



Key2Dream

Bárbara Muniz
Data Analytics Ironhack Bootcamp

Index



ABOUT THE PROJECT



DEMO TIME!



WORKFLOW



DATABASE



TECHNICAL APPROACH



NEXT STEPS

A surreal, dreamlike landscape where buildings and trees have faces and brains for heads, set against a cloudy sky.

About the Project

About the Project



Motivation



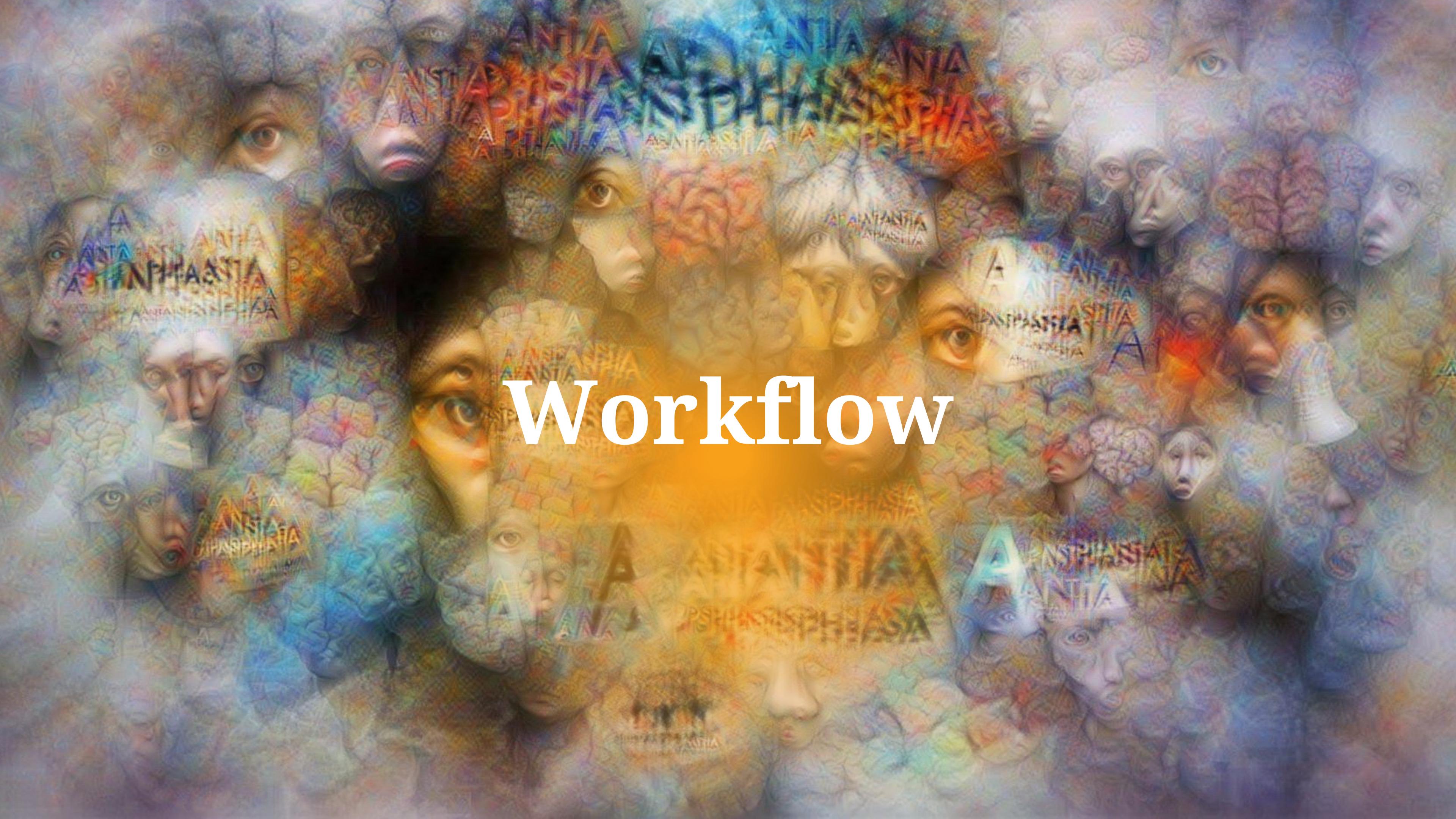
Main Idea: audio -> art!



