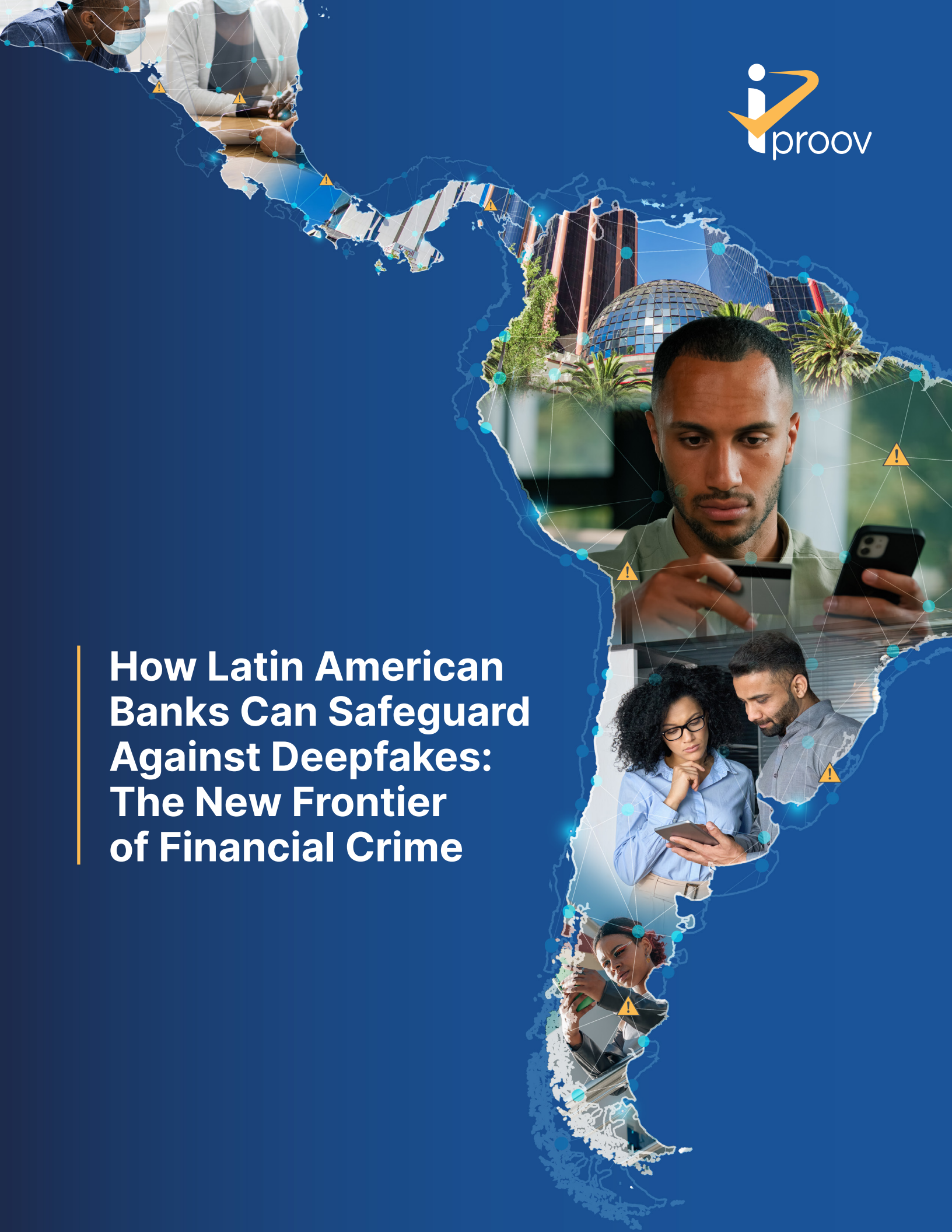


How Latin American Banks Can Safeguard Against Deepfakes: The New Frontier of Financial Crime

Content

Executive Summary

As the use of face biometrics by Latin American banks has grown, so too have criminal activities attempting to circumvent security systems and commit online fraud. The digital injection of synthetic media, such as deepfakes, is one of the latest attack methods criminals are using to exploit face verification and gain unauthorized access to banking applications.

This report explores the rising threat of digitally injected synthetic media to Latin American banks and provides recommendations on what defenses can be taken.

