Summarizing Clinical Notes using LLMs for ICU Bounceback and Length-of-Stay Prediction

Akash Choudhuri, Philip M. Polgreen, Alberto M. Segre, Bijaya Adhikari
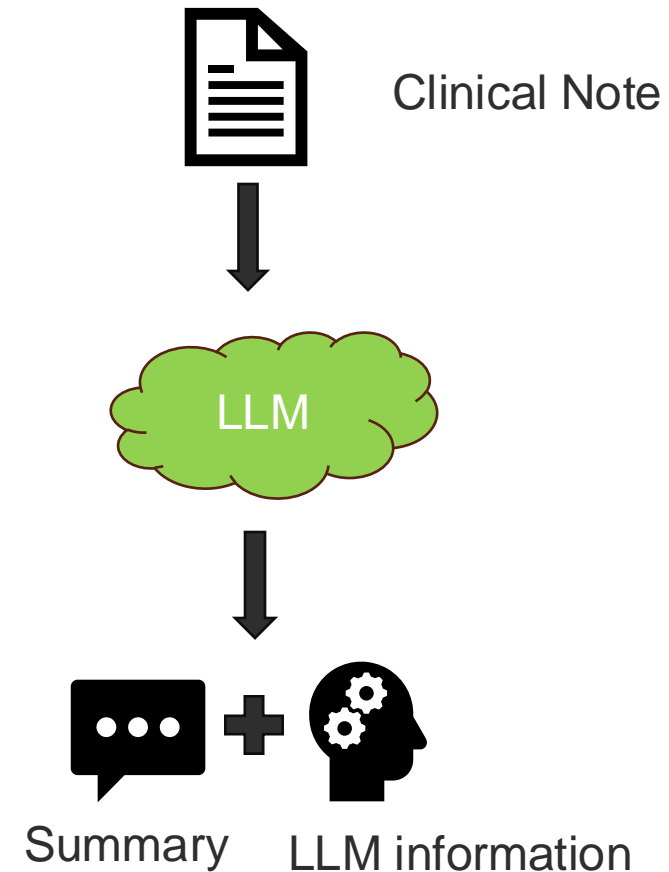
University of Iowa

For the CDC MInD Healthcare Network

University of Iowa
comp|epi
computational epidemiology research

CDC
CENTERS FOR DISEASE
CONTROL AND PREVENTION

THE
UNIVERSITY
OF IOWA

Contents

- Motivations

- Model and Components

- Experiments

- Results

- Discussion

THE UNIVERSITY OF IOWA

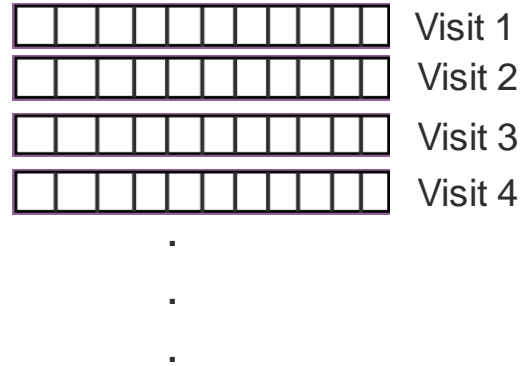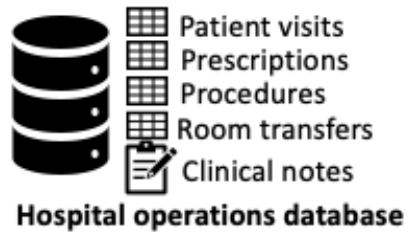# Motivation: LLMs Provide Additional Information

Motivation

- LLMs are trained on a wide variety of corpora

- They have general information about a wide variety of topics if prompted correctly

- Simple text encoders like BERT are not equipped with such vast knowledge and thus cannot add additional information to the knowledge that is already present
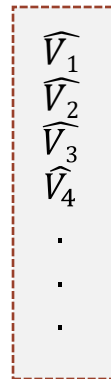
Clinical Note

LLM

Summary LLM information

Problem Formulation

Model and Components

Given:

Hospital operations database

Visit 1
Visit 2
Visit 3
Visit 4

Learn:

