Week 11 Homework 2: GenAI - Containerized video transcription and chat app

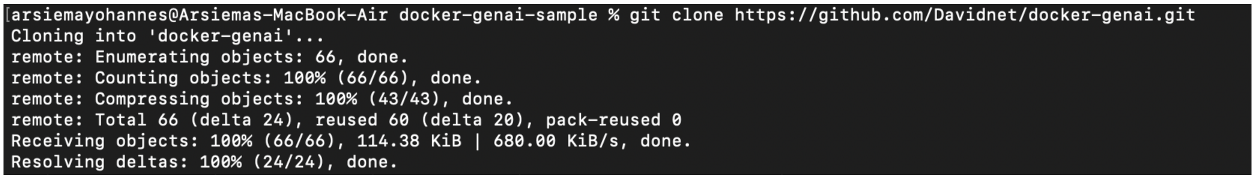
Name: Arsiema Yohannes

ID; 20039

1. Clone the Repository:

git clone https://github.com/Davidnet/docker-genai.git

This command clones the repository docker-genai from GitHub to your local machine.

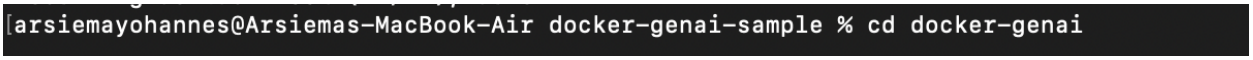


```
[arsiemayohannes@Arsiemas-MacBook-Air docker-genai-sample % git clone https://github.com/Davidnet/docker-genai.git
Cloning into 'docker-genai'...
remote: Enumerating objects: 66, done.
remote: Counting objects: 100% (66/66), done.
remote: Compressing objects: 100% (43/43), done.
remote: Total 66 (delta 24), reused 60 (delta 20), pack-reused 0
Receiving objects: 100% (66/66), 114.38 KiB | 680.00 KiB/s, done.
Resolving deltas: 100% (24/24), done.
```

2. Navigate to the Repository:

cd docker-genai

Move into the cloned repository directory.

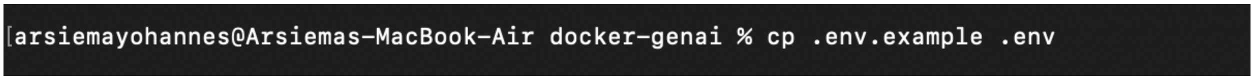


```
[arsiemayohannes@Arsiemas-MacBook-Air docker-genai-sample % cd docker-genai
```

3. Create the .env File:

cp .env.example .env

This command creates a copy of the .env.example file and names it .env.

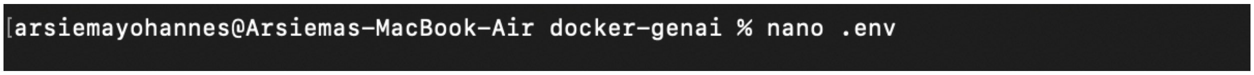


```
[arsiemayohannes@Arsiemas-MacBook-Air docker-genai % cp .env.example .env
```

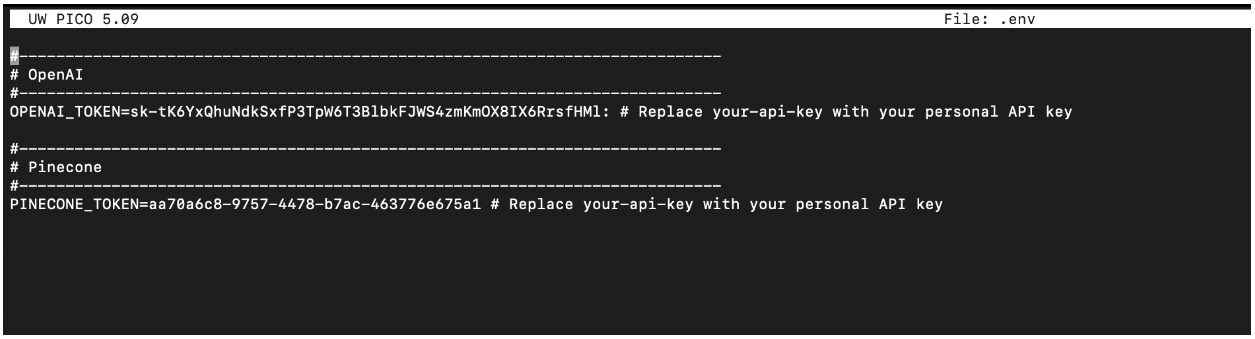
4. Open the .env File for Editing:

nano .env

This command opens the .env file in the nano text editor. If you prefer another text editor, you can replace nano with the command for your preferred editor.



```
arsiemayohannes@Arsiemas-MacBook-Air docker-genai % nano .env
```

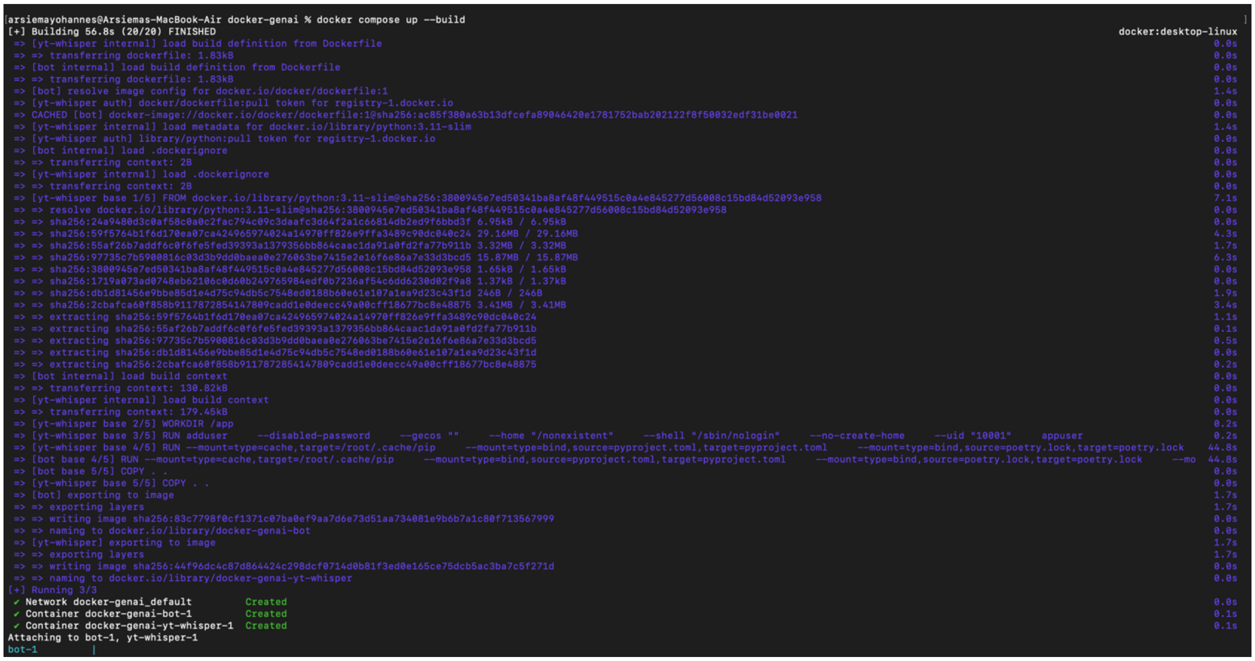


```
  UW PICO 5.09                                                                    File: .env

#-----------------------------------------------------------------------------
# OpenAI
#-----------------------------------------------------------------------------
OPENAI_TOKEN=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: # Replace your-api-key with your personal API key

#-----------------------------------------------------------------------------
# Pinecone
#-----------------------------------------------------------------------------
PINECONE_TOKEN=aa70a6c8-9757-4478-b7ac-463776e675a1 # Replace your-api-key with your personal API key
```

