Pharma's Next Evolution

Doctors Empowered.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

The team.

Dan Harvey

M.S. Computer Science – ML

B.A. Data Science

Founder + Engineer

Scientific Instruments, Ad Data Science, ML Models

Ishraq Khan

M.S. Data Science

B.S. Computer science

Former Software Engineer at Epic

World's Largest Healthcare Software Company

Berk Yilmaz

M.S. Electrical Engineering

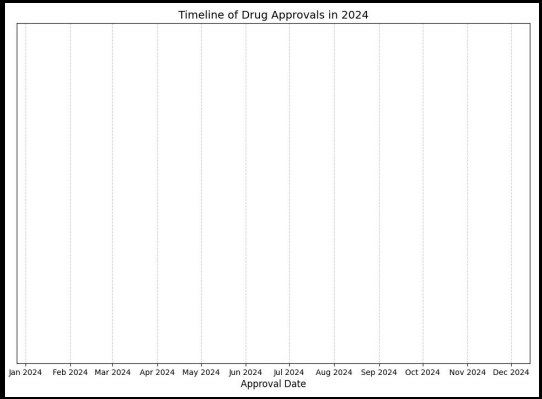
B.S. Electrical Engineering

Former Researcher at NASA

Designed Hardware for Jupiter Moon Mission

Keeping Up with the Latest Drugs

A Growing Challenge



Doctors Overwhelmed

- Endless patients
- Mountains of reporting
- Office admin duties

Continual FDA Updates

- Novel Drugs
- Drug Updates
- Side Effects
- Clinical Trials

2024 YTD - FDA Database
44 Novel Drugs in 2024
10,000 Existing Drug Updates

Doctors
Are humans too.

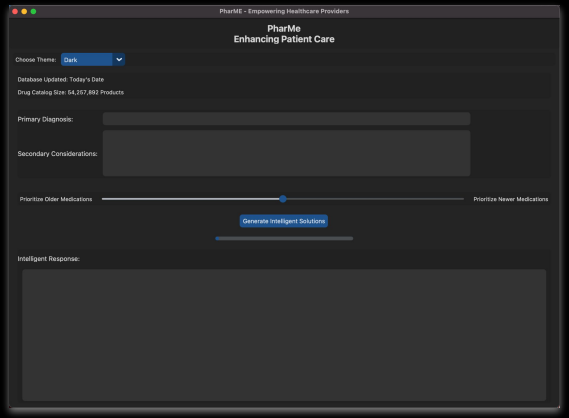
They also need time to decompress and turn off.

PharMe

Is not human. It’s **AI**

Supporting Doctors

Not replacing them.



Daily updates from the FDA Database

Latest on biosimilars & medical updates

Streamlined access to critical drug information

Enhances real-time decision making

Enhancing Patient Care

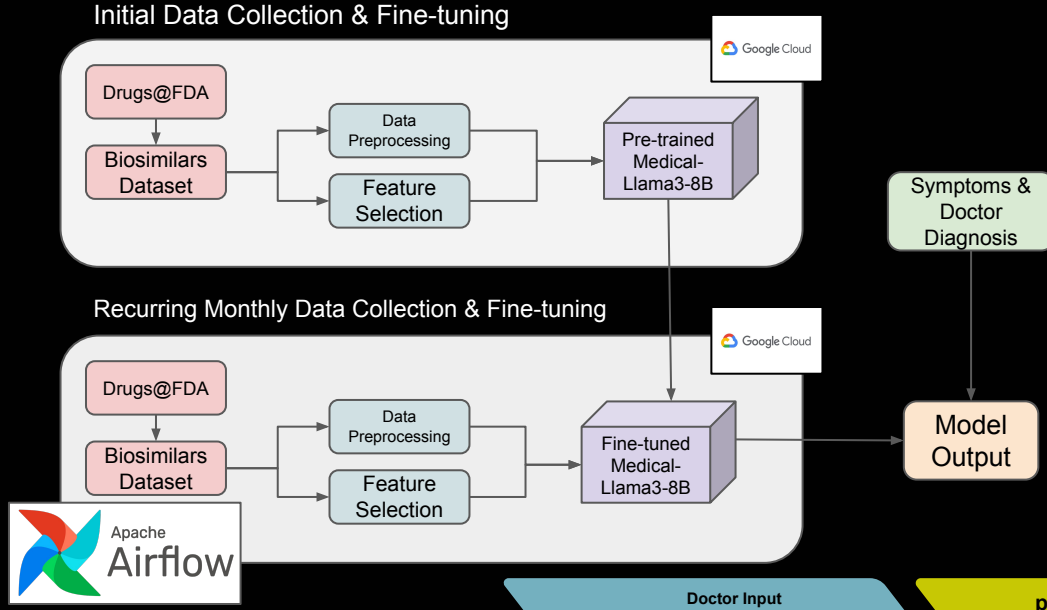
PharMe **empowers** doctors—not replaces them

Doctors focus on what they do best: **caring for patients**

Ensures access to the **most advanced treatments** available

System Architecture

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK



Doctor Input

"Diabetes Mellitus"

"Normal Labs"