Visit 1
Visit 2
Visit 3
Visit 4

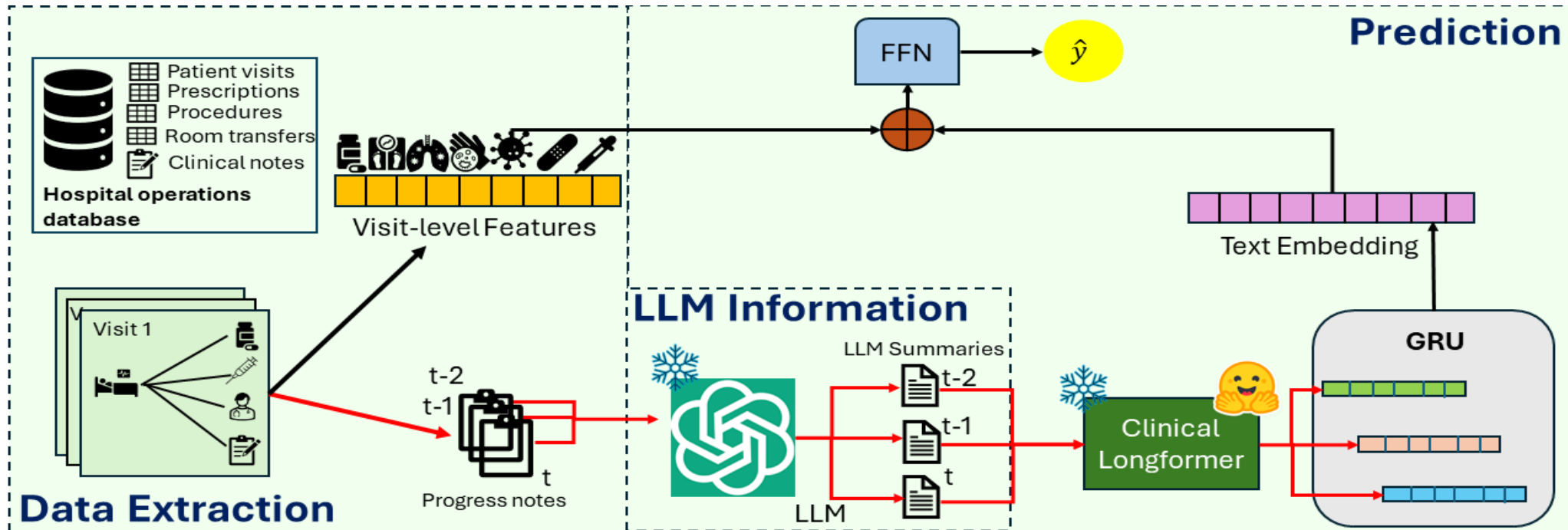Model → $\widehat{V_1}$ $\widehat{V_2}$ $\widehat{V_3}$ $\widehat{V_4}$
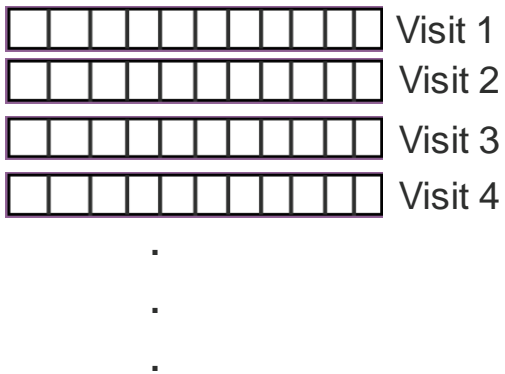
Such That:

A loss function $L(V_i, \hat{V}_i)$ is minimized

Overall Framework

Model and Components

- The steps denoted by red arrows are performed separately than the steps denoted by black arrows
- Data Extraction: Constructs visit-level data and progress notes for each individual visit
- The progress notes are sequentially inputted to the frozen LLM to generate summaries-> embeddings via Clinical Longformer [1] -> GRU gives overall embedding

[1] Li, Yikuan, et al. "Clinical-longformer and clinical-bigbird: Transformers for long clinical sequences." arXiv preprint arXiv:2201.11838 (2022).

