

Introducing GutGPT – An AI Chatbot to Provide Interpretable and Context-Guided Risk Assessment for Patients with Gastrointestinal Bleeding

Colleen Chan

Yale University, Dept. of Statistics and Data Science

JSM 2023

Yale SCHOOL OF MEDICINE



GUT-GPT

Artificial Intelligence and Large Language Models (LLMs) in Medicine

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

Jeffrey M. Drazen, M.D., *Editor*;
Isaac S. Kohane, M.D., Ph.D., and Tze-Yun Leong, Ph.D., *Guest Editors*

AI IN MEDICINE

Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine

Peter Lee, Ph.D., Sebastien Bubeck, Ph.D., and Joseph Petro, M.S., M.Eng.

JAMA Network

VIEWPOINT

AI-Generated Medical Advice—GPT and Beyond

Claudia E. Haupt, JSD, PhD

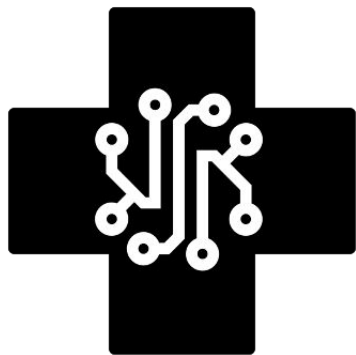
Northeastern
University School of
Law, Boston,
Massachusetts; and
Solomon Center for
Health Law & Policy,
Yale Law School, New
Haven, Connecticut.

Mason Marks, MD, JD

Florida State University
College of Law,
Tallahassee; and
Project on Psychedelics
Law and Regulation
(POPLAR), Petrie-Flom
Center for Health Law
Policy, Biotechnology,
and Bioethics at
Harvard Law School,
Boston, Massachusetts.

Few AI Clinical Tools Have Demonstrated Clinical Benefit

24%



of randomized controlled trials studying a **machine learning intervention** reported a statistically significant **improved primary clinical outcome**

51%

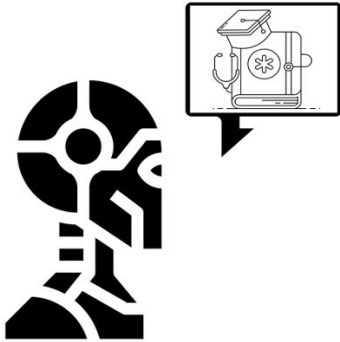


of randomized controlled trials studying a **machine learning intervention** were performed at a **single site**.

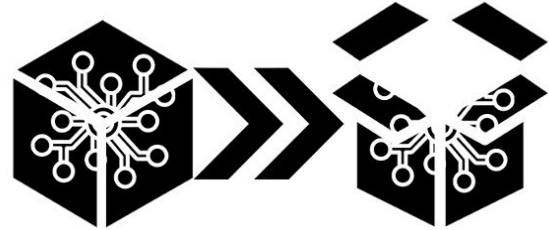
Improving Human Understanding

Blackbox nature: healthcare workers distrust these tools, leading to limited use

AI Chatbots may help

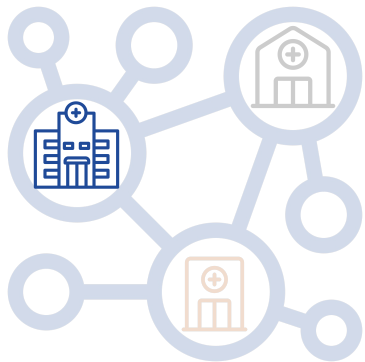


Point of care responses to relevant sections from sources of medical knowledge



Translate the inner workings of a model through natural language

Goals



We want to develop interpretable clinical risk machine learning models that are robust to distribution shifts and can transfer across hospitals



We want the clinical risk machine learning models to be implemented into the human clinical workflow effectively

Introducing

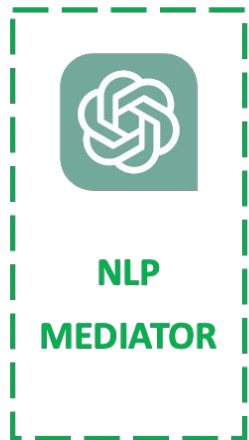
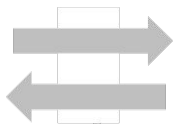


