RUFake

DJ Fajana, Jessica Jiang, Jasmine Lau,
Sanjiv Narayan, Tamar Brand-Perez

THE TEAM

SANJIV NARAYAN
Roles: Product Manager

TAMAR BRAND-PEREZ
Role: Project Manager

DJ FAJANA
Roles: Infrastructure & Application Developer

JESSICA JIANG
Role: ML Engineer

JASMINE LAU
Role: Data Lead & Support Engineer

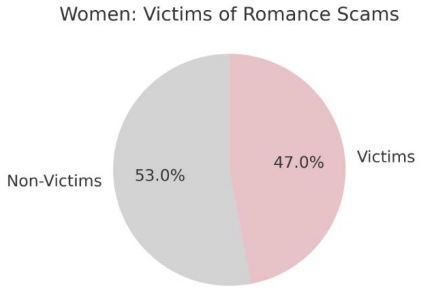
ONE TEAM MEMBER MISSING TODAY FOR . . .

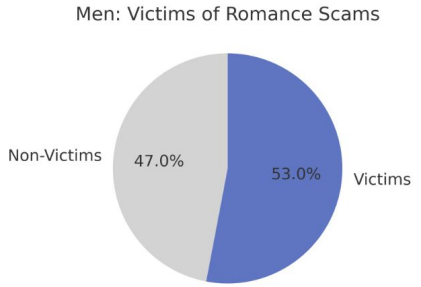
NIKKI MACLEOD

ALLA MORGAN

PROBLEM

- Online scamming is a $500B global challenge
  - Causes financial ruin and emotional distress
  - 53% of men & 47% of women are victims of romance scams
- Users require a trustworthy solution to detect scamming and protect them





images by ChatGPT, 2025

OVERVIEW

RUFake addresses several unmet needs

- An ethical online tool with > 85% accuracy to detect real vs fake photos of individuals
- Use cases:
  1. Online dating
  2. Online ID checks

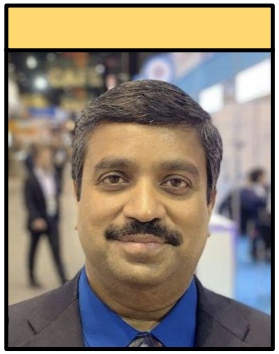
We prioritize privacy - Nobody will know you are checking
We prioritize security - We don't store your profile or photos

IMPACT

- Our AI was built from millions of photos in broad, diverse populations
    - Reduce catfish attempts, emotional distress, and financial losses
- Our market surveys showed that our value proposition was impactul in individuals of all ages including data scientists

