PARAVISION

AN INTRODUCTION **TO PARAVISION DEEPFAKE DETECTION**

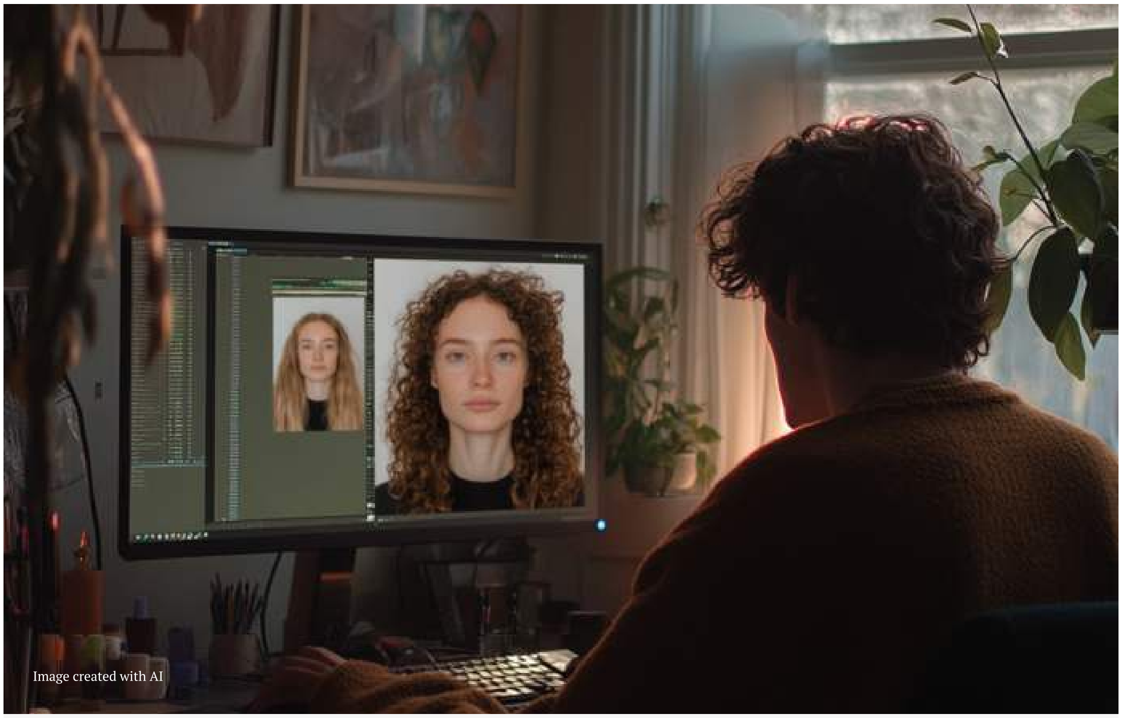


Image created with AI

⟶ paravision.ai

Trusted Vision AI

January 2025

Introduction

The increasing sophistication of deepfake technology has raised significant security concerns in government, traditional and social media, audio and video communications, and a broad range of identity-centric applications. Deepfake manipulations have become a particular concern in digital identity applications, including onboarding and identity verification as well as identification and authentication. They undermine trust in online identities, facilitate identity fraud, and disrupt security measures, making robust detection a critical need wherever digital identities are used.

Paravision's Deepfake Detection is a cutting-edge solution designed to help counter this threat, achieving high accuracy and low error rates across varied testing scenarios, protecting against the latest manipulation techniques including Face Swaps and Expression Swaps. This white paper delves into Paravision's approach to deepfake detection, examining how it works, its accuracy, and the key advantages it provides to users.

Paravision Deepfake Detection is a natural counterpart of Paravision Liveness, designed to work together seamlessly to form a robust foundation for ensuring Authentic Identity. While Liveness checks for the presence of physical presentation attacks like masks or high-resolution displays, Deepfake Detection adds an essential layer of protection by helping identify and mitigate the growing threat of synthetic imagery or digitally-altered faces. The powerful combination of Liveness and Deepfake detection addresses sophisticated fraud techniques, offering added security for identity verification processes. For further insights, see our blog post on the critical role these technologies play in combating modern identity fraud.

Table of Contents

Terminology

Deepfake

A class of technologies that leverage AI to create hyper-realistic synthetic media, including manipulated imagery, video, and audio, in which an identity is presented in an inauthentic context. A more complete description of deepfake technologies can be found here.

Deepfake Detection

Technology that identifies whether an image or video has been synthetically created or manipulated, to falsely resemble an authentic representation of an individual.