GUT-GPT

Our Vision – Artificial Intelligence as a Team Member

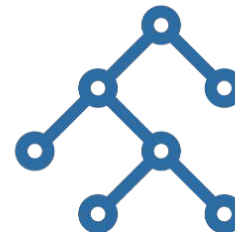


- Automatic extraction of structured and unstructured data from the electronic health record
- Direct interaction via natural language with model regarding risk stratification
- Dynamic, interpretable risk computation
- Dynamic interaction of patient management recommendations according to medical guidelines



Naturalistic Human-Machine Interaction

Risk Assessment with In-house
Machine Learning Model

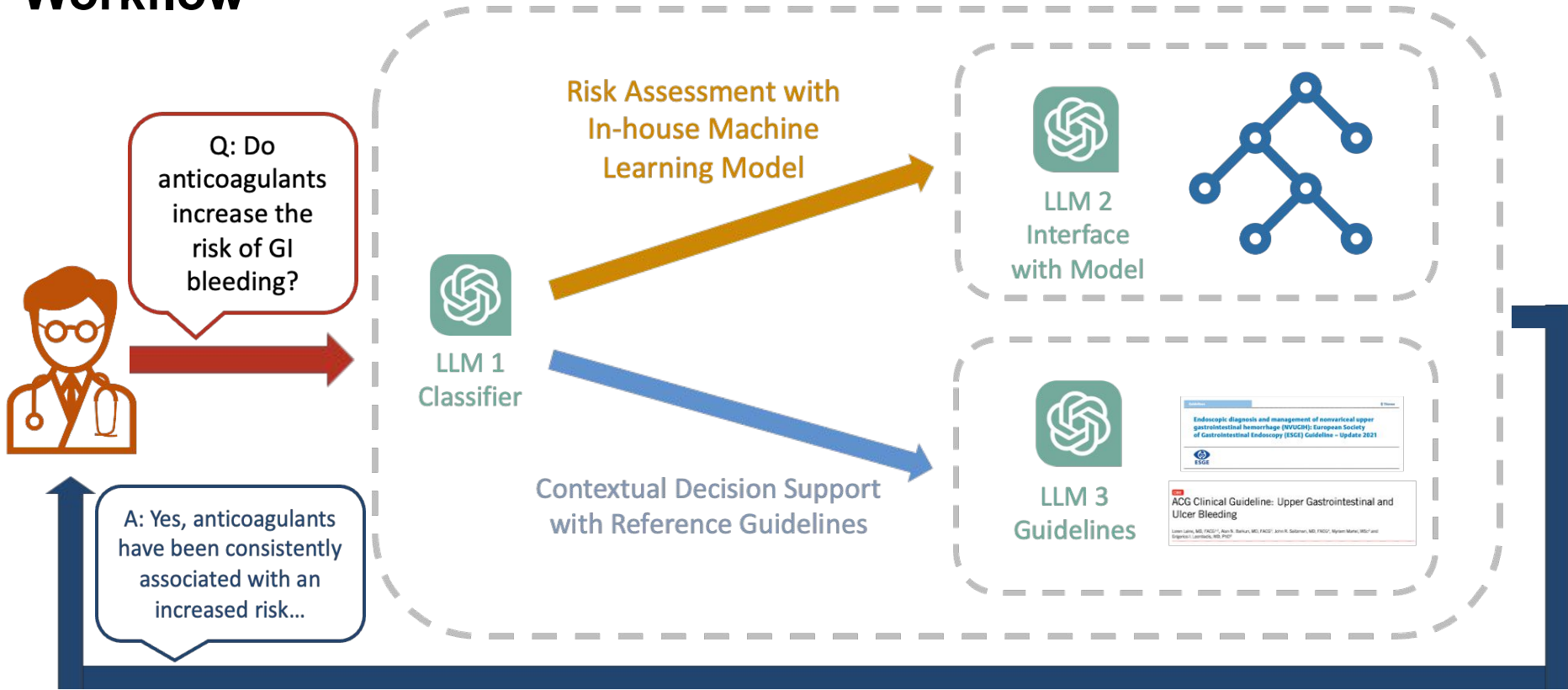


Contextual Decision Support
with Reference Guidelines



Two Type of Tasks

Workflow



LLM 4: Synthesizer

Adapting Large Language Models (LLM's)

