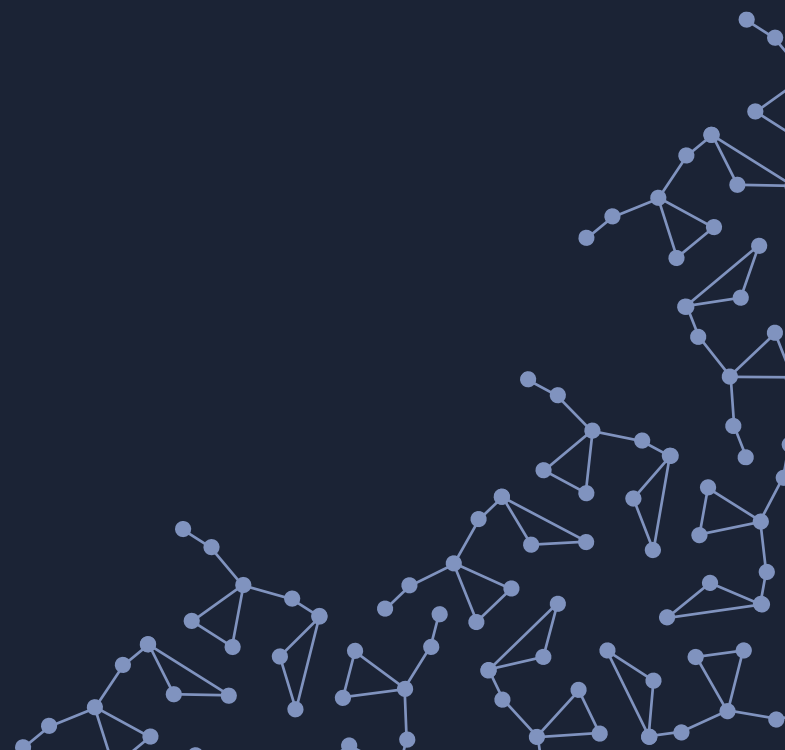
BioID
be recognized

# Deepfake Identity Fraud

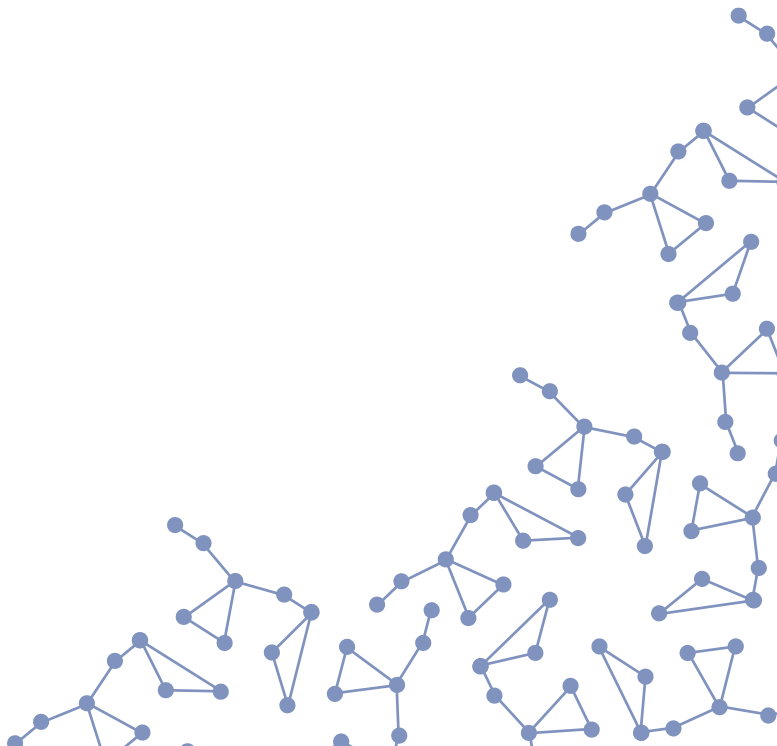
## A Guide to Mitigate AI Attacks on Identity Verification



BioID
be recognized

# Deepfake Identity Fraud

## A Guide to Mitigate AI Attacks on Identity Verification



Abstract ..... 5

1. Deepfakes – A Pandora's Box of Risks and Threats ..... 6

2. Identity Verification and Deepfake Attacks ..... 7

3. Creating Deepfakes – The Professional vs. Amateur Perspective ..... 9

3.1 Professional Deepfake Artistry with Autoencoders, GANs, and Diffusion Techniques ..... 10
3.2 Amateur Deepfake Creation with User-Friendly Apps and Platforms ..... 13

4. Deepfake Detection – The Battle AI vs. AI ? ..... 14

4.1 How to Detect Deepfakes Manually ..... 14
4.2 Automated Detection of AI Manipluation & Generation ..... 16

4.2.1 Detect Deepfakes with Presentation Attack Detection ..... 16
4.2.2 Detect Deepfakes with Anomaly Analysis ..... 17

5. Recommended IDV Proceeding to Stop Deepfake Impersonation ..... 19

6. Prevent Impersonation with BioID ..... 20

7. New threats Arising - An Outlook ..... 21

About BioID ..... 23

# Abstract

While it might not have happened knowingly, by today basically everyone has seen at least one deepfake medium online.

