Development and Seamless Integration of an On-premise AI Agent for Clinical Drafting: Insights from the Y-KNOT Project

Hanjae Kim, B.S.[1] ; So-Yeon Lee, M.D., Ph.D. [2,3] ; Seng Chan You, M.D., Ph.D. [1,2,3]

[1] Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea
[2] PHI Digital Healthcare, Seoul, South Korea
[3] Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, South Korea

Brief Summary

Large Language Models (LLMs) have shown promise in reducing clinical documentation burden, yet their real-world implementation faces significant challenges, particularly in non-English speaking countries. Here we present Your-Knowledgeable Navigator of Treatment (Y-KNOT) project, the first successful implementation of an on-premise bilingual LLM-based artificial intelligence agent integrated with electronic health records for automated clinical drafting. We successfully deployed the service in a tertiary hospital in South Korea, which seamlessly aligns with existing clinical workflows. This demonstrates a practical framework for implementing LLM-based systems for other healthcare institutions, paving the way for broader adoption of LLM solutions.

Introduction

Background

- Clinical documentation significantly burdens healthcare providers,[1,2] and there is growing optimism about Large Language Models' (LLMs) potential to alleviate this burden.[3]
- However, implementing existing LLM solutions in South Koreaa presents unique challenges, such as data sovereignty issues and the complexity of bilingual documents.
- Moreover, due to their separate interfaces, manually retrieving information from Electronic Health Records (EHRs) and typing it into LLMs may ironically be time-consuming.

Objectives

- To introduce Your-Knowledgeable Navigator of Treatment (Y-KNOT) project, aimed at developing a LLM-based artificial intelligence (AI) agent that seamlessly integrates a bilingual small LLM with EHR systems for automatic clinical drafting.

Methods

From June through November 2024 at Severance Hospital (Seoul, South Korea), we carried out three parallel processes.

Development of Medical Foundation LLM

Base architecture: Llama 3[4] (8B parameters)

First pre-training:

- general corpus (1.5 TB, Korean & English)

Second pre-training: (instruction pre-training[5])

- medical corpus (90 GB, Korean & English)
- general corpus (9 GB, Korean & English)