Fine-Tuning

- Involves additional training on model parameters to update pre-trained weights
- Requires access to model architecture
 - Ex: BloombergGPT, Medical Bert-QA

In-Context Learning

- Examples are provided in the prompt and can use LLM as is
- Done at inference time

Our choice

Why?

- Few shot learning has already been shown to perform well (Brown et al., NeurIPS 2020)
- Fine tuning is expensive

Task 1: Guidelines



What are medical guidelines?

Guidelines

Thieme

Endoscopic diagnosis and management of nonvariceal upper gastrointestinal hemorrhage (NVUGIH): European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2021



CME

ACG Clinical Guideline: Upper Gastrointestinal and Ulcer Bleeding

Loren Laine, MD, FACP^{1,2}, Alan N. Barkun, MD, FACP³, John R. Saltzman, MD, FACP⁴, Myriam Martel, MSc² and Grigorios I. Leontiadis, MD, PhD⁵

Guidelines



OPEN ACCESS

Asia-Pacific working group consensus on non-variceal upper gastrointestinal bleeding: an update 2018

Joseph JY Sung,¹ Philip WY Chiu,¹ Francis K L Chan,¹ James YW Lau,¹ Khean-lee Goh,² Lawrence HY Ho,³ Hwoon-young Jung,⁴ Jose D Sollano,⁵ Takuji Gotoda,⁶ Nageshwar Reddy,⁷ Rajvinder Singh,⁸ Kentaro Sugano,⁹ Kai-chun Wu,¹⁰ Chun-Yin Wu,¹¹ David J Bjorkman,¹² Dennis M Jensen,¹³ Ernst J Kuipers,¹⁴ Angel Lanas¹⁵

Gold standard in patient management recommendations
for each geographic macroregion

Pre-processing the Clinical Guidelines



Guidelines

Statement 7: Over the scope-clipping devices (such as Overcap) are useful in treating lesions refractory to conventional endoscopic haemostatic therapy

(Accept—agreement: 94.4%, level of evidence: moderate)

In contrast to haemostatic powder sprays, the over-the-scope-clip (OTSC), if successfully applied, appears to provide a firm and durable control of bleeding in NUGIB.

Made from nitinol alloy, the OTSC fits to the tip of the endoscope and can be deployed by tightening the thread with the hand wheel using a mechanism similar to rubber band vaginal ligators. After being released from the applicator, the shape-memory effect and elasticity of the alloy result in firm closure of the clip. Compared with conventional clips, the OTSC can take up much more tissue by grasping deeper layers of the gastrointestinal wall, and hence the device can be used to treat bleeding and bowel perforation.

Several case series reported promising results of successful haemostasis in the range of 0–100%.^{30–32} Recurrent bleeding within 7 days occurred in 3–33%. This device, however, is technically slightly more demanding than other through-the-scope haemostatic treatments. Deployment of the OTSC requires accurate positioning and adequate retraction of tissue (either by suction or retractor) into the tip of the OTSC before the clip can be released properly. The retractor or anchor device is used in hard fibrotic ulcers, especially those located in difficult positions such as the high lesser curvature of the stomach. It guarantees the base of the lesion and allows tissue to be pulled into the cap. At certain locations in the stomach (eg, proximal lesser curve of the stomach) and duodenum (antrum of the first and second part of the duodenum), this can be technically challenging.