DOMAIN EXPERTS

Dr. Vasha Dutell

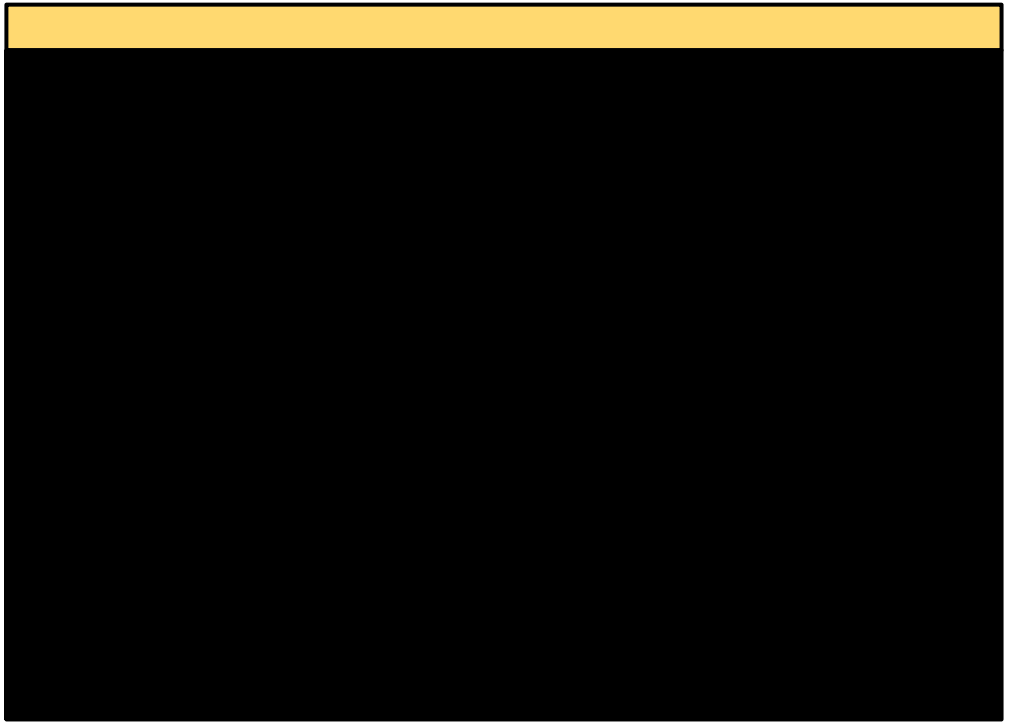


Dr. Senthil Periaswamy

Dr. Hany Farid

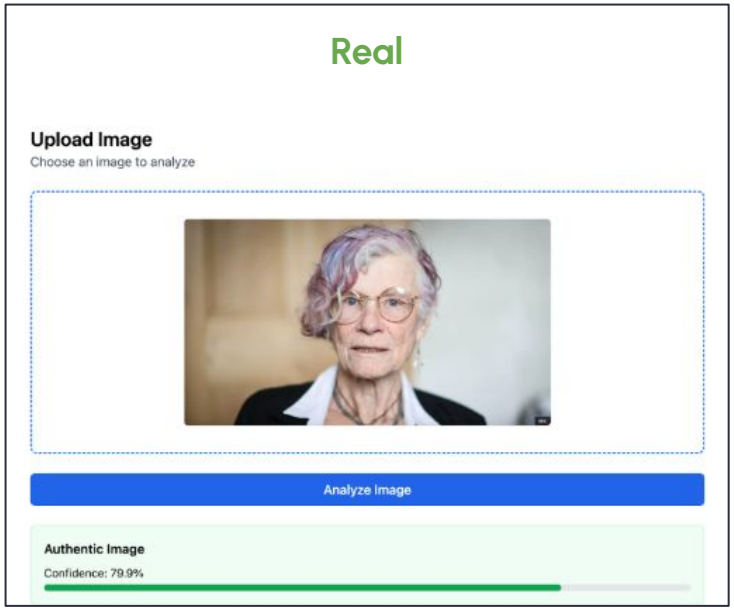
MVP

rufakeapp.com

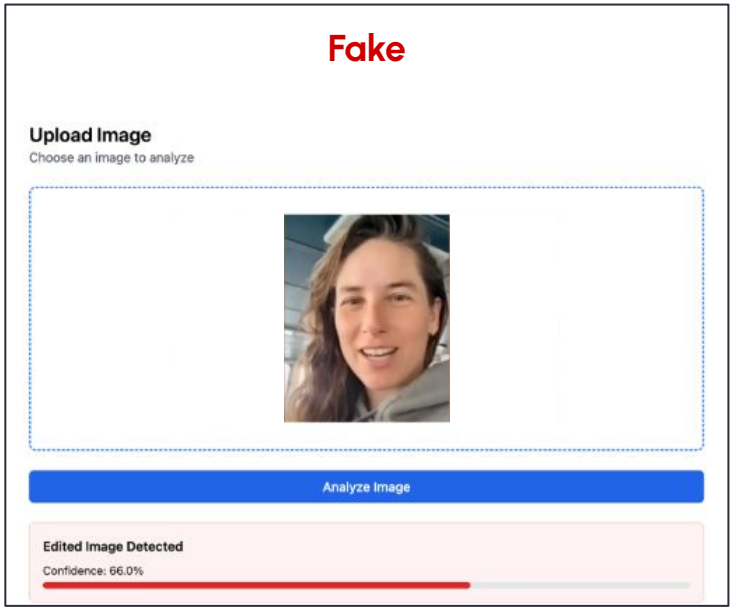


MVP

Real



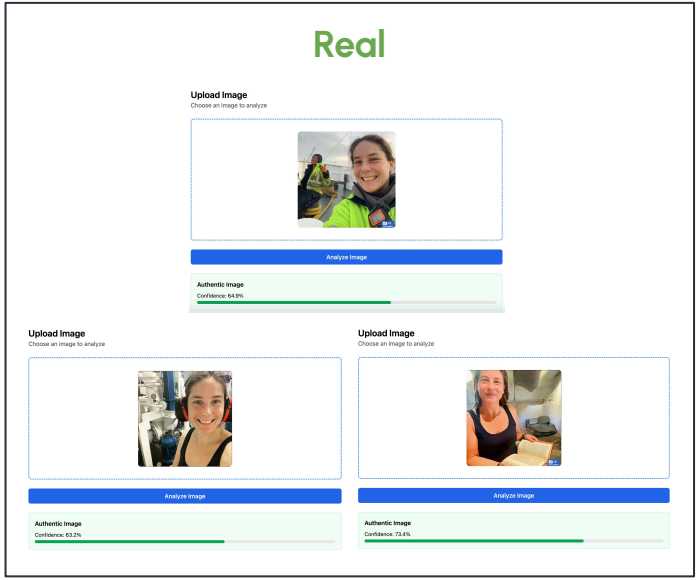
Fake



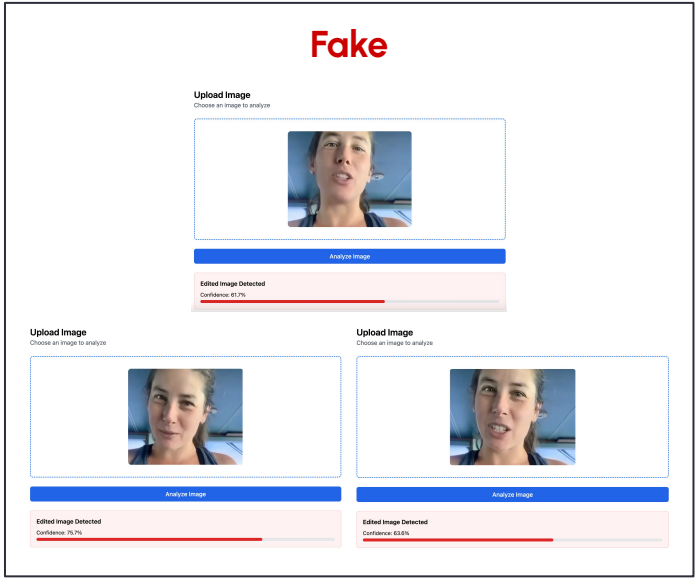
MVP

Scammers stole photos from an unsuspecting woman and generated deep fake videos

Real



Fake



MVP

Real

Upload Image
Choose an image to analyze

Analyze Image

Authentic Image
Confidence: 61.3%

Upload Image
Choose an image to analyze

Analyze Image

Authentic Image
Confidence: 61.3%

Upload Image
Choose an image to analyze

Analyze Image

Authentic Image
Confidence: 64.6%

Upload Image
Choose an image to analyze

Analyze Image

