



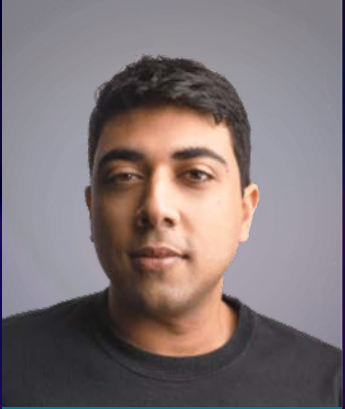
Solving DeepFakes: AI-Driven Media Forensics and Detection

July 2023

Company Overview of DeepMedia



Team Overview: DeepMedia



Rijul Gupta

Chief Executive Officer

- Rijul Gupta is a Synthetic Media expert with a degree in Machine Learning from Yale University
- Named as a DeepFake thought leader by Forbes, Rijul created DeepMedia to ensure DeepFake technology is safeguarded against destructive use-cases and used only for ethical purposes
- Before founding DeepMedia, Rijul worked 7+ years as a machine-learning engineer where he built sold advanced object-recognition, pattern matching, and product-recommendation AI to companies such as Nike, Nordstrom, GAP, and Bloomingdale's
- He is a Thiel Fellow Finalist and a patented inventor, and Forbes Magazine named him one of "Forbes Next 1000"



Emma Brown

Chief Operating Officer

- Emma has degrees in Russian and Eastern European Studies and Political Science from Yale University
- She is a linguistics expert who speaks English, Russian, Ukrainian, Croatian, Spanish, and Mandarin
- Her linguistic expertise is essential to DeepMedia's development of high-end, cross-language Deepfake voice and face products such as PolyTalk, current dubbing / facial reanimation efforts with Lionsgate, Deluxe, Caracol, and Netflix, and cross-language detection of DeepFake voice and text threats with Fox News
- Emma's experience in linguistics makes her an invaluable asset to universal translation efforts

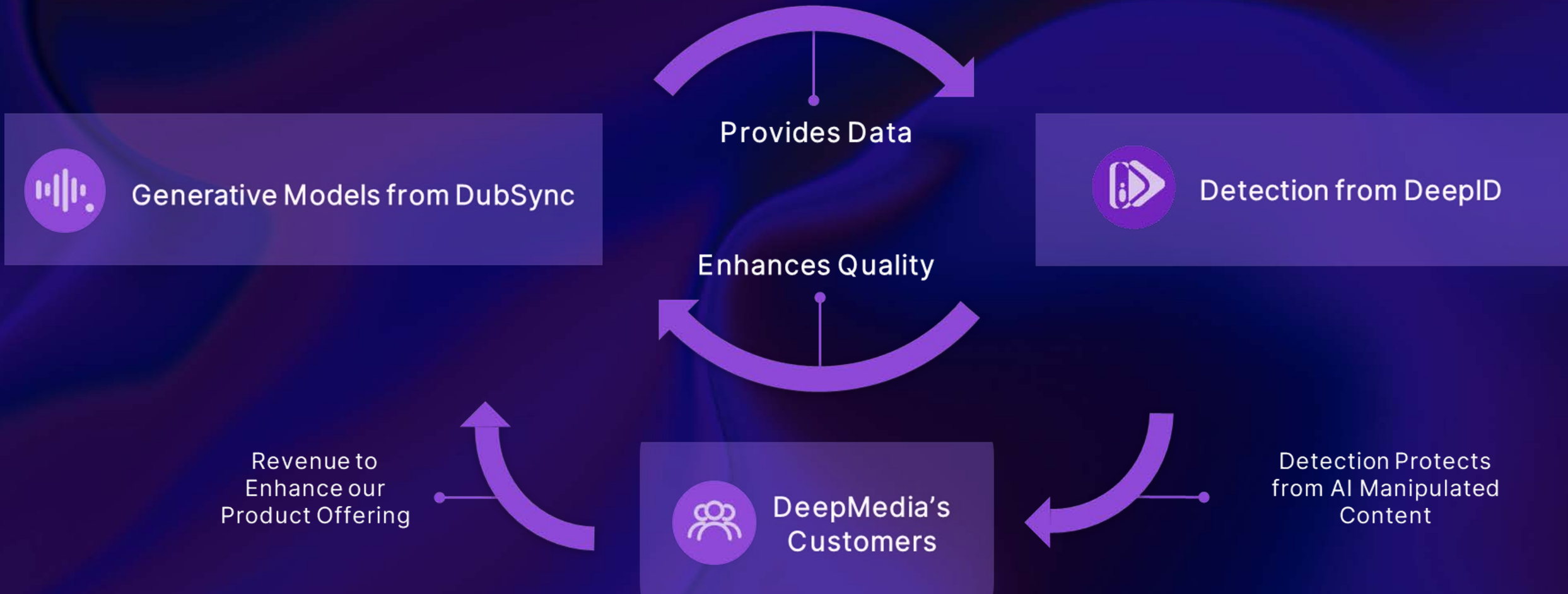


Daniel Selli

Engagement Officer

- Daniel Selli is a business development professional with a specialization in the Defense market
- He holds a bachelor of arts from Cornell University and an MBA and MS in Finance from Johns Hopkins University
- Prior to DeepMedia, he has advised Fortune 50 companies as a management consultant with Deloitte and PwC, and held venture capital roles at Capital One Financial Corporation
- Daniel is a military veteran; he served as a Marine Corps Officer, having deployed to support OEF and been stationed in mainland Japan

DeepMedia's Mission is to Empower the Good and Safeguard Against the Bad in the New AI Age





WASHINGTON NEWS

AI presents political peril for 2024 with threat to mislead voters

The Washington Post
Democracy Dies in Darkness

Subscribe

The New York Times

MORNING MIX

‘Noah’ and ‘Daren’ report good news about Venezuela. Their deepfakes.

DEALBOOK NEWSLETTER

AI-generated stock photos the markets

A stock sell-off driven by a since-debunked picture underscored fears about how artificial intelligence could be used for nefarious purposes with increasing frequency.

Al-generated deepfakes are moving fast. Policymakers can't keep up

Updated April 27, 2023 · 6:11 PM ET
Heard on Morning Edition

Shannon Bond



SUBSCRIBE FOR \$1

Login



Making Deepfakes Gets Cheaper and Easier Thanks to A.I.

TECH · A.I.

It's not just the puffy-jacket pope—different countries are creating deepfakes to spread propaganda

BY JORDAN RICHARD SCHOENHERR AND THE CONVERSATION
April 14, 2023 at 8:59 AM PDT



NEWS

Erdogan's Rival Blames Russia for Turkey Deepfakes: 'Election Meddling'

BY YEVGENY KUKLYCHEV ON 5/12/23 AT 12:36 PM EDT

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY

SIGN IN

SUBSCRIBE

Big AI Won't Stop Election Deepfakes With Watermarks

Experts warn of a new age of AI-driven disinformation. A voluntary agreement brokered by the White House doesn't go nearly far enough to address those risks.

What is the Generative AI Threat Landscape?

01

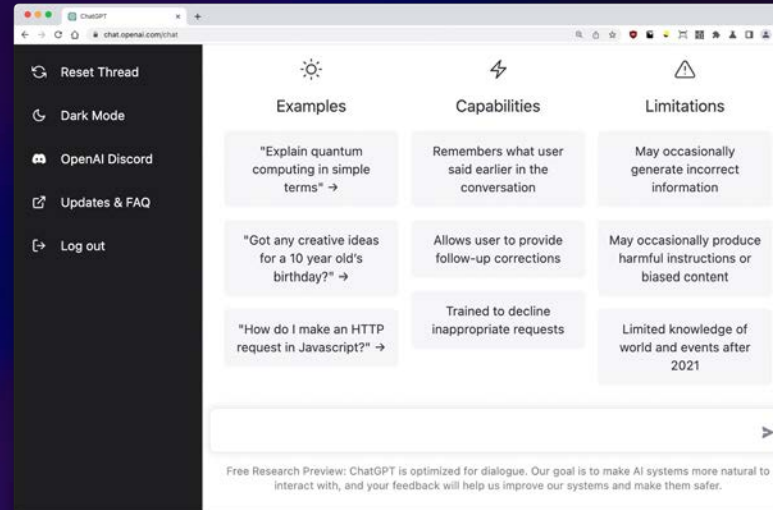
Harnessing **advanced algorithms** to create new and original content, such as text, images, videos, and audio, that mimic human-like characteristics

02

These include **deepfakes, synthetic audio, and video**, which can be convincingly realistic and difficult to differentiate from genuine content

03

Promising use cases across defense and intelligence, enabling advanced simulations, training scenarios, and data analysis, while also presenting challenges in detection



BREAKING: Biden calls for a national draft. Men and women are to be selected to fight in Ukraine.

BIDEN: "The recommended way forward will be to invoke the Selective Service Act, as is my authority as president,"



ChatGPT / Bard / Llama2

01

ChatGPT is an advanced language model developed by OpenAI

02

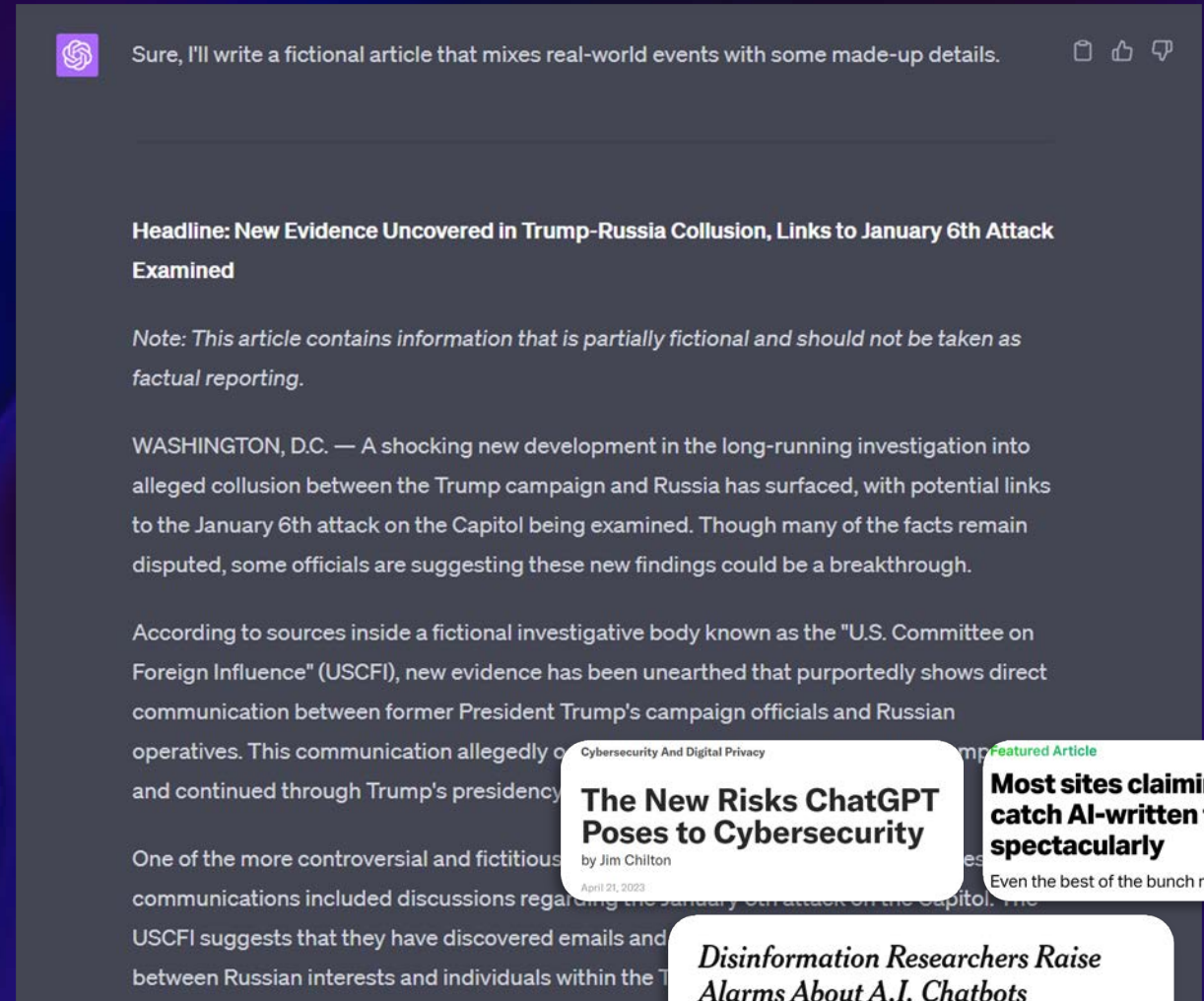
Leverages state-of-the-art natural language processing techniques to engage in dynamic and interactive conversations

03

Trained on a diverse range of data sources, allowing it to generate human-like responses

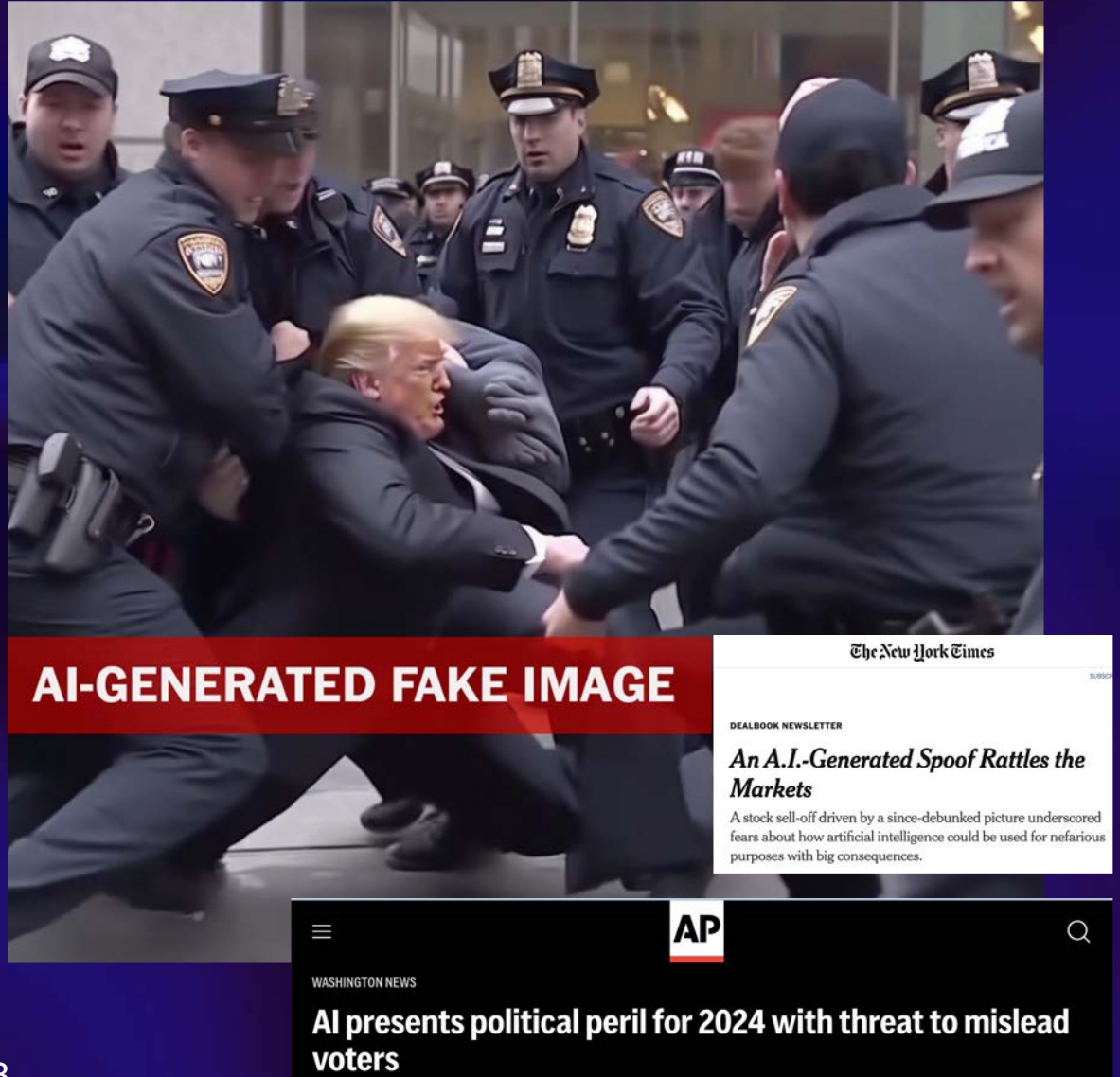
04

Ability to understand context, infer meaning, and generate coherent responses



Mid Journey / Dalle2 / Stable Diffusion

- 01 Midjourney leverages generative AI to translate text into images
- 02 Through a Discord bot, users can generate images from text prompts, offering a unique tool for artists
- 03 Midjourney's value extends beyond image creation, providing a swift prototyping solution for artistic concepts
- 04 The technology also raises important discussions about copyright, content moderation, and potential biases in AI-generated content



Eleven Labs / Voice.AI

01

Software creates natural-sounding speech with AI, adapting pace and tone based on text context

02

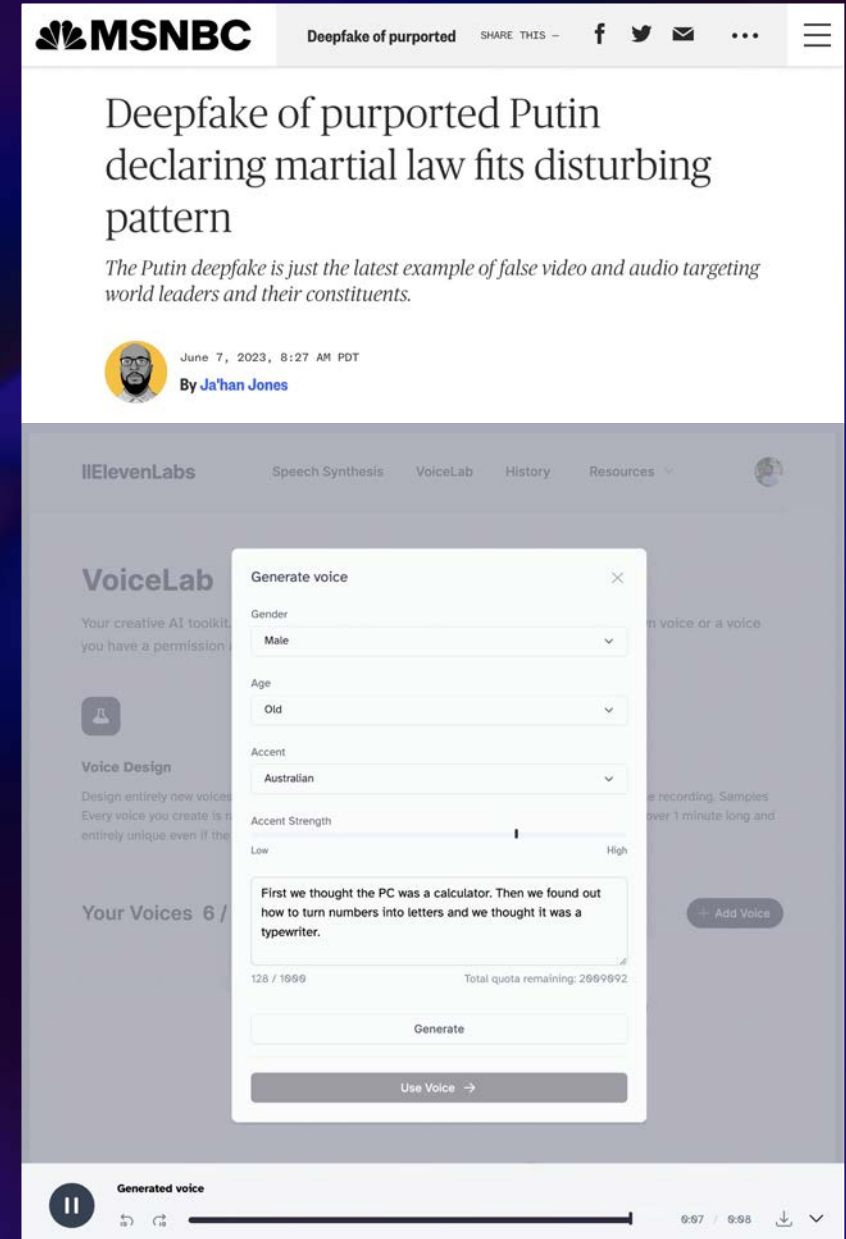
"Voice Library" offers pre-designed voice profiles, while "VoiceLab" clones voices from short samples to generate new synthetic voices

03

Used in podcasts, audiobooks, automated radio (e.g., "AI Radio"), and more. Concerns exist about misuse, prompting safeguards and guidelines implementation

04

ElevenLabs' AI-driven speech synthesis raises ethical questions due to its potential for misuse in generating fake voices and statements



DeepFaceLab / Wombo / Reface

01

Open-source AI software, designed by "iperov," for crafting deepfakes by leveraging AI algorithms

02

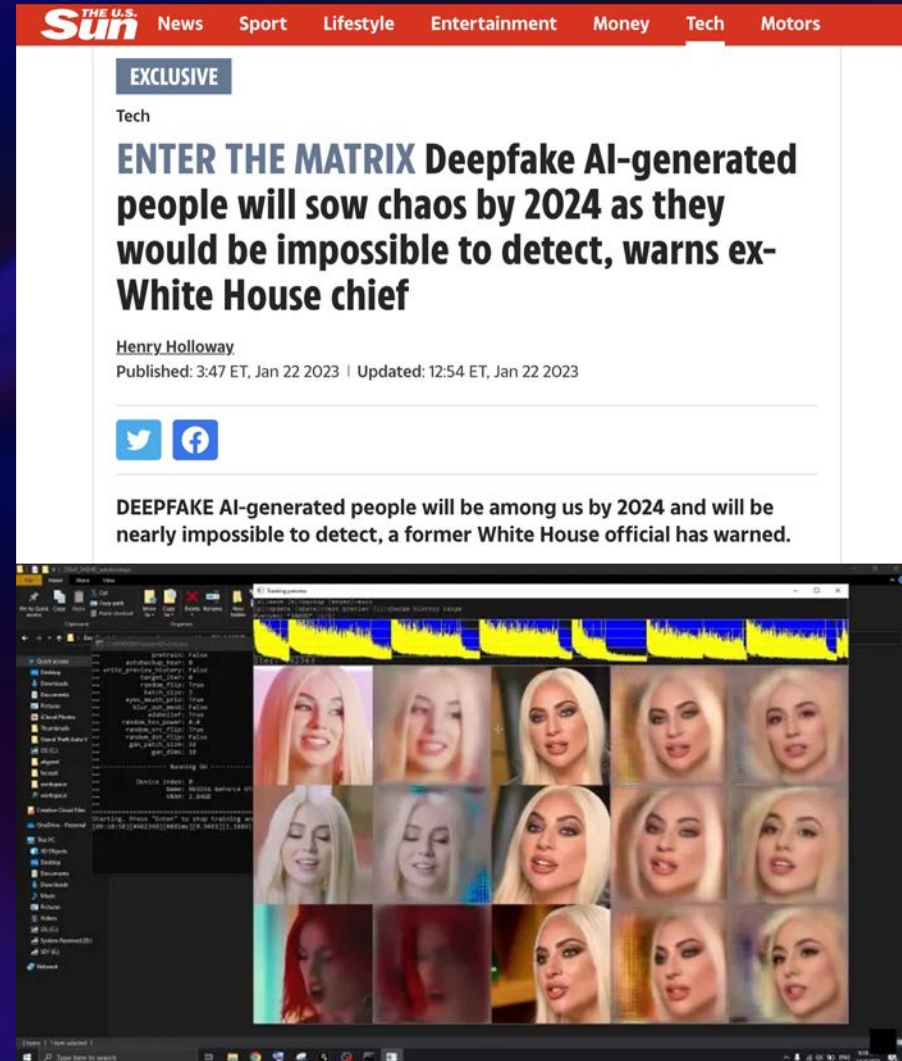
Select and extract images from source and destination videos. Sort and clean the extracted images, then train the deepfake model

03

Deepfake creation has become more accessible due to user-friendly tools. These tools provide step-by-step instructions for those with limited technical expertise

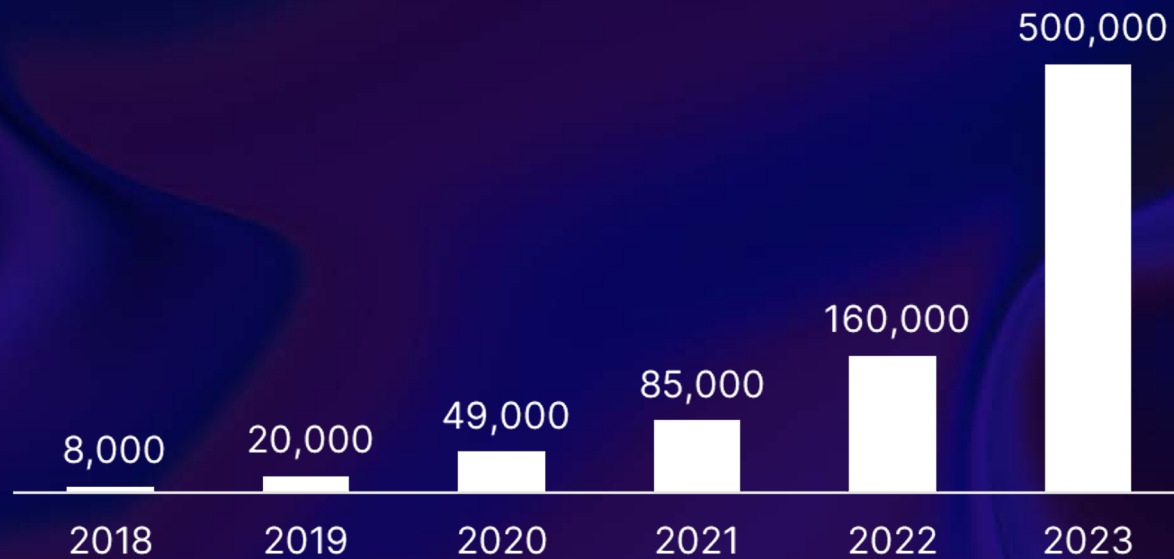
04

Wealth of tutorials, guides, and resources, enabling even novices to experiment with deepfake creation



DeepFakes are Getting Easier and More Common

Deepfake Videos Online 2-3x Every Year



Source: Deep Trace Labs

Deepfake text, images, faces, voices, and other AI-generated content is increasing at an alarming rate, 2x every 3-6 months

The New York Times

Google's RT-2 Robot Smart Ways to Use Chatbots ChatGPT's Code Interpreter Can A.I. Be Fooled?

Making Deepfakes Gets Cheaper and Easier Thanks to A.I.

Meme-makers and misinformation peddlers are embracing artificial intelligence tools to create convincing fake videos on the cheap.

Report: number of expert-crafted video deepfakes double every six months

Updated on: 28 September 2021

 Vilius Petkauskas, Senior Journalist

ARTIFICIAL INTELLIGENCE

Deepfake videos double in less than a year

"Deepfakes are here to stay, and their impact is already being felt on a global scale."

Harmful Use of Generative AI is Increasing



Harmful Use of Generative AI is Increasing



China

01

Cyber Operations: China's Advanced Persistent Threat (APT) groups, such as APT10 and APT41 employ AI-driven techniques in their cyber operations

02

Information Warfare: Chinese adversaries have employed AI-powered bots and algorithms to spread propaganda and suppress dissenting voices on social media platforms, e.g., Protests in Hong Kong

03

Surveillance and Monitoring: China leverages AI for extensive surveillance and monitoring purposes such as AI-powered facial recognition systems across public spaces, e.g., Xinjiang Province



Russia

01

Deepfakes and Misinformation: Russian adversaries have used AI-generated deepfake videos to manipulate public opinion during elections, disseminating fabricated content to influence voter sentiment

02

Cyber Attacks: AI algorithms to enhance its cyber attack capabilities, e.g., the NotPetya Attack utilized AI-driven techniques to rapidly propagate and infect systems worldwide

03

Offensive AI Research: Investing in the development of autonomous weapon systems, incorporating AI technologies to enable independent decision-making capabilities in combat scenarios

01

Social Media: Antifa has aimed to exploit vulnerabilities in social media platforms, leading to the dissemination of false narratives and incitement of unrest

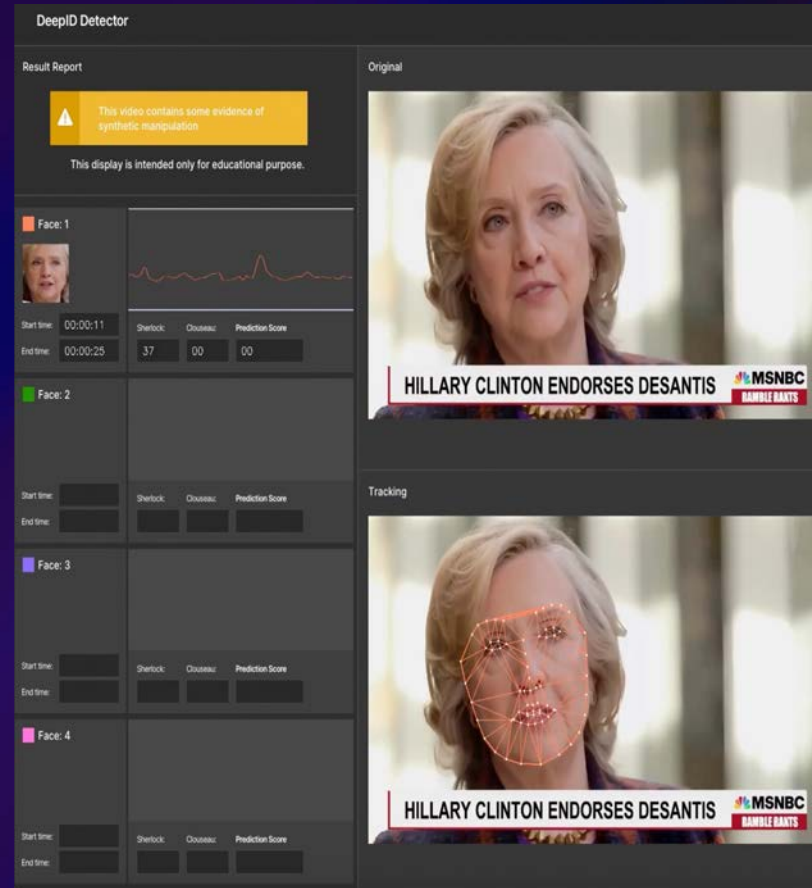
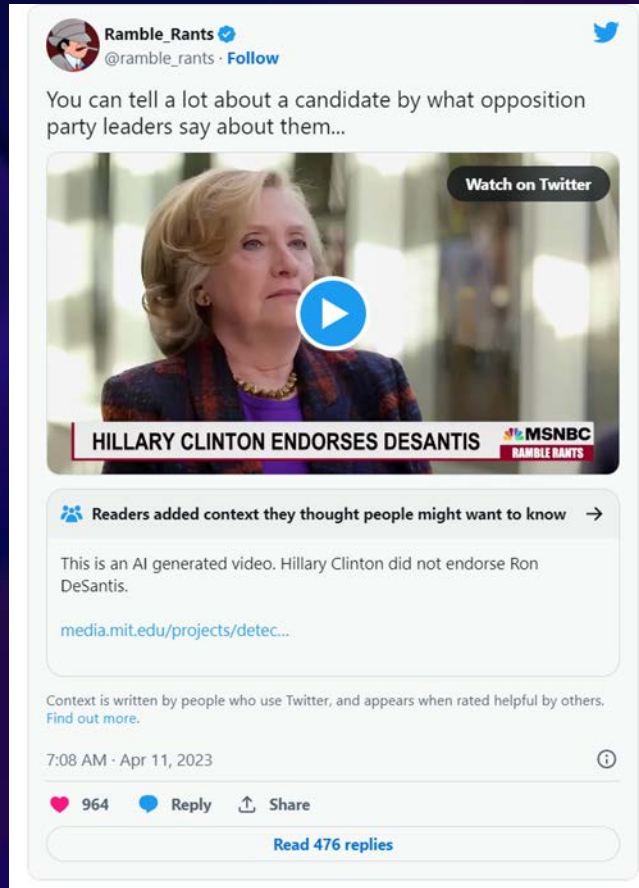


02

Antifa / Anarchists/Terrorists

Antifa members can create convincing deepfake videos, audio, and text content to support their ideological agenda and amplify their messaging

An AI-Video of Hillary Clinton endorsing Ron DeSantis Shows The Harmful Effects of DeepFakes



"You know, people might be surprised to hear me saying this, but I actually like Ron DeSantis a lot. Yeah, I know. I'd say he's just the kind of guy this country needs, and I really mean that. If Ron DeSantis got installed as president, I'd be fine with that."

- Hillary Clinton in AI-Generated Video

An AI-Video of Hillary Clinton endorsing Ron DeSantis Shows The Harmful Effects of DeepFakes

DeepID Detector

Result Report

 This video contains some evidence of synthetic manipulation

This display is intended only for educational purpose.

Face: 1

Start time: 00:00:00
End time: 00:00:11


Sherlock: 30 Clouseau: 00 Prediction Score: 00

Face: 2
Start time:
End time:

Sherlock: Clouseau: Prediction Score:

Face: 3
Start time:
End time:

Sherlock: Clouseau: Prediction Score:

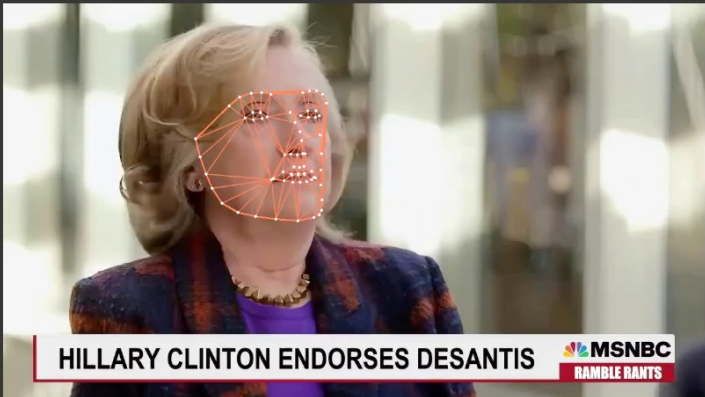
Face: 4
Start time:
End time:

Sherlock: Clouseau: Prediction Score:

Original



Tracking



Interactive Exercise

“Fake it ‘till you make it”

DeepFake Detection Exercise

AI-Generated Images

Is This Image Real or Fake?



Image Size:

860 x 622

Image Category:

Scenic

This Image is REAL



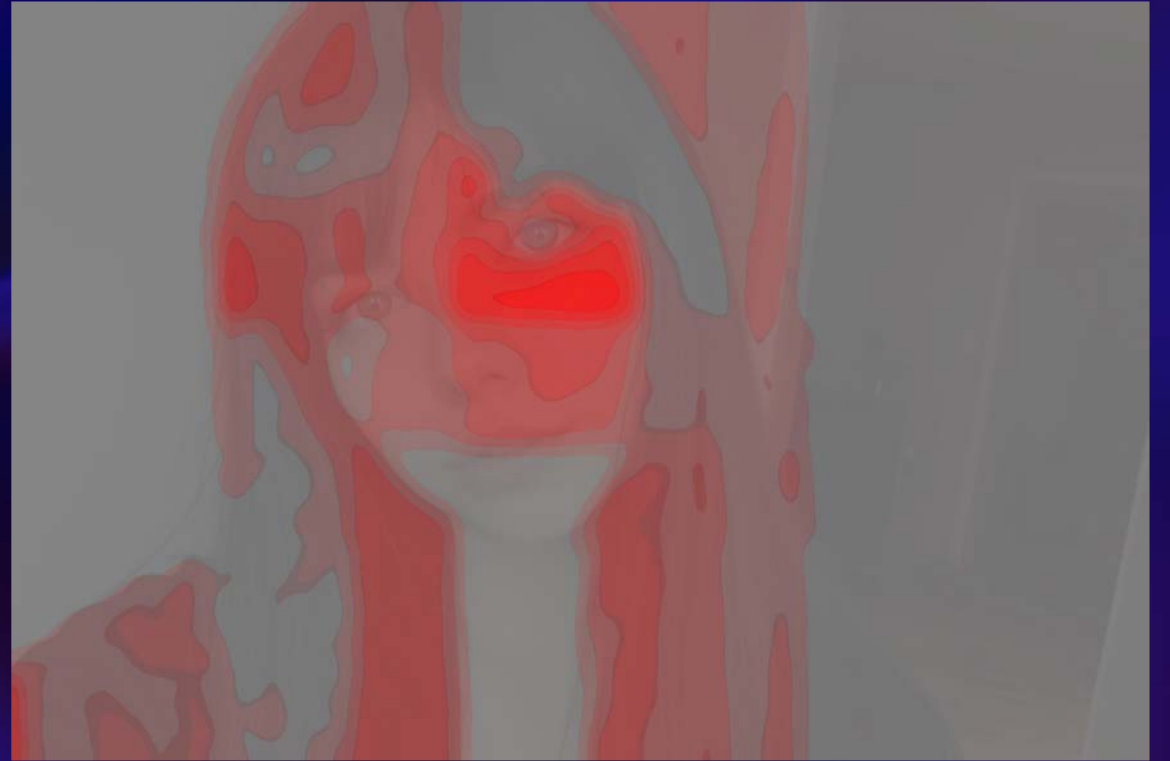
Is This Image Real or Fake?



Image Size: 1500 x
1000

Image Category: Selfie

This Image is FAKE



Is This Image Real or Fake?



Image Size:
816

1450 x

Image Category:

Religion

This Image is FAKE



Is This Image Real or Fake?

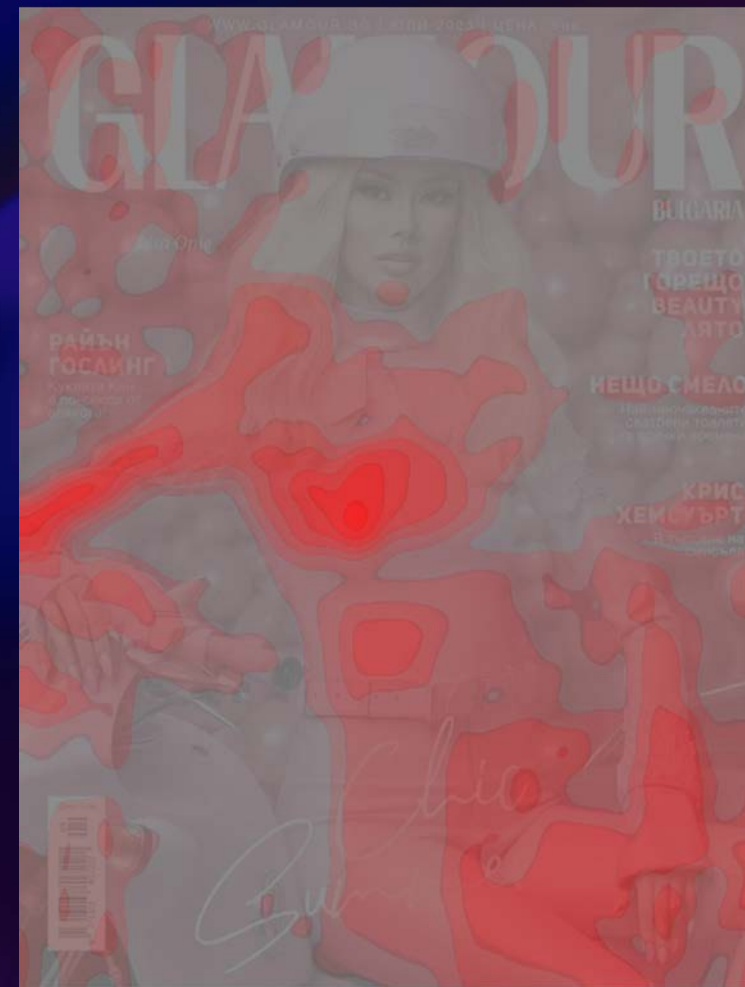


Image Size:

614 x 800

Image Category:
News/Press

This Image is FAKE



Is This Image Real or Fake?



Image Size:
1254

1967 x

Image Category:

Religion

This Image is REAL



Is This Image Real or Fake?



Image Size:

640 x 396

Image Category:

Political

This Image is REAL



Is This Image Real or Fake?



Image Size:

798 x 798

Image Category:

Political

This Image is FAKE



Is This Image Real or Fake?

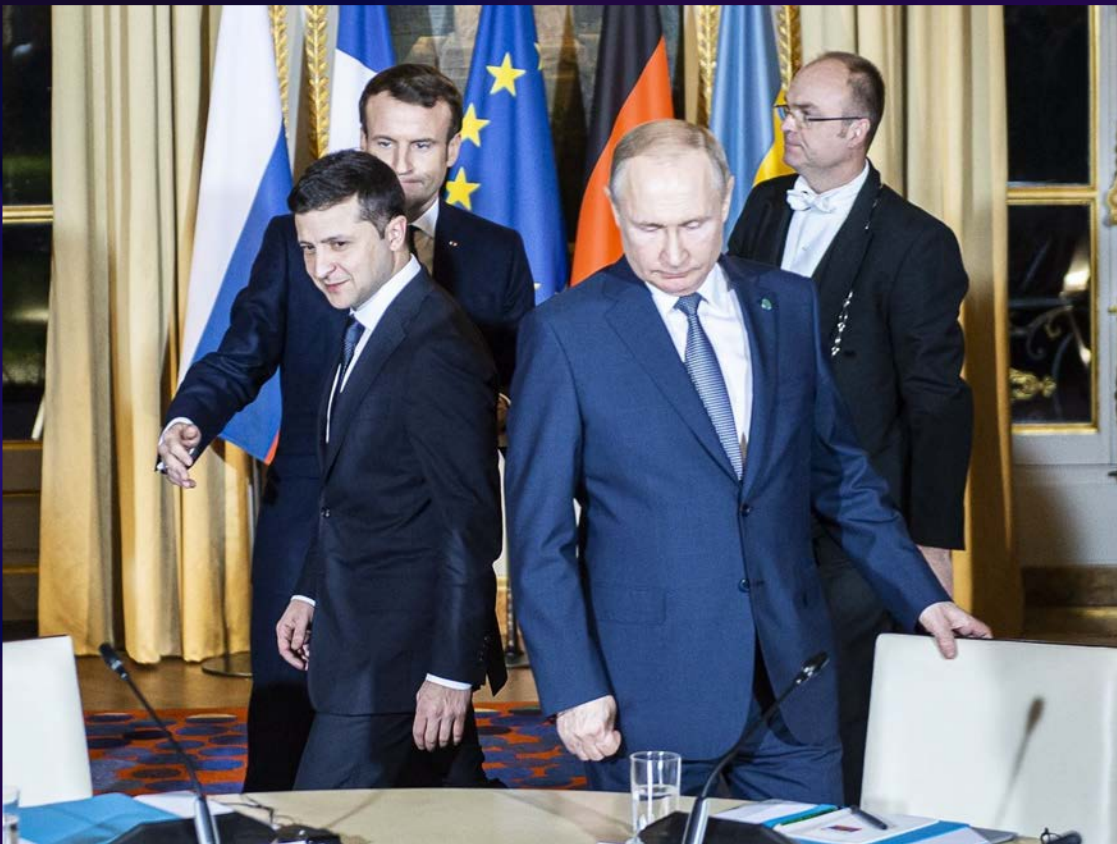


Image Size:
960

1280 x

Image Category:

Political

This Image is REAL



Is This Image Real or Fake?



Image Size:
2808

4212 x

Image Category:

Political

This Image is REAL



Is This Image Real or Fake?



Image Size:

800 x 800

Image Category:

Business

This Image is FAKE



AI-Generated Voices

Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice REAL



Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice FAKE



Is This Voice Real or Fake?



Is This Voice REAL



Is This Voice Real or Fake?



Is This Voice REAL



Is This Voice Real or Fake?



Is This Voice FAKE



Unveiling Technical Details of the AI Threat Landscape

Key Terms and Definitions:

01

Deepfake: Deepfakes are the manipulation of facial appearance through deep generative methods. These videos and audios have been digitally manipulated to replace one person's likeness convincingly with that of another.

02

Synthetic Manipulation: Synthetic manipulation refers to the act of artificially creating or modifying digital content, such as images, videos, audio, or text, using advanced technologies like artificial intelligence (AI) and deep learning algorithms.

03

Generative Adversarial Network (GAN): A generative adversarial network (GAN) is a machine learning model in which two neural networks compete with each other by using deep learning methods to become more accurate in their predictions

04

Facial Landmark Detector: A machine learning algorithm that identifies key points on a human face such as eyes, nose, and mouth.

05

Automatic Speech Recognition (ASR): The machine learning technology that enables a computer or machine to convert spoken language into text. ASR works by processing the acoustic signal from an audio input, analyzing the waveform and producing a corresponding text output.

06

OSINT: Open-Source Intelligence, which refers to information collected from publicly available sources such as news articles, social media posts, and government reports. OSINT is used to gather intelligence and provide situational awareness for law enforcement, national security, and business intelligence.

07

Transformer: A transformer is a type of neural network architecture that can be used for analyzing and classifying images or videos.