Face Swaps

Face swaps refer to a type of deepfake manipulation where the face of one individual is replaced with the face of another within an image or video. This creates a realistic appearance of the swapped face, often retaining the expressions and movements of the original subject, making it appear as if the swapped face is naturally present in the scene.

Expression Swaps

Expression swaps are a type of deepfake manipulation where the facial expressions of an individual are altered to match those of another person. Unlike face swaps, which replace one face with another, expression swaps retain the original person’s face while modifying their expressions, creating a realistic imitation of emotions or reactions that the person did not originally make.

APCER

APCER (Attack Presentation Classification Error Rate) measures the rate at which fake or manipulated inputs are mistakenly classified as real by the system. A lower APCER is crucial for security, as it indicates the system's effectiveness in identifying and rejecting deepfake content or synthetic media.

BPCER

BPCER (Bona Fide Presentation Classification Error Rate) represents the rate at which real, authentic inputs are incorrectly classified as fake or manipulated by the system. A lower BPCER is desirable to ensure legitimate users or real content are accurately recognized without unnecessary rejections, thereby improving usability.

EER

Equal Error Rate is the point where Attack Presentation Classification Error and Bona Fide Presentation Classification Error Rates are equal, with lower EER values indicating better overall model performance.

Use Cases for Deepfake Detection

Deepfake Detection supports a broad spectrum of industries and applications where confirming the authenticity of an individual's face image is critical. Its ability to seamlessly identify whether the presented image or video is authentic and free from manipulation makes it an excellent solution for use cases such as:

01

Financial Services

Deepfake Detection is vital in financial services for preventing identity fraud, particularly in remote verification processes such as loan applications or digital onboarding. Paravision's technology helps banks and financial institutions ensure that customers are authentic, safeguarding against identity manipulation attempts.

02

Social Media Platforms

Social media companies face challenges in policing manipulated content. Paravision's Deepfake Detection helps platforms proactively identify and flag deepfake media in both user profiles and posted content, reducing the spread of misinformation and protecting users from misleading or harmful content.

03

Media and Entertainment

Deepfake technology has significant implications in media and entertainment, where unauthorized digital likenesses can be misused or lead to reputational harm. Paravision's solution can help content producers verify authenticity, maintain brand integrity, and ensure ethical use of likeness in media.

04

Digital Onboarding

In digital onboarding, for example with gig economy platforms or HRIS systems, verifying the authenticity of a new user's identity is crucial. Paravision Deepfake Detection helps ensure that synthetic or manipulated identities cannot be used to bypass verification processes, allowing businesses to onboard legitimate users confidently. By detecting deepfake manipulations, the technology helps prevent fraudulent attempts during account creation and enhances the security of remote onboarding workflows.

05

Passwordless Authentication

As more organizations adopt passwordless authentication for secure, frictionless access, verifying user authenticity becomes critical to prevent identity spoofing and unauthorized access. Paravision Deepfake Detection can be integrated into passwordless and / or MFA authentication workflows to ensure that only legitimate users are granted access. This additional layer of protection against deepfake-based impersonation improves both security and user experience, allowing for seamless yet secure access control.