Data Extraction

Model and Components

For each visit, extract the following tabular visit information from EHR:

- Age
- Gender
- Length of Stay
- Body Mass Index (BMI)
- Glasgow Coma Scale Scores:
    - Eye
    - Motor
    - Verbal
- Vital sign values from Blood Culture:
    - Albumin
    - Bicarbonate
    - Bilirubin
    - BUN
    - Chloride
    - Creatinine
    - Glucose
    - Haemoglobin, etc.

LLM Information Extraction

Model and Components

- Extract all Clinical Progress Notes by physicians sequentially for each visit

- For each note in the sequence:
  - Generate LLM Summary from the note
  - Pre-process the generated summary
  - Get embedding of the pre-processed summary

- The sequence of embeddings are the inputs for the GRU

Prompt Template

Context: You are a Medical Specialist capable of analyzing complex medical cases. To do that, you will carefully review the physician's daily progress note for a patient in the ICU detailing the patient's health progress in the ICU.

Engineering: If a normal human medical professional has level 10 knowledge, you will have level 3000 of knowledge in this role. As it's crucial for the patient's well-being, please ensure accurate assessments because if you don't, the patient's condition could worsen. Take pride in your work and give it your best.

Task: The task involves analyzing the progress note documenting the patient's health progress in the ICU. Based on this information, you need to summarize the information in the note and determine the potential medical complications and conditions that can arise.

Noise Prevention: Do not tell your limitations. Do not echo my prompt. Please do not remind me what I asked you for. Do not apologize. Do not self-reference. Get to the point precisely and accurately.

THE UNIVERSITY OF IOWA

LLM Summary Example

Model and Components

Input Note: Chief Complaint: **Shortness of breath**. Fever. I saw and examined the patient, and was physically present with the ICU Resident for key portions of the services provided. I agree with his/her note above, including assessment and plan. 24 Hour Events: Admitted, did not required more Bipap overnight. Stable respiratory status on 6L NC and weaned to 3L (baseline O2 requirement)....Physical Examination General Appearance: Well nourished, No acute distress Eyes / Conjunctiva: No(t) PERRL. Head, Ears, Nose, Throat: Normocephalic Cardiovascular: (S1: Normal), (S2: Normal) Peripheral Vascular: (Right radial pulse: Not assessed), (Left radial pulse: Not assessed), (Right DP pulse: Not assessed), (Left DP pulse: Not assessed) Respiratory / Chest: (Breath Sounds: Crackles : bilateral symmetric) Abdominal: Soft, Non-tender. Extremities: Right lower extremity edema: Absent, Left lower extremity edema: Absent, Skin: Warm Neurologic: Attentive, Follows simple commands, Responds to: Verbal stimuli, Movement: Purposeful, Tone: Normal

...

Assessment and Plan H/O FEVER: High grade fever in the ER. follow up blood cultures, also follow up sputum and urine culture. Cover broadly with levo/vanco..H/O PNEUMONIA, OTHER; continue antibiotics, also apperas volume overloaded, HD today with fluid removal. Sarcoid and pulmonary HTN: on prednisone, bactrim, Oxygen supplementation Renal failure: on HD TTS schedule, HD today. DVT: Boots (on no anticoagulation due to severe prior GI bleeds.) Stress ulcer: PPI Code status: Full code Disposition : Transfer to floor

LLM Generation: The patient, a **short-of-breath individual with a fever**, has been admitted to the ICU. Overnight, they remained stable with a weaned oxygen requirement from 6L to 3L. They have a history of allergies, medications, and medical conditions, including **thrombocytopenia, sarcoidosis, pulmonary hypertension, and renal failure**. Laboratory results show elevated white blood cell count, hemoglobin, and creatinine levels, as well as low platelet count. The patient's vital signs are stable, with a temperature of 36.6°C, heart rate of 98 bpm, and blood pressure of 110/77 mmHg. They are receiving antibiotics, oxygen supplementation, and dialysis.

