iiDENTIFii

iproov

How Can Financial Institutions Safeguard Against Deepfakes: The New Frontier of Online Crime?

Content

Executive Summary

Adopted by industry leaders like UBS, biometric face verification has a high market profile. Yet as the technology becomes more ubiquitous, criminals are developing new ways to circumvent these cybersecurity systems and commit fraud, launder money, or engage in other illicit activities for financial gain.

This report explores the rising threat of digitally injected synthetic media to financial institutions globally and provides recommendations on what defenses can be taken to secure high-risk use cases.

Key findings include:

1 Digital injection attacks are challenging to detect and highly scalable, making them appealing to fraudsters. There is currently no industry-wide accredited testing for digital injection attack detection like there is for presentation attack detection (PAD).

2 Digital injection attacks are rapidly shared and tested from numerous locations worldwide, whether by the same criminal organization or – according to Europol – via a Crime-as-a-Service economy.

3 All liveness detection technologies are not created equal. While many solutions offer some level of presentation attack detection, most cannot detect digitally injected deepfake attacks.

4 Our recent Biometric Threat Intelligence Report report highlights that digital injection attacks occurred five times more frequently than persistent presentation attacks across web in the second half of 2022.

5 In the same report, the new occurrence of face swaps (a form of deepfake) emerged and grew exponentially in 2022, with a massive 295% increase in the space of a few months.

6 Although 57% of global consumers believe they can successfully spot a deepfake, research shows that only 24% of people can. This statistic proves that video calls are not a reliable means of identity verification.

Introduction: The Rise of Online Face Verification in Financial Services

In 1834, two thieves hacked the French telegraph system to steal information about financial markets – effectively carrying out the first known cyberattack.

A lot's changed in the centuries since, but there has been one constant. As new digital channels and cybersecurity measures emerge, criminals develop new methods and tools to breach them.

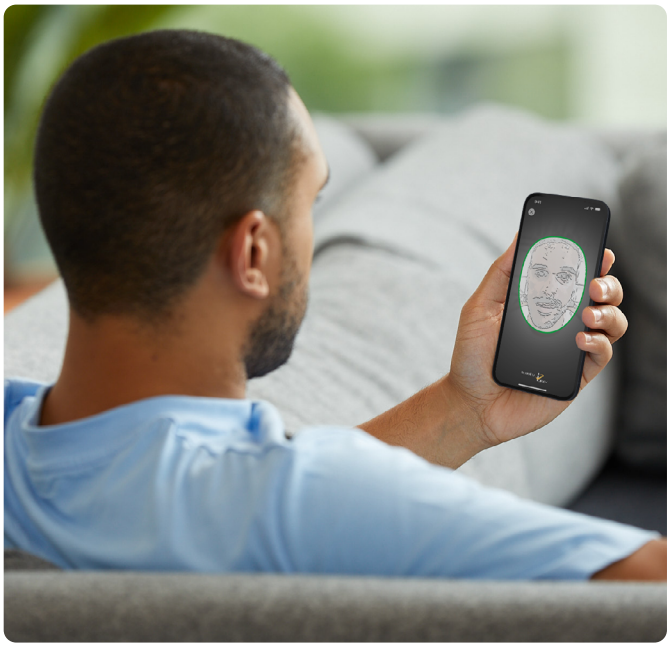
Accelerated by the global pandemic, global online banking participation has reached new heights with Juniper Research declaring that the number of digital banking users is expected to reach 3.6 billion worldwide by 2024, a 54% increase from 2020. Much of this recent growth has come from online banking adoption from previously unbanked people. In 2022, 71% of people had access to a bank account, up from 42% a decade before.

The move from in-person to remote financial services poses a challenge. When a remote user applies for an account, product, or service, how can an institution verify that they are the real owner of a genuine identity? What's more, how can they ensure that an existing remote customer is the same person each time they return – and not an imposter or a synthetic identity?

With the pivot to remote onboarding, financial crime, and cybercrime have become more inextricably linked than ever before. According to Interpol, not only are financial and cybercrimes the world's leading crime threats, but they are also projected to increase most in the future.

Traditional authentication solutions have become commoditized as threat actors innovate and advance their abilities. Passwords can be shared, stolen, or compromised – and have been leaked in countless widescale data breaches. Meanwhile, one-time passcode (OTP) authentication is impaired because devices can be lost or stolen. Plus, SIM swaps and the SS7 flaw are just some of the, now well-established, methods threat actors use to circumvent these technologies.