A multicentre randomised control trial comparing through-the-scope clips (TISC) with OTSC has recently been reported.³³ In that study, 52 patients received TISC and 33 received OTSC. Initial haemostasis was reported in 82.5% of those who received TISC and 96.8% received OTSC (P=0.007). Recurrent bleeding within 7 days after treatment occurred at the same rate in both groups (33.3% vs 24.6%). The interim results of this study suggested that OTSC is a better haemostatic device than haemoclips in the treatment of peptic ulcer bleeding. In view of the promising interim results while waiting for a full report, the working group recommends the use of OTSC in treating lesions refractory to conventional endoscopic therapy, such as through-the-scope haemoclips, thermal device or endoscopic injection. There may be a role for OTSC as primary therapy, especially in peptic ulcer bleeding with large vessels. This device will add to the armamentarium for NUGIB, with the level of evidence graded as moderate.

Statement 8: Endoscopic treatment of delayed bleeding after endoscopic mucosal resection or endoscopic submucosal dissection is similar to that for bleeding peptic ulcers.

(Accept—agreement: 89%, level of evidence: moderate)

An endoscopic mucosectomy (EM) and endoscopic submucosal dissection (ESD) are gaining popularity among tertiary centres worldwide, complications such as delayed bleeding require more guidelines from experts. A meta-analysis which pooled data from over 70 studies (15 RCTs, three prospective trials, five retrospective cohort studies, and 45 retrospective cohort and case-control studies) depicted clearly the risks associated with EM and ESD.³⁴ Post-ESD bleeding occurred in 5.10 (95% CI 4.59 to 5.76)% of patients. Risk factors identified included male gender, cardiac disease, the use of anti-thrombotic agents, chronic liver or kidney disease, tumour size >2 cm or resected specimen size >3 cm, lesions on the lesser curve, flat or depressed lesion and invasive carcinoma. Procedure time was not a clear risk factor for post-ESD bleeding but the need for endoscopic haemostasis was a factor. Postoperative time was not a clear risk factor for post-ESD bleeding but the need for endoscopic haemostasis was a factor. Postoperative time was not a clear risk factor for post-ESD bleeding but the need for endoscopic haemostasis was a factor. Postoperative time was not a clear risk factor for post-ESD bleeding but the need for endoscopic haemostasis was a factor.

Post-endoscopic management

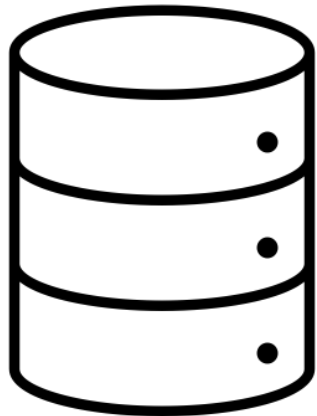
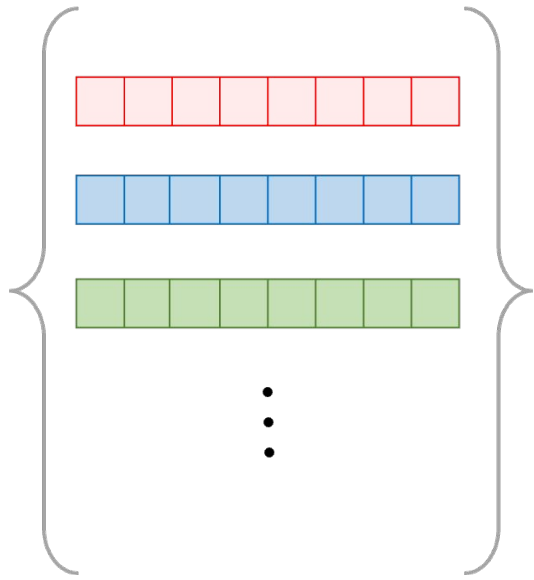
Statement 9: As an adjunct to endoscopic treatment, high-dose oral proton pump inhibitors can be used to prevent rebleeding

(Accept—agreement: 88.9%, level of evidence: moderate)

The use of intravenous high-dose proton pump inhibitors (PPIs) has become standard practice in the management of upper gastrointestinal bleeding. At least three randomised trials, all from South Asia, showed that oral PPIs given with or without endoscopic therapy, also reduce the risk of recurrent bleeding from peptic ulcers.^{35–37} New evidence suggests that high-dose oral PPIs may have a similar effect to delay action in preventing recurrent bleeding from peptic ulcers.

