
THE RAPID SPREAD OF DEEPFAKES
AND HOW TO PROTECT YOURSELF AGAINST THEM

## CONTENTS

## EBOOK OVERVIEW

Technology begets better technology, which has caused digital functionality to progress exponentially. Artificial Intelligence not only can write code, it can write its own code, and continually learn in order to make itself smarter. This has led to the meteoric rise of deepfakes, or AI-generated constructs that mimic human voices, faces, and even behavior. The potential benefits of AI are paralleled by the nefarious use of its deepfake children. Political misinformation, social disruption, and fraud attacks both minor and massive have been perpetrated through the use of deepfakes.

authID explains how deepfakes are spawned and continually improved, and how they are used for fooling both human and digital channels for committing fraud and other crimes. By understanding the arc of deepfakes, IT and other professionals can learn to exceed the most basic (and ineffective) defenses against deepfakes and achieve a truly capable approach to staying ahead of this self-perpetuating digital attack. By evolving our tech and how we deploy it, we can perfect our best practices to beat the deepfake threat.

## ABOUT AUTHID

authID (Nasdaq: AUID) ensures enterprises “Know Who’s Behind the Device” for every customer or employee interactions, through our easy-to-integrate, patented, biometric identity platform. authID quickly and accurately verifies a user’s identity, to protect accounts and other digital assets. By creating a biometric root of trust for each user, authID stops fraud at onboarding, detects and stops deepfakes, eliminates password risks and costs, and provides the fastest, frictionless, and most accurate user identity experience in the industry.


ADD RECORD\\
sys_exit)
add.citizen
t%10s\
//if(go)
FILE * old//
add.acc
%d %s
scanf("%d",& main_exit);
(main_exit=
0
if (main_exit==1)
scanf("
;if (choice==1)
return to main menu and 2 to exit:")

# CRIME ALSO ENJOYS PROGRESS

For as long as there have been objects precious to humans, there have been other humans willing to steal them outright. And then there have been others who were clever enough to steal through subterfuge. Instead of brute force, they have used a combination of dishonesty and trickery to take that which belongs to somebody else. Because they are clever, they have also taken advantage of all manner of inventions and innovations to further their underhanded cause.

> ***"For as long as there have been objects precious to humans, there have been other humans willing to steal them outright"***

Counterfeits are as old as currency. The ancient Egyptians punished by death the criminals who fabricated their own versions of painted pottery that served as receipts for grain kept at the central repository. Metallurgy brought us coinage, as well as fake coins. The printing press has spread scholarship as well as phony paper money.

Impersonation is nearly as old. Jacob disguised himself as his brother Esau in order to steal the blessing of their blind father Isaac. In the nineteenth century, Arthur Orton claimed to be the presumed-dead son of a wealthy woman in England, and managed to convince her of that fact and acquire a fat allowance. Only when he tried claiming the estates of the dead man was he found to be the son of a butcher and put in prison.

The internet brought about the age of mass scale remote interaction, making it all the easier for bad actors to act as others. They use the internet to steal identity data, then utilize that same web to leverage that data to pretend to be people with assets they wish to steal.

As technology begets itself, and innovation accelerates its own pace, counterfeiting has managed to reach all new depths.

Progress is a hockey stick, an exponentially accelerating graph. Hundreds of thousands of years of stone tools turned to thousands of years of metal tools, evolving into hundreds of years of steam power. Less than two centuries of electrical power, and the telegraph quickly giving away to phones and radios, and now just three decades of HTML have become chatbots. Now the capacity for impersonation is growing exponentially. We've always suffered from very clever fakes, in the forms of identities, faces, and voices. But now Gen AI has super-powered criminals, allowing them to create false presentations of identity on an industrial scale, and at breakneck speed.

Gen AI was itself barely an infant when we recognized, and were already falling victim to, its criminal misuse, in the form of Deep Fakes, which are constructs, meant to mimic faces, voices, or even video, for the purpose of committing fraud, spreading misinformation, or otherwise perpetrating nefarious activities. So how are these created?

- Images, voices, and video can be manipulated from source materials. They can be modified to make their sources appear to appear, say, or even do something that the sources had not.
- Images, voices, and video can be generated from scratch, based on templates that aren't necessarily based on real people.