Authentic Image
Confidence: 63.8%

Upload Image
Choose an image to analyze

Analyze Image

Authentic Image
Confidence: 70.1%

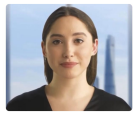
Fake

Upload Image
Choose an image to analyze

Analyze Image

Edited Image Detected
Confidence: 60.0%

Upload Image
Choose an image to analyze



Analyze Image

Edited Image Detected
Confidence: 52.0%

Upload Image
Choose an image to analyze

Analyze Image

Edited Image Detected
Confidence: 77.6%

Upload Image
Choose an image to analyze

Analyze Image

Edited Image Detected
Confidence: 57.9%

Upload Image
Choose an image to analyze

Analyze Image

Edited Image Detected
Confidence: 67.7%

USER FEEDBACK

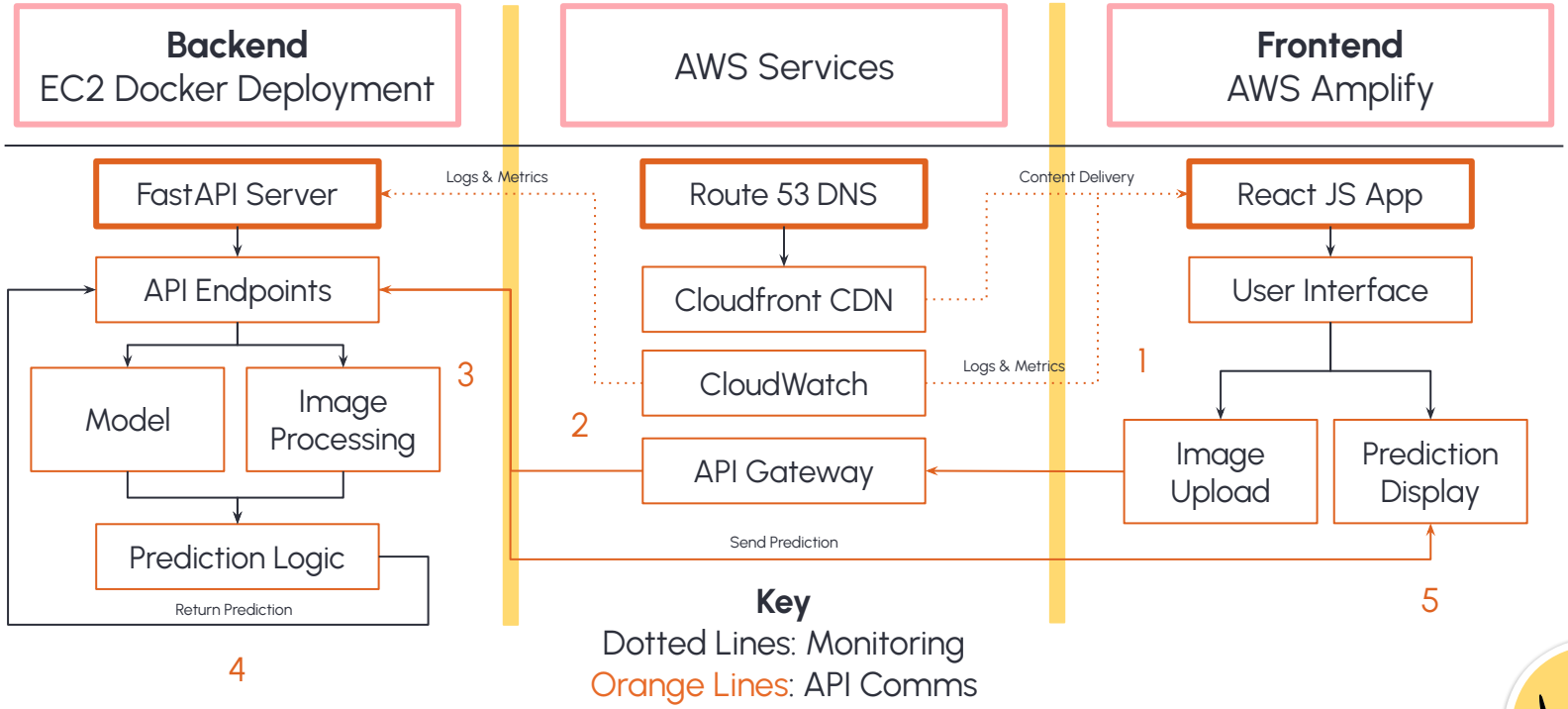
Usability
"The website is pretty easy and intuitive to use."