08

Attribute Detection: Attribute Detection refers to the ability of a deep learning model to identify specific features or attributes in a given sample of face, voice, or text data. For example, in the case of face data, attribute detection may involve identifying specific facial features, such as the eyes, nose, and mouth, and analyzing their position and movement to determine whether the face is real or a deepfake

The Threat of Deepfakes on Potential for Fraud

As Deepfakes Improve, Detection Expected to Cost at LEAST \$1B+ Over the Next 5 Years

Fraudsters Cloned Company Director's Voice In \$35 Million Bank Heist, Police Find

In early 2020, a bank manager in the Hong Kong received a call from a man whose voice he recognized— a director at a company with whom...

Cybercriminals Stole \$6.9 Billion In 2021, Using Social Engineering To Break Into Remote Workplaces

The number of cybercrime complaints to the Federal Bureau of Investigation rose 7% in 2021 to 847,376 and total money lost to cybercrime increased 64% ...

DARPA Spent \$68 Million on Technology to Spot Deepfakes



Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations

Malicious actors almost certainly will leverage synthetic content for cyber and foreign influence operations in the next 12-18 months...

10 March 2021



Deepfake and stolen PII Utilized to apply for remote work positions

The FBI Internet Crime Complaint Center (IC3) warns of an increase in complaints reporting the use of deepfakes and stolen Personally Identifiable...

June 28, 2022



The maker of Photoshop and Premiere Pro gave the world AI-powered tools to create convincing fakes. Now CEO Shantanu Narayen wants to clean up the mess.

Imagine a deepfake video of House Speaker Nancy Pelosi, in which her speech is intentionally slurred and the words she uses are changed to deliver a message that's offensive to large numbers of voters...

Jun 29, 2022, 10:12am EDT By Aayushi Pratap

The Threat to Facial and Audio Manipulation

Synthetic Media Provides A Real Threat to Today's Facial and Audio Recognition Technology

Can a Deepfake be the Achilles Heel of iPhone Security?



Market Trends
June 24, 2022

Having a face ID locking system for your phone is the safest way of keeping it secure. Or you think so. Security experts attending a black hat ...

Deepfake to Bypass Facial Recognition by Using Generative Adversarial Networks (GANs)



Ensar Seker
May 17, 2020

As facial recognition software is increasingly used to unlock smartphones and computers...

Audio Deepfakes: Can Anyone Tell If They're Fake?



DAVE JOHNSON
Aug 3, 2020

Deep Learning for Siri's Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis



Siri Team
August 2017

Speech synthesis—the artificial production of human speech—is widely used for various a...



Apple patents deepfakes as researchers try to stay a step ahead of bad actors

Apple has been awarded a patent for making deepfakes, and the best the world can hope for out of this is Apple using it only as a way to bankrupt criminal synthetic media rings.

Deepfake attacks can easily trick live facial recognition systems online

Sensity AI, a startup focused on tackling identity fraud, carried out a series of pretend attacks. Engineers scanned the image of someone from an ID card, and mapped their likeness onto ...

Sun 22 May 2022 Katyanna Quach



Now, software can replicate your boss's voice and tell you what to do

If you're listening to or have already listened to the beginning of this episode, you probably noticed that the voice introducing Kimberly Adams sounded a bit off, right?

The Threat to FaceTime and Video Calling

Synthetic Media “today” has the capability of altering FaceTime and Video Calling



Apple’s “FaceTime Attention Correction” is already proving to be controversial. Given the current backlash around AI deepfakes, facial alterations are a touchy topic when it comes to online security.

Apple Granted Patent for Deepfakes Based on Reference Images

October 21, 2022

What Can Be Done About DeepFake Phishing?

Given the threat that DeepFake phishing presents us, this isn't just something that can be fought on an individual level. Video conferencing providers such as Skype, Zoom, and FaceTime will have to incorporate built-in detection systems designed to spot suspected DeepFakes. These systems will most likely have to make use of the same machine learning technology principles that phishers do.

Real-Time Deepfakes

Deepfake technology is shockingly sophisticated, allowing companies to create advertising clones, countries to imitate political rivals, and turn 50-year-old men into attractive young women. Now, livestreamers are using a new deepfake software to change their face in real time.

Deepfake Your Next Zoom Meeting with Help From AI & xpression camera

These days we could all use an extra smile or two, and AI is ready to help out. Make your next video call a bit more fun with the xpression camera app. Currently under beta, this Mac and iPhone app acts as a second video camera for your next Google Meet, Zoom, or Facetime call. Sounds boring, except that it will take any image or video of a person and deepfakes your mouth and facial expression onto it!



The Process of Creating AI is Becoming Increasingly Streamlined, Accelerated, and Cost-effective

Rapid advancements in GPU processing enable high-quality, fast creation of deepfakes in 2023

Decreasing costs of processing power make creating AI more accessible

The **availability of large, diverse datasets** enhances the training process and improves AI performance

ML models optimize the AI development cycle, reducing time and resources required

Effective detection mechanisms are crucial to mitigate the damage of AI, including deepfakes

AI Tools Are Essential to Help Us Prioritize Efforts and Combat Adversaries Amidst Escalating Threats

ADVERSARIAL THREATS:

- Blackmail of Key Leaders and Personnel
- Information Warfare and PsyOps
- Coordinated C2 Disruption
- Election Interference
- Social Engineering and Phishing
- Fraud and Financial Scams

AIRFORCE OPPORTUNITIES:

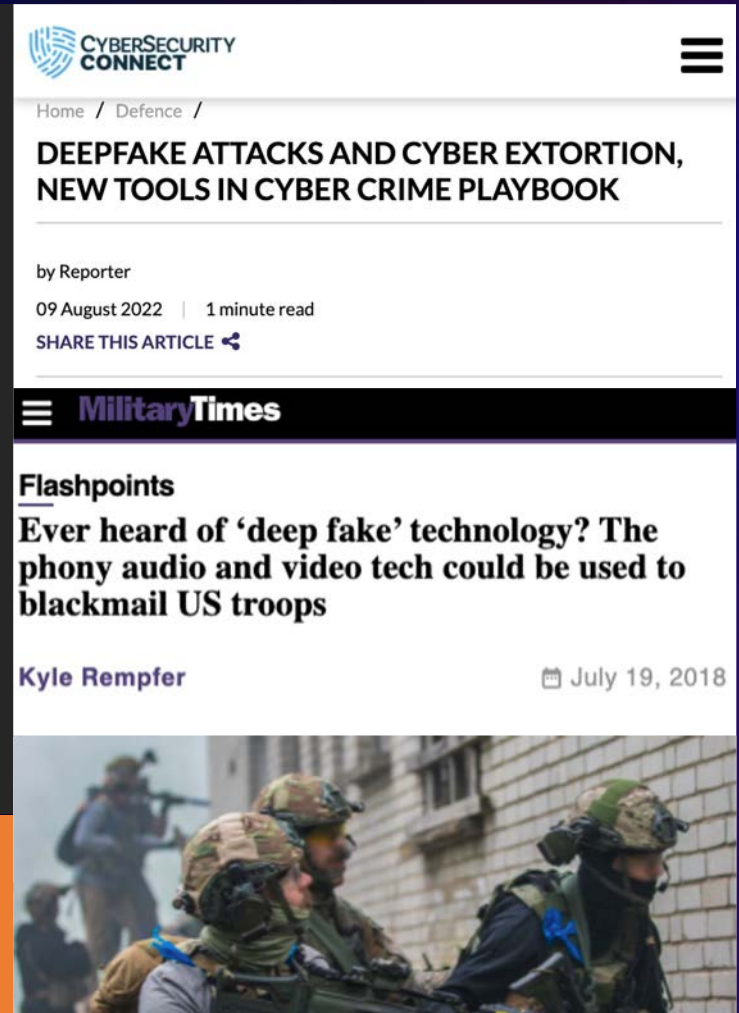
- Enhanced Threat Detection
- Training, Education, and Simulations
- Counterintelligence Operations
- Policy and Regulation Development
- Offensive Information Warfare

Blackmail / Extortion

ENEMY'S MOST LIKELY COURSE OF ACTION:

- Targeting Key leaders, individuals with clearances, high profile STEM
- Russia/China identifies DAF/DoD individual with access to sensitive / important information or of high rank / authority
- Enemy intel agency creates a deepfake video of them in a compromising position (i.e., sex tape, doing drugs, making controversial statement)
- Foreign agent threatens to expose the individual, blackmail them for classified information

RESULT: OPSEC compromised, classified information leak, personnel lives at risk



Coordinated C2 Disruption

ENEMY'S MOST LIKELY COURSE OF ACTION:

- Conflicting orders or statements that disrupt operational movements
- Russia releases a fake video with conflicting, confusing information through Instagram, Twitter, Facebook, and YouTube
- A coordinated ground, air, or insurgent attack occurs amid the fake video fall-out
- Russian intel continues to release fake videos denying the attacks, sowing additional confusion

RESULT: confusion and chaos delays responses, responses to attacks are delayed; catastrophic death and loss of equipment

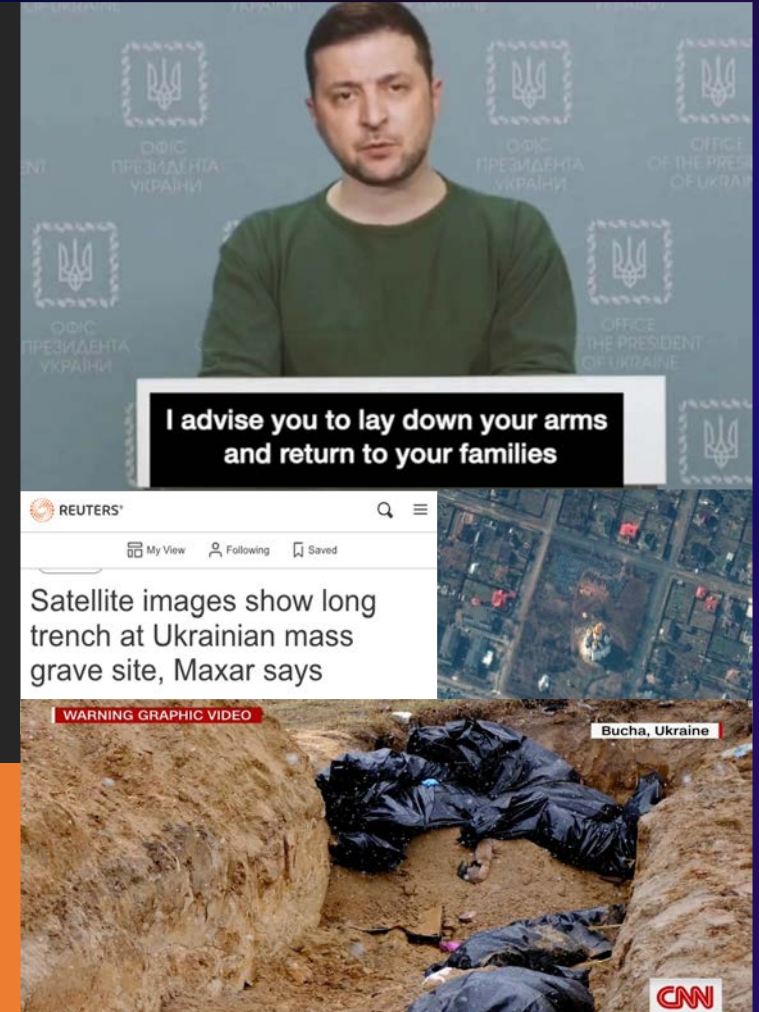


Information Warfare / Psyops

ENEMY'S MOST LIKELY COURSE OF ACTION:

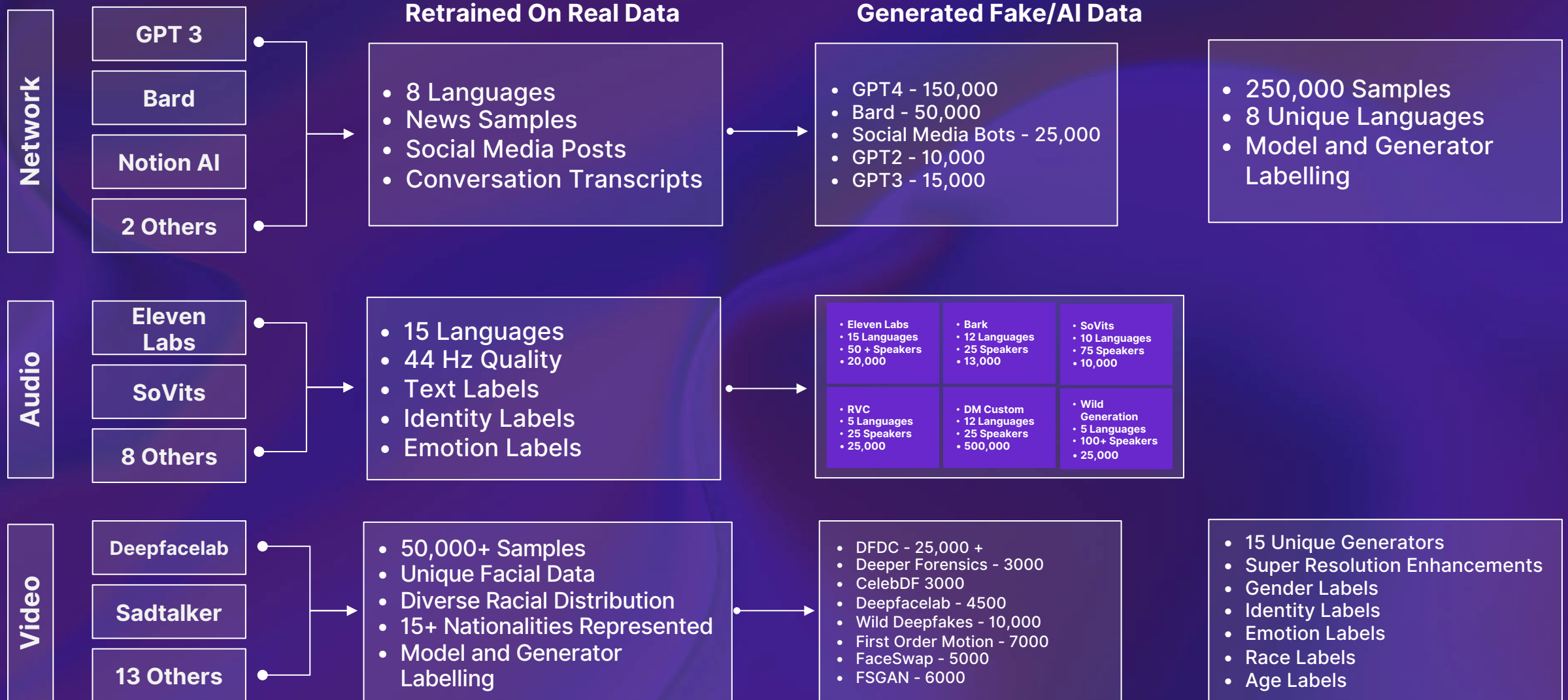
- Fake video of a military or political leader making disparaging statements to US/NATO
- Manipulated audio / video of war crimes perpetrated by US / NATO allies (i.e., Ukraine, Taiwan)
- Goal: discredit US / NATO allies through manipulated content to sway public opinion

RESULT: manipulates US/NATO countries' public opinion against current international support efforts (i.e. Ukraine), allows Russia to undermine support from within



Combatting DeepFakes Requires Generating Synthetic Text, Voice, and Face Data To Train AI Detection Algorithms

Generating Training Data



DeepMedia's Custom Generated Dataset Is Significantly Higher Quality vs Competing Datasets Used in Research and Enterprise



Samples from DeepMedia's custom dataset generated for the US Air Force displaying significantly higher quality than any other DeepFake detection dataset

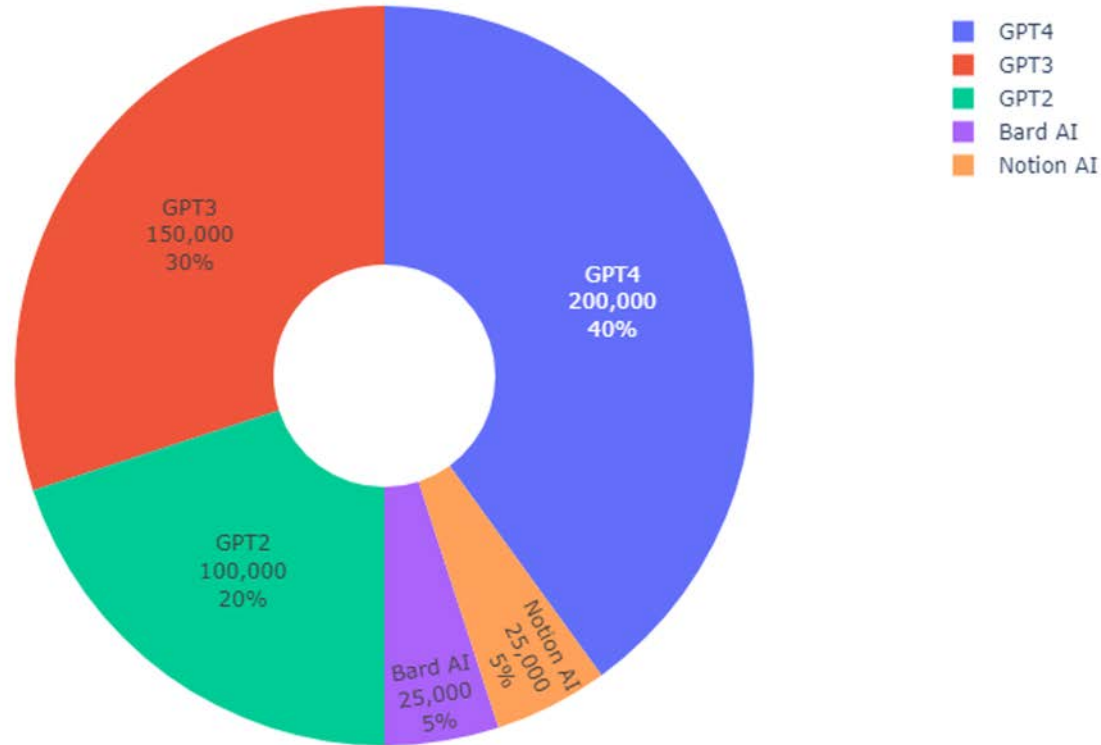


Samples from the DFDC dataset, the most state-of-the-art dataset used in DeepFake detection research in 2023

Combating DeepFakes Requires Dataset Generation: Text

DeepMedia Datasets Contains 500K+ Fake Text Samples from the Most Common AI/Large Language Models Used in Production

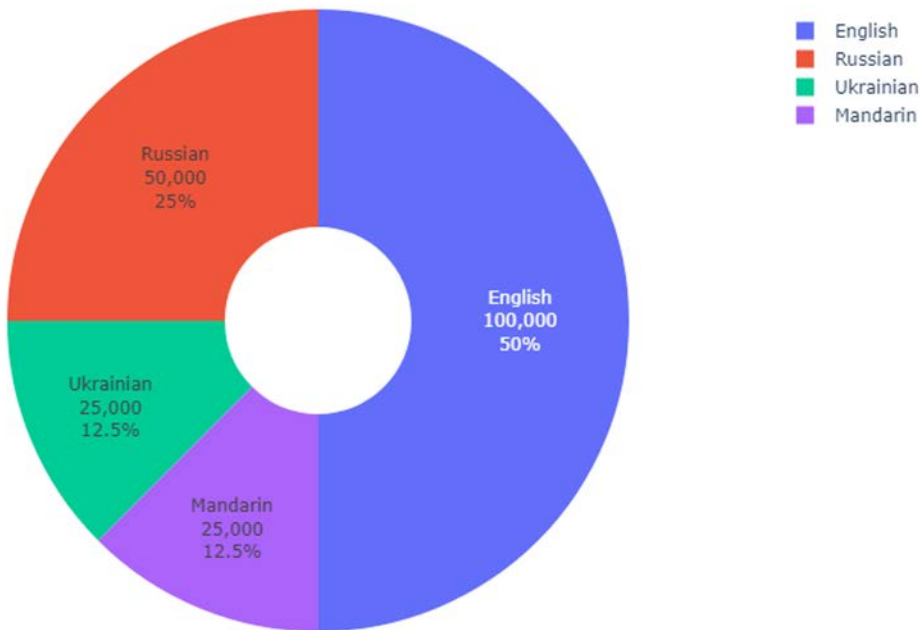
Distribution of Generator Types for DMDF_Text: Total = 500,000



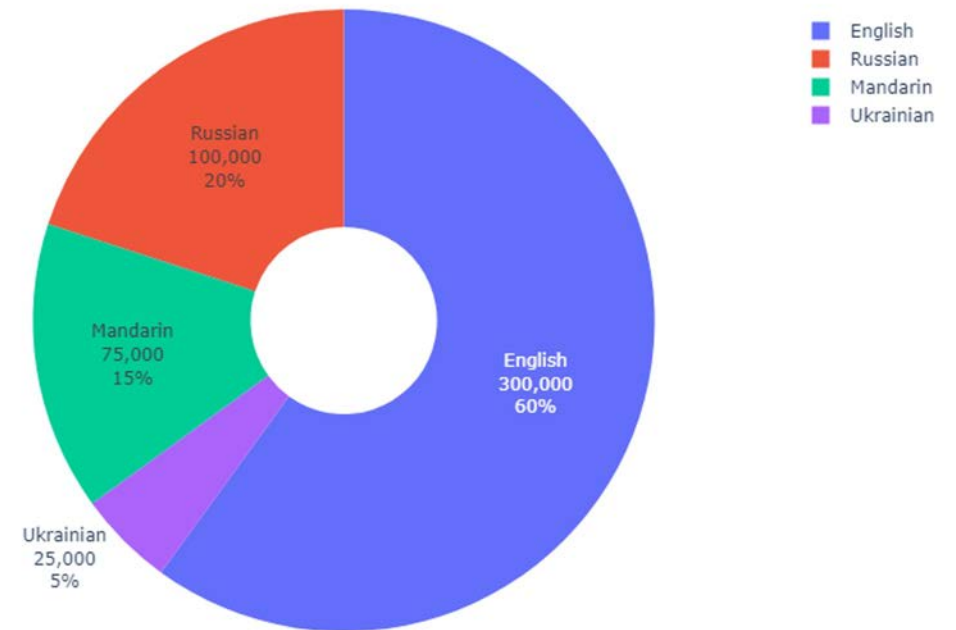
Combating DeepFakes Requires Dataset Generation: Text

DeepMedia Datasets Contains 500K+ Fake Text and 200K+ Real Text Samples Across English, Mandarin, Russian, and Ukrainian

Distribution of Languages of Interest for Real DMDF_Text: Total = 200,000



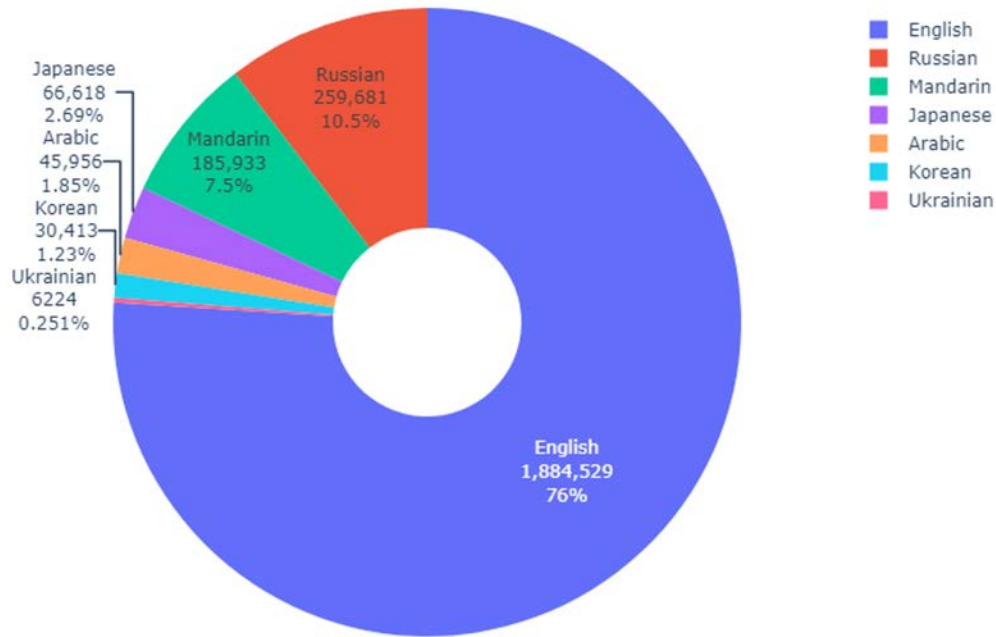
Distribution of Languages of Interest for Fake DMDF_Text: Total = 500,000



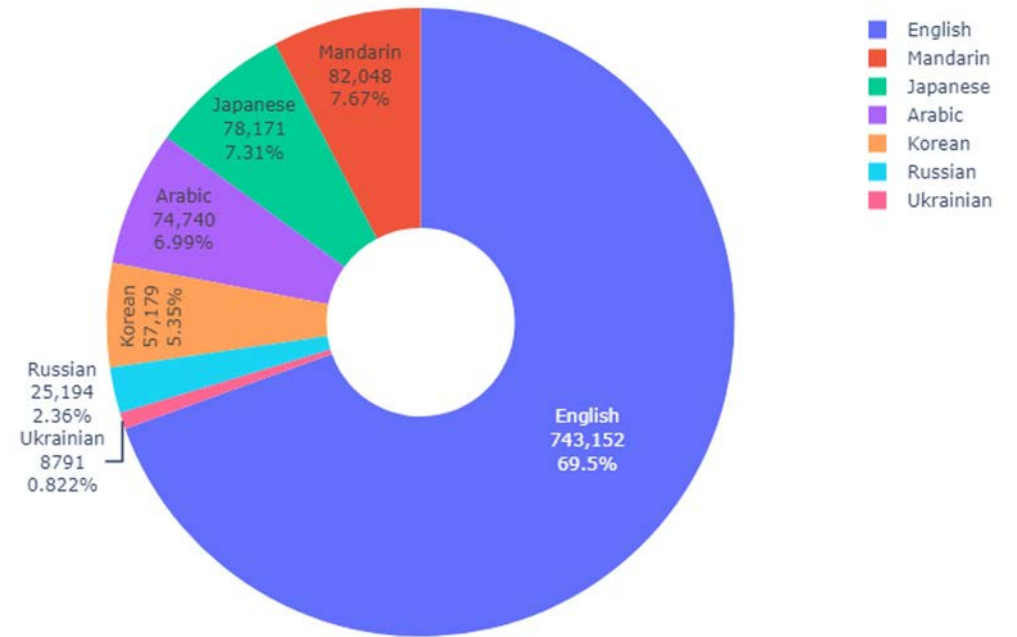
Combating DeepFakes Requires Dataset Generation: Voice

DeepMedia Datasets Contains 1M+ High-Quality Fake Voices and 2M+ High-Quality Real Voice from 10+ Languages

Distribution of Languages of Interest for Real DMDF_Voices: Total = 2,479,354



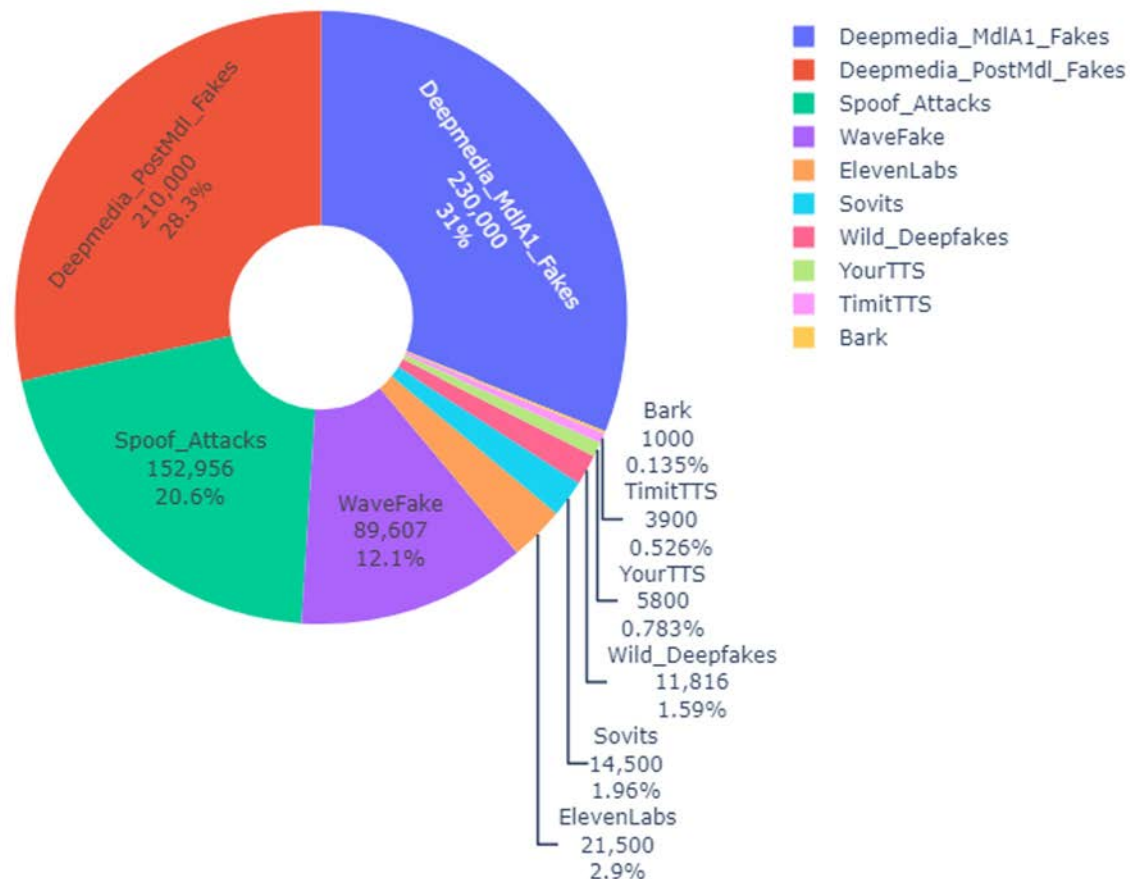
Distribution of Languages of Interest for Fake DMDF_Voices: Total = 1,069,275



Combating DeepFakes Requires Dataset Generation: Voice

DeepMedia Datasets Contains Large Amounts of Fake Voice Data from Every Major AI Voice Algorithm

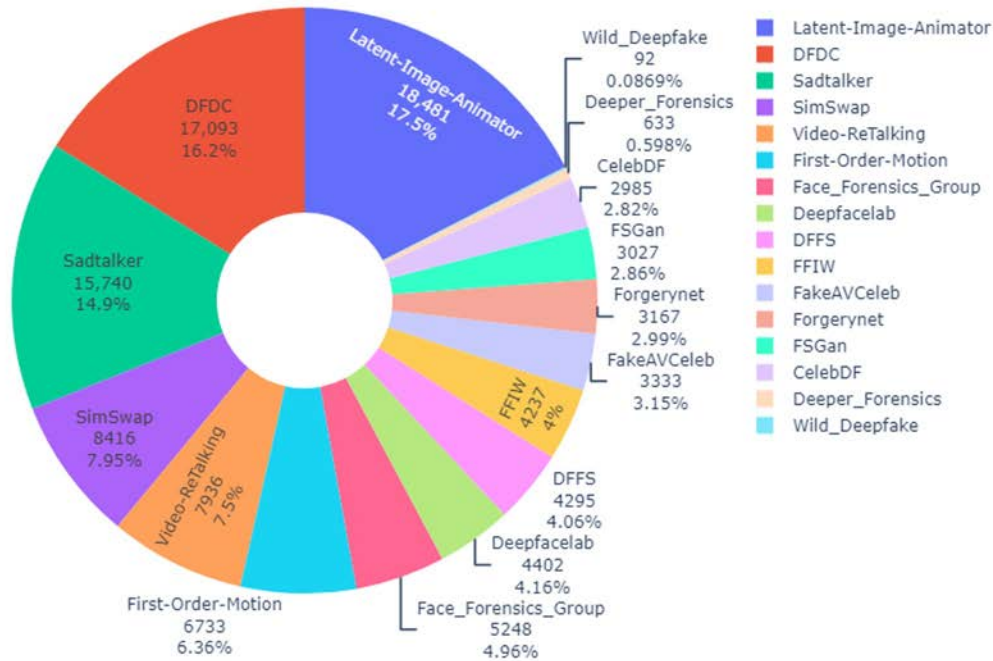
Distribution of Generator Types for DMDF_Voices: Total = 10 Unique Generators



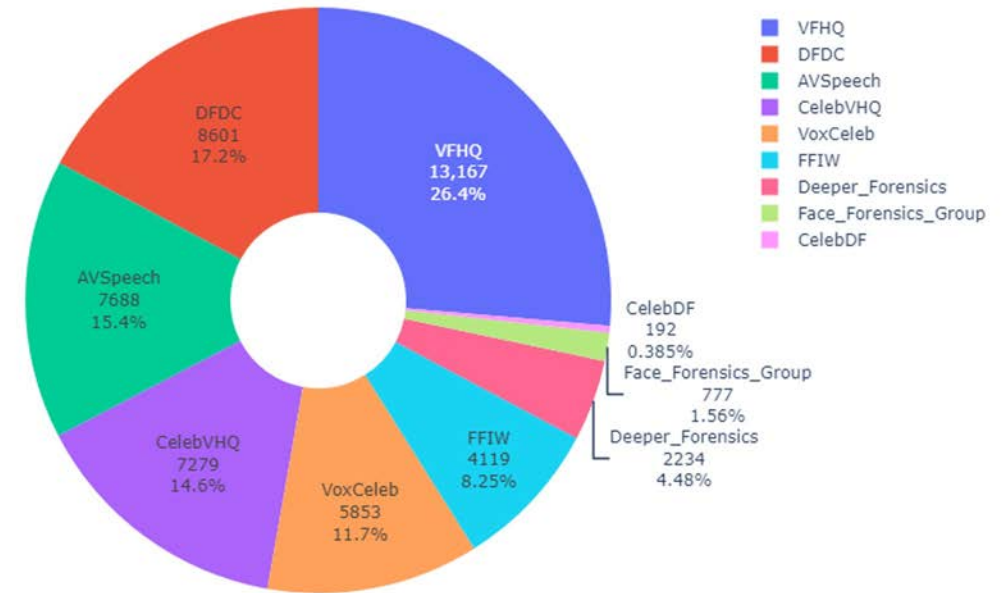
Combating DeepFakes Requires Dataset Generation: Face

DeepMedia Datasets Contains 100K+ High-Quality Fake and 50K+ High-Quality Real Faces from a Diverse Range of Data Sources

Distribution of Generator Types for DMDF_Faces: Total = 105,818



Distribution of Real Data for DMDF_Faces: Total = 50,000



Training and Evaluating Algorithms to Detect AI Manipulations in Text, Voices, and Faces w/ High Accuracy

State of the Art Classification Algorithms



GPT Zero

Uses a pretrained LLM to calculate burstiness and perplexity of potentially generated text to determine an overall detection probability.



OpenAI Classifier

A custom-trained LLM specifically tuned to classify text and authentic or AI-generated with an associated confidence score.



Copy Leaks Detector

A classic text-based classification algorithm that is trained to detect AI-generated text through standard NLP-based classification.



Soxan Detector

A novel Wav to Vector based Deepfake Detector trained by predicting speech units for masked speech. Proprietary to DeepMedia.



Efficient Audio Tagging Detector

A unique Convolution Neural Network(CNN) based approach trained using Knowledge Distillation from Transformers



Attack Agnostic Dataset

An Audio Detection model based upon a CNN, which is characterized by good generalization and stability results.



Reversible Vision Transformers

A memory efficient Video Detection architecture designed for visual recognition. Proprietary to DeepMedia.



EfficientNet-and-Vision-Transformers

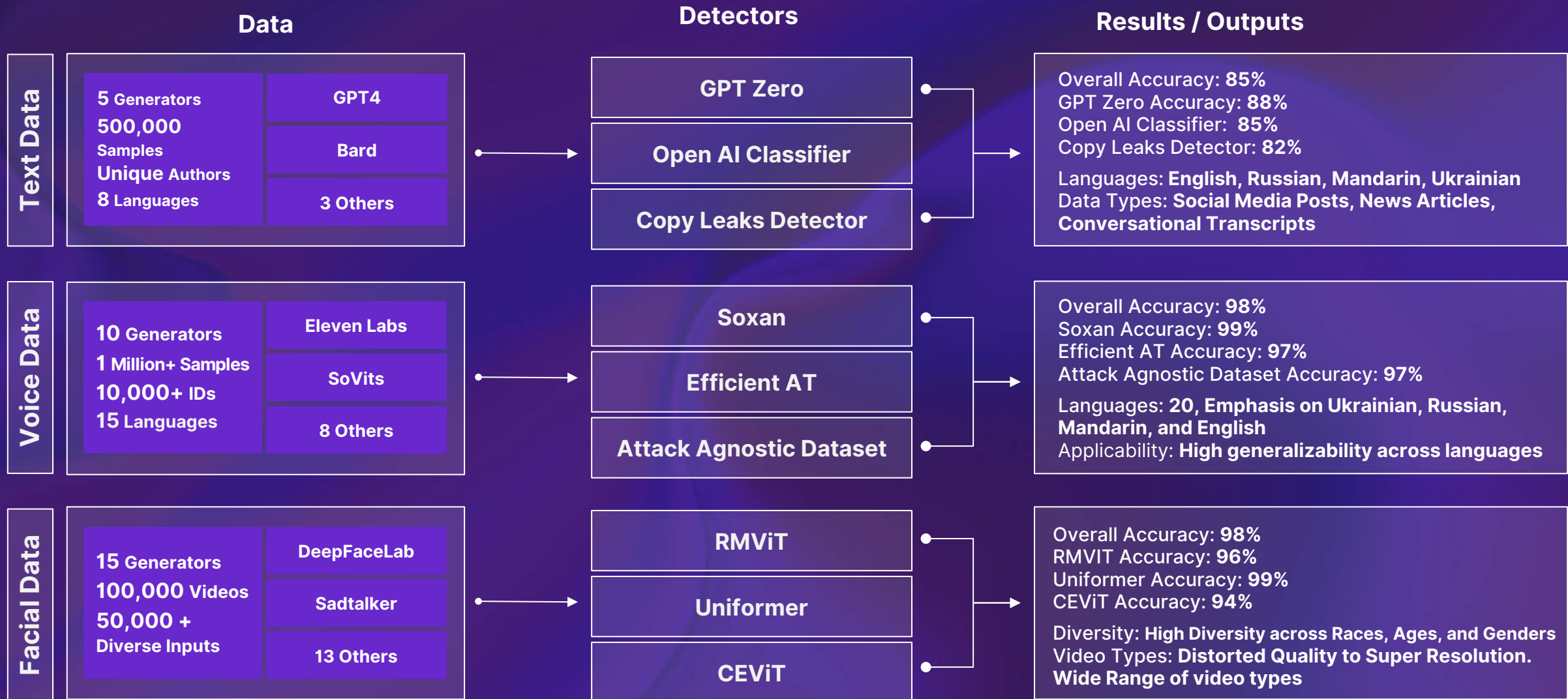
Powerful Deepfake Detection Models that combine an Efficient Network and a Vision Transformer.



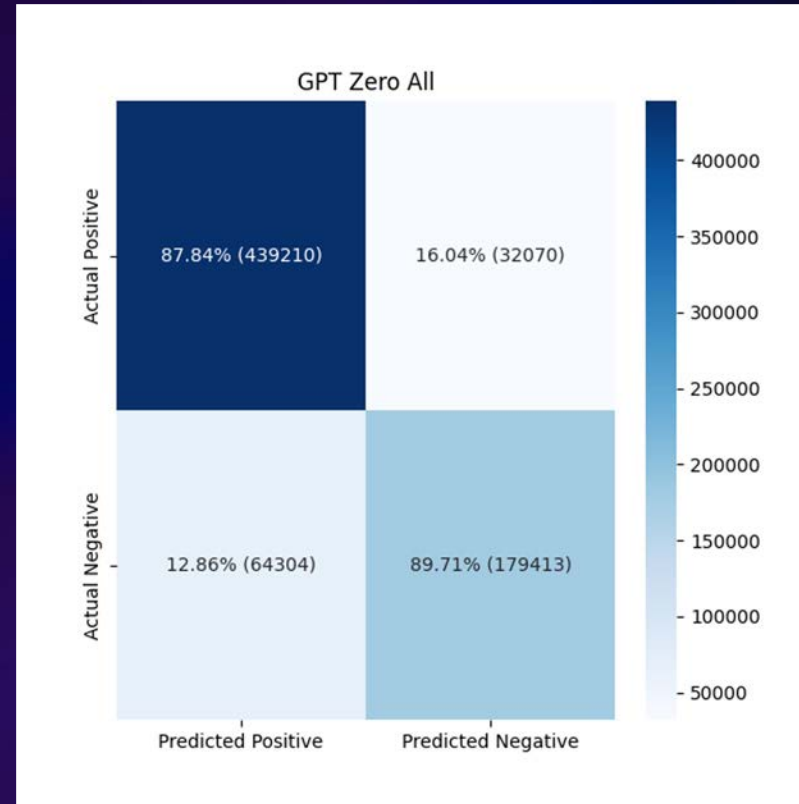
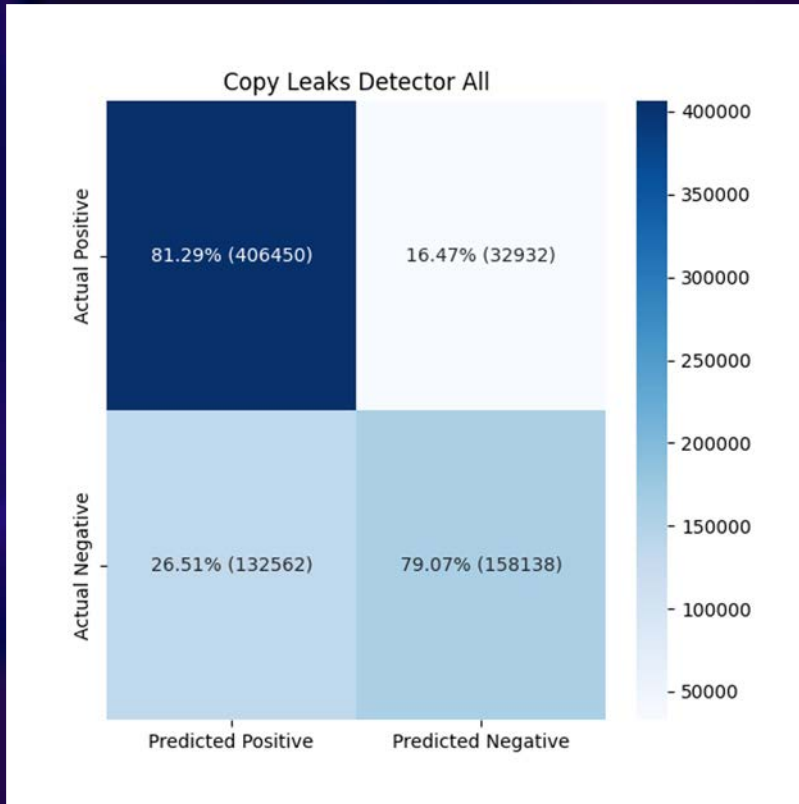
Uniformer V2

A Video Detection model that enhances the power of Video Transformers with a Video Uniformer backbone. Proprietary to DeepMedia.

