

Biometric systems 2025/2026

# **Evaluation of Vision Transformers for face verification under attacks and appearance variations**

Alessia Infantino 1922069



# Problem Statement

## Introduction

### Face Verification



**Face verification** is the process of confirming a person's identity by comparing two facial images and deciding if they belong to the same individual.

**Real-world conditions** can significantly affect verification performance. In addition, **spoofing attacks** represent a serious threat to system reliability.



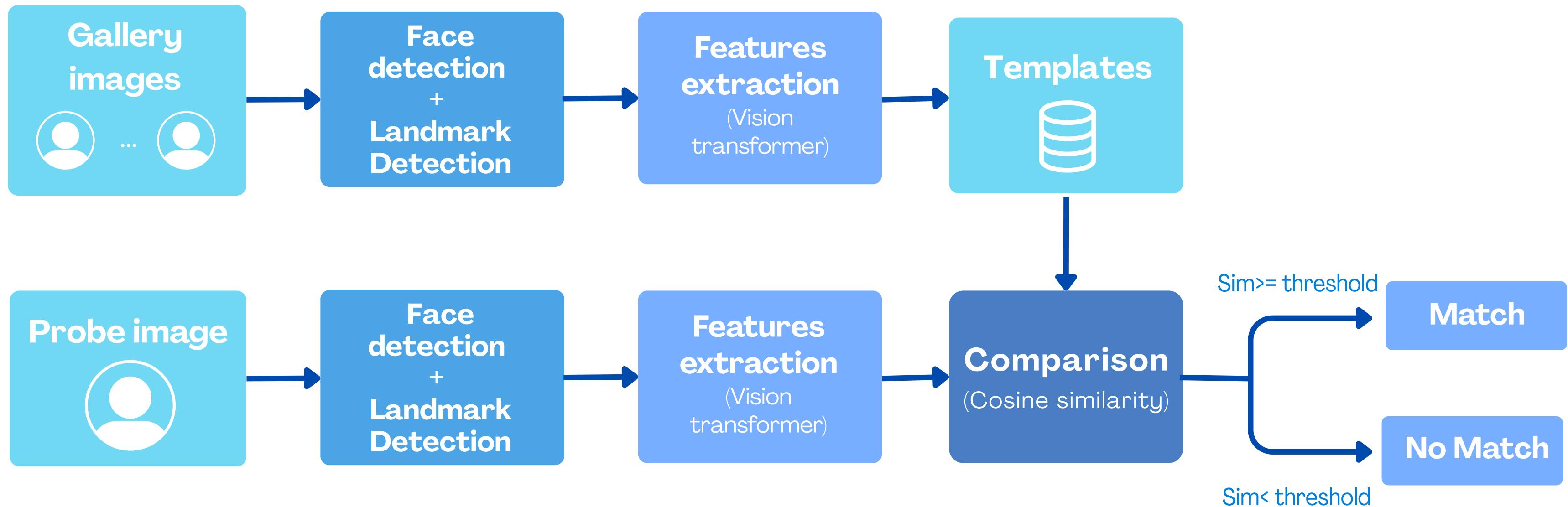
### Objectives



- **Analyze** a ViT-based verification model in real-world scenarios
- **Evaluate** performance on a custom dataset
- **Assess** resistance to attacks

# Project overview

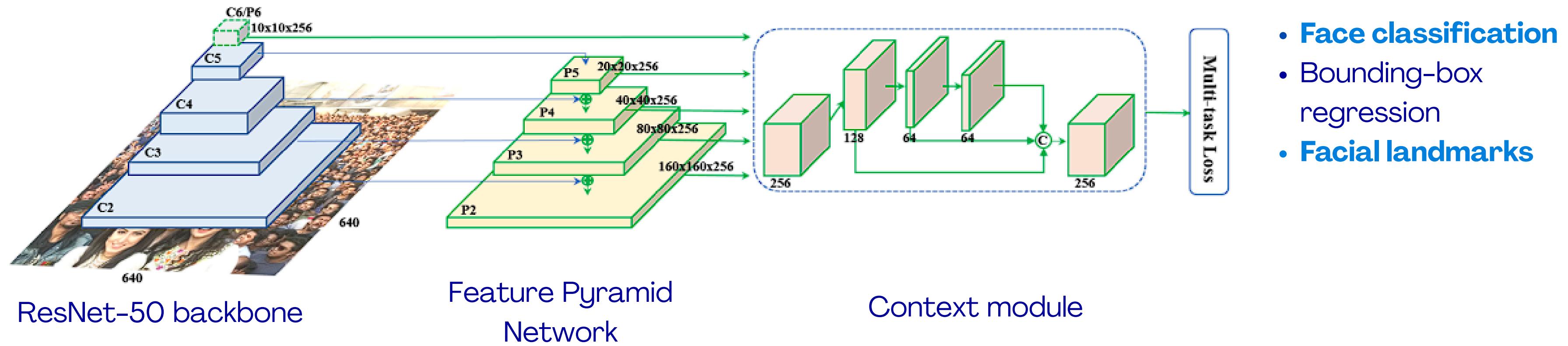
Full pipeline



# Model Architectures (1)

## RetinaFace

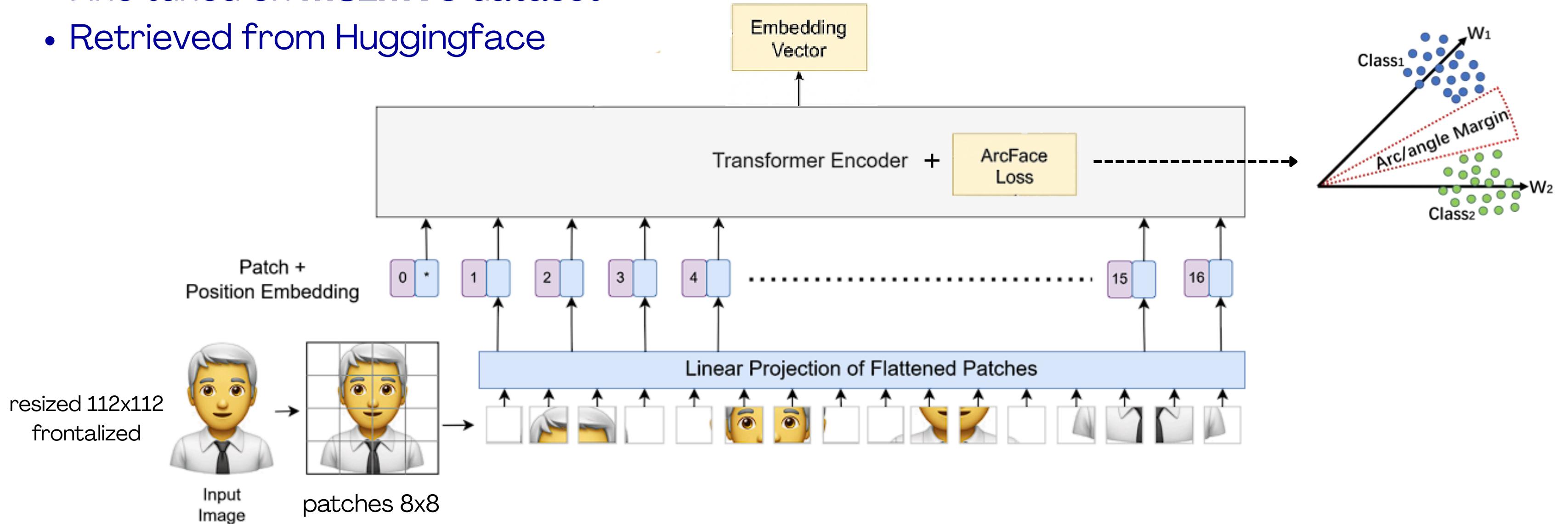
- Trained on **WIDER FACE** dataset
- **Anchor-based** training strategy



# Model Architectures (2)

ViT

- Fine tuned on **MS1MV3** dataset
- Retrieved from Huggingface



- **Arcface loss** for discriminative embeddings:

$$\mathcal{L} = \frac{-1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot \cos(\theta_{y_i} + m)}}{e^{s \cdot \cos(\theta_{y_i} + m)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}}$$

# Datasets (1)

## User variations

- Gallery: **212 images** (3 for each identity)
- Each category contains **12 images** (27 for interpersonal)
- Overlap categories