Key findings from the report include:

1 Digital injection attacks are difficult to detect and highly scalable, making them appealing to fraudsters. They are being rapidly shared and tested from numerous locations worldwide, whether by the same criminal organization or via a Crime-as-a-Service economy.

2 All liveness detection technologies are not created equal. In fact, many liveness detection solutions cannot defend against digitally injected deepfake attacks. One-time biometrics that assure both liveness and that the user is a real-live person, verifying in real-time, is essential in a bank's defense strategy against deepfakes.

3 Humans are ineffective at detecting deepfakes. Despite the fact that 57% of global consumers believe they can successfully spot a deepfake, research shows that only 24% can.

Introduction

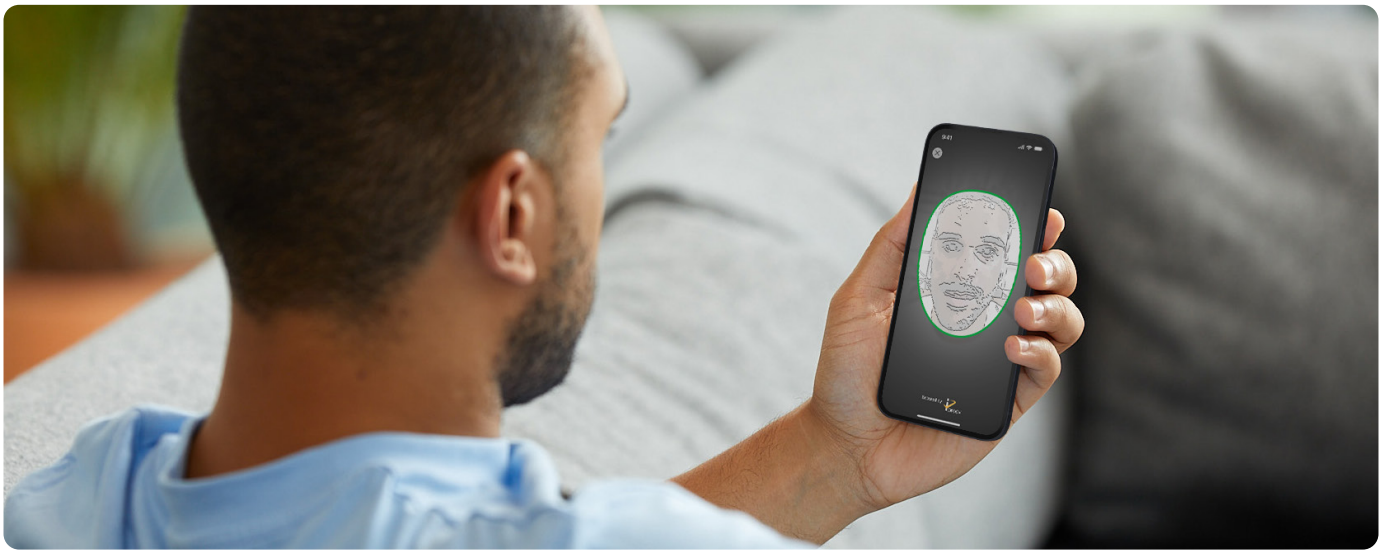
Online banking in Latin America has been a huge growth area in recent years – in large part driven by the global pandemic which introduced tens of millions of traditionally unbanked Latin Americans to online financial services. This growth has caught the attention of McKinsey, which has recently identified the region as a 'global growth leader in banking'.

However, the boom has not only led to revenue growth for the banking industry, but has also helped increase the revenue of fraudsters. It is currently estimated that 20% of the region's online revenue is lost to fraud. To combat this rise in fraudulent transactions, many Latin American banks have adopted or are in the process of adopting biometric verification as a secure method to facilitate Know Your Customer (KYC) practices, identify unknown customers at onboarding, and authenticate transactions – all remotely.

As banks continue to adopt face biometrics to thwart fraudulent digital activities, criminals are also innovating and developing new ways to circumvent security systems and commit online fraud.

