Appendix A Critical Findings Extraction Prompt

Task-based Prompting Template

REPORT: Textual content from the radiology report. This consists of both the impression and findings included in the report.
DEFINITIONS: This includes the definitions of Critical and Incidental findings that need to be extracted from the report.
EXAMPLES: If few-shot prompting option is selected, task-specific example(s) are included in the prompt.
TASK INSTRUCTION: This is the prompt or the task-based instruction to be followed by the LLM.

Zero-shot Prompting

"Findings are consistent with small-bowel obstruction likely secondary to a closed loop obstruction. Additionally, presence of abdominal ascites, mesenteric engorgement, and differential enhancement of the wall raise concern for intestinal ischemia. These findings were discussed with Dr. [**Redacted Name**] on [**2107-11-21**].
The definitions of CRITICAL and INCIDENTAL findings in a radiology report: CRITICAL findings are life threatening imaging findings that needs to be communicated immediately while INCIDENTAL findings are non-life-threatening findings, but significant enough that they need to be communicated within a short period of time.
Based on these above definitions, find the CRITICAL and INCIDENTAL findings mentioned in the report. Provide an explanation for each finding."

Few-shot Prompting

"Findings are consistent with small-bowel obstruction likely secondary to a closed loop obstruction. Additionally, presence of abdominal ascites, mesenteric engorgement, and differential enhancement of the wall raise concern for intestinal ischemia. These findings were discussed with Dr. [**Redacted Name**] on [**2107-11-21**].
The definitions of CRITICAL and INCIDENTAL findings in a radiology report: CRITICAL findings are life threatening imaging findings that needs to be communicated immediately while INCIDENTAL findings are non-life-threatening findings, but significant enough that they need to be communicated within a short period of time.
Here is an example of the above task: { 'report': 'Interval increase in the left lower lobe atelectasis and effusion. 2) Right lower lobe atelectasis. 3) Pneumoperitoneum consistent with the patient's history of recent surgery', 'explanation': 'Atelectasis and pleural effusions are not critical findings because they are not life threatening medical emergencies and do not need to be acted on immediately. Pneumoperitoneum is a critical finding because it is a medical emergency and would need to be acted on immediately. It is characterized as known/expected because it is described in the report to be expected due to a recent surgery.', 'mentions': ['pneumoperitoneum']. }
Based on these above definitions and provided few-shot examples, find the CRITICAL and INCIDENTAL findings mentioned in the report. Provide an explanation for each finding."

Appendix B Few-Shot Prompting Examples

Below are the reports, explanations, and critical findings used as task-specific examples for few-shot prompting along with the task instruction.

Example I

'category': 'Gastrointestinal',
'report': 'Findings are consistent with small-bowel obstruction likely secondary to a closed loop obstruction. Additionally, presence of abdominal ascites, mesenteric engorgement, and differential enhancement of the wall raise concern for intestinal ischemia. These findings were discussed with Dr. [**Redacted Name**] on [**2107-11-21**].',
'explanation': 'Small bowel obstruction due to a closed loop obstruction is a critical finding because it is life-threatening if not treated immediately. It is characterized as new, because it is the noted to be present for the first time. Intestinal ischemia is a critical finding, because it is a medical emergency, needing to be acted on immediately. It is characterized as an uncertain critical finding because the possibility is raised.',
'mentions': ['small bowel obstruction','closed loop obstruction', 'intestinal ischemia'].

Example II

'category': 'Chest',
'report': 'Interval increase in the left lower lobe atelectasis and effusion. 2) Right lower lobe atelectasis. 3) Pneumoperitoneum consistent with the patient's history of recent surgery',
'explanation': 'Atelectasis and pleural effusions are not critical findings because they are not life threatening medical emergencies and do not need to be acted on immediately. Pneumoperitoneum is a critical finding because it is a medical emergency and would need to be acted on immediately. It is characterized as known/expected because it is described in the report to be expected due to a recent surgery.',
'mentions': ['pneumoperitoneum'].

Example III

'category': 'Neurologic',
'report': 'As compared to the prior examination, there is extensive subarachnoid hemorrhage and possibly left-sided uncal herniation. Pneumocephalus is noted. Multiple calvarial and facial fractures. Ventricular shunt in place. Further follow up is recommended within 24 to 48 hours.',
'explanation': 'There are multiple new critical findings in this report, including subarachnoid hemorrhage and traumatic pneumocephalus, because they need to be acted upon immediately. They are characterized as new, since they are implied to be present for the first time. Uncal herniation is a critical finding, because it needs to be acted on immediately. It is characterized as uncertain, because the possibility or suspicion is raised. Calvarial and facial fractures are not critical findings, because they do not require immediate medical intervention.',
'mentions': ['subarachnoid hemorrhage', 'pneumocephalus', 'uncal herniation'].