A study from Hong Kong recruited 113 high-risk patients with Forrest I or IIa/b peptic ulcer bleeding to receive either IV esomeprazole plus oral placebo or oral esomeprazole (40 mg every 12 hours) plus IV PPI placebo.³⁸ Recurrent bleeding within 30 days was reported in 7.7% in the IV esomeprazole group and 6.1% in the oral esomeprazole group. There was no difference in the requirement for blood transfusion, repeated endoscopic therapy and hospital stay between the two groups. It was noted that the study was stopped prematurely and was not designed as an equivalent trial. The trend suggests that the action of high-dose oral PPIs on peptic ulcer bleeding is comparable to that of IV PPI. In Taiwan, IV esomeprazole was compared with oral lansoprazole (30 mg four times a day for 3 days) in patients with peptic ulcer bleeding.³⁹ There was no difference in all the clinical outcome parameters, except that those who received oral PPI had a shorter hospital stay. In Korea, when IV esomeprazole was compared with oral esomeprazole (20 mg twice daily), the recurrent bleeding rate, surgical intervention and mortality between the two groups were comparable.⁴⁰

There is no properly powered RCT to confirm that high-dose oral PPI is as effective as IV PPI. The working group accepted that high-dose oral PPI can be used to prevent recurrent bleeding, but emphasised that it has to be used as an adjunct to endoscopic therapy. Only after endoscopic haemostasis is achieved, can high-dose oral PPI be recommended to prevent recurrent bleeding. The definition of high-dose oral PPI has been stated as at least 80 mg of esomeprazole (or



1. Section chunking

2. Embedding Chunks

3. Save as vector stores

Retrieval of Guidelines

What do the Asia Pacific guidelines recommend to prevent post-endoscopic bleeding?



Query embedding



Post-endoscopic management
Statement 9: As an adjunct to endoscopic treatment, high-dose oral proton pump inhibitors can be used to prevent rebleeding

(Accept—agreement: 88.9%, level of evidence: moderate)
The use of intravenous high-dose proton pump inhibitors (PPIs) has become standard practice in the management of upper gastrointestinal bleeding. At least three randomised trials, all from South Asia, showed that oral PPIs, given with or without endoscopic therapy, also reduce the risk of recurrent bleeding from peptic ulcer.^{45,52} New evidence suggests that high-dose oral PPIs may have a similar effect to their action in preventing recurrent bleeding from peptic ulcers.

A study from Hong Kong recruited 118 high-risk patients with Forrest I or IIa/b peptic ulcer bleeding to receive either IV esomeprazole plus oral placebo or oral esomeprazole (40 mg every 12 hours) plus IV PPI placebo.²¹ Recurrent bleeding within 30 days was reported in 7.7% in the IV esomeprazole group and 6.4% in the oral esomeprazole group. There was no difference in the requirement for blood transfusion, repeated endoscopic therapy and hospital stay between the two groups. It was noted that the study was stopped prematurely and was not designed as an equivalent trial. The trend suggests that the action of high-dose oral PPI peptic ulcer bleeding is comparable to that of IV PPI. In Taiwan, IV esomeprazole was compared with oral lansoprazole (30 mg four times a day for 3 days) in patients with peptic ulcer bleeding.⁵² There was no difference in all the clinical outcome parameters, except that those who received oral PPI had a shorter hospital stay. In Korea, when IV omeprazole was compared with oral rabeprazole (20 mg twice daily), the recurrent bleeding rate, surgical intervention and mortality between the two groups were comparable.²⁴

There is no properly powered RCT to confirm that high-dose oral PPI is as effective as IV PPI. The working group accepted that high-dose oral PPI can be used to prevent recurrent bleeding, but emphasised that it has to be used as an adjunct to endoscopic therapy. Only after endoscopic haemostasis is achieved, can high-dose oral PPI be recommended to prevent recurrent bleeding. The definition of high-dose oral PPI has been stated as at least 80 mg of esomeprazole (or

Find “similar” vectors

Score: 0.1



Score: 0.3

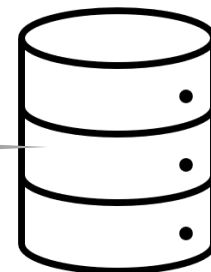


Score: 0.8



Retrieve the relevant portions of the guideline

⋮



Vector Store

What do the Asia Pacific guidelines recommend to prevent post-endoscopic bleeding?



