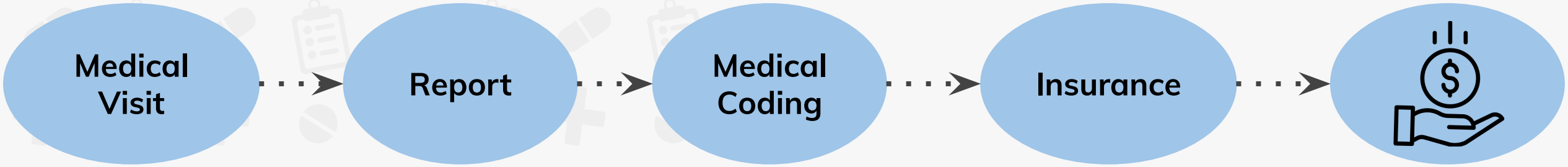
# Smart Compose for Medical Applications

Data-X Project by:
Fred Hertlein, Sophie Pealat, David Scanlan, Olivia Zhang

Partnership with:
Abboud Chaballout - Diagnoss

# Context



# Example Medical Report

Example Medical Codes Shown To The Right:

CC = Complications

PMH = Past Medical History

CAD = Coronary Artery Disease

HTN = Hypertension / High Blood Pressure

LVH = Left Ventricular Hypertrophy

SCH = Subchronic Haematoma

CC: 80 year old woman with chest pain

History of present illness: The patient does have a complaint of chest pain. It was the morning of. It was not worse with walking. No radiation. It was mild. She sat and it went away. BP was 140/90. There is no complaint of shortness of breath with exertion or at rest. There is no complaint of palpitations, dizziness, or lightheadedness. The patient has not passed out. The patient does not have orthopnea, paroxysmal nocturnal dyspnea or increased lower extremity edema..

PMH:
1.CAD, Chest pain: coronary calcium score = 35. Ct coronary angiogram on 7-24-17 showed mild non obstructive CAD with only minimal mixed plaque in prox LAD. Other vessels without significant disease.
2.Dyspnea, mildly dilated pulmonary artery at 32 mm.
3.HTN: moderate LVH, grade 2/4 diastolic dysfunction. LVEF = 65% on echo but 42% on stress test.
4.Meningioma: brain
5.Obesity
6.Paroxysmal atrial fibrillation: this was noted in 2000.
7.Hypothyroidism
8.Prolonged qt (mild, qtc 480 ms on 8-2-17).
9.Subendocardial fat deposition in RV apex and septum (LV and RV) seen on CT coronary angiogram 7-24-17 thought due to inflammation or ischemia. No other findings of ARVD on echo. No MRI.
10.Asthma.

Family history: Daughter had a heart arrhythmia. No MI or pacemaker. Son is ok. Brother has HTN. No SCD..

Social history: She doesn't smoke. She drinks rare alcohol. No drugs. She drinks 1-2 cups of caffeinated beverages today..

# Context

**30%** medical coding errors

**$210B** lost or delayed revenue

# Impact:

**“Doctors spending twice as much time filling out paperwork as visiting with patients”**

Develop better technologies to automate the paperwork process + reduce medical coding errors