Clinical Co-development

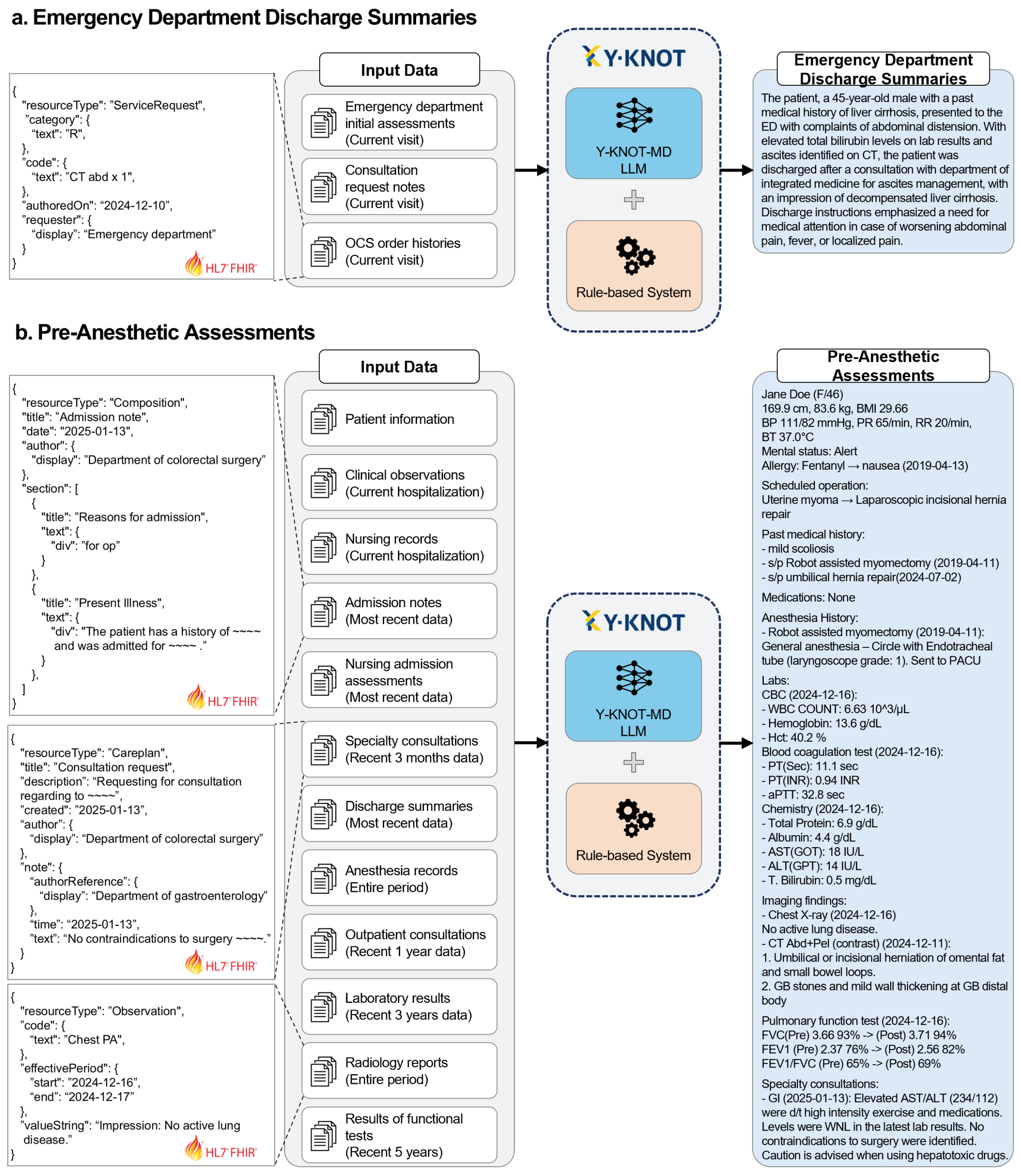
- In collaboration with physicians, data scientists, software engineers, and medical record specialists, we conducted iterative cycles of aligning clinical requirements, hospital data availability, documentation standards, and technical feasibility.
- Through this process, the LLM was instruction-tuned for specific documentation tasks: emergency department (ED) discharge summaries and pre-anesthetic assessments.

EHR Integration

We integrated the LLM into EHR by proceeding through three key components:

- Document standardization based on Fast Healthcare Interoperability Resource[6] (FHIR).
- Defining trigger points for the activation.
- Optimizing the user interaction framework for seamless clinical workflow incorporation.

Results



a. Emergency Department Discharge Summaries

```
{
  "resourceType": "ServiceRequest",
   "category": {
     "text": "R",
  },
  "code": {
     "text": "CT abd x 1",
  },
  "authoredOn": "2024-12-10",
  "requester": {
     "display": "Emergency department"
  }
}
```

HL7 FHIR
Input Data
Emergency department initial assessments (Current visit)
Consultation request notes (Current visit)
OCS order histories (Current visit)
Y-KNOT
Y-KNOT-MD LLM
Rule-based System
Emergency Department Discharge Summaries

The patient, a 45-year-old male with a past medical history of liver cirrhosis, presented to the ED with complaints of abdominal distension. With elevated total bilirubin levels on lab results and ascites identified on CT, the patient was discharged after a consultation with department of integrated medicine for ascites management, with an impression of decompensated liver cirrhosis. Discharge instructions emphasized a need for medical attention in case of worsening abdominal pain, fever, or localized pain.