Example IV

'category': 'Gastrointestinal, Chest',
'report': '1. Findings consistent with small bowel obstruction with a transition point at the anastomotic site in the pelvis. New marked edema within a loop of bowel in the left lower quadrant is potentially concerning for ischemia. 2. Stable pneumoperitoneum, both free and loculated in a lower abdominal/pelvic collection with rim enhancement. 3. Right-sided hydronephrosis and hydroureter with nephrostomy catheter in place. 4. Anasarca. 5. Bilateral pleural effusions and patchy opacities likely reflecting infectious or inflammatory process. 6. Probable severe fatty infiltration of the liver. Results were discussed with [**First Name8 (NamePattern2) 52259**] [**Last Name (NamePattern1) 13544**] at 5:30 p.m. on [**2180-2-18**].',
'explanation': 'Small bowel obstruction is a critical finding, because if not acted upon immediately, it can be life threatening. It is characterized as new because it is implied to be reported for the first time. Pneumoperitoneum is a critical finding because it needs to be acted upon immediately. It is characterized as known/expected because it is described as being unchanged or stable, and therefore implied to be known. Intestinal ischemia is a critical finding because it needs to be acted on immediately. It is characterized as an uncertain because the concern is raised.',
'mentions': ['small bowel obstruction', 'pneumoperitoneum', 'ischemia'].

Example V

'category': 'Non-critical report',
'report': '1. No evidence of pulmonary embolism or aortic dissection. 2. Small bilateral pleural effusions with dependent atelectasis. Patchy areas of ground glass opacity consistent with pulmonary edema. Clinical correlation is requested.',
'explanation': 'There are no critical findings in this report, because pulmonary edema and pleural effusions are not considered medical emergencies.',
'mentions': [None].

Appendix C LLM-based Evaluation Prompt

For human-annotated small-scale test set

'human label': 'subarachnoid hemorrhage', 'pneumocephalus'
'model output': 'subarachnoid hemorrhage',
'metric': 'correctness',
'metric definition': 'Given the human label, determine if the model output is correct. Consider partial matches.'
'score': 'Assign a continuous score of correctness between 0 to 1, 1 being the highest value.'

For large-scale test set

'report': ' Multifocal bilateral pneumonia with right lung cavitary lesions , right calcified granulomas , and right pleural plaques are very concerning for reactivation tuberculosis with a component of right upper lobe necrotizing pneumonia.',
'definition': 'The definition of CRITICAL findings in a radiology report: CRITICAL findings are life threatening imaging findings that needs to be communicated immediately.',
'model output': 'tuberculosis', 'pneumonia',
'metric': 'correctness',
'metric definition': 'The model output identifies the critical findings mentioned in the report. Given the report and the definition of critical findings, determine if the model output is correct. Consider partial matches.'
'score': 'Assign a continuous score of correctness between 0 to 1, 1 being the highest value.'

Appendix D Results on Test Datasets - Small-scale and Large-scale

Table D1: Results on human-annotated small-scale test datasets - (a) Internal hold-out (Mayo Clinic) and (b) External (MIMIC-III). R-BScore: RadBERTScore; PT-Pretrained; WFT: Weakly Fine-tuned.

