# Building a Diet and Wellness AI Agent: A Comprehensive Guide

## Introduction

In today's fast-paced world, maintaining a healthy lifestyle through proper diet and wellness is more important than ever. With an overwhelming amount of information available, it can be challenging to find accurate and personalized advice. This guide aims to build an AI agent that can answer queries related to fitness, health, calories, and more, leveraging reasoning capabilities to provide accurate and contextual responses.



## Problem Statement

Responding to health-related questions often requires more than just retrieving information; it necessitates reasoning and contextual understanding. For example, determining whether a particular diet is suitable for someone with specific health conditions involves evaluating multiple factors. Therefore, an AI agent with robust reasoning capabilities is essential to provide meaningful and accurate advice in the domain of diet and wellness.

## Code and Guide

### Code

https://github.com/heathbrew/Building-a-Diet-and-Wellness-AI-Agent-A-Comprehensive-Guide

Git clone this repo [1] and follow along for the set-up.

I have given the requrements.txt to setup this “or else Problems”.

### Dataset

For this guide, we will use the PUBHEALTH dataset. This comprehensive dataset is designed for explainable automated fact-checking of public health claims, making it ideal for training our AI agent. Each instance in the dataset includes a veracity label (true, false, unproven, mixture) and an explanation text field that justifies the claim's veracity label [2].

### Dataset Summary

**Dataset Card for PUBHEALTH**:

* **Dataset Summary**: PUBHEALTH is a comprehensive dataset for explainable automated fact-checking of public health claims. Each instance has a veracity label and an explanation text field.
* **Languages**: The text in the dataset is in English.

### Dataset Creation & Curation Rationale

The dataset was created to explore fact-checking of difficult to verify claims i.e., those which require expertise from outside of the journalistics domain, in this case biomedical and public health expertise.

It was also created in response to the lack of fact-checking datasets which provide gold standard natural language explanations for verdicts/labels.

<https://huggingface.co/datasets/health_fact>

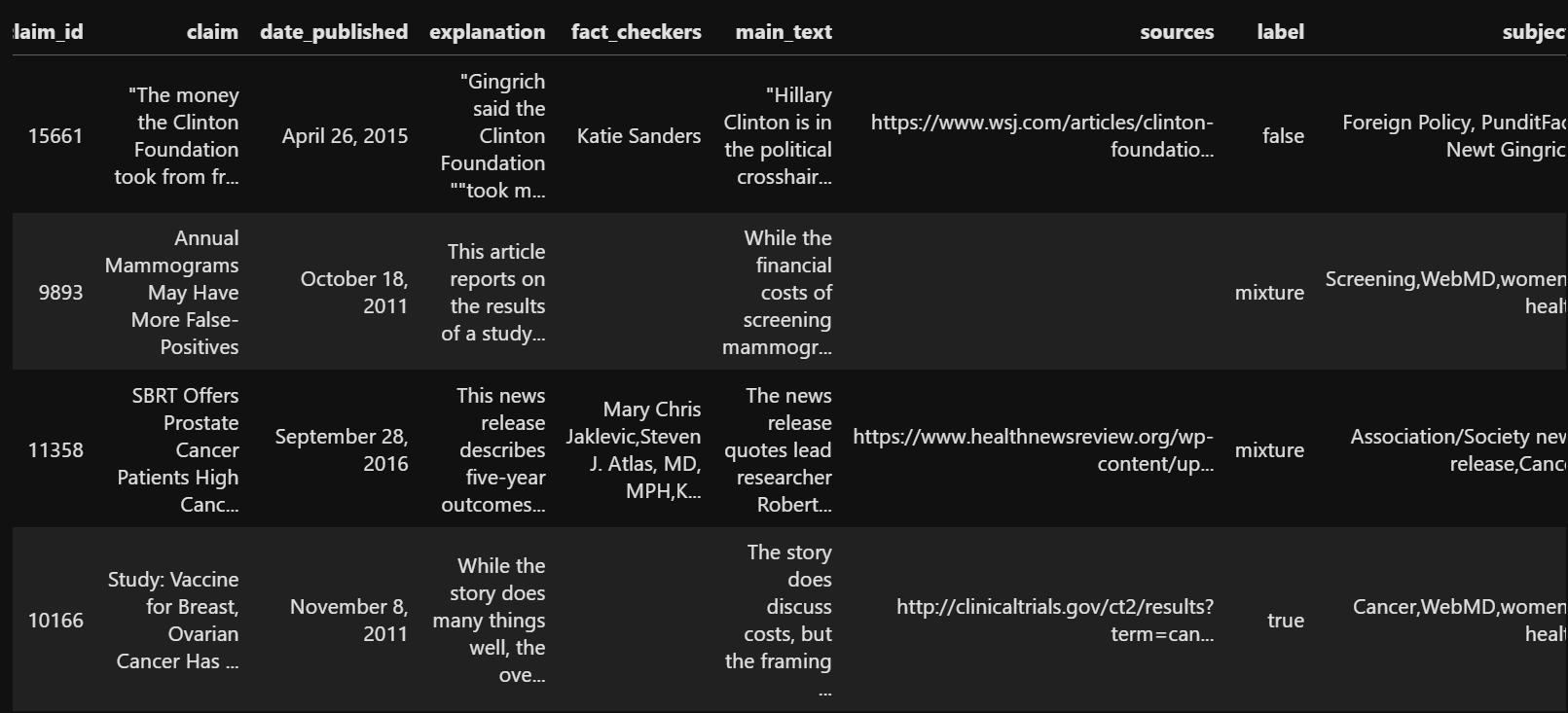
import pandas as pd

# Read the TSV file into a pandas DataFrame

df = pd.read\_csv("Dataset/train.tsv", sep='\t')

# Now you can work with the DataFrame as usual

df.head()



As we are using react agent so will consider claim and explanation from the dataset

# Fill NaN values with empty strings

df['claim'] = df['claim'].fillna('')

df['explanation'] = df['explanation'].fillna('')

df['merged\_data'] = df.apply(lambda row: row['claim'] + ' ' + row['explanation'], axis=1)

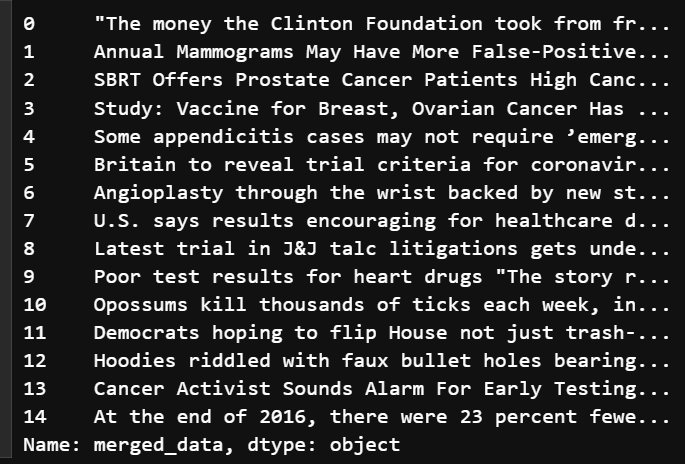
# Create dfcopy with 15 rows

dfcopy = df.head(15).copy()

# Display the first few rows

dfcopy.head()

dfcopy['merged\_data']



## Loading the Embedding Model

I will be using the MINI-LM12-V2 [3] to embed the Health Dataset.

from sentence\_transformers import SentenceTransformer

this code will download and save the model in a local directory   
from transformers import AutoTokenizer, AutoModel

from pathlib import Path

def download\_model\_and\_tokenizer(model\_name, save\_path):

    """

    Download and save both the model and the tokenizer to the specified directory.