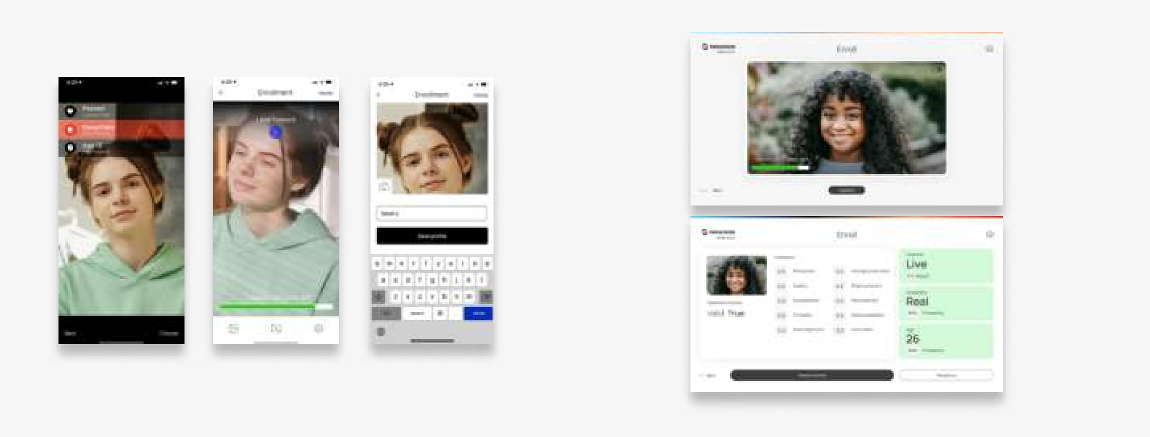
06

Government, Intelligence, and Law Enforcement

For government organizations, intelligence agencies, and law enforcement, the ability to verify the authenticity of images and videos is mission-critical. Deepfake Detection aids forensic investigations and authentication processes by helping ensure that digital evidence remains reliable and credible, and that deepfake content is correctly flagged and identified. Deepfake Detection helps safeguard the integrity of investigations and decision-making processes in an increasingly complex digital landscape.

How Paravision Deepfake Detection Works

Paravision Deepfake Detection includes the following main steps:



01

Image Capture

Paravision Deepfake Detection begins with capturing a facial image or video frame, often from a smartphone camera or a webcam. Paravision provides optional image validation metrics that evaluate quality, face size and position. These optional checks guide users in real time, prompting adjustments like “move camera closer” to optimize capture conditions.

02

Deepfake Detection and Scoring

Once an image is captured, it undergoes deepfake analysis on the partner server-side application. The AI model assesses the image, generating a score that reflects the likelihood the image is authentic. Images scoring below a chosen threshold are flagged as deepfakes, while those above are considered authentic. The system allows for tailored thresholds depending on the requirements of specific use cases.

Validity Metric	Description	Default Threshold
Face Size	Ensures the face is well-positioned, neither too close to the image’s edges nor excessively large or small.	Face size should be a minimum of 50 pixels wide.
Face Position	Measures the position of the face within the field of view.	The face must remain centered, with 100% of the face visible within the image’s field of view (FOV) to ensure accurate detection.
Face Quality & Acceptability	Face Quality and Acceptability assess a combination of face size and visibility of key facial features to determine the suitability of a face for detection and further processing.	The face should have a minimum face quality of 0.5, which is determined by factors like image quality, resolution, and facial occlusions; The face should have a minimum acceptability of 0.15, which is determined by multiple factors like focus and completeness of the face.

Considerations for Deepfake Detection

Ethical Development and Dataset Use

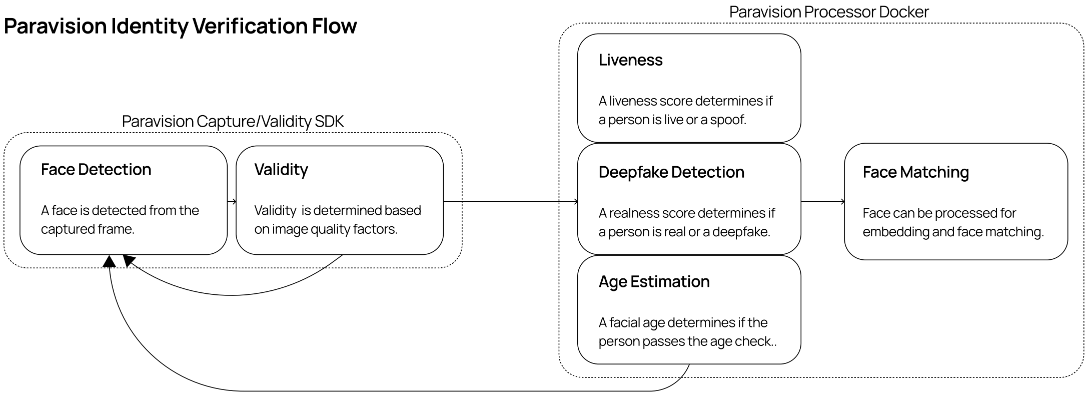
Paravision's Deepfake Detection has been developed and tested using extensive, ethically-sourced datasets that are properly consented and gathered with a commitment to user privacy. With a training dataset of over 1.4M and a benchmark dataset of over 500,000 images across multiple manipulation methods and demographic groups, Paravision ensures its models are trained with and tested against diverse manipulation types. This development and testing approach emphasizes both high accuracy and respect for privacy and data provenance.

! For more information about Paravision Deepfake Detection, including detailed data on accuracy and demographic performance, please reach out to info@paravision.ai or book a meeting at paravision.ai/contact.