Potential Medical Complications:

1. Sepsis: The patient's high-grade fever and elevated white blood cell count suggest the possibility of sepsis, which could lead to organ dysfunction and failure. **2. Pneumonia:** The patient's respiratory symptoms and imaging results indicate pneumonia, which could worsen their respiratory status and increase the risk of complications. **3. Renal Failure:** The patient's pre-existing renal failure could worsen with continued fluid overload, leading to electrolyte imbalances and acidosis. **4. Thrombocytopenia:** The patient's low platelet count increases the risk of bleeding complications, particularly in the setting of invasive procedures or anticoagulation therapy. **5. Pulmonary Hypertension:** The patient's underlying pulmonary hypertension could worsen with respiratory failure, leading to increased risk of cardiac complications. **6. Gastrointestinal Bleeding:** The patient's history of severe prior GI bleeds increases the risk of bleeding complications, particularly in the setting of anticoagulation therapy or invasive procedures. **7. Infection:** The patient's broad-spectrum antibiotic coverage and positive blood cultures suggest the possibility of infection, which could lead to antibiotic resistance and treatment failures.

Dataset

Experiments

- Used MIMIC-III EHR Dataset:
  - Only used information about the patients who were admitted to the ICU
  - Stayed there for more than 2 days for each admission to the ICU

- Additionally, there were 2,083,180 clinical notes are broadly divided into 15 categories:
  - Only consider those clinical notes under the category 'Physician' and the subcategories 'Physician Resident Progress Note' and 'Physician Attending Progress Note'
  - 53,321 and 17,771 clinical notes were under the sub-categories 'Physician Resident Progress Note' and `Physician Attending Progress Note' respectively

Application 1: ICU Bounceback Prediction

Experiments

- Utilize information from the patient's current ICU visit to predict whether a patient is at risk of being transferred back to the ICU after discharge

- Binary Classification Problem:

  - Instance: Features obtained for the current ICU visit

  - Label: Binary indicator of whether the patient gets readmitted to the ICU during the current hospital visit

- Evaluation Metrics: ROC-AUC, AUPRC

- Label Distribution:

Class	Count
Negative	2703
Positive	137
Total	2840

Application 2: ICU LOS Prediction

Experiments

- Predict the total length of stay (LOS) for each patient visit in the ICU

- Multi-Class Classification Problem:
  - Between 2-4 days: low
  - Between 4-7 days: medium
  - 7 days and above: high

- Evaluation Metrics: Macro F-1 score, AUPRC

- Label Distribution:

Class	Count
Low	1437
Medium	674
High	729
Total	2840

Models

Experiments

The models used in the study were as follows:

- **BASE:** Uses only the visit level features to make predictions (no notes)

- **NOTES:** Directly uses the clinical note embeddings and the visit level features to make predictions

- **LLAMA3:** The LLM used to create summaries is LLAMA3-8B[1]

- **MedLLAMA:** The LLM used to create summaries is a fine-tuned version of LLAMA3-8B[2]

- **LLAMA3-Meerkat:** The summaries uses a fine-tuned version of LLAMA3-8B on a synthetic dataset consisting of high-quality chain-of-thought reasoning paths sourced from 18 medical textbooks, along with diverse instruction-following datasets[3]

[1] https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md
[2] https://huggingface.co/johnsnowlabs/JSL-MedLlama-3-8B-v2.0
[3] Kim, Hyunjae, et al. "Small language models learn enhanced reasoning skills from medical textbooks." *arXiv preprint arXiv:2404.00376* (2024).

Performance of models on 2 Applications

Results

Model	ICU Bounceback Prediction		ICU Length of Stay Prediction	
	AUC-ROC	**AUPRC**	**Macro-F1**	**AUPRC**
BASE	62.49 ± 0.16	9.54 ± 0.14	34.38 ± 0.58	42.11 ± 0.23
NOTES	66.08 ± 0.73	10.38 ± 0.42	41.08 ± 0.46	44.93 ± 0.73
LLAMA3 [39]	69.79 ± 0.51	11.74 ± 0.15	40.94 ± 0.59	48.13 ± 0.64
MedLLAMA [40]	66.17 ± 0.05	10.86 ± 0.83	41.76 ± 0.52	47.86 ± 0.03
LLAMA3-Meerkat [41]	70.82 ± 0.76	11.85 ± 0.39	42.25 ± 0.56	46.75 ± 0.26

