Deepfake and bot detection
How to stop the most sophisticated attackers

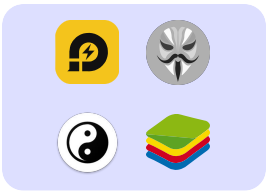
Persona combines injection attack detection with image and population-level analysis to capture deepfake selfies.

1

Submission

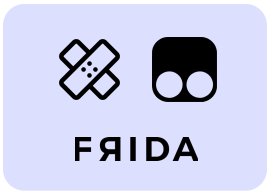
Injection attack detection

To submit deepfakes into a verification flow, bad actors often attempt to bypass or replace the native camera. We evaluate signals from multiple layers to detect injection attacks, including:

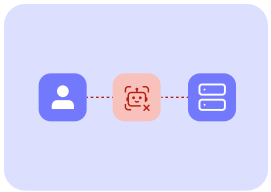


Device
We check whether a device is rooted, jailbroken, emulated, or simulated, conditions that allow bad actors to manipulate inputs.

Application
We look for signs of app tampering and repackaging using several methods, including native Google and Apple app tools.



Runtime
We inspect the runtime environment for signs of function hooking, where attackers redirect the native camera call to a video feed.



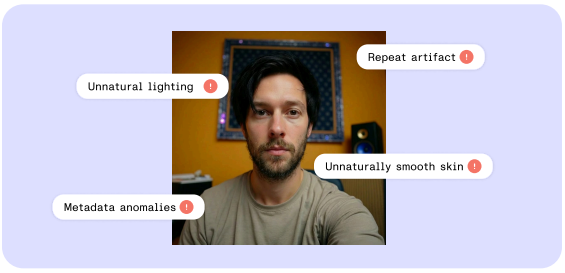
Network
We assess the risk of man-in-the-middle attacks, where bad actors intercept traffic to inject deepfake videos directly into our server.

2

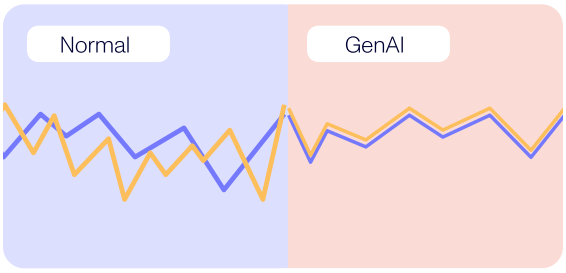
Verification

Image analysis

We evaluate images for GenAI indicators at two levels and continually update our approach to incorporate signals from the latest image generation models.



Visual tells
We continually train models on proprietary and public deepfake datasets to detect GenAI visual artifacts, such as blended items, malformed anatomy, and unnatural lighting.



Frequency domain analysis
We evaluate signal-to-noise (SNR) ratios for signs of AI image generation. Every real photo has pixel-level noise that comes from translating the real world to a digital image.

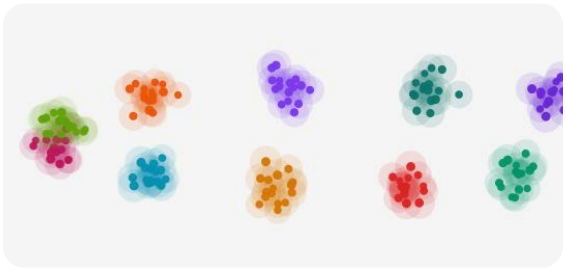
3

Population

Population analysis

In addition to focusing on individual submissions, we develop and improve detection models based on learnings from working in fraud prevention across industries and regions.

Known deepfake permutations
We consider similarities between images. Fraudsters often recycle templates and prompts to create new deepfakes, and image analysis allows us to detect these attempts at scale.



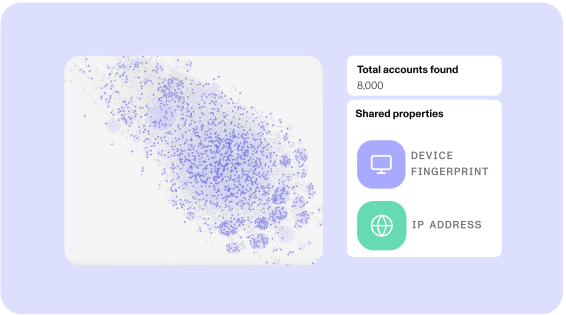
Cluster evaluations
We evaluate clusters of selfies to find which attributes deepfakes commonly share, detect organized fraud rings, and expand our understanding of anomalous factors.

4

Population

Graph

Graph is Persona's real-time link analysis investigation and visualization tool. Use signals from Persona, including image similarities, or create your own custom nodes to uncover bad actors. You can also add Graph results to verification flows and automate decisions.



Comprehensive risk signal capture

Behavioral signals
Find suspicious behavior, such as shortcut usage, that may indicate scripted flows.

Network signals
Uncover risky network traffic indicators, such as whether a VPN may be in use.

Device signals
Find signs of malicious activity using device and app runtime environments.

Learn more about fraud detection

withpersona.com/product/verifications/selfie →

Click here

Trusted by our partners

Etsy lyft Square OpenAI Brex