Retrieved Text

Post-endoscopic management
Statement 9: As an adjunct to endoscopic treatment, high-dose oral proton pump inhibitors can be used to prevent rebleeding

(Accept—agreement: 88.9%, level of evidence: moderate)
The use of intravenous high-dose proton pump inhibitors (PPIs) has become standard practice in the management of upper gastrointestinal bleeding. At least three randomized trials, all from South Asia, showed that oral PPIs, given with or without endoscopic therapy, also reduce the risk of recurrent bleeding from peptic ulcers.^{46,47} New evidence suggests that high-dose oral PPIs may have a similar effect to their action in preventing recurrent bleeding from peptic ulcers.

A study from Hong Kong recruited 118 high-risk patients with Forrest I or IIa/b peptic ulcer bleeding to receive either IV esomeprazole plus oral placebo or oral esomeprazole (60 mg every 12 hours) plus IV PPI placebo.⁴⁸ Recurrent bleeding within 30 days was reported in 7.7% in the IV esomeprazole group and 6.4% in the oral esomeprazole group. There was no difference in the requirement for blood transfusion, repeated endoscopic therapy and hospital stay between the two groups. It was noted that the study was stopped prematurely and was not designed as an equivalent trial. The trend suggests that the action of high-dose oral PPI peptic ulcer bleeding is comparable to that of IV PPI. In Taiwan, IV esomeprazole was compared with oral lansoprazole (30 mg four times a day for 3 days) in patients with peptic ulcer bleeding.⁴⁹ There was no difference in all the clinical outcome parameters, except that those who received oral PPI had a shorter hospital stay. In Korea, when IV esomeprazole was compared with oral rabeprazole (20 mg twice daily), the recurrent bleeding rate, surgical intervention and mortality between the two groups were comparable.⁵⁰

There is no properly powered RCT to confirm that high-dose oral PPI is as effective as IV PPI. The working group accepted that high-dose oral PPI can be used to prevent recurrent bleeding, but emphasized that it has to be used as an adjunct to endoscopic therapy. Only after endoscopic haemostasis is achieved, can high-dose oral PPI be recommended to prevent recurrent bleeding. The definition of high-dose oral PPI has been stated as at least 80 mg of esomeprazole (or

“You are a GutGPT, an expert in gastroenterology conversing with a medical expert regarding a patient. Your task is to answer a question based on the guidelines provided...”

“In your answer, summarize the evidence provided in the guidelines that supports your answer...”
{Retrieved Text}

“Use these to answer the following question: {query}

Format the reference according to the custom formatting strategy ...

Engineered Prompt



“According to the Asia-Pacific guidelines, the use of high-dose proton pump inhibitors (PPIs) may reduce the risk of recurrent bleeding...”

Ingest & Generate Response

Task 2: Risk Assessment



Machine Learning Model

- **Goal:** Predict whether patients require a hospital-based intervention or can be safely discharged to outpatient care
- (Composite) Outcome
 - 1: Intervention required (red blood cell transfusion, endoscopic or hemostatic intervention) or 30-day mortality
 - 0: No intervention required
- Metrics: AUC and TNR at a 99% sensitivity threshold
- Data: EHRs from patients suspected of acute GI bleeding in the Yale New Haven Health System (YNHH)
 - Training: St. Raphael, York St. campuses (n = 3,364)
 - Test: Random split from training (n = 1,127)
 - External Validation: Bridgeport campus (n = 1,197)



Challenge: EHR data is high-dimensional

- Code classes: ($p = 6,714$)
 - Demographics (age, sex)
 - Nursing assessment
 - Personal medical history (PHE, CCS, ICD codes)
 - Lab tests
 - Medications
- Embedding methods: principal components analysis, canonical correlation analysis, variational autoencoders, 2-layer neural net, LASSO
- Estimators: LASSO, XGBoost, random forests with honesty, 2-layer and 5-layer neural networks

$O(100\text{'s})$ of data-embedding-estimator combinations applied!