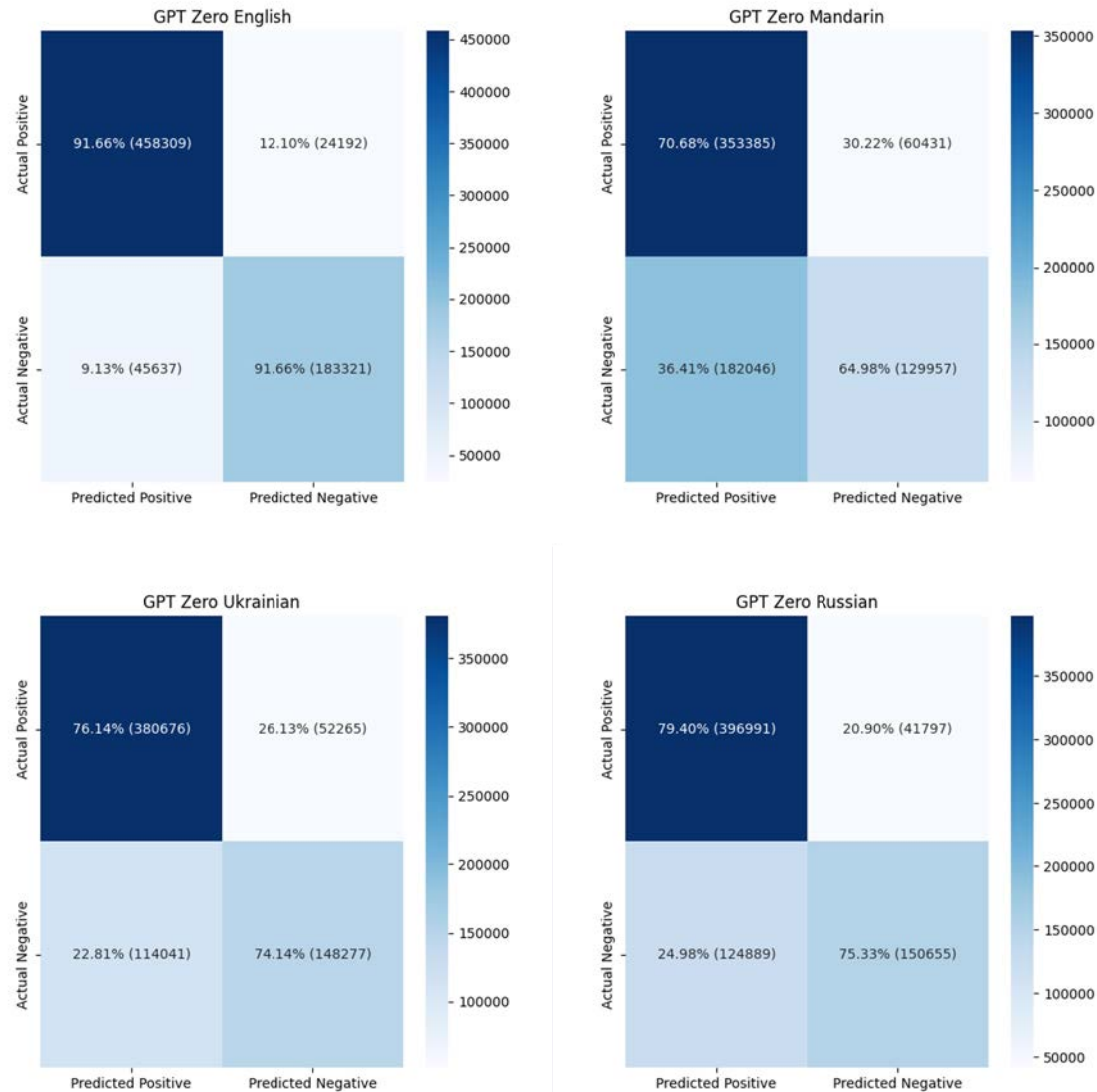
Training Details And Results



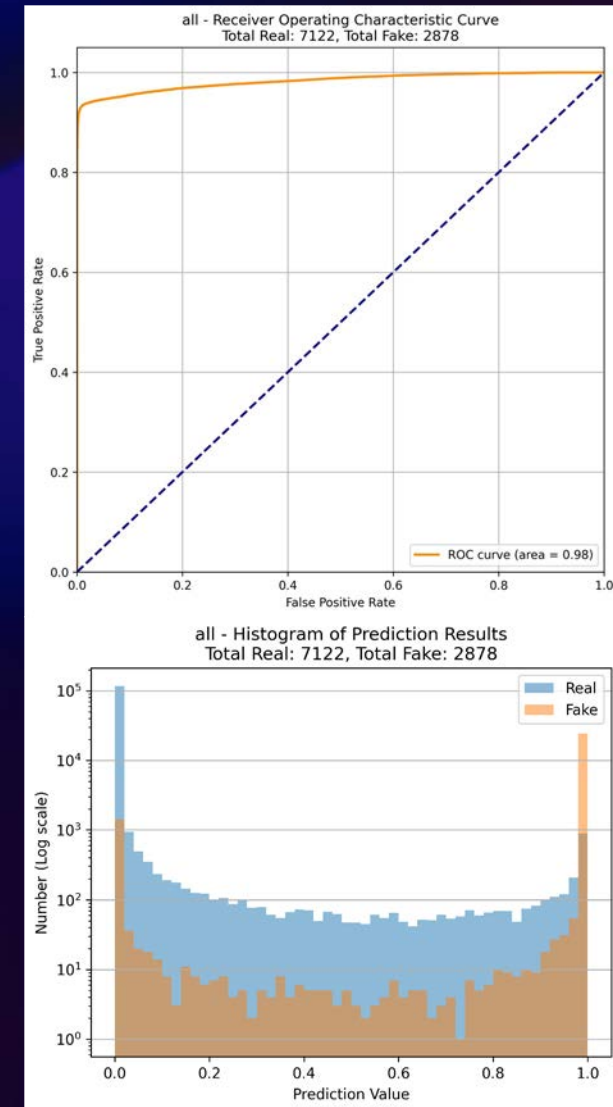
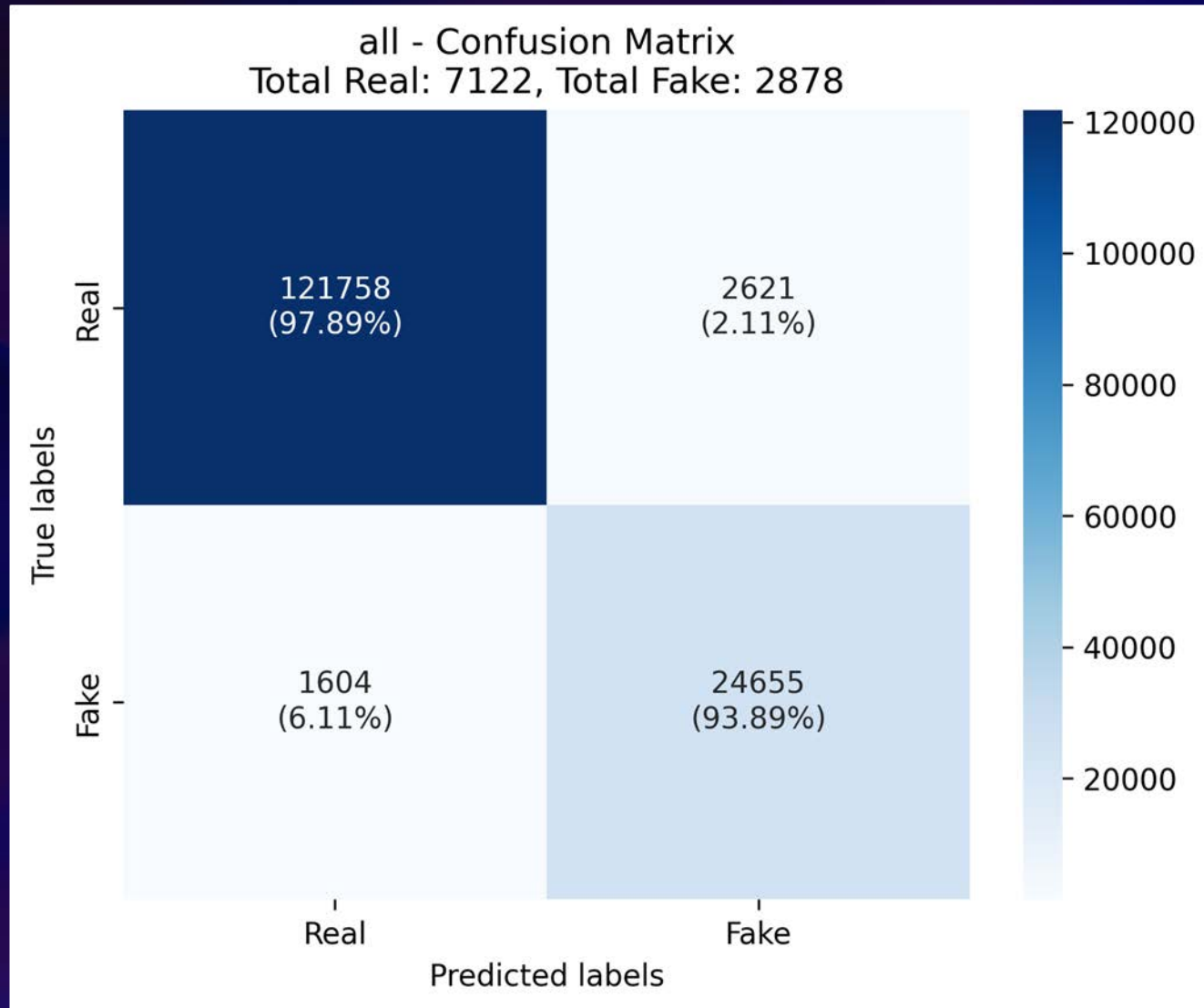
Text Detection Results



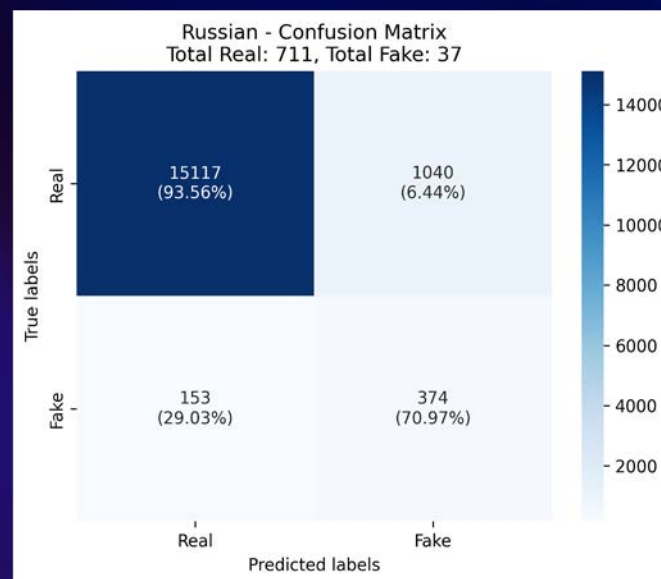
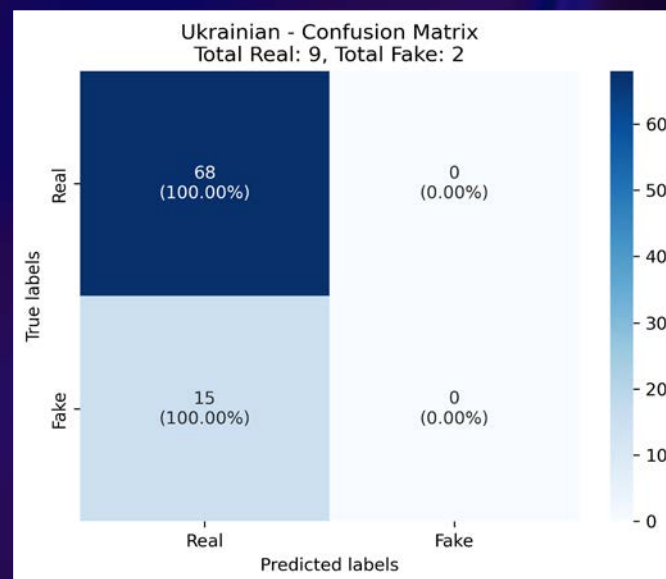
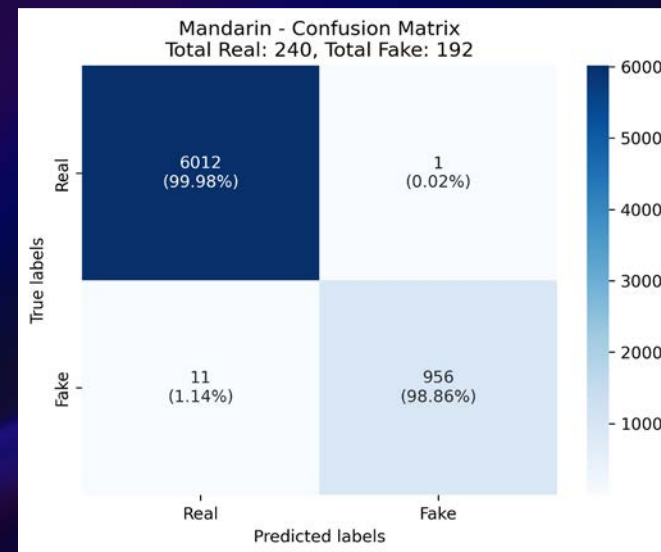
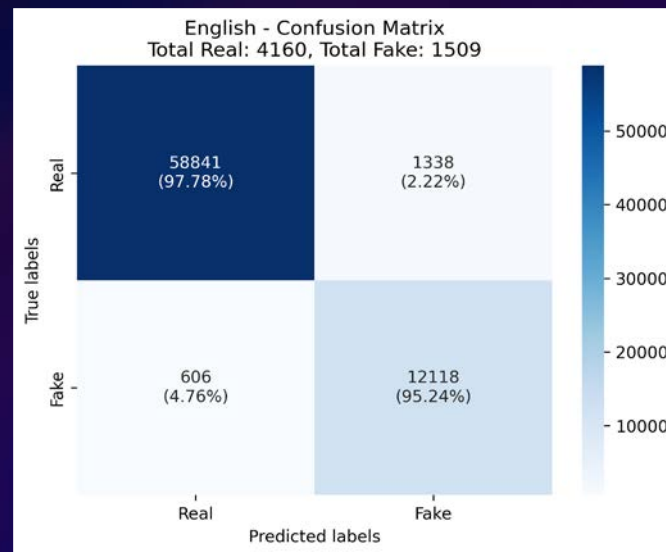
Text Detection Results



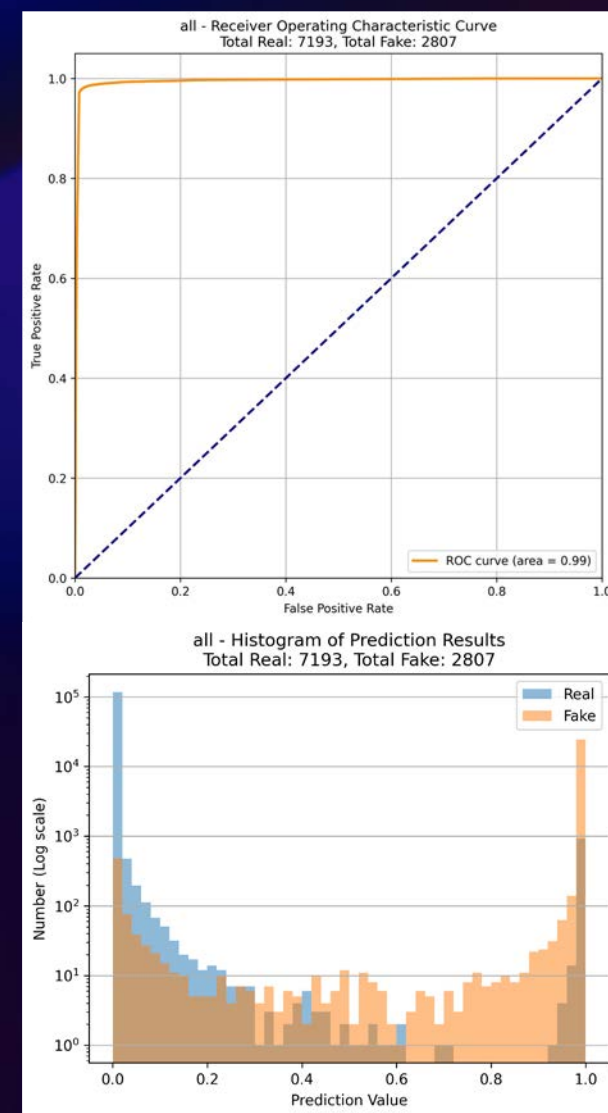
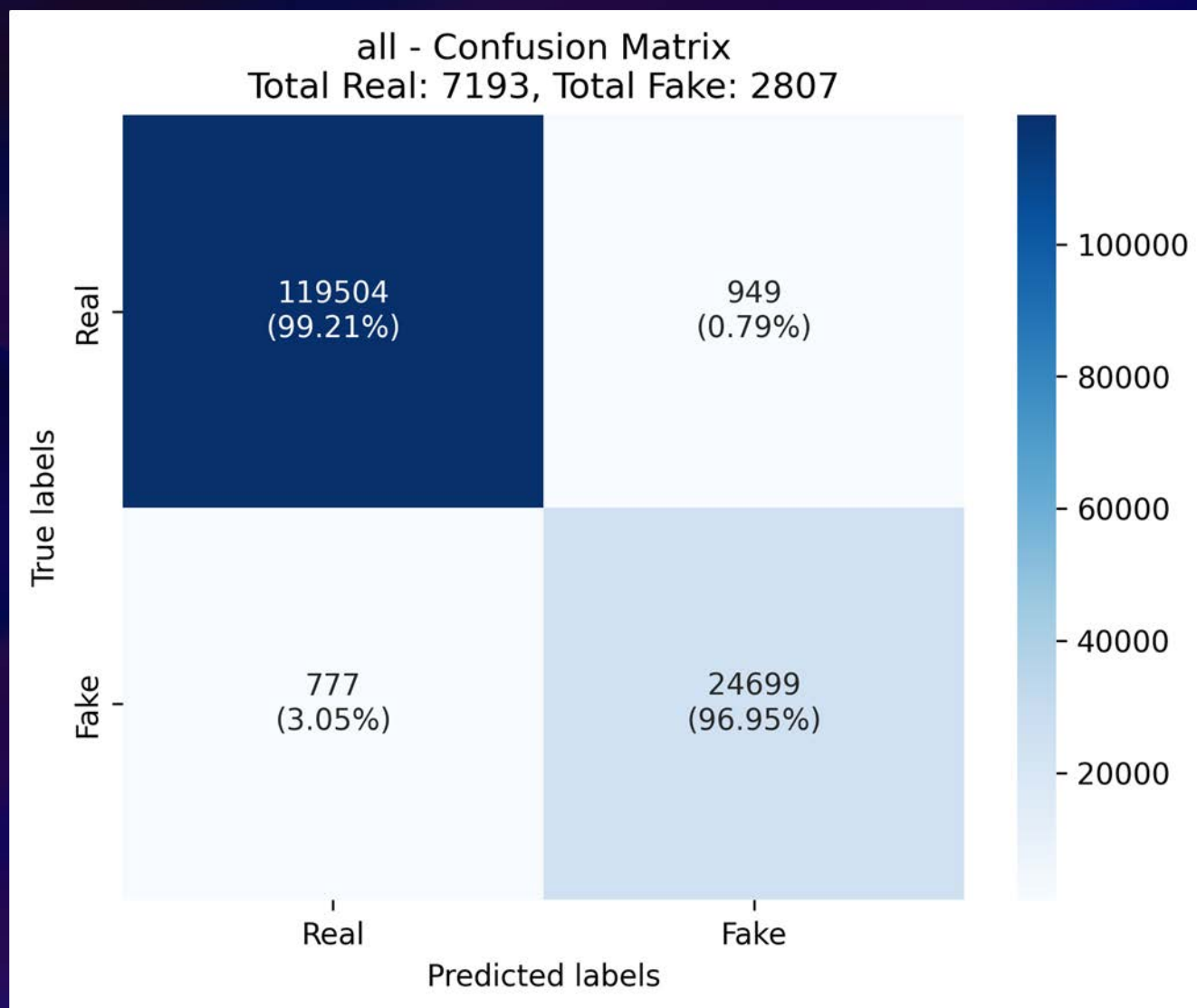
Voice Detection Results - AAD



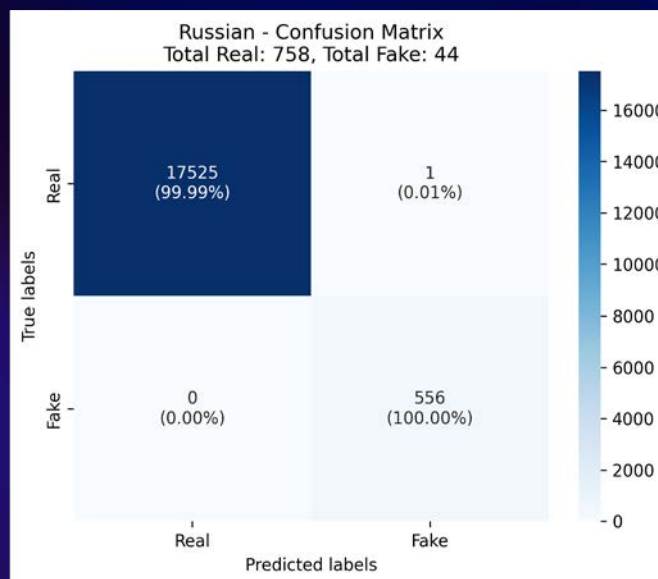
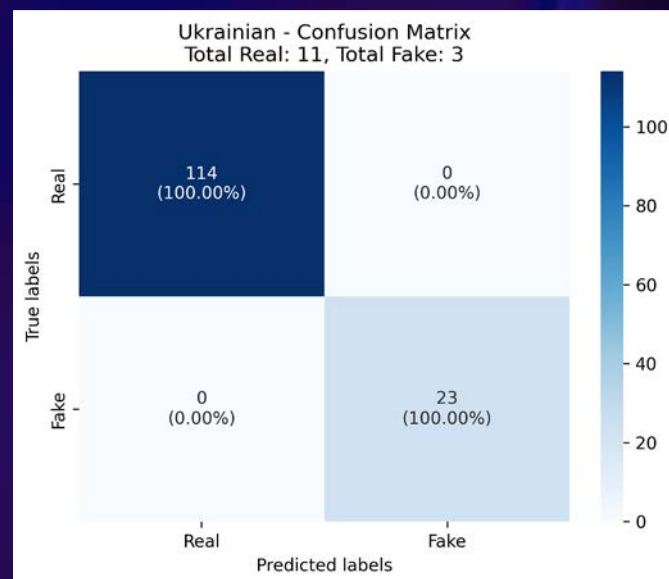
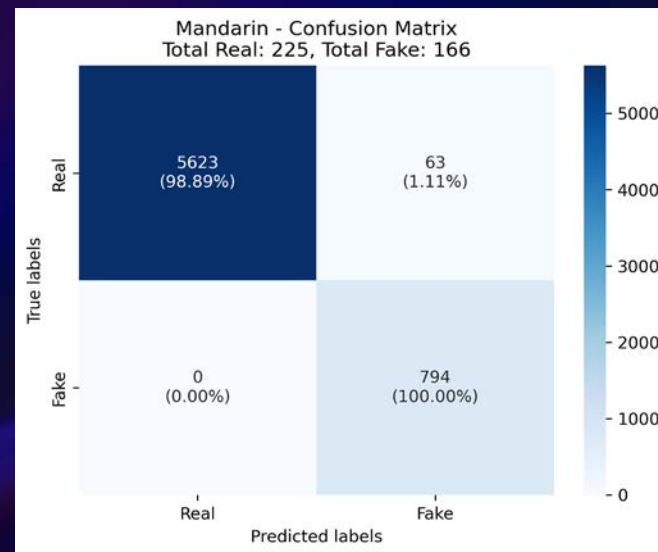
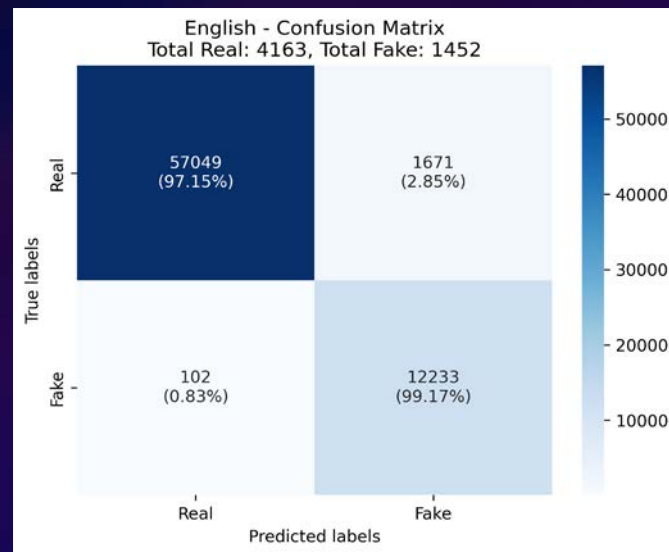
Voice Detection Results - AAD



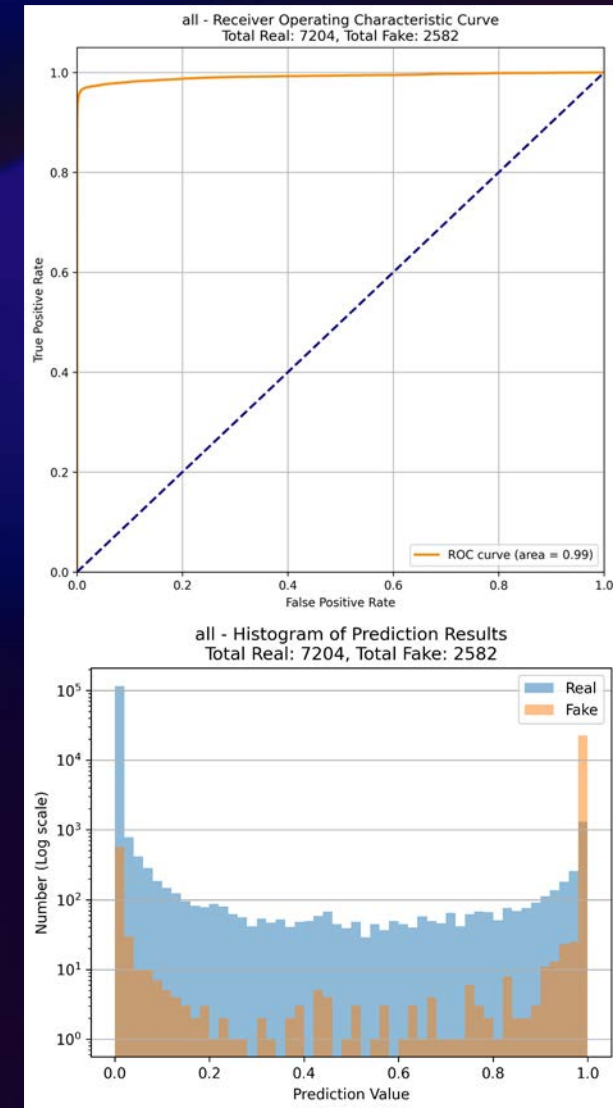
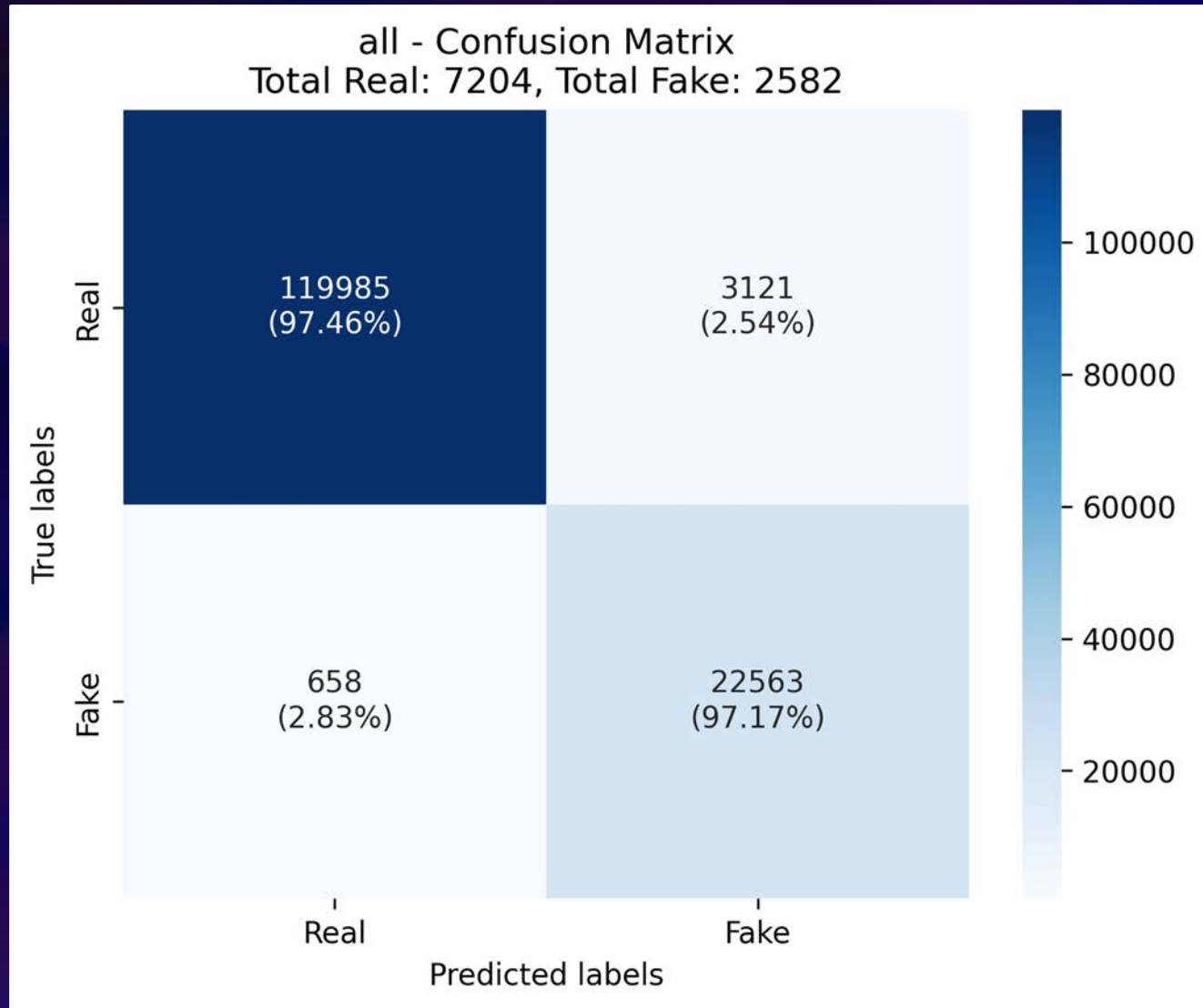
Voice Detection Results - EAT



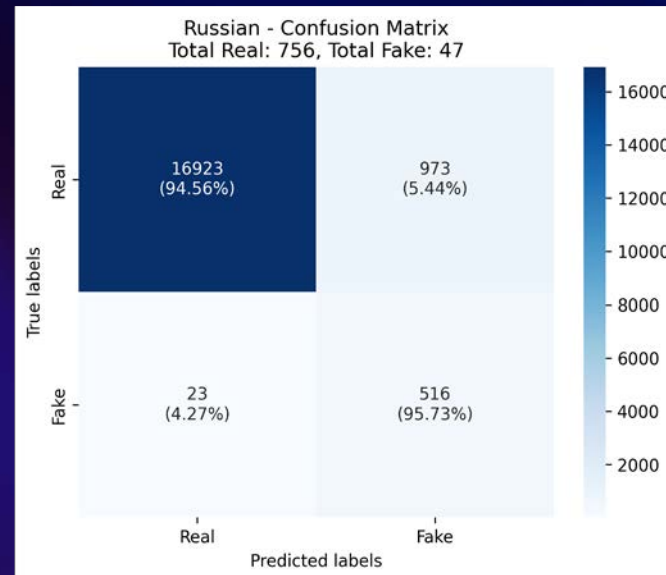
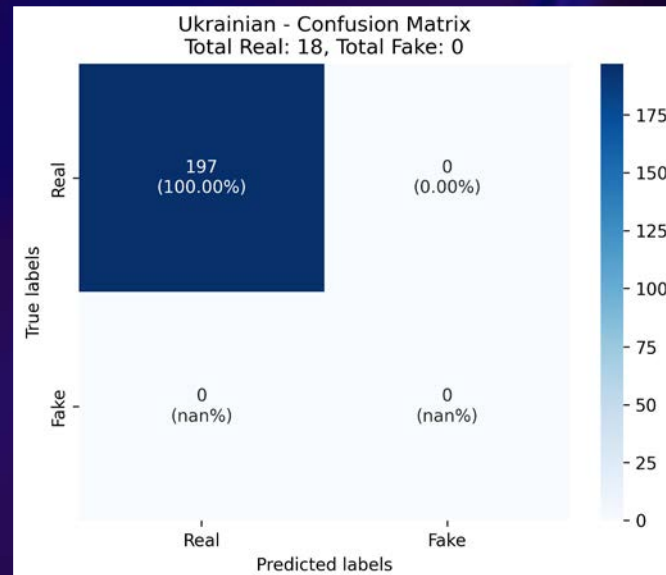
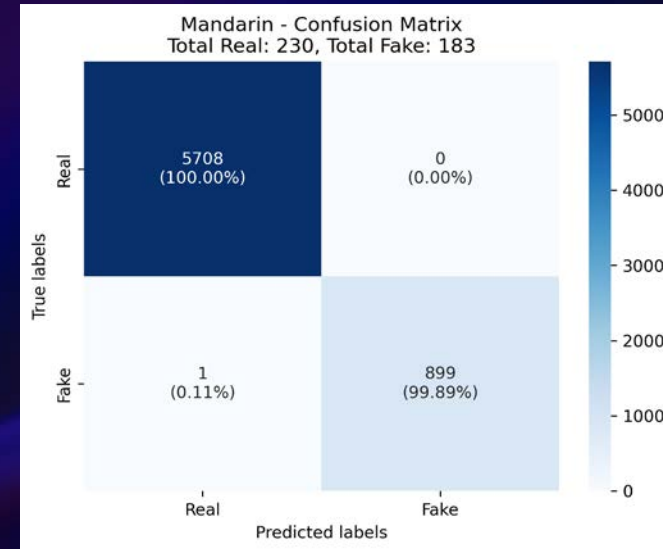
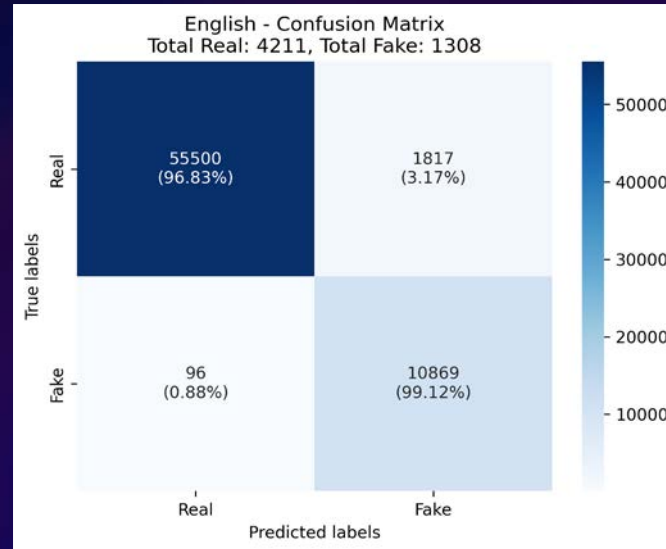
Voice Detection Results - EAT



Voice Detection Results - Soxan



Voice Detection Results - Soxan

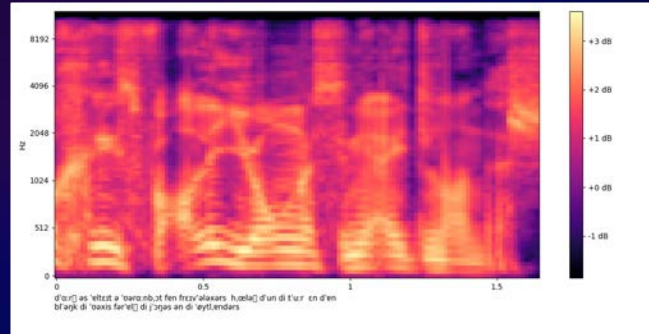


Adding Audio Augmentations

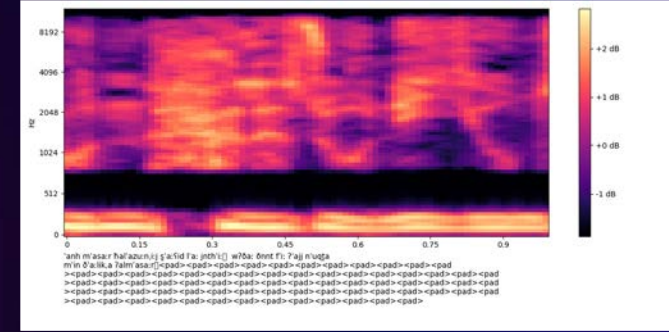
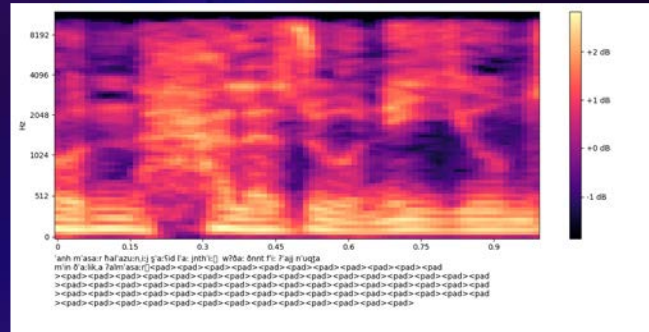
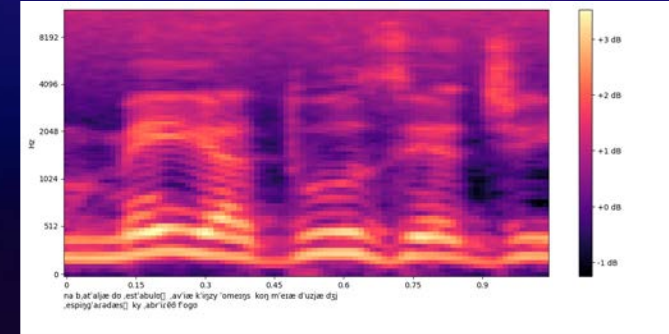
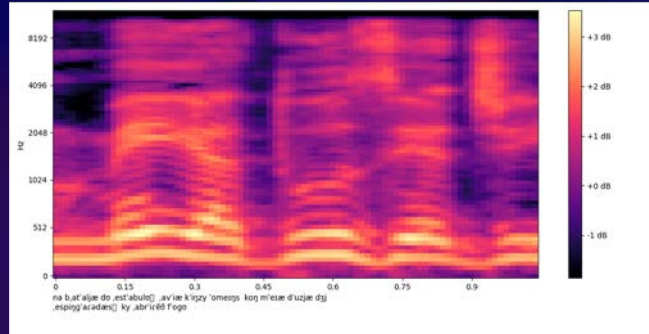
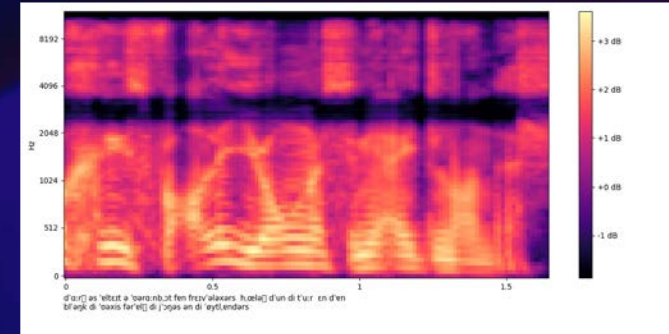
```
apply_augmentation = Compose(
    transforms=[
        Shift(
            min_shift=0.05,
            max_shift=0.05,
            shift_unit="seconds",
            p=0.2,
            sample_rate=24000,
            rollover=False,
            output_type="dict",
        ),
        Gain(
            min_gain_in_db=-3.0,
            max_gain_in_db=3.0,
            p=0.2,
        ),
        Padding(
            p=0.1,
            sample_rate=WAVE_FAKE_SR
        ),
        PolarityInversion(
            p=0.5,
        ),
        ApplyImpulseResponse(
            ir_paths="AudioFiles/RIRS_NOISES_16K/real_rirs_isotropic_noises",
            p=0.1,
        ),
        BandStopFilter(p=0.2),
        LowPassFilter(p=0.2),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/crowds_non_laughter_heavyvocals",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/DNSFreesound",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/MUSANmusicremovedvoice",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/MUSANnoiseS80only",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/music_no_voice",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/nature_ambient_sounds",
            min_snr_in_db=12.0,
            max_snr_in_db=40.0,
            p=0.1
        ),
        AddBackgroundNoise(
            background_paths="AudioFiles/Noise_wavs_organized_16K/nature_animal_sounds",

```

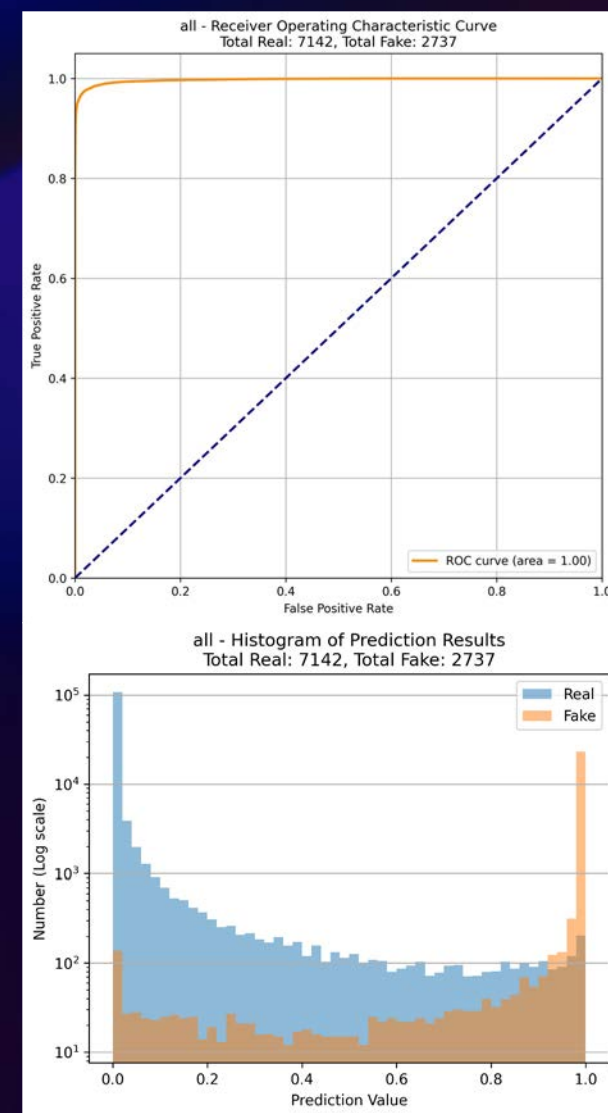
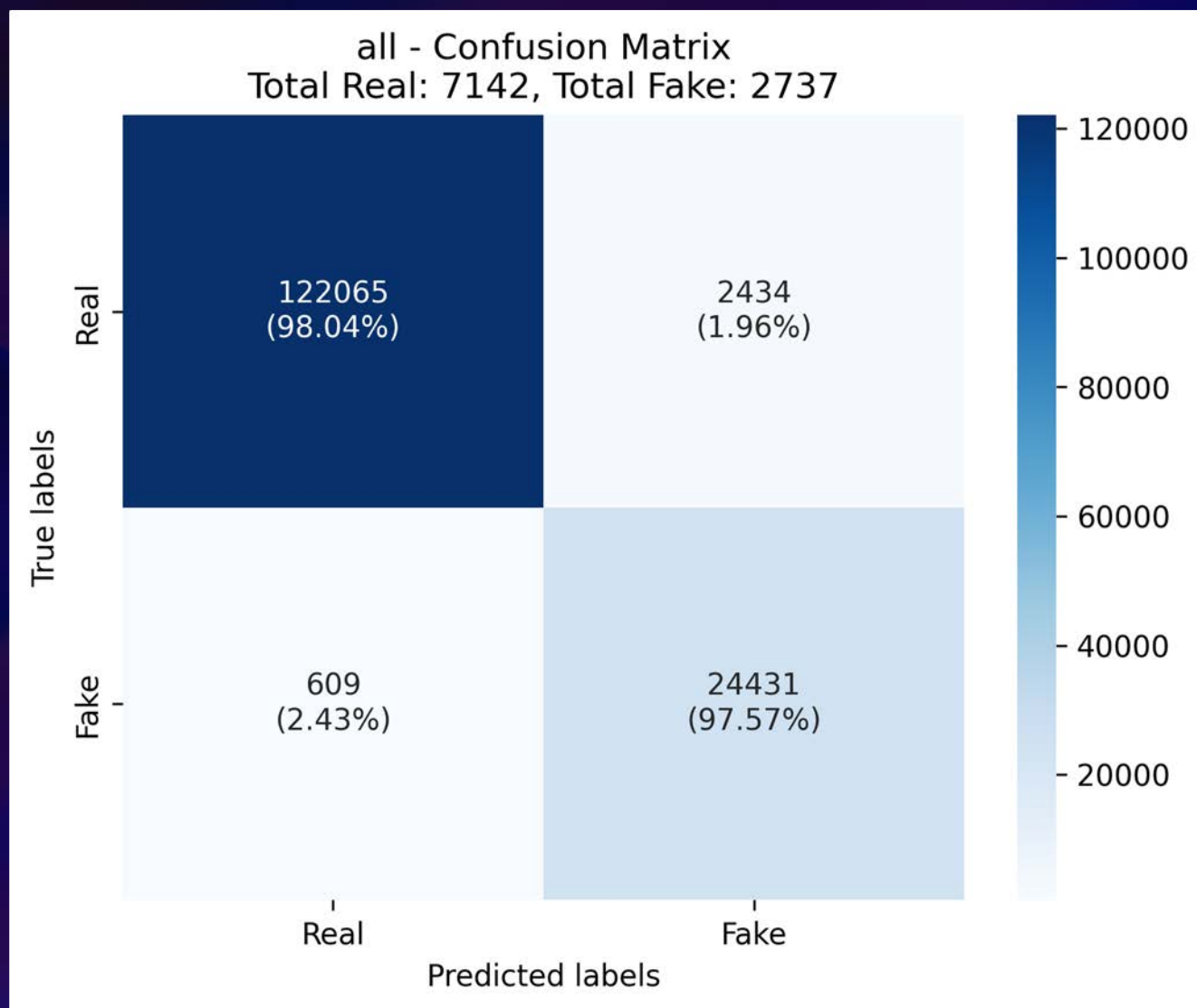
Original