Gen AI employs deep learning to assimilate models or templates so that it can generate variations on those models, even merging different kinds of images into a single, final package, the Deepfake.

***"Gen AI has super-powered criminals, allowing them to create false presentations of identity on an industrial scale"***

What are deepfakes? They are artificial images, videos, sounds, or other factors created by Generative Artificial Intelligence, which leverages machine learning in order to assimilate the necessary data to create variations on extant elements. In layman's terms, a deepfake can be a fake face, fake voice, fake video, or fake id. Gen AI learns from large sets of real faces, voices, etc. in order to make its own versions. Deepfake examples are all over the media, literally every single day.

How are deepfakes created? Deepfake apps take advantage of sophisticated technology. Generative Adversarial Networks (GANs) are two-headed approaches to developing deepfakes. One half generates a fake, and the other half attempts to identify it as fake, in an iterative process resulting in better and better fakes that can pass detection. Artificial Neural Networks (ANNs) train on large datasets for predictive modeling and serve as integral factors in the creation of deepfakes.


PASSWORD
LOCKED
ATTACK

# DEEP FAKES GO DEEP… AND WIDE

Teenagers have forever been trying to acquire fake ids so they can get into the bar. They would buy fake ids, or modify existing ones, in order to pass as older. But we're past that. Far, far past.

Deepfakes have been used to make it appear that politicians and celebrities have said things they didn't. The purpose has been to create disorder or generate phony dissent. Almost anyone can easily access deepfake apps, many of them for free, in order to create deepfake examples, using existing video, images, and voices of those famous persons. The potential to undermine elections and other civic activities is dangerous. In fact, when AI started truly grabbing headlines in a big way in 2023, the proclamations of doom were instant. Visions of Skynet and the Terminator were put forth by paranoiacs who read too much science fiction and not much science fact.

> ***"The underground site OnlyFake made it unbelievably easy to create ridiculously realistic fake ids"***

On its own, AI is still pretty dumb in that it often generates images with clear flaws (such as people with two left hands or elongated limbs). But what the average deepfake app does is make the jobs of bad guys far easier. If existing tools previously allowed criminals to perpetrate fraud at a high rate, AI enables fraudsters to commit their evil acts on an industrial scale. In the old days, we worried over "script kiddies" who could download open source hacking tools for infiltrating networks. It still took some level of skill to unlock and use these tools. But AI is a productivity tool for even the dumbest hackers. Deepfake apps can be operated by even those with no talent. With AI, literally any dummy can commit fraud.

Teenagers with fake ids? Amateurs. Almost any deepfake tool can create phony physical ids. Using simple prompts, fraudsters can tell AI-powered deepfake apps to generate phony drivers' licenses, passports, and other types of illicit docs containing the PII of their choosing.

The underground site OnlyFake made it unbelievably easy to create ridiculously realistic fake ids, available for 26 countries, and for only $15 each. Choose a document type, provide a picture (or have them do it for you), and you're a new you, or a new somebody else. The founder claimed that their fake ids could pass KYC (Know Your Customer / Patriot Act), and defenses at Coinbase, Binance, and several other bitcoin platforms.

Before the site went dark, its Telegram account declared that "the era of rendering documents using Photoshop is coming to an end." The script kiddie version of fake ids is replaced by the instantaneous and very realistic version. Forget modifying your

license so your underage self can get into the bar. Now you can deep-fake your way into someone's bank account.

The sophistication of deepfakes is increasing faster than most deepfake detection tools can keep up.

Deepfake identities can even replicate the barcodes on the backs of drivers' licenses or the MRZ on a passport. This makes for an even more convincing presentation. By presenting deepfake ids in conjunction with real or synthetic faces, standard identity verification platforms can be fooled. Provide images of the ids as well as matching faces, and you're in. And inevitably, fraud follows.

***"Deepfake identities can even replicate the barcodes on the backs of drivers' licenses or the MRZ on a passport"***

What kinds of fraud? Precisely how do deepfakes work in the practice of fraud? What are the deepfake examples behind fraud? Through the generation of deepfake ids, voices, and faces, Gen AI enables:

- Third party fraud – me pretending to be you
- Synthetic fraud – me pretending to be a completely artificial person

These types of fraud are easily affected through Gen AI, in the form of deepfake apps, which can be prompted to create authentic-looking fake ids with the desired face and data, or by generating the very authentic-looking face of a non-existent individual.