**Yet deepfakes still remain a mystery for many.**

The word itself is a fusion of "deep learning" and "fake". It entails manipulated media, predominantly videos, audio recordings or images, where the original content is replaced or altered using sophisticated machine learning algorithms and generative AI. These alterations are often unrecognizable to the human eye, rendering deepfakes a potent tool for misinformation, identity theft, and fraud.

Essentially, there is an ongoing arms race between content creators and detection methods as creators attempt to push the boundaries of AI models. This underscores the urgent need for robust deepfake detection solutions.

As a response, BioID introduces its deepfake detection software. It is a cornerstone in reducing the risks associated with digital identity verification processes. The SaaS solution is easy to integrate in applications spanning corporate enrollment, Know Your Customer (KYC), Anti-Money Laundering (AML) compliance, and more. By leveraging advanced technologies, BioID's Deepfake Detection offers unparalleled reliability in safeguarding against costly fraud incidents.

But what are the risks and dangers of deepfakes and how are they created? How does deepfake detection stand as a protection of authenticity in today's digital identity verification? Addressing IDV solution providers and users, this whitepaper aims to answer these questions in the following chapters.

# 1.
# Deepfakes – A Pandora's Box of Risks and Threats

The rise of deepfake technology introduces profound risks, explicitly in identity fraud. By convincingly replicating individuals' appearances, fraudsters proceed with unlawful activities, including financial scams, political manipulation, and criminal impersonation. Just recently for example, global news outlets have reported an incident involving a finance employee of a multinational corporation who fell victim to a sophisticated deepfake fraud scheme.

In this scam, the unsuspecting worker participated in a video call under the impression that fellow attendees were genuinely his colleagues. However, all participants were crafted deepfakes. Under the false pretence of authentic interaction, the finance worker consented to transfer a sum of $200 million Hong Kong dollars, equivalent to approximately $25.6 million USD (Source: CNN World).

Such manipulation not only poses significant challenges for law enforcement but also undermines trust in digital communication channels and identity verification systems.

Additionally, the emergence of "deep porn", involving the imposing of individuals' faces onto explicit content without consent, represents a violation of privacy and consent. Statistically speaking, alone in 2023 it is estimated that over 113.000 deep porn videos were uploaded on the internet (Source: Cryptopolitan). Consequently, there is an urgent imperative for the development and implementation of robust detection and prevention mechanisms.

Furthermore, deepfakes are one of the main causes of widespread fake news and misinformation. By fabricating convincing audiovisual content depicting public figures engaging in falsehoods or unethical behaviour, this manipulation undermines media literacy and democratic discourse. It necessitates proactive strategies such as the EU AI Act among many regulations to mitigate them. This noteworthy development signals a growing recognition of the potential dangers posed by deepfakes.

Having all these issues in mind, we need to focus even more on technical solutions for deepfake detection. These may employ sophisticated techniques like forensic analysis, biometric comparison, watermarking and machine learning models to identify manipulated media.

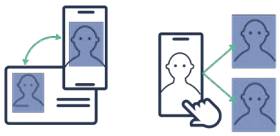
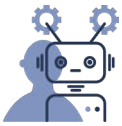
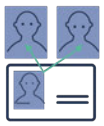
As deepfake technology continues to evolve, the need for reliable detection methods become a necessity, in particular when dealing with regulatory compliance and related fraud prevention at banks, governments or large enterprises.

## 2. Identity Verification and Deepfake Attacks

Digital identity verification plays a pivotal role in various sectors such as finance, healthcare, and online services. It is the act of verifying an individual's identity, typically using a photo identity document during the onboarding process.

For companies, banks and governmental facilities, digitalizing identification measures provide benefits such as saving time and cutting costs that would occur during the traditional laborious physical identity verification through employees. For KYC (Know Your Customer) and AML (Anti-Money-Laundering) compliance, such manual or (semi-) automated processes are a necessity for many industries.

The advanced technological methods rely on biometric data analysis. Because biometric data is unique to each individual it can easily be utilized for trained algorithms to accurately ensure that individuals are identified in digital transactions and interactions.

| Step 1 | Step 2 | Step 3 |
| --- | --- | --- |
|  |  |  |
| User submits identification document and/or biometric data. System verifies the authenticity of the submission. | Liveness detection process ensures the presence of a live person during verification. Deepfake detection algorithms analyse media content for signs of manipulation. | Facial matching and final verification result displayed to the user (e.g., successful verification or request for additional information). |

Biometrics, and facial verification in particular, can be used to bind a physical person to a digital identity and derive valuable identity information, such as in the context of KYC. However, established processes all around the world, both fully automated as well as agent-assisted identity verification (IDV) through online channels are now being challenged: Deepfakes are creating new attack vectors making digital identity verification vulnerable hence its integrity questionable.