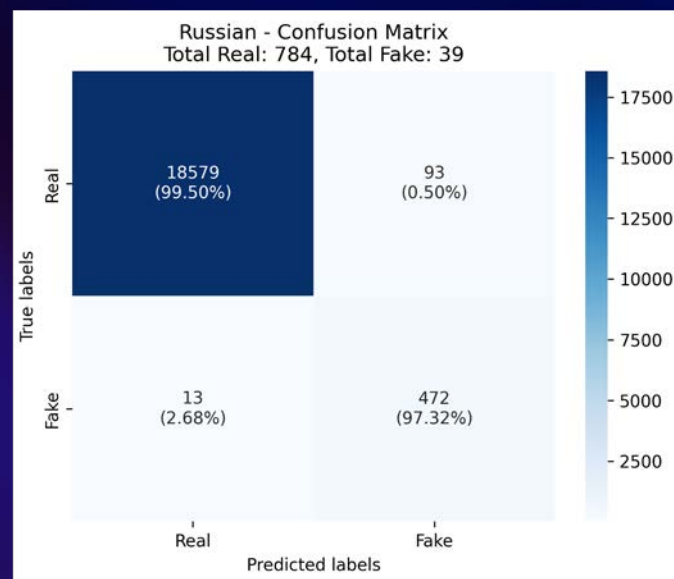
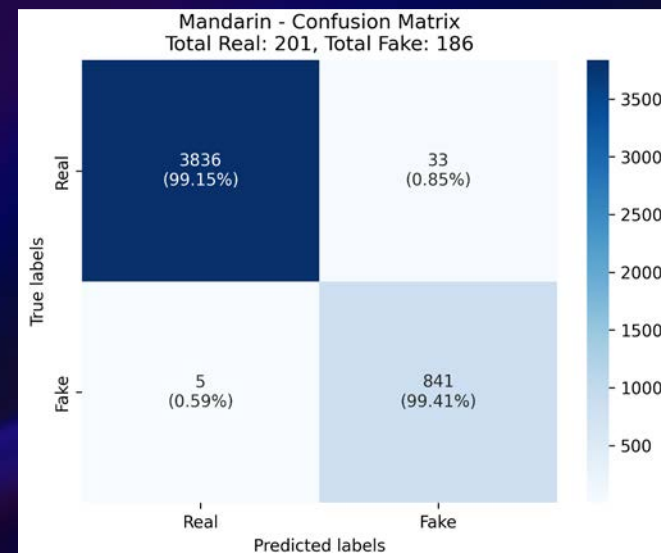
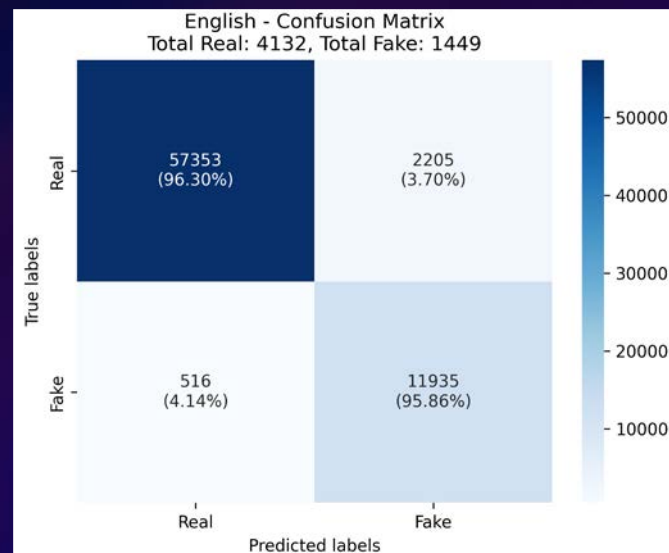
Augmented



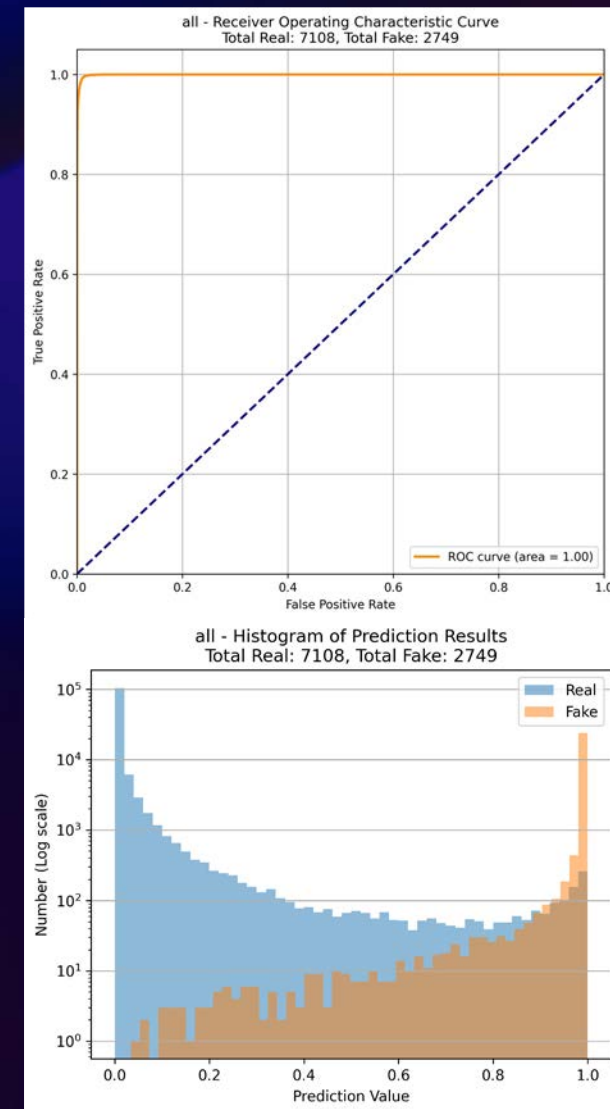
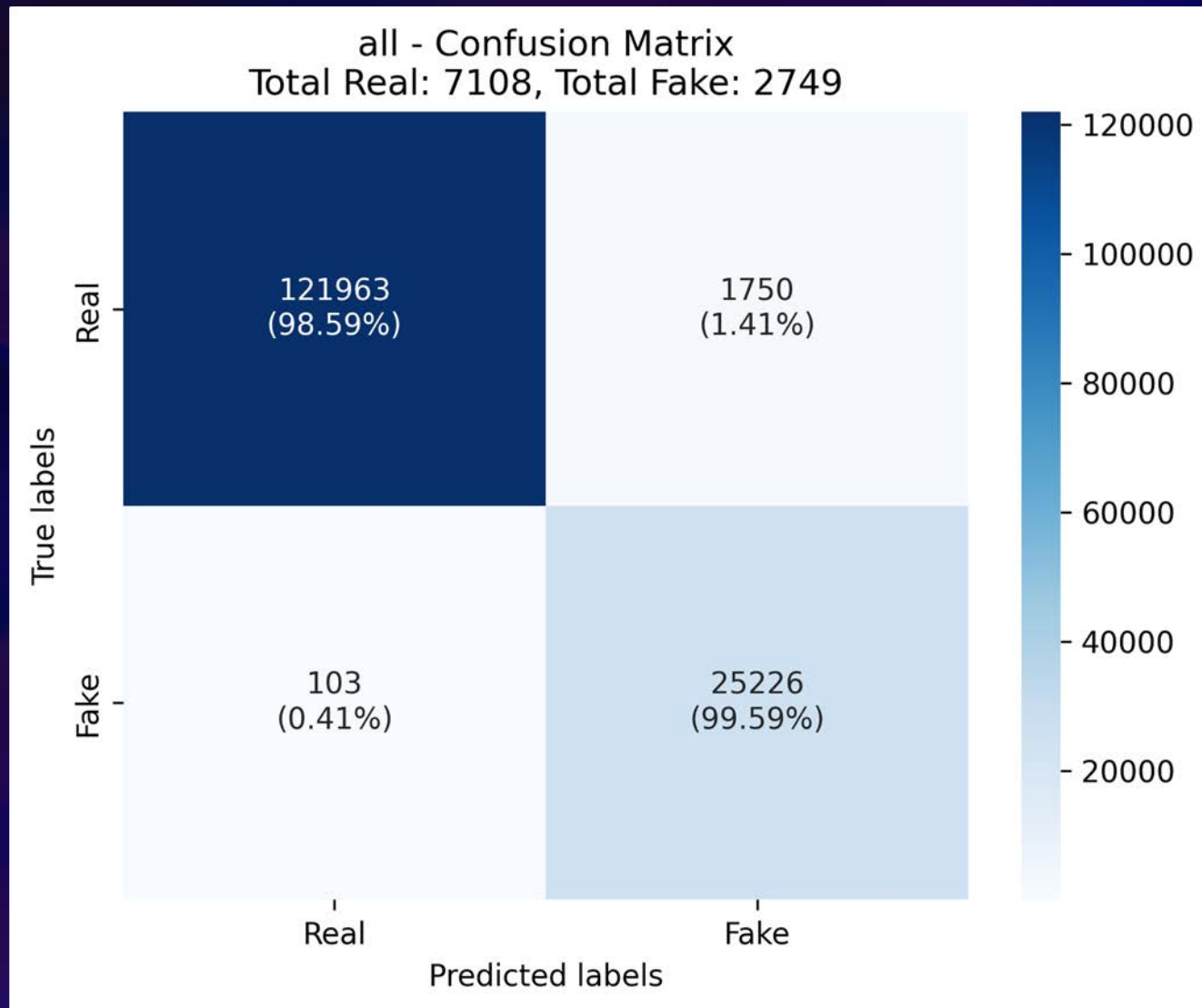
Voice Detection Results - AAD Augmented



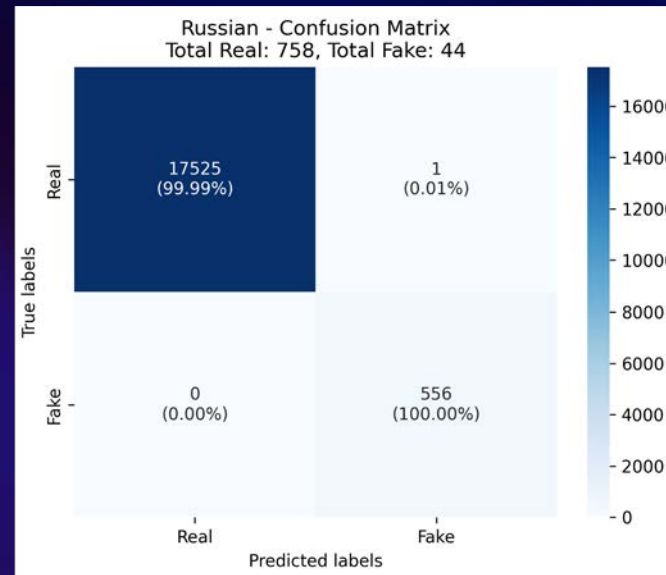
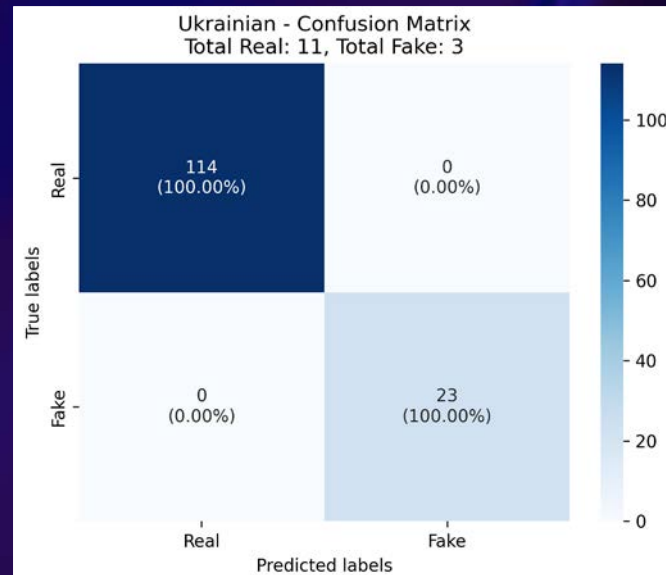
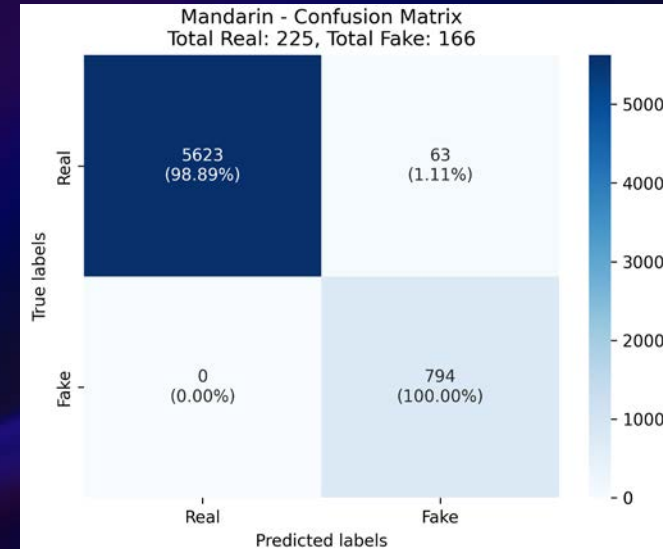
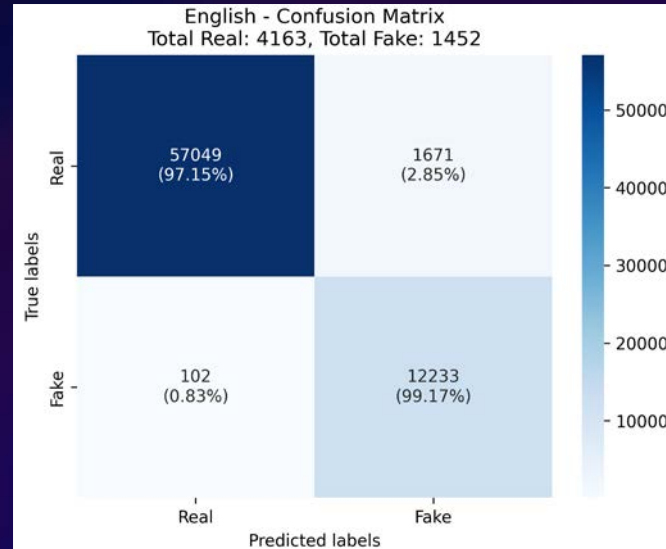
Voice Detection Results - AAD Augmented



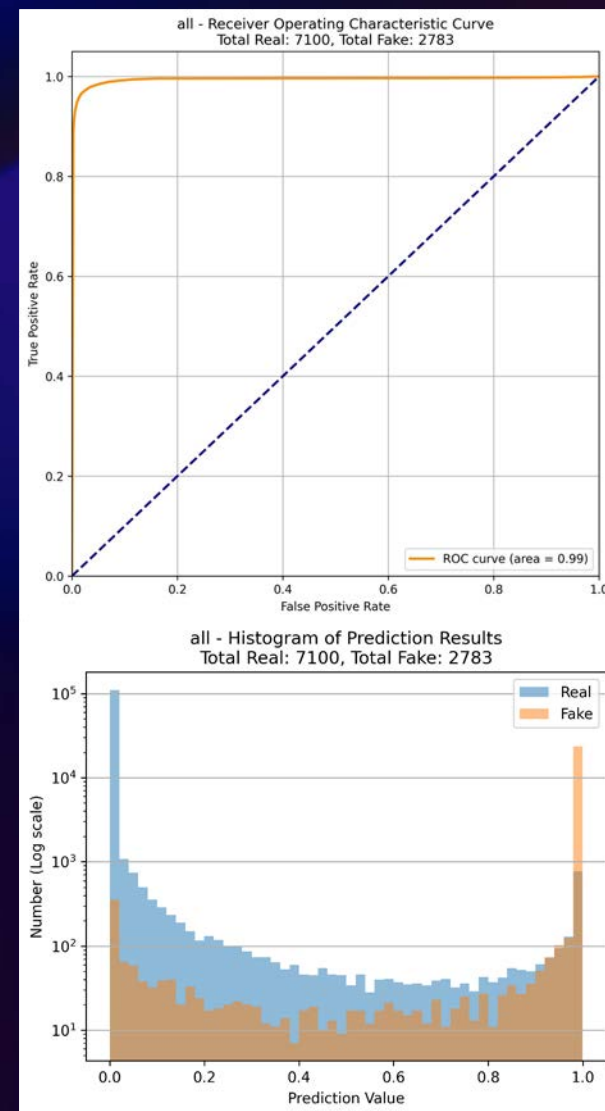
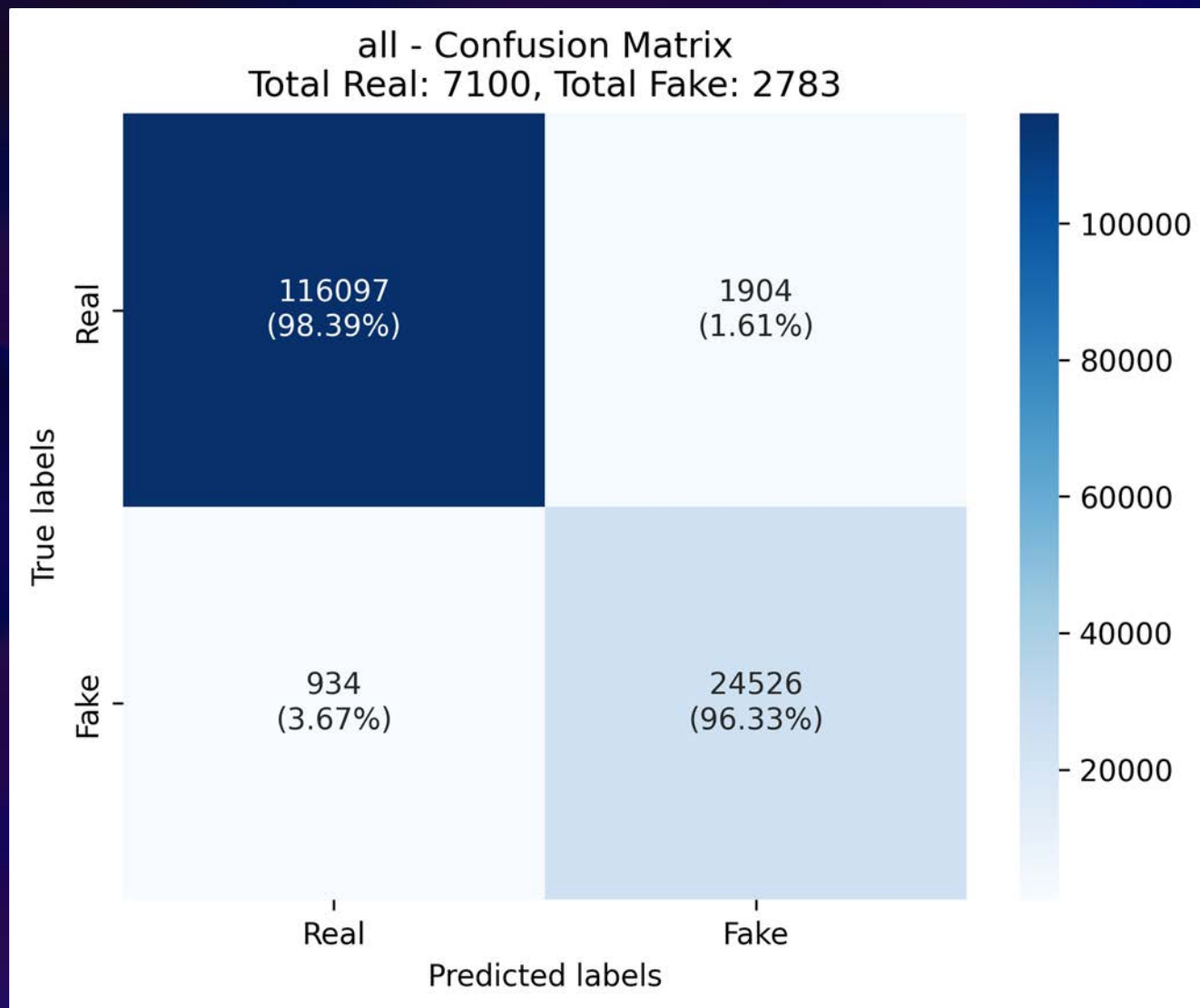
Voice Detection Results - EAT Augmented



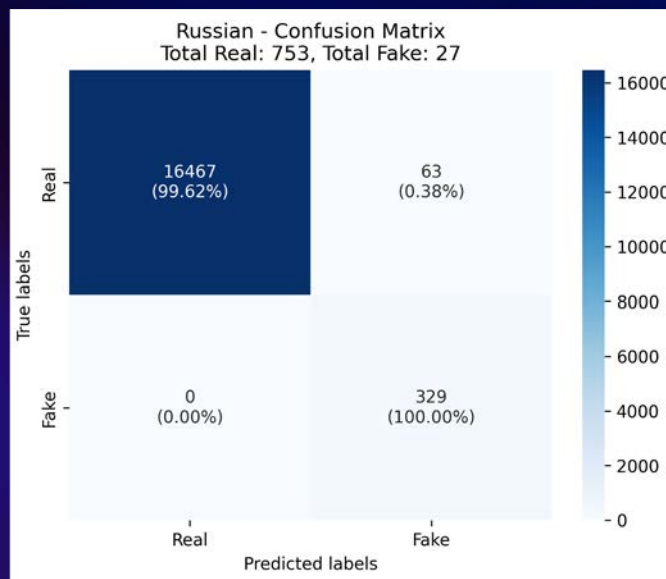
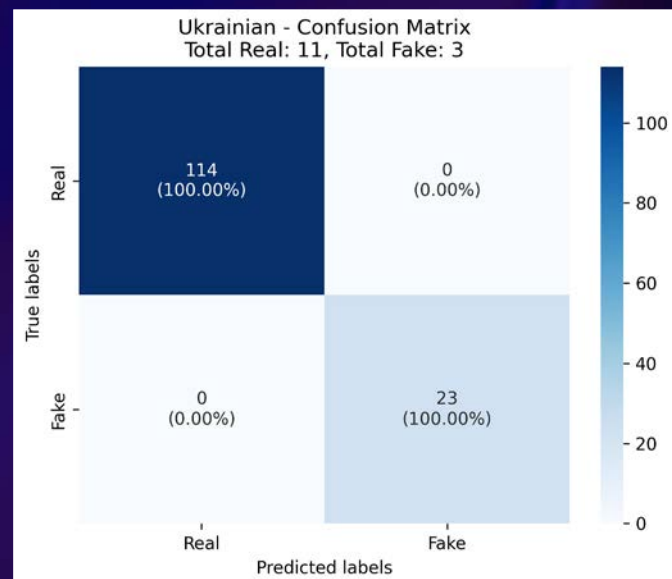
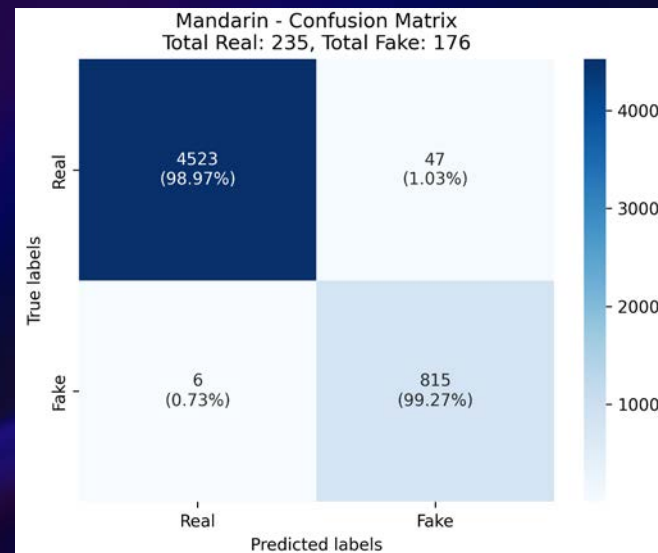
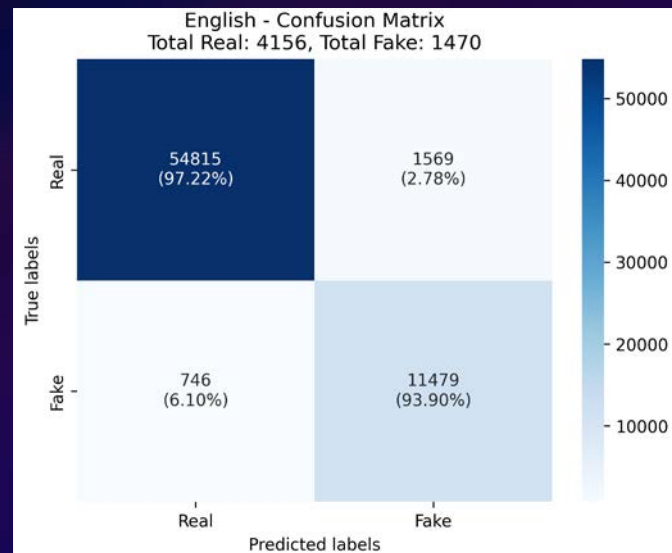
Voice Detection Results - EAT Augmented



Voice Detection Results - Soxan Augmented



Voice Detection Results - Soxan Augmented



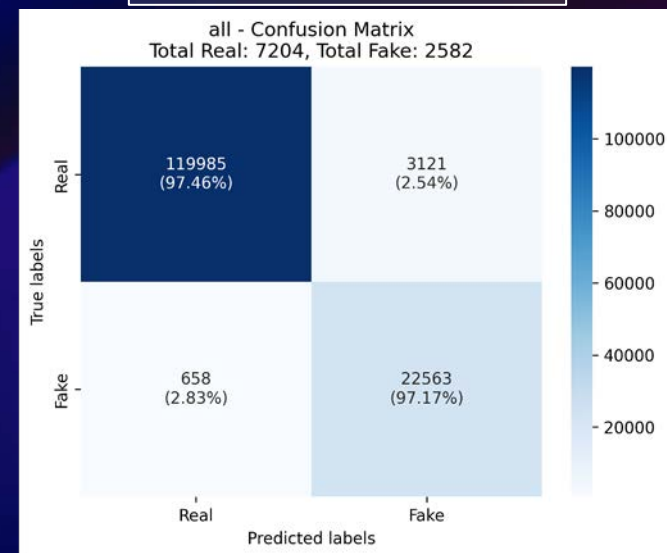
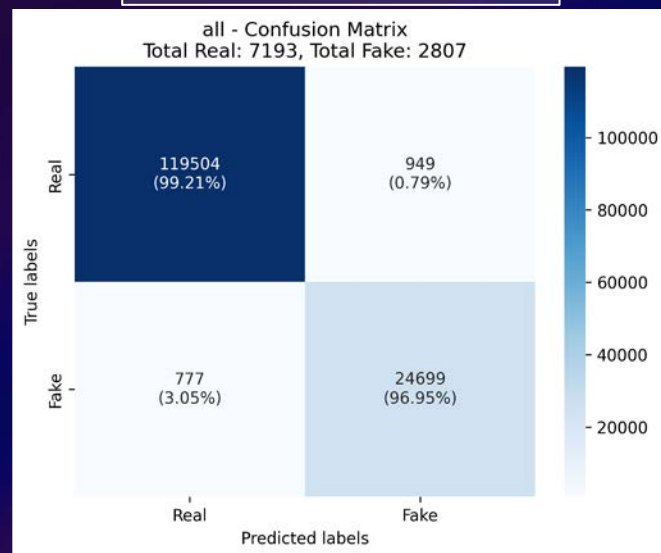
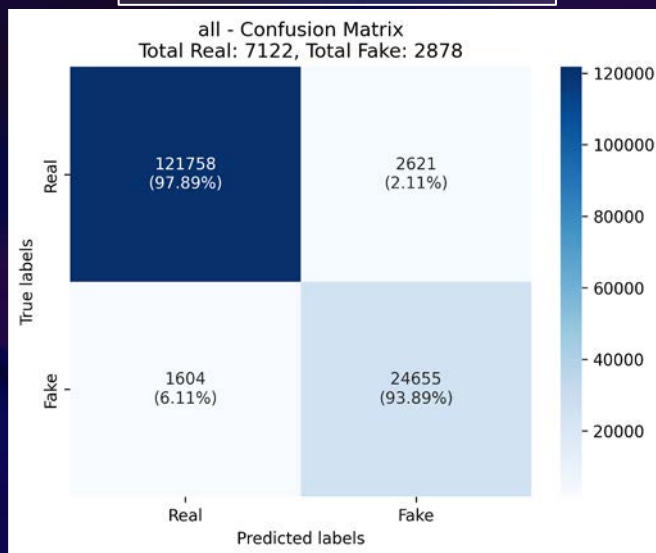
Voice Detection Results - Augmentation Effects

AAT

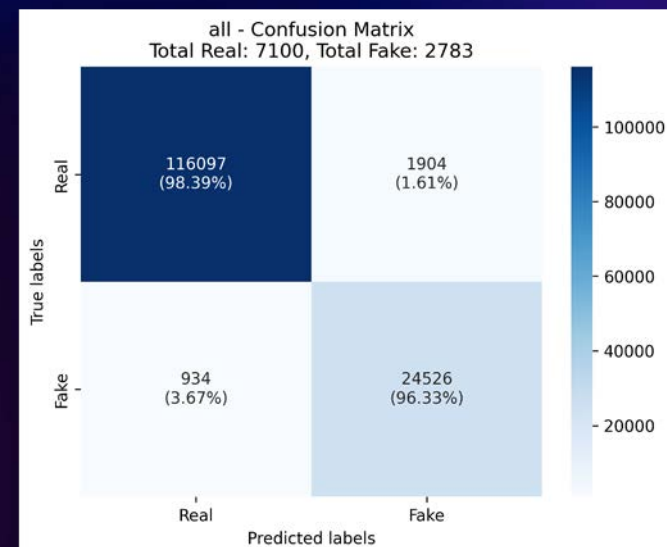
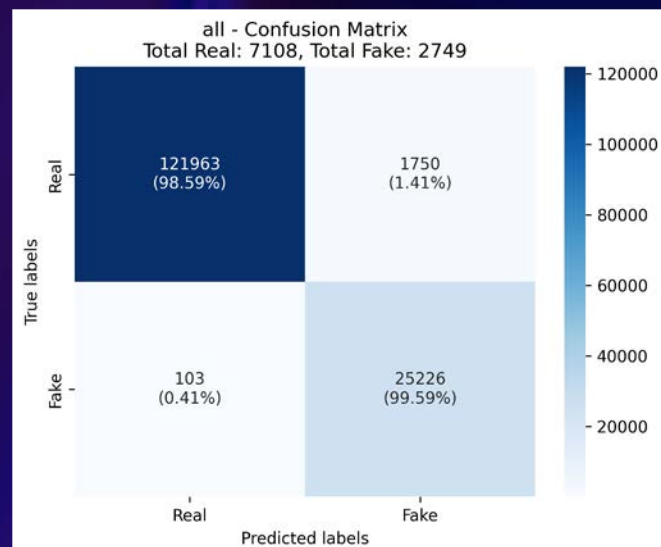
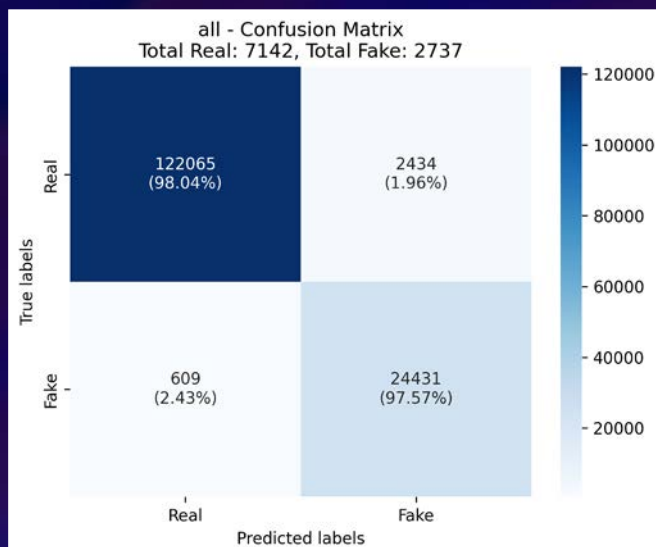
EAT

SOXAN

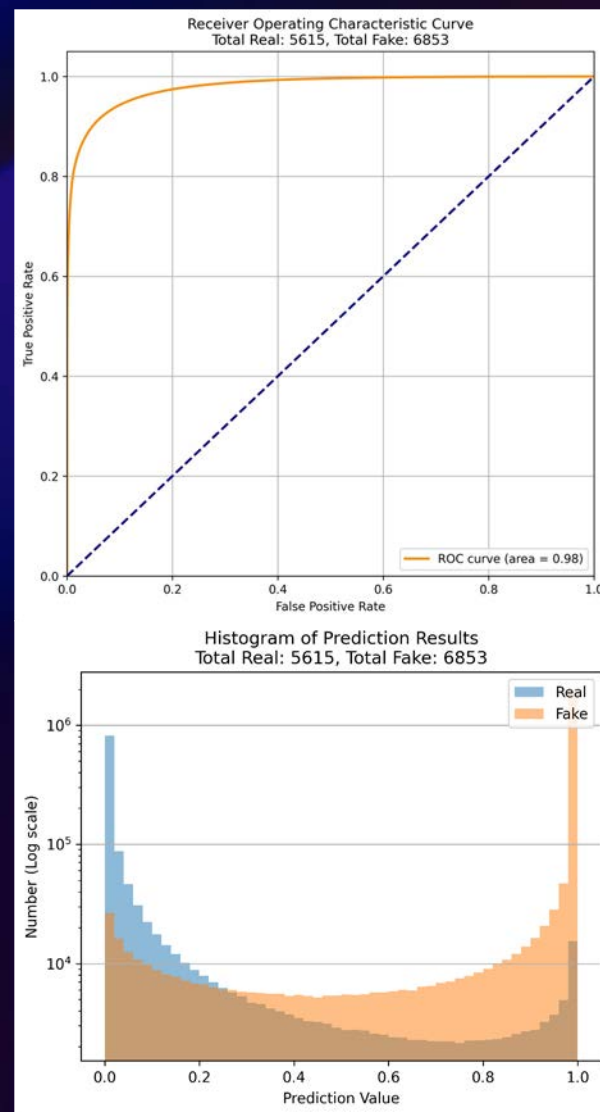
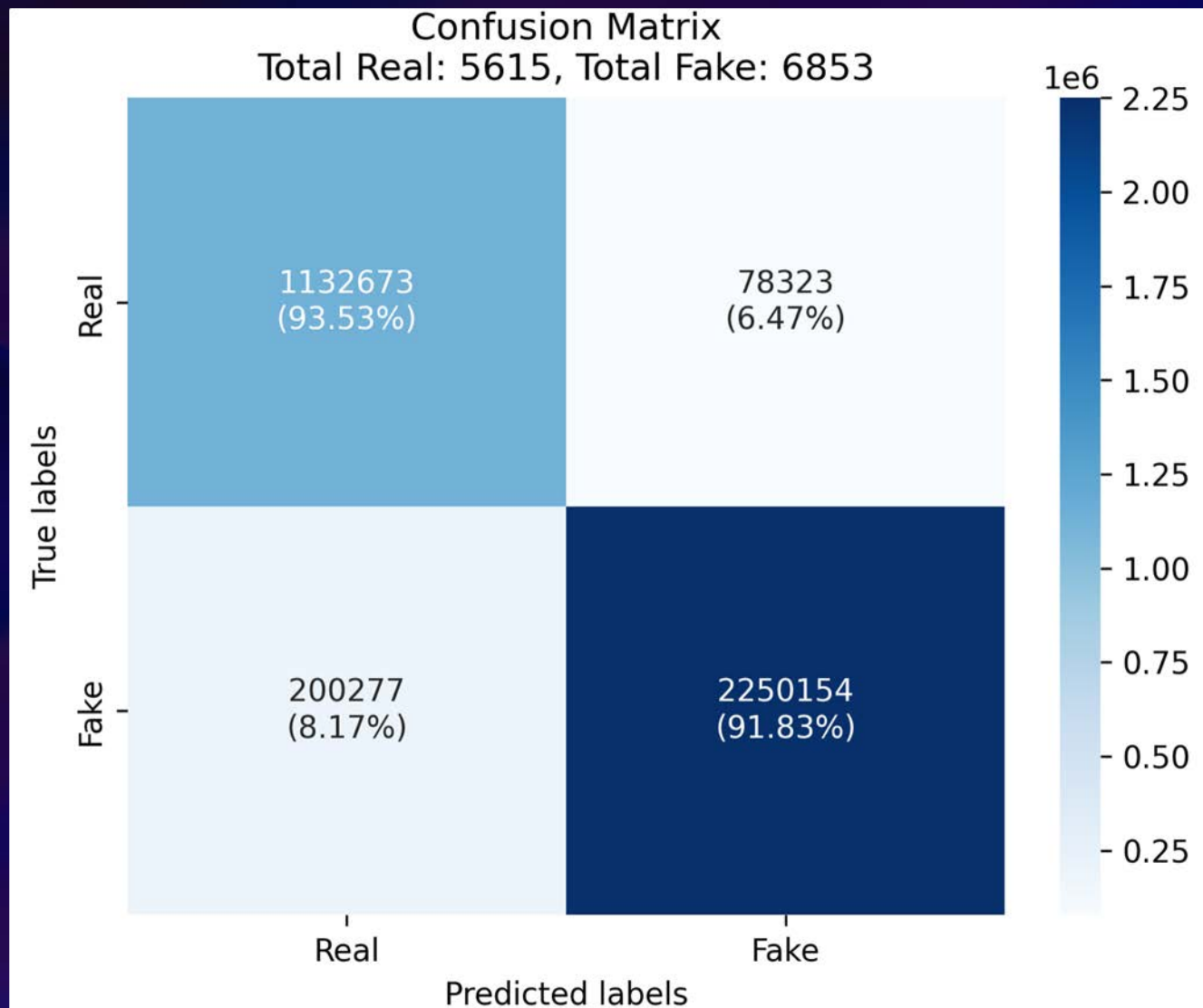
BASE



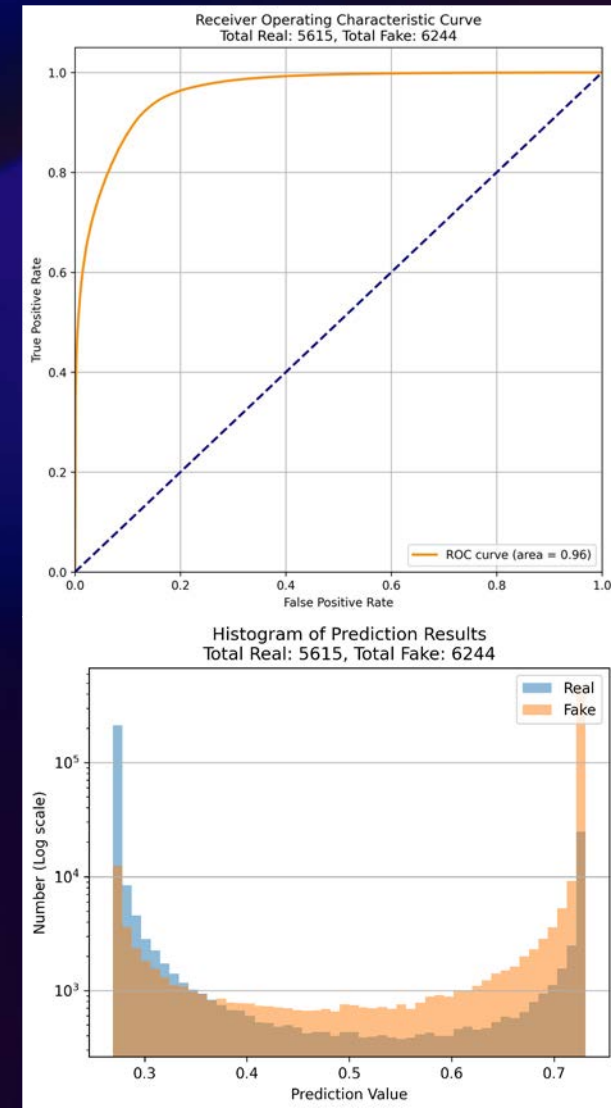
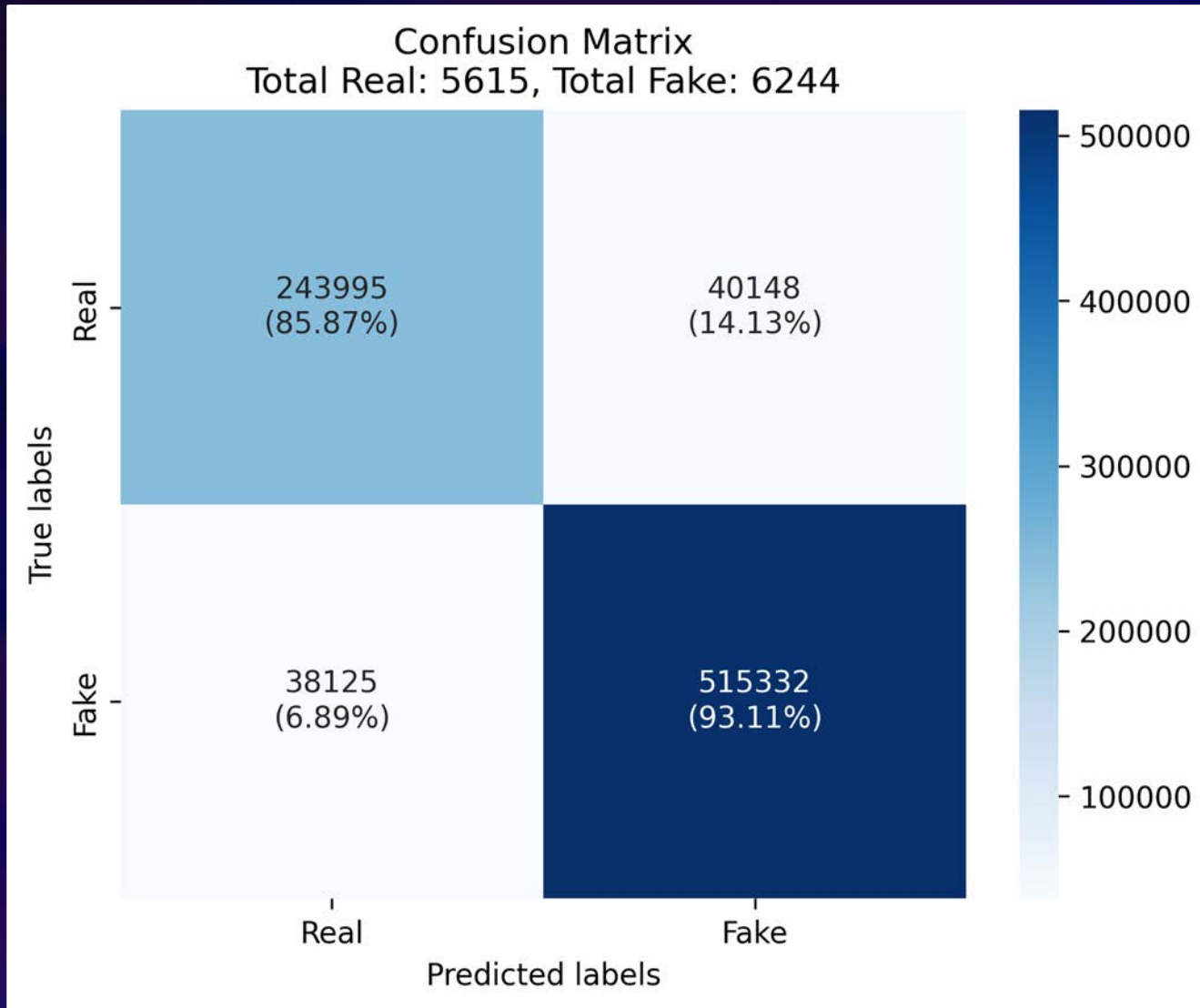
AUGMENTED