    Parameters:

        model\_name (str): Name of the model to download.

        save\_path (str or Path): Path to the directory where the model and tokenizer will be saved.

    """

    # Create the save path if it doesn't exist

    save\_path = Path(save\_path)

    save\_path.mkdir(parents=True, exist\_ok=True)

    # Initialize tokenizer and model

    tokenizer = AutoTokenizer.from\_pretrained(model\_name)

    model = AutoModel.from\_pretrained(model\_name)

    # Save tokenizer

    tokenizer.save\_pretrained(save\_path)

    # Save model

    model.save\_pretrained(save\_path)

# Example usage

model\_name = 'sentence-transformers/all-MiniLM-L12-v2'  # Model name to download

save\_path = Path("MiniLM-L12-v2/")  # Path where model and tokenizer will be saved

download\_model\_and\_tokenizer(model\_name, save\_path)

this code will load the model and tokenizer from the local directory

from transformers import AutoTokenizer, AutoModel

def load\_model\_and\_tokenizer(model\_path):

    """

    Load the model and tokenizer from the specified directory.

    Parameters:

        model\_path (str or Path): Path to the directory containing the saved model and tokenizer.

    Returns:

        tokenizer (transformers.PreTrainedTokenizer): Loaded tokenizer.

        model (transformers.PreTrainedModel): Loaded model.

    """

    model\_path = Path(model\_path)

    tokenizer = AutoTokenizer.from\_pretrained(model\_path)

    model = AutoModel.from\_pretrained(model\_path)

    return tokenizer, model

# Load the model and tokenizer

model\_path = Path("MiniLM-L12-v2/")

tokenizer, model = load\_model\_and\_tokenizer(model\_path)

these are the functions I will use use to use the embedding model and tokenizer

import torch

#Mean Pooling - Take attention mask into account for correct averaging

def mean\_pooling(model\_output, attention\_mask):

    token\_embeddings = model\_output[0] #First element of model\_output contains all token embeddings

    input\_mask\_expanded = attention\_mask.unsqueeze(-1).expand(token\_embeddings.size()).float()

    return torch.sum(token\_embeddings \* input\_mask\_expanded, 1) / torch.clamp(input\_mask\_expanded.sum(1), min=1e-9)

def generate\_embedding(text):

    # Tokenize input text

    encoded\_input = tokenizer(text, padding=True, truncation=True, return\_tensors='pt')

    # Compute token embeddings with model

    with torch.no\_grad():

        model\_output = model(\*\*encoded\_input)

    # Perform mean pooling

    sentence\_embedding = mean\_pooling(model\_output, encoded\_input['attention\_mask'])

    # Convert to numpy for FAISS compatibility and ensure it's 2D

    return sentence\_embedding.cpu().numpy().reshape(1, -1)

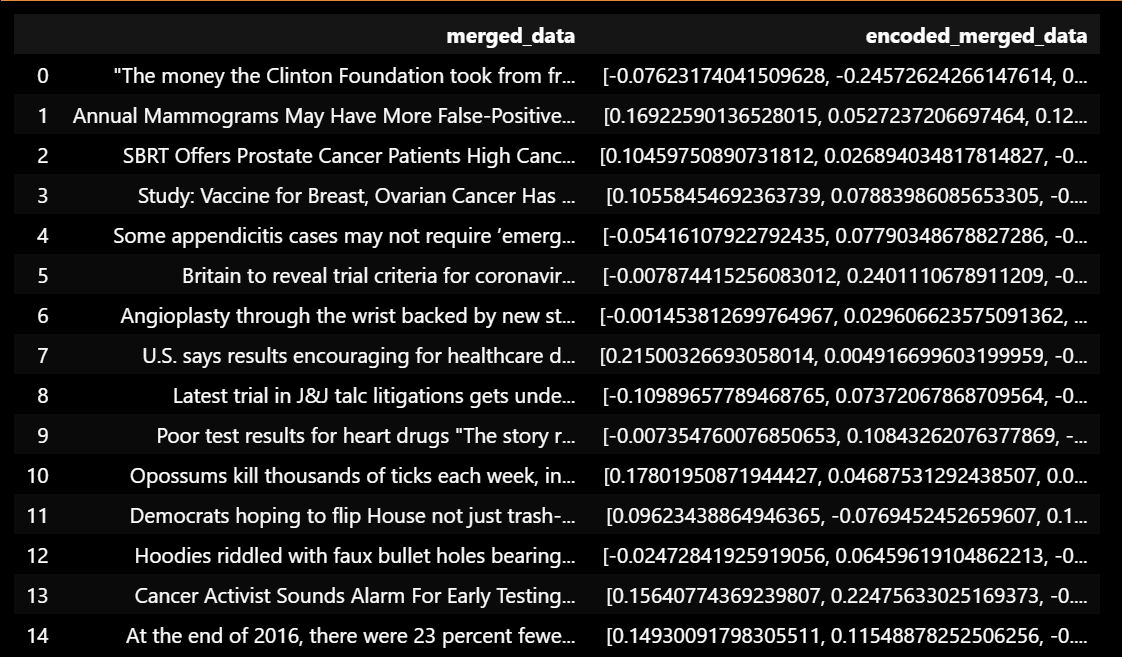
apply these embedding function on the dataset

# Generate embeddings for the 'input' column

dfcopy['encoded\_merged\_data'] = dfcopy['merged\_data'].apply(lambda x: generate\_embedding(x)[0].tolist())

dfcopy[['merged\_data', 'encoded\_merged\_data']]

Lets take a look at the Numbers



Lets take a look at the complete dataset



## Creating a Qdrant Database

In the last project [4], I used [Qdrant](https://qdrant.tech/) which is a locally supported vector store for RAG.

<https://medium.com/@AyushmanPranav/anomaly-detection-using-vector-search-c3f48f8e61eb>