- NOTES and LLAMA models have clear gain over BASE

- LLAMA models outperform NOTES in both Applications:
  - Fine-tuning may not always be beneficial, as indicated by the comparison between LLAMA3 and MedLLAMA in ICU Bounceback Prediction
  - For LOS prediction, LLAMA3-Meerkat is the best-performing LLM in terms of Macro-F-1 score. However, LLAMA3 and MedLLAMA cannot outperform NOTES in terms of Macro F-1 score. Evaluating the models on the AUPRC metric shows that LLAMA models have clear performance gain over NOTES

- Although LLAMA3 was not explicitly pre-trained to cover medical text, it performs competitively compared to the fine-tuned variants for both the tasks

The University of Iowa

Analyzing Similarities in LLM Generations

Discussion

- We wanted to compare the diversity of medical topics in the texts

  - We concatenated all the clinical notes/ LLM summaries throughout a visit

  - We then used the biomedical Named-Entity Recognition (NER) pipeline from ScispaCy to extract relevant medical terms from the texts

  - Used Jaccard Score to compare the similarity of terms appearing across the text

LLM	Jaccard Score
Notes:LLAMA3	0.1446
Notes:MedLLAMA	0.1226
Notes:LLAMA3-Meerkat	0.1543
LLAMA3:MedLLAMA	0.1977
LLAMA3:LLAMA3-Meerkat	0.2001
MedLLAMA:LLAMA3-Meerkat	0.2497