5. Build and Run the Application:

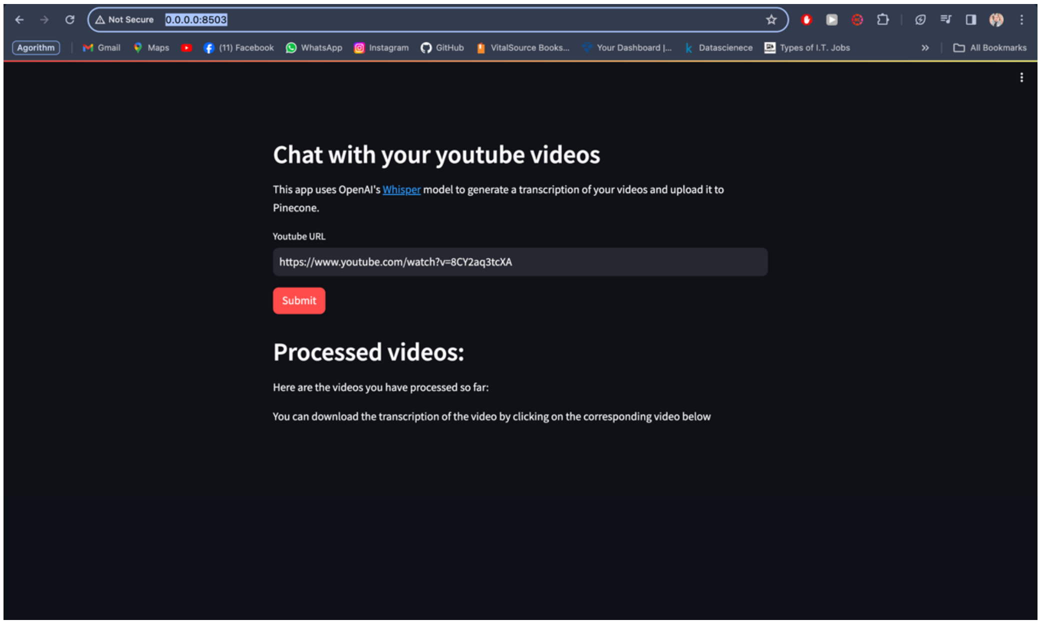
docker-compose up --build

This command builds the Docker images and starts the containers specified in the docker-compose.yaml file.



```
arsiemayohannes@Arsiemas-MacBook-Air docker-genai % docker compose up --build
[+] Building 56.8s (20/20) FINISHED                                                                    docker:desktop-linux
 => [yt-whisper internal] load build definition from Dockerfile                                                      0.0s
 => => transferring dockerfile: 1.83kB                                                                               0.0s
 => [bot internal] load build definition from Dockerfile                                                             0.0s
 => => transferring dockerfile: 1.83kB                                                                               0.0s
 => [bot] resolve image config for docker.io/docker/dockerfile:1                                                     1.4s
 => [yt-whisper auth] docker/dockerfile:pull token for registry-1.docker.io                                          0.0s
 => CACHED [bot] docker-image://docker.io/docker/dockerfile:1@sha256:ac85f380a63b13dfcefa89046420e1781752bab202122f8f50032edf31be0021  0.0s
 => [yt-whisper internal] load metadata for docker.io/library/python:3.11-slim                                       1.4s
 => [yt-whisper auth] library/python:pull token for registry-1.docker.io                                             0.0s
 => [bot internal] load .dockerignore                                                                                0.0s
 => => transferring context: 2B                                                                                      0.0s
 => [yt-whisper internal] load .dockerignore                                                                         0.0s
 => => transferring context: 2B                                                                                      0.0s
 => [yt-whisper base 1/5] FROM docker.io/library/python:3.11-slim@sha256:3800945e7ed50341ba8af48f449515c0a4e845277d56008c15bd84d52093e958  7.1s
 => => resolve docker.io/library/python:3.11-slim@sha256:3800945e7ed50341ba8af48f449515c0a4e845277d56008c15bd84d52093e958  0.0s
 => => sha256:24a9480d3c0af58c0a0c2fac794c09c3daafc3d64f2a1c66814db2ed9f6bbd3f 6.95kB / 6.95kB                       0.0s
 => => sha256:59f5764b1f6d170ea07ca424965974024a14970ff826e9ffa3489c90dc040c24 29.16MB / 29.16MB                     4.3s
 => => sha256:55af26b7addf6c0f6fe5fed39393a1379356bb864caac1da91a0fd2fa77b911b 3.32MB / 3.32MB                       1.7s
 => => sha256:97735c7b5900816c03d3b9dd0baea0e276063be7415e2e16f6e86a7e33d3bcd5 15.87MB / 15.87MB                     6.3s
 => => sha256:3800945e7ed50341ba8af48f449515c0a4e845277d56008c15bd84d52093e958 1.65kB / 1.65kB                       0.0s
 => => sha256:1719a073ad0748eb62106c0d60b249765984edf0b7236af54c6dd6230d02f9a8 1.37kB / 1.37kB                       0.0s
 => => sha256:db1d81456e9bbe85d1e4d75c94db5c7548ed0188b60e61e107a1ea9d23c43f1d 246B / 246B                           1.9s
 => => sha256:2cbafca60f858b9117872854147809cadd1e0deecc49a00cff18677bc8e48875 3.41MB / 3.41MB                       3.4s
 => => extracting sha256:59f5764b1f6d170ea07ca424965974024a14970ff826e9ffa3489c90dc040c24                            1.1s
 => => extracting sha256:55af26b7addf6c0f6fe5fed39393a1379356bb864caac1da91a0fd2fa77b911b                            0.1s
