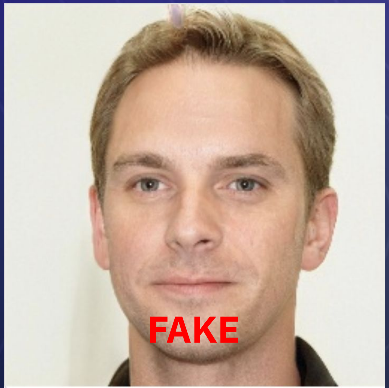
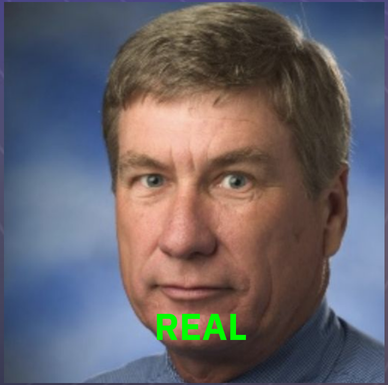
# Unmasking Deepfakes: How a Custom CNN Exposes AI-Generated Faces

Presented by Edric Ma

# DEVELOPING CRISIS...

- Rise of AI generated media
- Deep fake images
- Implications of such technology
- What are ways we can combat this issue?

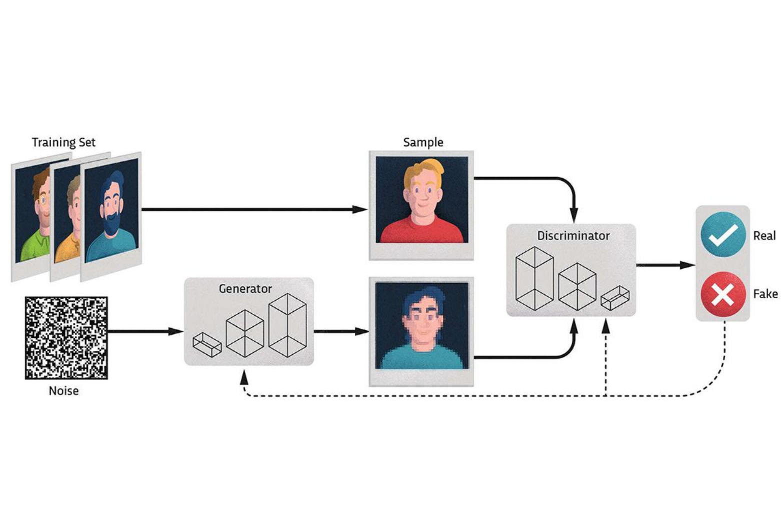


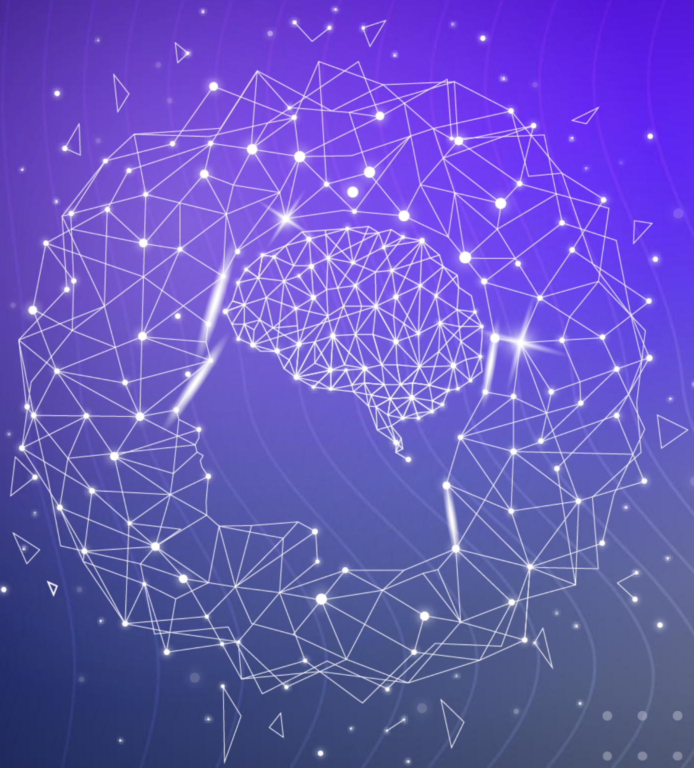


# Generative Adversarial Network (GAN)

GANs are a class of machine learning models that generate realistic synthetic data (e.g., images, audio) by pitting two neural networks against each other.