Purpose
"It's clear what you were going for with the website, it's meant to inform people about dating scams."

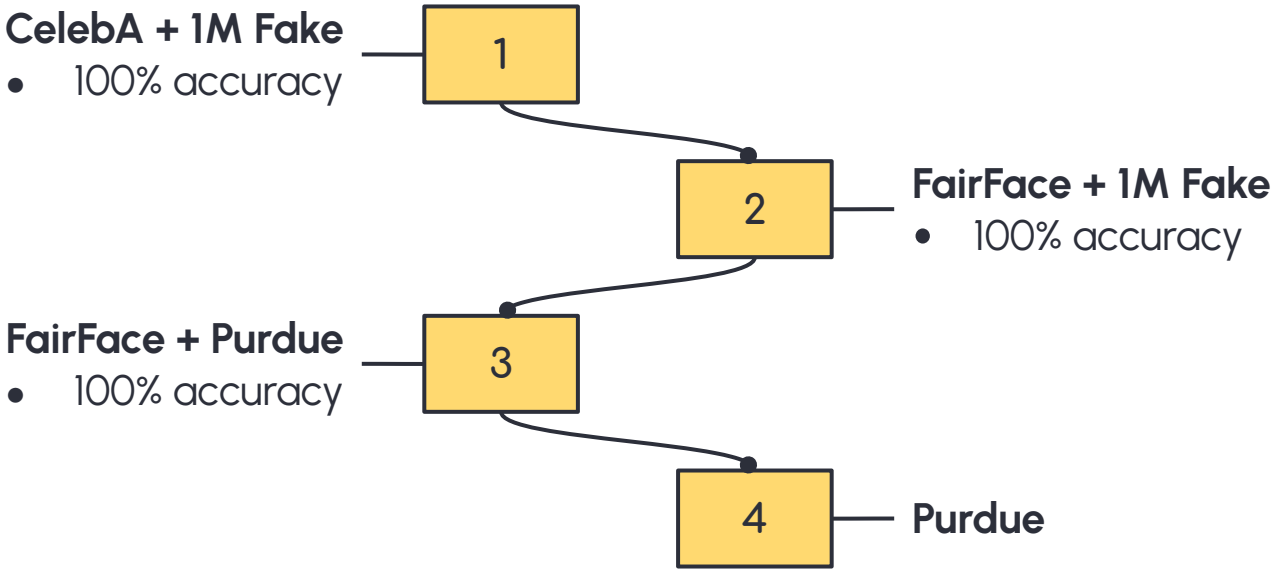
Improvements
"It would be great to be able to crop people to focus on one person in a picture at a time."

Design
"It's clean."