**Make-up**



**Age**



**Plastic surgery**

# Datasets (2)

Potential attacks



**Print attacks**



**Camouflage**



**Deepfake**



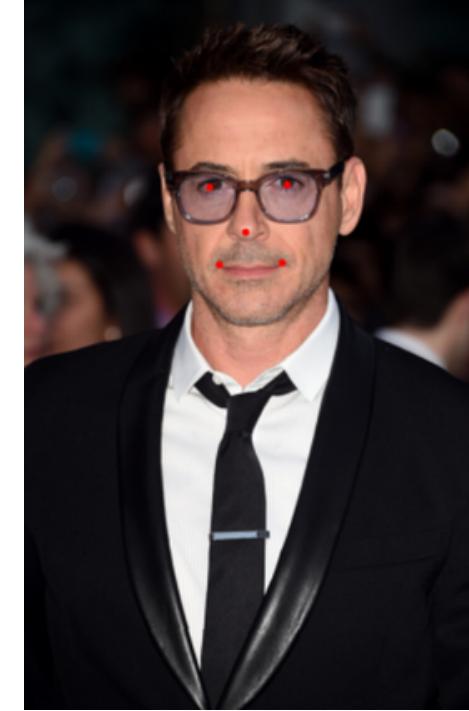
**Interpersonal**

# Proposed method (1)

## Step 1: Landmarks detection and Face alignment

- **Frameworks:** OpenCV, insightface library

1. **Face localization and landmarks extraction** (buffalo\_l model, RetinaFace-10GF)
2. **Matching and warping** with estimateAffinePartial2D and warpAffine
3. **Result:** normalised  $112 \times 112$  face

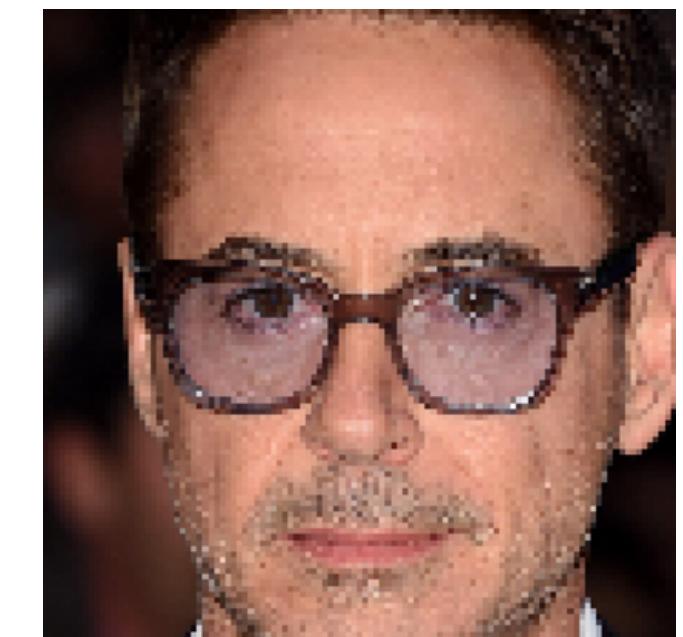


| LANDMARK    | X       | Y       |
|-------------|---------|---------|
| Left eye    | 38.2946 | 51.6963 |
| Right eye   | 73.5318 | 51.5014 |
| Nose tip    | 56.0252 | 71.7366 |
| Left mouth  | 41.5493 | 92.3655 |
| Right mouth | 70.7299 | 92.2041 |

Standard ArcFace 5-point template

Original image

Image with landmarks



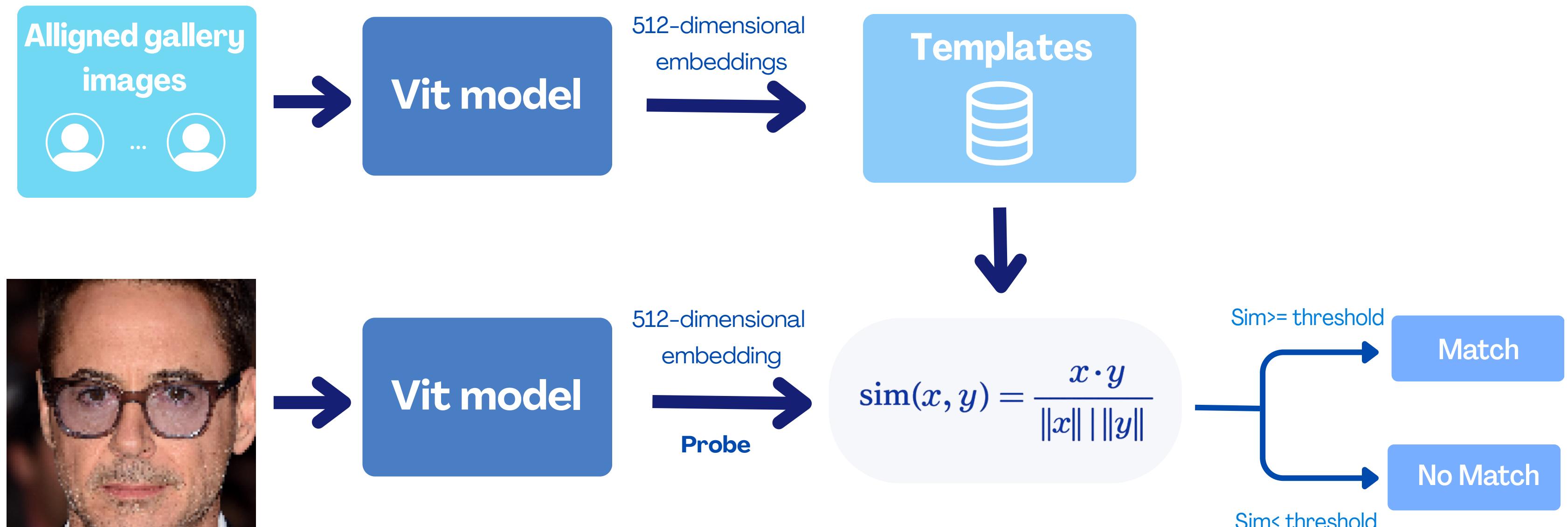
Aligned image (112x112)

\*same processing for gallery images

# Proposed method (2)

## Step 2: Features Extraction and Similarity Computation

- Thresholds from 0 to 1 (step of 0.5)



\*Probe vs Gallery comparisons are used to generate genuine and impostor verification trials

# Proposed method (3)

## Step 3: Evaluation

### Scores

- Genuine Acceptance
- False Acceptance
- Genuine Rejection
- False Rejection

### Graphics

- FAR
- FRR
- GAR
- GRR
- ERR
- ROC
- DET
- Margin

# Results (1)

## Detection and alignment

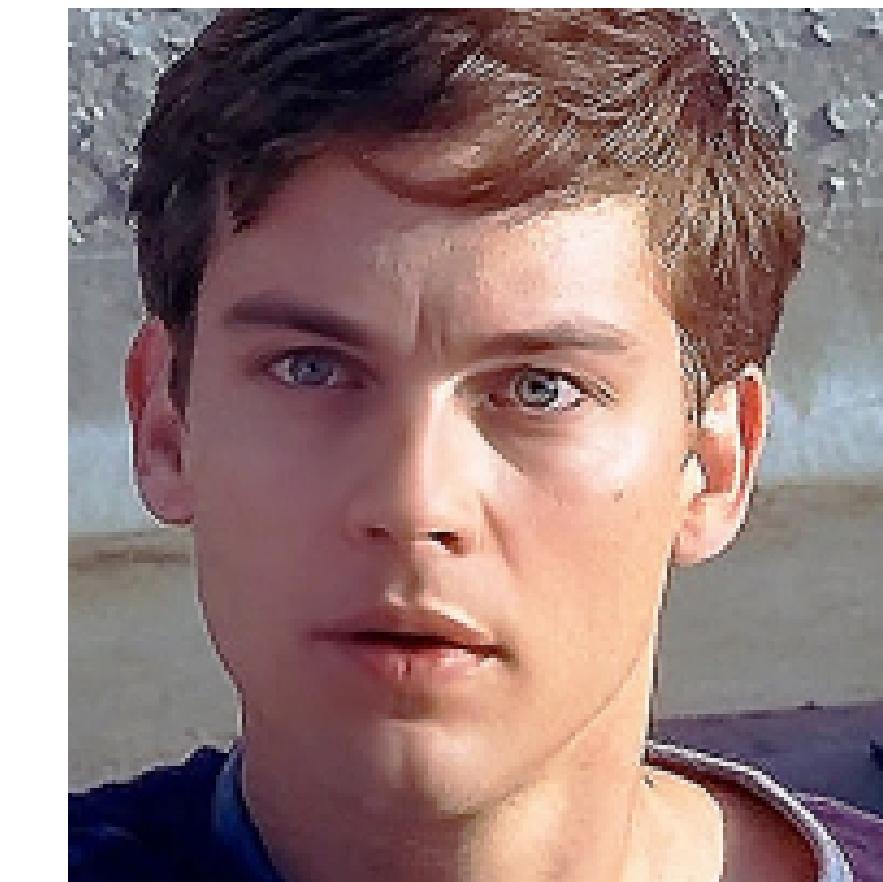
- Images where the face detector did not produce any bounding box:



(a)



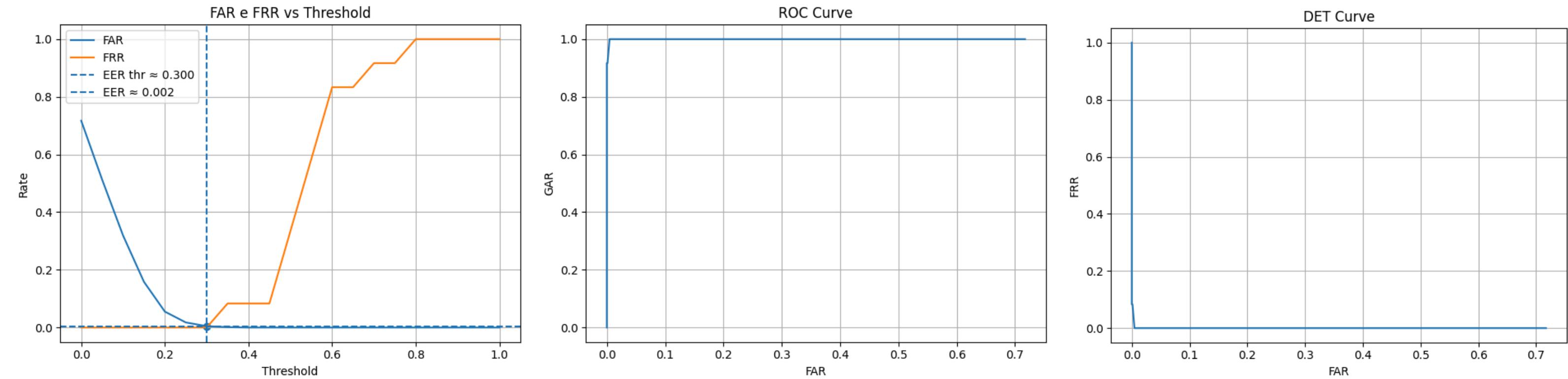
(b)



(c)

# Results (2)

## Age



Genuine acceptances:



Best Match: Zac Efron

Similarity score: 0.7523

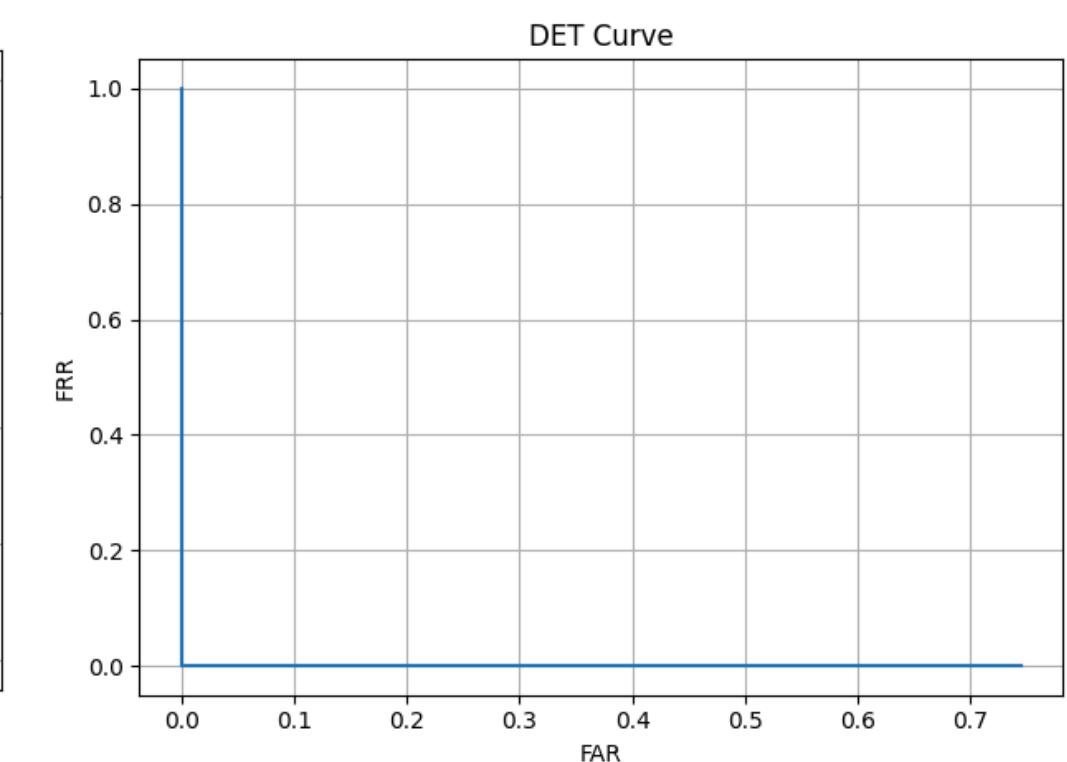
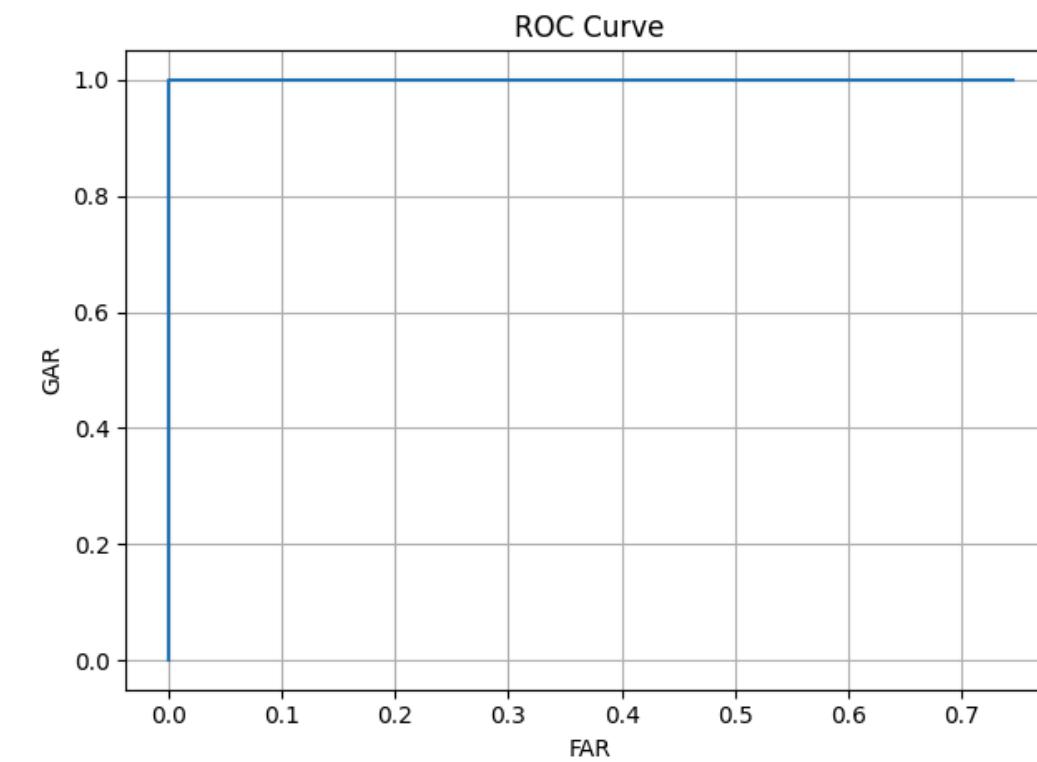
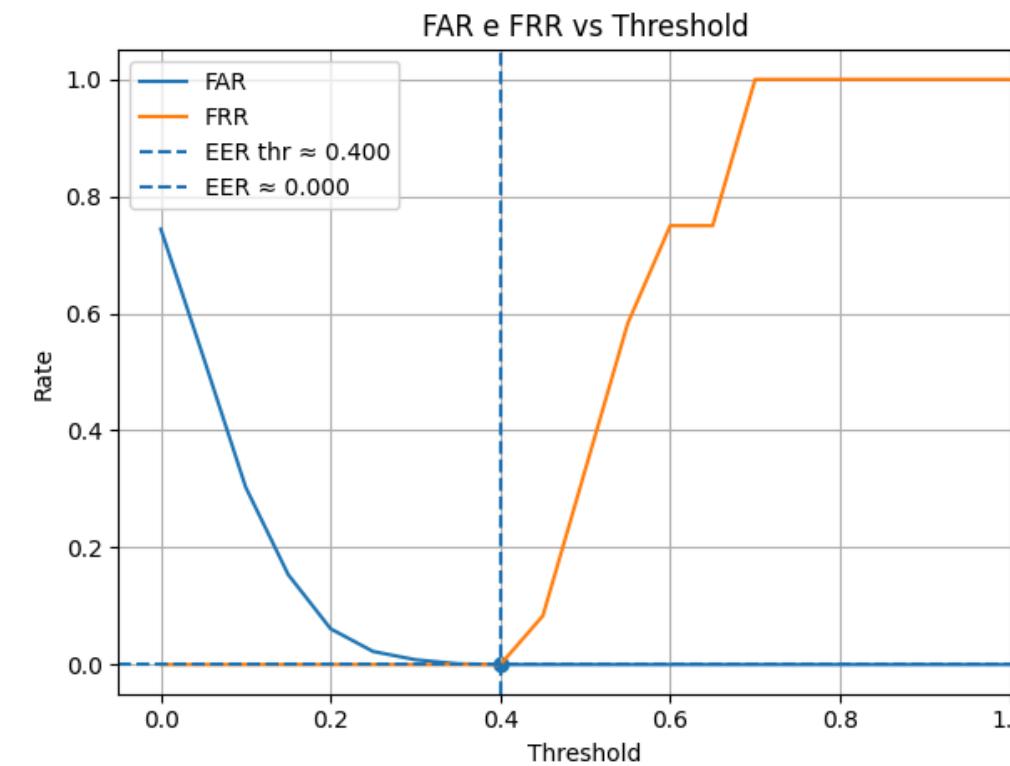