Deepfake examples include pornographic images and even video of celebrities, in one of the most disturbing practices of the new technology.

Banks and other financial institutions budget for a certain amount of fraud every year, some of them into nine figures, because

- A certain portion of their new applicants will be fraudulent
- A certain portion of those will get through, despite their defenses
- A certain portion of their existing accounts are sleepers who haven't activated yet, but will, and they're already in the door, in the form of synthetic identities.

Once again, those sleepers are easier to create now, using deepfake apps.

This is why it's critical to not only weed them out before they can be put into the system, it's also helpful to periodically sweep the existing rolls with more enhanced methods to find the bad guys, including the synthetic identities, before they hatch.

So the question is, how to spot deepfakes? Sometimes they can be detected by the human eye, especially in the form of video. Unnatural blinking, mismatched body parts or jewelry, distorted voices, odd expressions, these can all be earmarks of a deepfake.

For example, the Apple Vision Pro avatars are realistic-ish, but can be picked out by a discerning person.

But if all these deepfake examples were that easy to recognize with common deepfake detection tools, if we could train our platforms how to spot deepfakes, they could never be used for fraud.

***"AI enables fraudsters to commit their evil acts on an industrial scale"***

Too much data is out there that allows criminals to impersonate their victims. Create the profile, then create deep fakes to present the profile in the most realistic manner. Using easy chat prompts, fraudsters can feed this data into an AI generator and become their victim.

# HOW ARE DEEPFAKE ATTACKS LAUNCHED?

Quite commonly, criminals utilizing deepfake apps try to create bank accounts or other online presences by way of third-party fraud (pretending to be an actual victim) or synthetic fraud (pretending to be a phony identity they've created from scratch). They may also try to authenticate (log in) to existing accounts using deepfake facial versions of their intended victims. As organizations move to passwordless options for logging in, deepfake authentication is a true threat.

There are two standard attack vectors that are utilized by deepfake fraudsters who attempt to fool online systems such as account registration or authentication platforms.

The first is a Presentation Attack. The name is deceptively simple. I've got that fake person ready to go, and now I have to present it to the portal, the website, the application in a way that it's expecting a legit person to show up. The images are presented to the sensor, typically the camera on a smartphone or laptop, and the camera in turn presents the deepfake images to the systems or humans that judge the legitimacy of those images. If the image capture process isn't smart enough to detect the deepfake, then it is up to the backend processing system, or human reviewers, to catch the fake and prevent fraud from occurring. In a single factor login scheme, or even dual factor when one of those involves only a weak password, deepfake authentication can come into play through presentation attack.

The other typical path is an Injection Attack. Software that is meant to detect fakes at the point of presentation (i.e. camera) is bypassed by hackers who inject or insert the deepfake into the flow behind that sensor or camera. It is a form of man-in-the-middle attack. If the backend system or human reviewers aren't smart enough to detect the fake, then once again fraud occurs.

> ***"The ideal defense is a layered one, in which the organization employs multiple detection schemes"***

There are multiple approaches to preventing the success of deepfake software in these situations. Detecting the fake up front is of course the preferred method, but it requires sophisticated capabilities that often involve large, downloadable codebases. Such code is a hassle for consumers who will use that code only once, for account creation. Detecting the fakes on the back end is a risk, in that the fakes have penetrated beyond the first layer of the platform, and this allows for injection attack. The ideal defense is a layered one, in which the organization employs

multiple detection schemes that account for both presentation and injection attacks. More on this later.

One of the arts to deepfakes is the ability to feed Gen AI the appropriate prompts. For example, someone might tell their favorite deepfakes app to create a picture of “a smiling young child eating cotton candy.” But deepfake technology has evolved to the point where a user can prompt AI to generate entire deepfake videos from scratch.

# DEEP FAKES – THE HUMAN ELEMENT

Primarily the target of deepfakes are human. Fake faces, fake voices, fake ids that include fake faces. The faces can be completely faked, or even just manipulated, modified to express something that the original model did not intend.