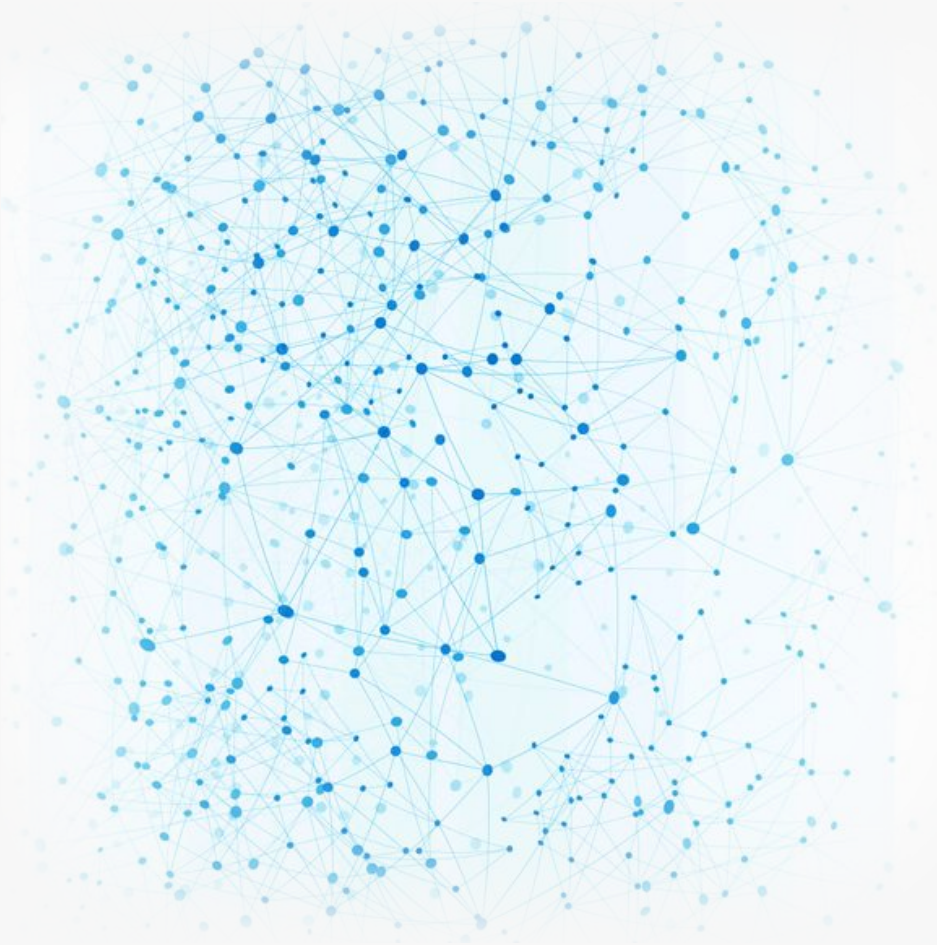
Save time for doctors and optimize hospitals’ efficiency in the current Covid-19 situation

*J-Marc Overhage, Annals of Internal Medicine*

# About Diagnoss:

- Company is using ML to drive a predictive healthcare analytics engine
  - Assigns medical coding +
  - Acts as a “medical version of Grammarly”

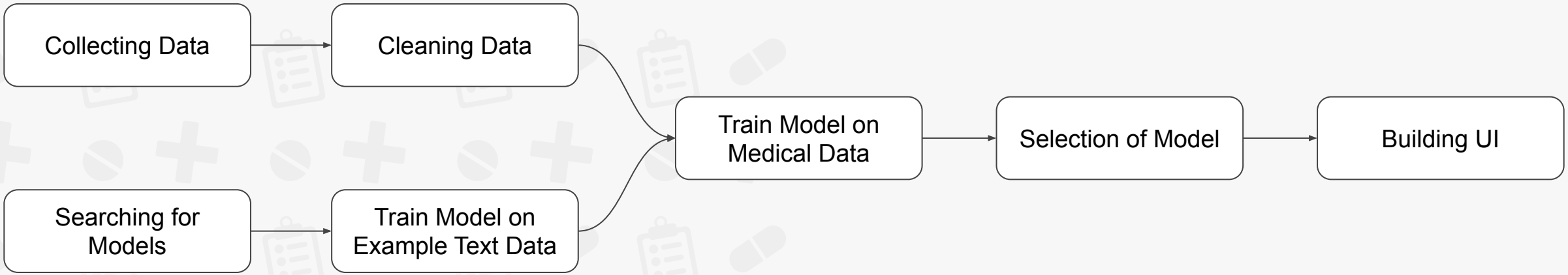
- The first application of Diagnoss is to use AI to reduce errors in medical coding