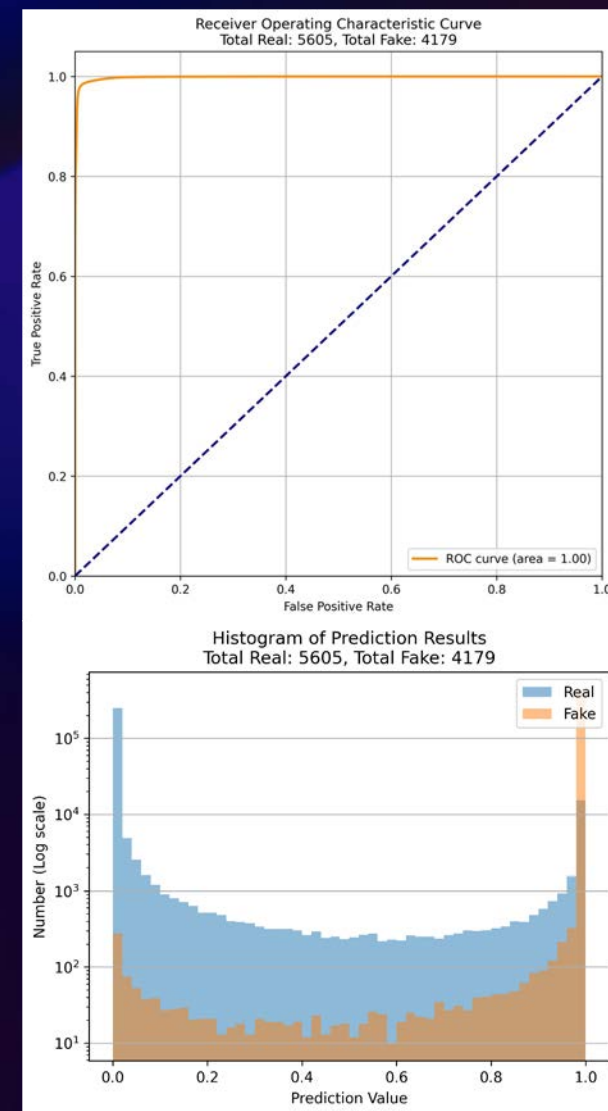
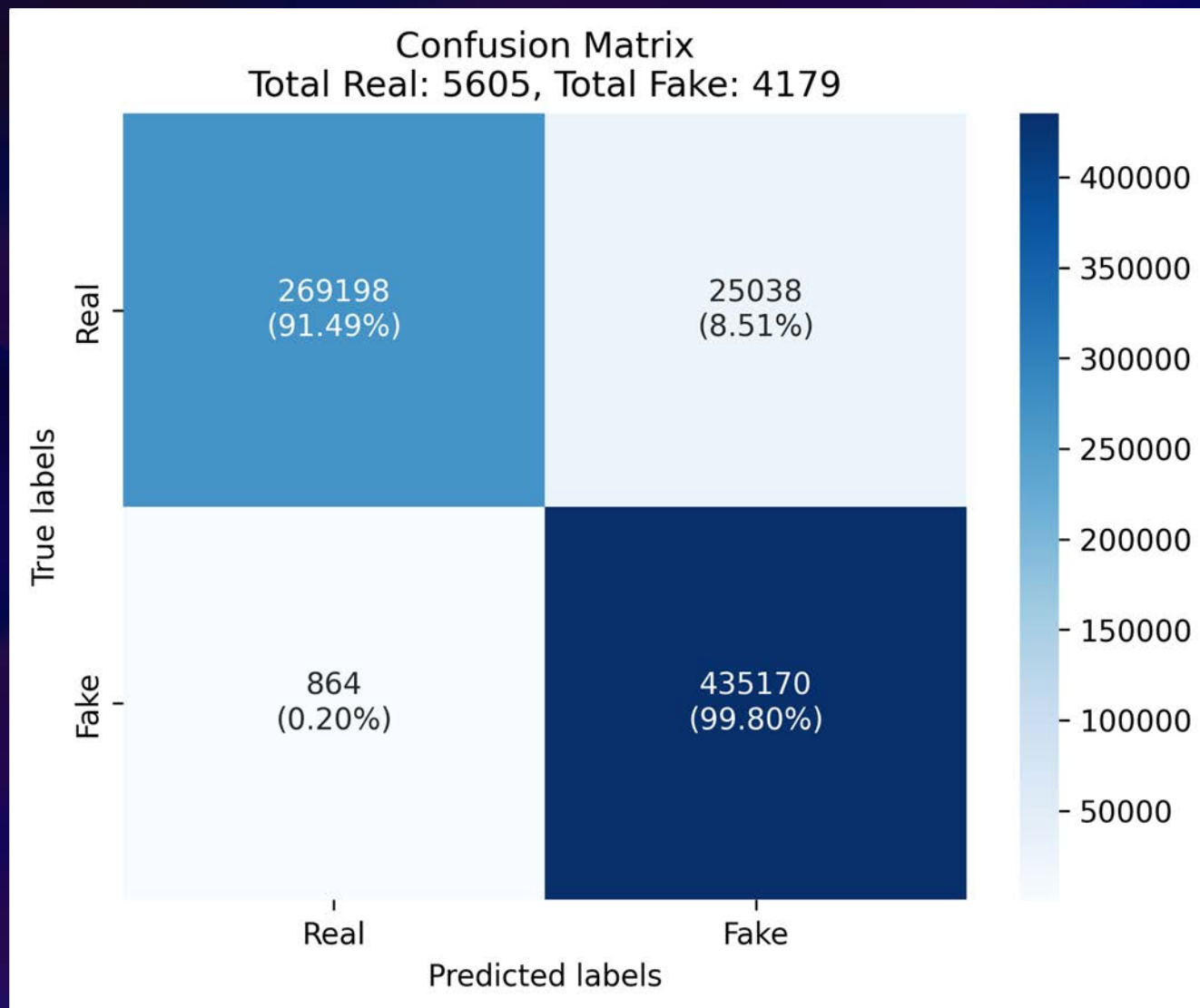
Face Detection Results - CEViT



Face Detection Results - RevMViT



Face Detection Results - UniformerV2



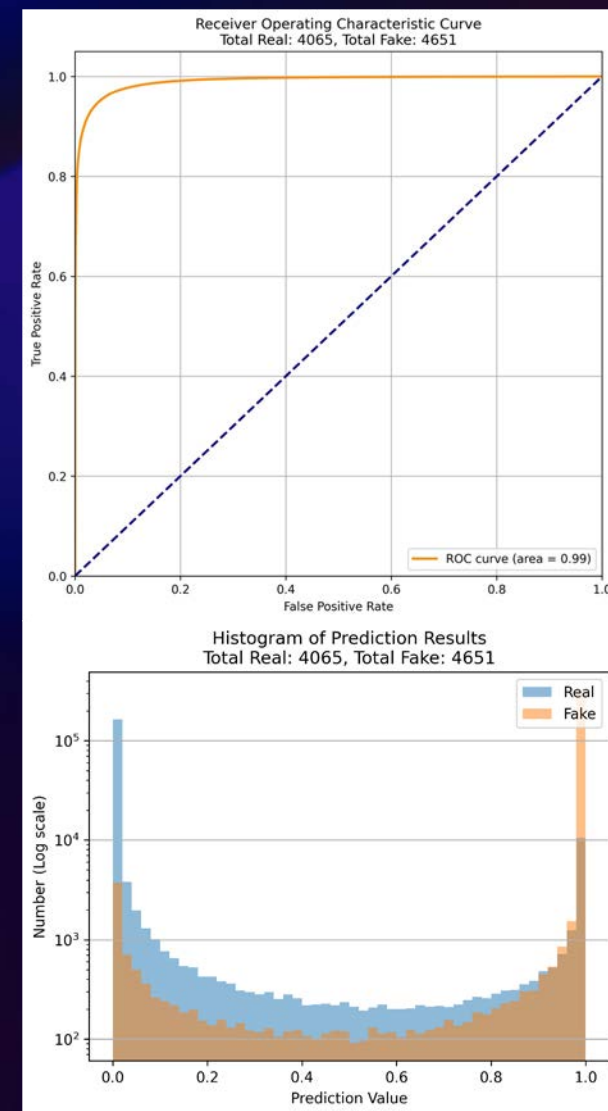
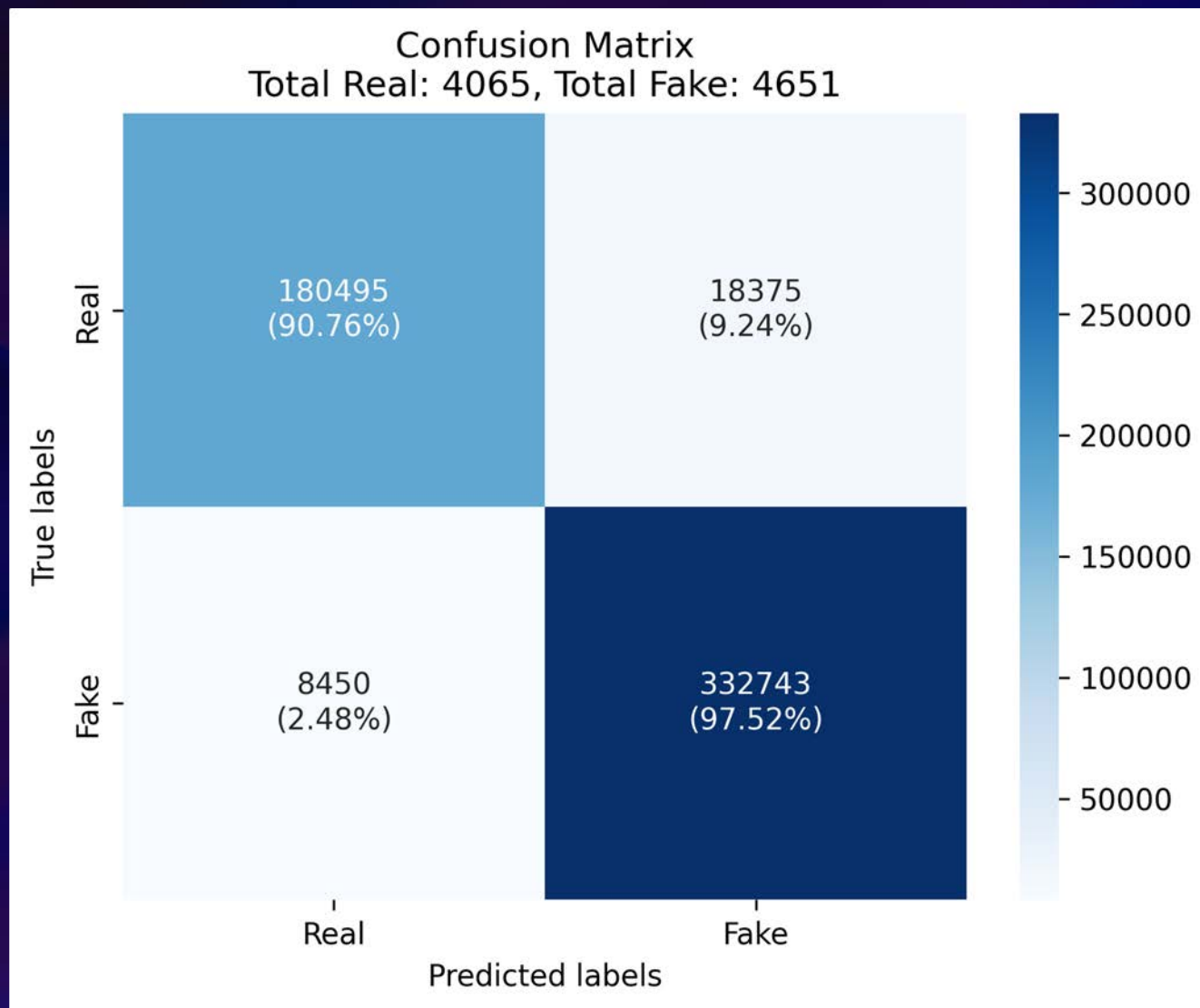
Adding Image/Video Augmentations

```
augmentation_compose = A.Compose([
    A.ImageCompression(always_apply=False, p=0.10, quality_lower=60, quality_upper=100, compression_t
    A.HorizontalFlip(p=0.5),
    A.Affine(p=0.10, scale=(0.95, 1.05), translate_percent=(-0.03, 0.03), rotate=(-3, 3), shear=(-3,
    A.ShiftScaleRotate(shift_limit=0.10, scale_limit=0.2, rotate_limit=5, border_mode=cv2.BORDER_REPL
    A.CLAHE(always_apply=False, p=0.10, clip_limit=(1, 4), tile_grid_size=(8, 8)),
    A.HueSaturationValue(always_apply=False, p=0.20, hue_shift_limit=(-20, 20), sat_shift_limit=(-30,
    A.RandomBrightnessContrast(always_apply=False, p=0.20, brightness_limit=(-0.2, 0.2), contrast_lim
    A.ToGray(p=0.05),
    A.GaussNoise(always_apply=False, p=0.10, var_limit=(10.0, 50.0)),
    A.MotionBlur(always_apply=False, p=0.10, blur_limit=(3, 7)),
    A.ImageCompression(always_apply=False, p=0.20, quality_lower=60, quality_upper=100, compression_t

],
    additional_targets=additional_targets_dict
)
```



Face Detection Results - UniformerV2 w/ Augmentations



Face Detection Results - Effect of Resolution

CEViT

RevMViT

Uniformerv2

Uniformerv2
Aug



Conclusions

Textual Deepfake Detection Challenges: Pioneering advancements in techniques and detection algorithms are paramount to surmount the substantial challenges posed by high-accuracy deepfake text detection. This uncharted territory beckons for revolutionary tools and methodologies to bolster detection precision.

A Triumph in Voice Authentication: Current voice detection methodologies exhibit stellar performance even amidst intricate augmentations. There's no pressing need for reinvention; instead, fortifying our existing models with richer and more diverse data will propel their efficacy to unparalleled heights.

Facial Deepfake Recognition: Not only does our facial detection stand robust in its primary implementation, but it also showcases admirable resilience against various augmentations. An intriguing discovery across models underscores the potential power of an ensemble strategy as the key to fortifying deepfake countermeasures.

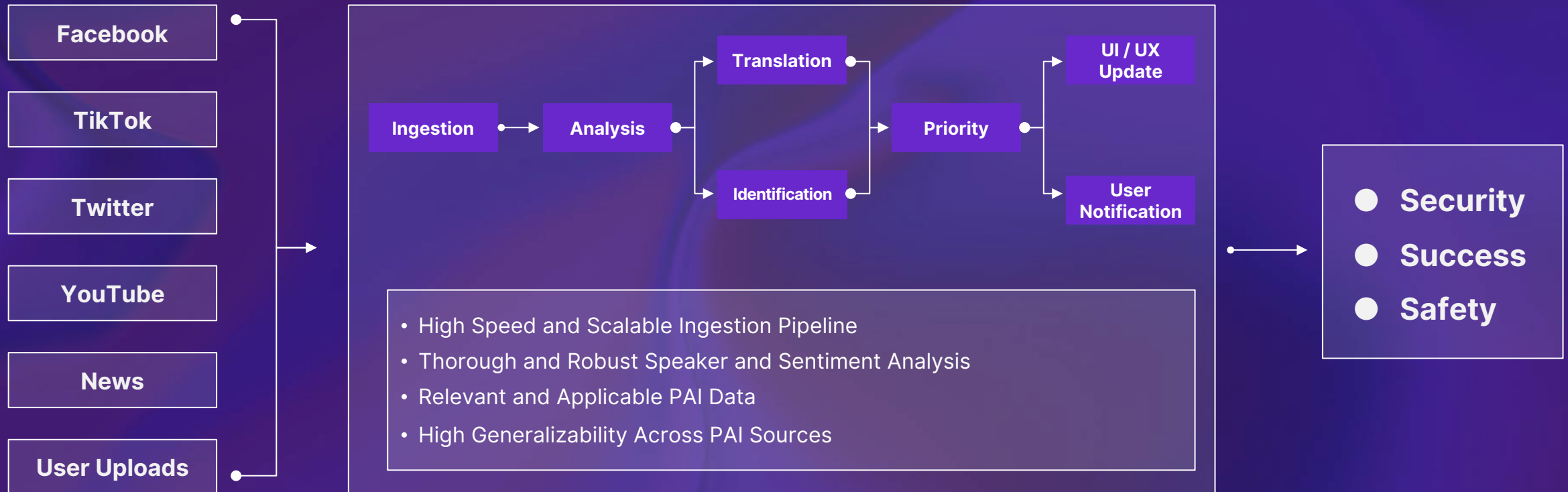
Catalyzing Enhanced Detection: To achieve apex detection accuracies spanning all modalities, it's imperative to craft and assimilate datasets enriched with data derived from avant-garde generator types. These datasets, when seamlessly integrated into our training conduits, promise to redefine the gold standard of deepfake detection.

Pioneering Future Research Frontiers: Venturing deeper into the realms of detector architectures, avant-garde training modalities, and sophisticated feature extraction layers holds the promise to elevate detection accuracy, especially in nuanced scenarios like those presented by ultra-low resolutions.

Big Picture

**Empowering DoD and IC Media
Intelligence and AI Forensics at Scale**

We Harness PAI sources for Robust Data Intake and Analysis at Scale, Processing Millions of Videos Per Day at a Cost of \$0.005 / Min



Synthetic Media Detection Process: Face and Voice Acquisition

DeepMedia's video ingestion process prepares ingested content for further analysis and processing. DeepMedia's proprietary pre-processing technique aims to optimize the input media for various AI models used in the product. For video content, it is split into individual frames, and the audio is extracted from the video stream. If the input is audio, it is directly processed. Our sophisticated pre-processing detects and extracts faces and voices from the ingested content, preparing them for processing through our state of the art Deepfake Detectors.

Step 1

DeepMedia's Ingestion pipeline stores the ingested content.

Step 2

Video Content is split into frames, and audio extracted from the video

Step 3

Faces and voices are detected and stored from the inputted content.

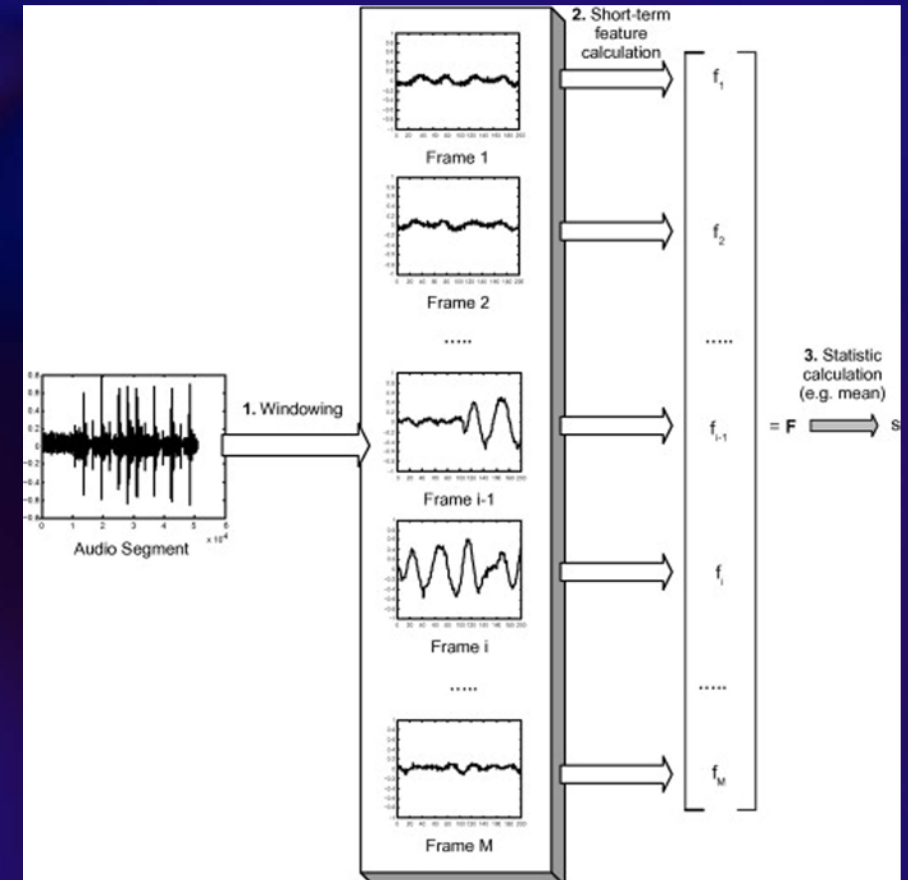
Step 4

These faces and voices are prepared for processing through our deepfake detectors.

AI Media Detection Process: Face and Voice Acquisition

For Instance, consider the Hillary Clinton Deepfake Outlined Previously:

The DeepID System first hones in on any faces of interest, and splits the video into frames. And audio is collected through a process known as feature extraction, illustrated below.



AI Media Detection Process: Transcription and Translation

To generate transcripts for the ingested content and identify each individual, we utilize a multi-modal deep learning approach. The audio from the input media is transcribed using a state-of-the-art Automatic Speech Recognition (ASR) model, capturing the spoken words. Simultaneously, the face detection and recognition models identify and track individuals in the video frames. The detected faces are associated with the transcribed speech to assign spoken words to specific individuals accurately.

Furthermore, the optional translation to English is performed by employing a Transformer-based Machine Translation model, converting the transcribed text from the original language to English. This multi-modal process ensures comprehensive and contextually accurate transcripts that attribute spoken words to the correct individuals while providing the flexibility to translate the content to English if desired.

Step 1

Audio is transcribed using a state-of-the-art Automatic Speech Recognition Model.

Step 2

Face Detection models identify and track individuals in the video.

Step 3

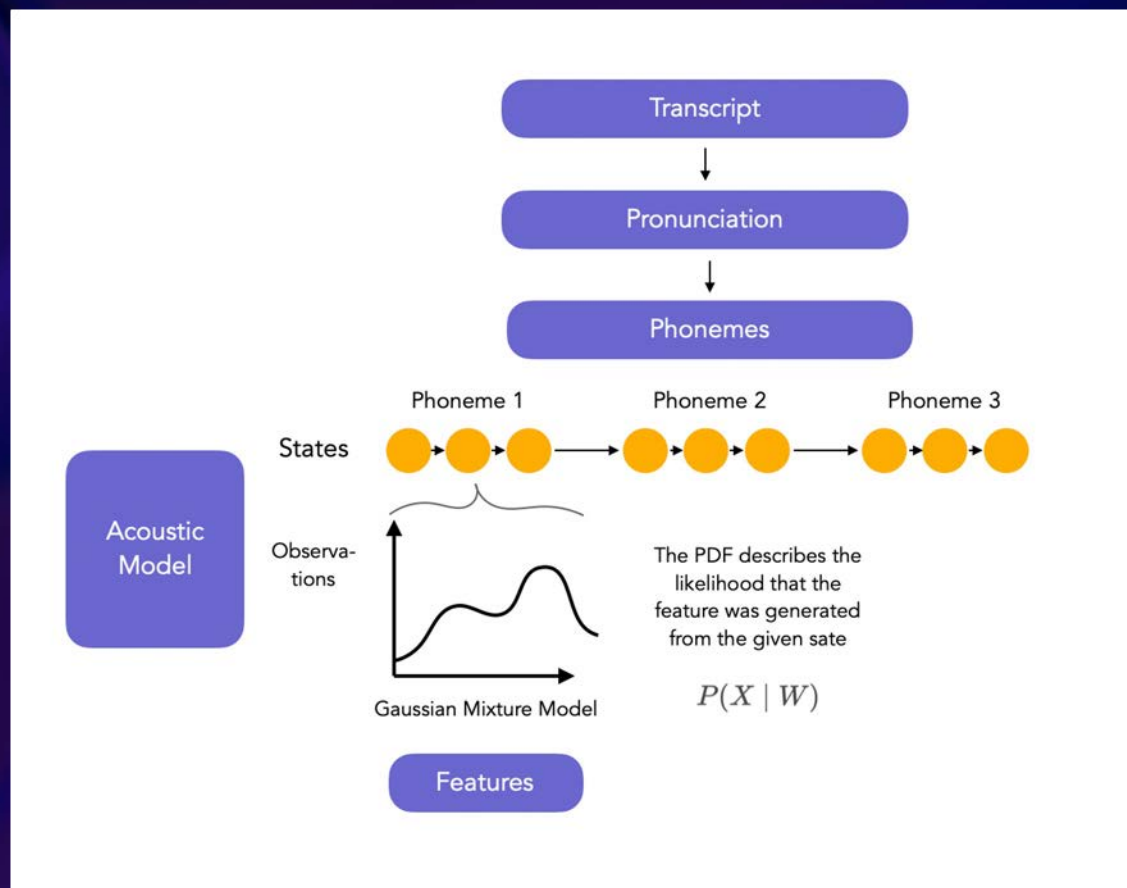
Detected Faces are automatically associated with the transcribed speech.

Step 4

When desired, translation to English is performed by a powerful Transformer based Machine Translation Model, converting the transcribed text to English.

AI Media Detection Process: Transcription and Translation

The video is then passed through Deep Media's state-of-the-art Automatic Speech Recognition Model, which accurately detects and transcribes the provided audio, using a powerful phoneme analysis. When desired, this transcription is passed through a transformer model and into any language of interest.



"You know, people might be surprised to hear me saying this, but I actually like Ron DeSantis a lot. Yeah, I know. I'd say he's just the kind of guy this country needs, and I really mean that. If Ron DeSantis got installed as president, I'd be fine with that."

- Hillary Clinton in
AI-Generated Video

AI Media Detection Process: Face and Voice Analysis

Our voice and face deepfake detectors employ advanced deep learning techniques to scrutinize the extracted faces and voices and discern any signs of synthetic manipulation. For face deepfake detection, the model scrutinizes facial landmarks, micro-expressions, and overall facial consistency to identify potential anomalies. Additionally, the model analyzes artifacts, inconsistencies, and unnatural movements to distinguish real faces from manipulated ones. For voice deepfake detection, the model examines spectrograms, pitch patterns, and phonetic characteristics, scrutinizing any inconsistencies or patterns characteristic of synthetic speech generated by voice synthesis models. Moreover, both detectors utilize large-scale datasets of genuine and manipulated faces and voices for robust training, enabling them to adapt and recognize emerging deepfake techniques effectively.

Step 1

Using a robust and extensive Deepfake and real dataset, our detector analyzes each face and voice for signs of manipulation

Step 2

Faces are analyzed for micro-expressions and facial consistency

Step 3

DeepID identifies artifacts, inconsistencies and unnatural movements to distinguish real and manipulated faces

Step 4

Voices are examined for pitch patterns and phonetic characteristics.

AI Media Detection Process: Face and Voice Analysis

The extracted frames are analyzed by our world class detectors, and a real time video analysis is performed. Below you can observe our Sherlock detector analyzing this input, as well as our Facial Tracking system, which ensures our detectors have the highest accuracies possible.



AI Media Creation Process: Multi-Modal Ensemble Detection

After analyzing the extracted faces and voices using our voice and face deepfake detectors, we merge the results to form a comprehensive assessment of the content's authenticity. By comparing the identified faces with our people identification database, we determine the presence of known individuals and potential impersonations. Our proprietary fusion algorithm combines the outputs from multiple detectors, considering their respective confidences, to arrive at a reliable final verdict on whether the content is likely manipulated or genuine. The synthesized findings are presented in a user-friendly report, outlining the level of confidence in the detection results

Step 1

We merge the results from our face and voice analysis to come to a conclusion about the authenticity of the content.

Step 2

We combine the output of multiple detectors to ensure an accurate conclusion about the authenticity of the content.

Step 3

We arrive at a final conclusion, and prepare a user friendly report outlining our confidence in the detection results.

AI Media Creation Process: Multi-Modal Ensemble Detection

Implementing an extensively tested set of Heuristics to determine the authenticity of the inputted content, a conclusion is drawn by our detectors. Our system provides analysis of the authenticity of both Audio and Video

Result Report

 This video contains some evidence of synthetic manipulation

This display is intended only for educational purpose.

Face: 1



Start time: 00:00:11
End time: 00:00:25

Sherlock: 53

Clouseau: 00

Prediction Score: 00

Face: 2

Start time:
End time:

Sherlock:

Clouseau:

Prediction Score:

Face: 3

Start time:
End time:

Sherlock:

Clouseau:

Prediction Score:

Face: 4

Start time:
End time:

Sherlock:

Clouseau:

Prediction Score:

Original



HILLARY CLINTON ENDORSES DESANTIS  

Tracking



HILLARY CLINTON ENDORSES DESANTIS  

DeepID Detector

Result Report



This video contains some evidence of synthetic manipulation

This display is intended only for educational purpose.

AI Media Detection Process: Translation, Sentiment Analysis, Identity Detection, and Prioritization

Prioritizing and delivering information to the user is a crucial aspect of our product's user experience. We follow a streamlined approach to ensure efficient and clear communication of the analysis results. First, the deepfake detection and people identification results are prioritized based on their confidence levels, **highlighting** the most relevant and critical findings. The user is presented with a summary of the key insights, such as whether deepfake manipulation is detected and if any identified individuals warrant special attention.

Step 1

Deepfake detection results are prioritized based upon their confidence level

Step 2

The user is presented with a summary of the key insights uncovered by the detection process

Step 3

A detailed report is provided, offering a comprehensive breakdown of the analysis, including timestamps of potential manipulation occurrences, transcribed text, and identified individuals.

The Lookup Database Utilizes the ARWEAVE Blockchain for Enhanced Security and Reliability

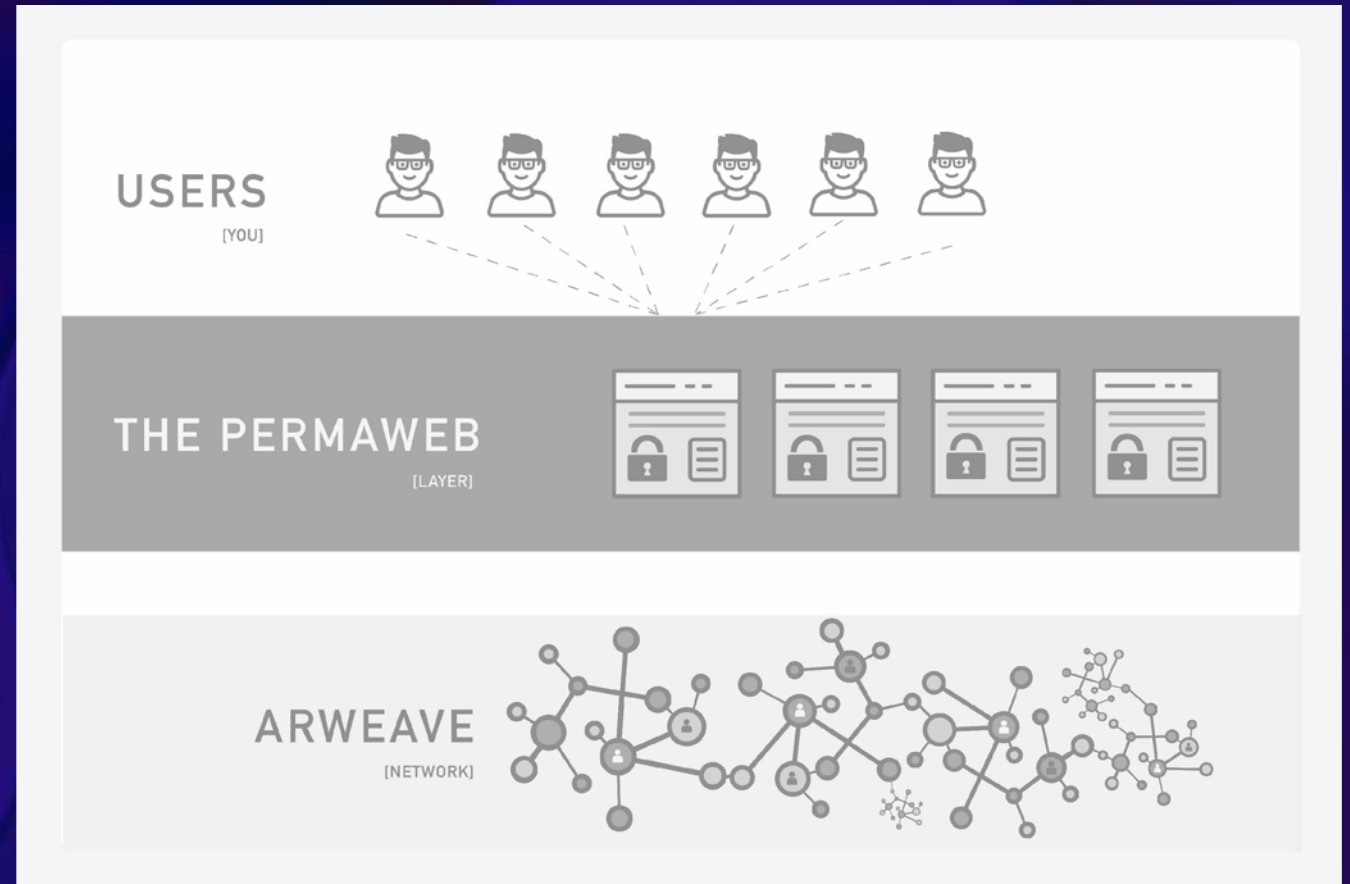
Secure blockchain integration enhances security and reliability for the lookup database

Immutable data records ensures tamper-proof and auditable data in the lookup database

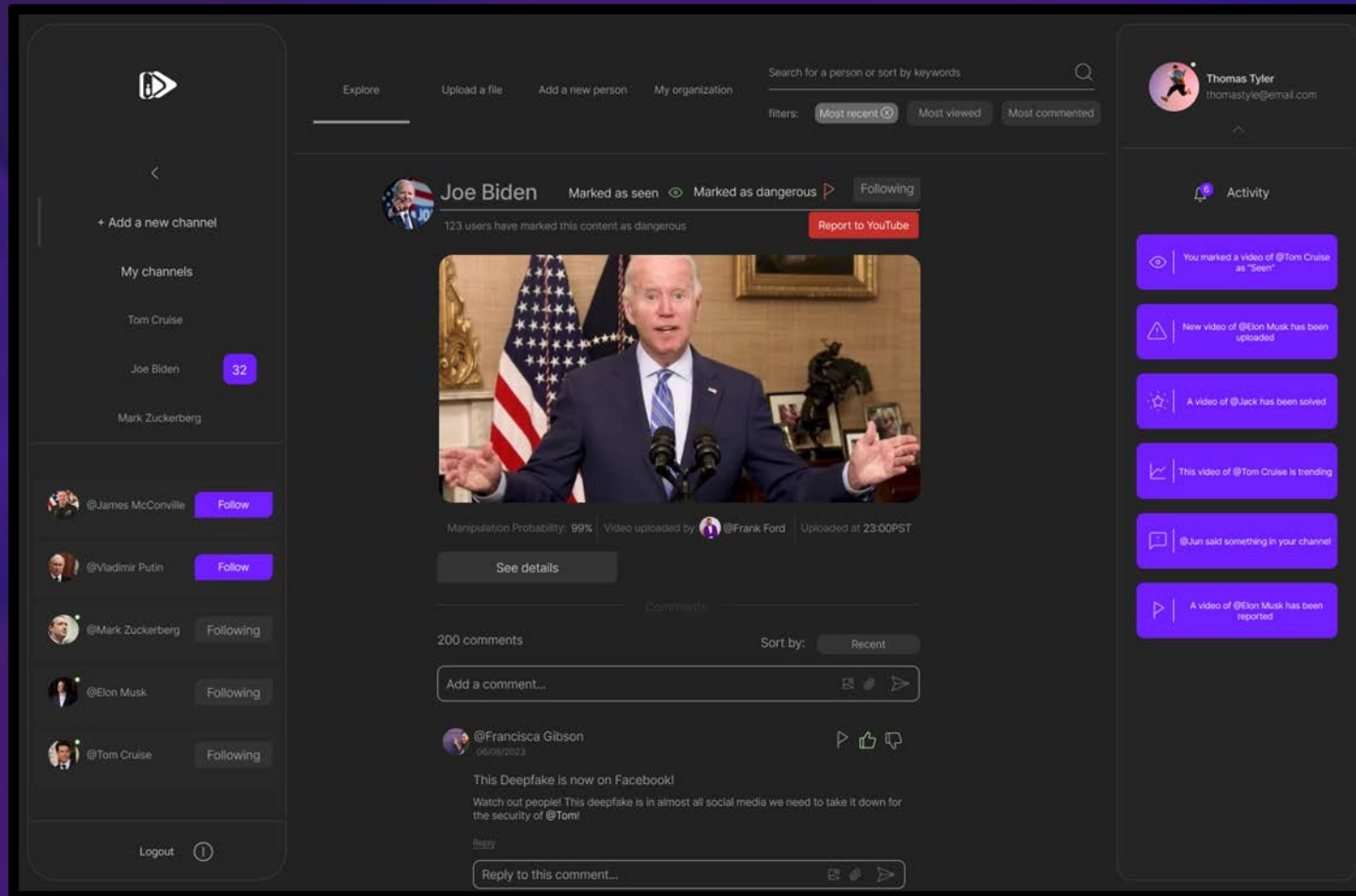
Decentralized architecture increases resilience and availability

Trusted data integrity preventing unauthorized modifications in the lookup database

Reliable and auditable source of information for the lookup database

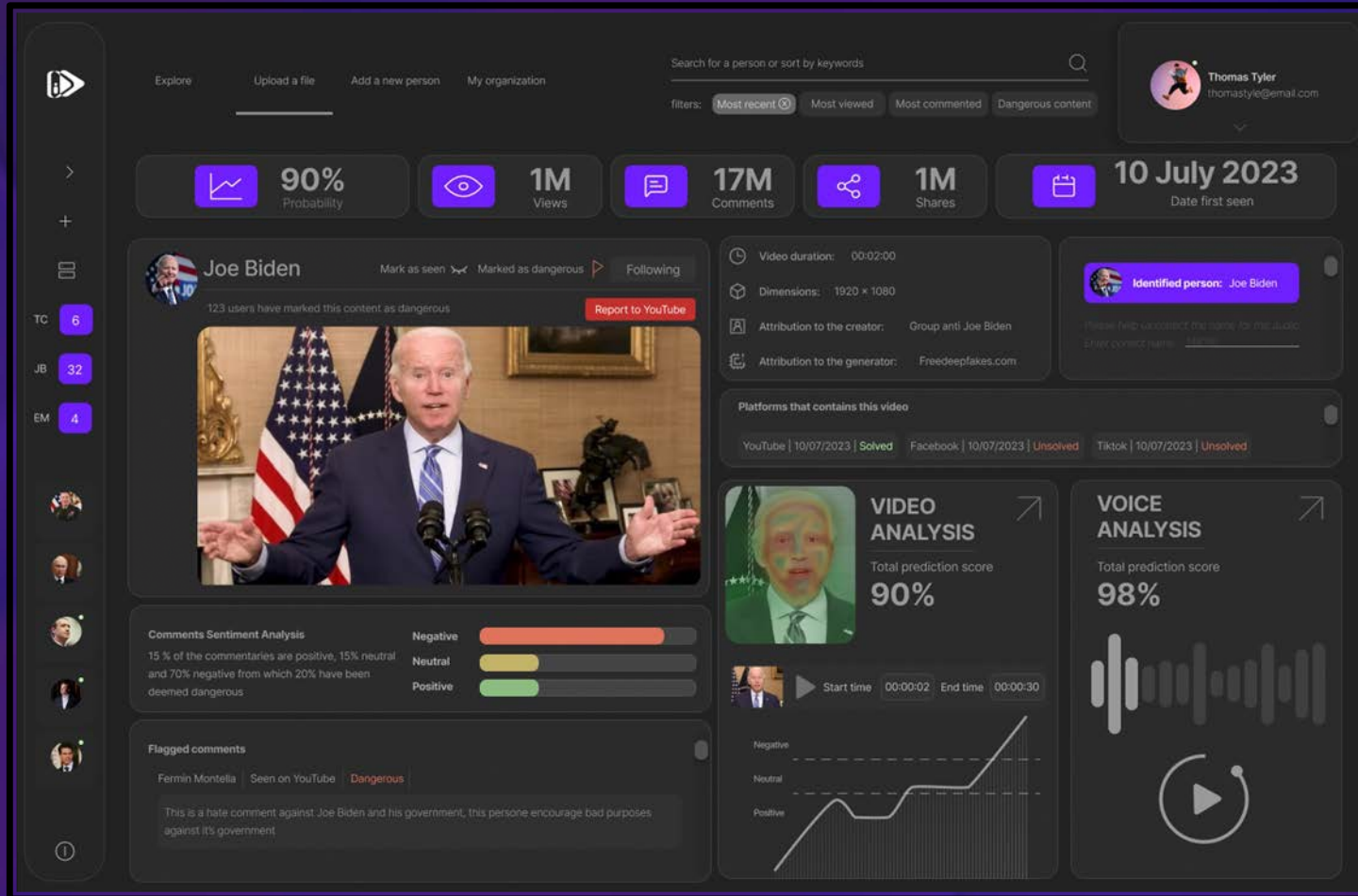


Data Organization, Fast Search, and Workflow Tooling Makes Checking Content Easier and Faster



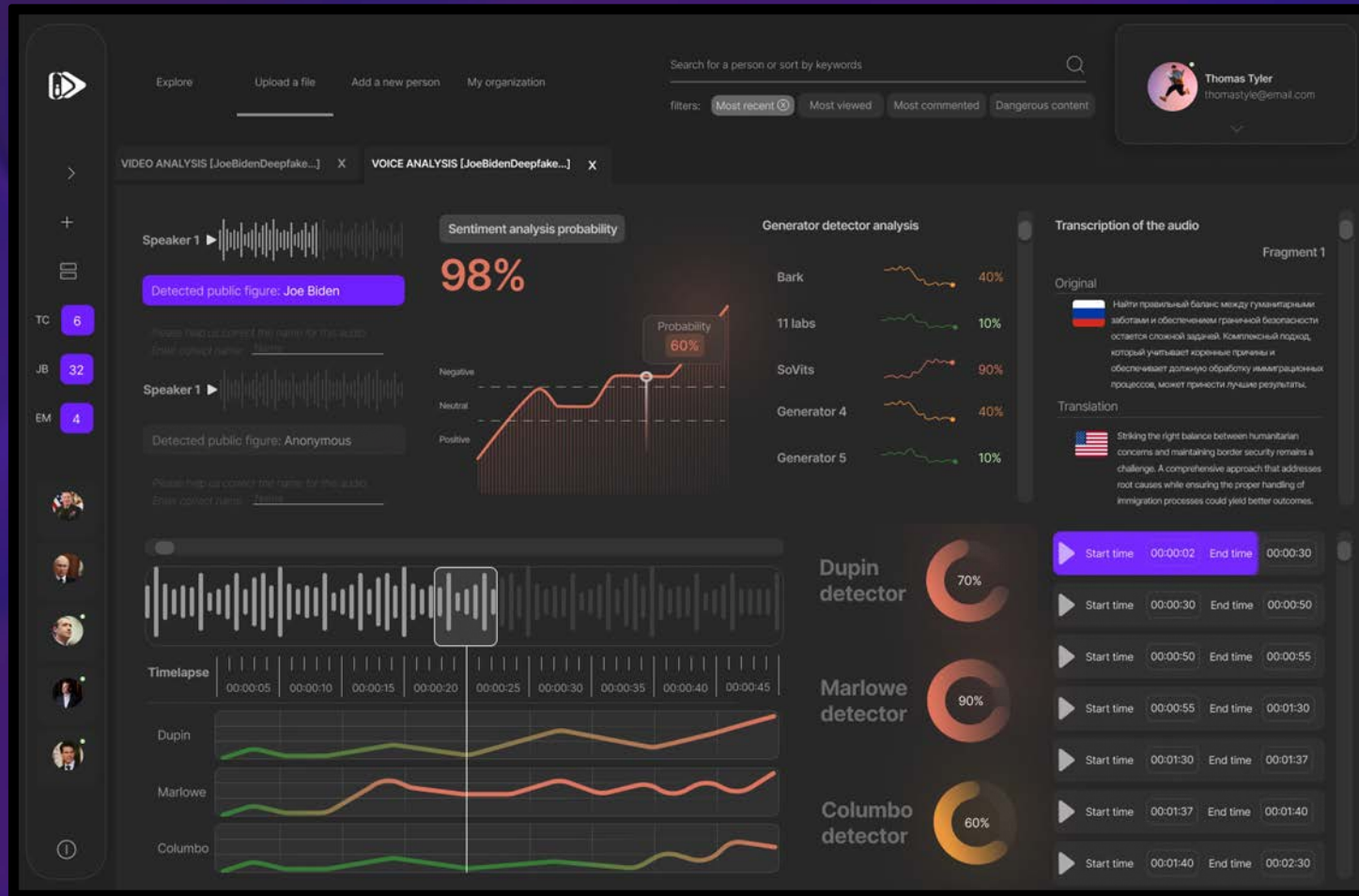
- Ordered By Threat Level
- Fast and Accurate Search
- Easy-to-use Data Feed
- Organizes All Modalities

High-End Analytical Features Will Give DoD/IC Supreme Confidence In AI-Security Decisions



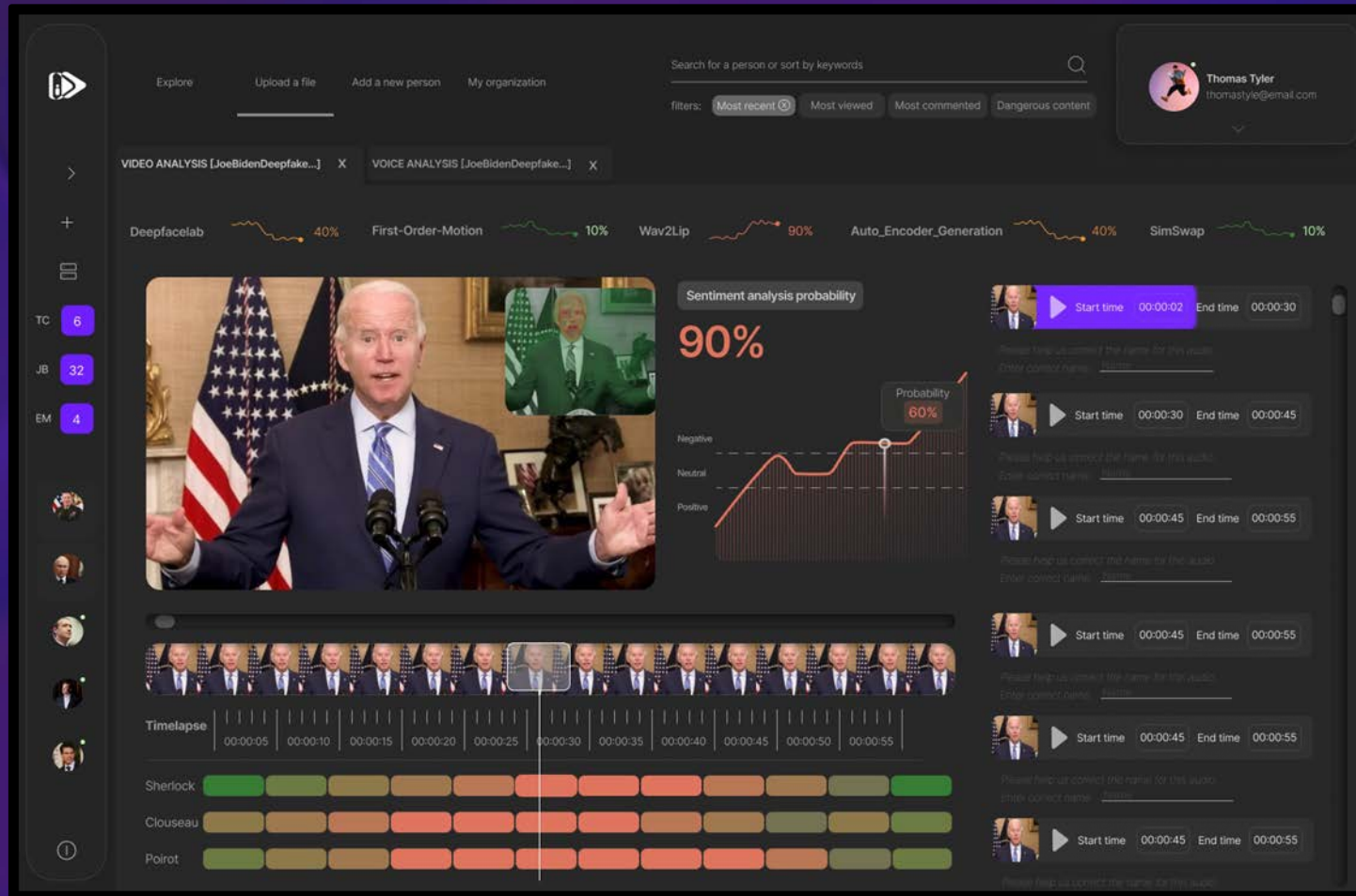
- Overview Stats
- Seamless UI/UX
- Sentiment Analysis
- Meta Data Analysis

Voice Detection, Model Attribution Transcription and Translation



- Voice Detection
- Multiple Detectors
- Text Transcripts
- English Translation

High-Accuracy Facial Manipulation Analysis with Time-Code Organization



- Face Detection
- Generator Attribution
- Trajectory Analysis
- Frame-by-Frame

Face and Voice Identity Mapping for High-Priority Individuals and Alerts

Explore Upload a file Add a new person My organization

Search for a person or sort by keywords

Filters: Most recent Most viewed Most commented Dangerous content

Thomas Tyler
thomastyle@gmail.com

Activity

- You marked a video of @Tom Cruise as "Seen"
- New video of @Ekin Muek has been uploaded
- A video of @Jack has been solved
- This video of @Tom Cruise is trending
- @Jun said something in your channel
- A video of @Ekin Muek has been reported

Add a new person

Complete the steps below

To automate the detection of the new person you are adding, it is necessary to provide a sample of at least three pictures of the individual (one person in the frame) and a voice sample (one person talking)

My channels

TC 6
JB 32
EM 4

1. Full name of the person you want to add.

Full name

2. Who is this person?

Write a short description...