These solutions also lack usability. It's unfeasible to expect users to remember complex and unique passwords for each application they use. Password resets are an inevitable, arduous process. OTPs meanwhile assume that people have two separate devices with them at any given time.



As these approaches have failed to reliably verify and authenticate remote individuals or provide a positive user experience, financial institutions have turned to biometric face verification. The security of biometrics relies not on the fact that faces are secret – they're not – but that they're unique, non-sharable, cannot be stolen, and never need to be reset. Face verification can bind digital identities to real-world users by matching a selfie image with a government-issued ID.

Face verification resolves the usability issues of passwords and OTPs – there's nothing to remember or forget, and a person carries their face wherever they go. The value of this technology has been recognized by consumers. According to iProov's survey, 64% of global consumers who use mobile banking either use face verification to access their accounts already or would do so if they could.

As we will explore, some biometric face verification technologies can provide a strong level of assurance that an individual is who they claim and are a 'live' human being. Yet, as face verification providers and adoption increase, criminals are finding new ways to circumvent these weaker systems.

Biometric attacks comprise presentation attacks (pg8) and digital injection attacks (pg9). While presentation attacks and presentation attack detection (PAD) are understood, many biometric systems are not equipped to defend against injection attacks.

The digital injection of synthetic imagery, in particular deepfakes, is one of the latest and fastest growing threat vectors, and not all liveness technologies are resilient to it.

Combining the iProov Security Operations Center (iSOC) research into the biometric threat landscape, customer experience, industry expertise, and consumer survey data, this report explores the fastest growing and scalable threat of injected synthetic imagery to banks worldwide. It seeks to illustrate:

- How novel attacks are targeting biometric face verification systems, exploiting the vulnerabilities within different technologies.
- The efficacy of methods, such as video calls, as a means to verify customer identity.
- Defenses banks can deploy against the deepfake threat.

1 iSOC utilizes state-of-the-art machine-learning computer vision systems in conjunction with complementary, multimodal approaches to monitor traffic in real-time to detect attack patterns across multiple geographies, devices, and platforms.

What are Deepfakes?

A deepfake is a video, visual, or audio recording that has been distorted, manipulated, or synthetically created using deep learning techniques to present an individual, or a hybrid of several people, saying or doing something that they did not say or do.

Not all deepfakes are created for nefarious purposes. Many are used for entertainment purposes. Yet, criminals have also adopted them to impersonate public figures online, spoof video conferencing calls and interviews, and gain unauthorized access to valuable online services.

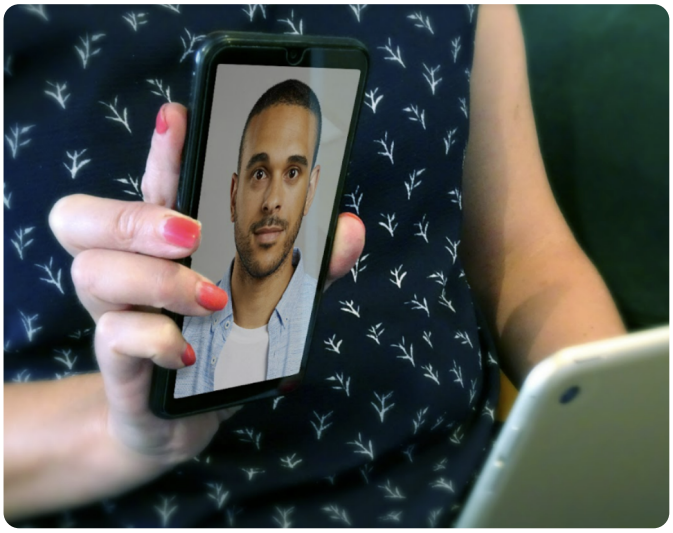
Figure 1: Example of how a deepfake is created

The New Risks: How Is Deepfake Technology Being Used to Exploit Biometric Face Verification?

As face verification gains traction and becomes more prevalent, threat actors are developing evermore sophisticated ways to circumvent these systems to commit fraud. Not all attacks carry the same threat level – some are more challenging to detect and scalable than others.

Presentation Attacks

A presentation attack is an act of holding up an artifact to the user-facing camera in an attempt to spoof the face authentication sequence. These artifacts can take the form of static images, videos (e.g. replays of previous authentication attempts), and high-quality masks. A deepfake video played on a device and held in front of the camera is another example of a presentation attack.