 => => extracting sha256:97735c7b5900816c03d3b9dd0baea0e276063be7415e2e16f6e86a7e33d3bcd5                            0.5s
 => => extracting sha256:db1d81456e9bbe85d1e4d75c94db5c7548ed0188b60e61e107a1ea9d23c43f1d                            0.0s
 => => extracting sha256:2cbafca60f858b9117872854147809cadd1e0deecc49a00cff18677bc8e48875                            0.2s
 => [bot internal] load build context                                                                                0.0s
 => => transferring context: 130.82kB                                                                                0.0s
 => [yt-whisper internal] load build context                                                                         0.0s
 => => transferring context: 179.45kB                                                                                0.0s
 => [yt-whisper base 2/5] WORKDIR /app                                                                               0.2s
 => [yt-whisper base 3/5] RUN adduser     --disabled-password     --gecos ""     --home "/nonexistent"     --shell "/sbin/nologin"     --no-create-home     --uid "10001"     appuser   0.2s
 => [yt-whisper base 4/5] RUN --mount=type=cache,target=/root/.cache/pip     --mount=type=bind,source=pyproject.toml,target=pyproject.toml     --mount=type=bind,source=poetry.lock,target=poetry.lock   44.8s
 => [bot base 4/5] RUN --mount=type=cache,target=/root/.cache/pip     --mount=type=bind,source=pyproject.toml,target=pyproject.toml     --mount=type=bind,source=poetry.lock,target=poetry.lock     --mo  44.8s
 => [bot base 5/5] COPY . .                                                                                          0.0s
 => [yt-whisper base 5/5] COPY . .                                                                                   0.0s
 => [bot] exporting to image                                                                                         1.7s
 => => exporting layers                                                                                              1.7s
 => => writing image sha256:83c7798f0cf1371c07ba0ef9aa7d6e73d51aa734081e9b6b7a1c80f713567999                         0.0s
 => => naming to docker.io/library/docker-genai-bot                                                                  0.0s
 => [yt-whisper] exporting to image                                                                                  1.7s
 => => exporting layers                                                                                              1.7s
 => => writing image sha256:44f96dc4c87d864424c298dcf0714d0b81f3ed0e165ce75dcb5ac3ba7c5f271d                         0.0s
 => => naming to docker.io/library/docker-genai-yt-whisper                                                           0.0s
[+] Running 3/3
 ✔ Network docker-genai_default         Created                                                                      0.0s
 ✔ Container docker-genai-bot-1         Created                                                                      0.1s
 ✔ Container docker-genai-yt-whisper-1  Created                                                                      0.1s
Attaching to bot-1, yt-whisper-1
bot-1         |
```

6. Can access the apps with http://0.0.0.0:8504/ and http://0.0.0.0:8503/



Chat with your youtube videos

This app uses OpenAI's Whisper model to generate a transcription of your videos and upload it to Pinecone.