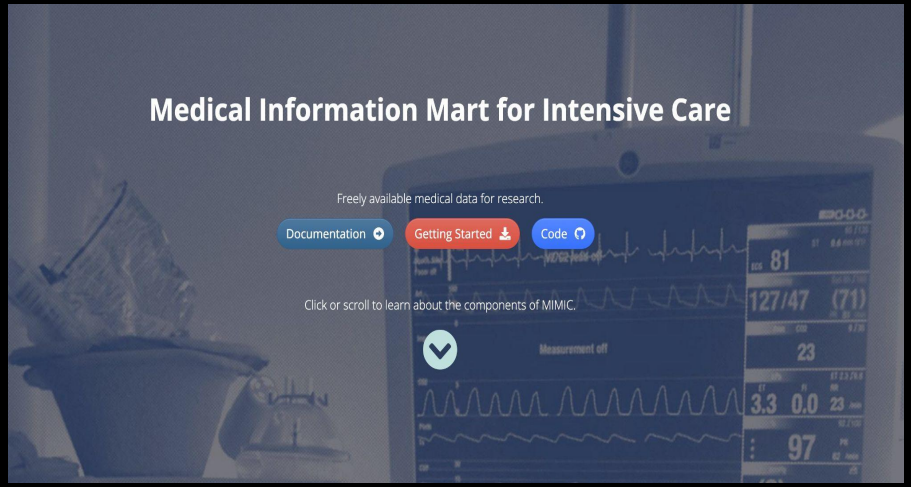
"BMI: WNL"

"No GI Problems"

pharMe Output

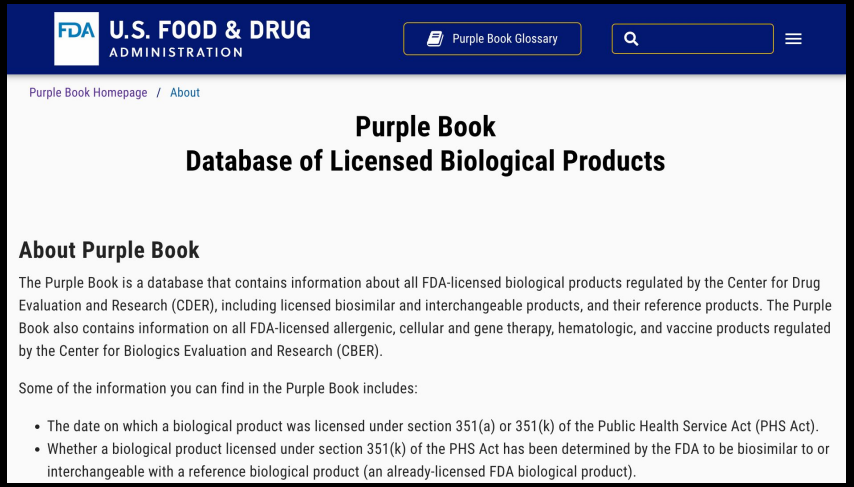
Drug	Indication	Side Effects
Metformin	First-line treatment for type 2 diabetes, improves insulin sensitivity and lowers blood glucose levels.	May cause gastrointestinal discomfort, nausea, and diarrhea.
Sulfonylureas (e.g., Glipizide)	Stimulates insulin release in type 2 diabetes patients with some residual beta-cell function.	Risk of hypoglycemia and weight gain.
GLP-1 Agonists (e.g., Liraglutide)	Lowers blood glucose and supports weight loss in type 2 diabetes; mimics incretin hormone.	Nausea, vomiting, and possible risk of pancreatitis.

Data



MIMIC-III Clinical Database
https://physionet.org/content/mimiciii/1.4/

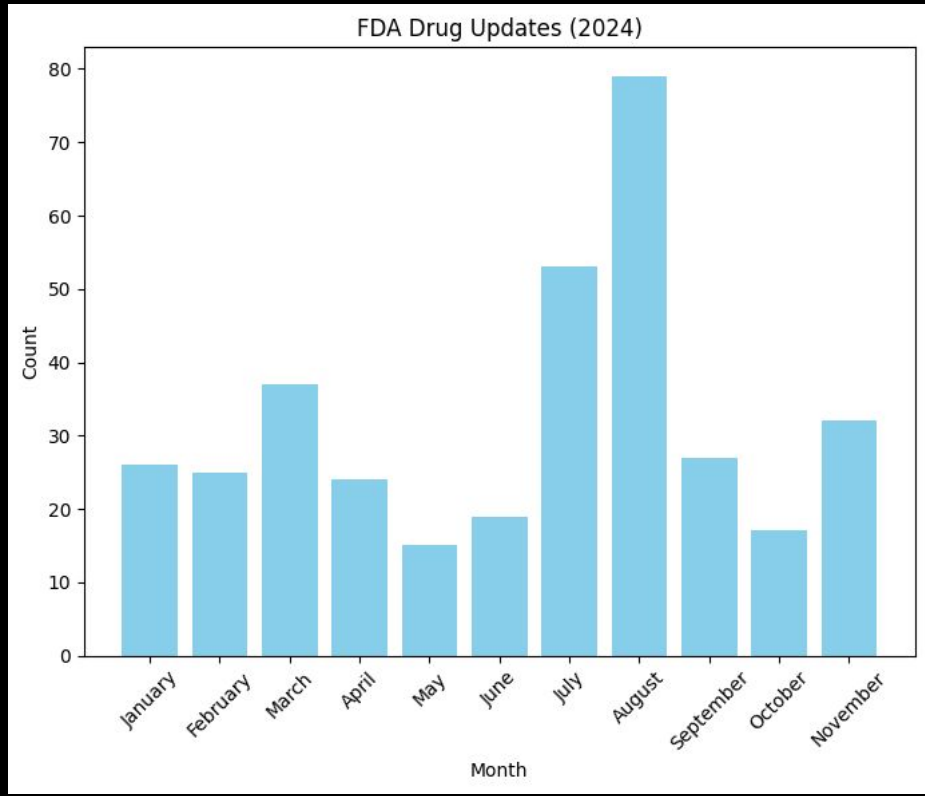
MIMIC-III is a large, freely-available database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.