b. Pre-Anesthetic Assessments

```
{
  "resourceType": "Composition",
  "title": "Admission note",
  "date": "2025-01-13",
  "author": {
    "display": "Department of colorectal surgery"
  },
  "section": [
    {
      "title": "Reasons for admission",
      "text": {
        "div": "for op"
      }
    },
    {
      "title": "Present Illness",
      "text": {
        "div": "The patient has a history of ~~~~
         and was admitted for ~~~~ ."
      }
    },
  ]
}
```

```
{
  "resourceType": "Careplan",
  "title": "Consultation request",
  "description": "Requesting for consultation
  regarding to ~~~~",
  "created": "2025-01-13",
  "author": {
    "display": "Department of colorectal surgery"
  },
  "note": {
    "authorReference": {
      "display": "Department of gastroenterology"
    },
    "time": "2025-01-13",
    "text": "No contraindications to surgery ~~~~."
  }
}
```

```
{
  "resourceType": "Observation",
  "code": {
    "text": "Chest PA",
  },
  "effectivePeriod": {
    "start": "2024-12-16",
    "end": "2024-12-17"
  },
  "valueString": "Impression: No active lung
   disease."
}
```

HL7 FHIR
Input Data
Patient information
Clinical observations (Current hospitalization)
Nursing records (Current hospitalization)
Admission notes (Most recent data)
Nursing admission assessments (Most recent data)
Specialty consultations (Recent 3 months data)
Discharge summaries (Most recent data)
Anesthesia records (Entire period)
Outpatient consultations (Recent 1 year data)
Laboratory results (Recent 3 years data)
Radiology reports (Entire period)
Results of functional tests (Recent 5 years)
Y-KNOT
Y-KNOT-MD LLM
Rule-based System
Pre-Anesthetic Assessments

Jane Doe (F/46)
169.9 cm, 83.6 kg, BMI 29.66
BP 111/82 mmHg, PR 65/min, RR 20/min,
BT 37.0°C
Mental status: Alert
Allergy: Fentanyl → nausea (2019-04-13)

Scheduled operation:
Uterine myoma → Laparoscopic incisional hernia repair

Past medical history:
- mild scoliosis
- s/p Robot assisted myomectomy (2019-04-11)
- s/p umbilical hernia repair(2024-07-02)

Medications: None

Anesthesia History:
- Robot assisted myomectomy (2019-04-11): General anesthesia – Circle with Endotracheal tube (laryngoscope grade: 1). Sent to PACU

Labs:
CBC (2024-12-16):
- WBC COUNT: 6.63 10^3/μL
- Hemoglobin: 13.6 g/dL
- Hct: 40.2 %
Blood coagulation test (2024-12-16):
- PT(Sec): 11.1 sec
- PT(INR): 0.94 INR
- aPTT: 32.8 sec
Chemistry (2024-12-16):
- Total Protein: 6.9 g/dL
- Albumin: 4.4 g/dL
- AST(GOT): 18 IU/L
- ALT(GPT): 14 IU/L
- T. Bilirubin: 0.5 mg/dL

