# Towards Accurate and Clinically Meaningful Summarization of Electronic Health Record Notes: A Guided Approach


University of Pittsburgh
School of Computing and Information

**Zhimeng Luo**[1](zhl123@pitt.edu), **Yuelyu Ji**[1], **Abhibha Gupta**[1], **Zhuochun Li**[1], **Adam Frisch**[2], **Daqing He**[1]

[1] School of Computing and Information, University of Pittsburgh
[2] Department of Emergency Medicine, University of Pittsburgh


## MOTIVATION

**Electronic Health Records** (EHR) contains detailed narrative information about a patient's current and past medical history.

**Text summarization** is a task in natural language processing (NLP) aimed at shortening a long text into a more concise summary while preserving its core information and meaning.

### Importance of summarizing EHR

- Facilitates quicker and more accurate access to critical patient information
- Aligns with clinical best practices for readability and comprehensiveness
- Improve communication among healthcare providers

### Challenges

- **Data Volume and Complexity** of EHR makes it difficult to sift through and identify the most relevant information, adding layers of complexity to the summarization task.
- **Alignment with Clinical Practice**: Many summarization approaches are not tailored to align with prevalent clinical practices

### Proposed Template

Patient is a Age year old Gender who presents with chief complaint. Insert important OPQRST information. Insert important diagnostics (labs, imaging). Insert treatments (medications, procedures). Patient was Disposition.

1) Basic Demographics;
2) The Chief complaints are the reason for the patient's visit;
3) "OPQRST" information refers to the onset, provocation, quality, region/radiation, severity, and timing of the patient's symptoms;
4) Diagnostics include diagnostic studies that were performed;
5) Treatments encompass any medications prescribed or procedures performed;
6) Disposition indicates the patient's status at the end of the visit.