Summaries

...cial than notes

Chief Complaint: Respiratory Distress\n 24 Hour Events:\n PICC LINE - START [**2144-10-13**] 01:18 PM\n INVASIVE VENTILATION - START [**2144-10-13**] 01:30 PM\n ARTERIAL LINE - START [**2144-10-13**] 02:00 PM\n EKG - At [**2144-10-13**] 05:00 PM\n NPO for aspiration risk and possible extubation today\n **Abnormal EKG compared to prior**; starting cycling enzymes; negative x1\n so far\n History obtained from Patient, Family / [**Hospital 56**] Medical records\n Allergies:\n History obtained from Patient, Family / [**Hospital 56**] Medical [**Name2 (NI) 4484**] Known\n Drug Allergies\n Last dose of Antibiotics:\n Metronidazole - [**2144-10-14**] 12:52 AM\n Infusions:\n Fentanyl - 25 mcg/hour\n Midazolam (Versed) - 1 mg/hour\n Other ICU medications:\n Furosemide (Lasix) - [**2144-10-13**] 06:00 PM\n Heparin Sodium (Prophylaxis) - [**2144-10-14**] 12:53 AM\n Other medications:\n Changes to medical and family history:\n Review of systems is unchanged from admission except as noted below\n Review of systems:\n Flowsheet Data as of [**2144-10-14**] 06:28 AM\n Vital signs\n Hemodynamic monitoring\n Fluid balance\n24 hours\nSince 12 AM\n Tmax: 37.6\nC (99.7\n Tcurrent: 37.2\nC (99\n HR: 67 (65 - 91) bpm\n BP: 118/49(69) {99/44(59) - 172/85(117)} mmHg\n RR: 14 (14 - 24) insp/min\n SpO2: 100%\n Heart rhythm: SR (Sinus Rhythm)\nTotal In:\n243 mL\n160 mL\n PO:\nTF:\n IVF:\n153 mL\n131 mL\n Blood products:\n Total out:\n1,965 mL\n 205 mL\n Urine:\n1,965 mL\n205 mL\n NG:\n Stool:\n Drains:\n Balance:\n-1,723 mL\n-44 mL\n Respiratory support\n O2 Delivery Device: Endotracheal tube\n Ventilator mode: CMV/ASSIST/AutoFlow\n Vt (Set): 550 (550 - 550) mL\n RR (Set): 14\n RR (Spontaneous): 0\n PEEP: 5 cmH2O\n FiO2: 40%\n RSBI Deferred: Sao2 < 90%\n PIP: 25 cmH2O\n Plateau: 20 cmH2O\n SpO2: 100%\n ABG: 7.39/38/90.[**Numeric Identifier 128**]/22/-1\n Ve: 8 L/min\n PaO2 / FiO2: 225\n Physical Examination .....
Assessment and Plan\n THIS IS THE PLAN FROM YESTERDAY; PRELIMINARY NOTE FOR ROUNDS USE\n ONLY!!!!\n 59 y/o F with h/o mental retardation, recent dx of Nodular sclerosing\n Hodgkins Lymphoma s/p chemo on [**2144-10-9**] **presented with acute respiratory\n distress likely due to aspiration & mucus plugging** now intubated.\n .\n # Hypoxia: Likely due to aspiration & mucus plugging in the setting of\n depressed mental status, coughing and emesis. A component of volume\n overload cannot be excluded given her I/O positive >7 liters over the\n past several days. A primary lung reaction to her chemo is much less\n likely. Other possibilities including a primary cardiac event or PE are\n less likely with a relatively unchanged EKG and no tachycardia. Given\n rapid improvement of CXR after intubation, suspect that right lung\n white out was due to mucus plugging. ABG post intubation showed\n **signficant acidosis** with pH of 7.27.\n - increase RR & recheck ABG\n - advance ETT & recheck placement\n - no need for Abx currently\n - f/u sputum Cx & low threshold to start ABx if spikes temp\n - wean sedation, SBT & RSBI in am\n - continue regular suction & albuterol MDI\n - diuresis goal neg 1-2L\n - consider thoracentesis for diagnosis & therapeutic\n - recheck EKG (given pseudonormalization on tracing from am)\n - cycle CE to ROMI\n # Hodgkins Lymphoma: Pt is s/p EACoPP protocol on [**2144-7-21**] with\n cytoxan, doxorubicin on day 1 and etoposide on days [**2-13**]. Pt has also\n had pericardial tamponade s/p pericardiocentesis on [**2144-6-24**] and\n pericardial window on [**2143-7-25**]. Pt initiated ICE chemo reg on [**2144-10-9**].\n This acute hypoxia more likely due to aspiration & mucus plug, unlikely\n due to chemo toxicity.\n - continue Acyclovir & Bactrim ppx\n - f/u Onc recs\n - transfuse to maintain hct>25\n .\n # H/o megacolon/Diarrhea: Pt is s/p volvulous & sigmoid resection with\n large dilated loops of colon visible on portable films. She has\n required intermittent rectal tubes, currently having alot of stool\n output with recent C. Diff toxin + [**2144-8-8**] & [**2144-9-1**].\n - will keep NPO for now (possible extubation in am)\n - continue po vanco & IV flagyl\n - place rectal tube now\n - OG tube to intermittent suction\n - f/u C diff toxin\n - repeat imaging in am\n .\n # Paroxysmal A-fib: Recent adm to ICU due to poor rate control (rates\n of 160 with relative hypotension). Rates have been better controlled on\n current regimen of Metoprolol, Amiodarone & Diltiazem. No plan for\n anticoagulation due to bloody pericardial effusion. Currently in NSR\n with rate in 60s.\n - continue Metoprolol & diltiazem & amiodarone\n - f/u repeat EKG\n .\n # UA: **Pt had positive UA** on [**10-11**], Urine Cx + fungus 4000-6000 CFUs.\n Received fluconazole 400mg x 1\n - change out foley now\n - resend UA & Urine Cx\n - hold off on Fluconazole for now\n .\n # F/E/N: IVF prn, NPO for now, OG tube to intermittent suction\n - replete lytes agressively\n .\n # PPX: heparin SQ, PPI\n .\n # CODE: Intubate/DNR\n .\n # ACCESS: L PICC, working on PIV\n .\n # CONTACT: HCP - brother [**Name (NI) 3080**] [**Telephone/Fax (1) 3081**] (mobile) and [**Telephone/Fax (1) 3082**]\n (home); brother also gave permission to speak with his wife [**Name (NI) **]\n ICU Care\n Nutrition:\n Glycemic Control:\n Lines:\n PICC Line - [**2144-10-13**] 01:18 PM\n Arterial Line - [**2144-10-13**] 02:00 PM\n Prophylaxis:\n DVT:\n Stress ulcer:\n VAP:\n Comments:\n Communication: Comments:\n Code status: DNR (do not resuscitate)\n Disposition:\n