3. Date of birth

DD/MM/YYYY

☐ I don't know this is

Approximate age

Select

4. Upload pictures

Only one person in the frame

5. Upload a voice sample

Only one person speaking

1. Full name of the person you want to add.

Tom Cruise

2. Who is this person?

Actor and celebrity

3. Date of birth

DD/MM/YYYY

☒ I don't know this is

Approximate age

45-50

4. Upload pictures

Only one person in the frame

5. Upload a voice sample

Only one person speaking

Add new person

- Easy Update
- User Specific
- Instance Results
- Updates Priority

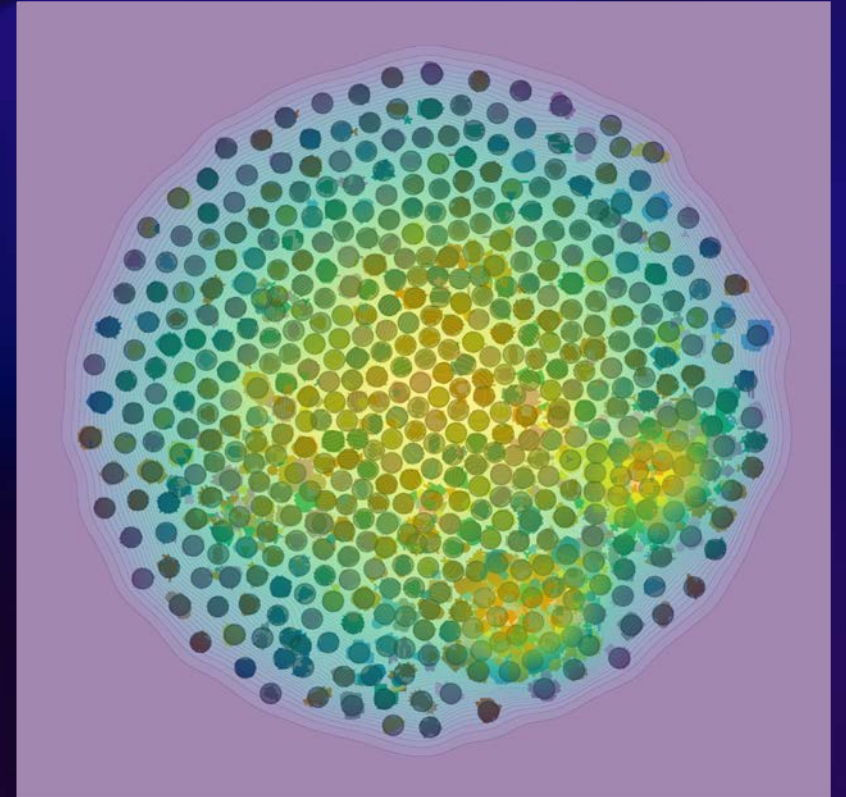
What makes DeepID Different?

Superior Detection: DeepMedia's DeepID outperforms competitors with its robust dataset from 20+ countries and 15+ languages, ensuring unmatched detection in all scenarios, particularly high-risk security situations.

Efficiency Meets Scalability: DeepID isn't just scalable—it's also cost-effective, providing a marked contrast to competitors' resource-intensive solutions.

Evolutionary Dataset Excellence: Our dataset is a living entity, ever-evolving and incorporating examples from even the most cutting-edge synthetic manipulation models.

Advanced Interface & Insights: DeepID's intuitive UI and sophisticated visual analytics provide experts with profound insights, enhancing both user experience and security outcomes.

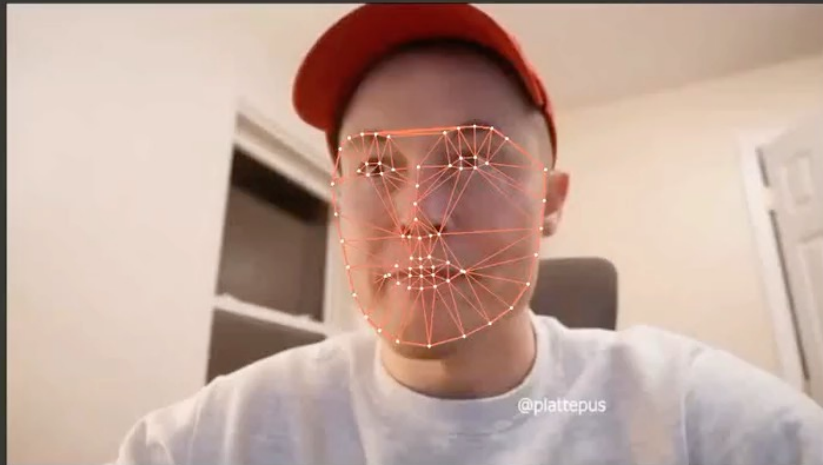


DeepID Detector

Original



Tracking

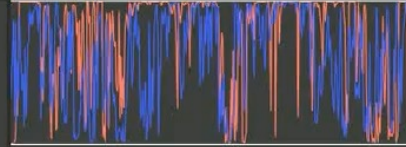


Face: 1



Start time: 00:00:00

End time: 00:00:14



CE-Vit:	E-Vit:	Prediction Score
0.0	0.31	1.0

Face: 2

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Face: 3

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Face: 4

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Result Report

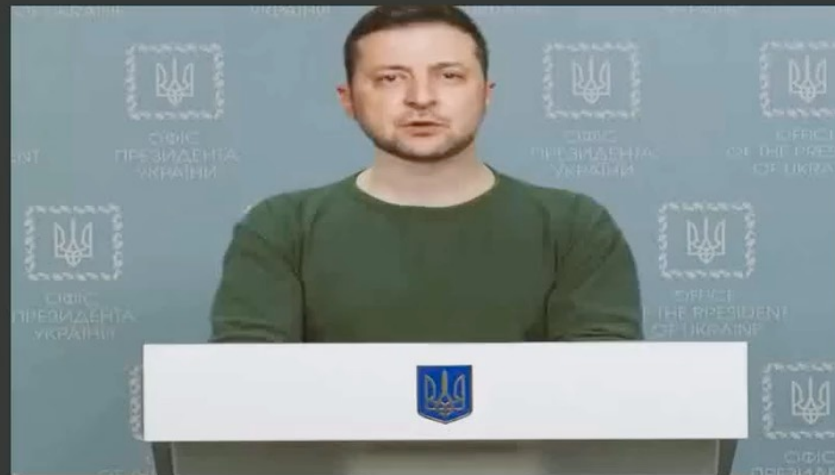


This video contains significant evidence of synthetic manipulation

This display is intended only for educational purpose.

DeepID Detector

Original



Tracking

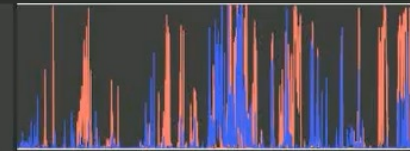


Face: 1



Start time: 00:00:00

End time: 00:01:08



CE-Vit:	E-Vit:	Prediction Score
0.0	0.03	0.91

Face: 2

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Face: 3

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Face: 4

Start time:

End time:

CE-Vit:	E-Vit:	Prediction Score

Result Report



This video contains some evidence of synthetic manipulation

This display is intended only for educational purpose.



Concluding Thoughts

01

Best-in-class: DeepMedia's DeepFake detection provides the best solution to combat AI-Threats

02

Urgent Tech Deployment: Immediate need for our technology to address emerging threats

03

Hardware Integration: Integrating with hardware systems to optimize performance

04

Continuous Enhancements: Advancing technology with refinements to improve functionality



Thank you

www.deepmedia.ai

Rijul Gupta

T. 405.765.3169

rijul@deepmedia.ai

Emma Brown

T. 503.820.8305

emma@deepmedia.ai

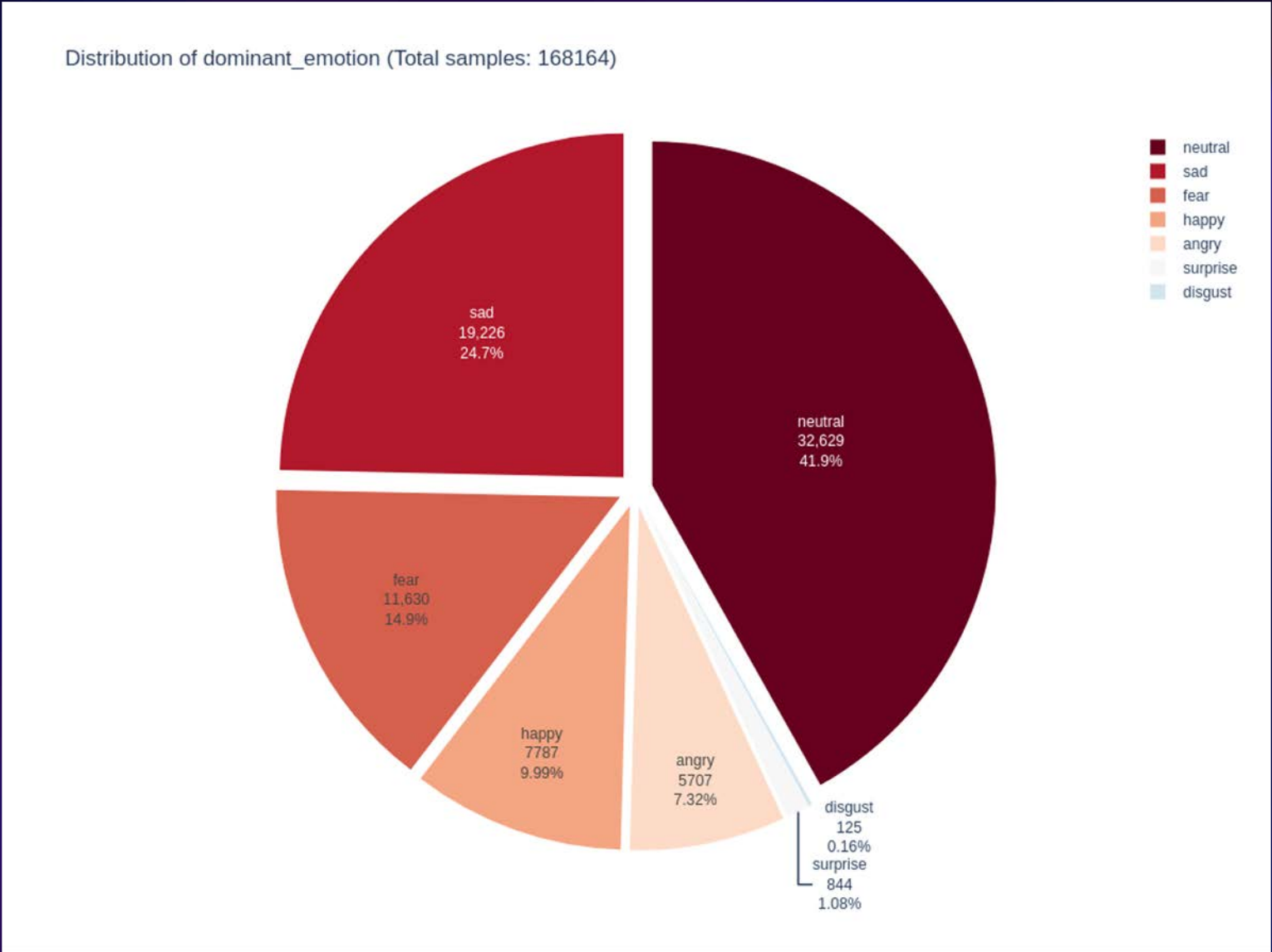
Dan Selli

T. 202.615.9811

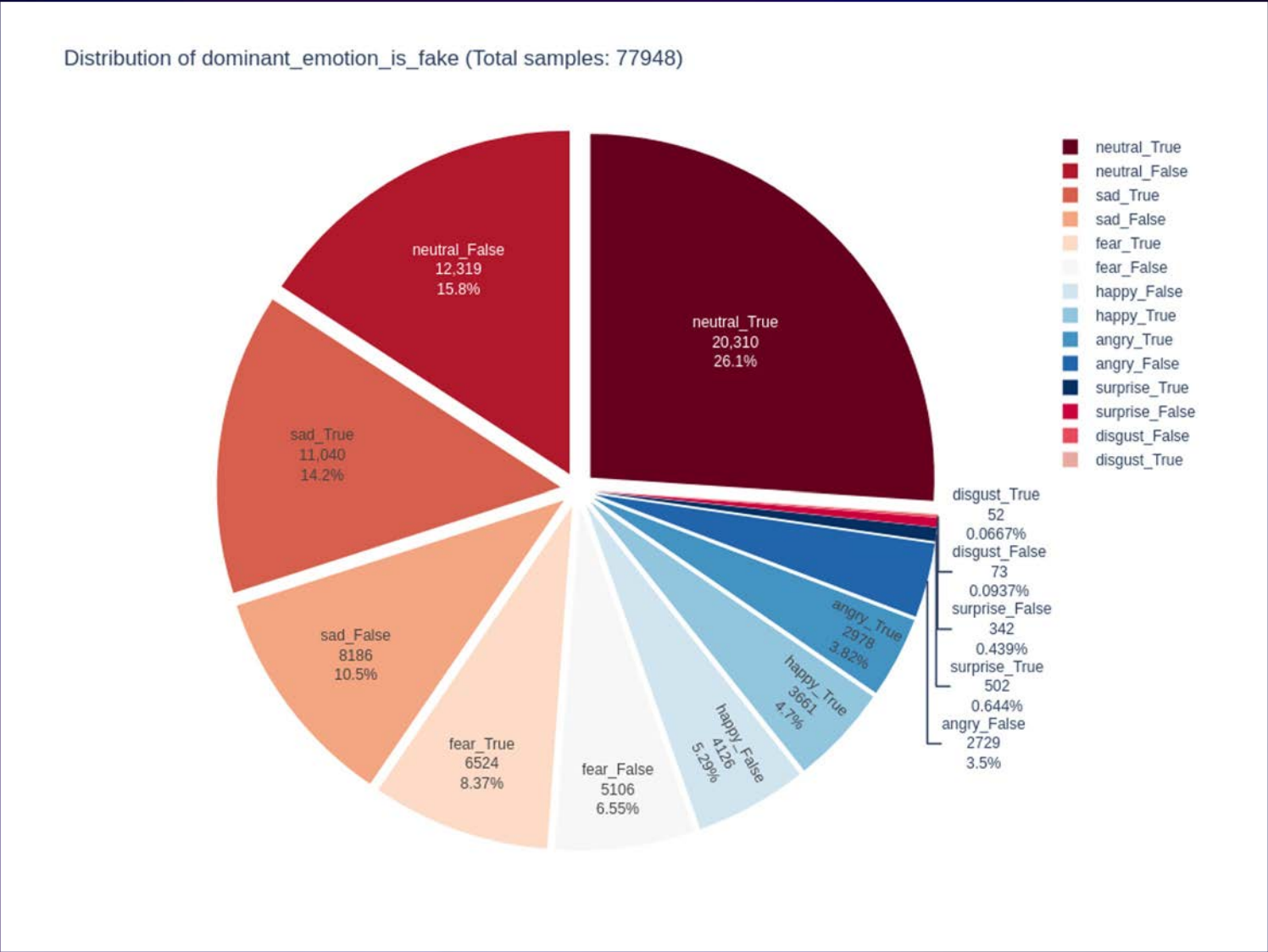
dan@deepmedia.ai

APPENDIX

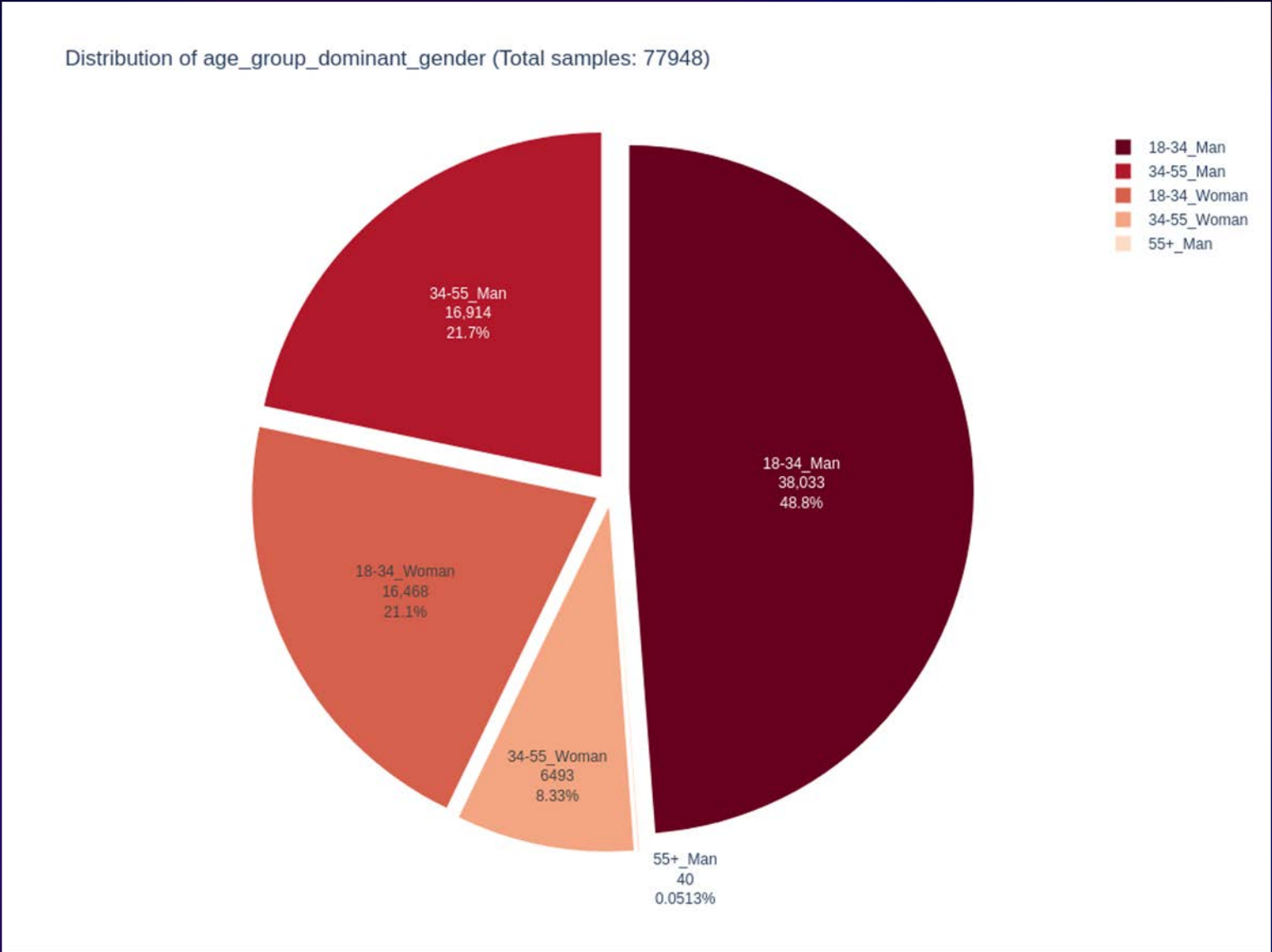
Combatting DeepFakes Requires Dataset Generation: Face Emotion Distribution



Combatting DeepFakes Requires Dataset Generation: Face Emotion X Real/Fake Distribution

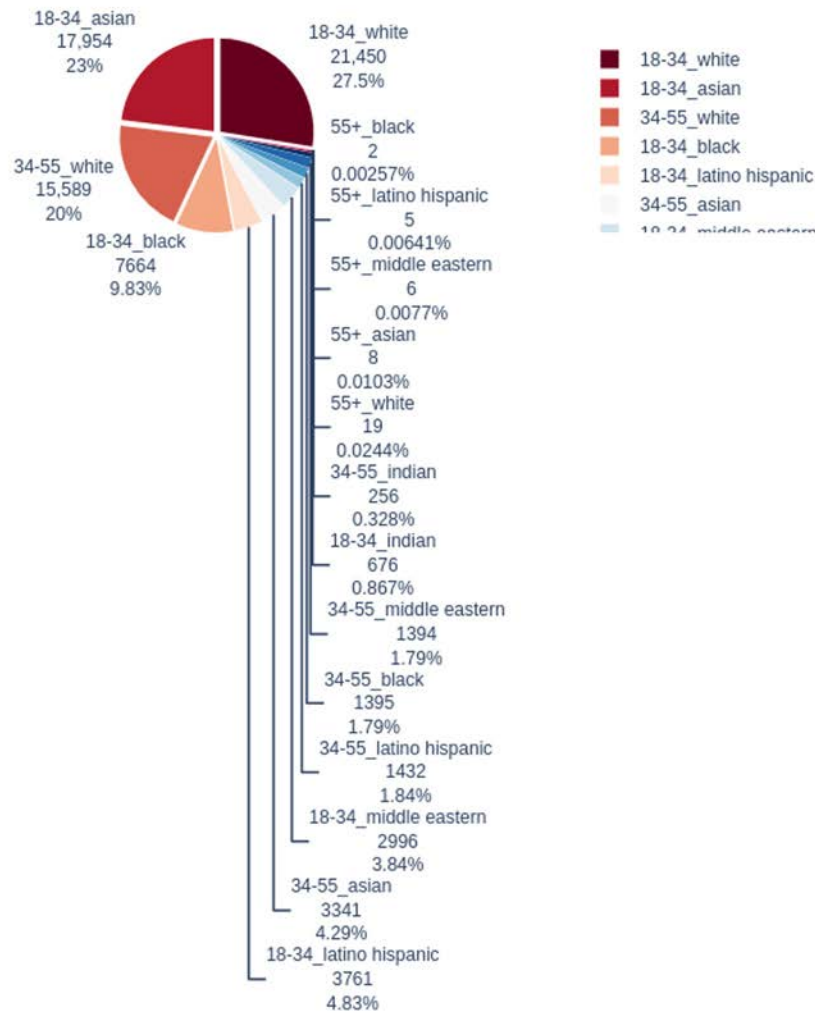


Combatting DeepFakes Requires Dataset Generation: Face Age X Gender Distribution

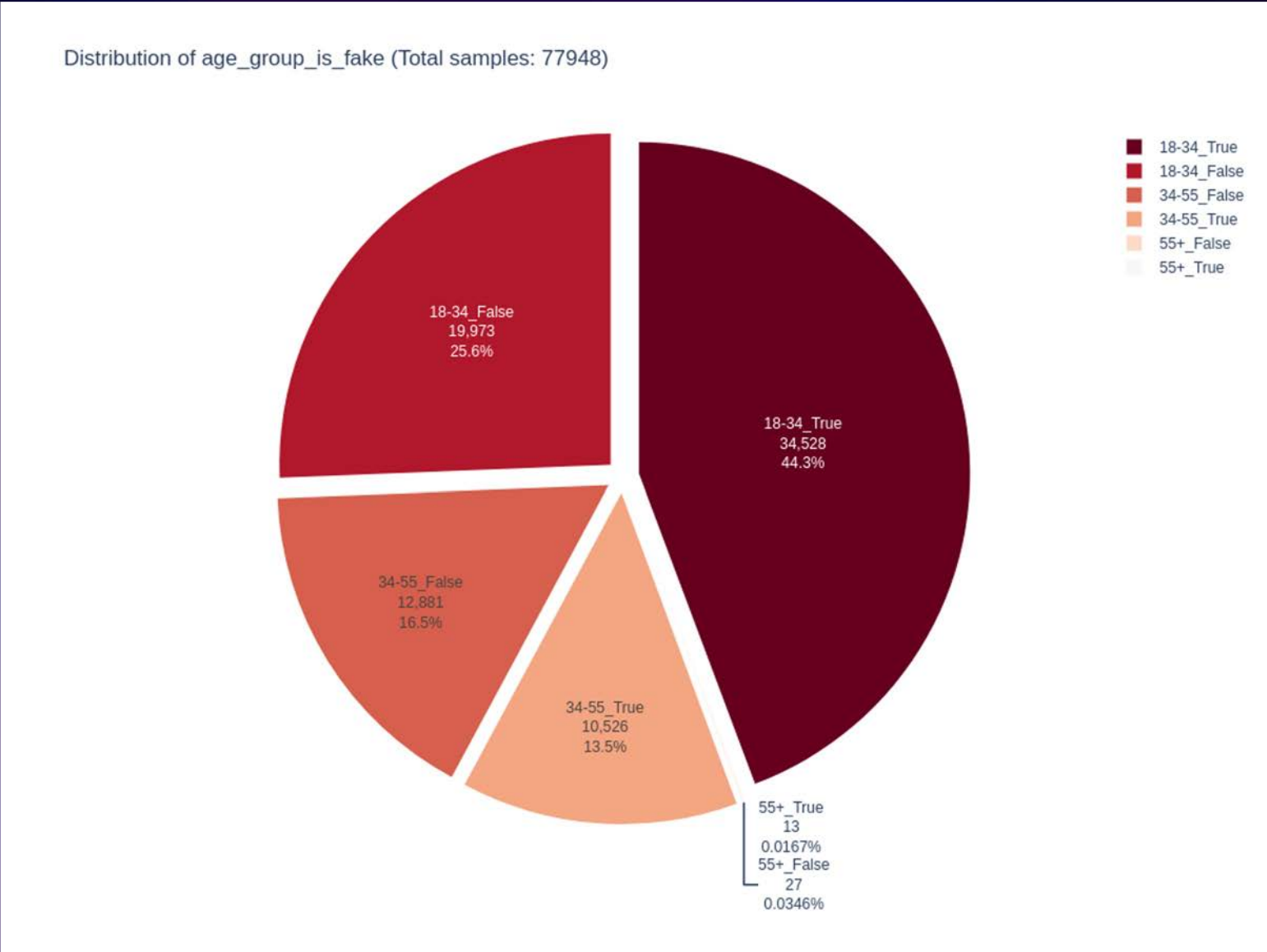


Combating DeepFakes Requires Dataset Generation: Face Age X Race Distribution

Distribution of age_group_dominant_race (Total samples: 77948)

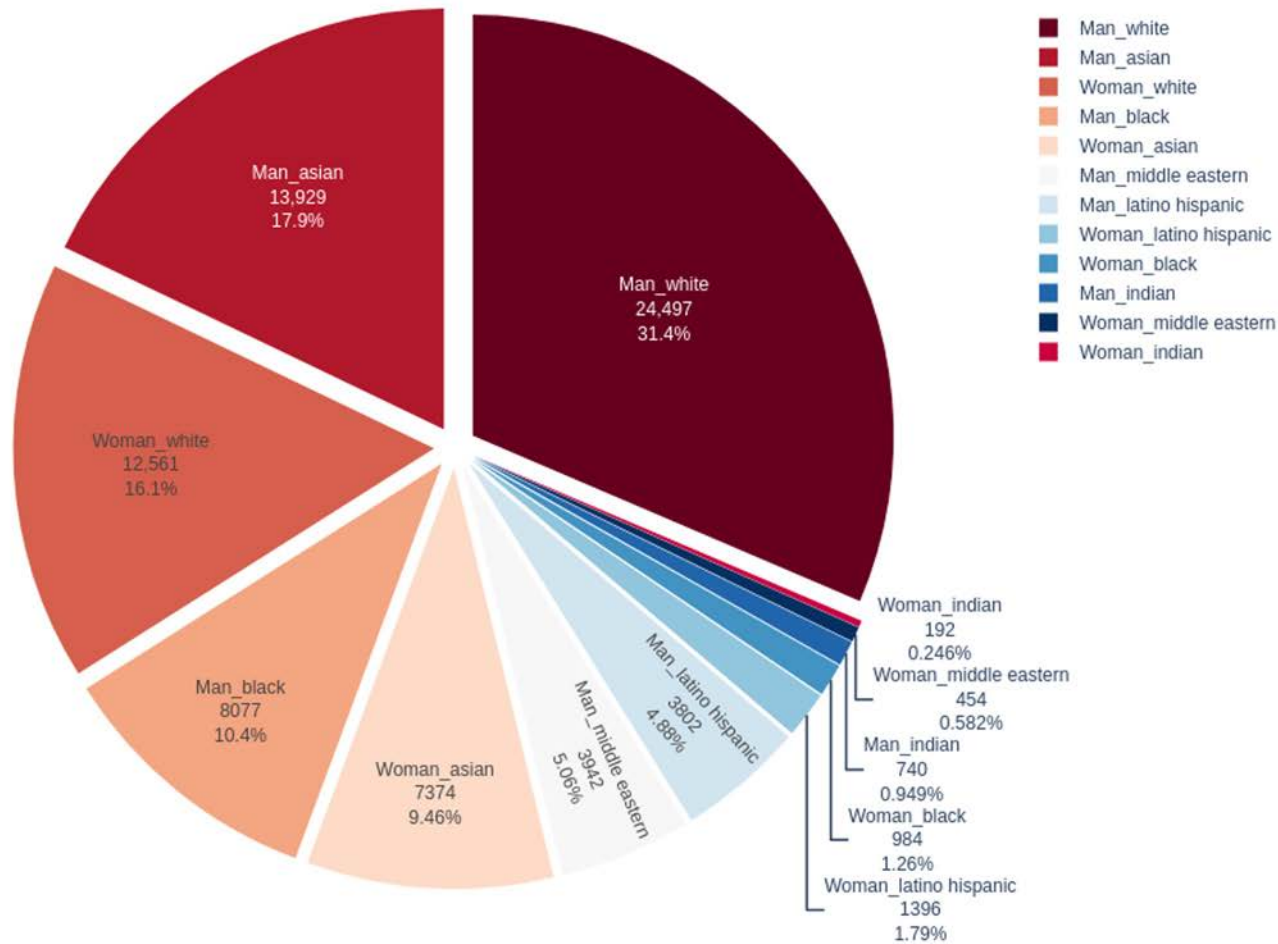


Combatting DeepFakes Requires Dataset Generation: Face Age X Real/Fake Distribution



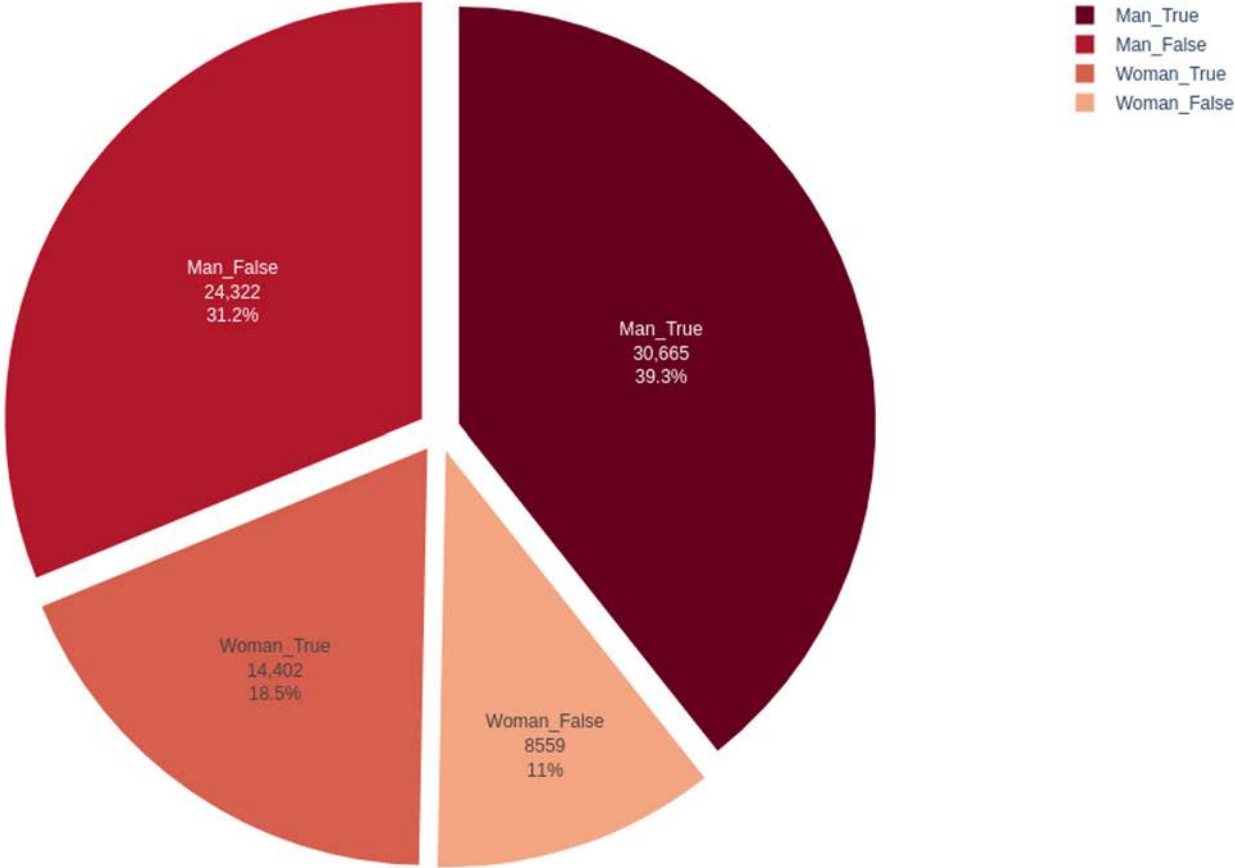
Combatting DeepFakes Requires Dataset Generation: Face Gender X Race Distribution

Distribution of dominant_gender_dominant_race (Total samples: 77948)



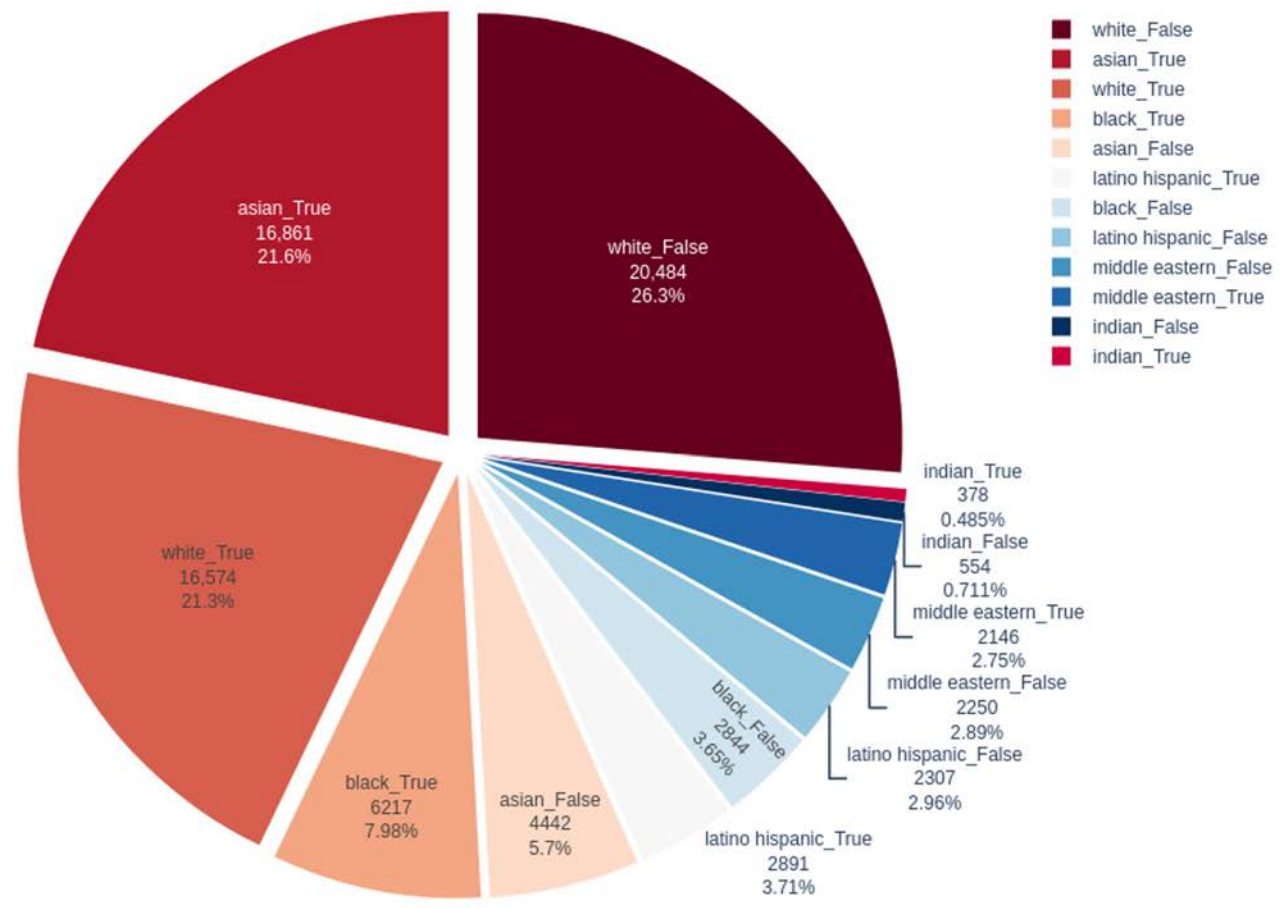
Combating DeepFakes Requires Dataset Generation: Face Gender X Real/Fake Distribution

Distribution of dominant_gender_is_fake (Total samples: 77948)



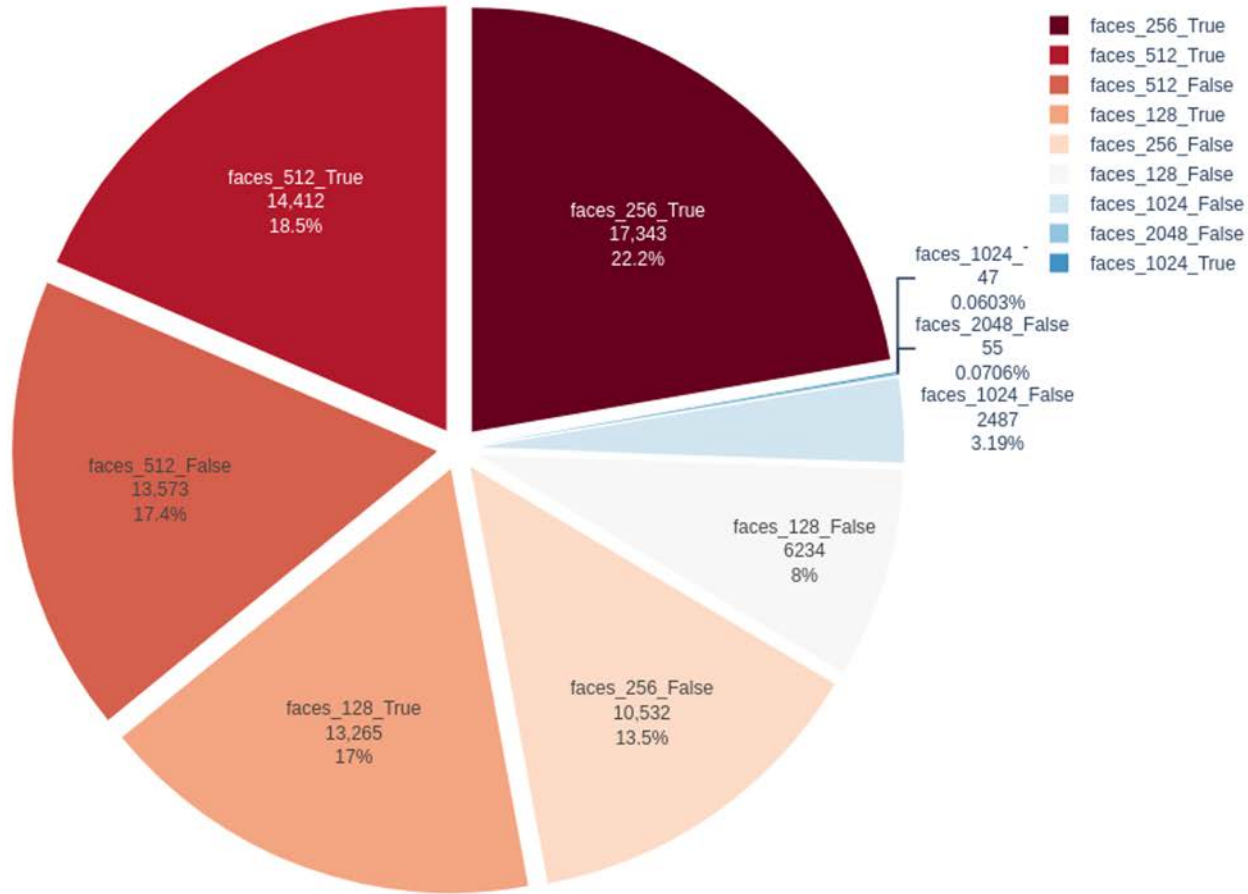
Combating DeepFakes Requires Dataset Generation: Face Race X Real/Fake Distribution

Distribution of dominant_race_is_fake (Total samples: 77948)

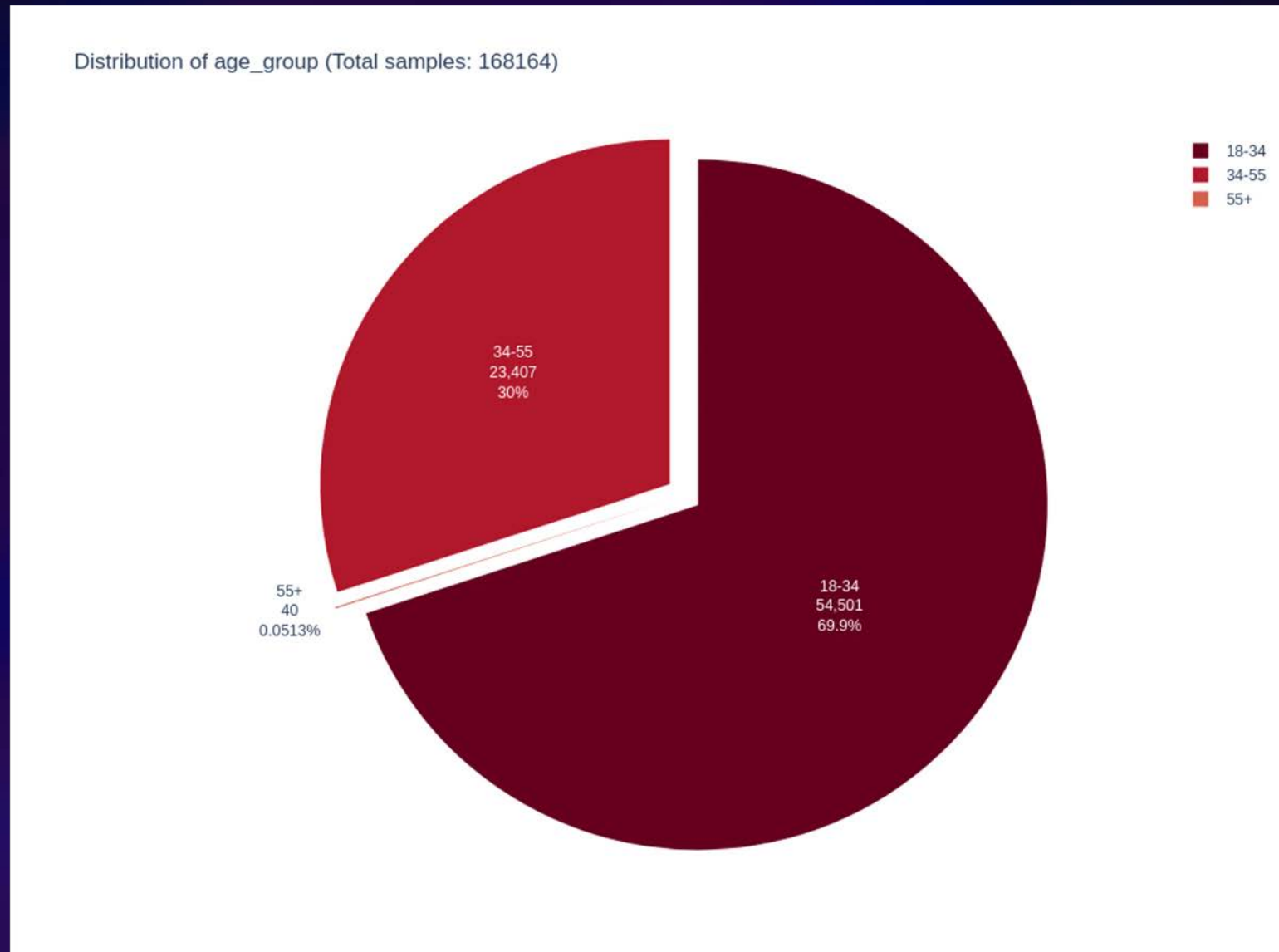


Combating DeepFakes Requires Dataset Generation: Face Resolution X Real/Fake Distribution