Presented deepfakes can be realistic and convincing. A non-reflective screen on a retina display makes images appear extremely crisp so that pixels are not visible to the naked eye or at viewing distance. To achieve presentation attack detection, and to detect presented deepfakes, biometric face verification applications must incorporate liveness detection, which we will explore later.

Many liveness solutions have got to the point where they can largely detect presentation attacks. But presentation attacks are not the primary threat to biometric systems, and achieving PAD on its own can lead to a false sense of security. As we will see, injection attacks have overtaken presentation attacks in sophistication and scale.

Digital Injection Attacks

The process of creating a deepfake and presenting it to a camera can be effective, but it is limited in scope: realistically, the criminal can only do this one at a time. Digital injection attacks, on the other hand, are far more scalable.

Digitally injected imagery enables criminals to inject deepfakes, either of synthetic or genuine individuals, directly into the data stream or authentication process.

Digital injection attacks are the most dangerous form of threat because they are more difficult to detect than presentation attacks and can be replicated quickly. They carry none of the clues that artifacts do when they are presented to the camera, making the more sophisticated attacks challenging for systems to distinguish and near impossible for humans.

iProov's recent threat intelligence report revealed that injection attacks were five times as frequent as persistent presentation attacks on the web throughout 2022.

The report also revealed these attacks being launched at scale. In one example, an indiscriminate attempt to bypass an organization's security systems, some 200-300 attacks were launched globally from the same location within a 24hr period.

The methods used to create and launch digital injection attacks are diverse:

1. Software-based camera is used to bypass the camera on their own device and injects a deepfake of a legitimate user. The bank's application on the criminal's device thinks it is receiving legitimate footage from the device camera.

2. Manipulation of the application on a user's device. This can be done by using malware or if the unsuspecting user has downloaded a genuine-looking copy of the banking app. The bank thinks it is receiving footage from the user's device but instead, it is receiving synthetic imagery from the app.

3. Injecting a deepfake into the data stream between the device and the organization's server. This is an example of a man-in-the-middle attack. It requires the criminal to understand the communication channel between the device and the organization.

4. Once the criminal understands the exchanges between the device and the organization's server, they can create software to pose as a legitimate device that injects and sends deepfakes to the organization. The same software can be run thousands of times in parallel to make it look like this synthetic imagery is coming from legitimate devices.

5. An emulator is used to mimic a user's device, such as a mobile phone. Emulators can enable threat actors to launch attacks across mobile web platforms, as well as native Android and iOS, which are traditionally seen as more secure than desktop web. The iProov Threat Intelligence Report 2023 witnessed these attacks increase 149% through 2022.

Deepfakes and Synthetic Identity Fraud

As the adoption of online banking has grown around the world, so too has identity-related fraud. While traditional identity theft is on the rise, reportedly accounting for $52 billion in losses in the US alone and affecting a staggering 42 million American adults in 2021, banks also face a more novel threat in the form of synthetic identity fraud (SIF).

Unlike identity theft – whereby a threat actor uses a real person's personal details to commit fraud – SIF is the act of creating a "person" who doesn't exist by using a mix of stolen, fictitious, or manipulated personally identifiable information (PII). This could include a person's name, address, and social security number.

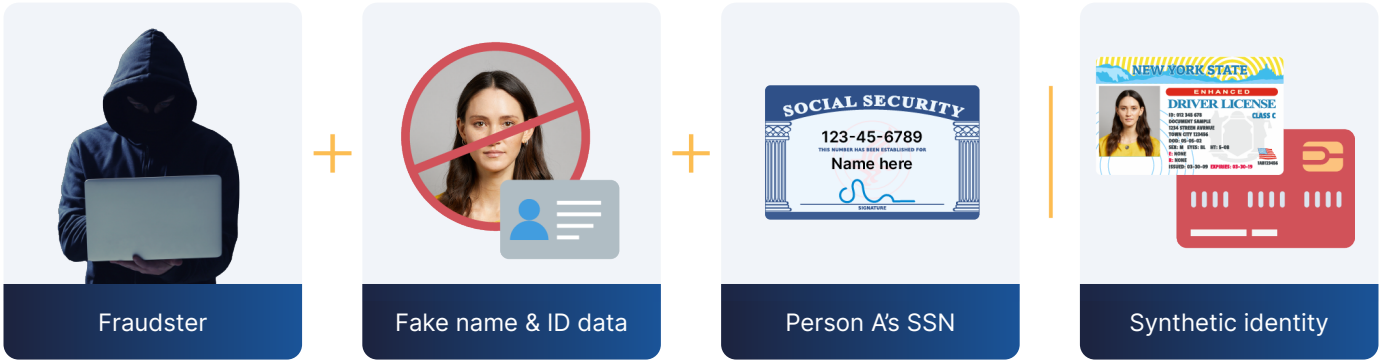
Synthetic identities can be used to establish accounts that behave like legitimate accounts and may not be flagged as suspicious using conventional fraud detection models. Plus, losses due to synthetic attacks are often written off as credit losses. SIF is on the rise: the Aite Group estimates that it will account for $2.42 billion in fraudulent funds obtained in the U.S. in 2023.