This report explores the rapidly evolving and scalable digital injection threat to Latin American banks, in particular the use of synthetic imagery such as deepfakes. Using iProov Security Operations Center's (iSOC) research into the threat landscape, customer experience, industry knowledge, and consumer survey data, it seeks to illustrate:

- How criminals use deepfakes and digital injection attacks to defraud banks
- How these attacks scale versus traditional attack methodologies
- Human ability to detect deepfakes
- How banks can protect themselves from this threat



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 saying or doing something that they did not say or do.

Not all deepfakes are created for nefarious purposes. They have been used for fun and entertainment. Yet, they have also been adopted by criminals to impersonate public figures online, spoof video conferencing calls and interviews, and gain unauthorized access to online services. In particular, we will discuss how criminals are using this technology to exploit face verification.

Figure 1: Example of how a deepfake is created

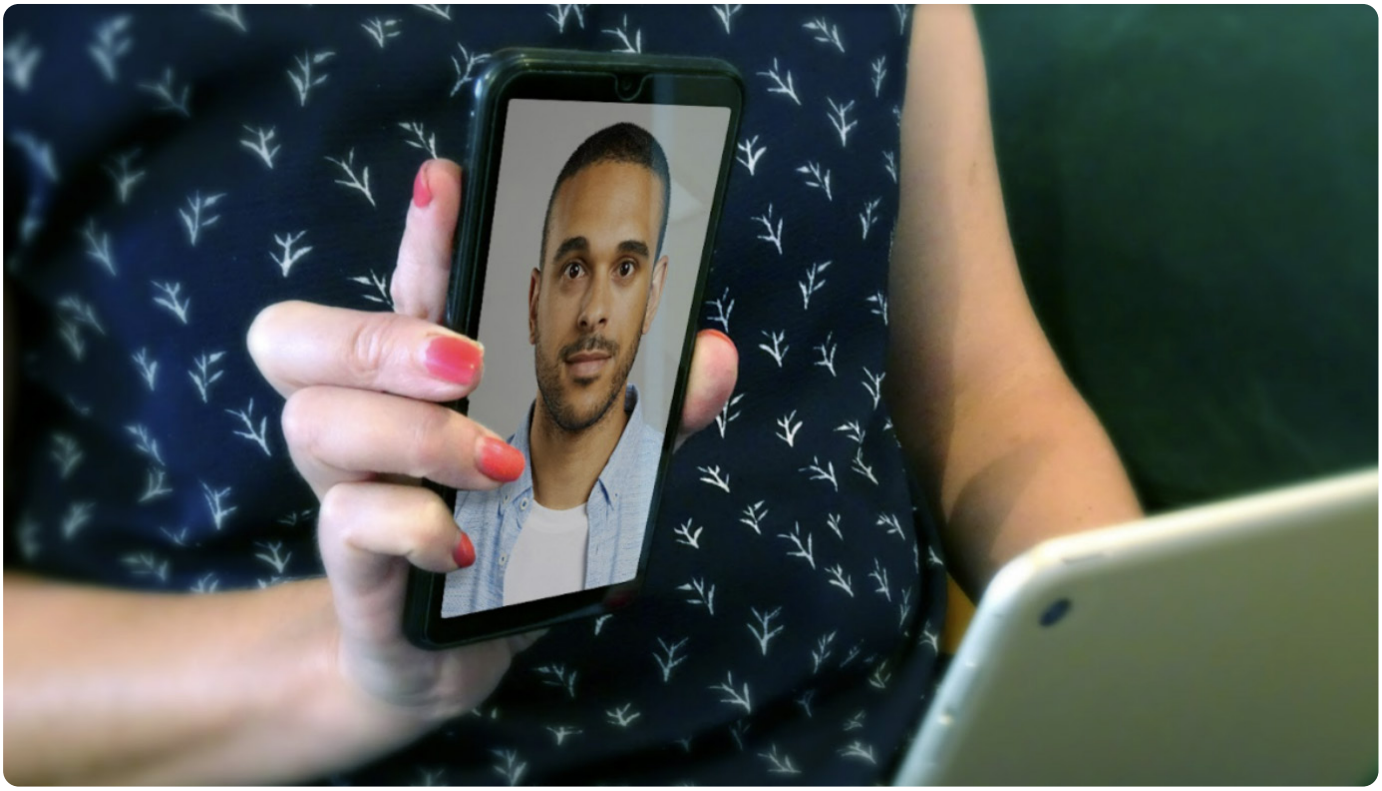
How Is Deepfake Technology Being Used to Exploit Face Verification?