# Our Project:

**Build a model that autocompletes doctors' input of examination reports for insurance billing purposes**

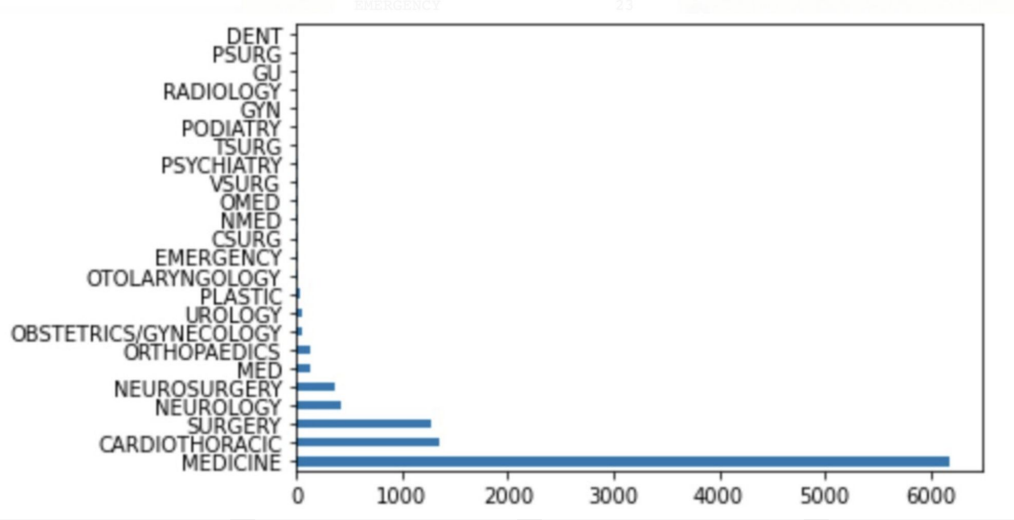
# Approach


Collecting Data
Cleaning Data
Searching for Models
Train Model on Example Text Data
Train Model on Medical Data
Selection of Model
Building UI

# Data Processing

- Access to MIT Database : **2 million** medical reports
  - MIMIC-III Clinical Database

- Selection of Categories:
  - Medicine
  - Cardiothoracic
  - Surgery
  - Neurology/Neurosurgery




* Based on 20,000 row training set

**Admission Date:** [**2172-3-5**] **Discharge Date:** [**2172-3-8**]
**Date of Birth:** [**2109-10-8**] **Sex:** F
**Service:** NEUROSURGERY

**Allergies:**
No Known Allergies / Adverse Drug Reactions

**Attending:**[**First Name3 (LF) 1835**]
**Chief Complaint:**
Meningioma

**Major Surgical or Invasive Procedure:**
Right Craniotomy

**History of Present Illness:**
[**Known firstname 622**] [**Known lastname 1836**] is a 62-year-old woman, with longstanding history of rheumatoid arthritis, probable Sweet's syndrome, and multiple joint complications requiring orthopedic interventions.
She was found to hve a right cavernous sinus and nasopharyngeal mass. She underwent a biopsy of hte nasopharyngeal mass by Dr. [**First Name4 (NamePattern1) **] [**Last Name (NamePattern1) 1837**] and the pathology, including flow cytometry, was reactive for T-cell lymphoid hyperplasia only.

**Past Medical History:**
She has a history of rheumatoid arthritis unspecified dermatosis, right knee replacement, left hip replacement, and fusion of subtalar joint, and resection of a benign left parotid gland tumor.

**Social History:**
She is married. She had smoked for approximately a year and a half when she was younger, but is not currently smoking. She has approximately one glass of wine per week. She is retired but was employed as a teacher.

**Family History:**
Cancer, diabetes, hearing loss, and heart disease.

# Data Processing cont.

**Brief Hospital Course:**
Patient presented electively for meningioma resection of [**3-5**].
She tolerated the procedure well and was extubated in the
operating room. She was trasnported to the ICU post-operatively
for management. She had no complications and was transferred to
the floor and observed for 24 hours. Prelim path is consistent
with meningioma.
She has dissolvable sutures, and will need to come to
neurosurgery clinic in [**6-28**] days for wound check only.
She will need to be scheduled for brain tumor clinic. She will
complete Decadron taper on [**3-10**] and then restart her maintenance
dose of prednisone. She will also be taking Keppra for seizure
prophlyaxis.
Her neurologic examination was intact with no deficits at
discharge. She was tolerating regular diet. She should continue
to take over the counter laxatives as needed.