DATA PIPELINE

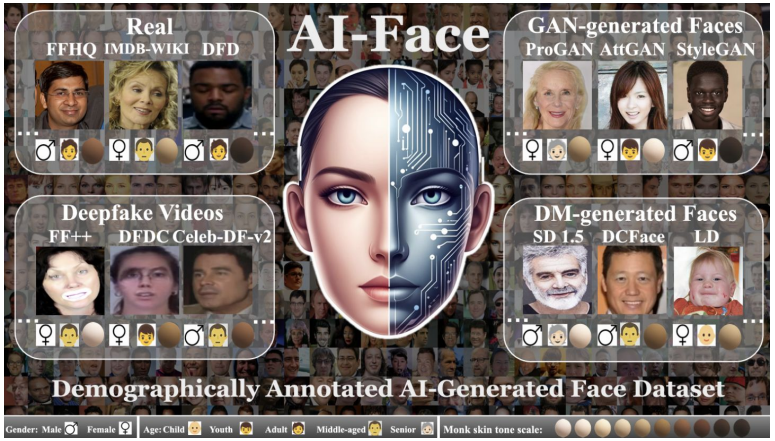


TIMELINE OF DATASETS & CHALLENGES



AI FACE FAIRNESS BENCH - PURDUE

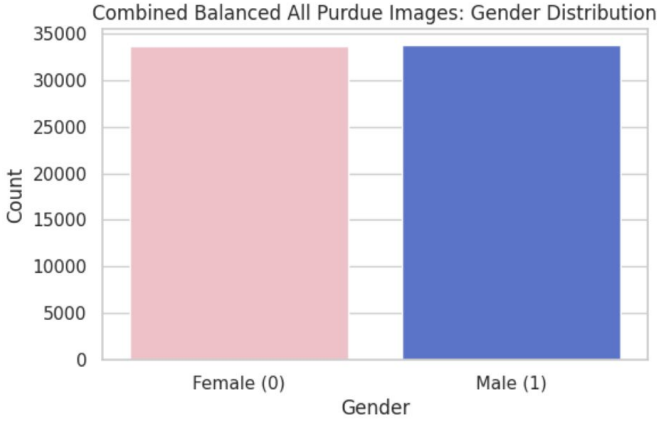
1. StyleGAN → 37 distinct generation methods
2. Pre-annotated for gender, age, and race
3. Real + Artificially generated images
4. Obtained permission from Purdue University

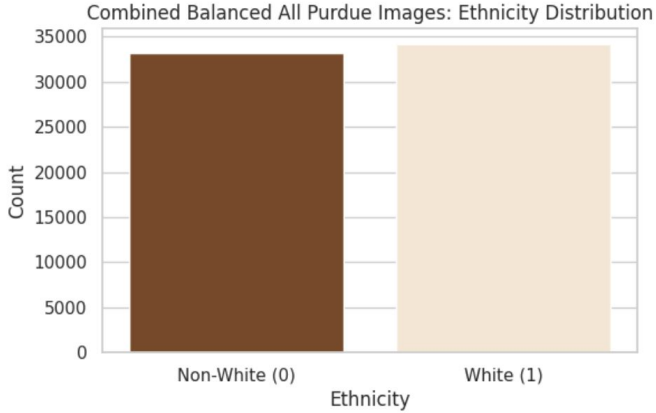


PRE-PROCESSING

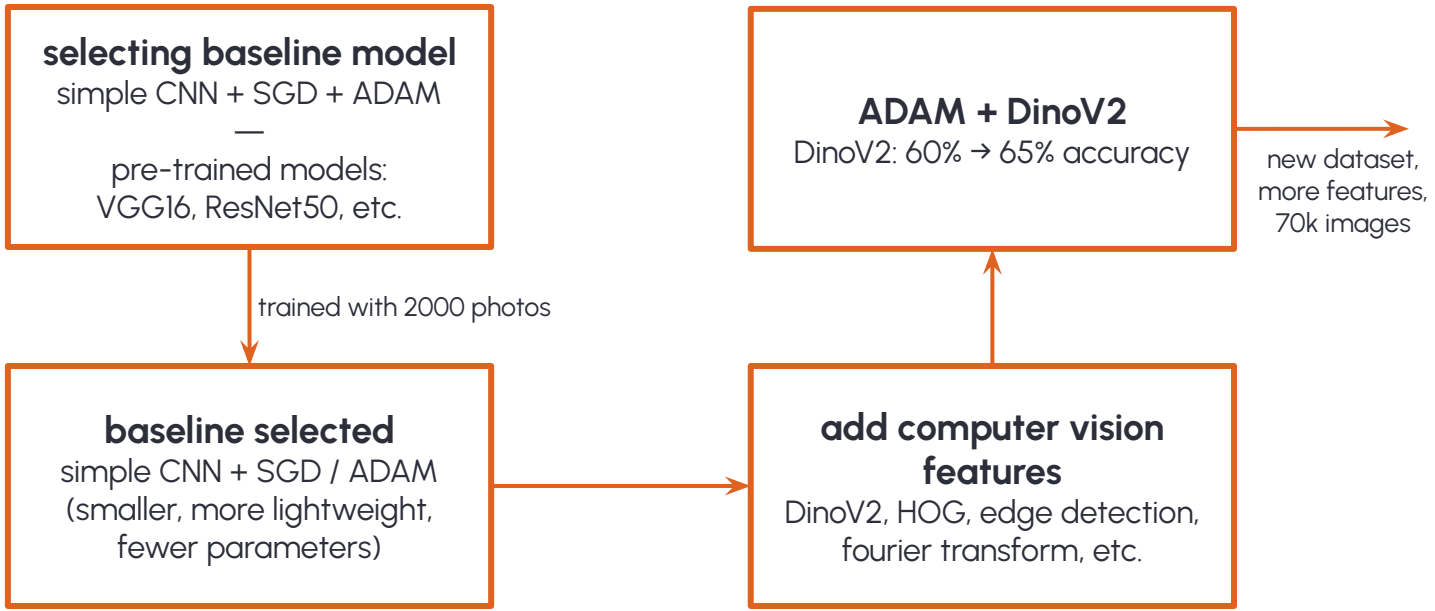
Cause	Samples Available (Post cause change)	Samples Changed for cause
total of 2 tar files	1,000,000	0
used 1 tar file	582,064	-417,936
random downsample to balance for label, gender and ethnicity	80,000	-502,064
images removed for path not found	67,350	-12,650
Final Real & Fake Images Count	**67,350**	