Why this topic?



Demo Time!

A dense, colorful collage of numerous faces and text, with a central bright yellow glow.

Workflow

Workflow



Data Collecting



AI Models

- Text -> Painting: no CUDA needed (GPU accelerator);
- Text -> Summarization;
- Audio -> Text



Functions

- Connecting the functions



Flask

- Incorporating the functions;
- HTML



Database

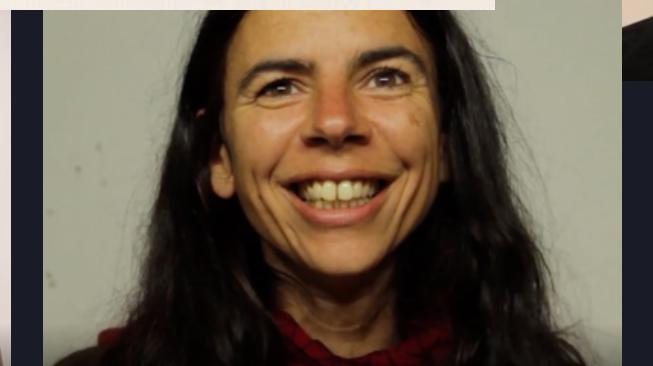
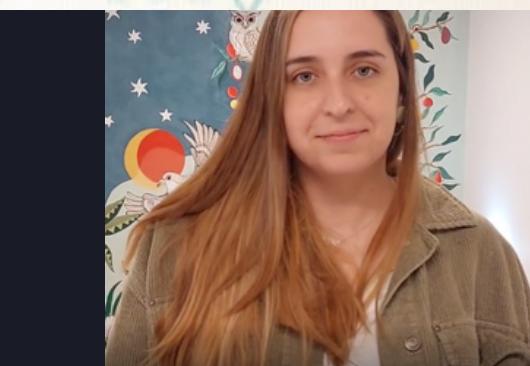
Database



Dreampire Website:
online video-based
dream;



Broader Public



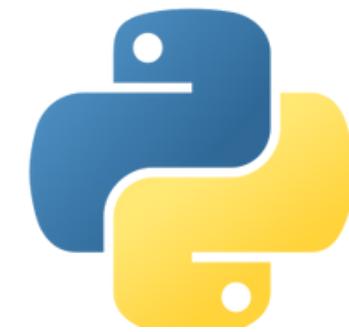


Technical Approach

AUTOMATION
ANIMATION

ANIMATION

Tools



Multiprocessing

GitHub



Flask



Transformers

Main Models/Libraries



AUDIO TO TEXT LIBRARY

Speech Recognition

TEXT SUMMARIZATION NLP MODEL

Abstractive Text Summarization
(ATS) : new sentences generated
HuggingFace Transformers library

TEXT TO IMAGE NEURAL NETWORK MODEL

- CLIP (Contrastive Language-Image Pre-training)
 - Aphantasia Repo
-

Why PTMs?

More flexible

Easier to implement

More general

Require less training time

High accuracy

What is CLIP?

Neural Network Model

CLIP is the first multimodal (in this case, vision and text) model tackling computer vision and was recently released by OpenAI on January 5, 2021

Trained on 400,000,000 (image, text) pairs

This is the data that is used in training the CLIP model

Predict the most relevant text snippet, given an image

Input an image into the CLIP model, and it will return the likeliest caption or summary of that image.