# Model Information: GPT-2

- Developed by Open.ai
- Open source model used to generate coherent paragraphs of text
- Trained on 40GB of Internet text
- Part of Transformers libraries

**Our goal :** Use transfer learning to train this model on a medical dataset

# Model Performance: GPT-2

- Models trained from scratch on medical report dataset :
- GPU
- Large memory
- Long time (1h - 5h)

- Main Parameters :
- Block Size : 50
- Number of training epochs : 1 or 2
- Save steps : 1000 or 3000
- Number of batch eval size : 1

| Models | Perplexity |
|---|---|
| Model Medicine | 14.57 |
| Model Cardiothoracic | 19.50 |
| Model Surgery | 26.48 |
| Model Neurology | 24.12 |
| Model “Raw data” (7000 complete reports) | 5.8 |

# Examples of predictions : prompt “The patient felt”

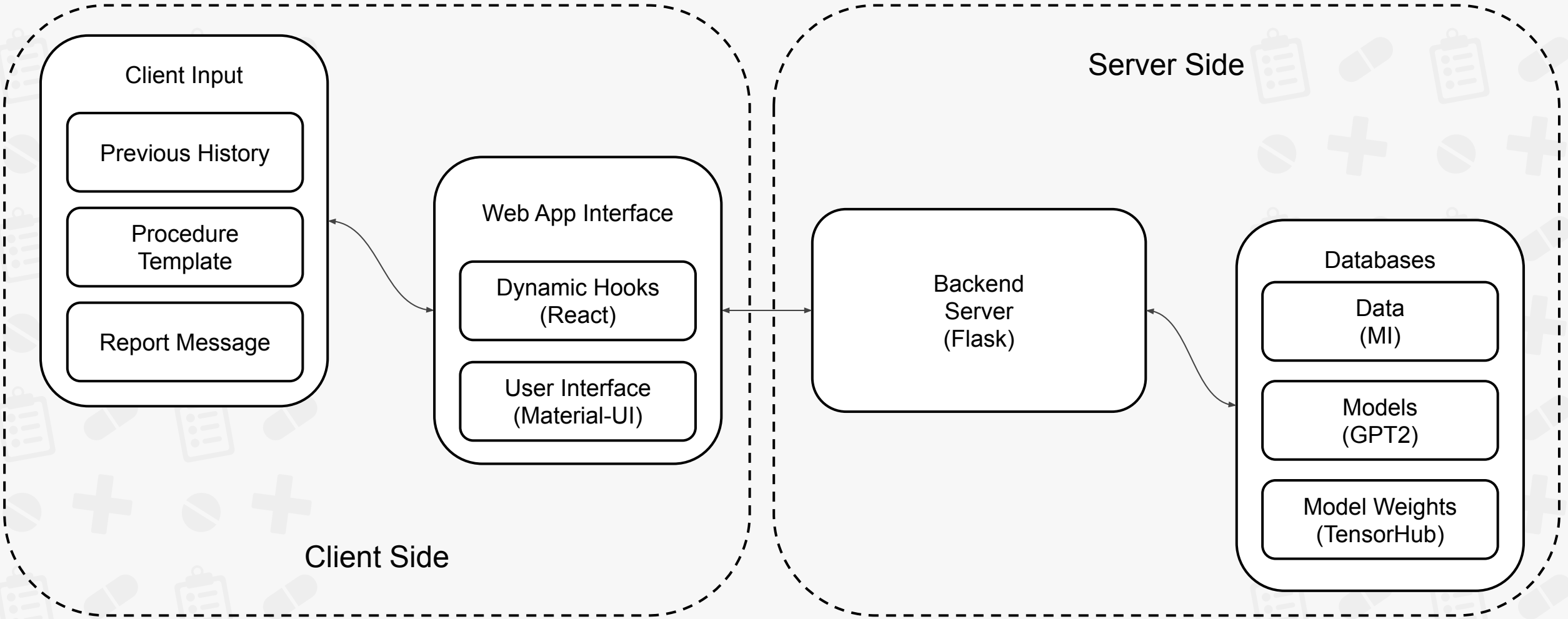
**Parameters for prediction**

- Temperature : 0.7
- Length : 10
- Top_K = 50
- Prompt
- Number of return predictions