There are several ways in which criminals use deepfakes to attack face verification systems. Not all attacks carry the same threat level – some are more challenging to detect and scalable than others. We will explore them here.

Presentation Attacks

A presentation attack is an act of holding up an artifact to the user-facing camera to impersonate a legitimate bank customer and spoof the face authentication sequence. These artifacts can take the form of static images, videos (e.g. replays of previous authentication attempts), and highly-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 defend against presentation attacks, including presented deepfakes, biometric face verification applications must incorporate liveness detection, which we will explore later.



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.

These types of attacks can be run from an attacker's computer. Or they can be done using a thousand cloned devices in a data center operated by a criminal network.

There are several ways criminals can digitally inject deepfakes to bypass a bank's authentication processes and gain unauthorized access:

1. The criminal uses a software-based camera 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. The criminal **manipulates the application on a user's device**. This could 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. The criminal injects a deepfake into the data stream between the device and the organization's server. This is what's known as 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.

Deepfakes and Synthetic Identity Fraud

As the adoption of online banking has grown in Latin America, so too has identity fraud. Brazil holds first place in the region for identity theft-related crimes, where it is estimated that 60bn reais (approximately $12bn) is lost yearly to this form of cybercrime.

The ability to digitally inject synthetic media into an authentication process has enabled criminals to amplify traditional identity fraud and enhance synthetic identity fraud.

Synthetic identity fraud (SIF) is the act of creating a "person" who doesn't exist by using stolen, fictitious, or manipulated personally identifiable information (PII). This could include a person's name, address, and social security number, for instance.

SIF can be used to exploit a bank's onboarding processes. The bank is presented with the synthetic identity and unknowingly onboards the "person" under the assumption they are genuine. These "people" can then max out credit cards, use the account to launder money, or commit other crimes.

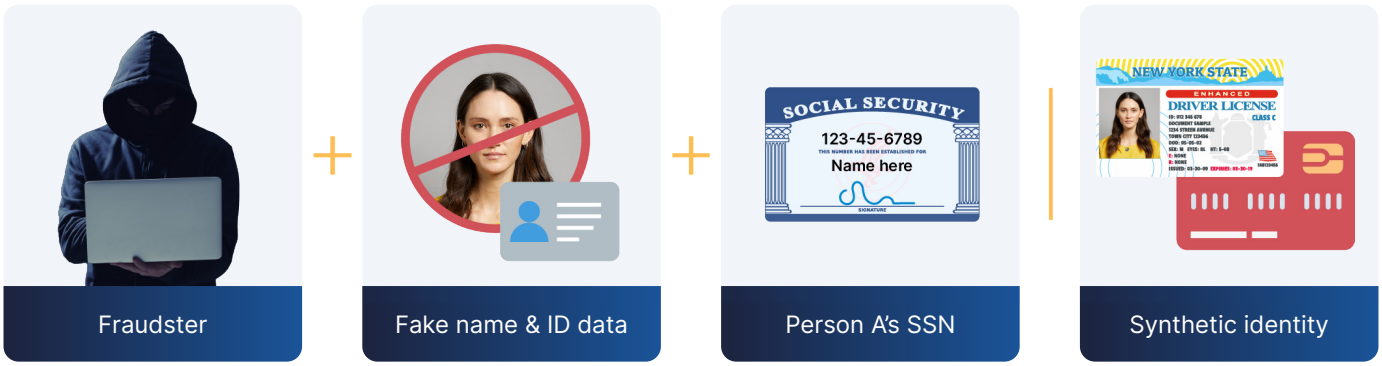
Synthetic imagery, such as deepfakes, can be used to enhance synthetic identity fraud. Generative Adversarial Networks (GANs) enable criminals to create deepfakes of individuals who do not exist. In this sense, the attacker can create a synthetic face to go with a synthetic identity.

For example, a criminal can forge an ID document with 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 then 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 Deepfakes Attacks Scale?

As we can see, if equipped with the ability to digitally inject synthetic media, criminals can launch thousands of attacks, 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, a thriving Crime-as-a-Service economy accelerates and widens the deepfake threat.