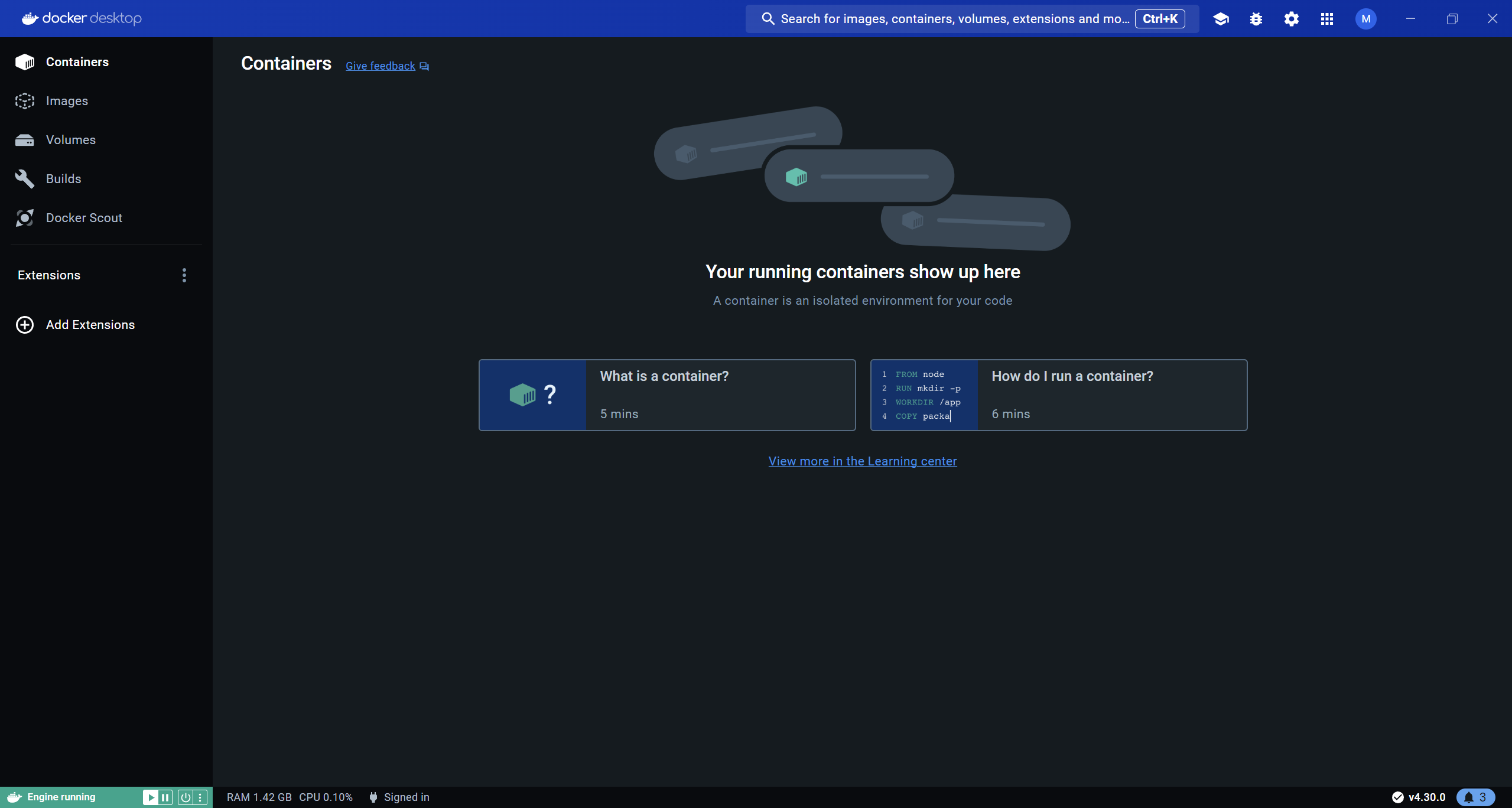
For this project, I will once again use Qdrant, but this time for financial data, to create a vector store.