**To mitigate the risk posed by deepfakes on IDV processes, it is crucial for businesses and facilities to adopt advanced detection mechanisms specifically tailored to identify synthetic or manipulated media. This includes the development of deepfake detection algorithms capable of distinguishing between genuine and fabricated data subjects.**

Additionally, implementing multi-layered security methods that combine various forms of (biometric) analysis enhances the robustness of identity verification processes. By applying a holistic approach to secure IDV applications, organizations can detect and prevent deepfake attacks to enhance overall security posture. For this, comprehensive knowledge about deepfake creation and use is beneficial and will be portrayed in the following chapter.

# 3.
# Creating Deepfakes – The Professional vs. Amateur Perspective

AI-generated face deepfakes are becoming increasingly sophisticated. Some of the methods are listed in the following:

**Face swaps:** replacing one's face with another.

**Face reenactment:** transferring facial expressions and movements from one person to another.

**Morphing:** blending the facial features of one individual onto another's.

**Attribute manipulation:** modifying facial attributes like age, gender, hair colour, or facial hair and thus altering a person's appearance.

**Face synthesis:** generative AI that creates a completely new face, crafts a lifelike and unique identity of a non-existent person.

Such techniques cause massive obstacles for digital identity authentication processes. As of today, many tools and techniques exist for both professionals and amateurs to create convincing deepfake material:

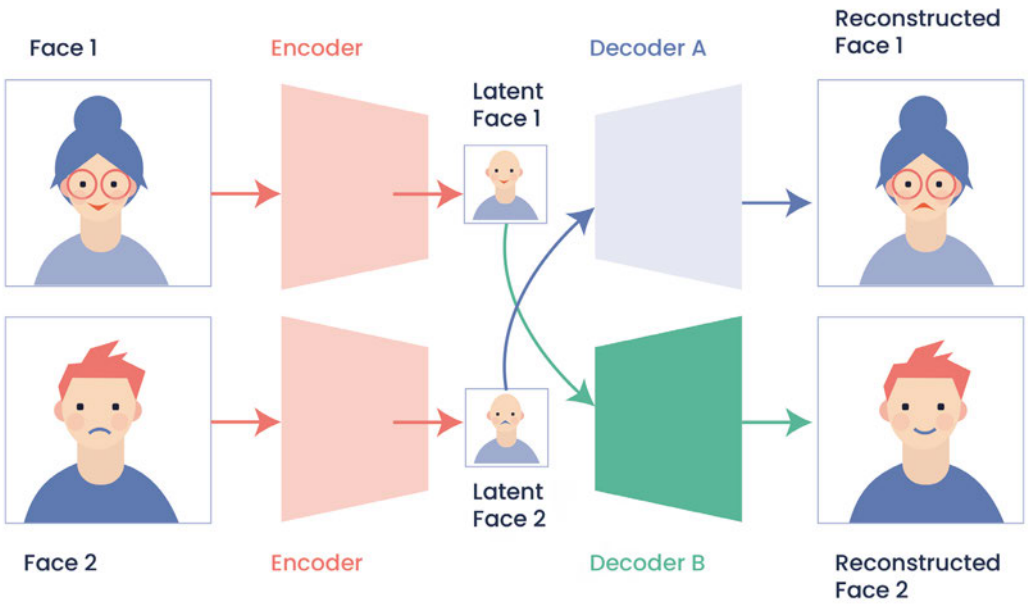
| PROFESSIONAL | AMATEUR |
|---|---|
| Autoencoder<br>GANs<br>Diffusion | **Text-to-Image Platforms:**<br>Dall-E, Midjourney, Leonardo AI, Stable Diffusion, DeepFaceLab<br><br>**Text-to-Video Platforms:**<br>Sora by OpenAI<br><br>**Fun Apps:**<br>Animafy, Animate, Avatarify, Copyface, DeepfakeStudio, FaceApp, Facefy, FacePlay, Impressions, Jiggy, MugLife, MyHeritage, Nostalgia, Reface, VidAvatar, Wombo, Xpression |

## 3.1 Professional Deepfake Artistry with Autoencoders, GANs, and Diffusion Techniques

Professionals often employ algorithms and software to produce high-quality deepfakes. The following technologies are commonly used, often as a combination:

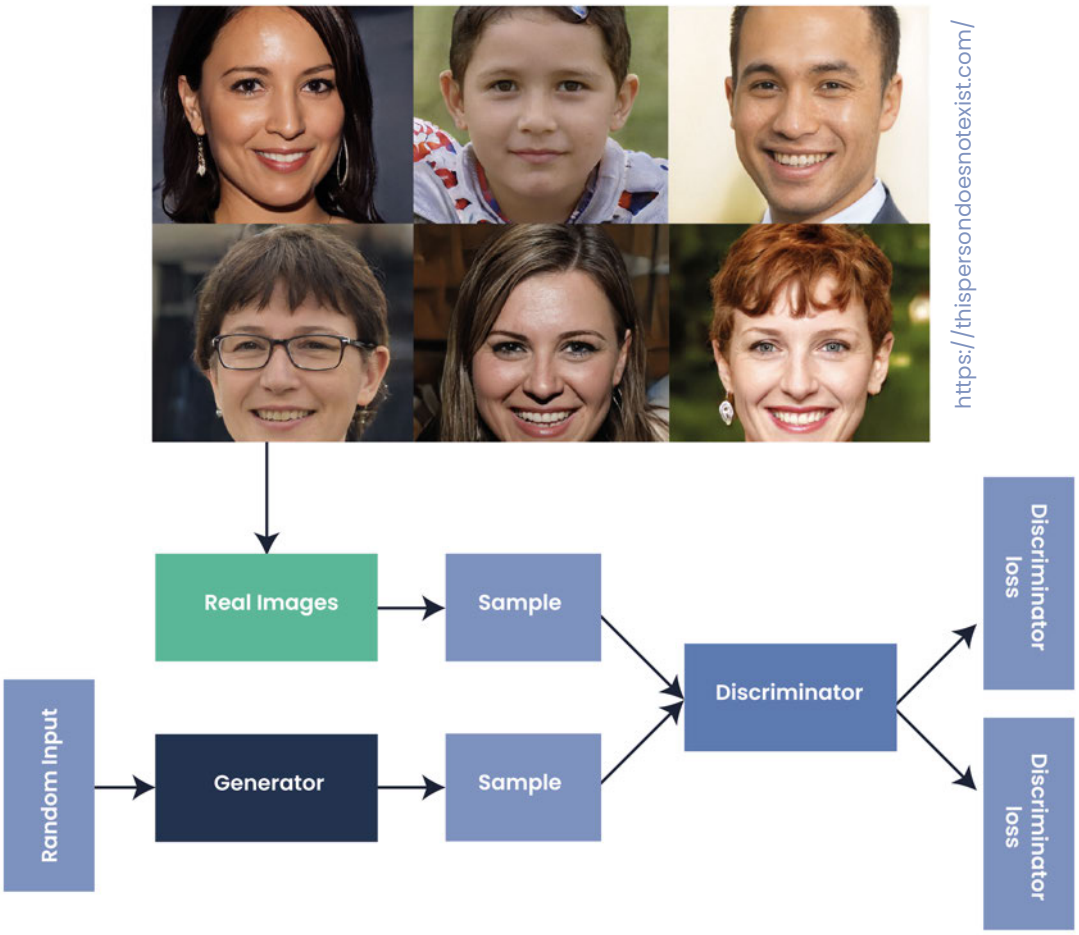
### Autoencoder:

Autoencoders are neural network architectures primarily used for data compression and feature learning. In the context of deepfakes, autoencoders are, so-to-say, the artists in the background that capture and recreate facial features with high fidelity. The process unfolds in two stages: compression and reconstruction. Initially, the autoencoder engages in dimensionality reduction, condensing the input data—such as facial images—into a compact representation. Then, the compressed data undergoes reconstruction by the decoder network, reverting it back to its original form. Along the way, the autoencoder studies the details and patterns in the data, learning what makes a face look like a face. This knowledge helps twist these features to create new, fake images, showing the power of deepfake technology.



### Generative Adversarial Networks (GANs):

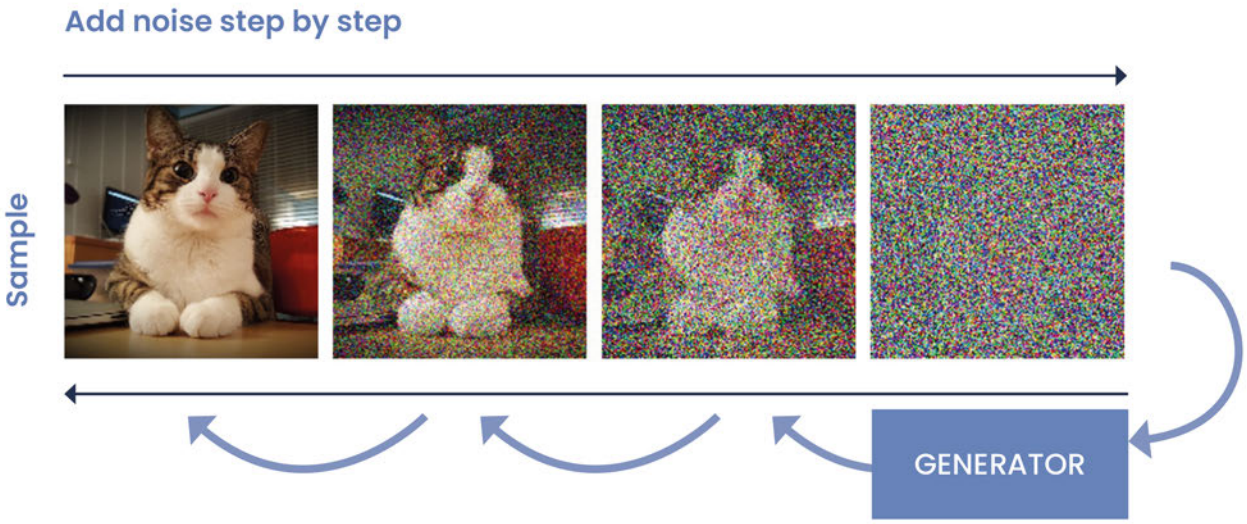
GANs consist of two neural networks, a generator and a discriminator, engaged in a competitive game. The generator creates synthetic data, while the discriminator distinguishes between real and fake data. GANs are widely employed in deepfake generation due to their ability to produce realistic outputs. This method laid the foundation for more sophisticated deepfake techniques, but it is now less preferred by deepfake creators.