Synthetic identity fraud can work in conjunction with deepfakes to exploit a bank's remote onboarding processes. For example, a criminal creates an identity document of a synthetic identity. They then create a deepfake that matches the photo on the ID document and use it to pass through a bank's remote face verification processes.

Once the deepfake has been enrolled into the service, the attacker is free to return using the same biometric credentials. As it is not the same as account takeover fraud, there is no genuine account holder to alert the bank; the synthetic person can continue to use it for criminal activity, potentially for years, without detection.

Figure 2: Synthetic Identity Fraud

Fraudster combines fake (or real) info to establish a credit record under the new synthetic identity.



How do Deepfake Attacks Scale?

If equipped with the ability to digitally inject synthetic imagery, criminals can build highly-automated attack machines, launching thousands of attacks, across multiple geographies, cheaply and in a short space of time.

This technology is not the only factor that contributes to the scalability of synthetic media attacks. As we discuss next, the availability of technology in a thriving crime-as-a-service economy accelerates and widens the deepfake threat.

Crime as a Service (CaaS)

It's a mistake to think that criminals act in isolation. Rather, there's a sophisticated cybercriminal network with extensive communication channels. Once an attacker infiltrates a bank's system, they often sell or share the profitable tools, techniques, and information (such as stolen identities) over the dark web.

In 2021, Europol warned that the CaaS economy continues to proliferate, stating that the 'availability of exploit kits and other services not only serves criminals with low technical skills but also makes the operations of mature and organized threat actors more efficient.' More recently, Europol reported that high demand has even led to a deepfake-as-a-service market, whereby criminal organizations create and deliver tailored deepfakes upon request. In one example, a threat actor was willing to pay $16,000 for the service.

iProov has witnessed similar indications of low-skilled criminals gaining the ability to create and launch advanced synthetic imagery attacks. In 2022, we saw the emergence and rapid growth of novel video face swaps. Face swaps are a form of synthetic imagery where the threat actor morphs more than one face to create a new fake 3D video output.

Through 2022, iProov saw sophisticated face swap attacks increase 295% from H1 to H2. This growth rate indicates that low-skilled criminals are gaining access to the resources necessary to launch sophisticated attacks.

CaaS and the availability of online tools accelerate the evolution of the threat landscape, enabling criminals to launch advanced attacks faster and at a larger scale. If attacks succeed, they rapidly escalate in volume and frequency, amplifying the risk of serious damage.

How Effective are Video Calls at Preventing Deepfakes?