Help desk and call center staff have been fooled into resetting passwords of legit accounts on behalf of illegitimate callers, and in some cases deepfake voices weren't even needed. But increasingly those fakes are being utilized. A study at the University of Chicago discovered that AI-generated deepfake voices could deceive three of the most widely-used voice recognition systems.

***"It's estimated that 500,000 deep fake videos were uploaded in 2023"***

Here's where fake faces and voices come together. In a too-crazy-to-be-believed story of the most egregious of all deepfake video examples, an unfortunate victim, an employee of a multi-national company in Hong Kong, was fooled over a Zoom call in which all the other participants, including a UK-based CFO, were generated with deepfake voices and faces. As a result of the call, he was directed to transfer over 25 million dollars to a criminal's account.

Deepfakes were already surging by 2022. It's estimated that 500,000 deep fake videos were uploaded in 2023, and it's further assumed that there will be an exponential rise in 2024 and 2025. Various deepfake laws have been enacted or proposed, including the Deepfakes Accountability Act (2019), although in some cases the First Amendment allows for satire and fair use that nullifies deepfake laws.

# OTHER PERILS POSED BY DEEPFAKE THREATS

Social media is naturally a breeding ground for deepfake technology. Manipulated images and videos are already in full force in attempts to sway voters. Starting in 2022, China began employing Spamoflauge, an influencing operation using fictitious AI figures in deepfake videos. We're even seeing deepfake politicians. The pope, as well as both current presidential candidates, have been targets of deepfake technology.

Face swapping is another product of deepfake software, in which Person A is filmed, but the face of Person B is overlaid, using video face replacement software. This was predicted in the movie The Running Man back in 1987 and has finally come to be. Through face swapping, some very deceptive content has been created for entertainment purposes, such as deepfake videos of Tom Cruise, Jerry Seinfeld, Spiderman actors, Emma Watson, Morgan Freeman, and others. Tom Hanks even appeared in a fake ad. Football star Tom Brady and others have appeared in fake endorsements via deepfakes. But it gets more nefarious from there.

Both buyers and sellers of real estate have been targeted by deepfake threats. A frightening example of deepfake scams is the use of deepfake voices to scare relatives into sending ransom money, after receiving calls from what they believed to be very real and very hysterical children. It's an old scam, but deepfake voices make it all the more convincing. Fake voices have also been leveraged for bank thefts. A Canadian CEO was scammed out of a quarter-million dollars via a deepfake voice.

***"Deepfake newspeople and popular figures can be used to amplify the dissemination of deepfake content"***

In another example of deepfake politicians, AI fakes have been used to try influencing people not to vote. Others have been used in trying to demonstrate fake political support. There has been deepfake war coverage of conflicts in Europe and the Middle East, as well as fake news anchors in Russia and China. In fact, deepfake newspeople and popular figures can be used to amplify the dissemination of deepfake content. Deepfakes are a peril for stock price manipulation, as well as brand reputation (via deepfake videos of company executives). Forrester has compiled a list of the various ways business organizations are endangered by deepfakes.

The United States federal government now recognizes deepfakes as true national security threats, as does the Department of Homeland Security. A popular TED talk also outlined how deepfakes pose a danger to democratic processes. Deepfake scams are one thing, but political deepfakes are a whole other matter.


Virus Detected

# WHO NEEDS PROTECTION FROM DEEPFAKE THREATS?

Virtually any organization that serves consumers needs defending from deepfake scams. While most decent-sized enterprises train their workforce on data protection and compliance, they don't teach them how to spot deepfakes.

They certainly can't wait for effective deepfake laws to be enacted on a timely basis. AI-generated deepfakes threaten privacy, financial assets, privileged access, brand reputation, and the integrity of the civic process. Not only must consumers and citizens be safeguarded from deepfakes that might be leveraged to compromise their accounts and assets, so too must the administrators whose access can be stolen by fraudsters who would use that access to commit all manner of crimes.

***"Thieves can steal your credentials… But they can't steal the real you"***

In general, any organization needs to both onboard users on Day Zero, and authenticate them everyday thereafter. But these use cases apply to any number of enterprises that are threatened by deepfakes. And their user bases include both employees and consumers.