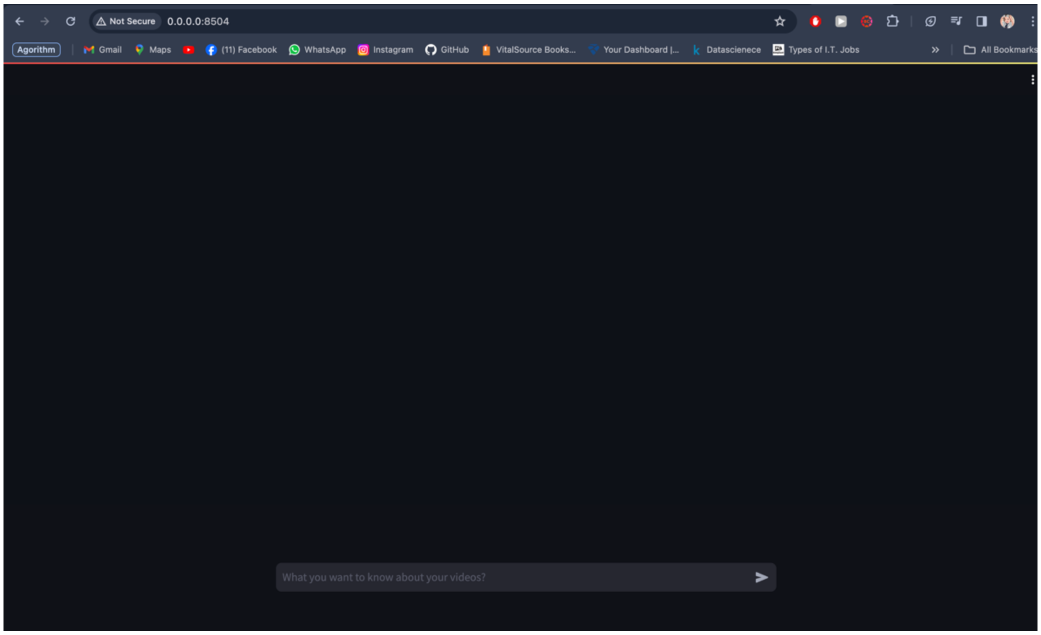
Youtube URL

https://www.youtube.com/watch?v=8CY2aq3tcXA

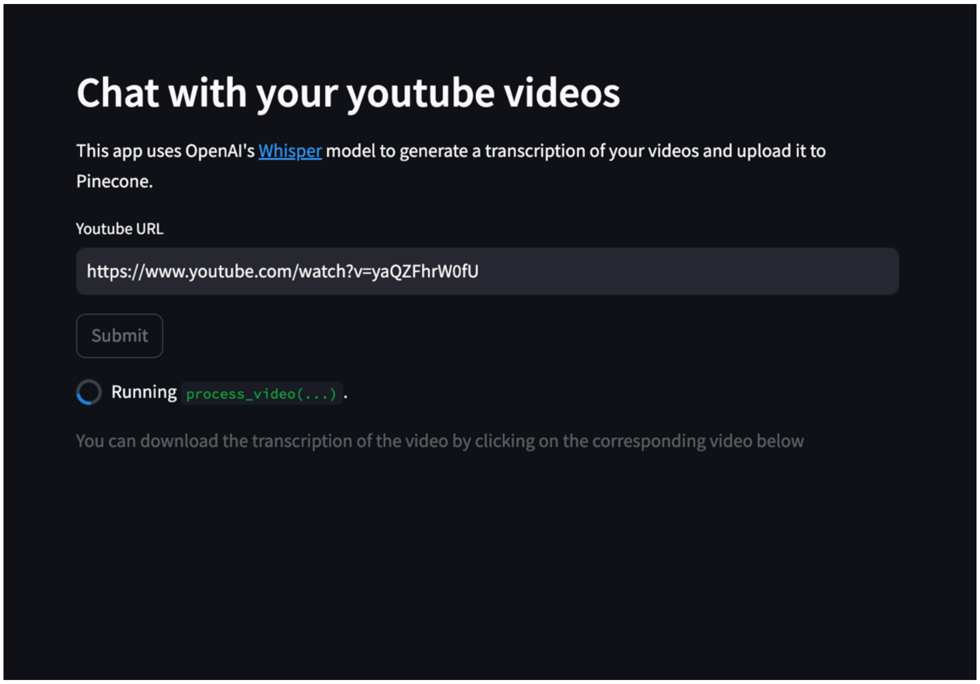
Processed videos:

Here are the videos you have processed so far:

You can download the transcription of the video by clicking on the corresponding video below



7. Once the application appears, in the Youtube URL field specify a Youtube video URL and select Submit. The following example uses https://www.youtube.com/watch?v=yaQZFhrW0fU



Chat with your youtube videos

This app uses OpenAI's Whisper model to generate a transcription of your videos and upload it to Pinecone.

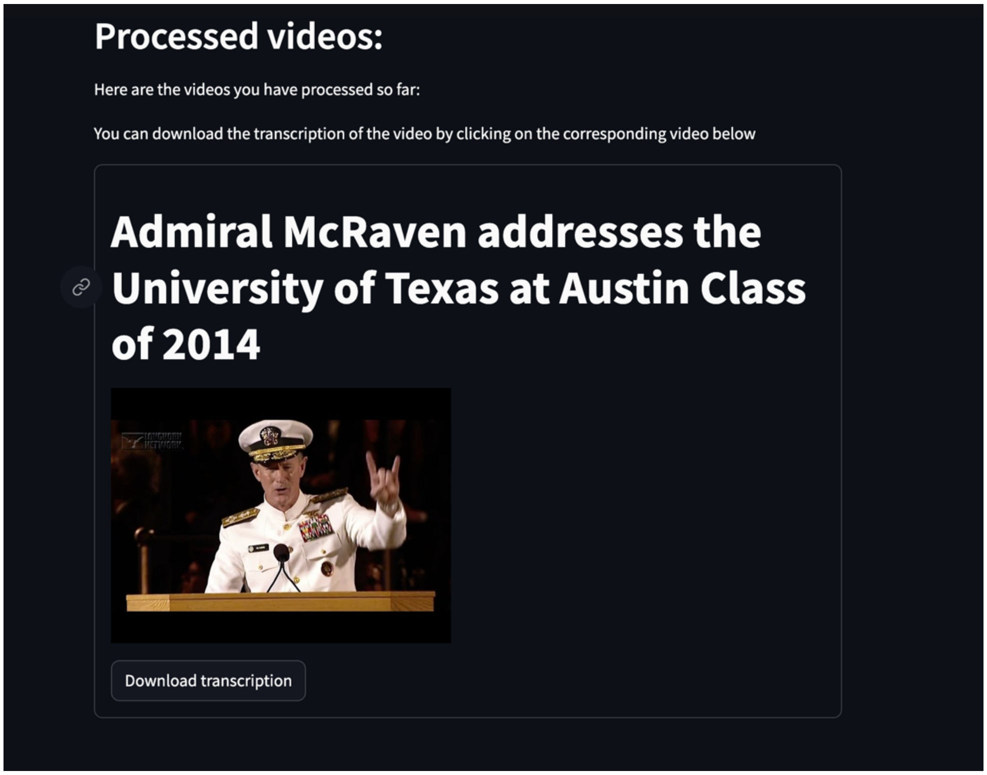
Youtube URL

https://www.youtube.com/watch?v=yaQZFhrW0fU

Running `process_video(...)`.

You can download the transcription of the video by clicking on the corresponding video below

After processing the video, a video list appears in the web app that informs you which videos have been indexed in Pinecone. It also provides a button to download the transcript.