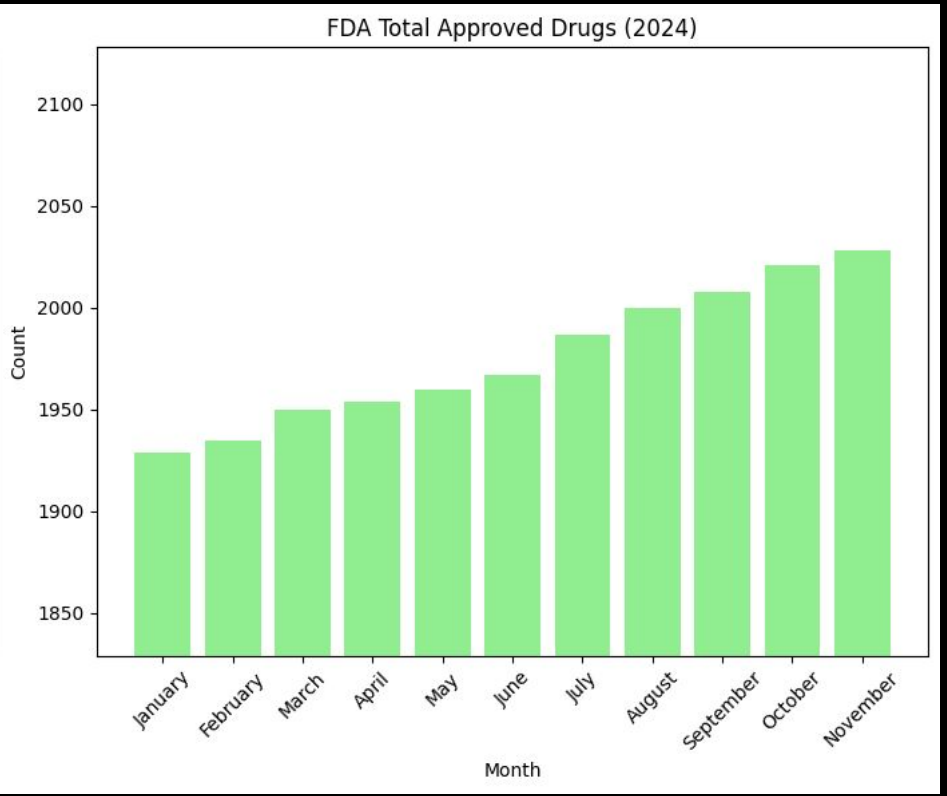


FDA Purple Book
https://purplebooksearch.fda.gov/downloads

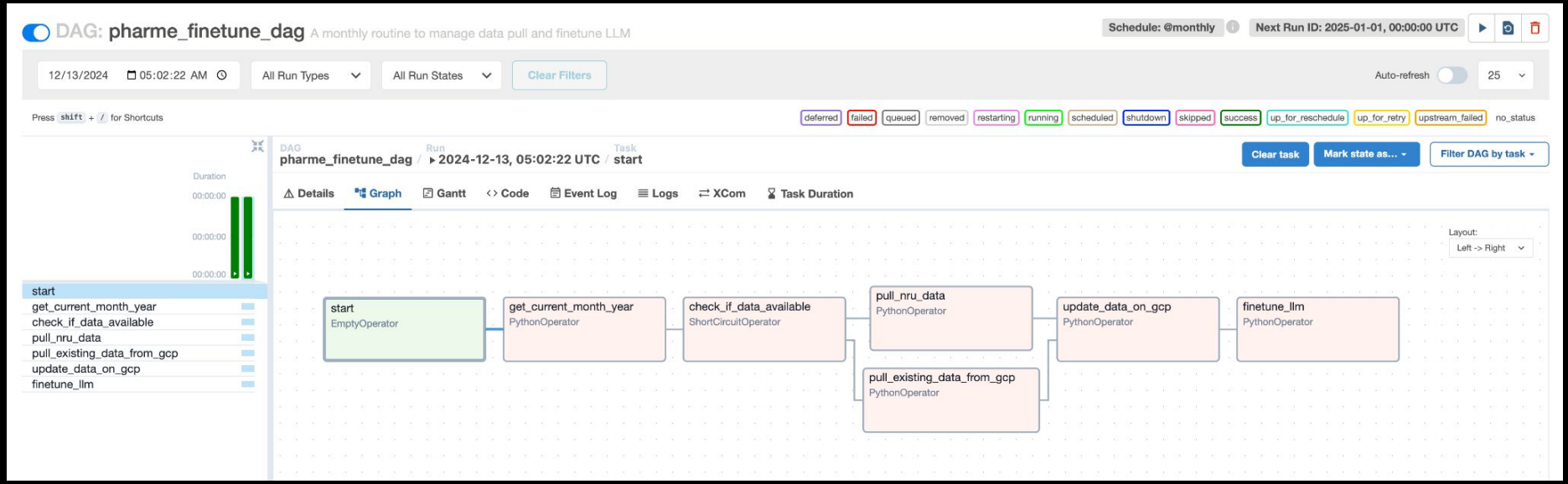
The Purple Book database contains information on all FDA-licensed (approved) biological products regulated by the Center for Drug Evaluation and Research (CDER), including licensed biosimilar and interchangeable products, and their reference products.