Dataset	Models	Prompting Setup	Human-aided Metrics					LLM-aided Metrics	
			ROUGE-1	ROUGE-2	BLEU	METEOR	R-BScore	G-eval	Prometheus
Internal hold-out test	Human	-	-	-	-	-	-	0.79 ±0.022	0.81 ±0.032
	Mistral-PT	Zero-Shot	0.17 ±0.022	0.29 ±0.013	0.27 ±0.021	0.31 ±0.014	0.11 ±0.010	0.06 ±0.033	0.15 ±0.025
		Few-Shot	0.23 ±0.025	0.32 ±0.005	0.31 ±0.017	0.33 ±0.039	0.23 ±0.033	0.12 ±0.033	0.23 ±0.011
	Mistral-WFT	Zero-Shot	0.25 ±0.029	0.31 ±0.019	0.29 ±0.037	0.35 ±0.033	0.24 ±0.032	0.11 ±0.040	0.17 ±0.058
		Few-Shot	**0.43** ±0.030	**0.48** ±0.006	**0.42** ±0.026	**0.39** ±0.034	**0.33** ±0.030	**0.12** ±0.034	**0.10** ±0.015
	BioMistral-PT	Zero-Shot	0.09 ±0.029	0.22 ±0.035	0.24 ±0.031	0.27 ±0.029	0.21 ±0.036	0.07 ±0.043	0.09 ±0.034
		Few-Shot	0.15 ±0.026	0.25 ±0.017	0.26 ±0.029	0.29 ±0.025	0.18 ±0.007	0.04 ±0.037	0.14 ±0.013
	BioMistral-WFT	Zero-Shot	0.19 ±0.028	0.27 ±0.030	0.25 ±0.007	0.28 ±0.034	0.17 ±0.015	0.12 ±0.031	0.09 ±0.040
		Few-Shot	**0.38** ±0.031	**0.41** ±0.036	**0.29** ±0.014	**0.41** ±0.037	**0.32** ±0.028	**0.13** ±0.030	**0.17** ±0.010
External test	Human	-	-	-	-	-	-	0.87 ±0.33	0.93 ±0.11
	Mistral-PT	Zero-Shot	0.29 ±0.15	0.40 ±0.19	0.39 ±0.17	0.40 ±0.15	0.41 ±0.21	0.15 ±0.37	0.17 ±0.32
		Few-Shot	0.41 ±0.12	0.43 ±0.09	0.49 ±0.11	0.55 ±0.11	0.50 ±0.15	0.19 ±0.27	0.21 ±0.33
	Mistral-WFT	Zero-Shot	0.36 ±0.18	0.42 ±0.23	0.39 ±0.19	0.47 ±0.29	0.38 ±0.37	0.21 ±0.19	0.18 ±0.21
		Few-Shot	**0.51** ±0.16	**0.59** ±0.26	**0.54** ±0.25	**0.61** ±0.18	**0.55** ±0.17	**0.35** ±0.24	**0.27** ±0.28
	BioMistral-PT	Zero-Shot	0.28 ±0.33	0.34 ±0.21	0.33 ±0.25	0.42 ±0.13	0.33 ±0.24	0.19 ±0.31	0.13 ±0.11
		Few-Shot	0.31 ±0.24	0.48 ±0.18	0.47 ±0.23	0.53 ±0.11	0.49 ±0.18	0.27 ±0.38	0.22 ±0.27
	BioMistral-WFT	Zero-Shot	0.31 ±0.27	0.38 ±0.18	0.37 ±0.24	0.40 ±0.17	0.21 ±0.16	0.23 ±0.31	0.19 ±0.23
		Few-Shot	**0.47** ±0.24	**0.51** ±0.38	**0.50** ±0.13	**0.52** ±0.24	**0.39** ±0.26	**0.31** ±0.43	**0.35**±0.21

Table D2: Results with the LLM-aided metrics on the large-scale unannotated dataset from MIMIC-IV.

Models	Prompting Setup	LLM-aided Metrics	
		G-eval	Prometheus
Mistral-PT	Zero-Shot	0.26 ±0.032	0.25 ±0.019
	Few-Shot	0.37 ±0.043	0.33 ±0.071
Mistral-WFT	Zero-Shot	0.24 ±0.029	0.38 ±0.049
	Few-Shot	**0.39** ±0.032	**0.42** ±0.061
BioMistral-PT	Zero-Shot	0.28 ±0.059	0.23 ±0.038
	Few-Shot	0.31 ±0.079	0.27 ±0.053
BioMistral-WFT	Zero-Shot	0.21 ±0.041	0.31 ±0.077
	Few-Shot	0.33 ±0.073	0.35 ±0.029

Appendix E Critical Findings Expanded Term List

Table E3: Critical Findings Terms.