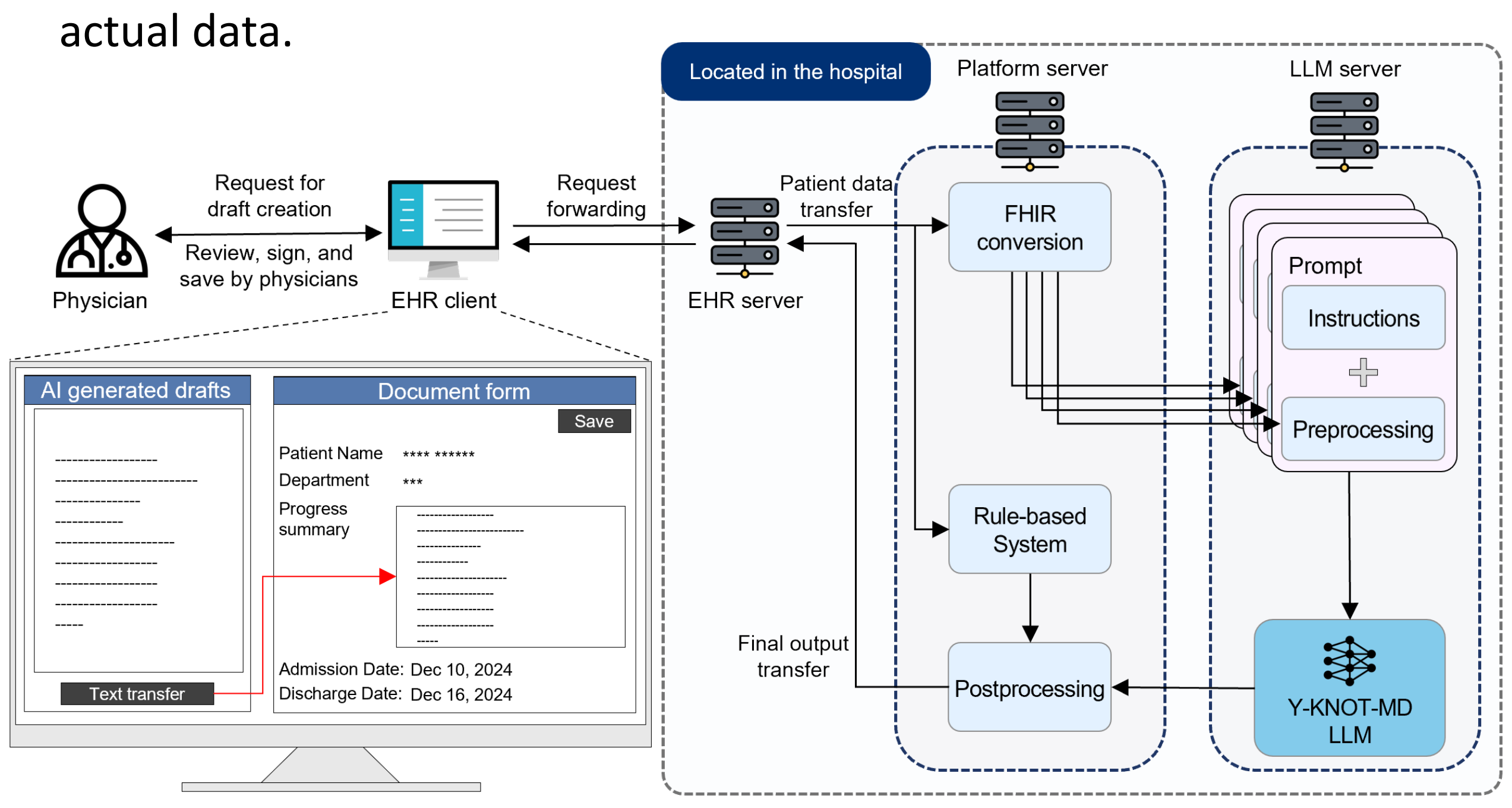
Imaging findings:
- Chest X-ray (2024-12-16)
No active lung disease.
- CT Abd+Pel (contrast) (2024-12-11):
1. Umbilical or incisional herniation of omental fat and small bowel loops.
2. GB stones and mild wall thickening at GB distal body

Pulmonary function test (2024-12-16):
FVC(Pre) 3.66 93% -> (Post) 3.71 94%
FEV1 (Pre) 2.37 76% -> (Post) 2.56 82%
FEV1/FVC (Pre) 65% -> (Post) 69%

Specialty consultations:
- GI (2025-01-13): Elevated AST/ALT (234/112) were d/t high intensity exercise and medications. Levels were WNL in the latest lab results. No contraindications to surgery were identified. Caution is advised when using hepatotoxic drugs.

Figure 1. Drafting clinical notes with LLMs: from FHIR input to clinical output.