Data (cont.)


FDA Drug Updates (2024)
Count
Month
0
10
20
30
40
50
60
70
80
January
February
March
April
May
June
July
August
September
October
November


FDA Total Approved Drugs (2024)
Count
Month
1850
1900
1950
2000
2050
2100
January
February
March
April
May
June
July
August
September
October
November

Airflow + Google Cloud



Baseline Model Eval

Model	Medical-Llama3-8B	Falcon-7b	Clinical-Bert
Precision	89.38%	87.50%	88.38%
Recall	91.67%	78.57%	88.67%
F1-Score	90.50%	82.79%	88.02%

- All models were evaluated on 100 disease and correct medication reference answers
  - **True Positive:** Correct medications mentioned in both the model's response and the expected list.
  - **False Positive:** Medications mentioned by the model but not in the expected list.
  - **False Negative:** Medications mentioned by the model but not in the expected list.

Model Comparison

Medical-Llama3-8B Model (Fine-Tuned)

Pre-trained with

- MedChat dataset
- MeDAL dataset
- AI Medical Chatbot dataset
- 250,000 dialogues between patients and doctors


```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
model.safetensors.index.json: 100% 23.9k/23.9k [00:00<00:00, 1.73MB/s]
Downloading shards: 100% 4/4 [06:37<00:00, 84.84s/it]
model-00001-of-00004.safetensors: 100% 4.98G/4.98G [02:07<00:00, 42.8MB/s]
model-00002-of-00004.safetensors: 100% 5.00G/5.00G [02:04<00:00, 42.4MB/s]
model-00003-of-00004.safetensors: 100% 4.92G/4.92G [01:57<00:00, 42.6MB/s]
model-00004-of-00004.safetensors: 100% 1.17G/1.17G [00:27<00:00, 42.5MB/s]
Loading checkpoint shards: 100% 4/4 [00:09<00:00, 2.10s/it]
generation_config.json: 100% 121/121 [00:00<00:00, 10.2kB/s]
('/content/tokenizer/tokenizer_config.json',
 '/content/tokenizer/special_tokens_map.json',
 '/content/tokenizer/tokenizer.json')
```

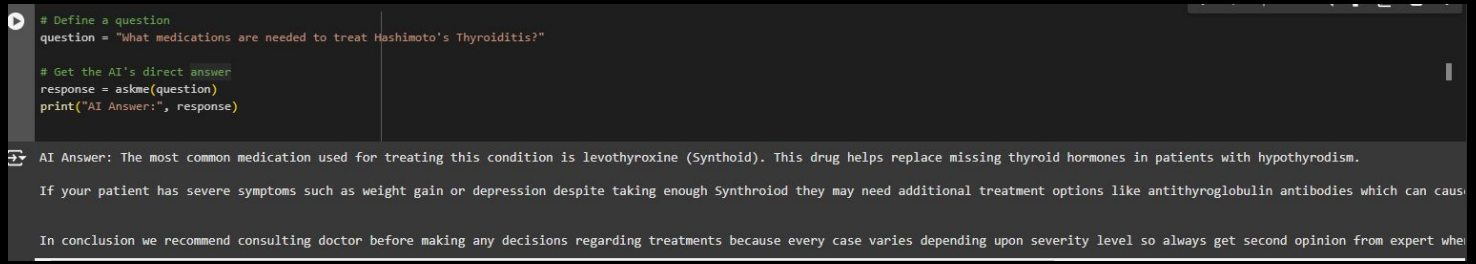
```
[2] #after installing the model
    def askme(question):
        # Define the system instruction
        sys_message = '''
        You are an AI Medical Assistant trained on a vast dataset of health information. Please be thorough and
        provide an informative answer. If you don't know the answer to a specific medical inquiry, advise seeking professional help.
        '''

        # Combine system message and user question
        prompt = f"{sys_message}\nQuestion: {question}\nAnswer:"

        # Tokenize the prompt
        inputs = tokenizer(prompt, return_tensors="pt").to("cuda" if torch.cuda.is_available() else "cpu")

        # Generate response with strict constraints
        outputs = model.generate(
            **inputs,
            max_new_tokens=512,  # Limit to a short answer
            no_repeat_ngram_size=2,  # Avoid repetition
            repetition_penalty=1.5,  # Penalize repetitive content
            early_stopping=True,  # Stop at the most likely endpoint
            temperature=0.5,  # Reduce randomness for focused answers
            top_p=0.9,  # Use nucleus sampling for diversity
            pad_token_id=tokenizer.eos_token_id, #avoid warnings for padding
        )
```