Generator vs. Discriminator





# Convolutional Neural Network (CNN)

A CNN is a machine learning model that analyzes images by detecting patterns like edges and shapes in small sections, then combines them to recognize objects or features.

# Technology Limitations

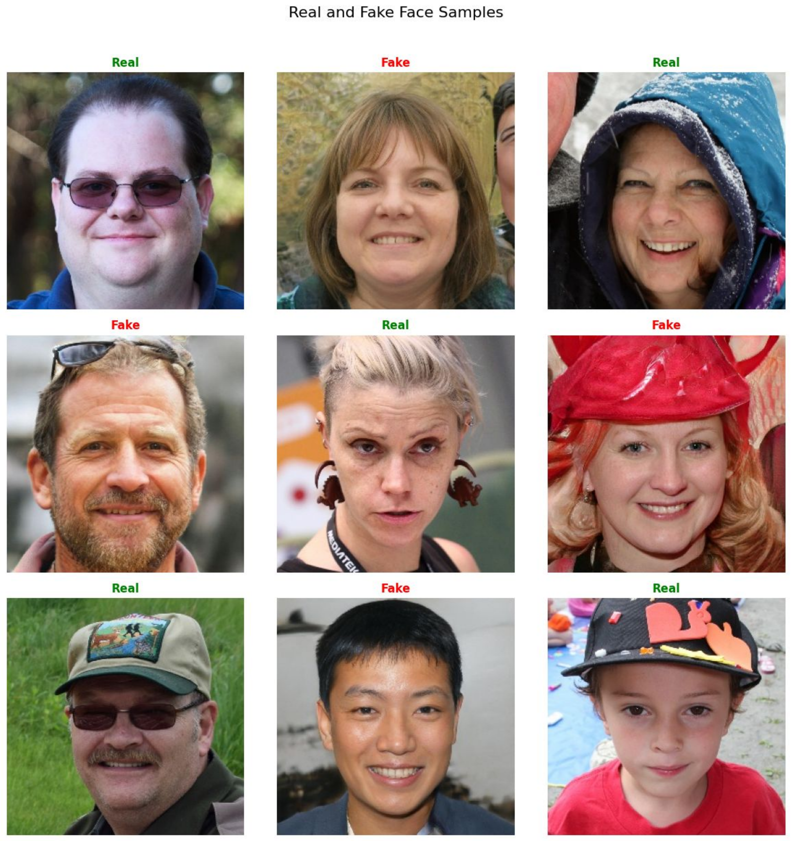
My specifications and applications:

- 2022 Macbook Air (Apple M2 Chip)
- 8GB Memory
- Google Colab (CPU Only)

# THE DATASET

70,000 real face images

70,000 fake face images



# TESTING ON SUBSETS

01 **EfficientNetB0**

50M to achieve 50% validation accuracy

02 **EfficientNetB5**

8H15M to achieve 54% validation accuracy

03 **Custom CNN**

1H45M to achieve 79% validation accuracy

# CUSTOM CNN

**3 convolutional layers**

Filter Numbers (32→64→128)