### Diffusion Models:

The evolution from traditional Generative Adversarial Networks (GANs) to newer methods, particularly diffusion models, marks a new era in deepfake creation. These models refine noisy data to create convincing deepfakes by understanding patterns in large datasets. This allows them to capture subtle details of facial expressions, gestures, and other features crucial for creating convincing deepfakes. Once they've grasped the essence of the data, diffusion models simulate changes, such as altering facial characteristics, expressions, or background elements. These changes are made in a way that preserves the realism of the original data, ensuring that the resulting deepfakes closely resemble the originals while incorporating the desired modifications.

Diffusion models strive to retain the overall structure and content of the original data, making adjustments while ensuring that the resulting deepfakes look as authentic as possible.




Generator gradually learns to create a proper image from noise

## 3.2 Amateur Deepfake Creation with User-Friendly Apps and Platforms

Amateurs typically rely on user-friendly applications and simplified AI tools created by professionals to make basic deepfake content. The technologies commonly used by amateurs include:

**Fun Apps:**
Various mobile and desktop applications offer intuitive interfaces for creating simple deepfake videos. These apps often provide pre-trained models and automated processes, enabling users with limited technical knowledge to generate basic deepfakes.

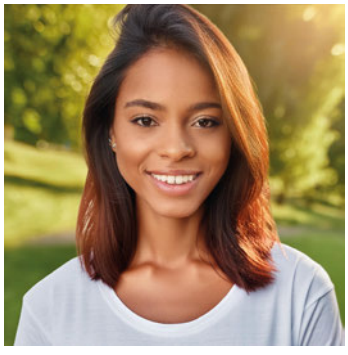
**Platforms:**
Amateurs may also leverage generative AI platforms that offer user-friendly interfaces and simplified workflows for generating deepfake content, mostly as text-to-image models. Upcoming are highly potent text-to-video platforms like SORA from Open AI. These platforms typically utilize pre-trained models and cloud-based processing to facilitate easy creation of deepfake photos and videos.

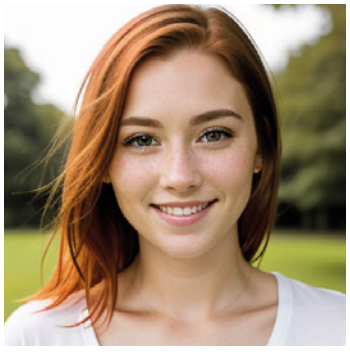
### One prompt - different AI platforms

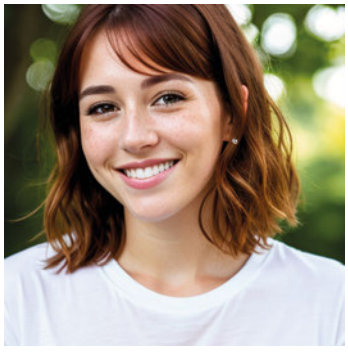

Leonardo Diffusion XL


Adobe Firefly Image 3


aZovyaPhotoreal_v2.safetensors


Adobe Express


aZovyaPhotoreal_v2.safetensors


realisticVisionV20_v13.safetensor

# 4.
# Deepfake Detection – The Battle AI vs. AI ?

The rise of deepfake technology has led to a chase between creators and detectors. As deepfake algorithms evolve, so must the methods employed to detect and combat them. This battle pits artificial intelligence against itself, with detection systems striving to stay one step ahead of increasingly convincing synthetic media. To counter the threat posed by deepfakes, the first step is to create awareness for deepfakes and how to expose them. For individuals, the following tips can help detect deepfakes with the human eye:

## 4.1 How to Detect Deepfakes Manually

**Facial features and body language:**
One should focus keenly on nuances in facial features such as eyes and teeth, as deepfake technology typically struggles to replicate these details convincingly. Deepfakes can sometimes have odd or exaggerated facial expressions that do not align with the rest of the body language or speech. Check for micro-expressions that seem off. Authenticity is often reflected in seamless emotional alignment with spoken content and the natural fluidity of movements, whereas deepfakes may betray themselves through awkward gestures and asynchronous facial cues.