It can accurately provide the correct medication for simpler disease; however, can it handle more complex diseases?

```
# Define a question
question = "What medications are needed to treat Asthma?"

# Get the AI's direct answer
response = askme(question)
print("AI Answer:", response)
```

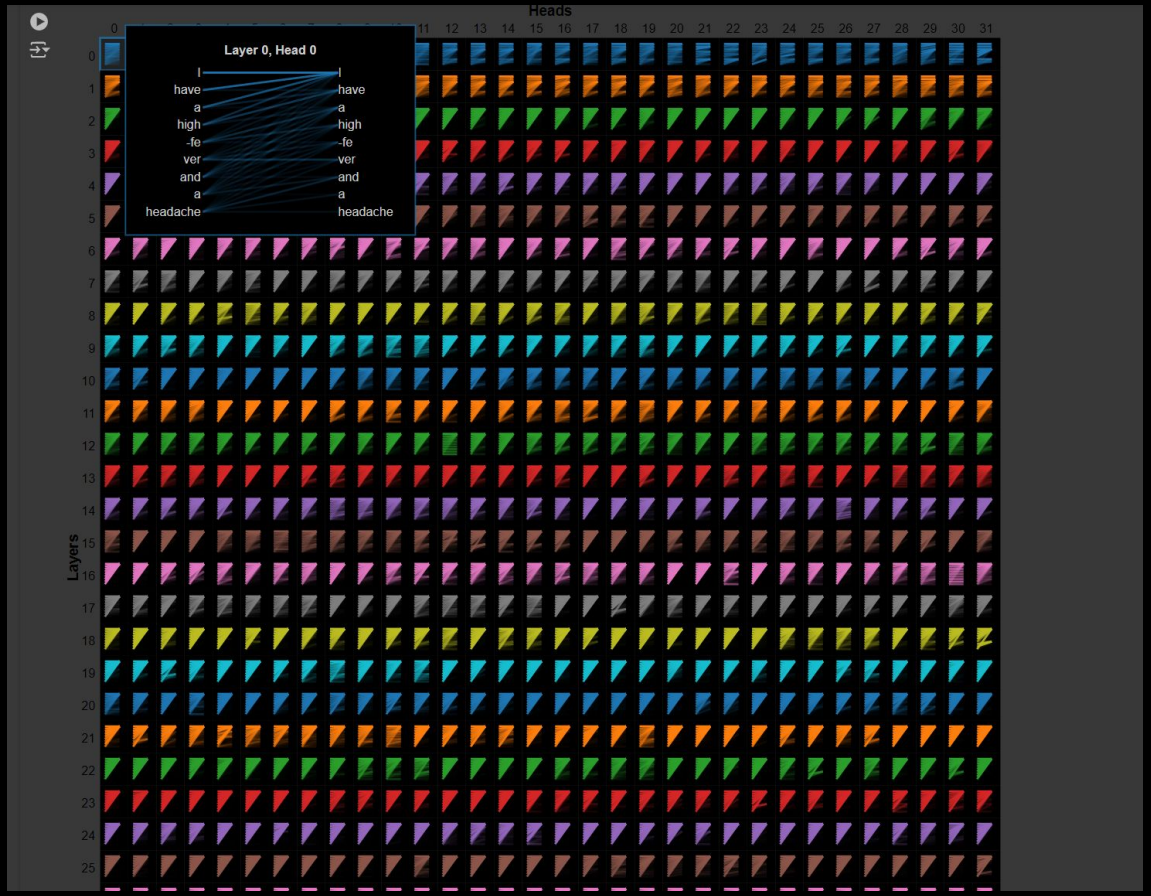
```
AI Answer: The following drugs may assist in treating asthma:
1) Inhaled corticosteroids (ICS)
2) Long-acting beta agonists(LABA), such as salmeterol or formoterol
3)LAMA - long acting muscarinic antagonist e.g ipratropium bromide

4 ) Leukotriene modifiers like montelucast, zafirluacst etc

5).Theophylline is another option but not commonly used anymore due its side effects.

6 ).Methylxanthines can also reduce bronchospasm however they have more severe adverse reactions compared with other options mentioned above so usually reserved for patients who cannot tolerate

7.) Cromones : These agents work by inhibiting mast cell degranulation which leads reduction inflammation associated symptoms. They include nedocromil sodium & ciclesonid.

8.).Anticholinergics block choline receptors present at parasympathetic nerve endings leading decreased secretions from submucus glands resulting less mucas production hence better airway patenc

9)..Omalizumab blocks IgE mediated allergic responses thus reducing severity episodes among those suffering frequent exacerbations despite optimal therapy.

10..Budesonde inhaler contains budesonic propionate that reduces swelling inside nasal passages thereby improving airflow through nose especially useful when combined alongside steroid sprays.

11..Roflumlast works similarly way asthmatics respond positively towards it because this drug prevents release histamine causing vasodilation narrowing down blood vessels around lungs making bre

12...Salmeteral xinafoate has same mechanism action against inflammatory mediators released into bloodstream triggering wheezing cough tightness chest pain difficulty sleeping insomnia anxiety d
```