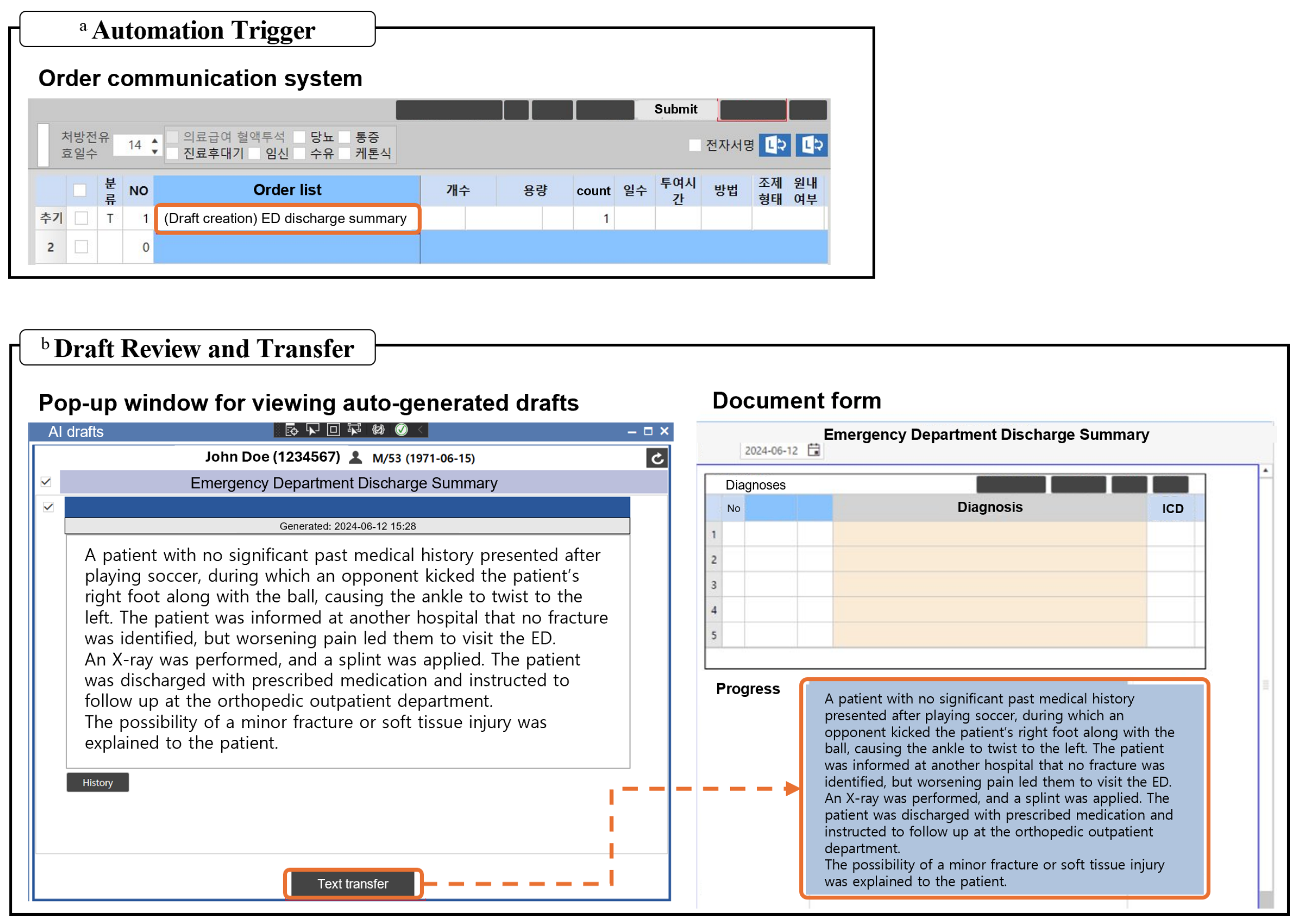
All medical records used as input data are converted into FHIR standards. Criteria for selecting input data are stated in parentheses. The examples provided in the figure are simplified versions of the actual data.



Patient Name **** ******
Department ***
Progress summary
Admission Date: Dec 10, 2024
Discharge Date: Dec 16, 2024

Figure 2. Overview of the automated drafting process with the AI agent in the EHR system.