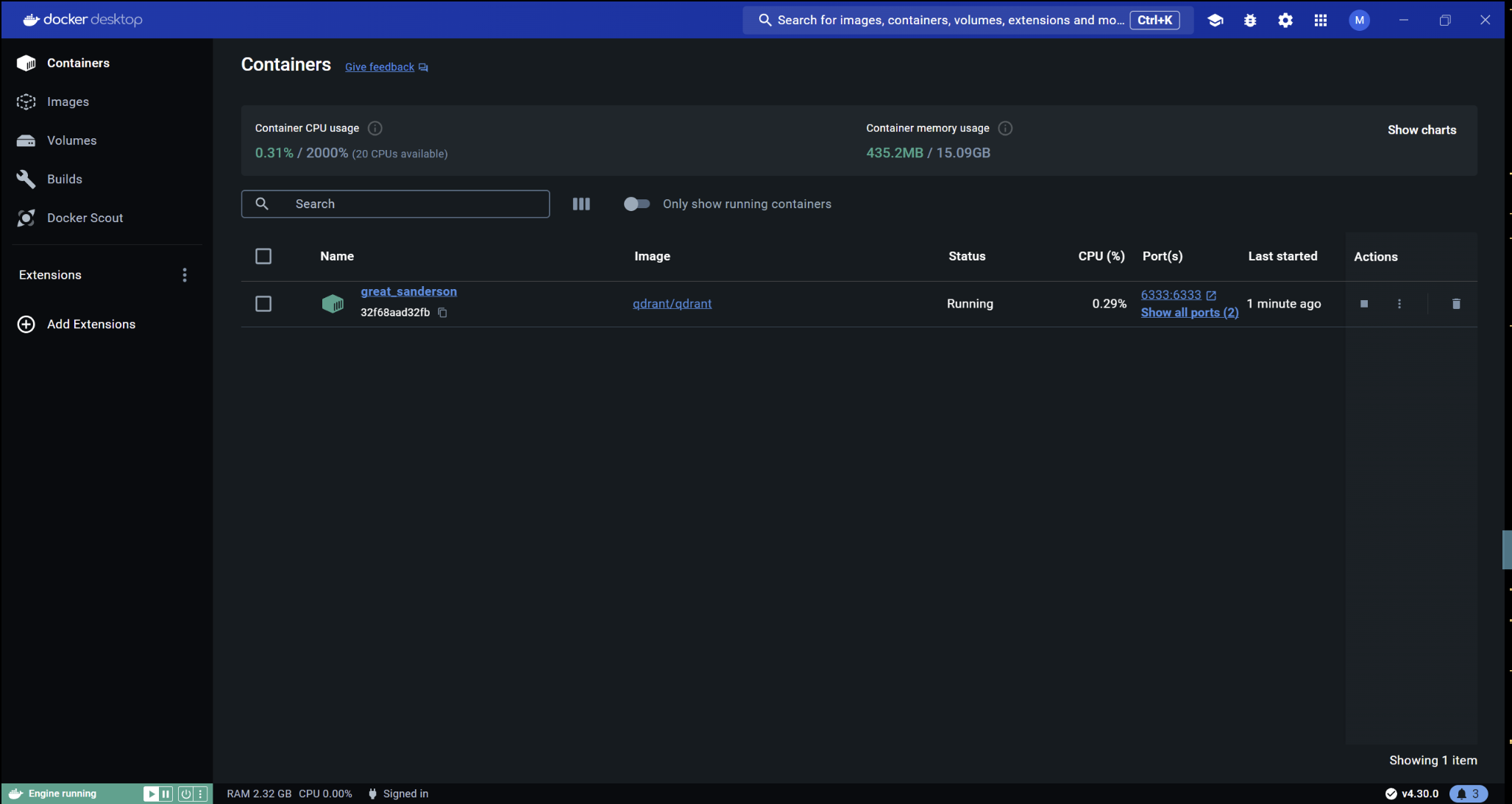
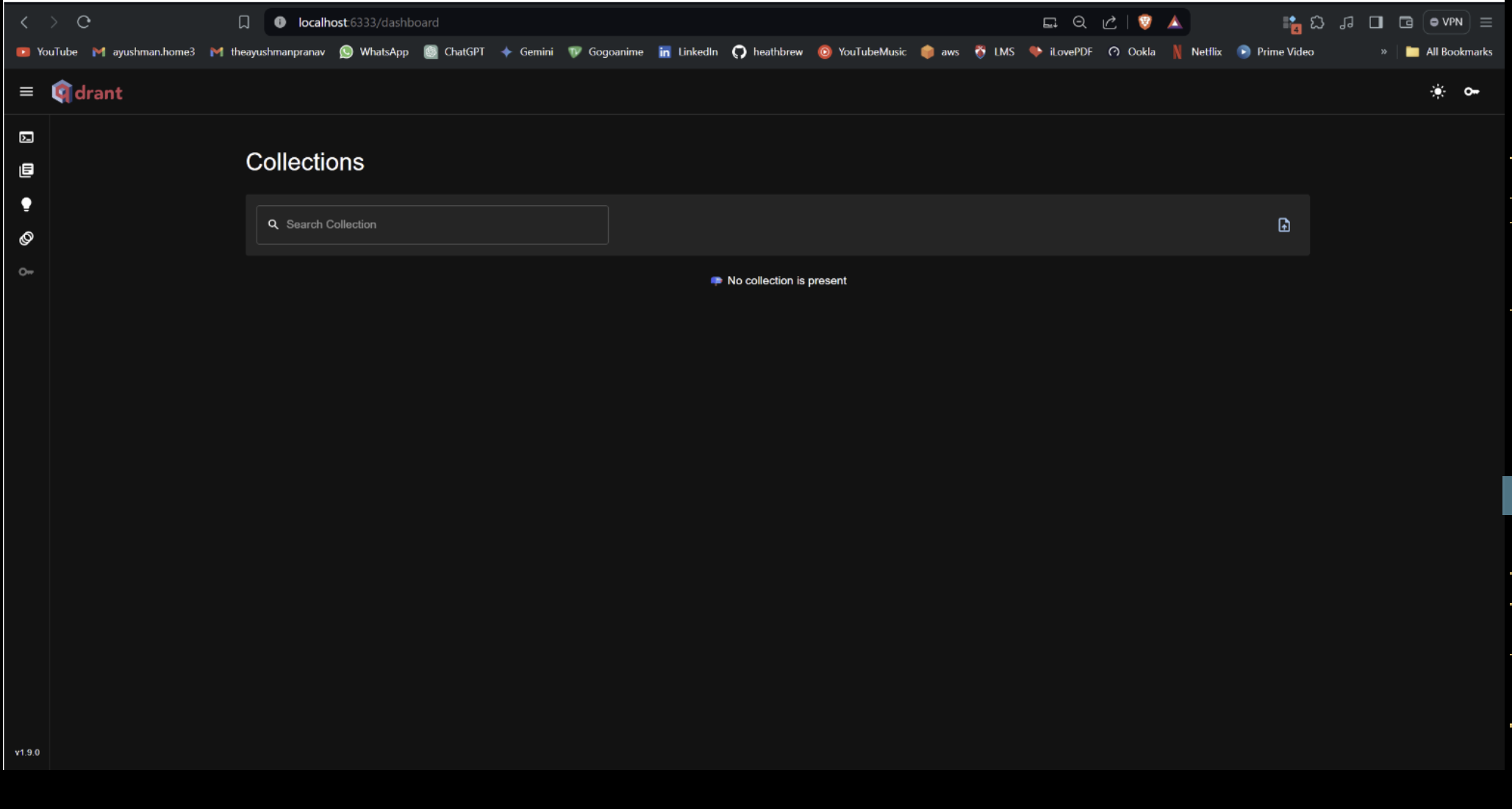
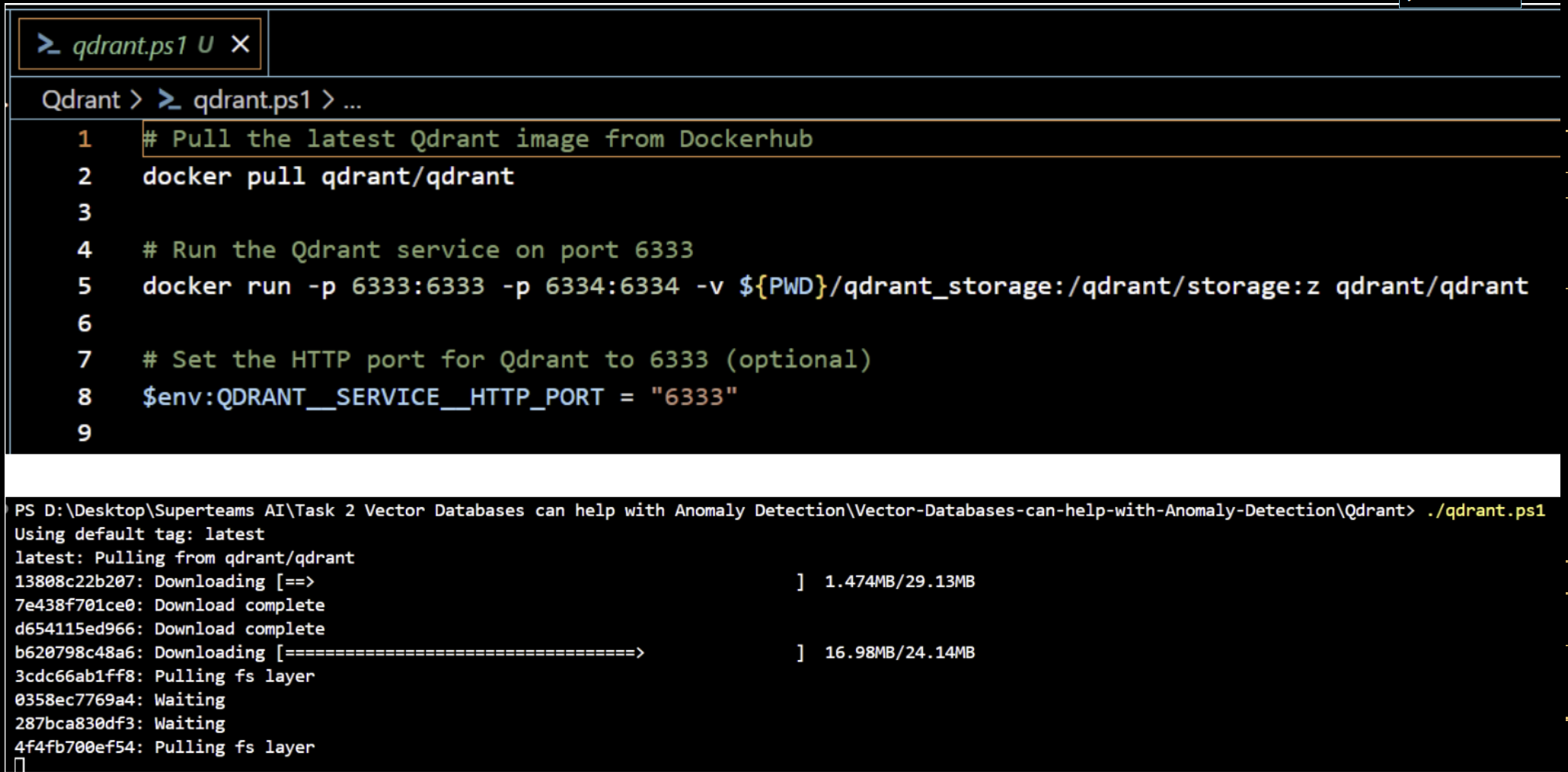
Docker Desktop must be installed.

In [1] I have included a file named Qdrant.ps1.

You can use this to pull the Qdrant image and then run it on port 6333. You will see the local UI, which displays the vector stores.

For this ensure you have docker desktop installed





This code creates the qdrant dataset with the name of pubheath

from qdrant\_client import QdrantClient

from qdrant\_client.http.models import Distance, VectorParams

# Initialize Qdrant client

qdrant\_client = QdrantClient(host='localhost', port=6333)

collection\_name = "pubhealth"

# Specify the vectors' configuration

vectors\_config = VectorParams(

    size=model.config.hidden\_size,  # The size of your embeddings

    distance=Distance.COSINE  # The distance metric for the vector space

)

# Create or recreate the collection with the specified configuration

qdrant\_client.recreate\_collection(

    collection\_name=collection\_name,

    vectors\_config=vectors\_config,

    # Optionally, you can specify other parameters for the collection

)

Why to use qdrant instead of faiss ?