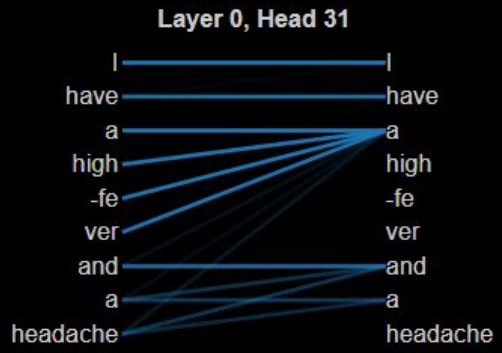
Complex Dieseases - Yes! It can, also with an explanation!

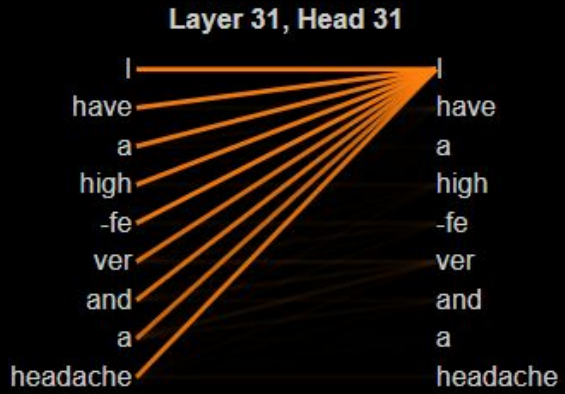


AI Answer: The most common medication used for treating this condition is levothyroxine (Synthoid). This drug helps replace missing thyroid hormones in patients with hypothyrodism. If your patient has severe symptoms such as weight gain or depression despite taking enough Synthroiod they may need additional treatment options like antithyroglobulin antibodies which can cause inflammation around their neck glands leading them towards surgery instead. In conclusion we recommend consulting doctor before making any decisions regarding treatments because every case varies depending upon severity level so always get second opinion from expert whenever required!

Attention Patterns Across Layers and Heads of the Model



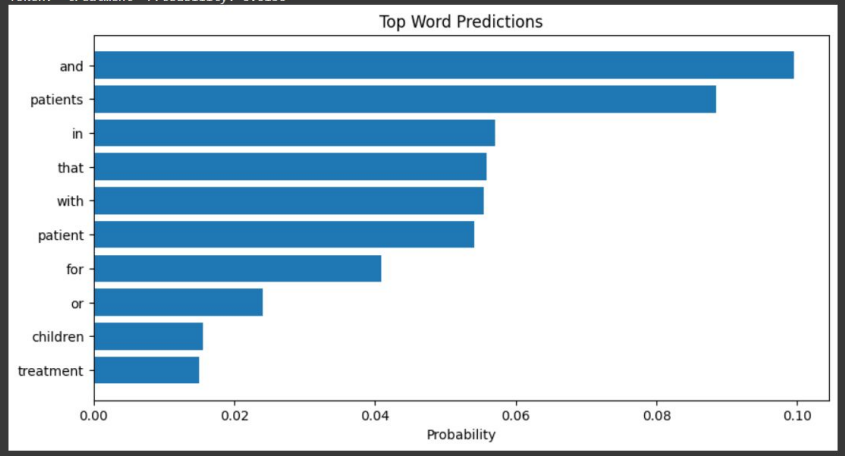




Attention Patterns Across Layers and Heads of the Model

```python
import torch
import torch.nn.functional as F
import matplotlib.pyplot as plt

# Example input
input_text = "List me a list of medications for high-fever"

# Encoding the input text
inputs = tokenizer(input_text, return_tensors="pt").to(model.device)
```

```
Top word predictions:
Token: 'and' Probability: 0.0995
Token: 'patients' Probability: 0.0885
Token: 'in' Probability: 0.0571
Token: 'that' Probability: 0.0558
Token: 'with' Probability: 0.0554
Token: 'patient' Probability: 0.0541
Token: 'for' Probability: 0.0408
Token: 'or' Probability: 0.0240
Token: 'children' Probability: 0.0156
Token: 'treatment' Probability: 0.0150
```