Worst Match: Queen Elizabeth

Similarity score: 0.2571

# Results (3)

## Camouflage



Genuine acceptances:



Best Match: Harry Styles

Similarity score: 0.6961

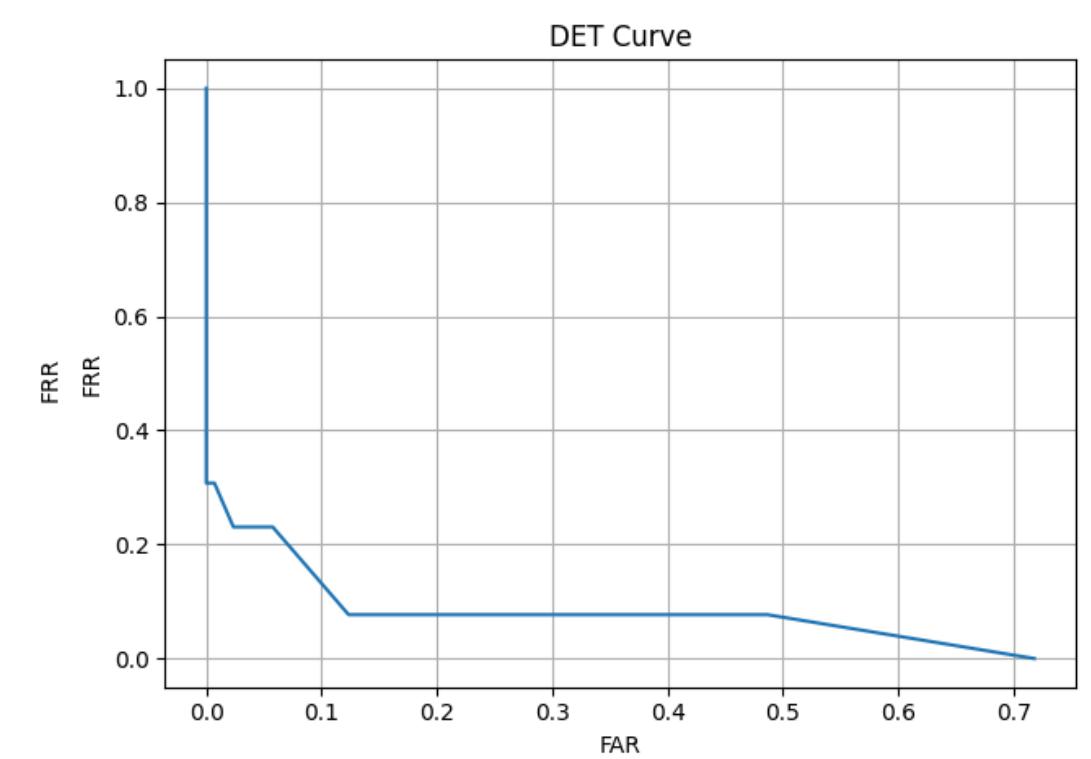
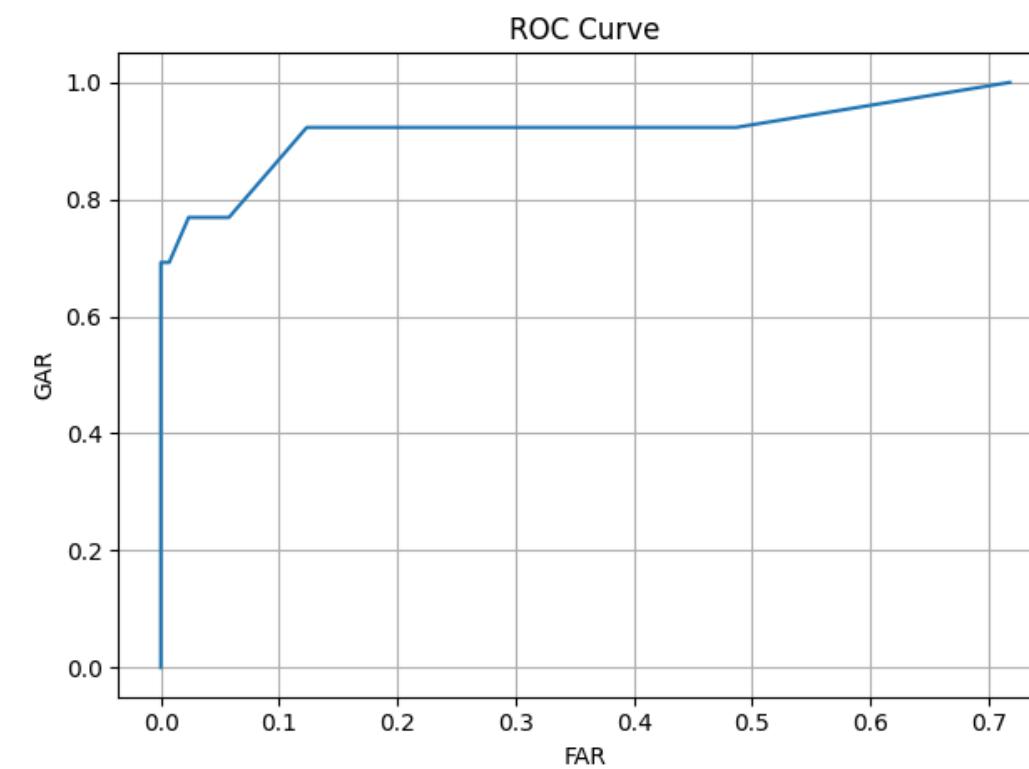
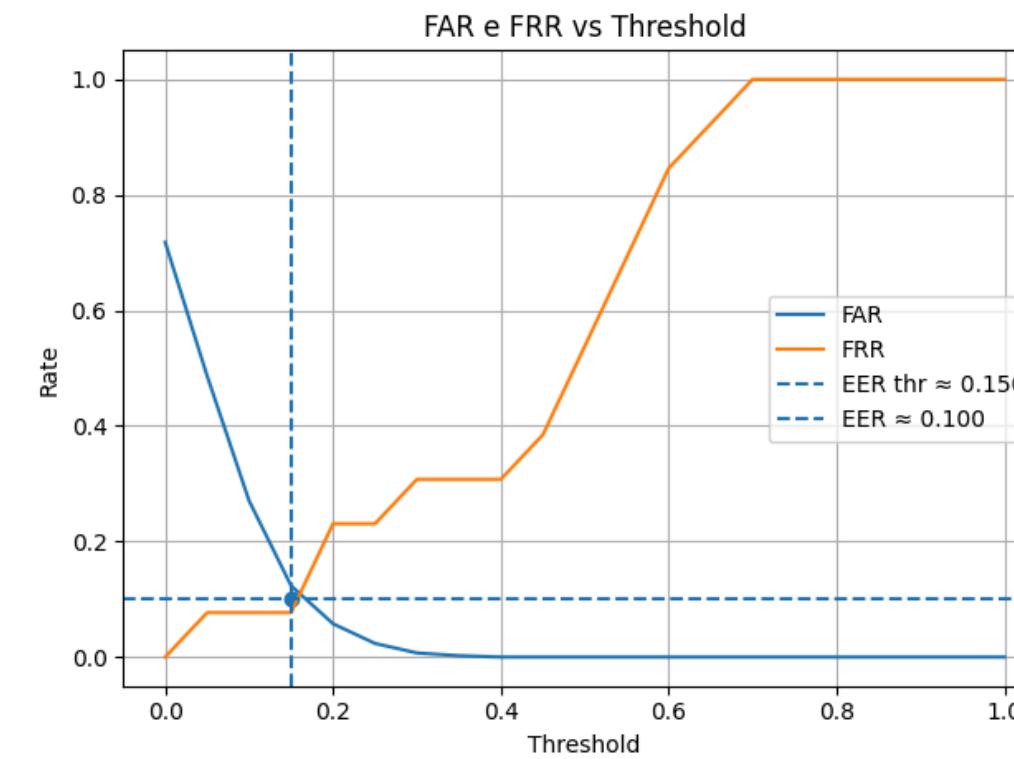