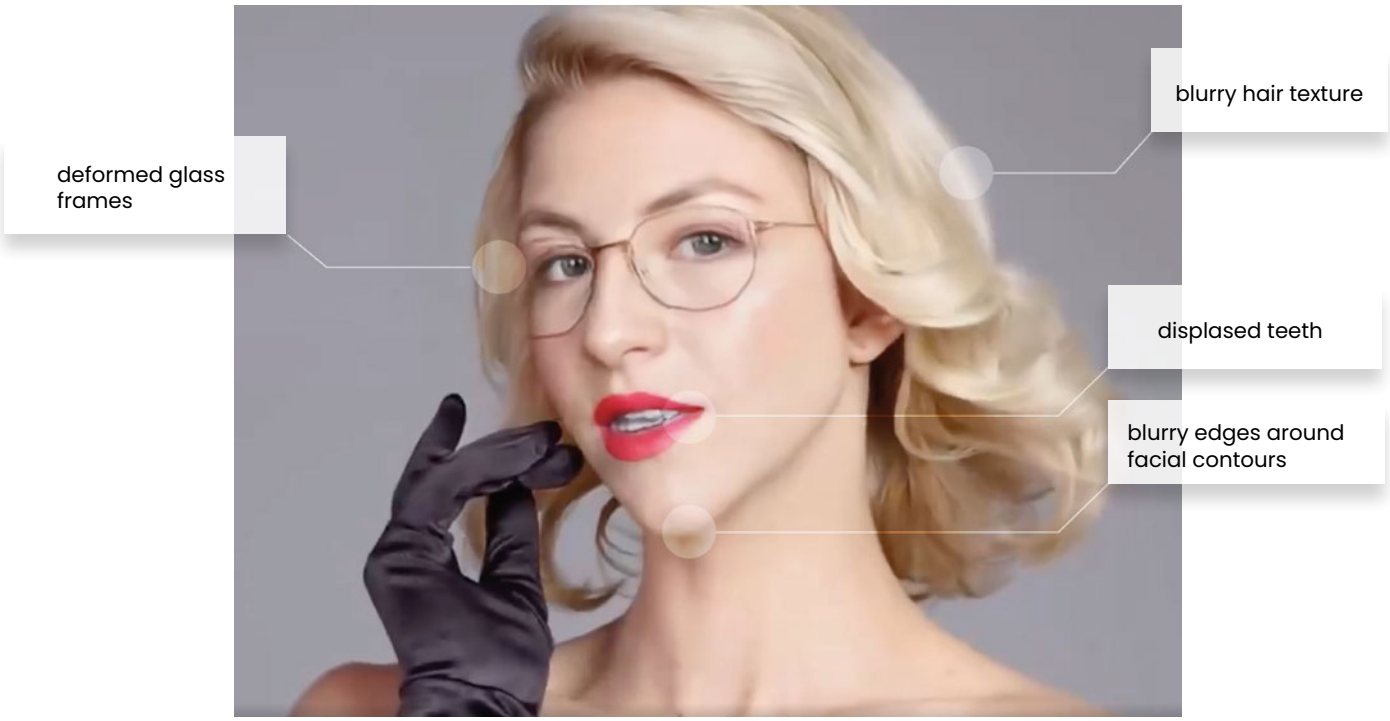
**Inconsistent blinking:**
Human blinking patterns are typically regular and natural. Deepfakes sometimes exhibit unnatural blinking patterns or a complete lack of blinking, though newer models are getting better at mimicking this.

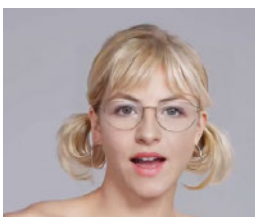
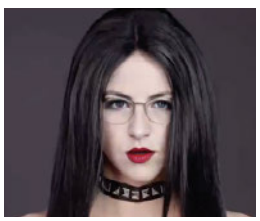
**Other telltale signs:**
One should be vigilant for signs of manipulation, such as blurry edges around facial contours, in particular during movements.

**Lighting and shadows:**
Look for inconsistencies in lighting and shadows on the face. Deepfakes might not have correctly aligned lighting, making it appear unnatural. The eyes can be particularly tricky for deepfake algorithms. In high-quality photos or videos, look for reflections in the eyes.

**Skin texture and anomalies:**
The texture of the skin may look overly smooth or too perfect. Real human skin has pores, wrinkles, and other small imperfections that deepfakes might not replicate accurately.

**Source:**
Last but not least, one should scrutinize the source – favouring content emanating from reputable platforms or known entities, as credibility often aligns with trusted sources.



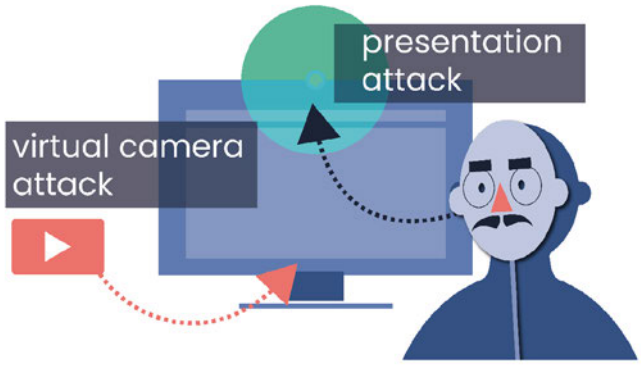

App: FacePlay

## 4.2 Automated Detection of AI Manipluation & Generation

Even with a 20/20 vision, one's abilities to detect deepfakes is naturally limited. Thus, training human agents involved in video identity verification is crucial but not sufficient. Therefore, artificial intelligence (AI) is the best way to offer effective deepfake detection in matters such as identity verification. Video agents can be assisted by AI to uncover AI manipulation or -generation. The same accounts for manual examination of photo- and video material in court or broadcast: assistance by technology is an unparalleled advancement concerning the assessment of genuineness. But how, in detail, can impersonation and fraud be stopped by technology?