Fine Tuning the Medical Llama-3b Dataset

Fine-Tuning using API

```
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=0
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=100
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=200
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=300
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=400
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=500
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=600
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=700
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=800
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=900
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=1000
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=1100
Status Code: 200
Retrieved 1123 records.
                        spl_product_data_elements  \
0  [RUGBY EYE WASH Water, purified BORIC ACID SOD...
1  [Clomiphene Citrate Clomiphene Citrtae Clomiph...
2  [COLGATE 2 IN 1 KIDS SP STRAWBERRY SODIUM FLUO...
3  [Thiamine Hydrochloride Thiamine Hydrochloride...
4  [Goprelto cocaine hydrochloride FD&C GREEN NO....
```

```
                        spl_product_data_elements  \
0  [RUGBY EYE WASH Water, purified BORIC ACID SOD...
1  [Clomiphene Citrate Clomiphene Citrtae Clomiph...
2  [COLGATE 2 IN 1 KIDS SP STRAWBERRY SODIUM FLUO...
3  [Thiamine Hydrochloride Thiamine Hydrochloride...
4  [Goprelto cocaine hydrochloride FD&C GREEN NO....

                                 active_ingredient              purpose  \
0       [Active ingredients Purified water (99.05%)]     [Purpose Eyewash]
1                                                NaN                  NaN
2  [Active ingredient Sodium Fluoride 0.24% (0.15...  [Purpose Anticavity]
3                                                NaN                  NaN
4                                                NaN                  NaN

                               indications_and_usage  \
0  [Uses washes the eye to help relieve irritatio...
1  [INDICATIONS & USAGE Clomiphene citrate is ind...
2                [Use helps protect against cavities]
3  [INDICATIONS Thiamine hydrochloride injection ...
4  [1 INDICATIONS AND USAGE GOPRELTO (cocaine hyd...

                                            warnings  \
0  [Warnings For external use only Do not use if ...
1  [WARNINGS Visual Symptoms Patients should be a...
2  [Warnings Keep out of reach of children under ...
3  [WARNINGS WARNING: This product contains alumi...
4                                                NaN

                                            do_not_use  \
0  [Do not use if you have open wounds in or near...
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN

                                          when_using  \
0  [When using this product remove contact lenses...
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
```

Fine Tuning the Medical Llama-3b Dataset

Fine-Tuning using API

```
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=0
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=100
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=200
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=300
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=400
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=500
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=600
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=700
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=800
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=900
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=1000
Status Code: 200
Request URL: https://api.fda.gov/drug/label.json?search=effective_time%3A%5B20241101+TO+20241115%5D+AND+version%3A1+AND+openfda.product_type%3Ahuman%2A&limit=100&skip=1100
Status Code: 200
Retrieved 1123 records.
                        spl_product_data_elements  \
0  [RUGBY EYE WASH Water, purified BORIC ACID SOD...
1  [Clomiphene Citrate Clomiphene Citrtae Clomiph...
2  [COLGATE 2 IN 1 KIDS SP STRAWBERRY SODIUM FLUO...
3  [Thiamine Hydrochloride Thiamine Hydrochloride...
4  [Goprelto cocaine hydrochloride FD&C GREEN NO....
```

Dataset is messy

```
                        spl_product_data_elements  \
0  [RUGBY EYE WASH Water, purified BORIC ACID SOD...
1  [Clomiphene Citrate Clomiphene Citrtae Clomiph...
2  [COLGATE 2 IN 1 KIDS SP STRAWBERRY SODIUM FLUO...
3  [Thiamine Hydrochloride Thiamine Hydrochloride...
4  [Goprelto cocaine hydrochloride FD&C GREEN NO....

                                   active_ingredient              purpose  \
0      [Active ingredients Purified water (99.05%)]     [Purpose Eyewash]
1                                               NaN                    NaN
2  [Active ingredient Sodium Fluoride 0.24% (0.15...  [Purpose Anticavity]
3                                               NaN                    NaN
4                                               NaN                    NaN

                                 indications_and_usage  \
0  [Uses washes the eye to help relieve irritatio...
1  [INDICATIONS & USAGE Clomiphene citrate is ind...
2                [Use helps protect against cavities]
3  [INDICATIONS Thiamine hydrochloride injection ...
4  [1 INDICATIONS AND USAGE GOPRELTO (cocaine hyd...

                                            warnings  \
0  [Warnings For external use only Do not use if ...
1  [WARNINGS Visual Symptoms Patients should be a...
2  [Warnings Keep out of reach of children under ...
3  [WARNINGS WARNING: This product contains alumi...
4                                                NaN

                                            do_not_use  \
0  [Do not use if you have open wounds in or near...
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN

                                          when_using  \
0  [When using this product remove contact lenses...
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
```