Deepfakes can be used for accessing bank accounts, payment applications, peer-to-peer networks, lenders, providers of digital wallets, credit unions, money transfer accounts, and other financial services.

Employees of gaming companies and physical facilities have been the targets of deepfakes as well as imposters who have used deepfakes to get privileged positions, with which they've committed large scale thefts.

Deepfake-based phishing and other attacks have allowed criminals to install malware and ransomware at hospitals, universities, and even grade schools, disrupting learning and even critical healthcare facilities.

The hospitality industry has been robbed of services, and individual consumers have seen their loyalty accounts breached and their points stolen because of deepfake technology. Those who trade in crypto need to protect an asset that promises to increase in value in the coming years.

Help desks / call centers are useful conduits for criminals who fool staff into granting access and resetting passwords. Healthcare organizations are honey pots full of personal data. Besides that, they are increasingly targeted by ransomware perpetrators who often infiltrate systems through compromised credentials of administrative users. Schools are also attacked with ransomware more and more.

# FIGHTING DEEP FAKES – THE OLD WAY

The techniques that have long been used to combat different kinds of fraud ultimately fail against the onslaught of very good and very voluminous deepfake software. Deepfake detection tools are not designed to keep up with the pace, nor the quality of deep fake identities. That's not to say that deepfake solutions are totally useless, because in the end a fake is a fake, and even the improved type won't always pass. But they definitely make it tougher to filter out the most insidiously constructed phonies.

Despite the advancements is automated identity checks, plenty of organizations still use manual checks of identity, including reviewing the data that's been submitted, and picture of physical ids and other documents (utility bills, credit cards, etc.). One might think that real eyeballs would be safer than putting one's faith in code. But there are plenty of drawbacks to this approach

- It's expensive
- It's time-consuming, since there is always a lag
- It exposes this data and these sensitive images to low-paid folks in a boiler room, often in a faraway country
- Humans can't always spot the fakes, because the signals are terribly subtle, and only getting subtler

(By the way, these same kinds of boiler rooms have long been used to torture consumers with calls claiming to be from Microsoft or the "department of taxes" or "the ministry of revenue.")

One well-known identity verification product is employed by a major airline for international check-ins online. And what do they provide? After electronically submitting an image of your passport, that image is examined by very real humans, meaning that 1) the process is subject to very real human error, and 2) your very precious documents are in the hands of strangers.

***"They all start with passwords, which are responsible for more than 3/4 of all individual breaches"***

Automated verification systems in general have less than overwhelming abilities, and often suffer from acceptance rates that are far below reasonable expectations. Deepfake physical ids, conspiring with deepfake faces and voices, often successfully pass such systems.

Data brokers whose stash of names, addresses, emails, social security numbers and other information is meant to be used for verifying identity, also serve as honey pots for thieves who breach those stores and use that data for the very theft it's meant to prevent.

Device-based authentication is meant to leverage something a creator of deepfakes does not have: something that is in the user's own hand. But there are multiple ways to compromise physical devices, and verification apps can be fooled by deepfake authentication by way of counterfeit faces and voices.

Multi-factor authentication is meant to provide extra layers of security to access rights, but in the end, they all start with passwords, which are responsible for more than three-quarters of all individual breaches.

If all these traditional methods of identity proofing and authentication were deepfake-proof, or if various deepfake solutions were as viable as advertised, we would not be seeing the many horror stories of large-scale breaches through AI.

# FIGHTING DEEP FAKES – NEW TECH THAT STAYS AHEAD OF THE BAD GUYS

Governments can't easily combat deepfake technology head-on, so they do the one thing they can do: enact legislation. It's only in the last couple of years that the government has stepped in to propose deepfake laws that recognize the new reality of Generative AI as a force for evil, but no unified legislation is being pushed. This means that business enterprises are now largely on their own for combatting deepfakes. So how to fight presentation attacks? How to guard the front door and say, "It looks and smells like the real thing, but it's not?" The id looks good, the face and voice look and sound good, but …

***"Business enterprises are now largely on their own for combatting deepfakes"***

Thieves can steal your credentials. They can steal your id or fake it. They can even steal your device. But they can't steal the real you. Still, how can they be sure it *is* the real you, in the age of counterfeits so easily created by deepfake apps? What are the best kinds of deepfake detection tools?