Worst Match: Chris Evans

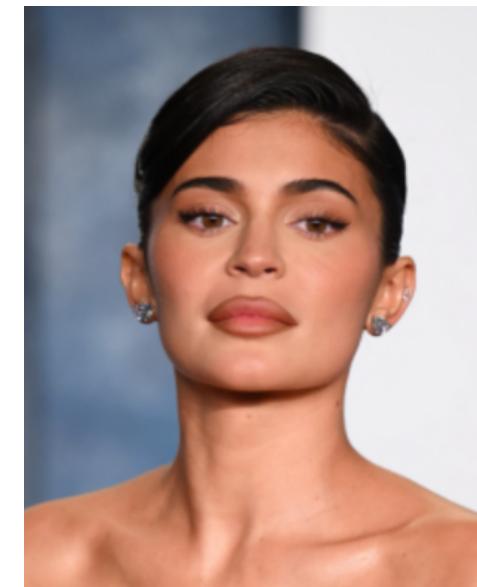
Similarity score: 0.3594

# Results (4)

## Make up



Genuine acceptances:



Best Match: Kylie Jenner

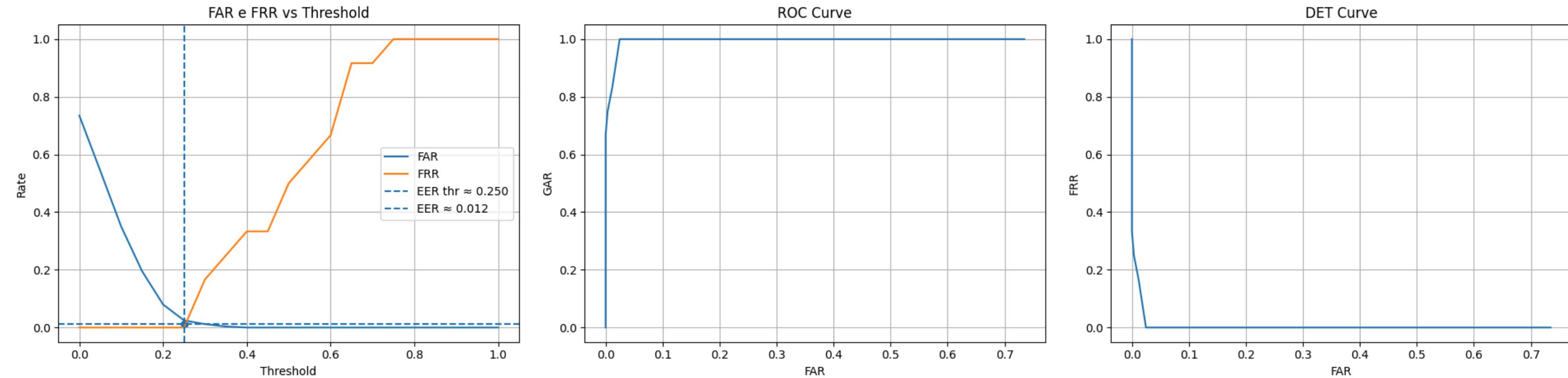
Similarity score: 0.6662

Worst Match: Robin Williams

Similarity score: 0.0232

# Results (5)

## Plastic Surgery



Genuine acceptances:



Best Match: Zac Efron

Similarity score: 0.7205

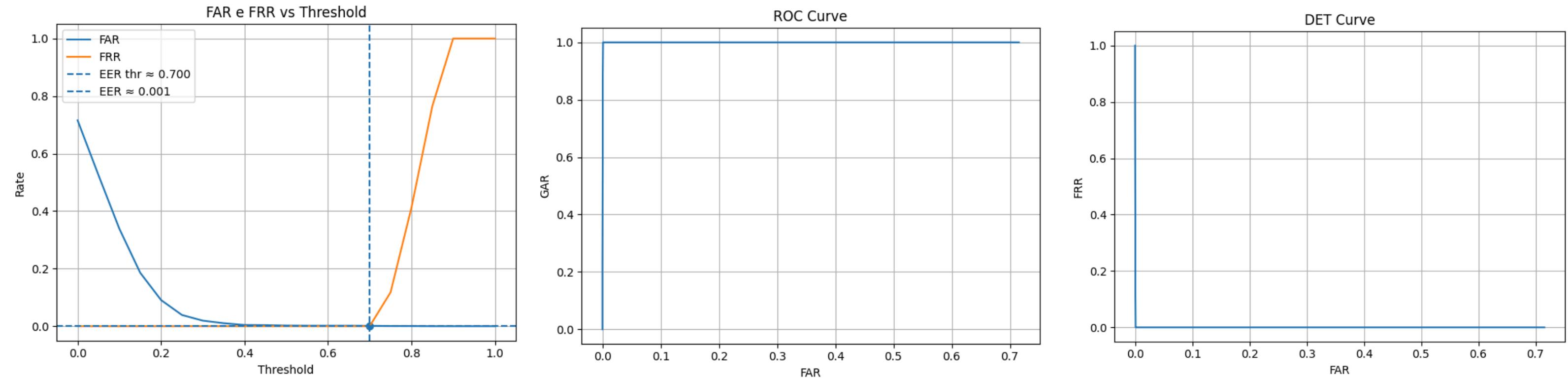


Worst Match: Lindsay Lohan

Similarity score: 0.1889

# Results (6)

## Inter-personal



Genuine acceptances and false acceptances:



Probe  
(In gallery)