AI FACE FAIRNESS BENCH





CNN MODELING PROCESS



CNN MODEL

training
loss
accuracy

result
performance
precision
recall

evaluating at each step

building simple CNN

Conv2D | MaxPooling2D | ReLU | Sigmoid

simple CNN with optimizer

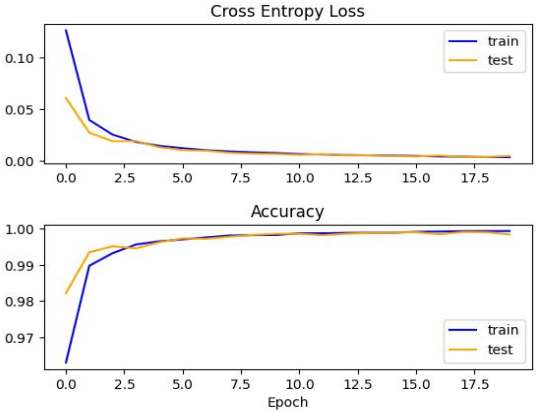
ADAM

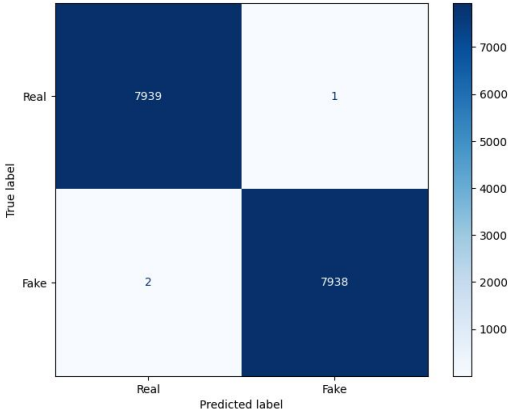
added CV features

DinoV2

EVALUATION & CHALLENGES

CNN training results with our 70k image dataset:





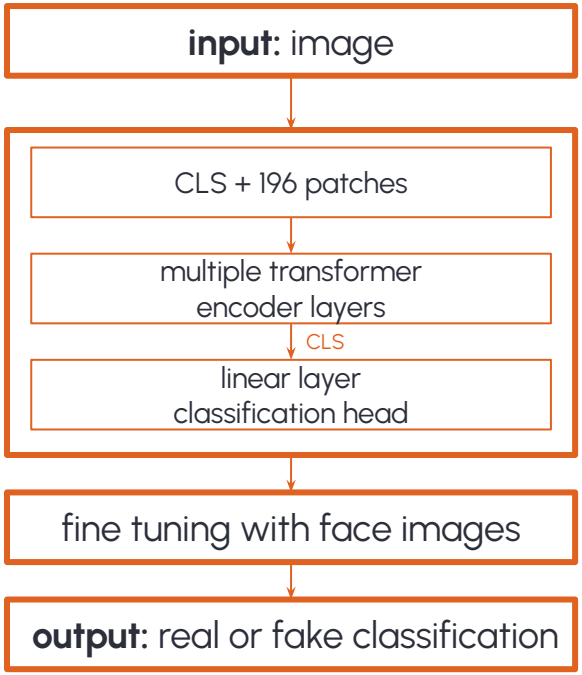
debug key steps recap:

1. avoid leakage by processing celebrity photos with k-means clustering
2. detect model logic issues by gradually increasing training sample size
3. new, large dataset from Purdue University
4. fingerprint injection worked, but time constraint

ViT BASE PATCH-16 224

Vision Transformer

- transformer encoder model
- captures global context
- patch size suitable for facial features
- pre-trained on ImageNet
- fine tuned on our data