This is done via biometrics, and liveness detection. And liveness not just for the person, but for the id. Biometrics involves evaluating the human factors, the things the thieves cannot take away. But these come with their own issues:

- Voice is now far too easily cloned. And it is easily injected at the actual point of entry, simply by playing that voice into the smartphone or laptop.
- Fingerprint is no longer supported on various platforms, including the manufacturer of the most ubiquitous smartphone on the planet.
- Some facial recognition works against large databases of known faces. But that means putting them all in one place, which provides yet another honeypot, just like all those breached data brokers. And faces can be manipulated as well, as in deepfake videos. But they're far harder to inject. And this is where liveness detection comes in.

# FIGHTING DEEP FAKES - WITH BIOMETRIC AUTHENTICATION FROM AUTHID

There's no getting around the fact that deep fakes will fool a lot of people - so the tech has to be even smarter. And that's what authID has concentrated on: a platform that is fast, accurate, user-friendly, and above all, smarter. This is how we service both onboarding and authentication, for consumers as well as the workforce. We give our customers the confidence of knowing that we will detect deepfake attempts on their portals.

***"Deep fakes will fool a lot of people - so the tech has to be even smarter"***

Our solution looks at every aspect of an identity transaction as an opportunity. We have several chances to enable a legit applicant to access the digital assets: validate that physical id document, validate the live person, then match them together. Of course, these are also opportunities to uncover fraud. And we do all of this through a multi-layered approach.

For first-time identity verification, or proofing:

1. authID prompts for photos of the physical document and the user
2. We perform both client side and back-end analysis of the images to detect deepfake attempts
3. We perform this analysis within 700 milliseconds, the fastest and most accurate process on the market
4. We detect deepfake images via digital signing of images, and pull apart physical documents multiple ways to determine their viability
5. We employ NIST PAD 2 liveness detection to determine if the selfie is provided by a live human, and prevent presentation attacks
6. We stop injection attacks, meaning preventing fraudsters from inserting fake images into the process thru hardware, software, or network attacks
7. And all this happens with a simplified, friendly user experience that accommodates even the least tech-savvy individual

To accommodate the various use cases across several verticals, authID provides a number of integration capabilities that are well-documented for our customers and partners.

For day-to-day authentication, authID requires only a selfie, and again employs the same liveness detection, defense against presentation and injection attacks, and of course this is all done through that same fast, accurate, and user-friendly interface.

We store only the biometric hash of each face. The entirety of data from each transaction is encrypted and accessed (only by authorized personnel) through our administrative portal. If preferred (for privacy or compliance purposes), that data can be pulled down by our customers for their own storage, and deleted from ours.

By locking out deepfake authentication, authID prevents account takeover, meaning situations where existing user accounts are compromised because a thief stole something the user knew (credentials) or something the user had (device). authID's solution is not device-bound. Biometrics are stored and evaluated in the cloud, so that a thief cannot leverage a purloined phone. This also means that the user can authenticate from any device and is not locked out when their phone or laptop is stolen, lost, broken, or upgraded. In the event an account is actually compromised, the user again leverages their own face and can recover their access.

authID is perfect for help desks. Some major intrusions, including two multi-million-dollar events, took place in 2023 because help desks reset passwords for imposters over the phone. Rather than simply fat-finger someone's access, the help desk sends a link to that user to initiate an authID transaction and use facial biometrics to recover their account. The help desk has helped, but the user still needs to prove their identity.

Many multi-factor authentication schemes still rely on at least one step utilizing passwords. This means that the entire process is subject to the weakest link, meaning compromised passwords that still account for the vast majority of breaches. By leveraging biometrics, organizations can eliminate that weak link while ensuring that only real, live, legitimate users are accessing their most valuable and sensitive digital assets.

***"Compromised passwords still account for the vast majority of breaches"***

Check out our blogs for a cross-section of how we think about the industry, about deepfakes, and about how authID solves the onboarding and authentication problems common to tech infrastructures.

authID truly provides the fastest, most accurate, and most frictionless solution for identity proofing and daily authentication on the market. Drop us a line, and we'll be happy to show you how we do this.

Have a look in the mirror. The future of identity verification is staring you in the face. ❖