Crime as a Service (CaaS)

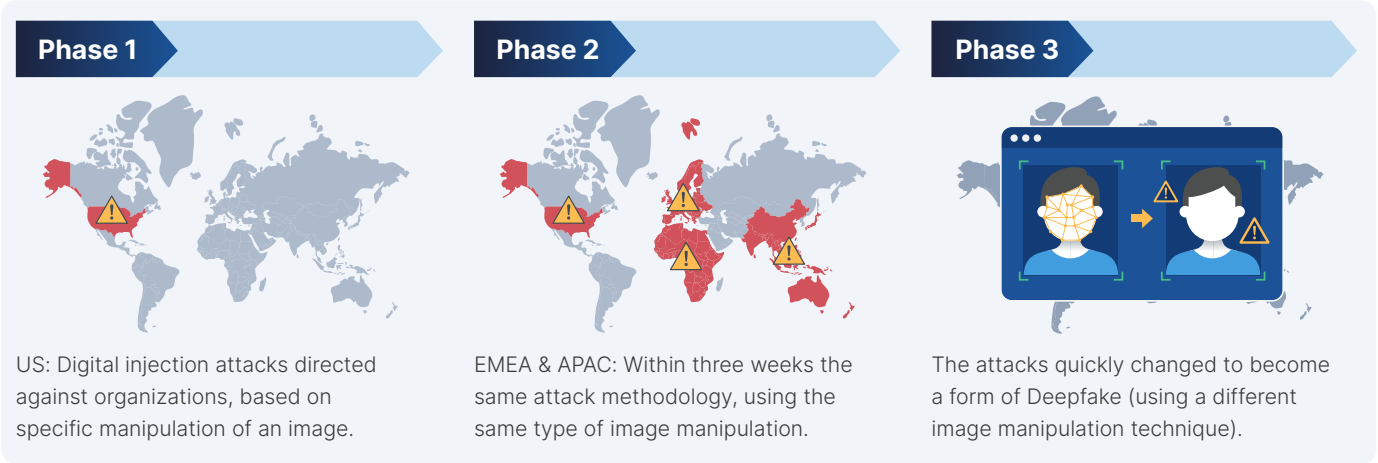
It's a mistake to think that criminals act in isolation. Rather, there's a sophisticated cybercriminal network with extensive communication channels. Once an attacker successfully infiltrates a bank's systems, they can use it for criminal purposes. But they can also quickly sell the effective tools, techniques and identities over the dark web.

This threat of CaaS is especially pertinent to Latin America. Research has shown that traditional violent-crime gangs are teaming up with an underground cybercriminal community to achieve common goals, primarily of the financial variety.

Throughout 2021, iProov's Security Operations Center (iSOC) witnessed clusters of similar methods of digitally injected attacks. At first, the attack took the form of a manipulated image. Eventually, this attack spread and developed into a more sophisticated deepfake. These occurred quickly, bounding from technology to technology and spreading across the globe.

This one example indicates that digital injection attack techniques are being rapidly shared and tested from numerous locations, whether by the same criminal organization or sold over the dark web. If attacks succeed, they rapidly escalate in volume and frequency, amplifying the risk of serious damage.

Likewise, CaaS means that if a criminal fails to break into your organization's system, they have very easy access to retool and try again.



US: Digital injection attacks directed against organizations, based on specific manipulation of an image.

EMEA & APAC: Within three weeks the same attack methodology, using the same type of image manipulation.

The attacks quickly changed to become a form of Deepfake (using a different image manipulation technique).

How Effective are Humans at Spotting Deepfakes?

Digitally injected synthetic media, such as deepfakes, is a clear and present danger to the Latin American financial services industry. It drives the question: how can a bank be sure that an unknown individual is who they say they are?