Complimentary Technologies

Integrating Deepfake Detection with complimentary technologies such as Liveness Detection elevates identity assurance to a new level. By pairing these solutions, organizations can guard against both traditional spoofing attacks and the advanced risks posed by deepfake-generated content. Together, these technologies enable a comprehensive defense strategy, ensuring that every verified identity is both live and authentic. This dual-layer approach is not only a significant leap forward in fraud prevention but also a critical step toward building trust in digital identity systems. To dive deeper into the importance of this integration, explore our recent blog post.

Paravision Identity Verification Flow



Face Detection

A face is detected from the captured frame.

Validity

Validity is determined based on image quality factors.

Liveness

A liveness score determines if a person is live or a spoof.

Deepfake Detection

A realness score determines if a person is real or a deepfake.

Age Estimation

A facial age determines if the person passes the age check..

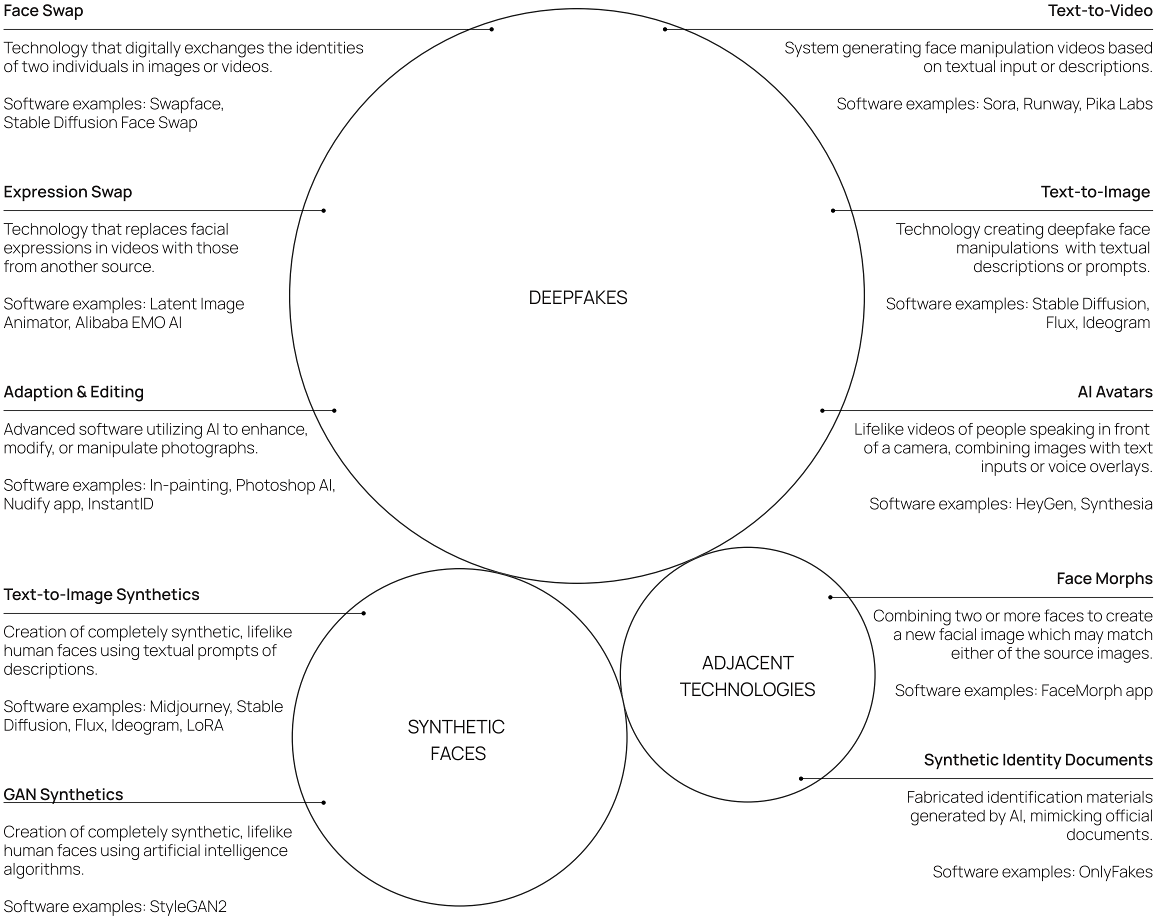
Face Matching

Face can be processed for embedding and face matching.