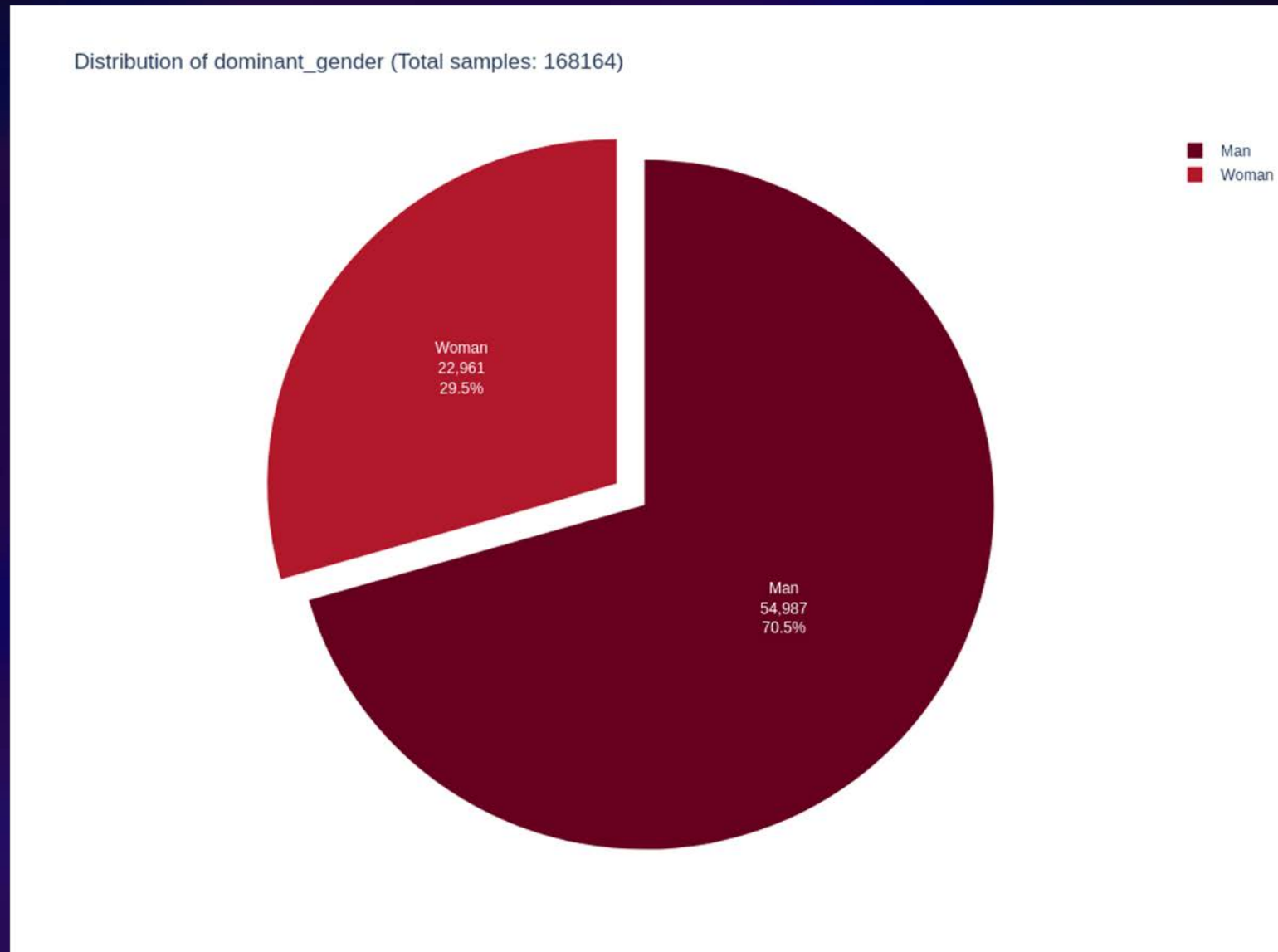
Distribution of face_size_is_fake (Total samples: 77948)



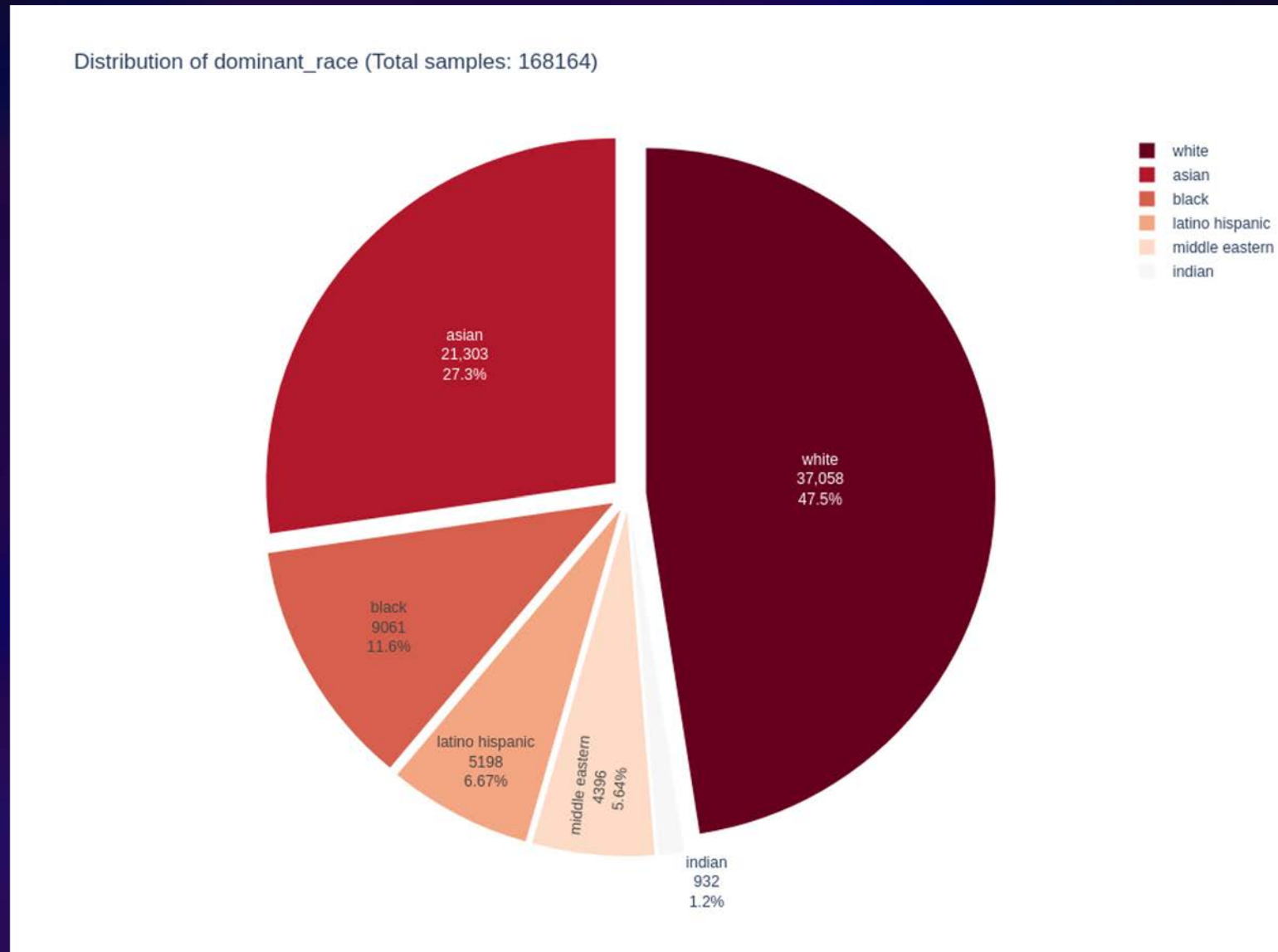
Combatting DeepFakes Requires Dataset Generation: Face Age Distribution



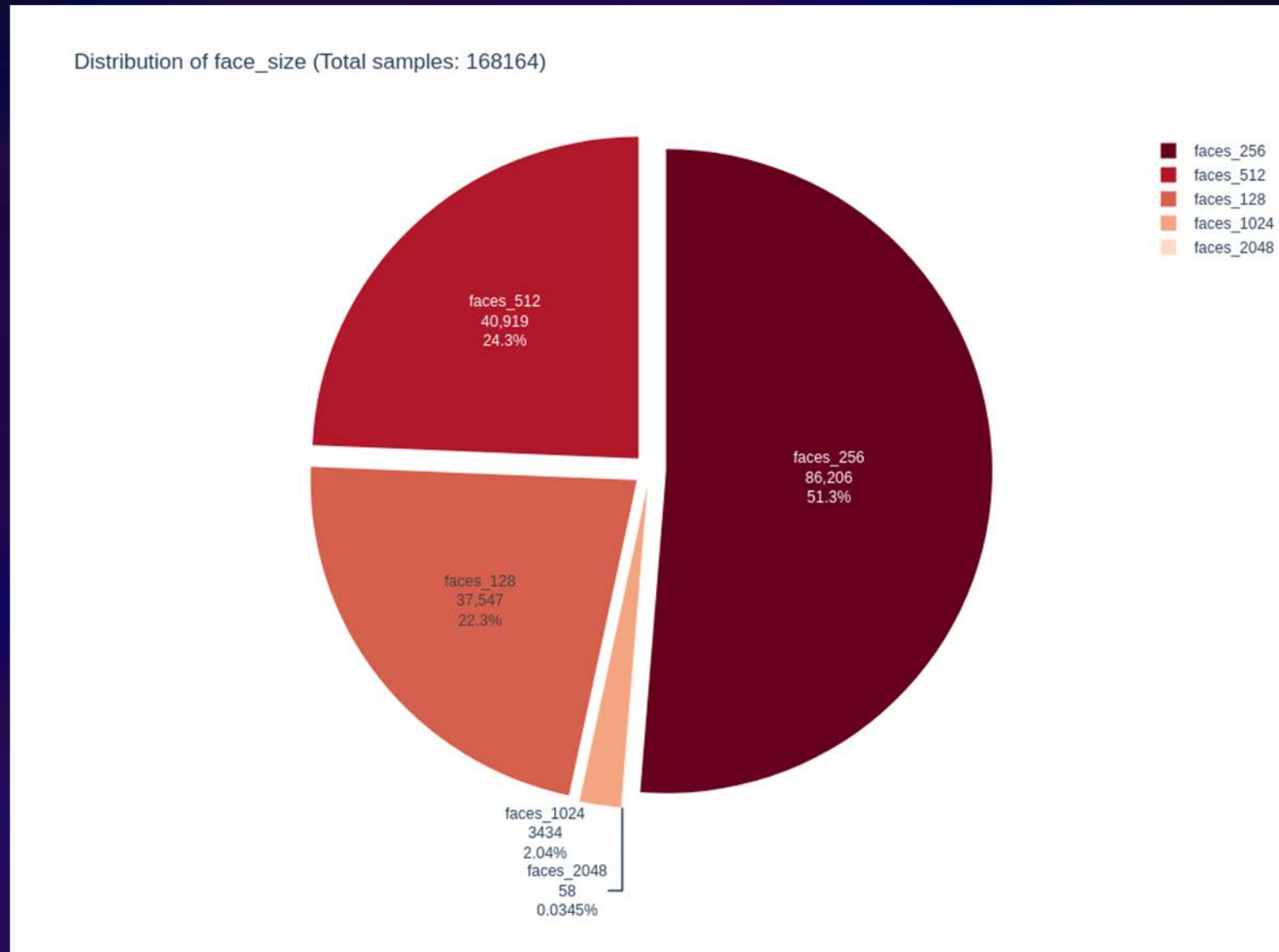
Combatting DeepFakes Requires Dataset Generation: Face Gender Distribution



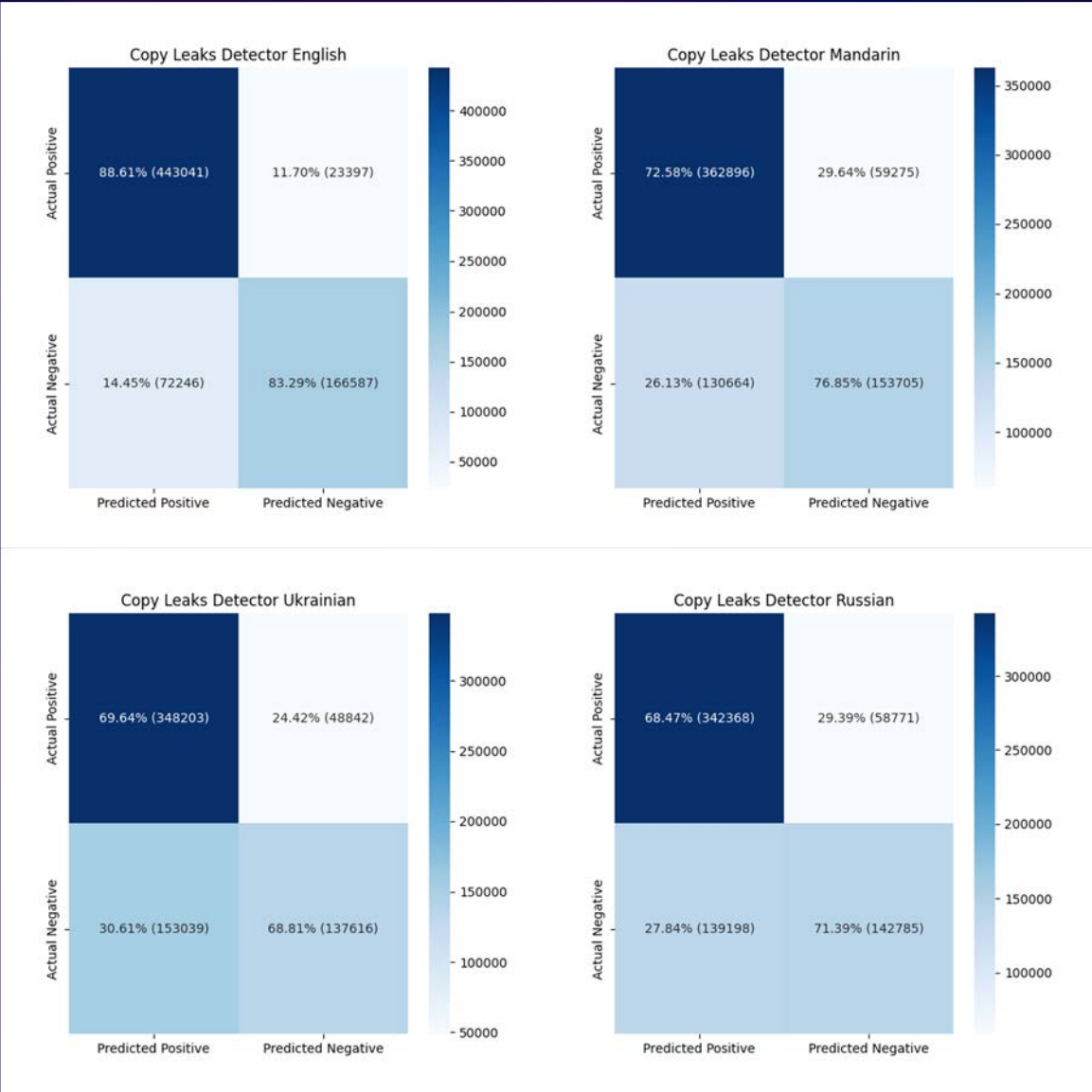
Combatting DeepFakes Requires Dataset Generation: Face Race Distribution



Combatting DeepFakes Requires Dataset Generation: Face Resolution Distribution

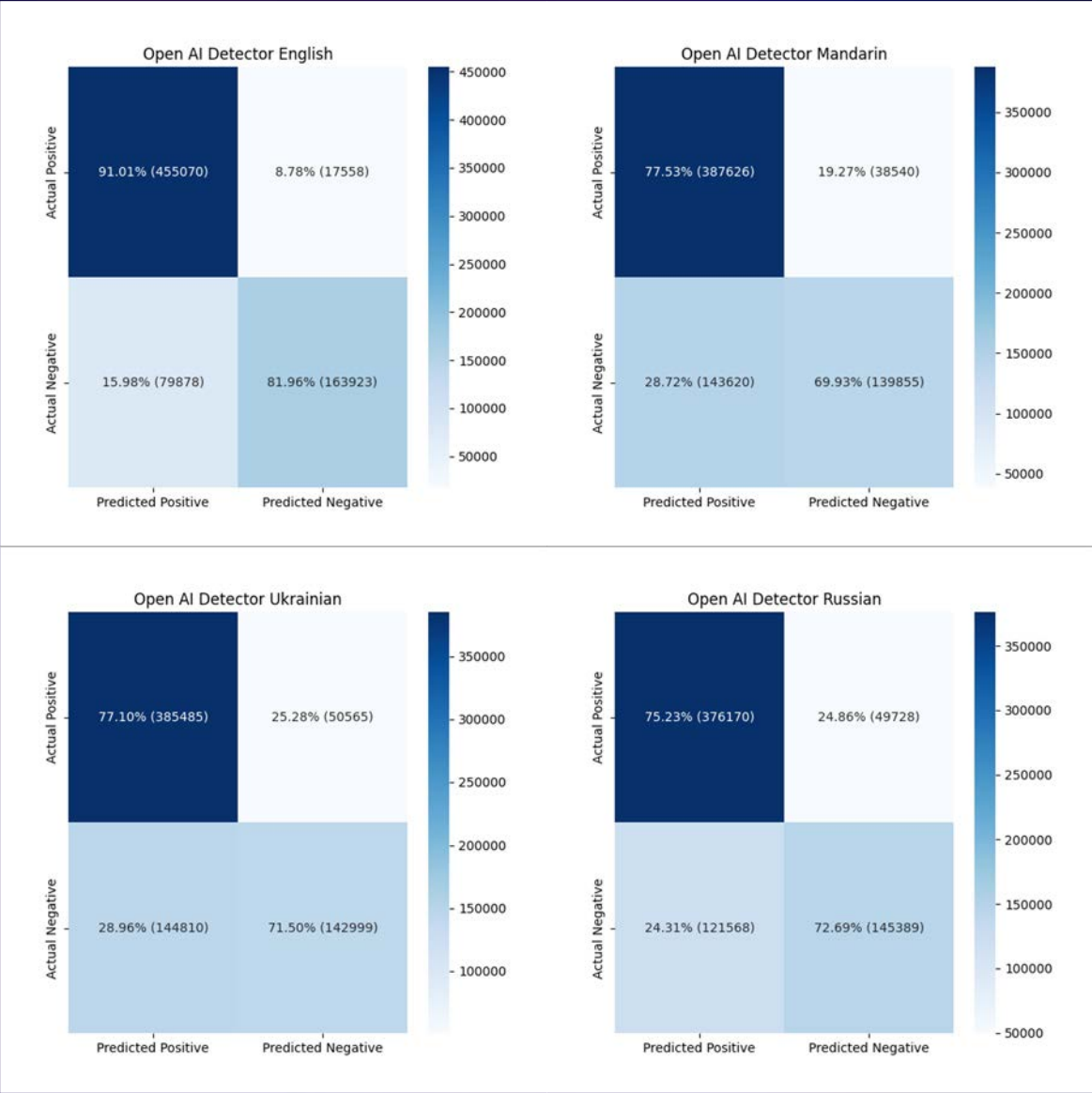


DeepFake Detection Results: Text Copy Leaks English, Mandarin, Russian, Ukrainian

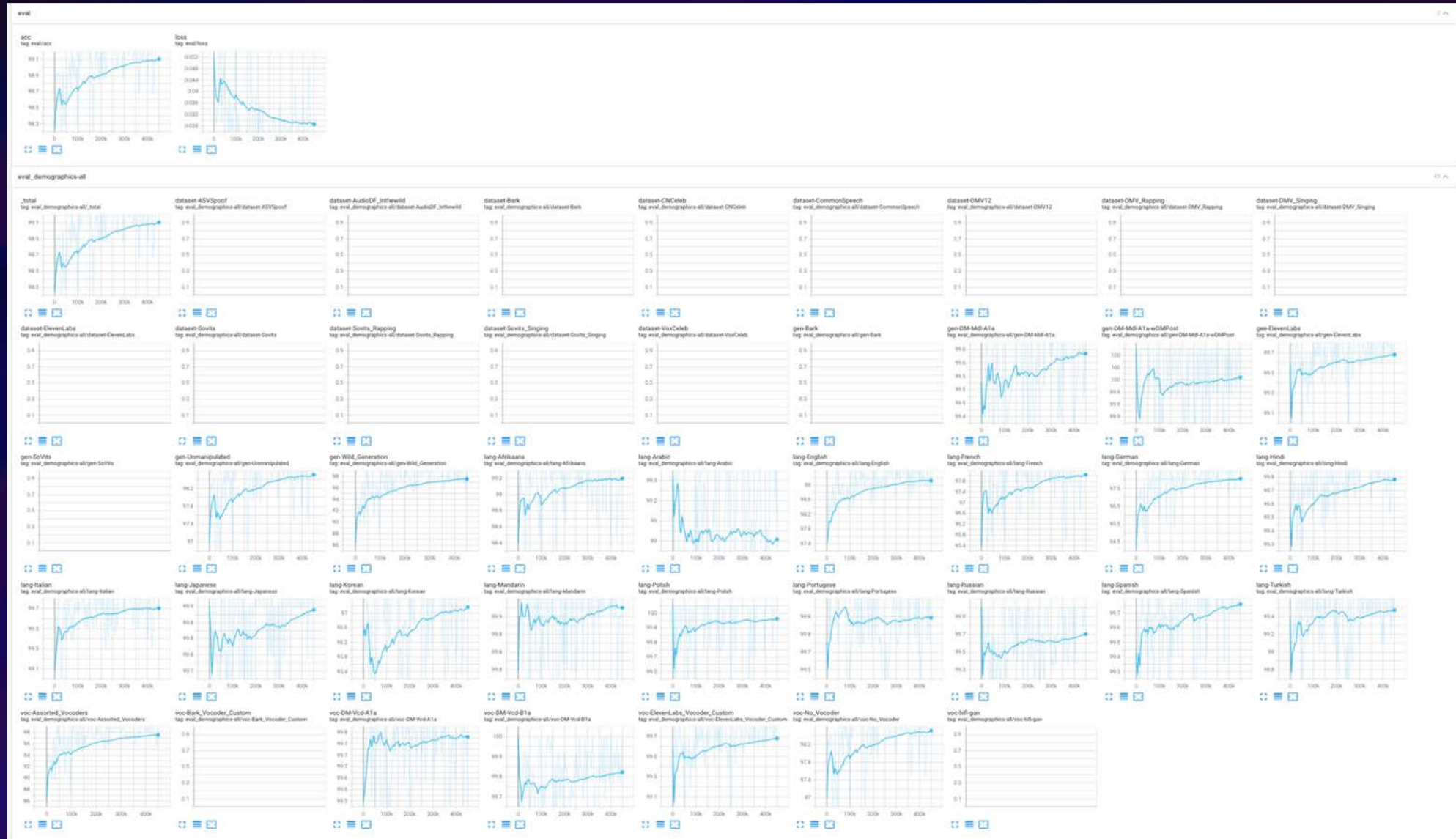


DeepFake Detection Results: Text

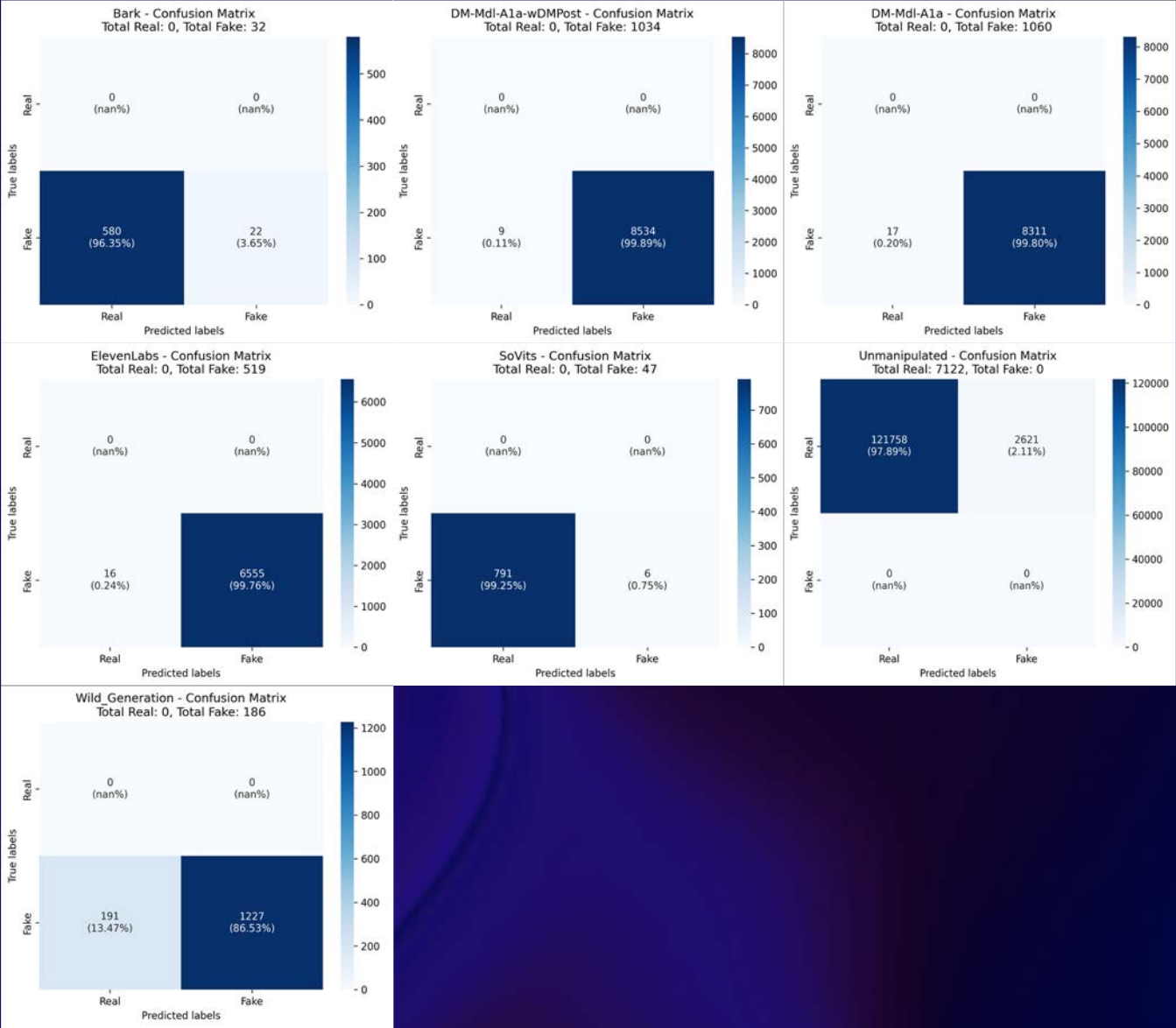
Open AI Detector English, Mandarin, Russian, Ukrainian



DeepFake Detection Results: Voice Attack Agnostic Detector Training Logs (No Augmentation)



DeepFake Detection Results: Voice Attack Agnostic Detector Results (No Augmentation)



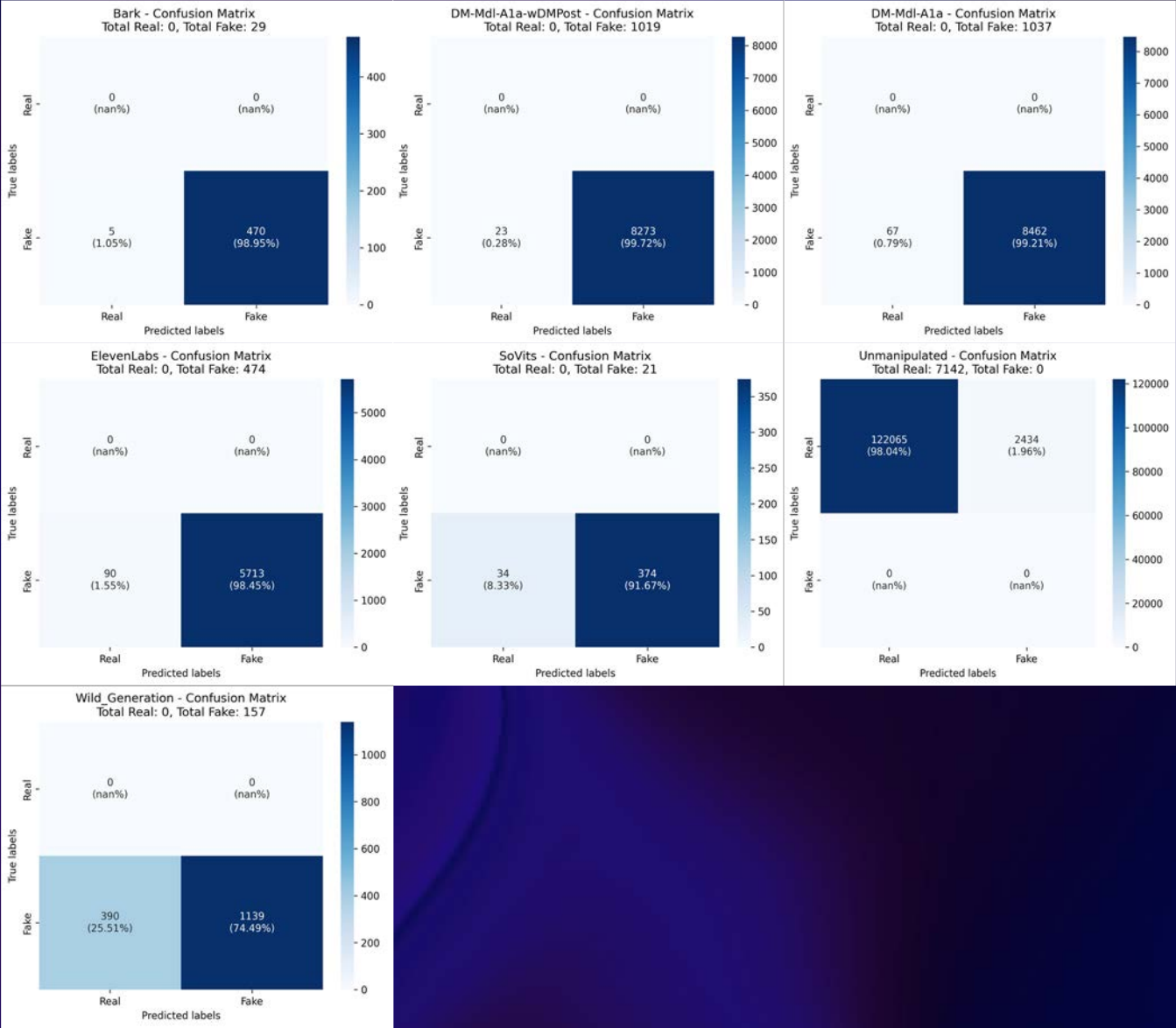
DeepFake Detection Results: Voice Attack Agnostic Detector Results (No Augmentation)



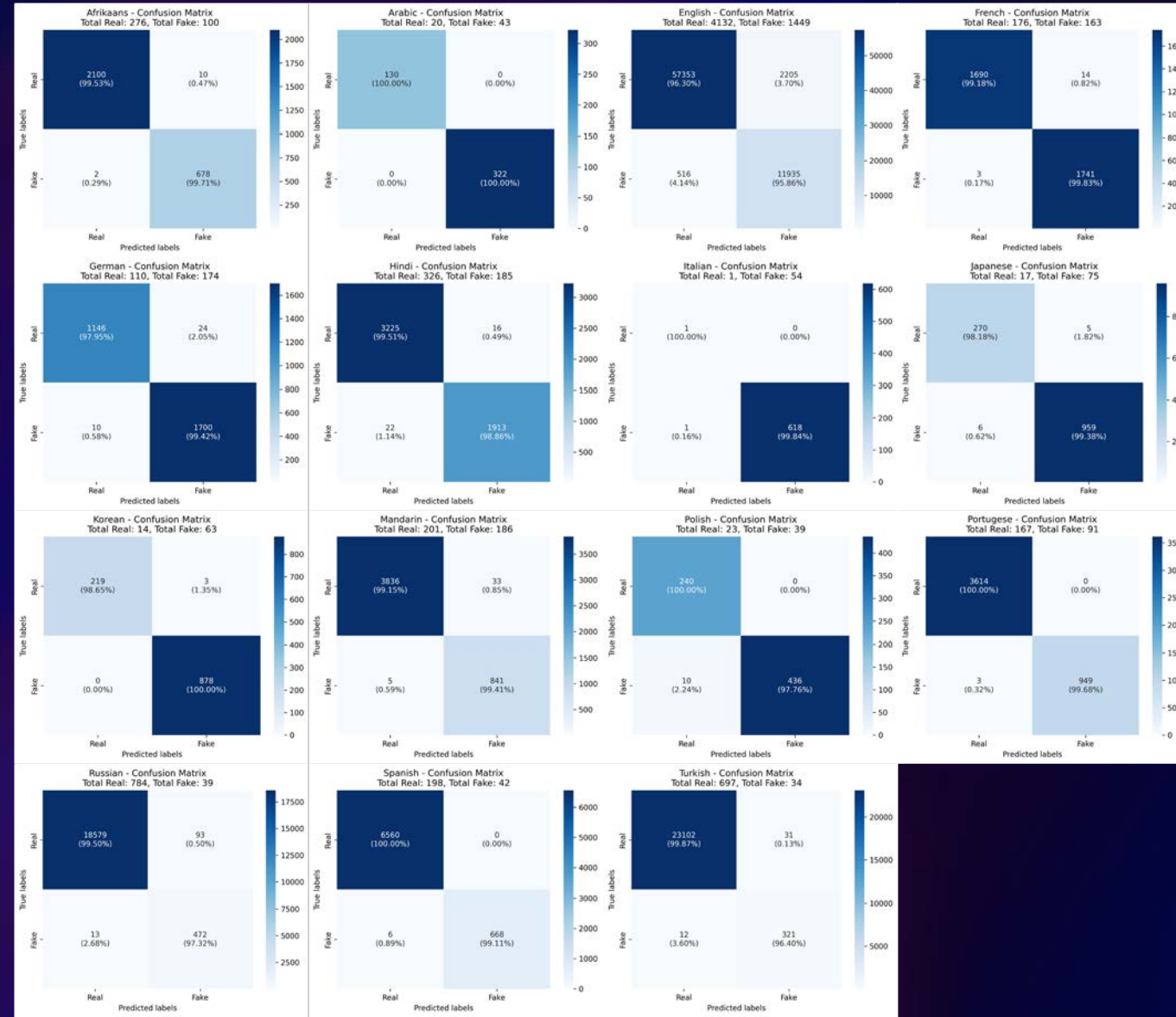
DeepFake Detection Results: Voice Attack Agnostic Detector Training Logs (With Augmentation)



DeepFake Detection Results: Voice Attack Agnostic Detector Results (With Augmentation)



DeepFake Detection Results: Voice Attack Agnostic Detector Results (With Augmentation)

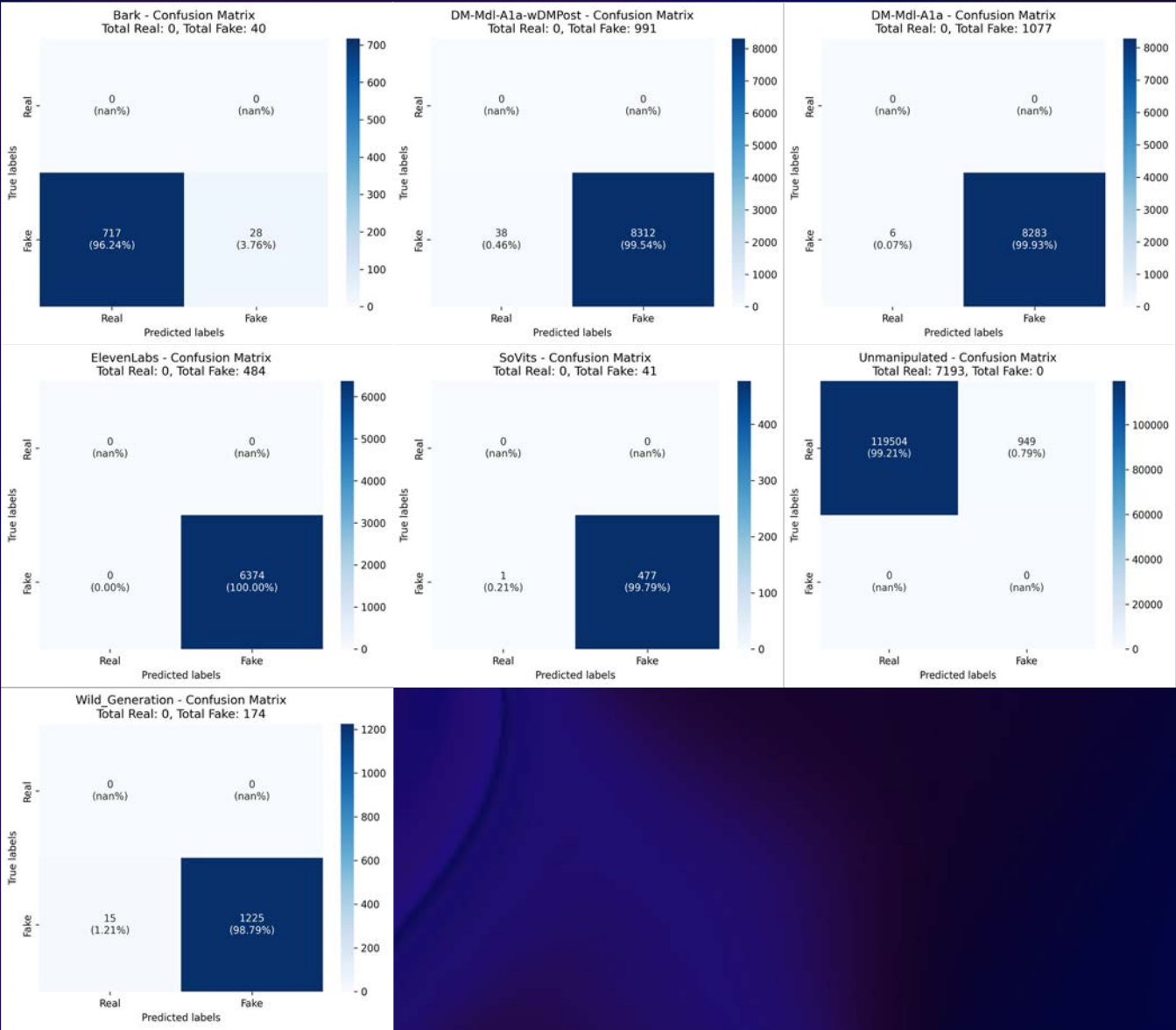


DeepFake Detection Results: Voice Efficient AT Detector Training Logs (No Augmentation)

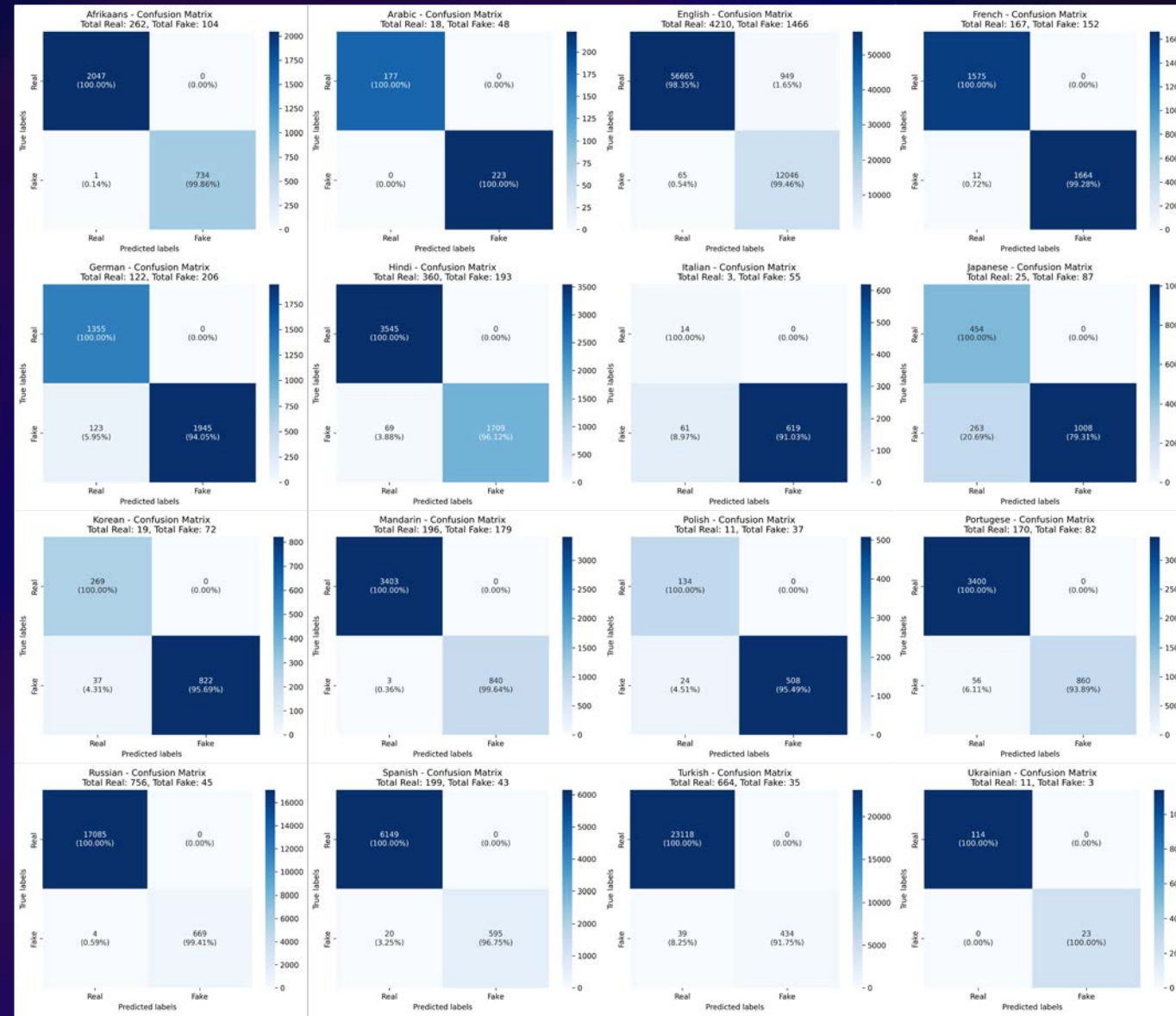


DeepFake Detection Results: Voice

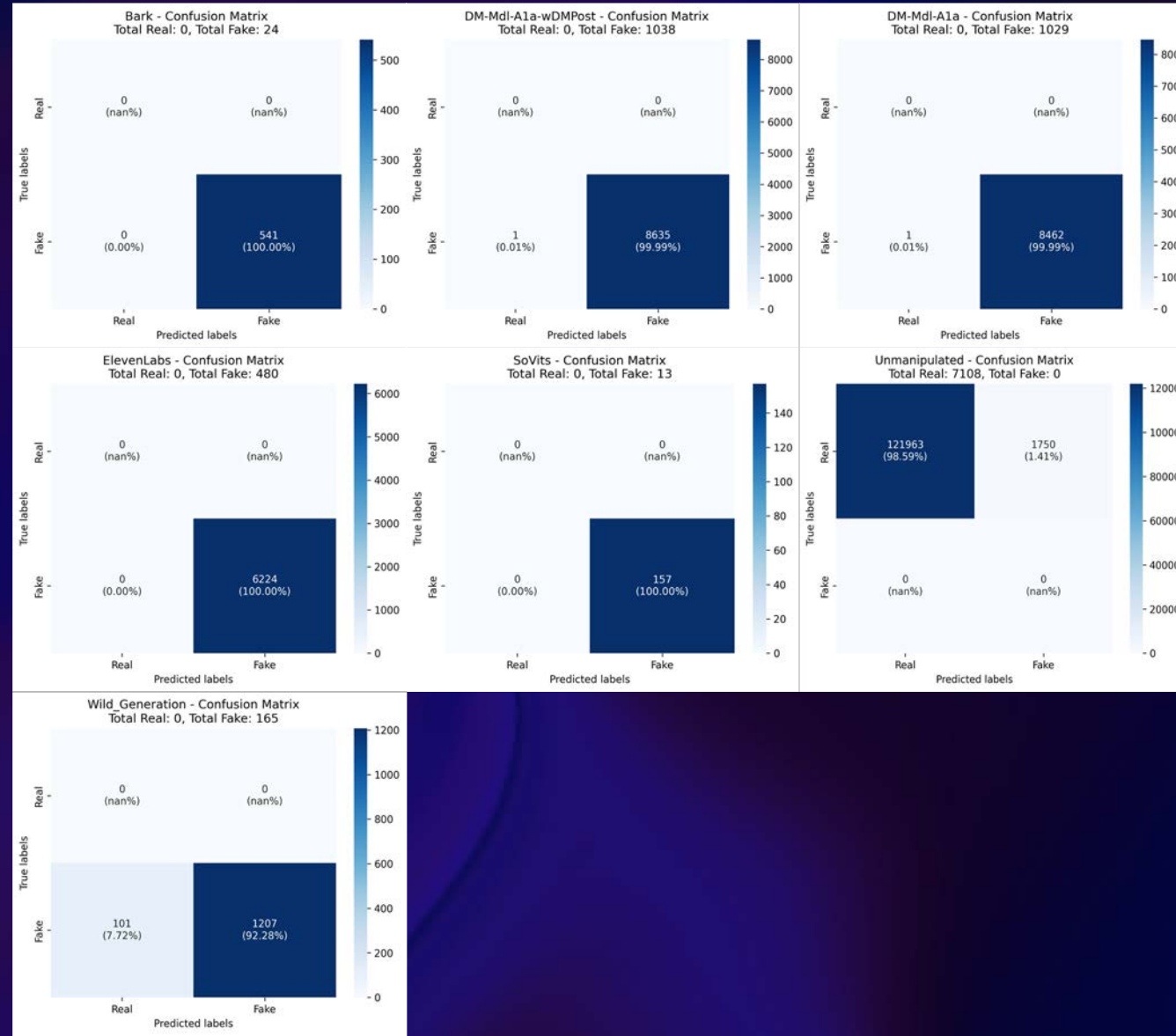
Efficient AT Detector Results (No Augmentation)