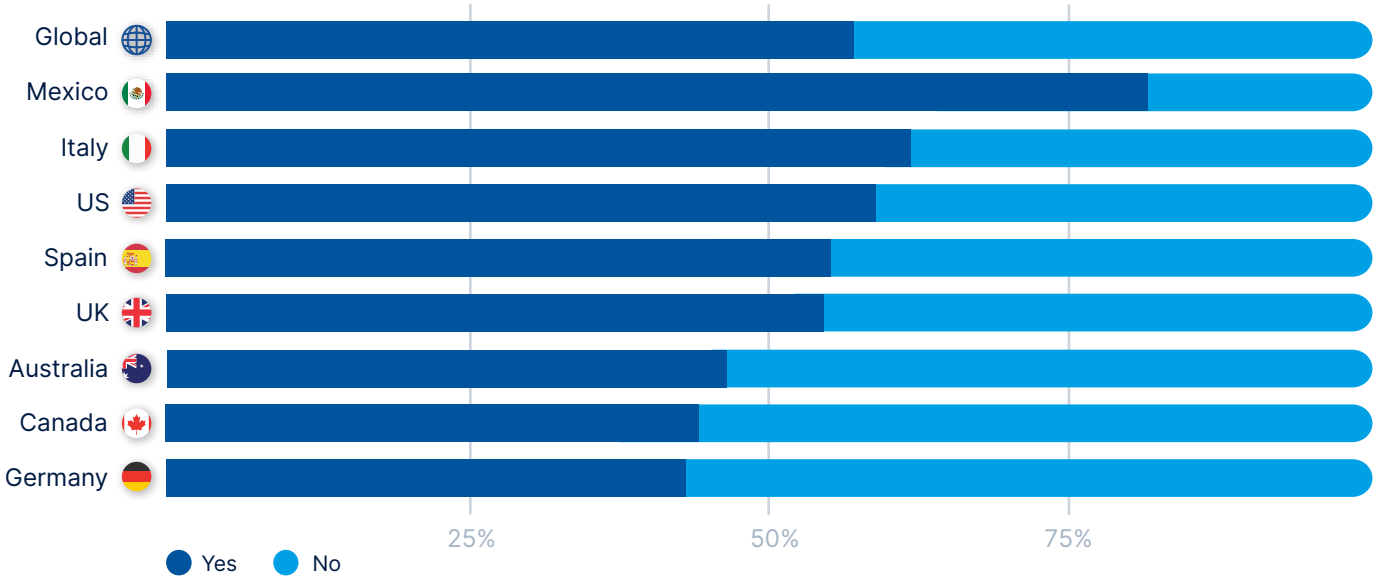
One approach is to verify customer identity and carry out KYC checks via a face-to-face video call between a member of staff and the user. Naturally, this relies on the staff member's ability to distinguish between a real person and synthetic imagery.

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 3: Do you think you would be able to tell the difference between a real video and a Deepfake?



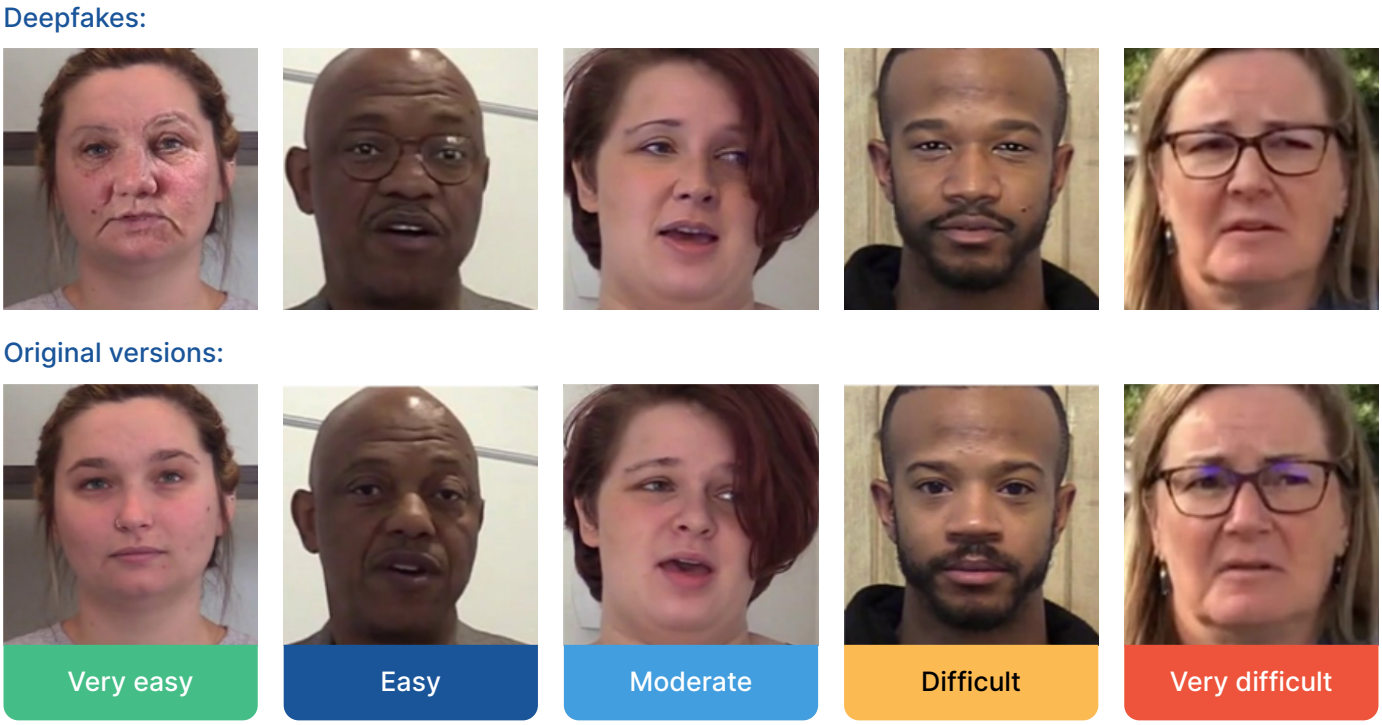
Mexican respondents were the most confident in their abilities to spot a deepfake, with 82% of those surveyed saying that they could do so.