3x3 Kernel Size

**Max pooling layer**

**Batch normalization**

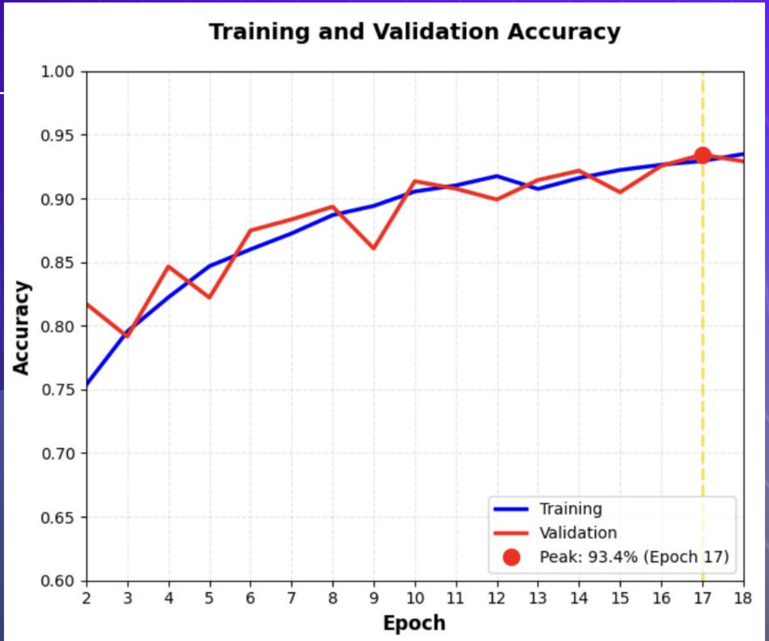
**Dense layer**

256 units

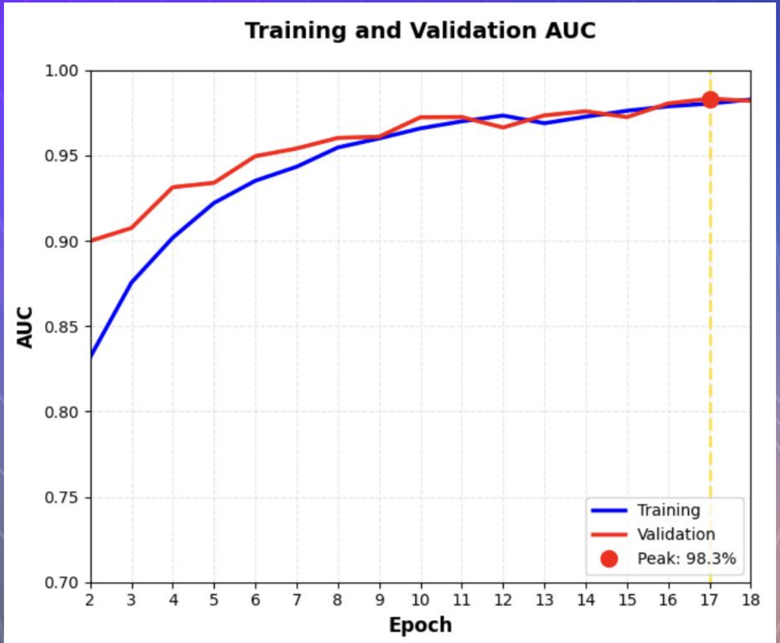
**Dropout layer**

50%

# TRAINING RESULTS



Training and Validation Accuracy

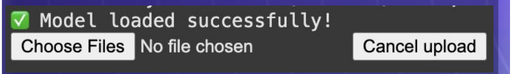


Training and Validation AUC

## 93.1% TEST ACCURACY

# Pending Product

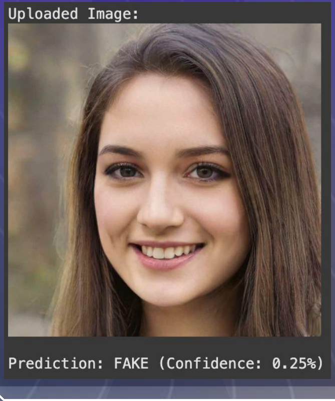
## UPLOAD



```
✅ Model loaded successfully!
```

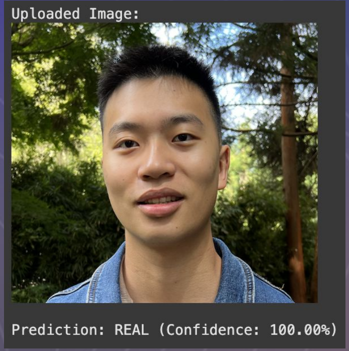
Upload a JPEG file for the model to classify whether it is real or fake

## RESULT: FAKE



```
Uploaded Image:
Prediction: FAKE (Confidence: 0.25%)
```

## RESULT: REAL



```
Uploaded Image:
Prediction: REAL (Confidence: 100.00%)
```

How Confidence Scores Work:
The model outputs a value between 0 (100% fake) and 1 (100% real)

# THANK YOU FOR LISTENING!

# External Resources

## Media:

- https://www.sciencefocus.com/future-technology/how-do-machine-learning-gans-work
- https://slidesgo.com/theme/korean-ai-agency-pitch-deck#search-ai&position-11&results-1421&rs=search
- https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces/data

## Additional Research:

- https://aws.amazon.com/what-is/gan/
- https://www.ibm.com/think/topics/convolutional-neural-networks