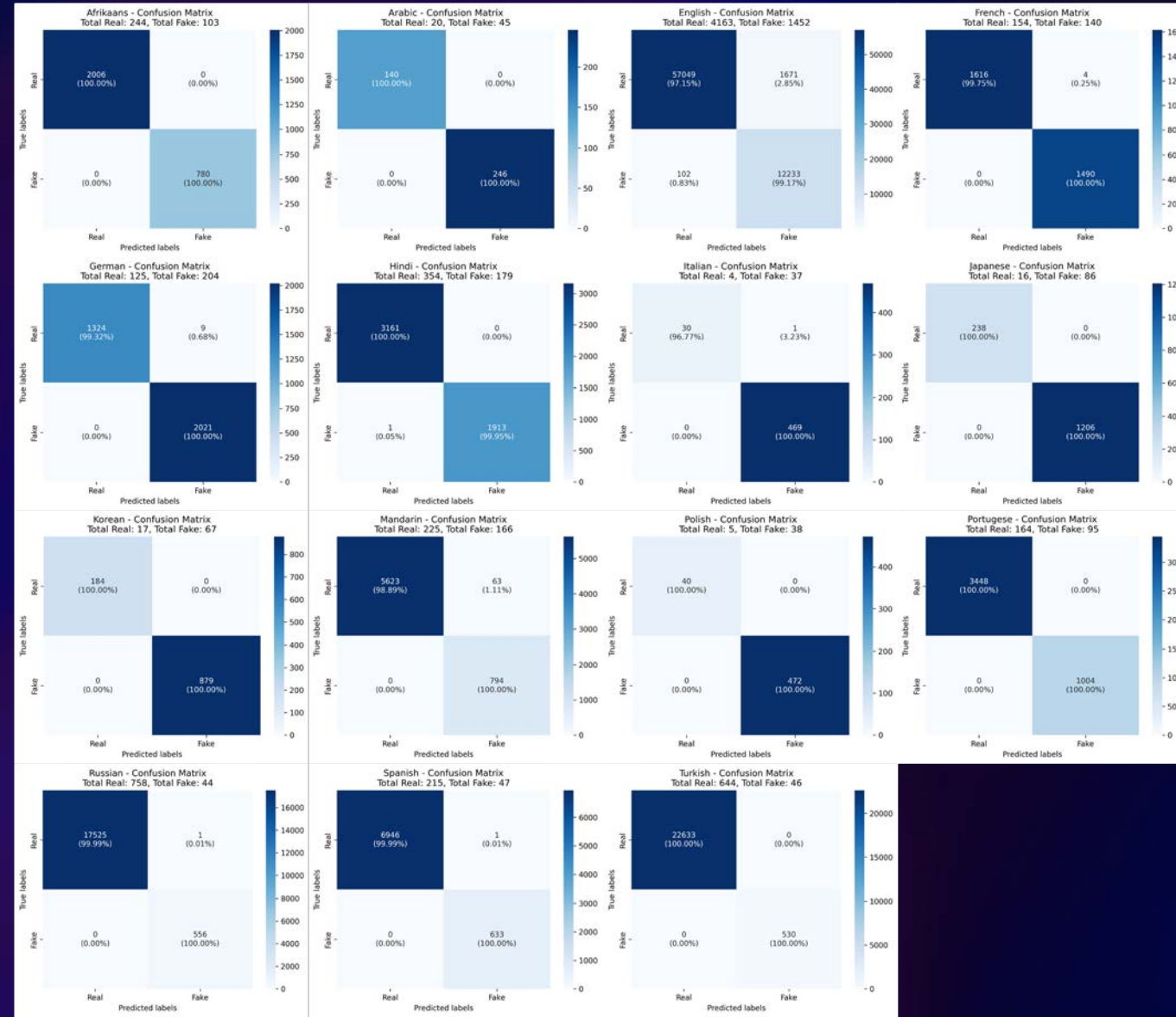
DeepFake Detection Results: Voice Efficient AT Detector Results (No Augmentation)



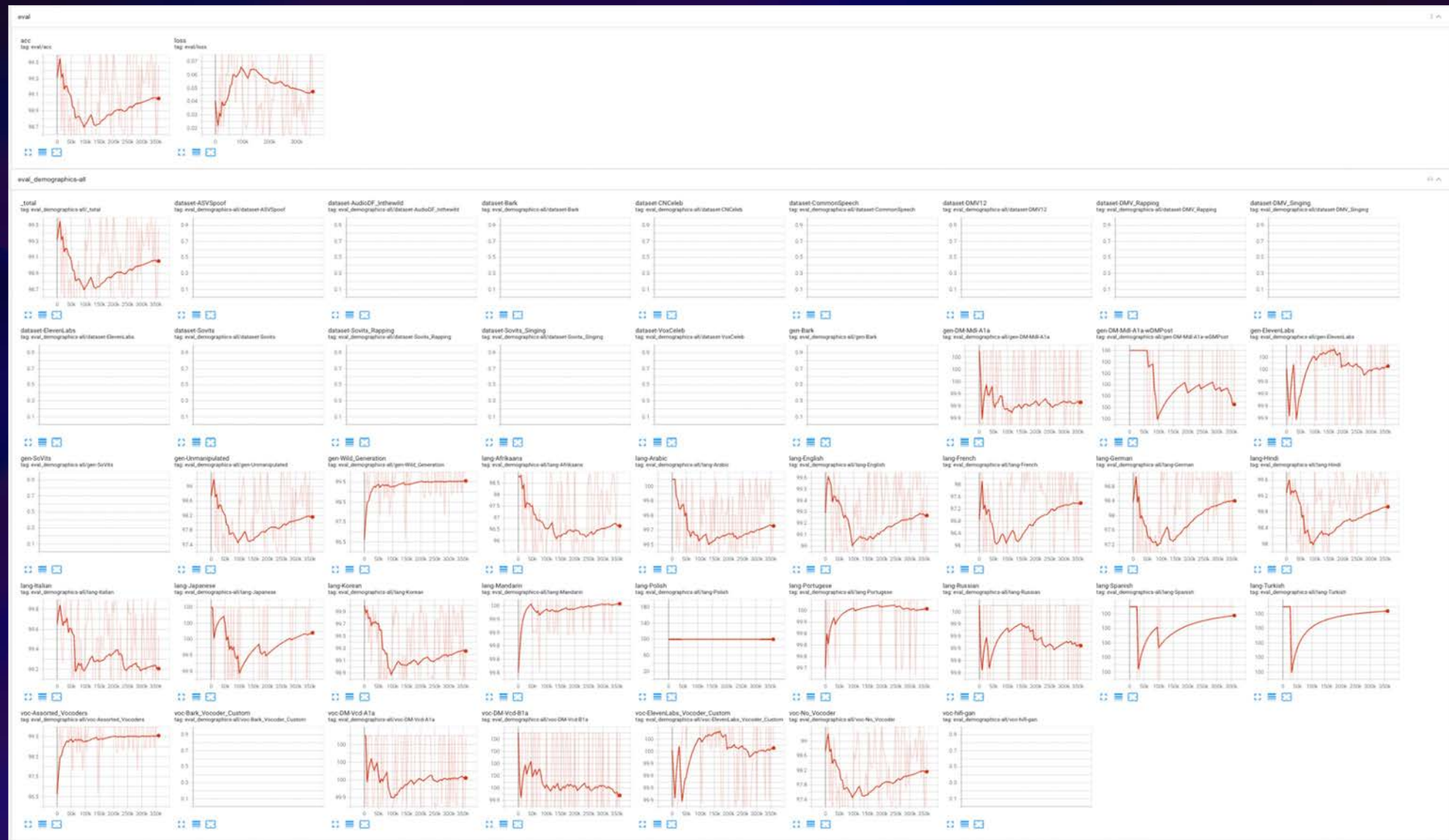
DeepFake Detection Results: Voice Efficient AT Detector Results (With Augmentation)



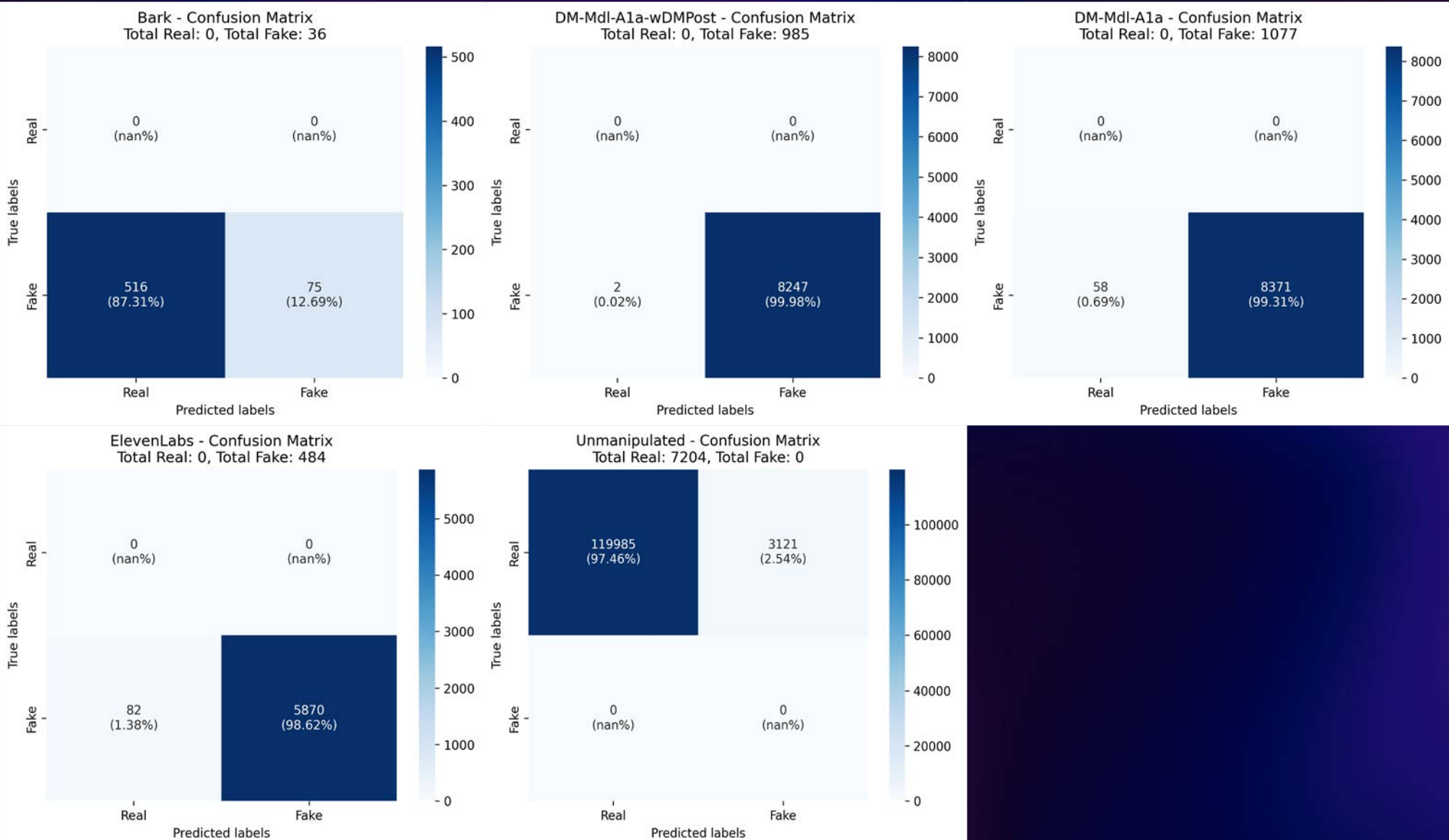
DeepFake Detection Results: Voice Efficient AT Detector Results (With Augmentation)



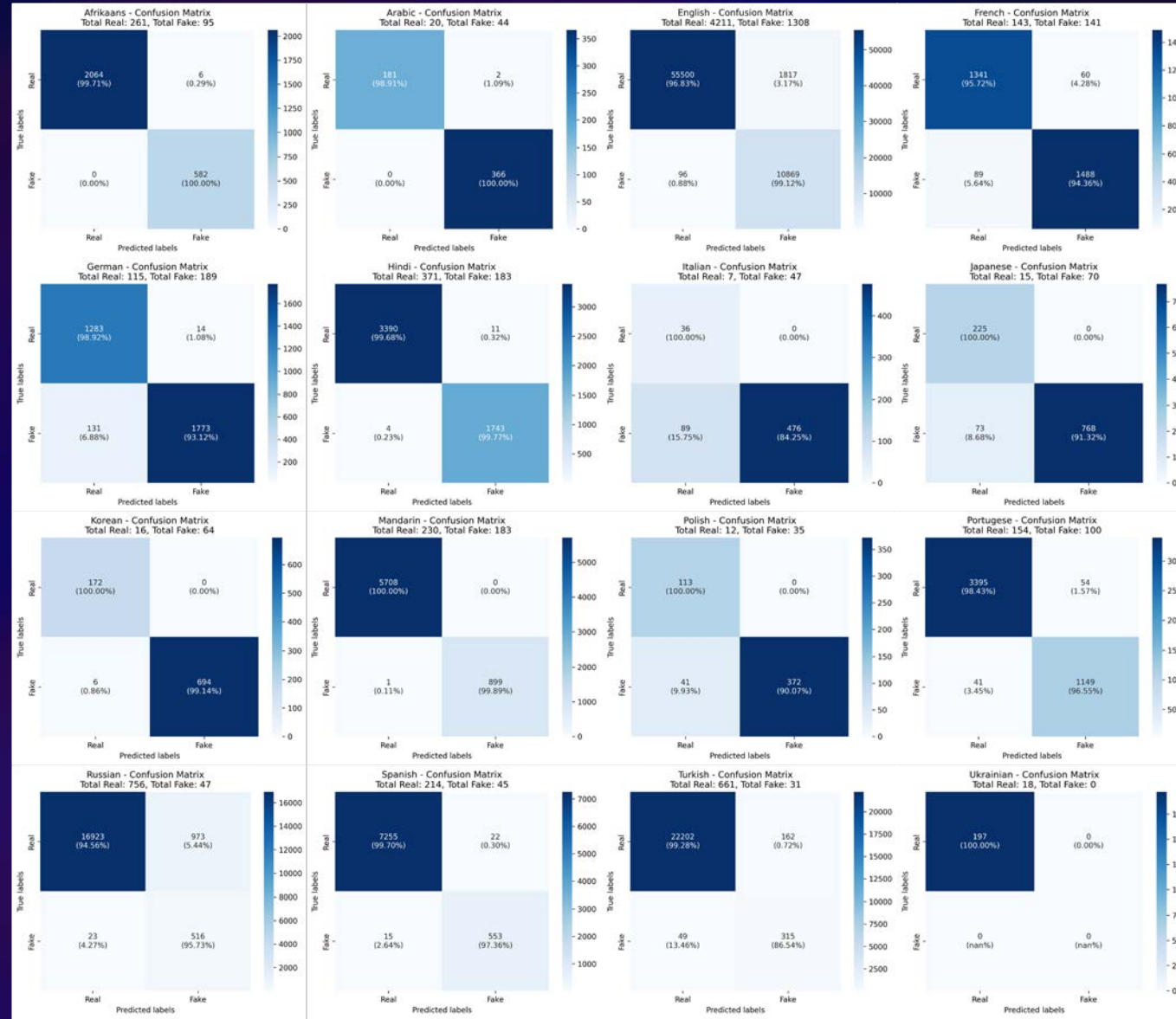
DeepFake Detection Results: Voice Soxan Detector Training Logs (No Augmentation)



DeepFake Detection Results: Voice Soxan Detector Results (No Augmentation)



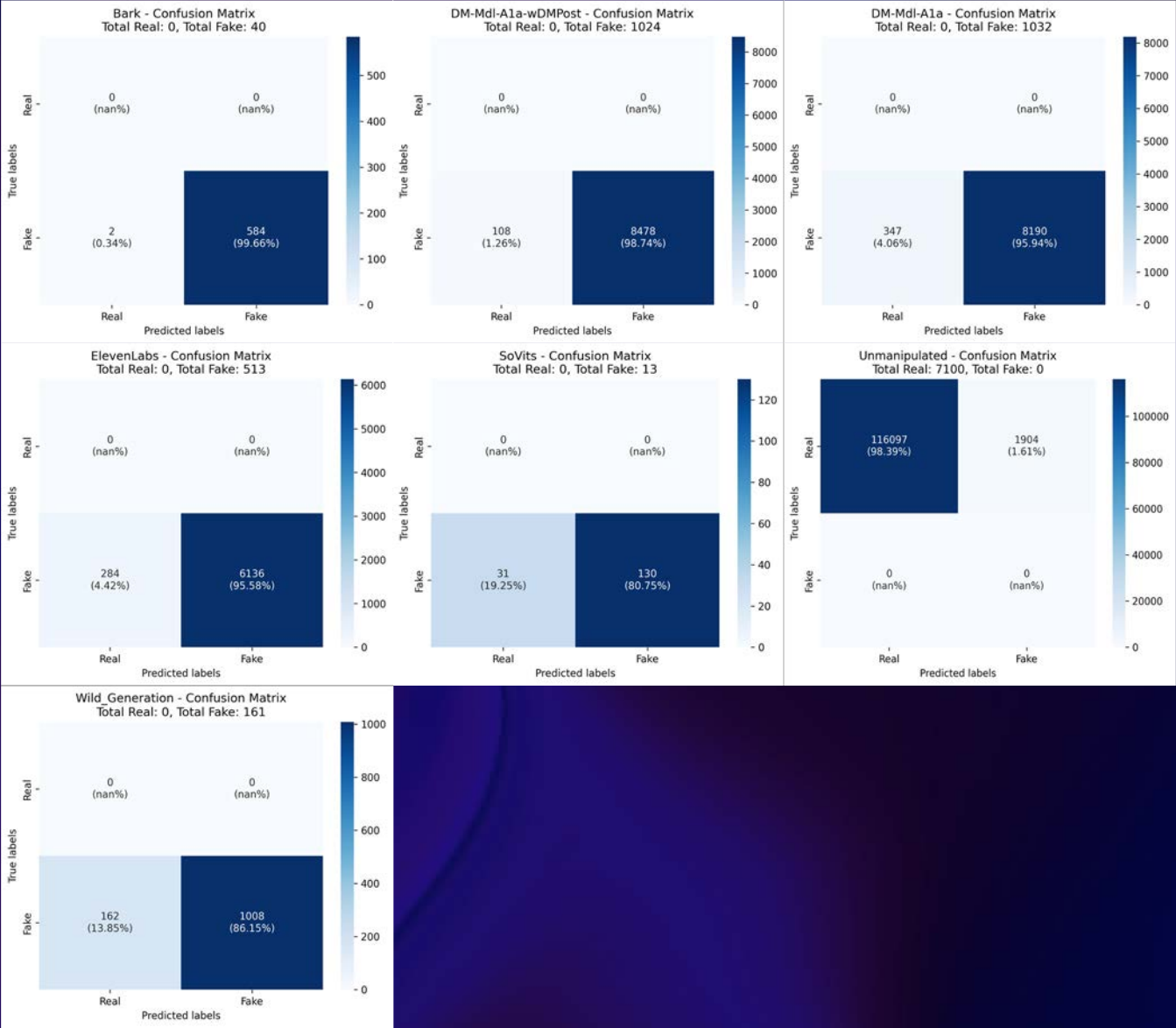
DeepFake Detection Results: Voice Soxan Detector Results (No Augmentation)



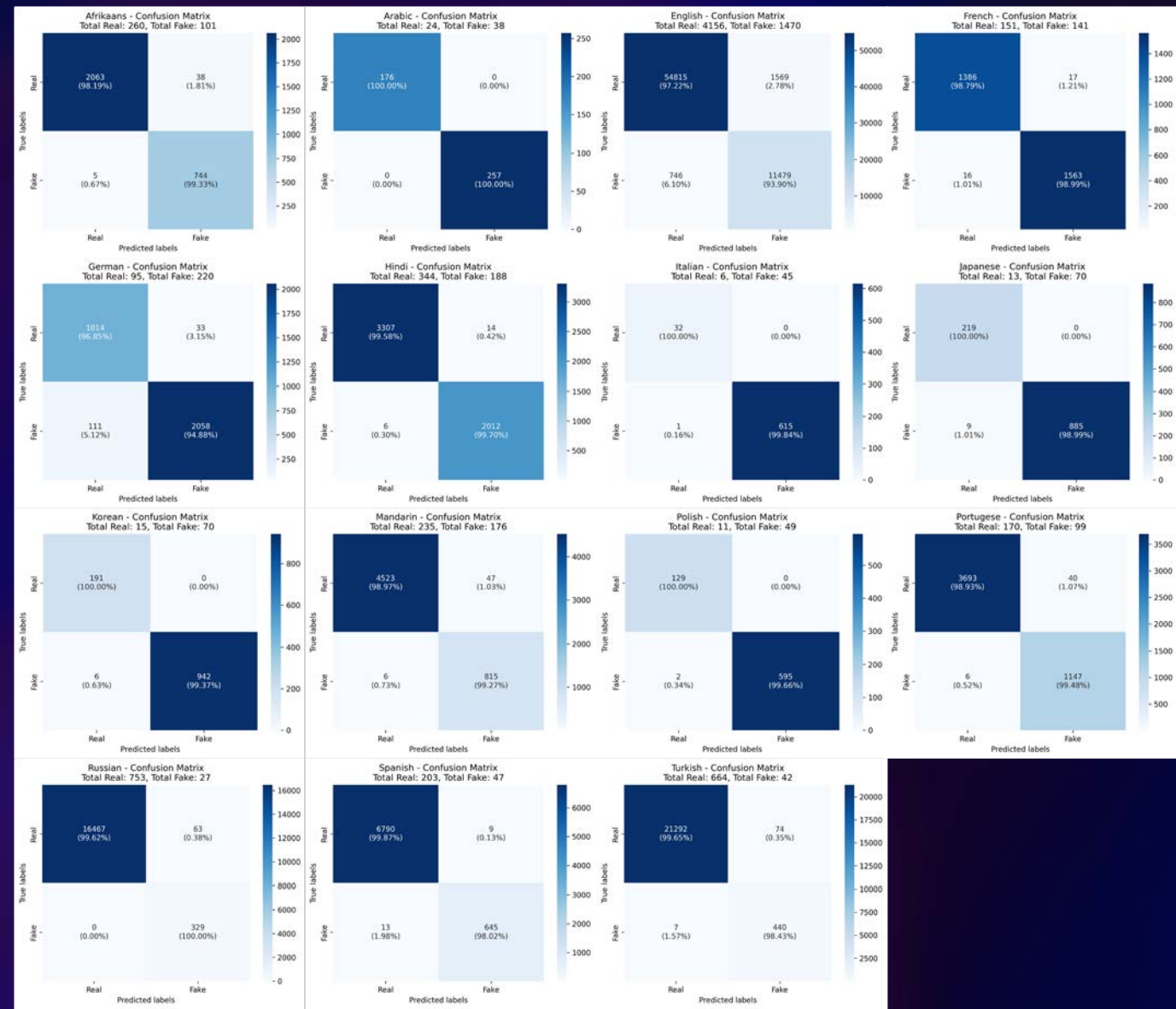
DeepFake Detection Results: Voice Soxan Detector Training Logs (With Augmentation)



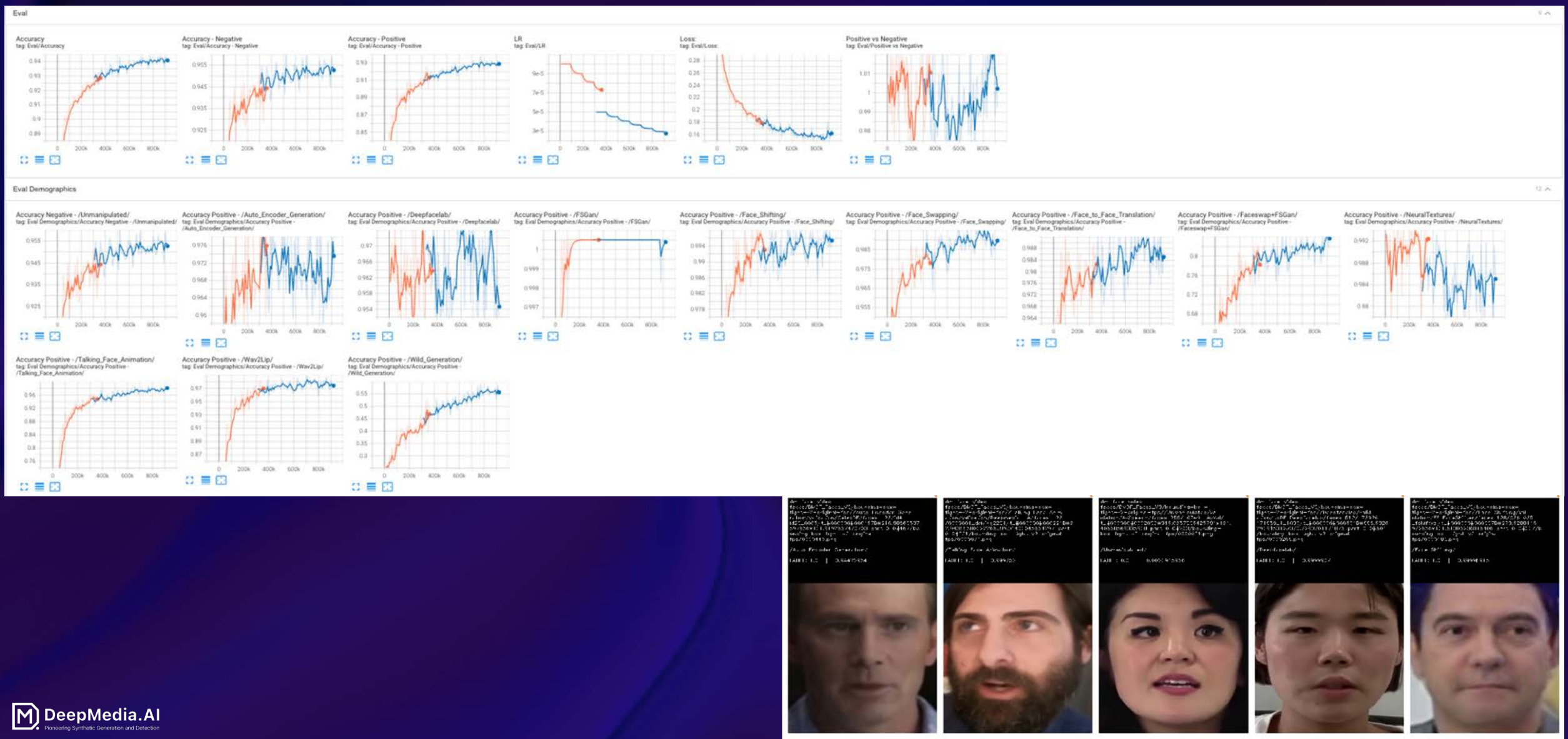
DeepFake Detection Results: Voice Soxan Detector Results (With Augmentation)



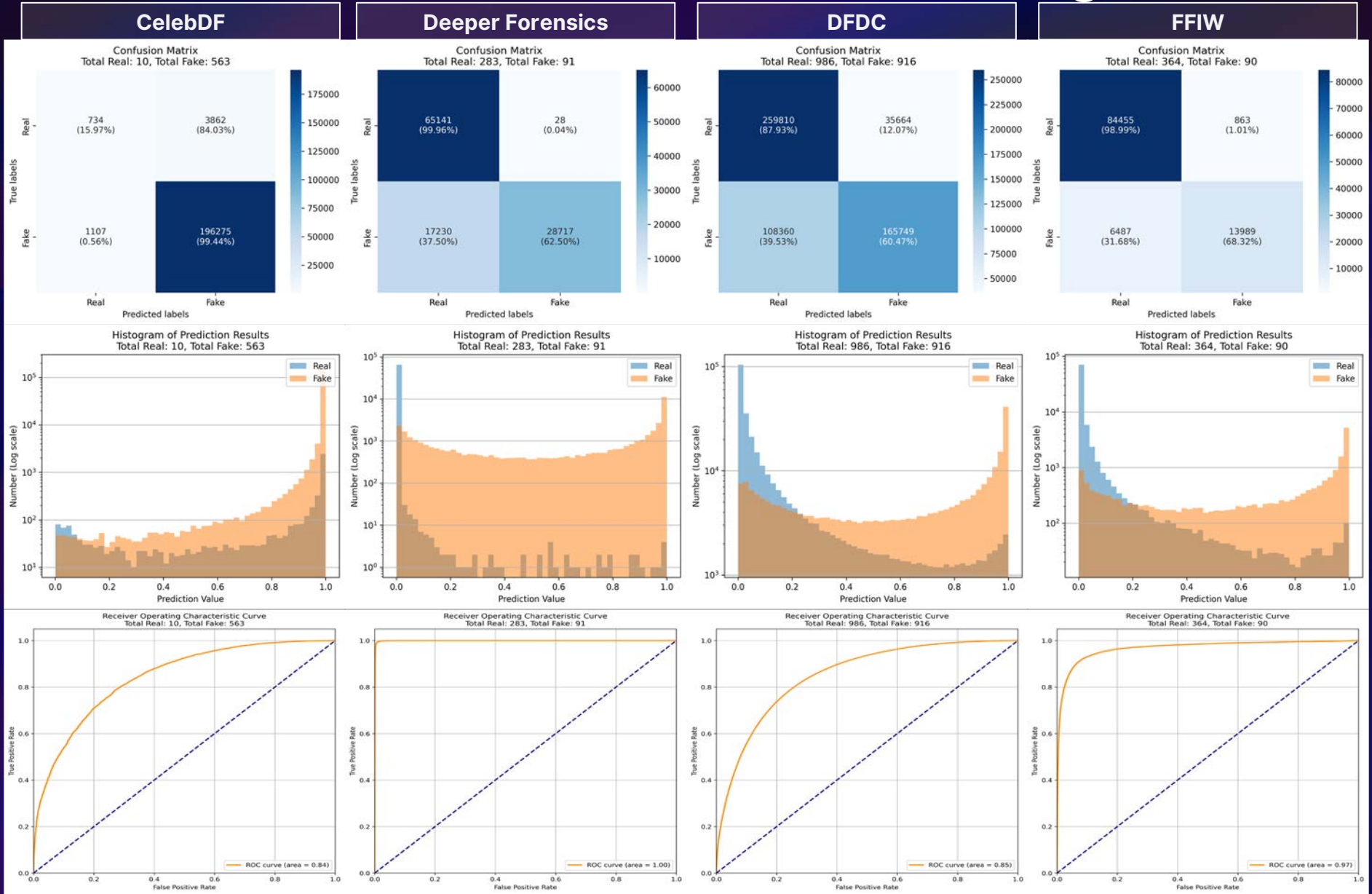
DeepFake Detection Results: Voice Soxan Detector Results (With Augmentation)



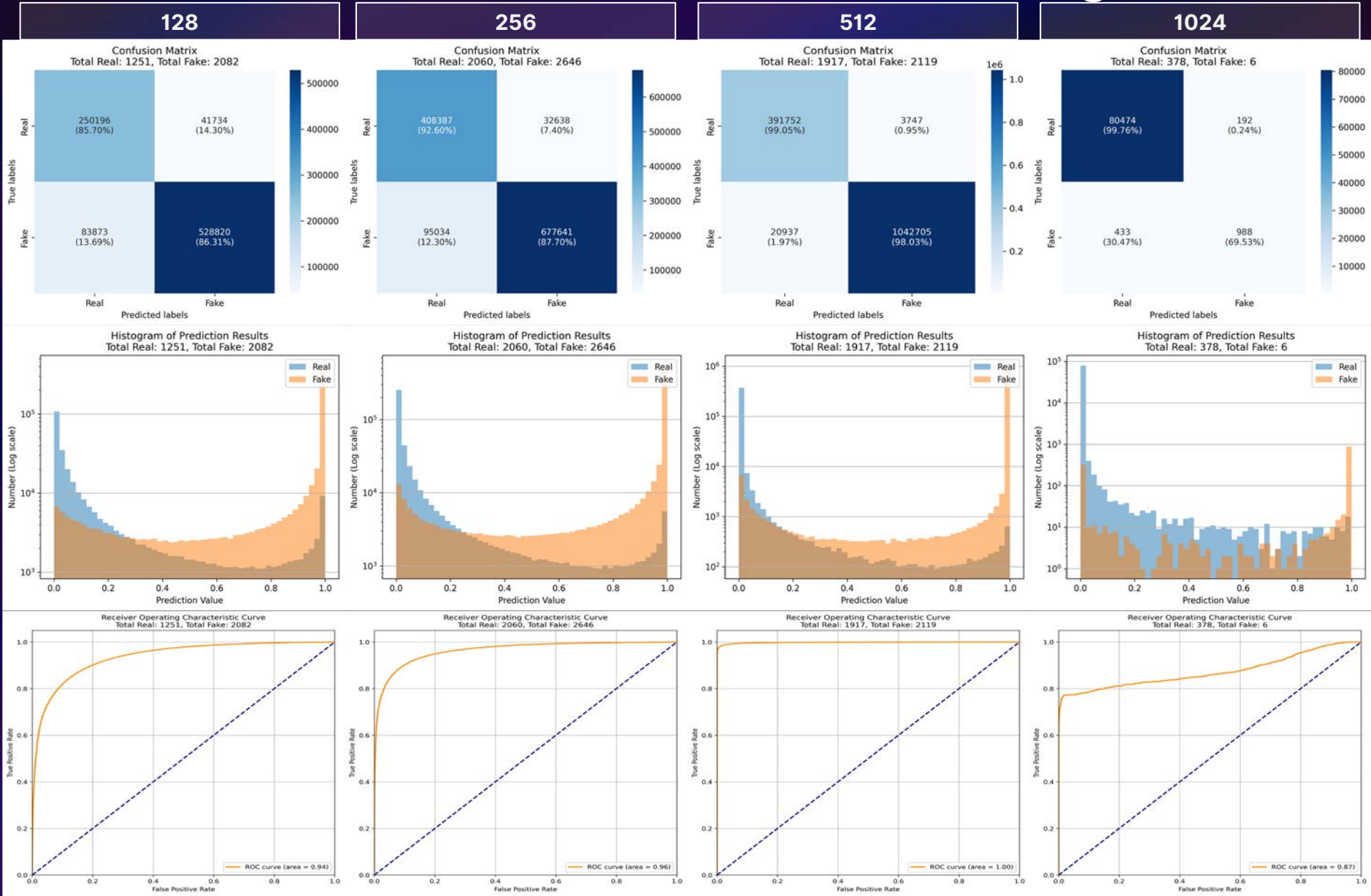
DeepFake Detection Results: Face CEViT Detector Training Logs (No Augmentation)



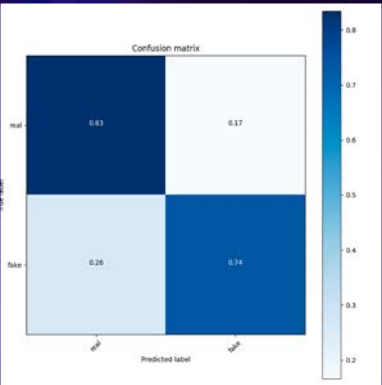
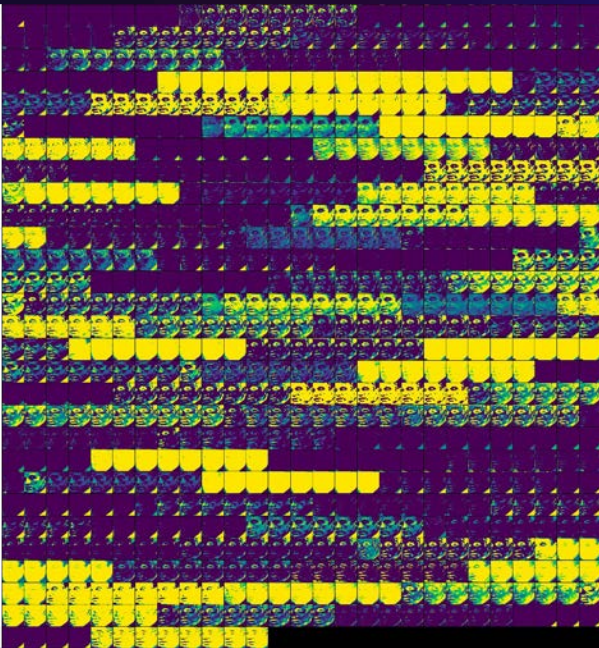
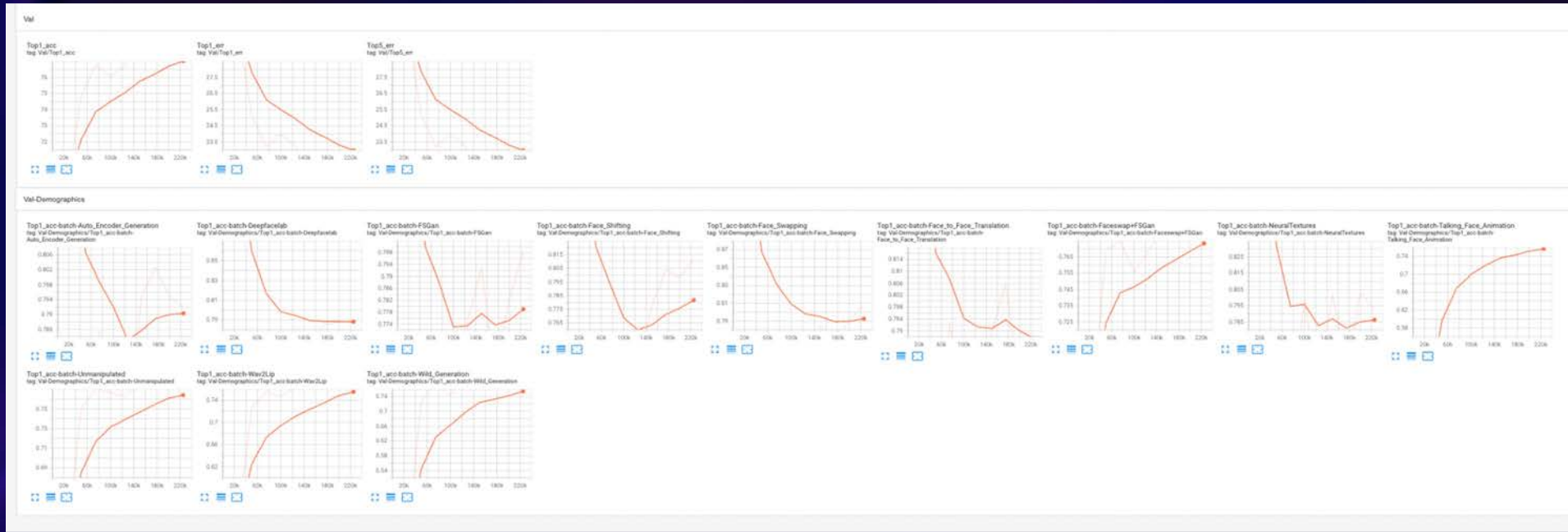
DeepFake Detection Results: Face CEViT Detector Results on Datasets (No Augmentation)



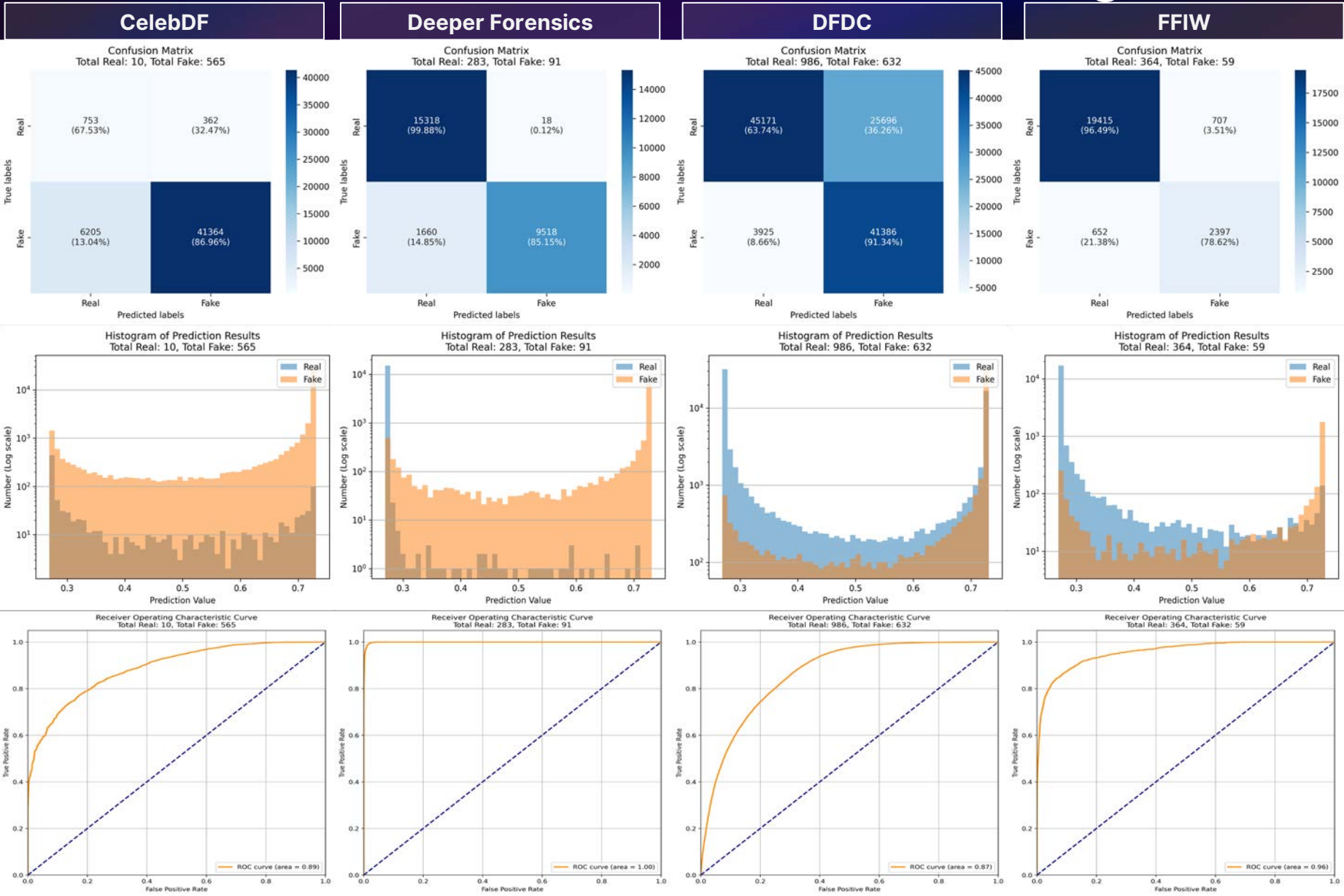
DeepFake Detection Results: Face CEViT Detector Results on Face Size (No Augmentation)



DeepFake Detection Results: Face RMViT Detector Training Logs (w/ Augmentation)



DeepFake Detection Results: Face RevMViT Detector Results on Datasets (No Augmentation)



DeepFake Detection Results: Face RevMViT Detector Results on Face Size (No Augmentation)