Zero-shot capabilities

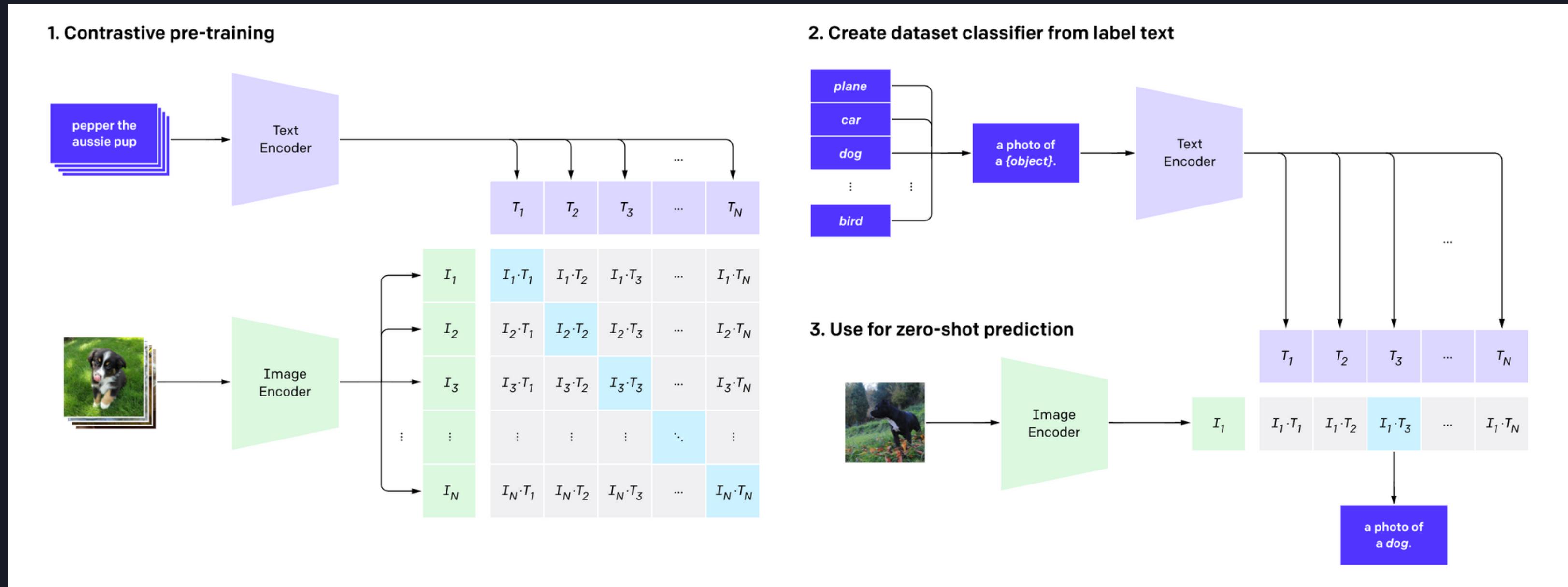
Perform well on tasks they aren't directly trained to do, which is called "zero-shot learning."

Computer vision + Natural Language Processing

"the ability for a computer to see and understand what it sees in manner similar to humans." + "the ability for a computer to understand language in manner similar to humans."

CLIP

Encoder Sub-models



CLIP - Art Style



A surreal, dreamlike landscape where human faces and trees are intertwined with a city skyline. The scene is filled with numerous faces of different sizes, some appearing to be made of trees or foliage. In the center, a large face with a yellow-to-orange gradient has a prominent brain-like texture on its forehead. The background features a dense cityscape with buildings and a park-like area with trees and a path.

Results



Next Steps



Next Steps

MAKE IT FASTER!

Have the computational power to make the model faster.

TRANSLATION

Make it possible to tell the dream in different languages.

GIF OF RESULTS

User can download the GIF at the website.

DEPLOY

Make it available for public access.

PROUD OF:

- Fun, artistic and innovative product
- Learning through struggle
- First time using Deep Learning and exploring HTML



Thank You!

Guys, that's (almost) it :)