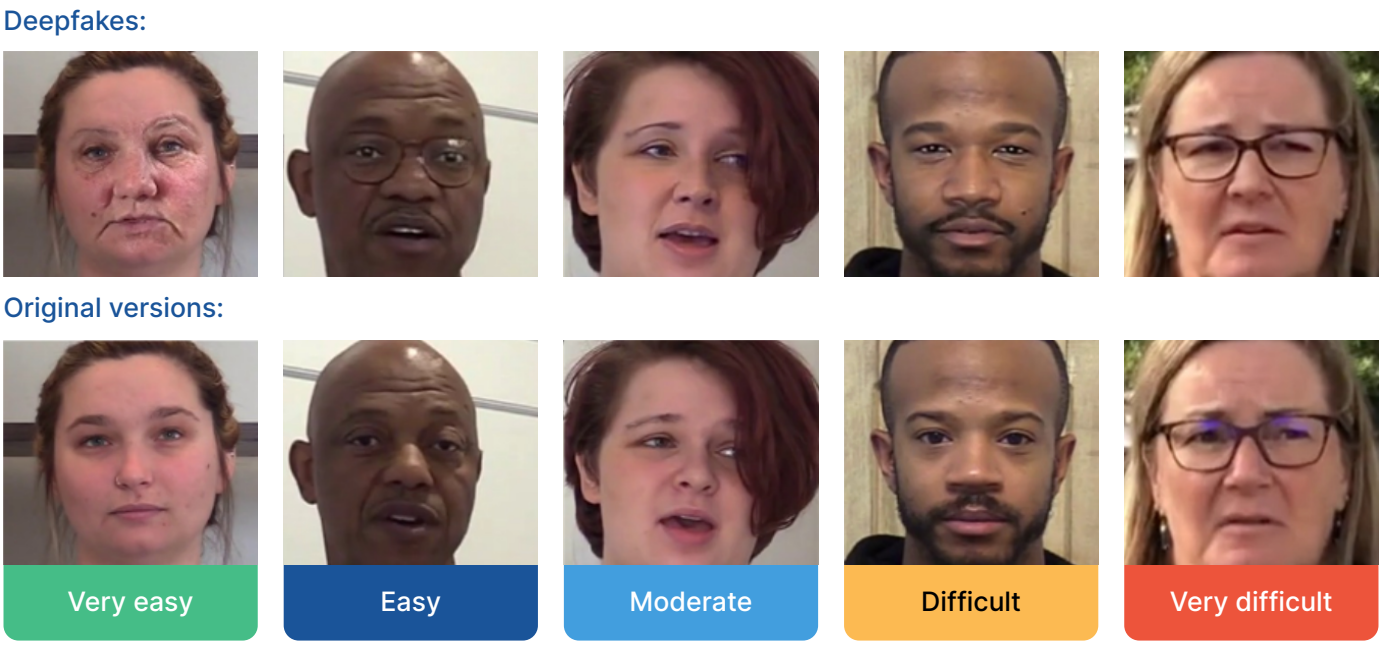
As the threat landscape evolves, how can banks be sure that a remote individual is who they say they are? One approach is to verify customer identity and carry out KYC checks via a face-to-face video call between a trained member of staff and the user. Naturally, this relies on the staff member's ability to distinguish between a real person and synthetic imagery.

Yet, how effective are humans at detecting deepfakes? We are born with an innate ability to recognize human faces. Surely, humans can tell between a real face and a deepfake.

Not the case. As deepfakes become more sophisticated, we can no longer rely on human ability to detect them. Yet, worryingly, most people have a false sense of confidence in their ability to detect a deepfake.

A study conducted by the IDIAP Research Institute, a facility built to examine artificial and cognitive intelligence, showed that human beings are wholly ineffective at detecting deepfakes. However, in a recent survey conducted by iProov, 57% of consumers were confident that they could tell the difference between a real video and a deepfake. This confidence is growing – in 2019, this figure was only 37%.

Figure 4: Cropped faces from different categories of deepfake videos of Facebook database (*top row*) and the corresponding original verisons (*bottom row*)



In contrast, the IDIAP research revealed that only 24% of their participants successfully detected a 'well-made' deepfake when shown progressively more convincing deepfakes interspersed with real videos and asked, 'is the face of the person in the video real or fake?'.

The researchers caveated the study, saying the number of people successfully detecting the fakes in real-world conditions would be 'significantly lower' as the laboratory conditions may have skewed the results.

Similarly, in a high-profile incident, the Mayor of Berlin and several other European public figures were duped into holding a video call with a deepfake of Vitali Klitschko, the Mayor of Kyiv and former professional boxer. It took the Mayor of Berlin and his aides over 15 minutes to realize they were talking to a fake.

A human's clear inability to distinguish between a real human being and synthetic imagery has raised the question as to the efficacy of video conferencing as a reliable means to verify the identities of unknown users.

How Can Banks Defend Against Deepfakes and Digital Injection Attacks?

Banks should be concerned about the threat of digitally injected synthetic media, but there are technology and processes they can employ to safeguard against them.