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 results may have been skewed due to laboratory conditions. They believed that if a deepfake were to be delivered to an unsuspecting audience, the number of people successfully detecting them to be fake would be 'significantly lower.'

For banks, this skewed overconfidence strongly indicates that humans cannot be depended upon to identify deepfakes with the naked eye if presented with one during an identity check or video conversation. Technology is the only way to provide a level of detection assurance.

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



How Can Organizations Defend Against Deepfakes and Digital Injection Attacks?

Banks should be concerned about the threat of digitally injected synthetic media, but there is 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 liveness detection can offer presentation attack detection (PAD), many solutions are unable to safeguard against digital injection attacks, as advanced attacks circumvent liveness detection technology.

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, 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. 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. Learn more about this here.

The most effective method to defend against digital injection attacks is to create a one-time biometric, which we will describe next.

One-Time Biometrics

For high-risk use cases, such as opening a new account or transferring a large sum of money, most liveness detection technology does not provide a high enough level of assurance. Deepfakes can emulate a person verifying themselves, which some liveness technology cannot spot. Advanced verification methods are needed to secure against advanced threat types.

One-time biometrics that assure both liveness and that the user is a real-live person, verifying in real-time, is essential in a banks defense strategy against deepfakes

A one-time biometric is an authentication method that takes place in real-time to assure that a user is 'live' and genuinely present. A one-time biometric is never repeated in a user's lifetime and has a limited time duration, which cannot be reused or recreated and is worthless if stolen.

One way to achieve this with a standard device is to use the screen to project controlled illumination onto the user's face. 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.

Another advantage is that if it's stolen, it's worthless because it's one-time and obsolete as soon as it's used.

Summary

Face verification remains the most secure and usable method for banks to verify customer identities and authenticate transactions. However, digitally injected synthetic media, such as deepfakes, is a present and growing attack vector that is defrauding Latin American 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 Crime-as-a-Service economy and iProov research into the threat landscape show us that this threat is growing and scalable.

As research shows, the human eye cannot be trusted to successfully detect high-quality synthetic media. Instead, specialized technology is required. The deployment of 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, 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 genuinely present at the point of authentication (not digitally injected synthetic media, such as a deepfake).

Example of using Genuine Presence Assurance®

Next steps

For a demo: to see how Genuine Presence Assurance can detect digital injection attacks, please visit www.iproov.com/book-my-demo and submit your details.

For more information: visit our website at www.iproov.com to find out more about how we use face verification to help organizations to complete secure processes online.

Methodology

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.

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**

contact@iproov.com

iproov.com

IPLATReen-USLE01/23