Because qdrant allows a searchable part of the dataset all the while retaining the jsonic structure of the rest of the dataset .

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Id | Vector (this is the part we search )  This contains the minilm-12 encoded merged data  Currently setup with Cosine distance | | This is the payload that will be returned   |  |  |  |  | | --- | --- | --- | --- | | claim | explanation | label | Main\_text | |

import ast

from qdrant\_client import QdrantClient

from qdrant\_client.http.models import PointStruct

# Connect to Qdrant

client = QdrantClient(host="localhost", port=6333)

# Convert encoded\_merged\_data from string to list of floats if necessary

df['encoded\_merged\_data'] = df['encoded\_merged\_data'].apply(lambda x: ast.literal\_eval(x) if isinstance(x, str) else x)

# Insert data into Qdrant

for index, row in df.iterrows():

    qdrant\_client.upsert(

        collection\_name=collection\_name,

        points=[{

            "id": index,  # Using the dataframe index as the ID

            "vector": row['encoded\_merged\_data'],  # Assuming row['encoded\_merged\_data'] is a list of floats

            "payload": {

                "claim": row['claim'],

                "explanation": row['explanation'],

                "label": row['label'],

                "main\_text": row['main\_text']

            }

        }]

    )

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:16 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:17 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:17 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:17 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

2024-05-22 16:04:17 - INFO - HTTP Request: PUT http://localhost:6333/collections/pubhealth/points?wait=true "HTTP/1.1 200 OK"

The data has reached the qdrant database

## Qdrant basics of a querying client

The code connects to the qdrant client and database

from langchain.vectorstores import Qdrant

url = "http://localhost:6333"  # URL where the Qdrant service is running

collection\_name =  "pubhealth"  # Name of the collection in Qdrant

# Initialize the Qdrant client with the specified URL

client = QdrantClient(

    url=url,

    prefer\_grpc=False  # Indicates whether to use gRPC for communication

)

logging.info(f"QdrantClient initialized: {client}")  # Prints the client information

logging.info(f"#################################")  # Prints a separator line

this code below

def similarity\_search\_with\_score(query, k=2):

    query\_embedding = generate\_embedding(query)[0].tolist()

    search\_results = qdrant\_client.search(

        collection\_name=collection\_name,

        query\_vector=query\_embedding,

        limit=k,

        with\_payload=True,

        with\_vectors=False

    )

    return search\_results

query = "What are the key strategies for improving public health in urban areas?"

search\_results = similarity\_search\_with\_score(query=query, k=5)

for result in search\_results:

    doc\_id = result.id

    score = result.score

    payload = result.payload  # The payload should contain your text or a reference to it.

    # Assuming the payload contains a field 'input' where the text is stored

    claim = payload.get('claim', 'No content available')

    explanation = payload.get('explanation', 'No content available')

    explanation = payload.get('explanation', 'No content available')

    label = payload.get('label', 'No content available')

    main\_text = payload.get('main\_text', 'No content available')

    # Print the similarity score and document content

    logging.info({"score": score, "doc\_id": doc\_id, "claim": claim, "explanation": explanation, "label": label, "main\_text": main\_text})

in this code we have asked for one query and then set the k=5 that means it ill return the top 5 docs stored (here the top 5 rows of the dataset)

2024-05-22 17:03:24 - INFO - HTTP Request: POST http://localhost:6333/collections/pubhealth/points/search "HTTP/1.1 200 OK"

2024-05-22 17:03:24 - INFO - {'score': 0.3489651, 'doc\_id': 7, 'claim': 'U.S. says results encouraging for healthcare delivery reforms.', 'explanation': 'The Obama administration on Thursday reported what it called encouraging results from efforts to reduce healthcare costs and improve the quality of care for more than 5 million Medicare beneficiaries under Obamacare', 'label': 'true', 'main\_text': 'As part of President Barack Obama’s healthcare reform law, the efforts center around more than 360 accountable care organizations (ACOs), which are networks of doctors, hospitals and other providers specially organized to help move Medicare away from traditional fee-for-service medicine. The U.S. Centers for Medicare and Medicaid Services (CMS) said preliminary data show that the ACOs produced $380 million in savings vis-a-vis traditional Medicare in 2012 by giving doctors and other healthcare providers the incentive to focus on improved outcomes for patients instead of fees from tests and services. Medicare, the $575 billion government healthcare system for 51 million elderly and disabled beneficiaries, faces growing financial pressures as a result of America’s aging population. A mainstay, the trust fund that pays for hospitalization, is expected to be exhausted in 2026. Deficit hawks view Medicare as a future driver of the federal debt and have called for major systemic reforms. But the Obama administration has pursued gradual changes including the reform of care delivery systems. So-called fee-for-service medicine is widely viewed as a cause of rising healthcare costs, because it calls for paying healthcare providers for tests and services that are sometimes unnecessary. Obamacare seeks to tackle costs by exploring ACOs and other new healthcare business models intended to find savings that do not jeopardize care. A main goal is to generate savings large enough to be shared between Medicare and providers. But some experts are skeptical, saying significant cost reductions could be hard to maintain over time. But CMS, an agency within the U.S. Department of Health and Human Services, runs two different ACO programs. In its largest, 54 of 114 ACO networks achieved lower than expected expenditures. But only 29 saw savings big enough to share with providers. All told, the program produced $128 million in net savings for Medicare’s trust funds. “Overall, the ACO program’s a net saver to the Medicare program,” CMS principal deputy administrator Jon Blum told reporters in a conference call. “It’s giving us great confidence that this is the right course for the Medicare program and we are confident that it will continue to show quality improvement and cost savings.” Officials said the ACOs also achieved a wide range of quality goals. But CMS released no quality statistics. Thursday’s government release drew some cautious optimism from the healthcare industry. “Today’s report reflects important steps. More work is needed to modernize our antiquated Medicare payment system and base payment on evidence-based quality measures and proven patient outcomes,” said Dr. John Noseworthy, chief executive of the Mayo Clinic in Rochester, Minnesota, which is not part of the government’s program. “As results of the team-based care models are analyzed, those most effective in driving down health care costs without compromising safety and quality should become part of the healthcare system,” he said.'}