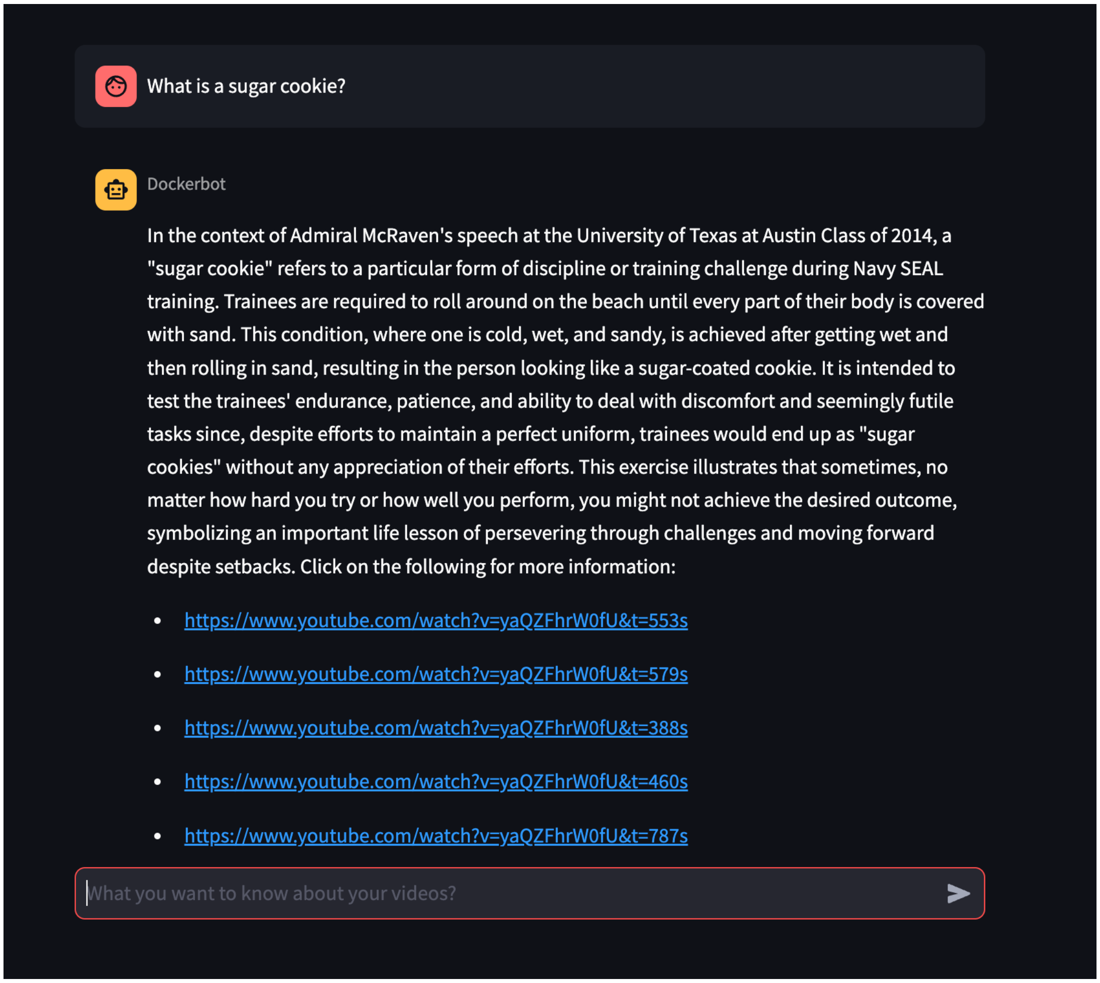
Processed videos:

Here are the videos you have processed so far:

You can download the transcription of the video by clicking on the corresponding video below

Admiral McRaven addresses the University of Texas at Austin Class of 2014

8. Ask the Dockerbot a question about a video that was processed by the yt-whisper service.



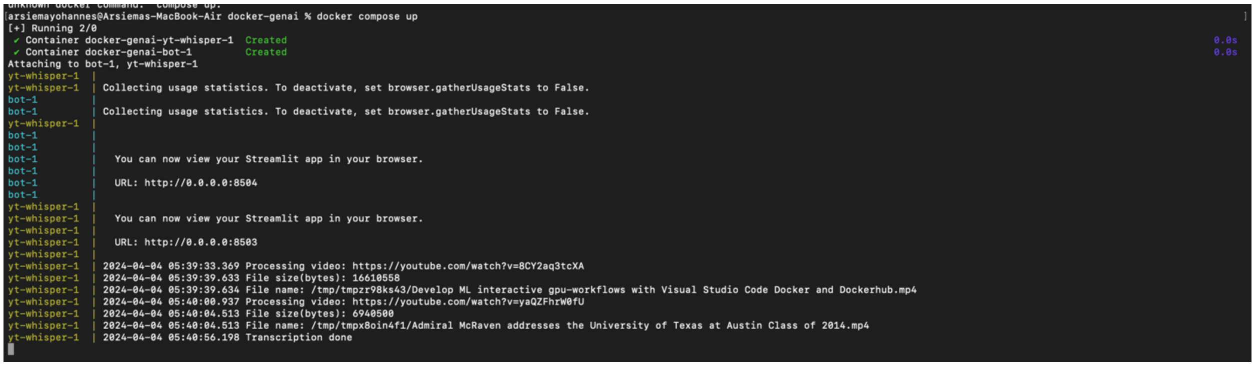
What is a sugar cookie?