## METHODS

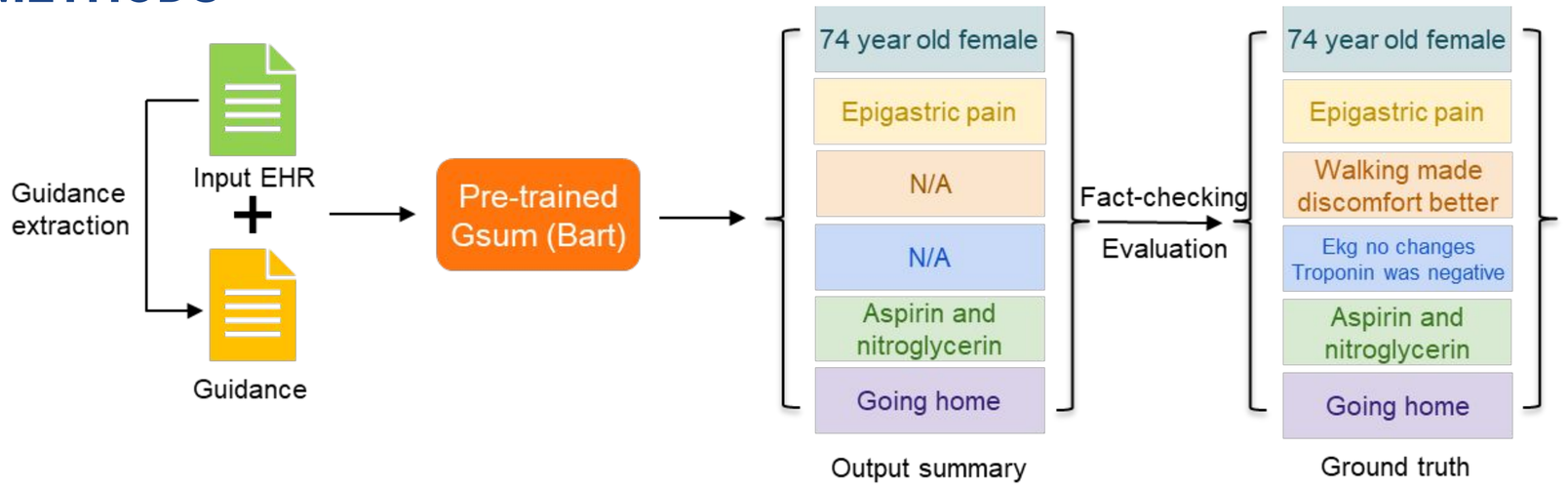


**Figure 1.** Proposed framework of summarization pipeline and evaluation metrics

- Developed a **standardized template** serves as a guideline for the summarization process
- **Augmented guidance** by automatically extracted critical information corresponding to aspects from the template
- Designed a **fact-checking metric** based on the aspects outlined in the template by checking for the presence and accuracy of clinically meaningful phrases

### Dataset

A large de-identified EHR dataset from UPMC. This dataset contains 93,258 EHRs (55,954 for training, 18,652 for validation, and 18,652 for testing) from emergency department, includes sections of **chief complaints**, **histories of present illness (HPI)**, **medical decision-making (MDM)**, and **diagnosis**.

### Guidance extraction

**Medical Entity Extraction**: NER with biomedical-ner-all model to extract medical entities related to the aspects outlined in proposed template

**OPQRST Extraction**: Bio_ClinicalBERT sentence classification model was fine-tuned for HPI to classify whether a sentence mentions OPQRST information

**Other Extraction**: another Bio_ClinicalBERT was fine-tuned for the MDM to classify whether a sentence belongs to treatment, diagnostic results, summarization, or is unrelated

## RESULTS

| Models | Guidance Settings | ROUGE | Fact-checking metrics Precision | Recall | F1 |
|---|---|---|---|---|---|
| BART | - | 63.8 | 51.7 | 27.2 | 33.5 |
| BioBART | - | 42.6 | 38.3 | 24.0 | 27.3 |
| Gsum (CNN) | CC | 59.7 | 51.7 | 23.4 | 30.0 |
| Gsum (CNN) | CC + NER + HPI guide + MDM guide | **75.7** | **58.6** | **42.1** | **46.4** |
| Gsum (CNN) | CC + NER subset + HPI guide + MDM guide | <u>73.8</u> | <u>56.9</u> | <u>40.3</u> | <u>44.8</u> |
| GSum (HPI) | CC | 42.1 | 31.3 | 26.1 | 26.3 |
| GSum (HPI) | CC + NER + HPI guide + MDM guide | 43.6 | 32.5 | 29.2 | 28.8 |
| GSum (HPI) | CC + NER subset + HPI guide + MDM guide | 43.7 | 32.5 | 29.3 | 28.8 |
| GSum (HPI + MDM) | CC | 49.9 | 44.3 | 27.7 | 31.8 |
| GSum (HPI + MDM) | CC + NER + HPI guide + MDM guide | 53.1 | 46.1 | 31.8 | 35.3 |
| GSum (HPI + MDM) | CC + NER subset + HPI guide + MDM guide | 53.2 | 46.1 | 30.8 | 34.6 |

Table 1. Summarization performance of models under different experiment settings. The **best**/<u>2nd-best</u> scores are in bold/underlined.

### Summarization Example

#### Source EHR

**History of Present Illness**: patient is a 74-year-old female presenting for evaluation of epigastric pain. … she states she has a history of gerd and had similar symptoms in the past when her gallbladder acted up. … she was given aspirin and nitroglycerin by ems, ... she states that actually walking around today made the epigastric discomfort better.

**Medical Decision Making**: patient presents for evaluation of epigastric pain. her ekg here shows no ischemic changes. troponin was negative. … she feels comfortable going home to follow-up with her primary care doctor.

#### Output Summary

A 74-year-old female presents for evaluation of epigastric pain. She has a history of gerd and had similar symptoms in the past when her gallbladder acted up. She was given aspirin and nitroglycerin by ems. She feels comfortable going home to follow-up with her primary care doctor.