2024-05-22 17:03:24 - INFO - {'score': 0.22870876, 'doc\_id': 4, 'claim': 'Some appendicitis cases may not require ’emergency’ surgery', 'explanation': 'We really don’t understand why only a handful of mainstream news organizations reported this story. (At least in what we found.) The most common emergency surgery in the world. Rushing to emergency surgery may not carry any benefit. Waiting a few hours may be safer and less expensive. Why is that not a story? We applaud USA Today for finding time and space – and clearly it didn’t need to free up much space to do a good job telling the story. The story explains that as many as 300,000 appendectomies are done each year in the US. That figure alone explains why this is an important study to report.', 'label': 'true', 'main\_text': '"Although the story didn’t cite the cost of appendectomy – emergency or urgent surgery – and we wish it had, we nonetheless will give it a satisfactory score because it at least cited what the editorial writer wrote, ""A secondary benefit is the savings to the hospital generated by minimizing staff and anesthesiologist presence late in the evening and during the wee hours of the morning."" As with our harms score above, although the story didn’t give absolute numbers, in this case we think it was sufficient for it to report that ""The scientists found no significant difference among the groups in the patients’ condition 30 days after surgery or in the length of their operation or hospital stay."" Although the story didn’t give absolute numbers, in this case we think it was sufficient for it to report that ""The scientists found no significant difference among the groups in the patients’ condition 30 days after surgery or in the length of their operation or hospital stay."" Despite running less than 300 words, this story did an adequate job in explaining the quality of the evidence, including pointing out limitations. No disease-mongering here. The story meets the bare minimum requirement for this criterion in that it at least cited what an editorial stated. The focus of the story was on a study comparing emergency appendectomy with surgery done up to 12 hours later or beyond. This is the whole focus of the story – and one we applaud – when it begins:\xa0 ""Appendectomy is the most common emergency surgery in the world, but it doesn’t have to be."" There were no claims made about the novelty of this research, and we may have wished for a bit more context on this. Nonetheless, the potential for guiding future care decisions was made clear. Not applicable. Given that the story only pulled excerpts from the journal article and the accompanying editorial, and didn’t include any fresh quotes from interviews, we can’t be sure of the extent to which it may have been influenced by a news release."'}

2024-05-22 17:03:24 - INFO - {'score': 0.21610679, 'doc\_id': 3, 'claim': 'Study: Vaccine for Breast, Ovarian Cancer Has Potential', 'explanation': 'While the story does many things well, the overall framing of the story is that the vaccine “shows promise,” when the evidence actually points in the other direction. Because only one patient in the study remains cancer free and because that patient may very well have benefited from an earlier cancer vaccine and other complicating factors, we question the decision to write this story in the first place. Right now, there more than 10,000 cancer-related clinical trials recruiting patients. Cancer has foiled scientists repeatedly with treatments that initially seemed promising in the laboratory or in a very small group of people and later proved unworkable on a larger scale. It’s a difficult task — but a crucial one — for reporters to ask tough questions of the evidence and a wide range of sources before deciding whether one of these thousands of experimental treatment options merits coverage.', 'label': 'true', 'main\_text': 'The story does discuss costs, but the framing is problematic. The story, based on a conversation with one source, the study’s lead investigator, says, “It’s difficult at this point to predict costs. However, he expects costs will not approach those for Provenge, the pricey treatment vaccine for prostate cancer approved by the FDA in 2010. Provenge costs $93,000 for the one-month, three-dose treatment. Medicare covers it.” This tells readers that, no matter what the drug costs, Medicare likely will cover it. We appreciate the effort to bring cost information into the story, but this type of information is misleading. The story does explain that only one patient remains cancer free following the study. It then details how for most of the patients cancer continued to progress after 2 months. It says that the median overall survival in both the breast cancer and ovarian cancer patients was less than 16 months. But the story is framed in such a way to highlight the one potentially positive outcome of the study and to downplay the negative. We read more sooner about the one patient who may have responded well to the vaccine than we do about the 25 other patients who did not. The story mentions side effects in a satisfactory way. Technically, the story provides readers with much of the information they would need to assess the validity of the study, but it comes out in bits and pieces. For example, we only find out near the end of the story that “The woman, who remains disease-free, had a previous treatment with a different treatment vaccine. ‘That might have primed her immune system,’ Gulley speculates. She also had only one regimen of chemotherapy, perhaps keeping her immune system stronger.” This casts much doubt on the study’s design, and it would have been nice to have seen some outside expertise brought in to either discuss those design problems or to torpedo the story altogether. Again, the story deserves high marks for being very specific in the lead and throughout the story. It says, that the vaccine is “for breast and ovarian cancer that has spread to other parts of the body” in the lead and later details the particular circumstances of the study cohort. It says, “The patients had already undergone a variety of treatments but the cancer was progressing. Twenty one of the 26 had undergone three or more chemotherapy regimens.” This is the root of the story’s main shortcoming. Almost all of the information in the story comes from one source: Dr. James\xa0Gulley, who oversaw the study. Gulley is quite enthusiastic about this vaccine, despite the evidence, and the story needed more perspectives to put this vaccine into a broader context. At the very end, there are a few comments from Dr. Vincent K. Tuohy, who also is working on a breast cancer vaccine. Because of his competing research, he seems to have a conflict, but even putting that aside, his comments were not used to their best effect. There was no comparison in the story to existing alternatives. The median survival, for example, is presented without the context of how long these patients might have lived had they been undergoing standard chemotherapy and radiation treatments. We give high marks to the story for saying right in the lead that the findings are from “a preliminary study in 26 patients.” That tells readers both that the findings need to be interpreted with caution and that the treatment is not available to most people. The concept of vaccines for breast/ovarian cancer is indeed novel, and the story acknowledges that other vaccines are being studied. The story does not rely on a news release.'}

2024-05-22 17:03:24 - INFO - {'score': 0.17250797, 'doc\_id': 5, 'claim': 'Britain to reveal trial criteria for coronavirus antibody tests.', 'explanation': 'British regulators will this week reveal approval criteria for firms offering new coronavirus antibody tests, touted by governments in Britain and elsewhere as critical to easing nationwide lockdowns without helping the virus to spread.', 'label': 'true', 'main\_text': 'Antibody tests show whether whether people have been infected with the novel coronavirus and developed immunity - potentially allowing them to return to their places of work. The British government has provisionally ordered 17.5 million of them, but health minister Matt Hancock has said some of those already being trialled work poorly, and that one test even missed three out of four cases. Brigette Bard, chief executive of the diagnostics firm BioSure, said that her firm’s at-home test could not get as far as a formal trial as she had received no details of the approval criteria. “How they fail tests when there is no specification, I literally have no idea. We have been begging to be told what to do. We have a test ready to submit,” she said. “We need to get our test approved, and they’re not giving us an option of how to get it approved.” BioSure has been producing an at-home HIV test since 2015, and Bard met Prime Minister Boris Johnson in March to discuss her COVID-19 test. At-home tests, in this case of blood obtained by pricking a finger, are designed to be read by the person taking the test, whereas others need results to be read in a lab. Bard said laboratory trials were producing good data, but that the firm needed to know what size and scope of trial and what accuracy level would be accepted by regulators. Doris-Ann Williams, chief executive of the British In Vitro Diagnostic Association, said that clarity would be coming soon. “There will be specifications released for tests this week by MHRA (the Medicines and Healthcare products Regulatory Agency) which will help companies understand the technical performance expected for their kits,” she said in an email. The MHRA itself said it was developing the specifications but declined to give a timeline. Once criteria were established, Bard said the trials would take one or two weeks, and then BioSure could begin production, subject to approval, with the aim of making 1 million tests a month. “If our test fails the validation, then it’s not fit to go to market,” she said. “I’m happy to go through whatever validation they want us to go through.”'}

