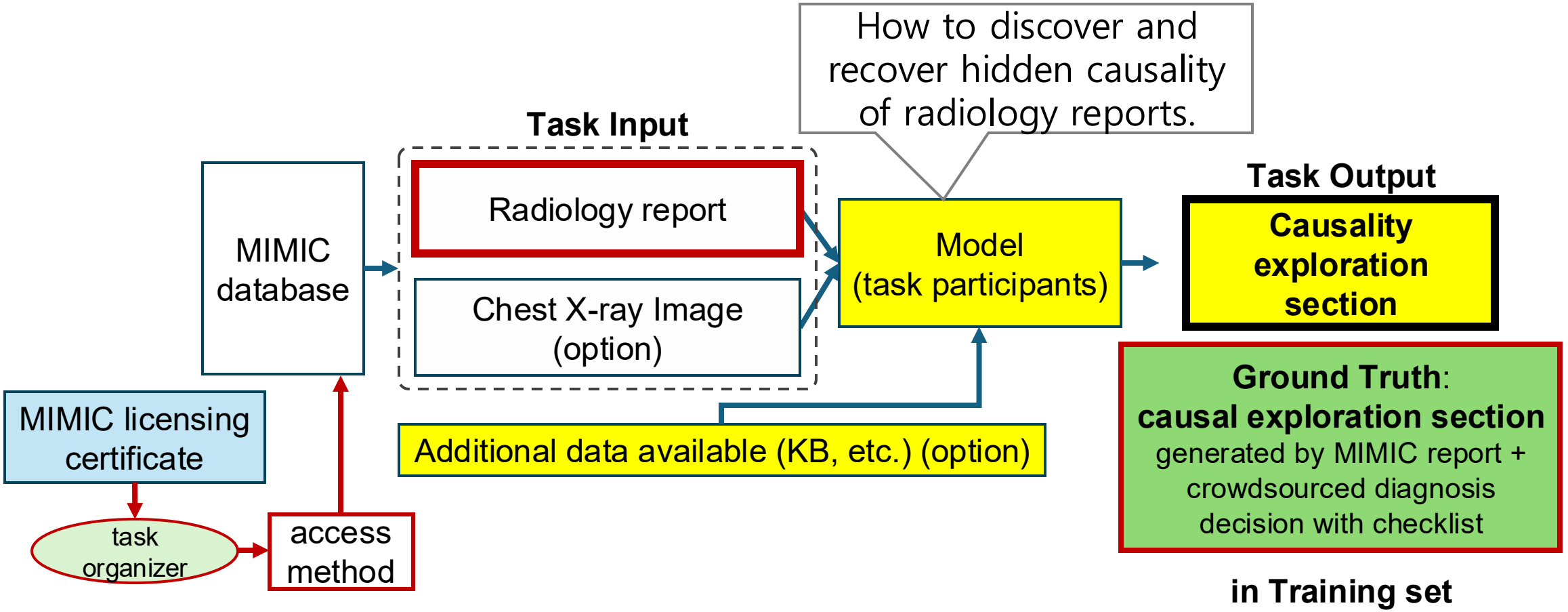
Hidden-Rad Task-1 Definition

Generating causality section by discovering hidden (missing) causality in radiology reports.



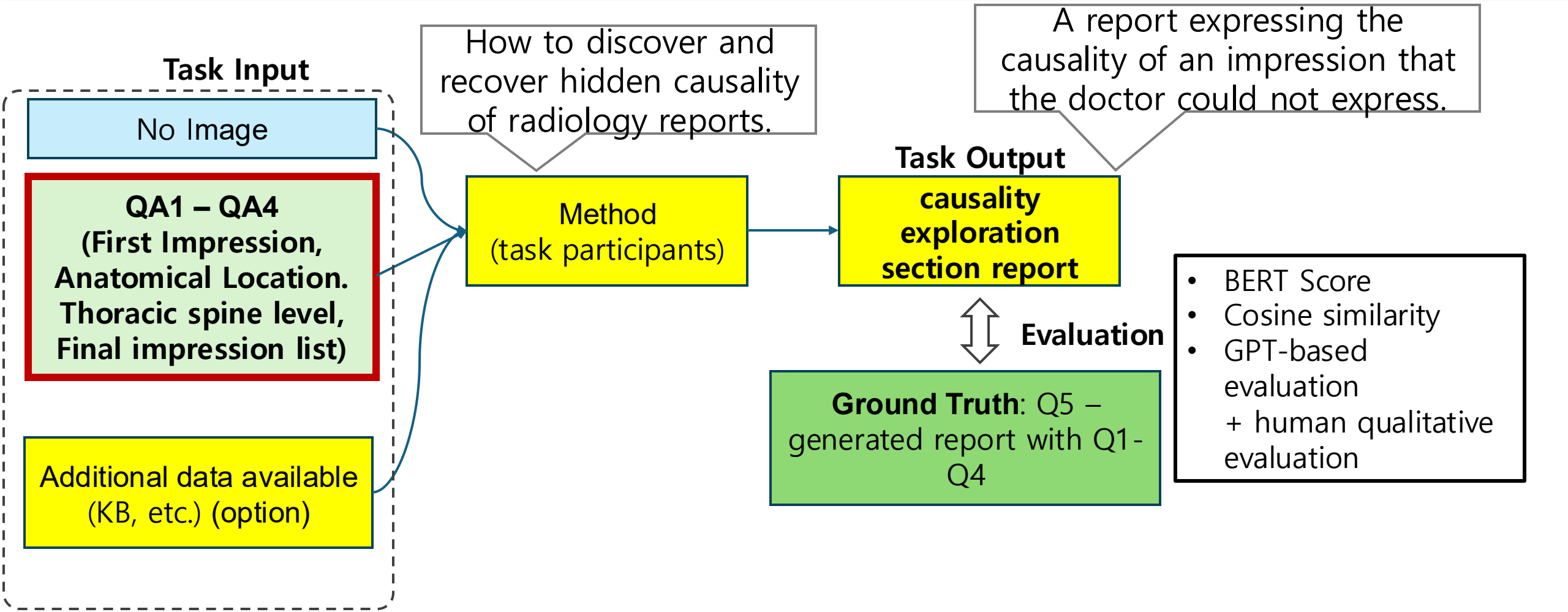
Task1 Example: Generating causality exploration section from a radiology report

