



Chameleon Dataset- Large Synthetic Radiology Report Dataset: A Pilot Study

Satvik Tripathi, Researcher, Department of Radiology , University of Pennsylvania

Dana Alkhulaifat, MD; Rithvik Sukumaran; Charles Chambers, MCIT; Darco Lalevic; Emiliano Garza, MD; Meghana Muppuri, MD; Dania Daye, MD, PhD; Tessa Cook, MD, PhD, CIIP, FSIIM

Background/Problem Being Solved

The availability of open-access radiology report datasets remains a critical limitation in academic radiology research. With the growing adoption of large language models (LLMs), compliance with protected health information (PHI) regulations poses significant challenges, restricting the scope of research leveraging models like GPT. This limitation necessitates the development of PHI-free, high-quality text datasets to advance natural language processing (NLP) and LLM research in radiology.

Intervention(s)

We introduce Chameleon, a large-scale synthetic radiology report dataset generated entirely using the GPT-4 API. The dataset was constructed without any PHI, providing a resource for training and evaluating NLP methods and LLM-based applications in radiology.

Barriers/Challenges

Key challenges encountered during dataset development included ensuring consistency in linguistic style, maintaining medical and anatomical accuracy, and adhering to structured reporting formats characteristic of clinical radiology.

Outcome

500 synthetic radiology reports were generated, encompassing diverse imaging modalities and pathologies, including head CT, thoracic CT (with and without contrast), and abdominal CT. Four expert medical reviewers systematically evaluated thoracic reports based on content accuracy, format integrity, and stylistic coherence. Four prompting strategies were tested, with the optimal approach selected based on a cumulative scoring metric derived from the reviewers' evaluations. The mean time for generating a single report was 0.23 seconds, highlighting the efficiency and scalability of the methodology.

Conclusion/Statement of Impact/Lessons Learned

The Chameleon dataset addresses a critical gap in radiology research by providing a scalable and PHI-compliant resource for LLM and NLP applications. This synthetic dataset ensures ethical data usage while enabling advancements in radiology-

specific AI research. Future efforts will aim to expand the dataset's scope, incorporating additional imaging modalities and refining generation techniques to enhance clinical applicability and translational potential.

Figure(s)

Organ/Pathologies	Generated Report
Head CT without contrast	<p>Patient Information:</p> <ul style="list-style-type: none"> - Age: 42 - Sex: Female - Relevant Medical History: Migraine with aura. <p>Study Information:</p> <ul style="list-style-type: none"> - Type of CT scan: Head CT without contrast - Date of Scan: 09/26/2023 - Comparison: No previous scans. <p>Clinical Information:</p> <ul style="list-style-type: none"> - Reason for examination: Severe headache following a head strike during sports. - Referring Physician: Dr. Anita Gomez <p>Technique: The head CT was conducted without contrast in axial, coronal, and sagittal views.</p> <p>Findings: There is a convexity subdural hematoma on the right, measuring 8 mm in maximal thickness. There is a slight mass effect with a 2 mm shift of the midline structures towards the left. No other acute intracranial abnormalities are detected.</p> <p>Impression: Right convexity subdural hematoma with mild mass effect. Given the patient's history of migraine, differentiation from migraine symptoms is important. Conservative management with follow-up imaging and neurosurgical consultation is advised.</p>
Lung Nodules	<p>Patient Information:</p> <ul style="list-style-type: none"> - Age: 57 - Sex: Female - Relevant Medical History: Asthma, previous tuberculosis <p>Study Information:</p> <ul style="list-style-type: none"> - Type of CT scan: CT of the chest with contrast - Date of Scan: September 14, 2023 - Comparison: Previous CT scan dated March 10, 2023 <p>Clinical Information:</p> <ul style="list-style-type: none"> - Reason for examination: Persistent cough and evaluation of previously noted lung nodules - Referring Physician: Dr. Jonathan Hughes <p>Technique: A contrast-enhanced CT of the chest was performed. Images were obtained at multiple phases post-intravenous contrast administration and were reviewed in axial, sagittal, and coronal reconstructions.</p> <p>Findings:</p> <ul style="list-style-type: none"> - Lungs and Pleura: There are three nodules in the right lung measuring up to 10 mm with partially calcified edges, likely indicative of prior granulomatous infection. No new nodules. No pleural effusion or thickening. - Cardiovascular, Mediastinum, Thyroid: Normal cardiac silhouette and mediastinal contours. No pathologic lymphadenopathy. - Upper Abdomen: No abnormalities detected in the visualized portions of the upper abdomen. <p>Impression: Stable partially calcified nodules in the right lung consistent with sequelae of prior granulomatous disease. No signs of active disease. Annual surveillance recommended unless new symptoms arise.</p>
Thoracic CT w Contrast	<p>Patient Information:</p> <ul style="list-style-type: none"> - Age: 55 - Sex: Female

	<ul style="list-style-type: none"> - Relevant Medical History: Rheumatoid arthritis <p>Study Information:</p> <ul style="list-style-type: none"> - Type of CT scan: Contrast Enhanced Chest CT - Date of Scan: 09/03/2023 - Comparison: CT scan dated 04/10/2022 <p>Clinical Information:</p> <ul style="list-style-type: none"> - Reason for examination: Evaluate for vasculitis - Referring Physician: Dr. Neil Gordon <p>Technique: Contrast-enhanced CT of the chest was performed with a focus on assessing the thoracic vessels for signs of vasculitis.</p> <p>Findings: No signs of active vasculitis are identified. The thoracic aorta and its major branches appear normal with no aneurysms or significant stenosis. There is a small calcified nodule in the left lower lobe, likely benign. No lymphadenopathy or pleural effusions are noted.</p> <p>Impression: No evidence of thoracic vasculitis. The small calcified nodule in the left lower lobe should be followed with a follow-up CT in 12 months to ensure stability.</p>
--	---

Table 1. Example reports generated for various body parts and pathologies using few shot learning prompting method.

Keywords

Artificial Intelligence/Machine Learning; Emerging Technologies; Imaging Research