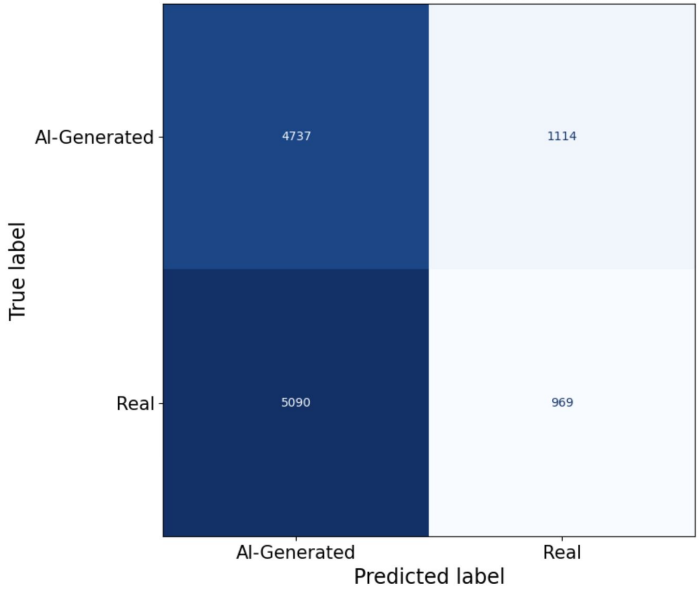
PERFORMANCE & ETHICAL EVALUATION

Pre Fine-Tuning Overall Accuracy on Validation Set: 0.4401

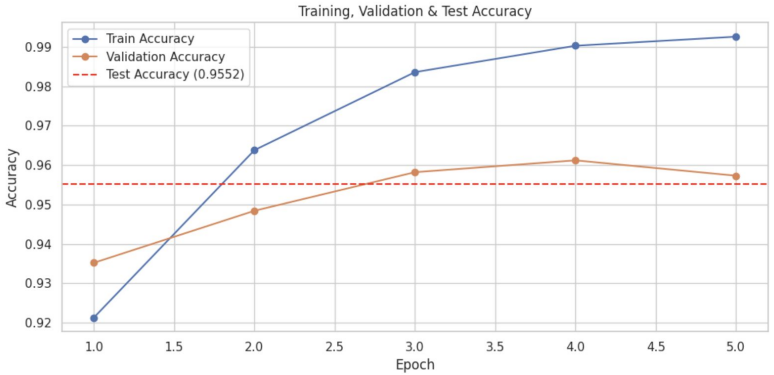
Accuracy by Gender:
Gender 0: 0.4787
Gender 1: 0.4051

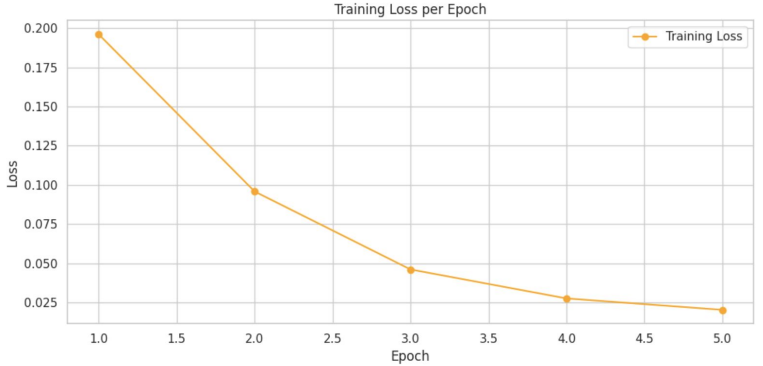
Accuracy by Ethnicity:
Ethnicity 0: 0.4390
Ethnicity 1: 0.4411

Combined Gender & Ethnicity Accuracy:
Gender 0 & Ethnicity 1: 0.4760
Gender 1 & Ethnicity 0: 0.4090
Gender 1 & Ethnicity 1: 0.4000
Gender 0 & Ethnicity 0: 0.4823



ViT HYPERPARAMETERS

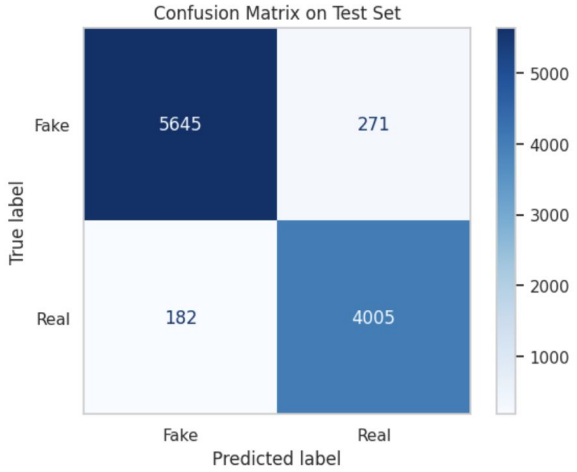




Learning Rate: 2e-5
Loss Function: Cross Entropy Loss
Epochs: 5
Batch Size: 32 images
Optimizer: AdamW (handles weight decay)