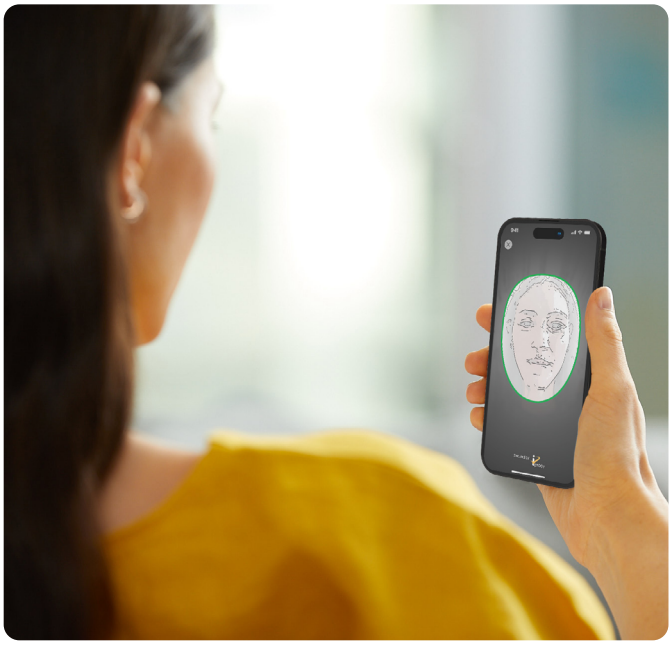
Most face biometric technology incorporates some form of liveness detection to verify and authenticate customers. Liveness detection uses biometric technology to determine whether the individual presenting is a real human being and not a presented artifact. Therefore, this technology can detect a deepfake if it were to be played on a device and presented to the camera.

While many liveness detection technologies can offer presentation attack detection (PAD), many solutions are unable to detect digital injection attacks.

There are several methods to defend against digital injection attacks. Some methods rely on preventing them from happening in the first place. In having totally secure hardware and an entirely secure communications channel from a trusted device, fake imagery cannot be injected. This is the approach adopted by the FIDO standards. However, this method is unrealistic and inherently not inclusive, as not every customer will have access to this hardware.

For general-purpose devices, such as smartphones or laptops, one approach is to instruct the user to do something different every time they authenticate, known as active authentication. The user performs actions, such as turning their head or reading out a sequence of characters to verify they are 'live' at the time of authentication. Yet, deepfakes can be coded to perform these actions just as well. Plus, it raises concerns regarding accessibility and inclusivity for those with physical or cognitive disabilities. Learn more about this here.

The most effective method to defend against digital injection attacks and still have high levels of customer usability is through passive authentication, such as one-time biometrics, where the technology does the hard work and is not dependent on an individual having to follow complex instructions, enhancing both security, user experience, and inclusivity.