Dockerbot

In the context of Admiral McRaven's speech at the University of Texas at Austin Class of 2014, a "sugar cookie" refers to a particular form of discipline or training challenge during Navy SEAL training. Trainees are required to roll around on the beach until every part of their body is covered with sand. This condition, where one is cold, wet, and sandy, is achieved after getting wet and then rolling in sand, resulting in the person looking like a sugar-coated cookie. It is intended to test the trainees' endurance, patience, and ability to deal with discomfort and seemingly futile tasks since, despite efforts to maintain a perfect uniform, trainees would end up as "sugar cookies" without any appreciation of their efforts. This exercise illustrates that sometimes, no matter how hard you try or how well you perform, you might not achieve the desired outcome, symbolizing an important life lesson of persevering through challenges and moving forward despite setbacks. Click on the following for more information:

- https://www.youtube.com/watch?v=yaQZFhrW0fU&t=553s
- https://www.youtube.com/watch?v=yaQZFhrW0fU&t=579s
- https://www.youtube.com/watch?v=yaQZFhrW0fU&t=388s
- https://www.youtube.com/watch?v=yaQZFhrW0fU&t=460s
- https://www.youtube.com/watch?v=yaQZFhrW0fU&t=787s

9. Docker Compose is a tool for defining and running multi-container applications. Compose makes it easy to run this application with a single command.



```
unknown docker command: "compose up"
arsiemayohannes@Arsiemas-MacBook-Air docker-genai % docker compose up
[+] Running 2/0
 ✔ Container docker-genai-yt-whisper-1  Created                                                              0.0s
 ✔ Container docker-genai-bot-1         Created                                                              0.0s
Attaching to bot-1, yt-whisper-1
yt-whisper-1  |
yt-whisper-1  | Collecting usage statistics. To deactivate, set browser.gatherUsageStats to False.
bot-1         |
bot-1         | Collecting usage statistics. To deactivate, set browser.gatherUsageStats to False.
yt-whisper-1  |
bot-1         |
bot-1         |
bot-1         |   You can now view your Streamlit app in your browser.
bot-1         |
bot-1         |   URL: http://0.0.0.0:8504
bot-1         |
yt-whisper-1  |
yt-whisper-1  |   You can now view your Streamlit app in your browser.
yt-whisper-1  |
yt-whisper-1  |   URL: http://0.0.0.0:8503
yt-whisper-1  |
yt-whisper-1  | 2024-04-04 05:39:33.369 Processing video: https://youtube.com/watch?v=8CY2aq3tcXA
yt-whisper-1  | 2024-04-04 05:39:39.633 File size(bytes): 16610558
yt-whisper-1  | 2024-04-04 05:39:39.634 File name: /tmp/tmpzr98ks43/Develop ML interactive gpu-workflows with Visual Studio Code Docker and Dockerhub.mp4
yt-whisper-1  | 2024-04-04 05:40:00.937 Processing video: https://youtube.com/watch?v=yaQZFhrW0fU
yt-whisper-1  | 2024-04-04 05:40:04.513 File size(bytes): 6940500
yt-whisper-1  | 2024-04-04 05:40:04.513 File name: /tmp/tmpx8oin4f1/Admiral McRaven addresses the University of Texas at Austin Class of 2014.mp4
yt-whisper-1  | 2024-04-04 05:40:56.198 Transcription done
```