Deepfake Types

The current version of Paravision Deepfake Detection is specially designed to address Identity Swap and Expression Swap manipulations, common forms of deepfakes used to impersonate or alter an individual's facial expressions. The model is trained on a proprietary training dataset of over 1.1 million deepfake images generated with properly-consented imagery, representing a broad range of racial and gender diversity. By leveraging over 30 model types and eight leading deepfake generation approaches, Paravision's deepfake detection achieves exceptional accuracy across different manipulation methods, making it robust against new and emerging deepfake technologies.

Common types of face deepfakes and related digital face manipulations:



Face Swap

Technology that digitally exchanges the identities of two individuals in images or videos.

Software examples: Swapface, Stable Diffusion Face Swap

Expression Swap

Technology that replaces facial expressions in videos with those from another source.

Software examples: Latent Image Animator, Alibaba EMO AI

Adaption & Editing

Advanced software utilizing AI to enhance, modify, or manipulate photographs.

Software examples: In-painting, Photoshop AI, Nudify app, InstantID

Text-to-Video

System generating face manipulation videos based on textual input or descriptions.

Software examples: Sora, Runway, Pika Labs

Text-to-Image

Technology creating deepfake face manipulations with textual descriptions or prompts.

Software examples: Stable Diffusion, Flux, Ideogram

AI Avatars

Lifelike videos of people speaking in front of a camera, combining images with text inputs or voice overlays.

Software examples: HeyGen, Synthesia

Text-to-Image Synthetics

Creation of completely synthetic, lifelike human faces using textual prompts of descriptions.

Software examples: Midjourney, Stable Diffusion, Flux, Ideogram, LoRA

GAN Synthetics

Creation of completely synthetic, lifelike human faces using artificial intelligence algorithms.

Software examples: StyleGAN2

Face Morphs

Combining two or more faces to create a new facial image which may match either of the source images.

Software examples: FaceMorph app

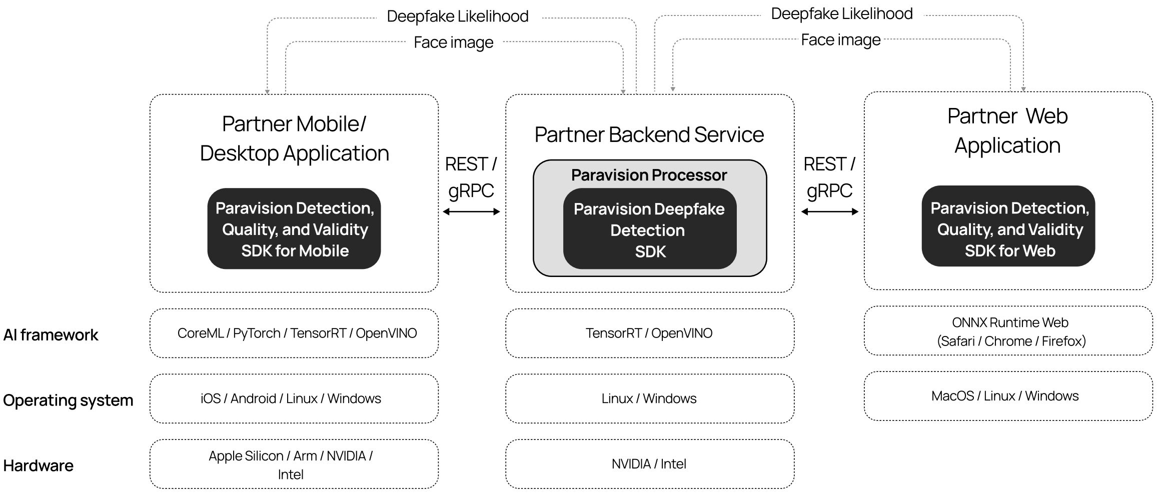
Synthetic Identity Documents

Fabricated identification materials generated by AI, mimicking official documents.

Software examples: OnlyFakes

System Architecture

Paravision Deepfake Detection is designed with a flexible, scalable architecture that supports a wide range of deployment environments to accommodate diverse partner needs. The solution can be deployed either as a standalone SDK or through a Docker-based backend container, allowing for flexible integration into existing infrastructures. Deepfake detection processing occurs server-side, while capture and validity checks can be performed on mobile, desktop, or web applications as well as on the server.