### 4.2.1 Detect Deepfakes with Presentation Attack Detection

Face Liveness Detection is an anti-spoofing method for facial biometrics. Scientifically, it is called Presentation Attack Detection (PAD). The core function of a PAD mechanism is to determine whether a biometric feature (e.g. a picture), was captured from a live person in front of a camera. According to the international PAD standard ISO/IEC 30107, fraudulent biometric traits can for instance be derived from photo printouts, paper masks or video replays presented to a camera. These attacks are referred to as a "Presentation Attack", as formerly captured facial traits are replayed to a sensor, namely a camera.



It is important to differentiate between two types of fraud attacks, namely presentation attacks (sensor level) and application-level attacks.

Presentation Attack Detection (PAD) prevents presentation attacks that occur at the sensor level, such as in front of a camera. It blocks fake biometric data from being presented to the camera by identifying and blocking various types of spoofing attempts, such as video replays, 3D papers, silicon masks, and deepfakes on displays. Liveness detection algorithms automatically reject any type of replays on displays – a deepfake presented as such – is no high-risk attack. It would be detected with common methods such as forensic texture detection and AI.

Stopping biometric fraud and deepfakes is only a problem if the camera source is attacked, e.g. with a virtual webcam. This attack is considered an application-level attack or video injection attack. While presentation attacks involve manipulating images to create deceptive content, application-level attacks target the software or hardware components of a biometric system. Attacks at the application level, e.g. using a virtual camera driver to feed a video into the application directly, are not subject of the international PAD standard ISO/IEC 30107.

That requires additional mechanisms to directly identify deepfakes within the video content.

### 4.2.2 Detect Deepfakes with Anomaly Analysis

Advanced deepfake detection algorithms leverage machine learning models trained on large datasets of authentic and manipulated media to identify patterns indicative of deepfake manipulation. As typical for training artificial intelligence in the form of DNNs, the quality and representativeness of the data is crucial. It needs to include a broad variety of different deepfakes to be able to generalize properly.

This means that it can also detect unknown types of AI generation or manipulation reliably, if set up carefully and with balanced datasets. In addition, a deep understanding and long-term experience of pattern recognition, forensic analysis and real-world challenges are the baseline for delivering a trusted solution.

Automated deepfake detection can use a combination of the following approaches to unmask AI manipulation and generation in photos and videos:

**Physical movement inconsistencies:**
Deepfake videos may exhibit unnatural facial expressions, lip movements, or body gestures that deviate from normal human behaviour.

**Inconsistencies in facial features:**
Deepfake algorithms may struggle to accurately replicate subtle details in facial features, such as eye movements, blinking patterns, or reflections in the eyes.

**Timely inconsistencies:**
Deepfake videos may contain inconsistencies in the timing or synchronization of audio and visual elements.

**Artifacts and distortions:**
Deepfake generation processes often leave behind artifacts or distortions in the video, such as blurring, pixelation, etc.

**Forensic analysis:**
Advanced forensic techniques can be used to analyze the digital footprint of a video, including metadata, compression artifacts, and other traces left by editing software.

BioID's Deepfake Detection is a real-time AI manipulation/ generation fake detector that only requires one image for analysis. As it is much easier to detect deepfakes using multiple frames as provided through a video, naturally speaking BioID also offers video deepfake detection.

# 5.
# Recommended IDV Proceeding to Stop Deepfake Impersonation

At BioID, while we offer technical elements to combat fraud, we also prioritize the security of our customers' end-user applications. Thus, our recommended approach for highly trusted identity verification processes is multi-layered:

## Biometric Components as Provided by BioID

**Presentation Attack Detection:**
Offering liveness detection since 2004, BioID has advanced its PAD to cover all types of presentation attacks including printed photos on various materials, 2D and 3D masks made from plastic, clay, resin, silicon, etc. In addition, animations, projections, videos and deepfakes are covered, as confirmed through BioID's ISO/IEC 30107-3 compliance and multiple other certifications.

**Deepfake Detection:**
Utilize BioID's advanced algorithms to identify and differentiate between authentic and manipulated facial images or videos, including those generated or altered by AI.

**Challenge-Response:**
Enhance security measures with optional challenge-response mechanisms. A BioID patent, this adds an additional layer of defense to thwart the use of pre-recorded videos or injected deepfake content.

**AI Based Injection Attack Detection:**
E.g. through video signal analysis. Please contact us to learn more about BioID's current proceedings in this field.

## Additional Recommendations:

**Native App Integration:**
Integrate secure (native) applications to ensure robust protection against unauthorized video manipulation, safeguarding video captions and preventing the infiltration of virtual or modified camera signals.

**Blacklisting Mechanisms**:
Implement measures to block virtual camera drivers, particularly within browser-based applications such as OBS, ManyCam, and Avatarify, fortifying defences against potential security breaches.