Genuine match  
Sim score: 0.8281



Top impostor match  
Sim score: 0.4920



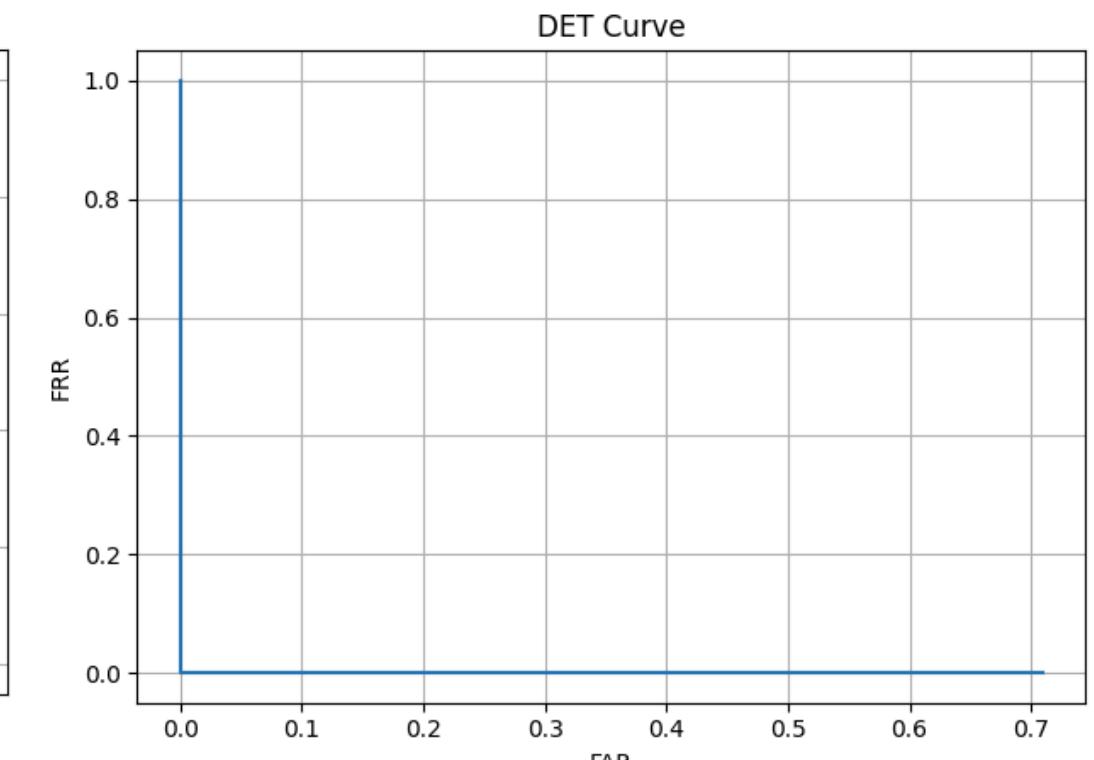
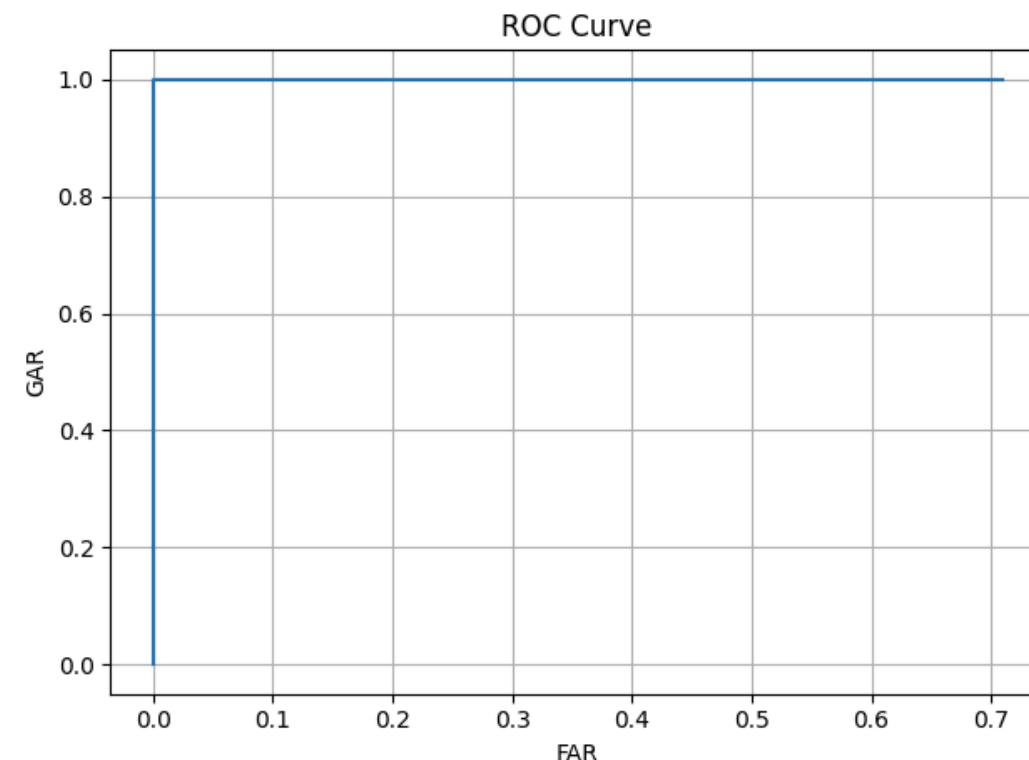
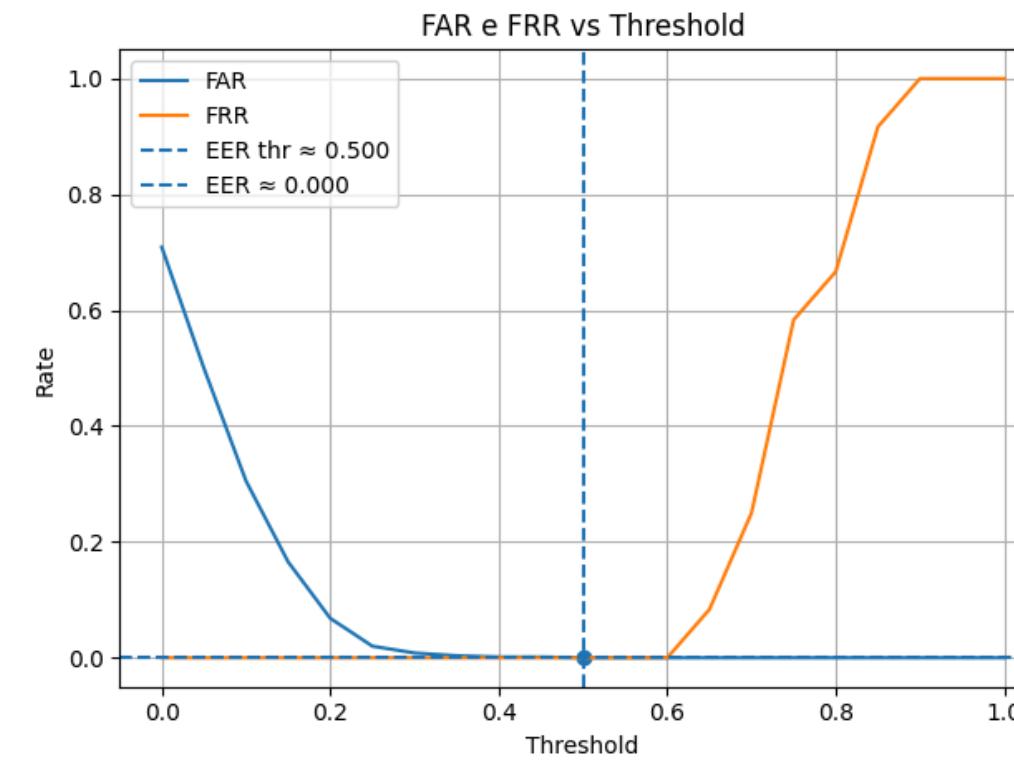
Probe  
(Not in gallery)



Impostor match  
Sim score: 0.6659

# Results (7)

## Print attack



Genuine acceptances:



Best Match: Leonardo Di Caprio

Similarity score: 0.8629.

