TODOs

**Data Collection**

* ~~Obtain new generated images~~
* ~~Obtain new artist images~~

**Coding**

* ~~Train model on more epochs (Jensen)~~
* ~~Process new data into 256x256 with modified crop script (Parnian)~~
* Feed new data into model and get results (Parnian)

**Screen Recording / Screenshots**

* ~~Intro pictures (AI or Real?) (Riley)~~
  + ~~Pick handful of good pictures~~
* ~~New data being loaded (Jensen)~~
  + ~~GPT-4 Prompt and image~~
  + [~~https://artmatch.ca/~~](https://artmatch.ca/) ~~browsing~~
* ~~Model being run on new data (Jensen)~~
* ~~Results (Riley)~~
  + ~~Confusion Matrix~~
  + ~~Results vs benchmark (study)~~
* ~~Saliency Maps (Riley)~~

**Script Writing**

* ~~Intro/motivation (Parnian)~~
* ~~Data collection & processing walkthrough (Jensen)~~
* ~~New data demo (Jensen)~~
* ~~Results (Riley)~~
* ~~Interpretation & Conclusion (Riley)~~

**Voiceovers**

* Intro/motivation (Parnian)
* Data collection & processing walkthrough (Jensen)
* New data demo (Jensen)
* Results (Riley)
* Interpretation & Conclusion (Riley)

SCRIPT

**Introduction**

Hi!  
I’m Parnian, and here are my teammates, Riley and Jensen. We’re here today to introduce our handy model for evaluating human versus AI artwork

With recent AI developments in the world of art, some of which are so good that many experts have trouble telling whether a work was created by humans or generated by AI,

the boundaries are so vague. Let’s say you wanna support an artist… are you even sure those paintings are actually done by him or her?

So what makes human made paintings so valuable? More than the artwork itself, its value comes from the great details put into the work, the story behind it, the time that the artist took to create it; its originality.

We need a tool for preserving Authenticity and Integrity in Art. something to protect the value of a human artist against some generated codes by AI…  
  
That's exactly why we asked AI for help!

So, we literally made an AI model to detect whether an artwork is created by a human or is AI-generated.

To show you how difficult it can be, let’s look at some examples here. So, how was your performance?

Now, let’s see how our model actually works.

**Background: Data**

Now, let’s talk about data collection. We sourced these artworks from WikiArt, a well-known and comprehensive online art source. WikiArt was chosen for its vast and diverse collection of paintings made by artists across many styles, periods, and groups.  
  
Our AI-generated images are from DiffusionDB, the first large-scale repository of AI-generated images. These artworks were created with stable diffusion, a state-of-the-art generative model, presenting a challenging counterpart to the wikiArt.

However, diffusion db contains 14 million images, and many of them do not look like traditional artworks. Luckily, diffusionDB also contains the prompts used to generate the images, allowing us to change the dataset to something more comparable to wikiArt. To achieve this, we only chose images that had a wikiart artist name in the prompt, and filtered out images with prompts containing words like “photoreal”, Or “3D”.

Given the diverse nature of our datasets, these artworks vary in resolution and aspect ratio, so we needed to standardize them. To do this, we cropped and resized them to get a consistent 256x256 image size.

And now we will demonstrate the exciting capabilities of our model on data from brand new sources. We’ve selected two new images,

**Human art demo:**

First is an artwork from an independent Canadian Artist in Edmonton, found on the Canadian art purchasing website ArtMatch.ca.

**ChatGPT4 demo:**

The second one is generated by ChatGPT-4 in Van gogh style, using the state-of-the-art image generation of Dall E 3.

**Live demo:**

First up, we uploaded these new paintings to our platform on Google Colab.

Now, let's look at the first painting, the one by the artist from Edmonton.

And there we have it! It;s identified as human-created art, with accuracy of over 99%.

Next, we'll test our model with the painting generated by ChatGPT-4.

And the result is in! The model successfully recognizes this image as AI-generated, with a remarkable confidence level of over 98%.

**Performance**

To properly understand performance, usually it’s good to establish a baseline to compare to. In ML tasks, human performance can be a good benchmark, and we have exactly that thanks to a study on AI art conducted last year. They asked participants to identify whether a piece of art was made by a human or an AI.

As you can see, we do pretty well at identifying human made art pieces, but AI generated art is able to fool us the majority of the time. Overall, we’re left with a benchmark accuracy of just 63%.

So how does this compare to our model? In unseen data from the testing set, our model is just as good at identifying human-made art as us, and almost never mistakes AI art as human art. This leaves us with an overall test accuracy score of 86.5%, far surpassing human performance.

While this performance is great, what’s more interesting is what the model really learned, and if there’s anything that we can learn from it?

**Analysis**

To answer this question, we’re going to use saliency maps. The basic idea is to figure out which parts of the input have the most influence on the output. Say with this image we want to know why our model predicted a dog and not a cat, or a tree for example. We start by taking our input image, feeding it into our model, and then getting back our output like normal.

Now we back-propagate this output through our network, all the way back to the input image, where we get the gradients of each input pixel.

This is our saliency map, with the bright spots showing us which parts of the input had the most influence on the output. Here it’s pretty clear that the model identifies the shape of the dog.

So what do our model’s saliency maps show us?

We have our input image on the left, the saliency map in the middle, and the map overlaid on the input over on the right.

Our model correctly identifies this picasso painting, and it seems to be paying attention to the edges of the object, like the bottom of this fruit bowl, and the contrasting colors of this other object on the right.

Next, looking at this correctly identified AI art, our model seems to be picking up on the curves of the vase, and the bottom edge of the tablecloth.

So maybe our model is picking up on the different ways to draw shapes. But why the vase and not the apple, or something else?

Take this picture for example. It’s probably pretty obvious to us this is AI-generated.

Just look at Kermit's face, his missing hand, or this strange car thing in the background.

What about our model?

From the saliency map, it picks up on completely different parts of the input, basically ignoring kermit altogether and focusing instead on the car's hood, its wheels, and some other seemingly random details.

So our model can distinguish between real and AI art, and we can explain how it does it.

However, understanding the why, the interpretation, is much more challenging.

**Conclusion**

This experiment highlights a crucial lesson about relying on neural networks.

Although we have ways to explain the model’s decisions and make them more transparent, interpreting their decisions using our own perspective is tricky.

Models are not humans, and they do not see the world the way we do.

So while problems like telling human and AI art apart may be simple for a neural network, determining whether or not neural nets truly understand the problem is much more complex.