2024-05-22 17:03:24 - INFO - {'score': 0.17075726, 'doc\_id': 9, 'claim': 'Poor test results for heart drugs', 'explanation': '"The story reports no additional cardiovascular protection with the cholesterol lowering drug ezetimibe (Zetia) alone, or in combination with the statin simvastatin (trade name Zocor). The report is unfortunately yet another example of an attempt to explain a very complicated story in a short TV chat. The study design, patient population, methodology and support for the surrogate endpoint are absent from the discussion. So, the context of ""failure"" is totally lost on the audience. This failure to provide any context is especially unfortunate because the report was accompanied by two excellent editorials. The editorial by Brown and Taylor in the NEJM highlights the study design, patient cohort studied and puts the results into an objective context. This report does little to inform its listeners and a wonderful opportunity was lost in the process. The story does not note the cost of a typical statin regimen, Zetia or the combination treatment, Vytorin. This is an important oversight as many people who are prescribed these medications take them for life. The story does mention that these new drugs ""raked in 5 billion dollars last year"". The failure of the combination product to be better than simvastatin alone has significant financial implications. Simvastatin (the generic name for Merck’s Zocor) is available for about $1.25 a day. Vytorin costs about $3.35 a day. Most clinicians assumed that the combined product was an advantage if you could not get a patient to a LDL goal with just the statin alone or side effects with the simvastatin prevented an adequate dose. The price for the combination is less than the cost of the components. Also, stating that taking more statins is the answer is incomplete at best. Lifestyle changes along with other drugs may be options as well. Interestingly however, this study raises questions about the simple notion that lowering total cholesterol, and LDL are important to lowering cardiovascular risk. People who are not able to lower their cholesterol enough with statins alone (or cannot tolerate the dose of statins needed to lower it) may wish to discuss with their doctor whether other medications (i.e., rather than ezetimibe) may be appropriate. These include niacin, fibrates, and bile acid resins. When added to statins, they can effectively lower cardiovascular risk. There are no interviews with the study authors or with practicing clinicians. The story reported on data presented at a recent American College of Cardiology Meeting, so there were several thousand cardiologists and other specialists available who could have been interviewed for clinical perspective on the results of this study."', 'label': 'false', 'main\_text': '"The story does not note the cost of a typical statin regimen, Zetia or the combination treatment, Vytorin. This is an important oversight as many people who are prescribed these medications take them for life. The story does mention that these new drugs ""raked in 5 billion dollars last year"". The failure of the combination product to be better than simvastatin alone has significant financial implications. Simvastatin (the generic name for Merck’s Zocor) is available for about $1.25 a day. Vytorin costs about $3.35 a day. Most clinicians assumed that the combined product was an advantage if you could not get a patient to an LDL goal with just the statin alone or side effects with the simvastatin prevented an adequate dose. The price for the combination is less than the cost of the components. The story provides no quantitative data and no information on the number needed to treat to show benefit (i.e. prevention of a heart attack or stroke through lowered LDLs) in one patient with the newer medications, with traditional statins or with combination treatment. The story says that there are no real dangers to continuing the drugs Zetia and Vytorin based on the lower drop-out rate and reported side effects in the study. The comments from both parties cast an extraordinarily negative shadow on the study drug and on the company sponsors. Without an appropriate description of the study design, the results cannot be fairly discussed to the public. While the primary endpoint was not obtained with the combination, it did lower LDL and inflammatory markers in excess of that achieved with simvastatin alone. Suggesting the drug did not ""work"" is a bit of a mis-statement. The story encourages patients to keep taking these drugs. The story also does not mention the potential harm of taking statins long-term, especially in high does. Some of these harms are very rare, however, they include: muscle pain and kidney or liver problems. The story provides no real discussion of the data presented at the American College of Cardiology meeting and published in the peer-reviewed New England Journal of Medicine. And the story didn’t assess the quality of the evidence. The study followed people (average age mid-40s) who had an inherited condition (called familial hypercholesterolemia) that is associated with very high cholesterol levels and greatly increased risk of early coronary artery disease. Many of them had been taking statins and other cholesterol-lowering medicines for years. All were randomly assigned to take either a statin (simvastatin, trade name Zocor) alone or a statin combined with another cholesterol-lowering medication, ezetimibe. The study was designed to find out if the combination of the two drugs could slow the growth of plaque in carotid arteries supplying the brain more than the statin alone. Plaque in these arteries is associated with an increased risk of stroke and heart attack. The two drugs together were more effective at lowering cholesterol than simvastatin alone, but adding ezetimibe did not change plaque measurements in the carotid arteries. The story does mention that lowering cholesterol may have other benefits, but we are not sure how those translate to fewer cardiovascular events or increased survival from these events. No overt disease-mongering. There are no interviews with the study authors or with practicing clinicians. The story reported on data presented at a recent American College of Cardiology Meeting, so there were several thousand cardiologists and other specialists available who could have been interviewed for clinical perspective on the results of this study. There was also no mention of the two accompanying editorials in the NEJM. The editorial by Brown and Tayor nicely identifies the issues related to the study and helps put the results into perspective. Unfortunately, this editorial was ignored by both parties involved in the story. Stating that taking more statins is the answer is incomplete at best. Lifestyle changes along with other drugs may be options as well. Interestingly however, this study raises questions about the simple notion that lowering total cholesterol, and LDL are important to lowering cardiovascular risk. People who are not able to lower their cholesterol enough with statins alone (or cannot tolerate the dose of statins needed to lower it) may wish to discuss with their doctor whether other medications (i.e., rather than ezetimibe) may be appropriate. These include niacin, fibrates, and bile acid resins. When added to statins, they can effectively lower cardiovascular risk. It’s clear from the story that the drugs in question are still available to patients, though they should not be considered first-line therapy. The story focuses on new information that there is little to no benefit of newer cholestrol-lowering medications on the prevention of arterial plaque, which translate to little benefit for prevention of heart disease, stroke and cardiovascular-related death. We can’t be sure if the story relied solely or largely on a news release. No researcher or cardiologist is interviewed. Quantified data from the American College of Cardiology presentation or from the New England Journal of Medicine article are not directly cited."'}

## Building the Retrieval-Augmented Generation (RAG) React Agent

We will be using langchain agent with this to use different tools for langchain agents

To create a multihop react agent .

from langchain.agents import initialize\_agent, Tool

from langchain.llms import OpenAI

from langchain.agents import AgentType

from langchain.tools import BaseTool

def similarity\_search\_with\_score(query, k=1):

    query\_embedding = generate\_embedding(query)[0].tolist()

    search\_results = qdrant\_client.search(

        collection\_name=collection\_name,

        query\_vector=query\_embedding,

        limit=k,

        with\_payload=True,

        with\_vectors=False

    )

    return search\_results

def lookup\_tool(search\_quality\_reflection, search\_quality\_score):

    search\_results = similarity\_search\_with\_score(search\_quality\_reflection)

    lookup\_str = "\n".join([f"Doc: {r.payload.get('claim', 'No text available')}" for r in search\_results])

    return lookup\_str

search\_tool = Tool(

    name="Search",

    func=similarity\_search\_with\_score,

    description="Searches for relevant information using a vector similarity search.",

    return\_direct=True

)

lookup\_tool = Tool(

    name="Lookup",

    func=lookup\_tool,

    description="Looks up the actual text for the search results."

)

llm = OpenAI(temperature=0)

agent = initialize\_agent([search\_tool, lookup\_tool], llm, agent=AgentType.REACT\_DOCSTORE, verbose=True)

In this code I have used 2 custom functions

1. Search tool ( this tool does the qdrant vector search for you )
2. Look up tool ( this tool does the payload get from the returned data )

query = "What are the key strategies for improving public health in urban areas?"

result = agent.run(query)

print(result)

2024-05-22 18:45:08 - INFO - Retrying request to /completions in 0.837140 seconds

[1m> Entering new AgentExecutor chain...[0m

2024-05-22 18:45:10 - INFO - HTTP Request: POST https://api.openai.com/v1/completions "HTTP/1.1 200 OK"

2024-05-22 18:45:10 - INFO - HTTP Request: POST http://localhost:6333/collections/pubhealth/points/search "HTTP/1.1 200 OK"

[32;1m[1;3mThought: I need to search strategies for improving public health in urban areas and find the key strategies.