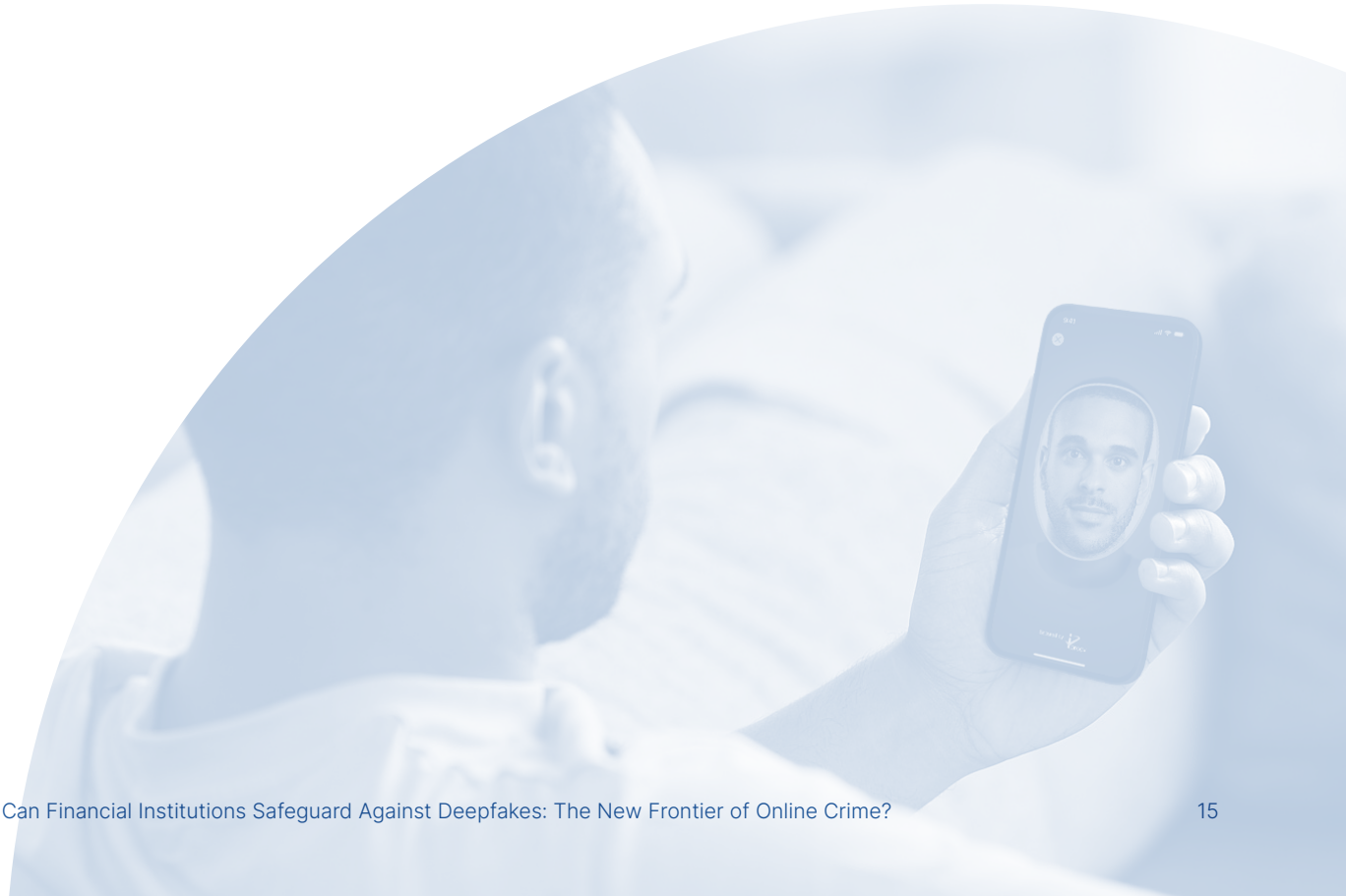
One-Time Biometrics

For high-risk remote use cases, such as opening a new account or transferring a large sum of money, most liveness detection technology does not provide an adequate level of assurance.

One-time biometrics assure both liveness and that a user is verifying in real-time, which is essential in a bank's defense strategy against deepfakes

A one-time biometric is a method that takes place in real-time to assure that a user is 'live' and present at the time of authentication. It's never repeated in a user's lifetime and has a limited duration. The unpredictability of one-time biometrics makes it extremely challenging for threat actors to replicate or reverse-engineer the authentication process. It also means it is worthless if stolen, mitigating the risk of a data breach.

One way to achieve this with a standard device is to use the screen to project controlled illumination onto the user's face. The color sequence of the illumination changes each time a user authenticates. The feedback from the illumination creates a one-time biometric. Once used, it can't be replayed by a person attempting to use a previous authentication to spoof the system.



Summary

Biometric face verification remains the most secure and convenient way to verify unknown customer identities at onboarding, grant returning users access to accounts, and authenticate transactions. However, digitally injected synthetic media, such as deepfakes, is a present and growing attack vector that is defrauding banks and their customers.

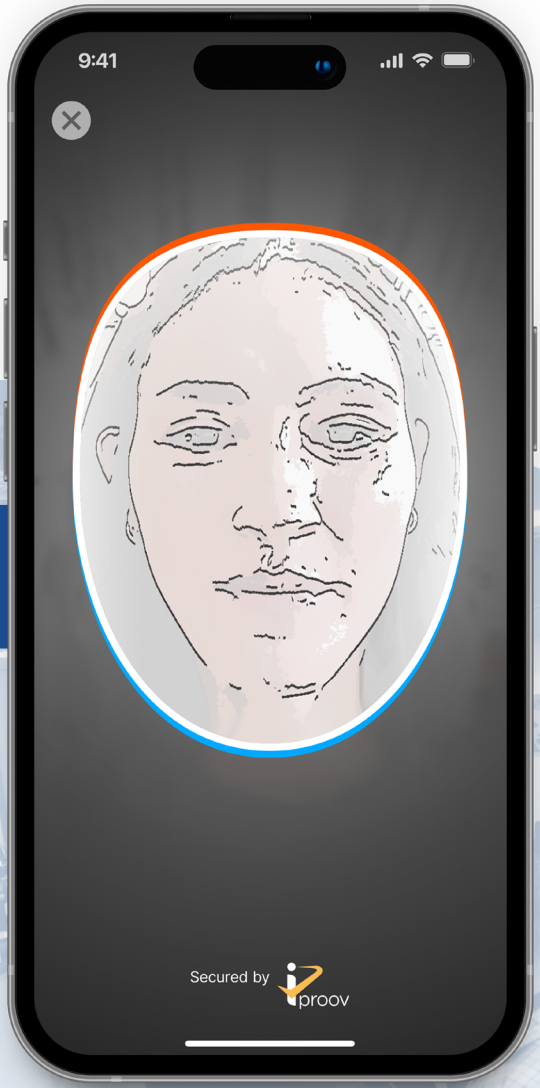
Technology that enables bad actors to move beyond presentation attacks, circumventing face verification technology by digitally injecting synthetic media into the authentication process, is widely available and used globally. Moreover, the CaaS economy and iProov research into the threat landscape show that this threat is growing and scalable.