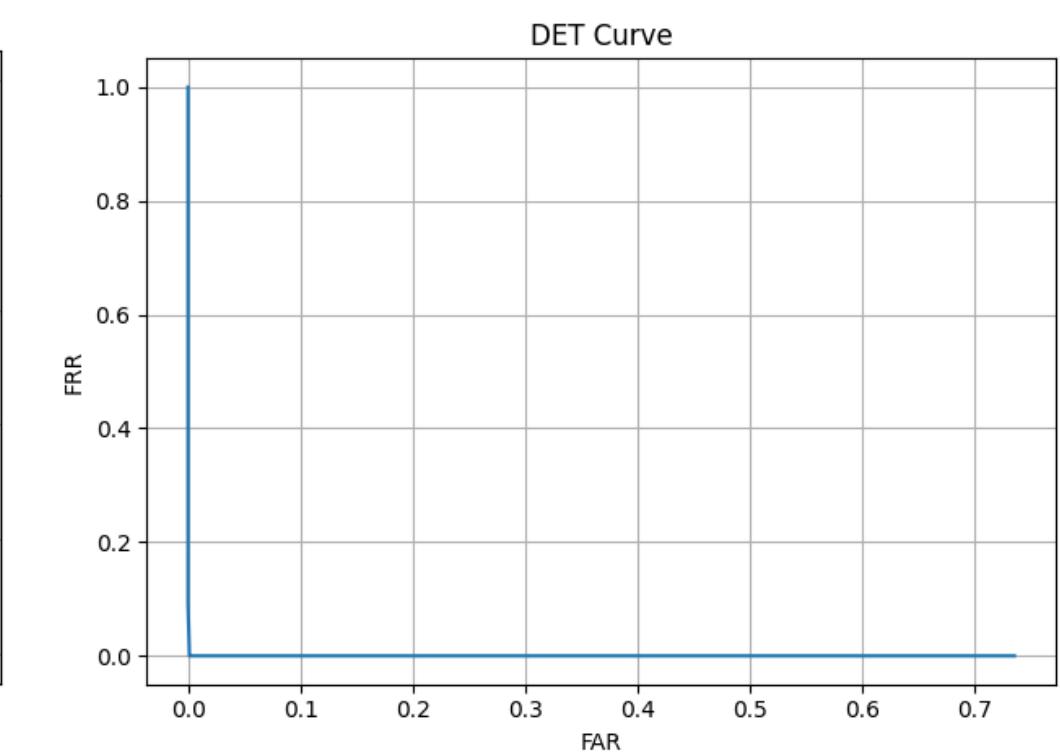
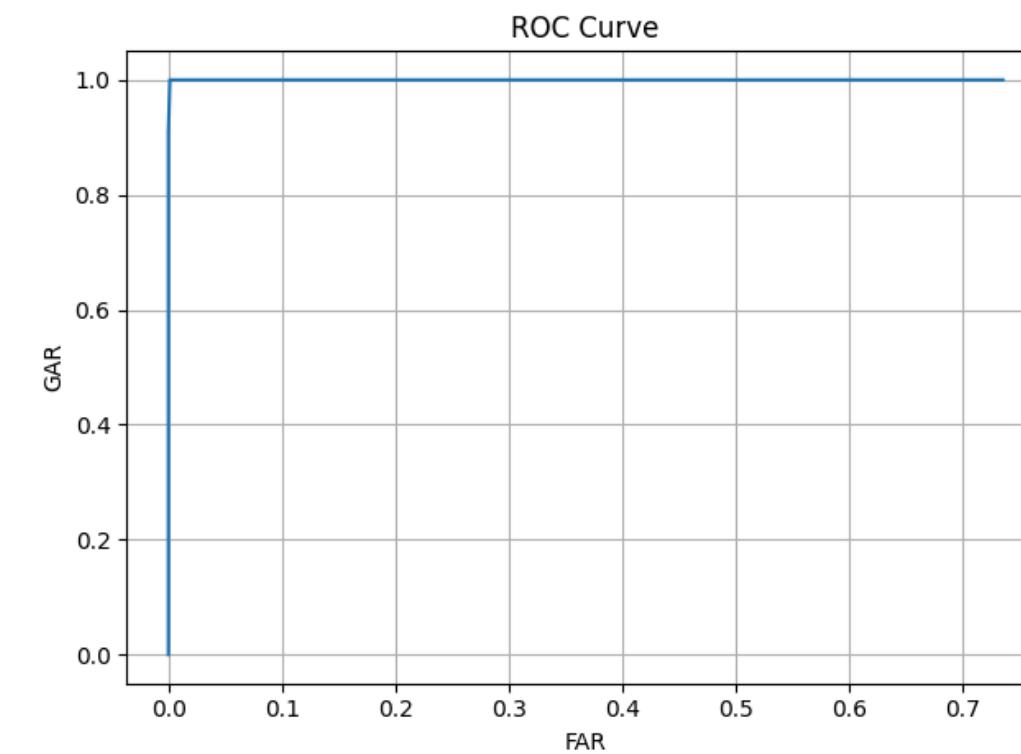
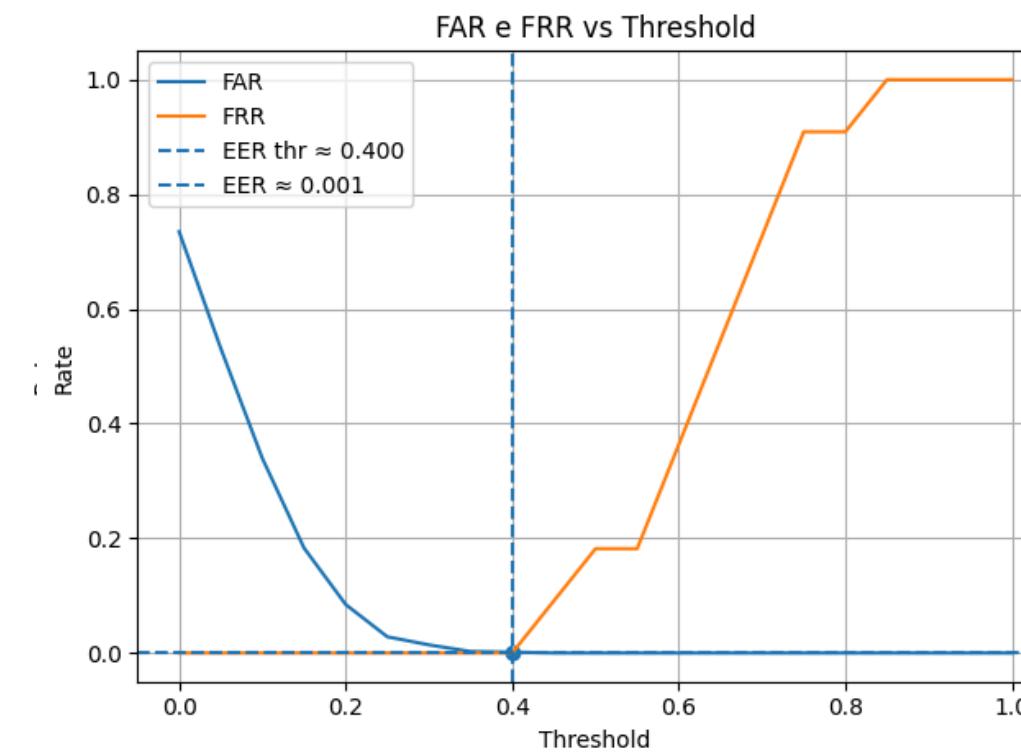


Worst Match: Johnny Depp

Similarity score: 0.4295

# Results (8)

## Deep fake

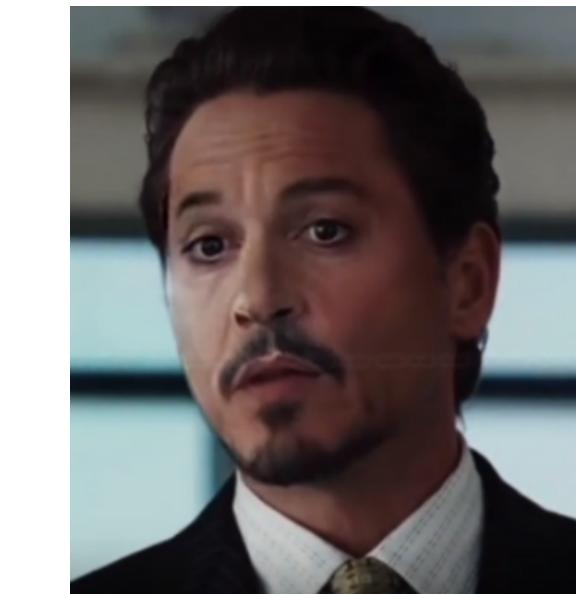


Genuine acceptances:



Best Match: Jennifer Lawrence

Similarity score: 0.8629

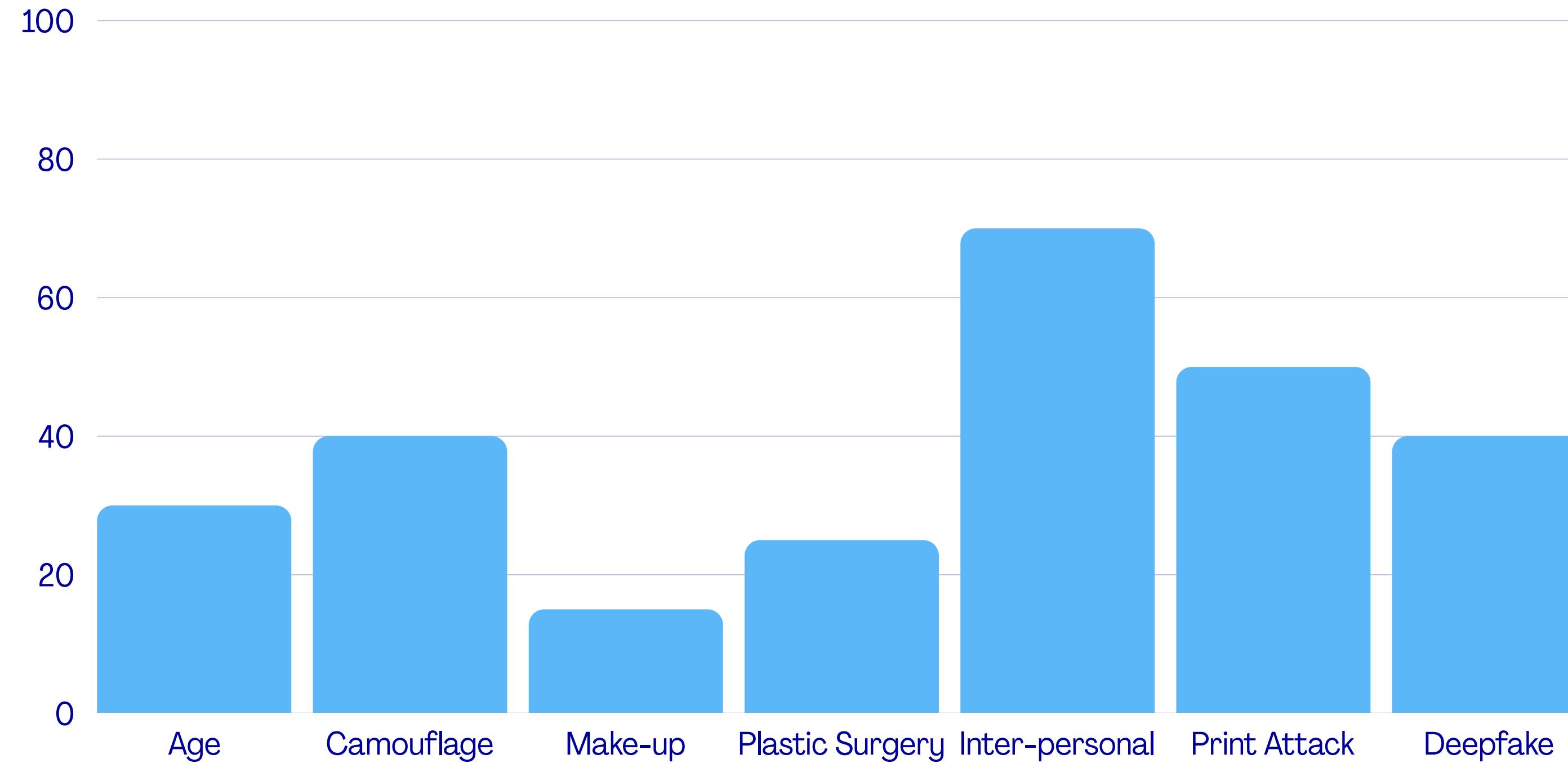


Worst Match: Johnny Depp

Similarity score: 0.4295

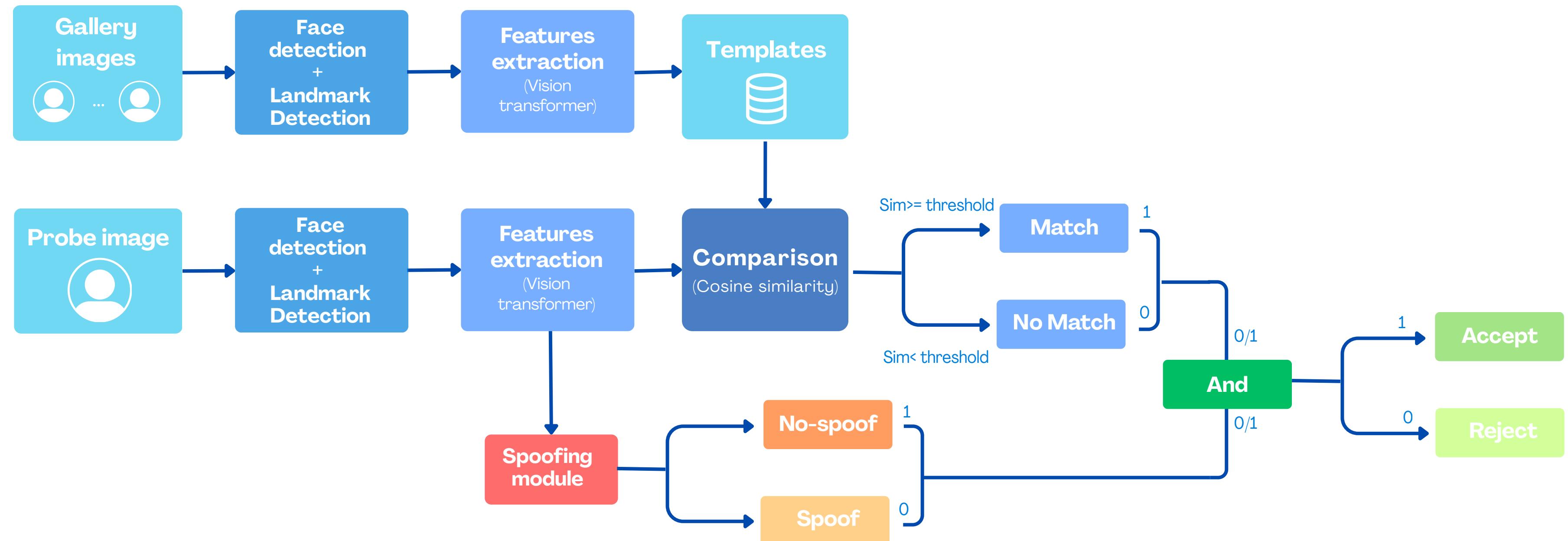
# Future works

## Adaptive Thresholds



# Future works

What's next?



# Conclusions

- ViT + ArcFace achieves **high face verification accuracy**
- Strong performance under natural variations
- **Performance decreases** under **extreme changes**
- **Spoofing attacks** are **not rejected** → require dedicated modules

# References (1)

- Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. arXiv preprint arXiv:1905.00641, 2019.
- Gau-Nernst. vit tiny patch8 112.arcface ms1mv3. [https://huggingface.co/gaunernst/vit\\_tiny\\_patch8\\_112.arcface\\_ms1mv3](https://huggingface.co/gaunernst/vit_tiny_patch8_112.arcface_ms1mv3), 2023. Accessed: 2025-11-20.
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- Ander Galván, Mariví Higuero, Ane Sanz, Asier Atutxa, Eduardo Jacob, and Mario Saavedra. Comparing cnn and vit for open-set face recognition. Electronics, 14(19), 2025
- Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi. Lightweight face recognition challenge. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 2638–2646, 2019.

# References (2)

- Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, October 2022.
- Tee Kai Feng. Understanding arcface loss: Intuitive insights and its application for representation learning. <https://medium.com/@teekafeng/understanding-arcface-loss-intuitive-insights-and-its-application-for-representatApril 2024>. Accessed: 2025-11-20.
- DeepInsight and Contributors. Insightface: 2d and 3d face analysis project. <https://github.com/deepinsight/insightface>, 2023. Accessed: 2025-11-20.
- Gau-Nernst. test\_ijb.py from the timm-face repository. [https://github.com/gau-nernst/timm-face/blob/main/test\\_ijb.py](https://github.com/gau-nernst/timm-face/blob/main/test_ijb.py), 2021. Accessed: 2025-11-20.

**Thank you  
for your attention!**