Category	Critical Terms	Expansions
General	Non-accidental trauma	
	Allergic reaction	
	Malpositioned line or tube	Malposition of central line or ET tube or NG tube, Malpositioned line or tube
	Foreign body	FB, Foreign material, Exogenous material
Neurologic/head and neck	Intracranial or spinal haemorrhage	Traumatic spinal cord haemorrhage or hemorrhage, Haemorrhage or hemorrhage (parenchymal, extradural, subarachnoid, subdural, epidural), Intracranial haemorrhage or hemorrhage (subdural, subarachnoid, epidural), Spinal hemorrhage
	Suspected stroke	Nonhemorrhagic stroke, Thrombolytic stroke
	Intracranial mass or lesion	Intracranial space-occupying lesion, Intracranial lump, Mass of intracranial structure, Midline shift mass, Intracranial mass with midline shift
	Brain herniation	Brain stem herniation, Herniation of brainstem
	Symptomatic hydrocephalus	Malfunctioning shunt
	Depressed skull fracture	Fracture of skull and facial bones, Basilar skull fracture, Depressed fracture of skull, Depressed skull fracture
	Post-traumatic pneumocephalus	
	Arterial dissection	Dissection of artery
	Brain death	Brainstem death, Death of brain
	Severe spinal cord compression of any cause	Compression of spinal cord, Cord compression, Spinal cord compression
	Unstable spine fracture	Spine injury, Fracture dislocation of spinal vertebra, Fracture dislocation of spine
	Traumatic spine injury	Traumatic injury of spinal cord or spine or vertebral column, Traumatic injury of spinal cord and vertebral column, Traumatic spinal cord haemorrhage, Traumatic spinal cord hemorrhage
	Cord hemorrhage or infarct	Infarction of spinal cord, Spinal cord infarct
	Airway obstruction or impending obstruction	Total airway obstruction, Totally obstructed airway, Facial fracture, Epiglottis obstruction, Tonsillitis, Retropharyngeal abscess
Gastrointestinal (GI)	Pneumoperitoneum	PP
	Intestinal obstruction	Obstruction of intestine, Bowel obstruction
	Intestinal ischemia and/or portal/mesenteric venous gas	Acute intestinal ischaemia/ischemia due to obstruction of vasculature of intestine, Acute intestinal ischaemia/ischemia with obstruction of intestinal vasculature, Intestinal ischaemia/ischemia, Ischaemia/ischemia bowel, Mesenteric organ thrombus, Transplat organ thrombus, Mesenteric/portal venous gas, Ischaemic/ischemic bowel disease, Ischaemic/ischemic disease of gut, Vascular insufficiency of intestine
	Pseudoaneurysm	Post-trauma GI bleed
	Intra-abdominal organ injury and/or bowel injury	Injury of liver/spleen/pancreas, Liver/spleen/pancreas injury, Post-trauma bowel injury, Injury of intra-abdominal organ, Hepatic injury/trauma, Injury of abdominal viscus, Internal injury of abdomen or abdominal organ

Table E4: Critical Findings Terms (Continued).

Category	Critical Terms	Expansions
Genitourinary/Obsteric (GU/OB)	High-grade GU injury	
	Testicular torsion	Torsion of testicle/testis
	Ovarian torsion	Torsion of ovary
	Ectopic pregnancy	EP
	Placental abruption	Ablatio placentae, Abruptio placentae, Placental abruption, Premature detachment of normally implanted placenta, Premature detachment of placenta
	Uterine rupture	UTI, Rupture of uterus
	Kidney injury and/or ureteral or bladder injury post-trauma	Ureteric injury, Injury of kidney/ureter/bladder, Injury to kidney/ureter/bladder, Renal injury
	Absent perfusion post-op kidney	Absent perfusion kidney, Absent perfusion of post-op kidney
Musculoskeletal (MSK)	Nonspinal fracture and/or dislocation	Non-spinal dislocation, Non-spinal fracture
	Necrotizing fasciitis	Necrotizing infection, Necrotising soft tissue infection, Necrotizing soft tissue infection
Chest	Tension pneumothorax	
	Pulmonary embolus	PE, Extensive emboli, History of pulmonary embolus, Pulmonary embolism
	Lung lesion with possibility of active TB	Lesion of lung, Active tubercolosis, Tubercolosis
	Superior vena cava occlusion (including SVC syndrome)	SVC occlusion/syndrome, Superior vena cava syndrome, Occlusion of superior vena cava, Occlusion of vein
	Pericardial effusion and/or suspected tamponade	Pericardial effusion, Tamponade, Tamponade - action
	Active post-traumatic hemorrhage	Post-trauma hemorrhage
	Tracheal obstruction or impending obstruction	Obstruction of trachea, Obstructive tracheal disease
Cardiac/Vascular	Ruptured/leaking arterial aneurysm	AAA, Abdominal aortic aneurysm, Arterial aneurysm, Ruptured thoracic aortic aneurysm, Thoracic aortic aneurysm, Thoracic aortic aneurysm which has ruptured, Torn aortic cusp, Rupture of aortic valve cusp
	Limb-threatening arterial or venous occlusion or stenosis	Arterial occlusion, Arterial occlusive disease, Occluded arterial bypass graft, Occlusion of artery, Blocked artery, Obstruction of artery, Venous occlusion, Occlusion of vein, Complete stenosis, Stenosis
	Arterial dissection or intramural hematoma	Intramural haematoma, Major vessel dissection
	Acute myocardial infarction	AMI