Deepfake Likelihood
Face image
Deepfake Likelihood
Face image
Partner Mobile/ Desktop Application
Paravision Detection, Quality, and Validity SDK for Mobile
REST / gRPC
Partner Backend Service
Paravision Processor
Paravision Deepfake Detection SDK
REST / gRPC
Partner Web Application
Paravision Detection, Quality, and Validity SDK for Web
AI framework
CoreML / PyTorch / TensorRT / OpenVINO
TensorRT / OpenVINO
ONNX Runtime Web (Safari / Chrome / Firefox)
Operating system
iOS / Android / Linux / Windows
Linux / Windows
MacOS / Linux / Windows
Hardware
Apple Silicon / Arm / NVIDIA / Intel
NVIDIA / Intel

Integration with Docker and SDK

Paravision Deepfake Detection's backend can be deployed using either a server SDK or Docker-based container, providing compatibility with cloud services like Google Cloud Platform, AWS, and Microsoft Azure, as well as private cloud or on-premises servers or workstations. The Docker container supports REST and gRPC APIs for efficient communication between partner applications and Paravision Deepfake Detection. The Paravision Processor Docker option helps enable streamlined updates, rapid deployment, and scalability, while the SDK facilitates a more controlled integration.

Frontend SDK for Validity Checks and Image Capture

The optional frontend SDK can be used on mobile, desktop, and web applications to perform image capture and validity checks before sending data to the backend for deepfake detection. This SDK can provide near real-time feedback to users to ensure optimal image quality, increasing the reliability and accuracy of deepfake detection.

This versatile architecture makes Paravision Deepfake Detection easily deployable in various environments, enhancing security through robust deepfake detection capabilities that adapt to diverse technical requirements.

Accuracy Analysis

Extensive Benchmarking for Real-World Reliability

Currently, there is no standardized industry benchmark for deepfake detection, such as those provided by NIST or iBeta for face recognition and liveness. This makes reliable benchmarking challenging. However, Paravision Deepfake Detection has been rigorously evaluated using an internal benchmark dataset comprising over 500,000 images, carefully selected to ensure a broad representation of demographic groups and manipulation types. This extensive dataset enables Paravision to assess the model's equitable performance across varied user groups, effectively mitigating biases. In terms of scale and diversity, Paravision's benchmark substantially surpasses industry references like the NIST FATE PAD benchmark, which includes only 51,000 images. Through this comprehensive testing, Paravision demonstrates the model's resilience and adaptability, establishing it as a best-in-class solution for deepfake detection in diverse real-world applications.

Accuracy Benchmarks

Paravision Deepfake Detection - Software Version 1.1.0

Paravision Deepfake Detection achieves leading accuracy, providing robust protection against a wide range of deepfake manipulations.

	EER (Equal Error Rate)	BPCER @ APCER = 0.1	BPCER @ APCER = 0.01	BPCER @ APCER = 0.001
Result	1.75%	0.09%	3.90%	18.45%

Model Improvements and Expanded Training Data

The newly released Paravision Deepfake Detection model (v1.1) represents a significant advancement over the previous version (v1.0) due to the addition of extensive proprietary training data, as shown in the latest model's performance metrics:

	EER	BPCER @ APCER = 0.1	BPCER @ APCER = 0.01	BPCER @ APCER = 0.001
New model 1.1	1.75%	0.09%	3.90%	18.45%
Old model 1.0	5.61%	2.05%	23.80%	47.67%
Change	-69%	-96%	-84%	-61%

! For more information about Paravision Deepfake Detection, including detailed data on accuracy and demographic performance, please reach out to info@paravision.ai or book a meeting at paravision.ai/contact.

In Conclusion

Paravision Deepfake Detection delivers a powerful solution to the growing threat of digital identity manipulation. With advanced model architecture and a proprietary dataset, it achieves exceptional accuracy, resilience, and adaptability. The latest iteration (v1.1) significantly reduces Equal Error Rate (EER) by 69% and enhances performance across a wide range of manipulation techniques. Extensive benchmarking against over 500,000 images ensures equitable performance across demographic groups, providing reliable protection in real-world scenarios.

Our ongoing investment continues to drive error rates lower while maintaining demographic equity. We are also expanding detection capabilities beyond identity and expression swaps to include emerging deepfake methods like GenAI and StyleGAN2, reinforcing our commitment to staying ahead of evolving threats.

In the absence of an industry-wide deepfake benchmark, Paravision sets new standards through rigorous internal testing. As deepfake threats continue to advance, our technology remains a critical asset for safeguarding digital identities and maintaining trust. Whether for digital onboarding, document verification, content moderation, or law enforcement, Paravision Deepfake Detection empowers industries with forward-looking, dependable protection—meeting today’s challenges and preparing for those of tomorrow.

PARAVISION

Trusted Identity AI

For more information or to schedule a demo, please contact us at:

info@paravision.ai