Ce Liu, Michael Zeng, and Lijuan Wang. Mm-react: Prompting chatgpt for multimodal reasoning and action. *arXiv preprint arXiv:2303.11381*, 2023.

[193] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikoł aj Bińkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karén Simonyan. Flamingo: a visual language model for few-shot learning. *Advances in Neural Information Processing Systems*, 35:23716–23736, 2022.

[194] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023.

[195] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol

Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.

[196] Tao Gong, Chengqi Lyu, Shilong Zhang, Yudong Wang, Miao Zheng, Qian Zhao, Kuikun Liu, Wenwei Zhang, Ping Luo, and Kai Chen. Multimodal-gpt: A vision and language model for dialogue with humans. *arXiv preprint arXiv:2305.04790*, 2023.

[197] Bo Li, Yuanhan Zhang, Liangyu Chen, Jinghao Wang, Jingkang Yang, and Ziwei Liu. Otter: A multi-modal model with in-context instruction tuning. *arXiv preprint arXiv:2305.03726*, 2023.

[198] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023.

[199] Chaoyi Wu, Jiayu Lei, Qiaoyu Zheng, Weike Zhao, Weixiong Lin, Xiaoman Zhang, Xiao Zhou, Ziheng Zhao, Ya Zhang, Yanfeng Wang, and Weidi Xie. Can gpt-4v(ision) serve medical applications? case studies on gpt-4v for multimodal medical diagnosis. *arXiv preprint arXiv:2310.09909*, 2023.

[200] Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. Instruction tuning for large language models: A survey. *arXiv preprint arXiv:2308.10792*, 2023.

[201] Jiaqi Wang, Zhengliang Liu, Lin Zhao, Zihao Wu, Chong Ma, Sigang Yu, Haixing Dai, Qiushi Yang, Yiheng Liu, Songyao Zhang, et al. Review of large vision models and visual prompt engineering. *arXiv preprint arXiv:2307.00855*, 2023.

[202] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021.

[203] Peng Wang, An Yang, Rui Men, Junyang Lin, Shuai Bai, Zhikang Li, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. In *International Conference on Machine Learning*, pages 23318–23340. PMLR, 2022.

[204] Zhiyang Xu, Ying Shen, and Lifu Huang. Multiinstruct: Improving multi-modal zero-shot learning via instruction tuning. *arXiv preprint arXiv:2212.10773*, 2022.

[205] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[206] Batu Ozturkler, Nikolay Malkin, Zhen Wang, and Nebojsa Jojic. Thinksum: Probabilistic reasoning over sets using large language models. *arXiv preprint arXiv:2210.01293*, 2022.

[207] Zhengliang Liu, Aoxiao Zhong, Yiwei Li, Longtao Yang, Chao Ju, Zihao Wu, Chong Ma, Peng Shu, Cheng Chen, Sekeun Kim, et al. Tailoring large language models to radiology: A preliminary approach to llm adaptation for a highly specialized domain. In *International Workshop on Machine Learning in Medical Imaging*, pages 464–473. Springer, 2023.

[208] NR Bhatt, E Dunne, M Faraz, AE Gillis, KC Conlon, S Paran, and PF Ridgway. Trends in the use of laparoscopic versus open paediatric appendicectomy: A regional 12-year study and a national survey. *World Journal of Surgery*, 42:3792–3802, 2018.

[209] Zijian Zhao, Tongbiao Cai, Faliang Chang, and Xiaolin Cheng. Real-time surgical instrument detection in robot-assisted surgery using a convolutional neural network cascade. *Healthcare technology letters*, 6(6):275–279, 2019.

[210] Thomas M Ward, Pietro Mascagni, Yutong Ban, Guy Rosman, Nicolas Padoy, Ozanan Meireles, and Daniel A Hashimoto. Computer vision in surgery. *Surgery*, 169(5):1253–1256, 2021.

[211] Beatrice van Amsterdam, Matthew J Clarkson, and Danail Stoyanov. Gesture recognition in robotic surgery: a review. *IEEE Transactions on Biomedical Engineering*, 68(6), 2021.