Action: Search[strategies for improving public health in urban areas][0m

Observation: [36;1m[1;3m[ScoredPoint(id=7, version=7, score=0.2911795, payload={'claim': 'U.S. says results encouraging for healthcare delivery reforms.', 'explanation': 'The Obama administration on Thursday reported what it called encouraging results from efforts to reduce healthcare costs and improve the quality of care for more than 5 million Medicare beneficiaries under Obamacare', 'label': 'true', 'main\_text': 'As part of President Barack Obama’s healthcare reform law, the efforts center around more than 360 accountable care organizations (ACOs), which are networks of doctors, hospitals and other providers specially organized to help move Medicare away from traditional fee-for-service medicine. The U.S. Centers for Medicare and Medicaid Services (CMS) said preliminary data show that the ACOs produced $380 million in savings vis-a-vis traditional Medicare in 2012 by giving doctors and other healthcare providers the incentive to focus on improved outcomes for patients instead of fees from tests and services. Medicare, the $575 billion government healthcare system for 51 million elderly and disabled beneficiaries, faces growing financial pressures as a result of America’s aging population. A mainstay, the trust fund that pays for hospitalization, is expected to be exhausted in 2026. Deficit hawks view Medicare as a future driver of the federal debt and have called for major systemic reforms. But the Obama administration has pursued gradual changes including the reform of care delivery systems. So-called fee-for-service medicine is widely viewed as a cause of rising healthcare costs, because it calls for paying healthcare providers for tests and services that are sometimes unnecessary. Obamacare seeks to tackle costs by exploring ACOs and other new healthcare business models intended to find savings that do not jeopardize care. A main goal is to generate savings large enough to be shared between Medicare and providers. But some experts are skeptical, saying significant cost reductions could be hard to maintain over time. But CMS, an agency within the U.S. Department of Health and Human Services, runs two different ACO programs. In its largest, 54 of 114 ACO networks achieved lower than expected expenditures. But only 29 saw savings big enough to share with providers. All told, the program produced $128 million in net savings for Medicare’s trust funds. “Overall, the ACO program’s a net saver to the Medicare program,” CMS principal deputy administrator Jon Blum told reporters in a conference call. “It’s giving us great confidence that this is the right course for the Medicare program and we are confident that it will continue to show quality improvement and cost savings.” Officials said the ACOs also achieved a wide range of quality goals. But CMS released no quality statistics. Thursday’s government release drew some cautious optimism from the healthcare industry. “Today’s report reflects important steps. More work is needed to modernize our antiquated Medicare payment system and base payment on evidence-based quality measures and proven patient outcomes,” said Dr. John Noseworthy, chief executive of the Mayo Clinic in Rochester, Minnesota, which is not part of the government’s program. “As results of the team-based care models are analyzed, those most effective in driving down health care costs without compromising safety and quality should become part of the healthcare system,” he said.'}, vector=None, shard\_key=None)][0m

[32;1m[1;3m[0m

[1m> Finished chain.[0m

[ScoredPoint(id=7, version=7, score=0.2911795, payload={'claim': 'U.S. says results encouraging for healthcare delivery reforms.', 'explanation': 'The Obama administration on Thursday reported what it called encouraging results from efforts to reduce healthcare costs and improve the quality of care for more than 5 million Medicare beneficiaries under Obamacare', 'label': 'true', 'main\_text': 'As part of President Barack Obama’s healthcare reform law, the efforts center around more than 360 accountable care organizations (ACOs), which are networks of doctors, hospitals and other providers specially organized to help move Medicare away from traditional fee-for-service medicine. The U.S. Centers for Medicare and Medicaid Services (CMS) said preliminary data show that the ACOs produced $380 million in savings vis-a-vis traditional Medicare in 2012 by giving doctors and other healthcare providers the incentive to focus on improved outcomes for patients instead of fees from tests and services. Medicare, the $575 billion government healthcare system for 51 million elderly and disabled beneficiaries, faces growing financial pressures as a result of America’s aging population. A mainstay, the trust fund that pays for hospitalization, is expected to be exhausted in 2026. Deficit hawks view Medicare as a future driver of the federal debt and have called for major systemic reforms. But the Obama administration has pursued gradual changes including the reform of care delivery systems. So-called fee-for-service medicine is widely viewed as a cause of rising healthcare costs, because it calls for paying healthcare providers for tests and services that are sometimes unnecessary. Obamacare seeks to tackle costs by exploring ACOs and other new healthcare business models intended to find savings that do not jeopardize care. A main goal is to generate savings large enough to be shared between Medicare and providers. But some experts are skeptical, saying significant cost reductions could be hard to maintain over time. But CMS, an agency within the U.S. Department of Health and Human Services, runs two different ACO programs. In its largest, 54 of 114 ACO networks achieved lower than expected expenditures. But only 29 saw savings big enough to share with providers. All told, the program produced $128 million in net savings for Medicare’s trust funds. “Overall, the ACO program’s a net saver to the Medicare program,” CMS principal deputy administrator Jon Blum told reporters in a conference call. “It’s giving us great confidence that this is the right course for the Medicare program and we are confident that it will continue to show quality improvement and cost savings.” Officials said the ACOs also achieved a wide range of quality goals. But CMS released no quality statistics. Thursday’s government release drew some cautious optimism from the healthcare industry. “Today’s report reflects important steps. More work is needed to modernize our antiquated Medicare payment system and base payment on evidence-based quality measures and proven patient outcomes,” said Dr. John Noseworthy, chief executive of the Mayo Clinic in Rochester, Minnesota, which is not part of the government’s program. “As results of the team-based care models are analyzed, those most effective in driving down health care costs without compromising safety and quality should become part of the healthcare system,” he said.'}, vector=None, shard\_key=None)]

## References

[1] “heathbrew/Building-a-Diet-and-Wellness-AI-Agent-A-Comprehensive-Guide: his guide aims to build an AI agent that can answer queries related to fitness, health, calories, and more, leveraging reasoning capabilities to provide accurate and contextual responses.” Accessed: May 22, 2024. [Online]. Available: https://github.com/heathbrew/Building-a-Diet-and-Wellness-AI-Agent-A-Comprehensive-Guide

[2] “health\_fact · Datasets at Hugging Face.” Accessed: May 21, 2024. [Online]. Available: https://huggingface.co/datasets/health\_fact

[3] “sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 · Hugging Face.” Accessed: Apr. 18, 2024. [Online]. Available: https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

## How to connect with me

https://www.linkedin.com/in/ayushman-pranav-20b49ba3/

## **Some Other articles by Me**

<https://medium.com/@AyushmanPranav/anomaly-detection-using-vector-search-c3f48f8e61eb>

<https://blog.devops.dev/steps-to-monitoring-dspy-qdrant-powered-rag-with-prometheus-or-grafana-b642335cbd50>

<https://medium.com/@AyushmanPranav/perplexity-calculation-in-nlp-0699fbda4594>

<https://medium.com/@AyushmanPranav/utilizing-aws-rds-with-python-pymysql-for-efficient-chatgpt-api-calls-management-c5fca30aac37>

<https://medium.com/@AyushmanPranav/build-a-mindblowing-3d-portfolio-website-three-js-beginners-tutorial-556c7b772989?sk=42186a0a7c4940e2a3e9a359348f9f6e>