Argon AI Code Screen

Exercise Time: <2 hours

Problem Overview

Sarah, an executive at Pharma Co., is in charge of a drug called *Luminarex,* an immunotherapy which has not yet been approved by the FDA. Sarah's drug is currently in phase 3 of Clinical Trials, specifically targeting a disease called *Non Small Cell Lung Cancer (NSCLC)*, and Sarah hopes that her drug will be approved soon.

Sarah wants to develop a better understanding of her competitive landscape. Specifically, she wants to know about other pharma companies with clinical trials in the NSCLC space.

She currently uses ClinicalTrials.gov to do her research, but is frustrated by how difficult it is to find the clinical trials she is looking for.

Task

Your task is to build an application that can help Sarah do research on her competitive space.

Using this dataset, your task is to build a part of a web application that allows Sarah to search through the dataset and return a list of corresponding clinical trials.

Your focus should be one of the following:

1. **[Backend/AI Engineers]** Search quality and thoroughness/accuracy of results. Get creative with the search algorithm. We want to strive for near perfect recall, even if it's at the expense of precision.
2. **[Frontend Engineers]** Presentation of the results in table format. The ability to manipulate, slice, and dice the data. Clean use of components, states, hooks, etc.
3. **[Fullstack Engineers]** Some combination of the previous two.

You are allowed to use any framework you want that gets the job done, but bonus points for using our stack (React, NextJS, Python). Feel free to load the dataset directly into memory so that you don't have to set up a DB. You do not need to deploy your code. We value clean code and clean project structure.

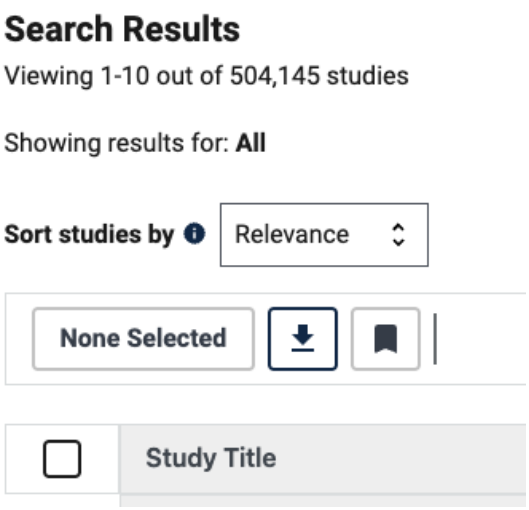
Once completed, email over your code in a .zip format or share your github repo with recruiting@argon-ai.com.

Bonus points:

- *NSCLC has many different representations in the dataset. For example, it could be "non small cell lung cancer", "non small cell lung carcinoma", "NSCLC", "carcinoma of the lungs, non small cell", etc. How do we capture all the relevant clinical trials for searches on any disease?*
- *How do we allow her to search for NSCLC trials -AND- immunotherapy related drugs?*
- *How would you deploy your software?*

- *What are the alternatives to loading the dataset into memory, and why would you want to use those alternatives?*
- *How do we evaluate completeness of results?*

Dataset

1. Go to https://clinicaltrials.gov/search
2. Click on the download icon (in between “None Selected” and “Bookmark”)
    a. 

3. Select either CSV or JSON, whichever your choice
4. Select all data fields
5. Download the **top 10,000** trials by scrolling to the bottom and clicking “Download”