[212] Carly R Garrow, Karl-Friedrich Kowalewski, Linhong Li, Martin Wagner, Mona W Schmidt, Sandy Engelhardt, Daniel A Hashimoto, Hannes G Kenngott, Sebastian Bodenstedt, Stefanie Speidel, et al. Machine learning for surgical phase recognition: a systematic review. *Annals of surgery*, 273(4):684–693, 2021.

[213] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*, 2023.

[214] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. Tidybot: Personalized robot assistance with large language models. *arXiv preprint arXiv:2305.05658*, 2023.

[215] Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. Building cooperative embodied agents modularly with large language models. *arXiv preprint arXiv:2307.02485*, 2023.

[216] Samir Yitzhak Gadre, Mitchell Wortsman, Gabriel Ilharco, Ludwig Schmidt, and Shuran Song. Cows on pasture: Baselines and benchmarks for language-driven zero-shot object navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23171–23181, 2023.

[217] Sai Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *Microsoft Auton. Syst. Robot. Res*, 2:20, 2023.

[218] Siyuan Huang, Zhengkai Jiang, Hao Dong, Yu Qiao, Peng Gao, and Hongsheng Li. Instruct2act: Mapping multi-modality instructions to robotic actions with large language model. *arXiv preprint arXiv:2305.11176*, 2023.

[219] Andrew Szot, Max Schwarzer, Harsh Agrawal, Bogdan Mazoure, Walter Talbott, Katherine Metcalf, Natalie Mackraz, Devon Hjelm, and Alexander Toshev. Large language models as generalizable policies for embodied tasks. *arXiv preprint arXiv:2310.17722*, 2023.

[220] Kevin Lin, Christopher Agia, Toki Migimatsu, Marco Pavone, and Jeannette Bohg. Text2motion: From natural language instructions to feasible plans. *arXiv preprint arXiv:2303.12153*, 2023.

[221] Tianbao Xie, Siheng Zhao, Chen Henry Wu, Yitao Liu, Qian Luo, Victor Zhong, Yanchao Yang, and Tao Yu. Text2reward: Automated dense reward function generation for reinforcement learning. *arXiv preprint arXiv:2309.11489*, 2023.

[222] Jingkang Yang, Yuhao Dong, Shuai Liu, Bo Li, Ziyue Wang, Chencheng Jiang, Haoran Tan, Jiamu Kang, Yuanhan Zhang, Kaiyang Zhou, et al. Octopus: Embodied vision-language programmer from environmental feedback. *arXiv preprint arXiv:2310.08588*, 2023.

[223] Abhishek Padalkar, Acorn Pooley, Ajinkya Jain, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anikait Singh, Anthony Brohan, et al. Open x-embodiment: Robotic learning datasets and rt-x models. *arXiv preprint arXiv:2310.08864*, 2023.

[224] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

[225] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.

[226] Wenhao Yu, Nimrod Gileadi, Chuyuan Fu, Sean Kirmani, Kuang-Huei Lee, Montse Gonzalez Arenas, Hao-Tien Lewis Chiang, Tom Erez, Leonard Hasenclever, Jan Humplik, et al. Language to rewards for robotic skill synthesis. *arXiv preprint arXiv:2306.08647*, 2023.

[227] Jinjie Mai, Jun Chen, Bing Li, Guocheng Qian, Mohamed Elhoseiny, and Bernard Ghanem. Llm as a robotic brain: Unifying egocentric memory and control. *arXiv preprint arXiv:2304.09349*, 2023.

[228] Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *arXiv preprint arXiv:2305.20050*, 2023.

[229] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*, 2023.

[230] Yuanzhen Xie, Tao Xie, Mingxiong Lin, WenTao Wei, Chenglin Li, Beibei Kong, Lei Chen, Chengxiang Zhuo, Bo Hu, and Zang Li. Olagpt: Empowering llms with human-like problem-solving abilities. *arXiv preprint arXiv:2305.16334*, 2023.

[231] Yifan Zhang, Jingqin Yang, Yang Yuan, and Andrew Chi-Chih Yao. Cumulative reasoning with large language models. *arXiv preprint arXiv:2308.04371*, 2023.

[232] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.

[233] Jiawei Zhang. Graph-toolformer: To empower llms with graph reasoning ability via prompt augmented by chatgpt. *arXiv preprint arXiv:2304.11116*, 2023.

[234] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15619–15629, 2023.

[235] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.