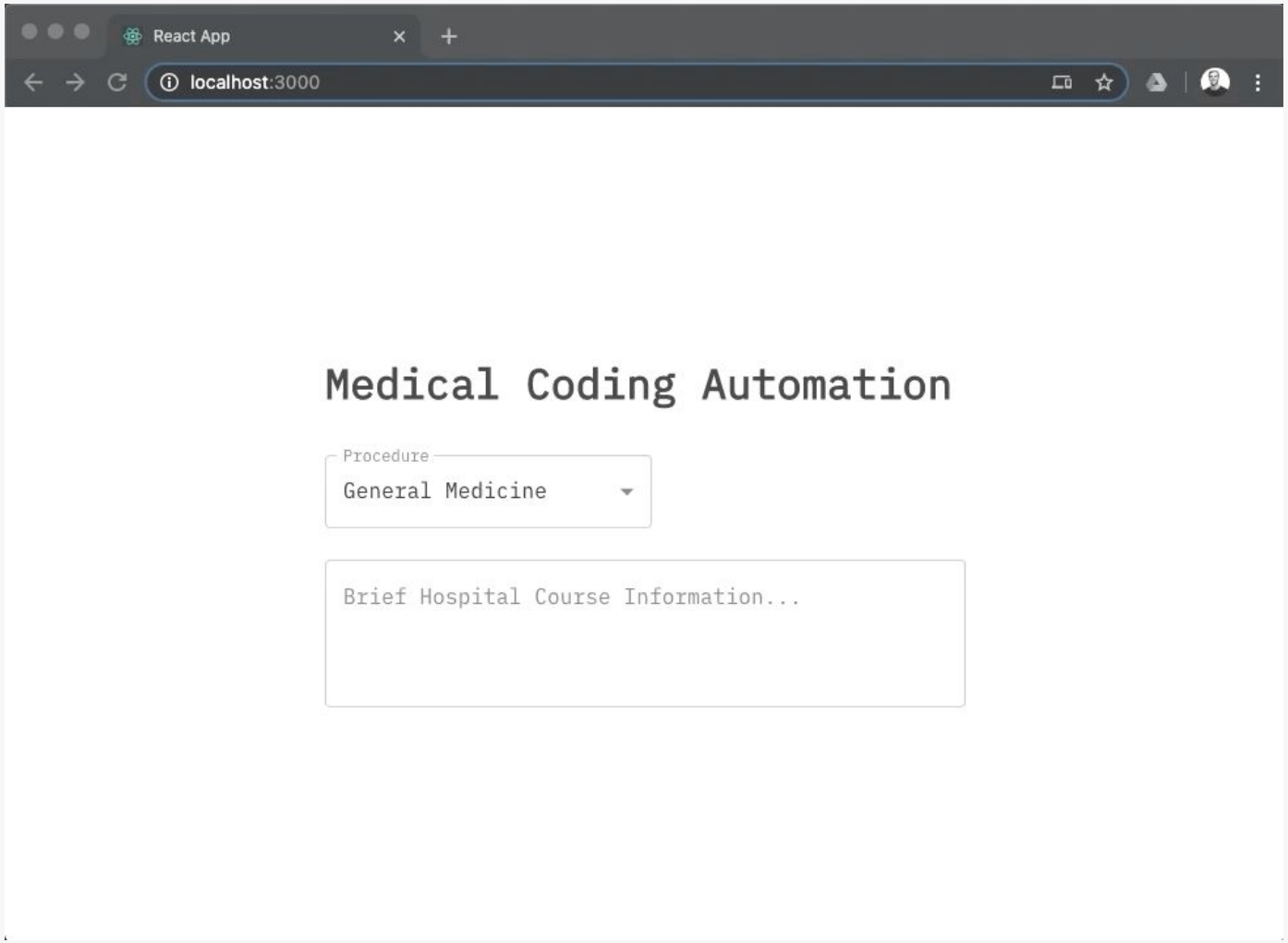
**Time :** 8s-12s

| Models | Examples of Predictions |
|---|---|
| Model Medicine | The patient felt well after extubation.  He was started |
| Model Cardiothoracic | The patient felt more stable from a cardiac standpoint.  He was |
| Model Surgery | The patient felt that she had been left with a low grade of |
| Model Neurology | The patient felt that she had been given a good dose of PR |
| Model “Raw data” (7000 complete reports) | The patient felt well until she had a respiratory arrest |

# System Architecture:



# UI Demo:




Github Repository: https://github.com/gfhertlein/medical_smart_compose

# UI Demo:


Medical Coding Automation
Procedure
Neurology
Brief Hospital Course Information...

# Limitations and Further work

1 - **Char Ngram Models via GPT2 Transfer Learning**
2 - **Faster Server or Static Model Predictions**
3 - **UI Simplicity + Formatting**
4 - **Training model on Diagnoss Data**

# Thank You !

# Q&A

# Github Repository

https://github.com/gfhertlein/medical_smart_compose