When the drafting is initiated, patients' data is transferred from the EHR server to the Y-KNOT system. The data is converted into FHIR structures and processed in combination of LLM and rule-based approaches. Multiple prompts are created for the LLM, each designed to extract specific aspects of the document. Data exchange between system components operates through predefined APIs. All servers and databases are located within the hospital's secure on-premise environment.



a Automation Trigger

Order communication system

	□	분류	NO	Order list	개수	용량	count	일수	투여시간	방법	조제형태	원내여부
추기	□	T	1	(Draft creation) ED discharge summary			1					
2	□		0									

b Draft Review and Transfer

Pop-up window for viewing auto-generated drafts

John Doe (1234567) M/53 (1971-06-15)

☑ Emergency Department Discharge Summary

☑ Generated: 2024-06-12 15:28

A patient with no significant past medical history presented after playing soccer, during which an opponent kicked the patient's right foot along with the ball, causing the ankle to twist to the left. The patient was informed at another hospital that no fracture was identified, but worsening pain led them to visit the ED.
An X-ray was performed, and a splint was applied. The patient was discharged with prescribed medication and instructed to follow up at the orthopedic outpatient department.
The possibility of a minor fracture or soft tissue injury was explained to the patient.

Document form

Emergency Department Discharge Summary

2024-06-12

Diagnoses

No			Diagnosis	ICD
1				
2				
3				
4				
5				

Progress

A patient with no significant past medical history presented after playing soccer, during which an opponent kicked the patient's right foot along with the ball, causing the ankle to twist to the left. The patient was informed at another hospital that no fracture was identified, but worsening pain led them to visit the ED.
An X-ray was performed, and a splint was applied. The patient was discharged with prescribed medication and instructed to follow up at the orthopedic outpatient department.
The possibility of a minor fracture or soft tissue injury was explained to the patient.

Figure 3. Example of user interaction with the EHR system for automatic clinical drafting: ED discharge summaries

a: An automatic clinical drafting is triggered by ordering a draft creation from the order communication system. In case of pre-anesthetic assessments, the drafting processes are triggered in batch according to procedure schedules.

b: As a physician opens a form for documentation, auto-generated drafts show up in the pop-up window. Selected draft is directly transferred to the form if the physician clicks the 'text transfer' button. Drafts then can be edited and saved on the document form.

Implications

- Y-KNOT project demonstrates the first seamless integration of an AI agent into an EHR for clinical drafting in routine clinical practice.
- Small LLM ensured minimal latency of the EHR, cost-efficiency, and environmental sustainability.
- The on-premise approach addressed the Korean medical regulation[7] requiring all medical records to be stored on domestic servers.
- Collaboration with related departments ensured that the AI enhances, rather than disrupts, the existing clinical workflows.
- This study highlights a practical and scalable approach to utilizing LLM-based AI agents for other institutions.

References

1. Tajirian T, Stergiopoulos V, Strudwick G, et al. The Influence of Electronic Health Record Use on Physician Burnout: Cross-Sectional Survey. *J Med Internet Res.* 2020;22(7):e19274.
2. Gaffney A, Woolhandler S, Cai C, et al. Medical Documentation Burden Among US Office-Based Physicians in 2019: A National Study. *JAMA Intern Med.* 2022;182(5):564-566.
3. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs*). NPJ Digit Med.* 2024;7(1):183.
4. Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, et al. The llama 3 herd of models. *arXiv preprint arXiv*:2407.21783
5. Cheng D, Gu Y, Huang S, Bi J, Huang M, Wei F. Instruction pre-training: Language models are supervised multitask learners. *arXiv preprint arXiv*:2406.14491
6. HL7 FHIR. HL7 International. Accessed December 1, 2024. https://hl7.org/fhir/
7. Ministry of Health and Welfare (South Korea), Korea Health Information Service. Guidelines for the Standards on Facilities and Equipment Required for the Management and Preservation of Electronic Medical Records. 2022.

Contact information

oneash082498@yuhs.ac (Hanjae Kim) ; imipenem@phidigital.co.kr (So-Yeon Lee) ; chandryou@yuhs.ac (Seng Chan You)