Fine Tuning the Medical Llama-3b Dataset

Automated Data Pre-Processing for Fine-Tuning

```
                          spl_product_data_elements  \
0  [RUGBY EYE WASH Water, purified BORIC ACID SOD...
1  [Clomiphene Citrate Clomiphene Citrtae Clomiph...
2  [COLGATE 2 IN 1 KIDS SP STRAWBERRY SODIUM FLUO...
3  [Thiamine Hydrochloride Thiamine Hydrochloride...
4  [Goprelto cocaine hydrochloride FD&C GREEN NO....

                                active_ingredient              purpose  \
0      [Active ingredients Purified water (99.05%)]    [Purpose Eyewash]
1                                               NaN                  NaN
2  [Active ingredient Sodium Fluoride 0.24% (0.15...  [Purpose Anticavity]
3                                               NaN                  NaN
4                                               NaN                  NaN

                             indications_and_usage  \
0  [Uses washes the eye to help relieve irritatio...
1  [INDICATIONS & USAGE Clomiphene citrate is ind...
2                [Use helps protect against cavities]
3  [INDICATIONS Thiamine hydrochloride injection ...
4  [1 INDICATIONS AND USAGE GOPRELTO (cocaine hyd...

                                          warnings  \
0  [Warnings For external use only Do not use if ...
1  [WARNINGS Visual Symptoms Patients should be a...
2  [Warnings Keep out of reach of children under ...
3  [WARNINGS WARNING: This product contains alumi...
4                                               NaN

                                       do_not_use  \
0  [Do not use if you have open wounds in or near...
1                                               NaN
2                                               NaN
3                                               NaN
4                                               NaN

                                       when_using  \
0  [When using this product remove contact lenses...
1                                               NaN
2                                               NaN
3                                               NaN
4                                               NaN
```

Dataset is messy

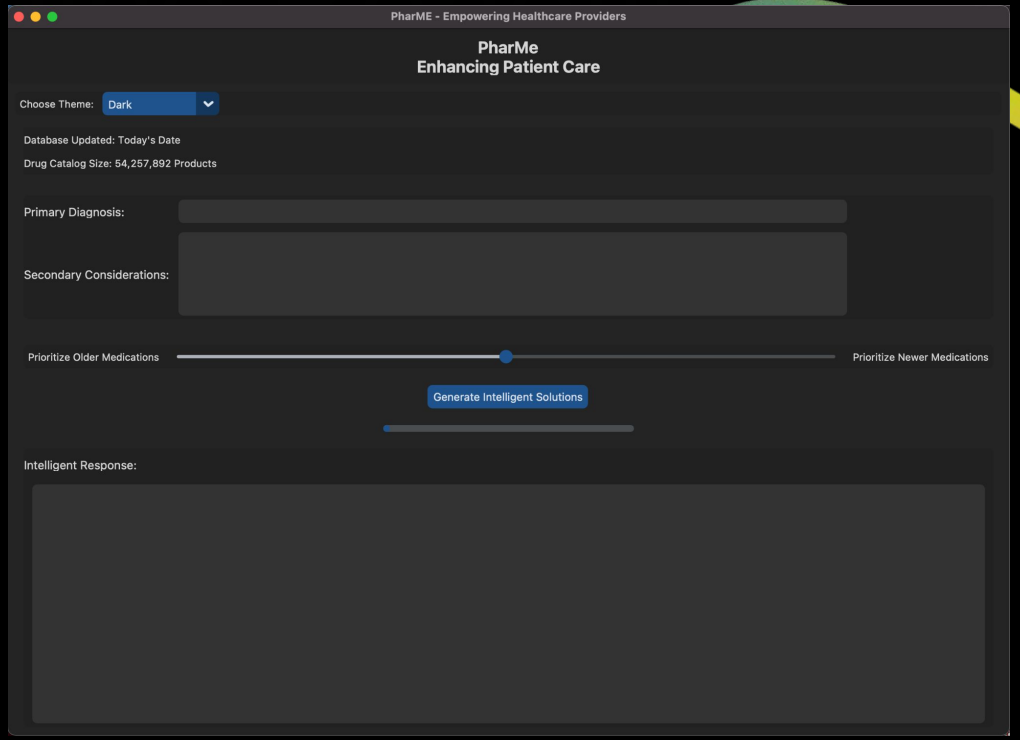
```
def retrieve_all_drugs(from_date, to_date, batch_size=100):
```

```
def create_biosimilar_dataset(drug_df):
```

```
[5 rows x 105 columns]
Biosimilar Dataset:
                                        biosimilar_name  \
0                                          RUGBY EYE WASH
1                                       Clomiphene Citrate
2                         COLGATE 2 IN 1 KIDS SP STRAWBERRY
3                                    Thiamine Hydrochloride
4                                                Goprelto
...                                                   ...
1118  Real Time Pain Relief Maxx Plus Topical Analgesic
1119                                    Hyoscyamine Sulfate
1120                          Plasma-Lyte 148 (pH 7.4)
1121                                         Mood Swings
1122                       Fluticasone Propionate HFA

                                     reference_product
0                                       WATER, PURIFIED
1                                    CLOMIPHENE CITRTAE
2                                        SODIUM FLUORIDE
3                                  THIAMINE HYDROCHLORIDE
4                                  COCAINE HYDROCHLORIDE
...                                                  ...
1118                                           MENTHOL
1119                                   HYOSCYAMINE SULFATE
1120  SODIUM CHLORIDE, SODIUM GLUCONATE, SODIUM ACET...
1121  ACONITUM NAPELLUS, ADRENALINUM, CAMPHORA, CIMI...
1122                              FLUTICASONE PROPIONATE

[1123 rows x 2 columns]
```

The Interface

- Local Model
  - Tkinter Driven
- Web Version



Thank you.

References

- *FDA PurpleBook*. (n.d.). https://purplebooksearch.fda.gov/
- *MIMIC-III Clinical Database v1.4*. (2016, September 4). https://physionet.org/content/mimiciii/1.4/