Machine Learning Model

Selected model: LASSO “embedding” with random forest with honesty

Still a black box model

Embed predictive model in a dashboard:

- 1) GutGPT chat interface
- 2) Interpretability metrics to guide model logic
- 3) Query model with hypothetical patient values

Our model

	AUC	TNR
Test	0.9134	0.3156
Ext. Valid.	0.9120	0.3316



GUT-GPT

What is this patient's risk for GI bleeding?

The patient's predicted risk for a hospital-based intervention is 40.0%. Of the 100 most similar patients in the training data set, 47 required a hospital based intervention. The American College of Gastroenterology (ACG) guidelines recommend a sensitivity threshold of 99%. Since the predicted risk is above 10.5% (the 99% sensitivity threshold), then the patient may require admission. Note this algorithm is unable to access any findings you may find on history and physical exam; furthermore, it can only use the diagnoses that are indicated via ICD codes already in the patient's chart. This algorithm also only takes the first recorded vital signs in the chart and does not incorporate any vital signs taken after the first values are measured and recorded.

What are the most important features that led to this prediction?

The important features for predicting the risk of gastrointestinal bleeding are age, sex, and pulse. These features are ranked in order of importance, with age being the most important, followed by sex, and then pulse. Age: The value of age for a particular patient is 60, which is lower than the average age of patients in the training data, which is 63.1. This means that the patient is relatively younger compared to the average patient in the training data. Younger age may suggest a lower risk of gastrointestinal bleeding. Sex: The value of sex for the patient is 0, which is the same as the average value in the training data. In this case, a value of 0 represents female, and a value of 1 would represent male. Since the value is the same as the average, it means that the patient is female, and gender does not contribute significantly to predicting the risk of gastrointestinal bleeding. Pulse: The value of pulse for the patient is 90, which is slightly lower than the average pulse of patients in the training data, which is 91.5. Pulse is a measure of heart rate. In this case, a

Ask questions about the model predictions, guidelines, or GI bleeding risk in general.

Submit

Use the dropdowns to select a plot type and what factors to examine.