...ntal retardation, recently diagnosed with Nodular Sclerosing ...ntricular response. The patient has a history of chemotherapy, and

...bpm, blood pressure of 102/48 mmHg, and respiratory rate of 28

...T service. Monitoring of daily ANC counts and maintenance of

...cefepime has been started due to neutropenia. ...mponade physiology, and further monitoring of vital signs is

...history of c. diff colitis and oral/genital herpes. ...d will be followed up with repeat UCx. ...ng it due to chemotherapy. ...upper limb and is not receiving anticoagulation due to

...**2144-9-21**] and is experiencing abdominal distention. Frequent ...l. ...comycin. ...on of ACV prophylaxis dose is recommended.

1. Sepsis: Given the patient's neutropenia and possible LUL infiltrate, there is a risk of developing sepsis.
2. Tamponade: If the patient's pericardial effusion worsens, there is a risk of developing tamponade physiology.
3. Hemorrhage: The patient's history of hemopericardium and neutropenia increases the risk of bleeding.
4. **Respiratory failure: The patient's atrial fibrillation and possible LUL infiltrate increase the risk of respiratory failure.**
5. **Renal failure: The patient's history of chemotherapy and possible UTI increase the risk of renal failure.**
6. **GI complications: The patient's history of c. diff and possible ileus/abdominal distention increase the risk of GI complications.**
Overall, the patient requires close monitoring and management of her medical complications to prevent further deterioration of her condition.

Future Directions

- Fine-tuning a model may be a very tricky option:
  - May not generalize to all tasks
  - There is a chance that fine-tuning data may be recovered from the LLM

- Retrieval Augmented Generation (RAG) methods may be a good step forward

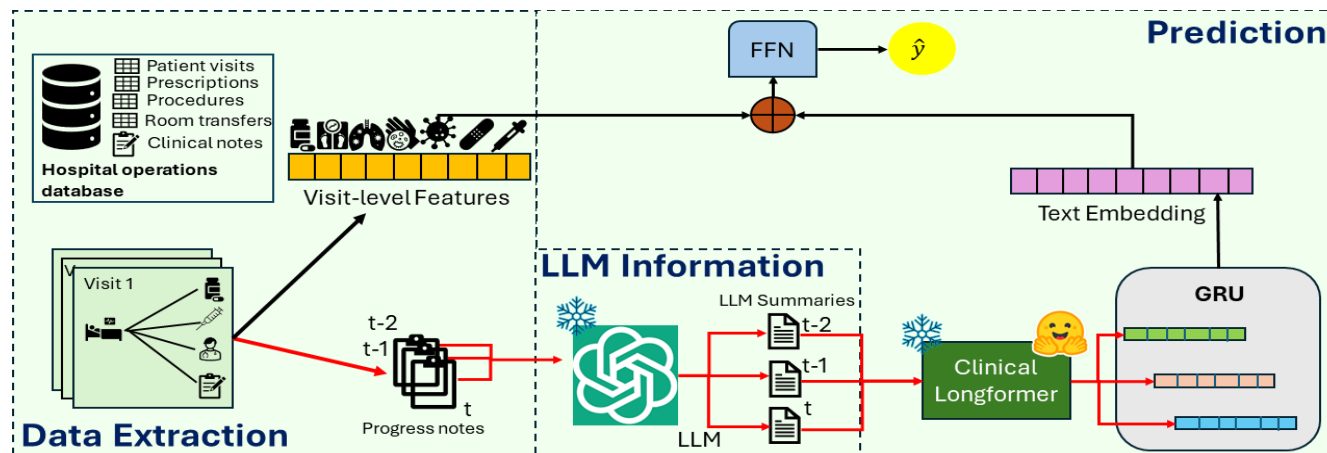# Team

Akash Choudhuri   Philip M. Polgreen   Alberto M. Segre   Bijaya Adhikari

Special Thanks

LLM Outputs: https://github.com/Soothysay/LLM-Outputs.git

Contact:
akash-choudhuri@uiowa.edu