As research shows, face-to-face video calls fall short as a defense against synthetic imagery, as the human eye can be spoofed. Instead, specialized technology is required. The deployment of passive one-time biometrics during verification and authentication sequences has proven to be the most effective, usable, and inclusive way to safeguard against the threat of digital injection attacks.

About iProov

iProov is used by leading organizations worldwide to reduce the risk of identity fraud. Financial services clients include UBS, ING, Rabobank, and Knab. Government clients include the U.S. Department of Homeland Security, the UK Home Office, the National Health Service, and GovTech Singapore.

Genuine Presence Assurance®, iProov's flagship technology, is the only way to detect whether a user is the right person (not an impostor), a real person (not a presented artifact), and the genuinely present at the point of authentication (not digitally injected synthetic media, such as a deepfake).



Example of using Genuine Presence Assurance®

Methodology

The iProov Threat Report 2023, frequently mentioned throughout this report, used data and insights from the iProov Security Operations Center (iSOC). iSOC uses technology, people, and processes to monitor traffic in real-time to detect attack patterns across multiple geographies, devices, and platforms.

This report is supported by research carried out by an independent agency on behalf of iProov in April-May 2022. Eight countries were included in the research (the US, Canada, Mexico, Germany, Italy, Spain, the UK, and Australia) with 2,000 consumers surveyed in each country.

References

Executive Summary

'UBS Partners with iProov for Automated Online Identity Verification', iProov, 2022

'Facing Reality? Law Enforcement and The Challenge of Deepfakes', Europol, 2022

The iProov Threat Intellgence Report 2023, iProov, 2023

Introduction

'The French Telegraph System Was Hacked in 1834. What Does the Incident Teach Us About Modern-day Network Security', Slate, 2018

'Digital Banking: Market Forecasts For Banking-As-A-Service, Open Banking & Digital Transformation 2021-2026', Juniper Research, 2022

'COVID-19 Boosted the Adoption of Digital Financial Services', The World Bank Group, 2022

'INTERPOL Global Crime Trend Report 2022', INTERPOL, 2022

'RockYou2021: Largest Password Compilation of All Time Leaked Online With 8.4 Billion Entries', Cybernews, 2022

'One-Time Passcode (OTP) Authentication: What Are the Risks?', iProov, 2022

'Digital Identity Report: What Consumers Want and How Governments, Banks and Other Enterprises can Deliver', iProov, 2022

Deepfakes and Synthetic Identity Fraud

'2022 Identity Fraud Study: The Virtual Background', Javelin Research, 2022

'Synthetic Identity Fraud: Diabolical Charge-Offs on the Rise', Aite Group, 2021

Crime-as-a-Service (CaaS)

'Internet Organised Crime Threat Assessment 2021', Europol, 2021

'Facing Reality? Law Enforcement and The Challenge of Deepfakes', Europol, 2022

How Effective are Video Calls at Preventing Deepfakes?

'Staring Us in the Face? An Embodied Theory of Innate Face Preference', National Library of Medicine, 2014

'Subjective and Objective Evaluation of Deepfake Videos', IDIAP Research Institute, 2021

'European Politicians Duped Into Deepfake Video Calls With Mayor of Kyiv', The Guardian, 2022

How Can Banks Defend Against Deepfakes and Digital Injection Attacks?

'COVID Drives Explosive Market Growth of Face Verification and Liveness Detection for Remote Digital Onboarding', FinTech News, 2021

'Breaking Down Active vs Passive Liveness', iProov, 2022

For more information on how to assure the genuine presence of the **right** person, **real** person, authenticating **right now** contact us at **contact@iproov.com**

iproov.com

iiDENTIFii is a premier partner of iProov in Africa. iiDENTIFii protects organisations and individuals against identity fraud. We are a proven enterprise-grade facial biometric authentication and automated onboarding platform. We verify an individual's digital identity and authenticate their physical presence in an online world. Our technology meets governance and legislative requirements, and plugs seamlessly into existing infrastructures, including mobile and web-based platforms.

https://iidentifii.com/
info@iidentifii.com

IPGloReen-USLE02/23