Chart Type

PDP



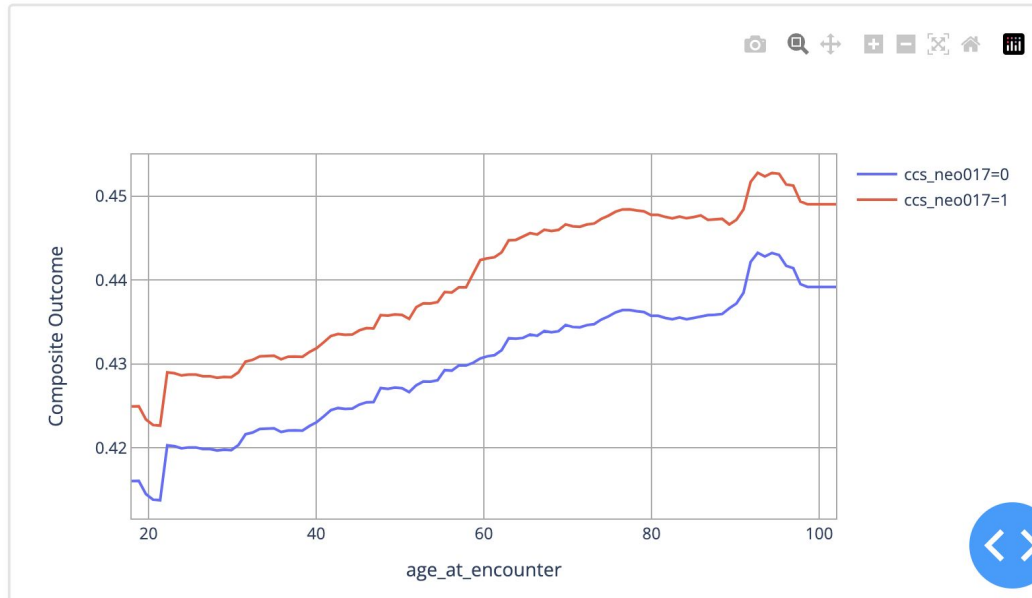
Factor 1

Age



Condition on Other Factors

Gastrointestinal cancers - liver



Risk assessment

Of the

100

most similar patients,

69

required a hospital based intervention

Overall risk of Needing Hospital-
Based Intervention:**0.642**

Adjust model inputs

Demographics ^

Age

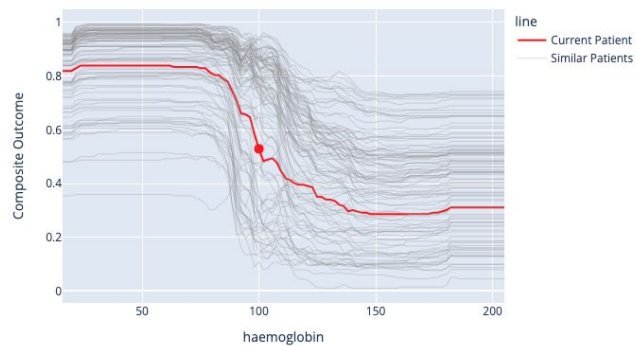


Sex

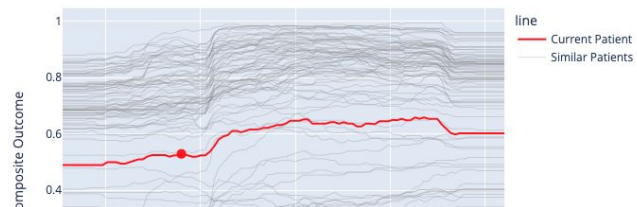
Female x vPhysical Exam vRelevant Laboratory Values v

Change in predicted risk with important features

Feature 1: haemoglobin

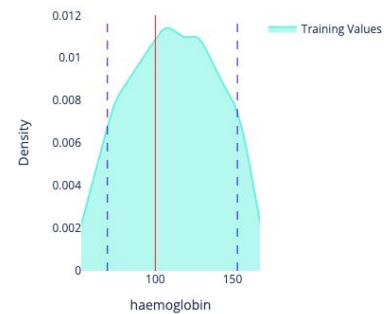


Feature 2: age

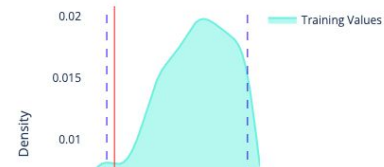


Feature Distributions

Feature 1: haemoglobin with value: 100



Feature 2: age with value: 36



Gut-GPT as a Team Member – Validation (Trust)

Simulation Center Study (Clinical Trial #NCT05816473)

- Design clinical scenarios to address current recommendations of the guidelines
- Population: Residents (emergency medicine and internal medicine), 3rd/4th year medical students
- Randomized controlled trial
 - Arm 1: Access to GutGPT + Dashboard Interface
 - Arm 2: Access to Dashboard Interface
- Outcomes:
 - Trust in Usability Attitude (UTAUT)
 - Qualitative Usability Study



Gut-GPT as a Team Member – Validation (Accuracy)

Comparison with World Experts

- Design prompts (questions, clinical scenarios) that address current guideline recommendations (e.g., risk stratification, transfusion thresholds, endoscopy timing, hemostatic intervention, etc.) based on the Issuing Medical Society (ACG, European, Asia-Pacific)
- Ask the prompts to world experts and to GutGPT and record responses



- Similarity index: compare answers provided by GutGPT and by world experts

Thank you!

Colleen Chan
colleen.chan@yale.edu

Yale SCHOOL OF MEDICINE



GUT-GPT