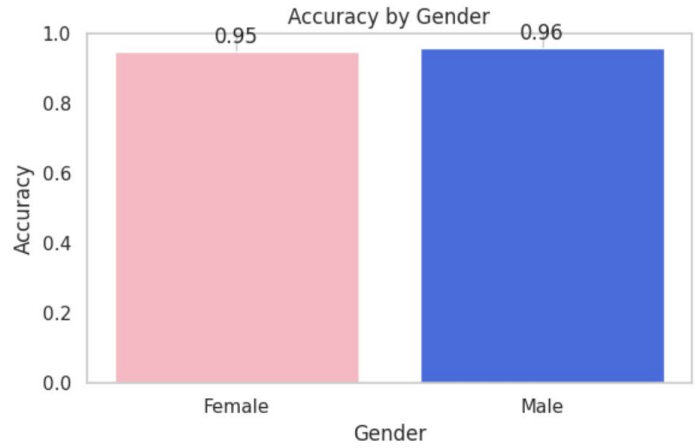
ViT RESULTS

```
Classification Report:
              precision    recall  f1-score   support

        Fake       0.97      0.95      0.96      5916
        Real       0.94      0.96      0.95      4187

    accuracy                           0.96     10103
   macro avg       0.95      0.96      0.95     10103
weighted avg       0.96      0.96      0.96     10103

Confusion Matrix:
[[5645  271]
 [ 182 4005]]
```

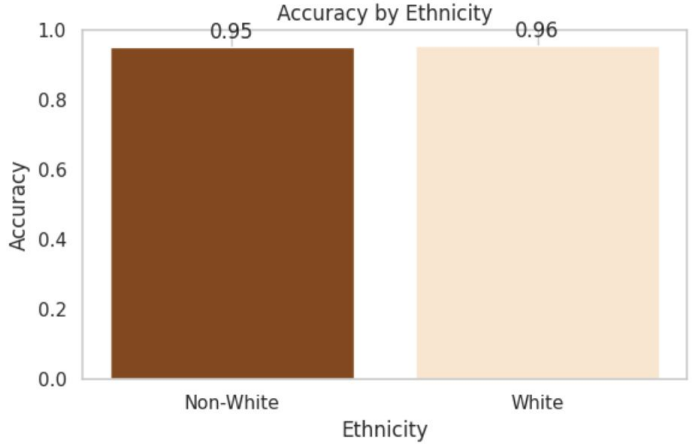


POST FINE-TUNING

```
Accuracy by Gender:
Gender 0: 0.9498
Gender 1: 0.9607
```

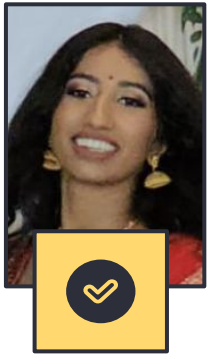



```
Accuracy by Ethnicity:
Ethnicity 0: 0.9526
Ethnicity 1: 0.9576
```

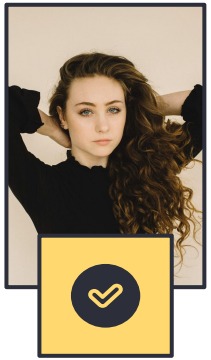


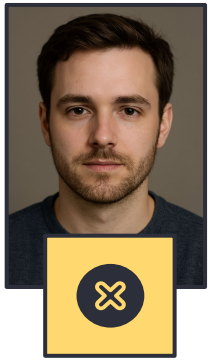
GENERALIZABILITY

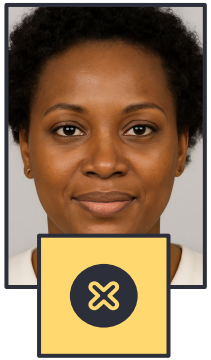
Real



Fake







RESULTS

1. Detection Accuracy & Fairness

- Achieved ≥ 85% accuracy in detecting AI-generated face images
- ≤ 10% difference in key performance metrics across demographic groups

2. User Experience & Speed

- Enable real-time image verification within 3 seconds for 95% of users, while ensuring no image storage

MISSION

To provide the detection of altered facial images with a user-friendly tool that embeds ethical safeguards and transparency, providing trust and confidence to users.

FUTURE WORK

- Use data with injected artificial fingerprints into real images (revisit FairFace)
- ↑ explainability of the model with attention maps
- ↑ generalizability of the model:
  - More diverse training data
  - Stopping rule
  - Changing model architecture to a sparse network
- User feedback & performance testing

THANK YOU

Capstone Professors
Computer Vision Professors
Dataset & Model Creators
Peers

https://www.rufakeapp.com/

Questions?