Input: Radiology report (MIMIC database) + optional data		Output: text for causality
Report	FINAL REPORT EXAMINATION: CHEST (AP upright AND LAT) INDICATION: ___M with AMS and R hip pain s/p recent fall COMPARISON: CT chest ___ FINDINGS: The imaging findings indicate signs of compromised lung integrity, with evidence of pneumothorax and pleural effusion, likely stemming from trauma or an underlying lung pathology. Additionally, the presence of subcutaneous emphysema suggests air leakage into the subcutaneous tissue, possibly due to rib fractures or a breach in the lung tissue. The stable cardiomediastinal contours, along with the placement of a right chest tube, support the current management approach for these conditions. IMPRESSION: Large mass in the right upper lung. Otherwise, unremarkable.	**Causal Exploration:** The presence of pneumothorax and pleural effusion indicates compromised lung integrity, likely resulting from trauma or underlying lung pathology. The subcutaneous emphysema suggests air leakage into the subcutaneous tissue, possibly due to rib fractures or a breach in lung tissue. The stable cardiomediastinal contours and the right chest tube in place support ongoing management of these conditions.
Image (Option)		
Additional resources from task participant (Option)		

Task-2 Definition (w/o MIMIC-licensing)

input: Crowdsourced data about Q1-Q4 (first/final impression, anatomic loc.)

output: text generated from Q5 (checklist answer for each case of Q4 decision)



Task2: from Radiologist questionnaire for Chest X-ray image to causal exploration text generation

Input		Output
A1 (First Impression)	Emphysema pleural effusion Pneumothorax	**Causal Exploration:** 1. Emphysema: The presence of subcutaneous emphysema as indicated in A5 can be caused by a rupture in the airways or lung tissue, which allows air to escape into the subcutaneous tissues. This can be linked back to A1~A4 as the condition was identified through these processes. 2. Pleural Effusion: The absence of certain findings in A5 such as lack of blunted costophrenic angles or no abnormalities in the hilar vessels could indicate that there is no pleural effusion. However, further investigations are needed as these signs are not definitive. 3. Pneumothorax: The fact that both lungs are well expanded and similar in volume as stated in A5, alongside other normal findings, suggests there is no pneumothorax present. This is in accordance with the findings from A1~A4. Thus, the Causal Explorations for these conditions are derived from the findings in A1~A4 and are further supported by the details in A5.
A2 (Anatomical Location)	'Subcutaneous tissue', 'Parenchyme', 'RUL(Right Upper Lobe)', 'Parenchyme', 'RML(Right Middle Lobe)', 'Parenchyme', 'RLL(Right Lower Lobe)', 'Parenchyme', 'LUL(Left Upper Lobe)', 'Parenchyme', 'LLL(Left Lower Lobe)', 'Pleural', 'Right Pleural'	
A3 (Thoracic Spin Level)	{'begin': 1, 'end': 12}	
A4 (Final Impression)	'emphysema' 'pleural effusion' 'pneumothorax'	

Evaluation Process Overview

1. Quantitative Evaluation (80 points)

- BERTScore (5%) – Ensures **causal explanations** align with the **original report.**

- COS Similarity (5%) – Measures **semantic coherence** in generated explanations.

- BioSentVec (20%) – Validates **medical accuracy** using **MIMIC embeddings.**

- GPT-based Score (White) (25%) – Rewards **structured** and **logical explanations.**

- GPT-based Score (Black) (25%) – Penalizes **inconsistencies** to enhance **reliability**.

2. Qualitative Evaluation (20 points)

- Top 5 models from each metric

- Human evaluation based on predefined criteria

Participating teams

Task 1

- Teddysum
- RADPHI3
- Nash

Task 2

- Teddysum
- Nash

Join the session of Hidden-Rad today!

- Poster session (Lunch time, 12:45-14:15, JST)
- Round table break-out session (16:15-17:15)
  - 16:15-16:30: [Invited talk] Future aspect on Hidden-Rad: Developing Clinical Report Processing Benchmarks for Enhanced Explainability and Verifiability (in-Dong Kim (DBCLS) and Konwoo Kim (NII))
  - 16:30-16:45: Evaluation scheme of NTCIR18 Hidden-Rad task (TO of Hidden-Rad)
  - 16:45-16:50: Teddysum team's issues and key insights
  - 16:50-16:55: RADPHI3 team's issues and key insights
  - 16:55-17:00: Nash team's issues and key insights
  - 17:00-17:15: [Discussion] Hidden-Rad2, Workshop PRICAI and Others