## 6.
## Prevent Impersonation with BioID

With more than 25 years of biometric development, BioID is an active technology leader in foreward-looking research. Partnering with renown German research institutes and institutions like the German Bundesdruckerei (responsible for issuing ID documents) BioID has been actively engaged in research aimed at real-time detection of deepfakes in photos and videos, since 2020. This ongoing endeavour is conducted in collaboration with the FAKE-ID initiative, a research consortium funded by the German Federal Ministry of Education and Research (BMBF). Leveraging the power of AI, BioID seeks to identify and mitigate potential attacks on visual media. BioID's participation in the FAKE-ID deepfake detection research showcases its expertise in biometrics and proprietary anti-spoofing technologies.

Moreover, with a strong dedication to high-end security solutions and privacy-centric biometrics, BioID continuously enhances the efficacy of our technology to secure our customers' applications and users. As a result, BioID's solutions guard digital identities against impersonation and identity fraud, and prevent fake news.

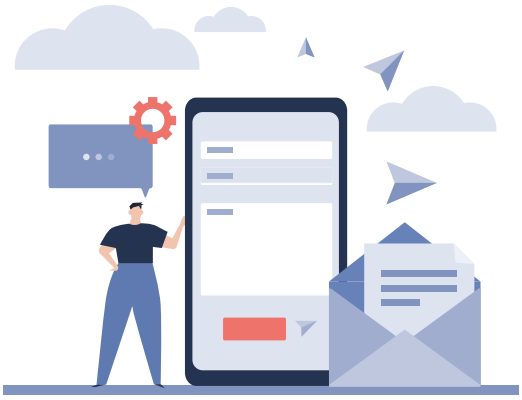
Today, as part of BioID's ISO/IEC 30107-3 compliant liveness detection, the latest deepfake detection mechanisms are available to all BioID Web Service users.

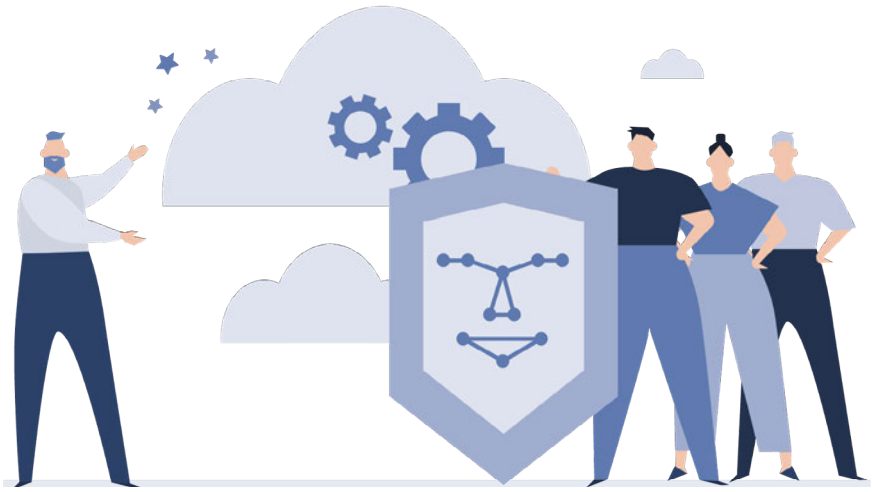
# 7.
# New threats Arising - An Outlook

While AI detectors are available today, no vendor nor company can rely on its existing security implementation being sufficient tomorrow. Constant updates and further development are required to keep up with the immense pace deepfake technology evolves in. State-of-the art deepfake detection should be able to uncover manipulated material from all major available deepfake apps as well as the many plattforms and text-to-image generators. At the same time, new threats are already emerging, one of them being the upcoming SORA from Open AI. An impressive text-to-video tool that creates realistic AI videos within seconds with only a short prompt.

At BioID, we are specialists in developing and detecting deepfakes. Our mechanisms are, already today, shielded against the above threats. With continuous efforts, BioID stays at the forefront of AI development, not only in the field of deepfakes but also in reliably preventing all kinds of identity fraud, while seamlessly verifying genuine users.

Contact us to add BioID's comprehensive and robust solutions to your services for combatting synthetic media manipulation of today and tomorrow.





# About BioID

BioID is a leader and pioneer in cloud-based biometric services. With its renowned BioID Web Service (BWS), BioID sets the highest standards for a secure biometric SaaS particularly for online face recognition.

BioID's mission is rooted in the belief that anonymous biometric authentication empowers users to fortify their online identities discreetly. Guided by this vision, BioID seamlessly connects real-world individuals with their digital personas.

Via its groundbreaking patented and revolutionary technologies, BioID makes self-service unattended identity validation a reality.

With operations in Switzerland and the USA, BioID is privately held and has its research centre in Germany. Years of successful implementation across businesses, banks, and governmental organizations have demonstrated the strength of our technology.

BioID Web Service can be tested free-of-charge on its playground.

**www.bioid.com**

BioID® GmbH
Bartholomaeusstr. 26D
90489 Nuremberg
Germany

+49 911 999 9898-0
e-mail: info@bioid.com