Deepfake It ‘til You Make It: A Comprehensive View of the New AI Criminal Toolset

Deepfake It Till You Make It: A Comprehensive View of the New AI Criminal Toolset

This report takes a comprehensive look at how deepfakes are used to support criminal business processes, what are the toolkits criminals are exploiting to power their deepfake creation, and what the deepfake underground looks like.

By David Sancho, Salvatore Gariuolo, Vincenzo Ciancaglini

Tools for creating deepfakes are now more powerful and more accessible by being cheaper and easier to use. Criminals can easily generate highly convincing deepfakes with very little budget, effort, and expertise, and deepfake generation tools will only become more affordable and more effective in the future.

While criminals have been attempting to extract value from deepfake-enabled social engineering and misinformation well before generative AI applications hit the mainstream, this technology is helping malicious actors make money in new and increasingly effective ways from consumer extortion to enterprise scams. We discuss both how companies are being targeted and how deepfakes are being used against individuals.

We also examine the toolkits powering deepfake creation, which shows how criminals no longer need to rely on custom-built underground services. Instead, they can simply exploit commercial tools that were not intended to be used maliciously. This report also explores the ongoing conversations in the criminal underground: from tool recommendations to the playbooks that threat actors actively use to deceive, manipulate, and defraud. This will provide the reader with a clear perspective of where the criminal focus currently lies.

Deepfake-enabled cybercrimes

Deepfakes in all their incarnations have been around for a relatively long time now: in 2017, the technology community experimented with tools like DeepFaceLab, which became a staple the following year. Since it became accessible to the general public, we have witnessed cybercriminals using deepfake technology, from fake propaganda videos spreading misinformation to cryptocurrency scams, and even attacks targeting specific individuals or organizations, as in the case of the first deepfake audio-enabled CEO scam in 2019.

Criminals leverage deepfakes differently to attack enterprise targets compared to individuals. While cybercriminals often tailor their attacks to specific organizations, they tend to use more generic

deepfakes when targeting individuals, making them more adaptable and accessible to a wider range of targets. In this section we compare how criminals customize their use of deepfakes depending on their targets and identify the different modus operandi in use.

Deepfake-enabled enterprise attacks

Deepfake-enabled attacks against corporations are typically highly targeted, requiring prior knowledge of the victim, extensive reconnaissance, and meticulous planning that takes time. Such attacks usually require either finding the right victim employee to enact the fund transfer, or learning about the HR department's hiring practices.

Deepfake-enabled cybercrime targeting consumers

Unlike enterprise attacks, consumer attacks tend to be broad and indiscriminate, requiring little to no prior information about the target victim. However, we have also started to observe a rise in more personalized attacks against individuals, which we will discuss in this section.

Click through the red plus symbols in the mobile to learn about different deepfake-enabled cybercrimes targeting consumers

In summary, criminals can easily generate a nude picture of a person, indiscriminate of age, with readily available and very affordable tools. It's worth noting how challenging it can be for these images to be taken down by victims, more so for celebrities. Also, many parts of the world lag behind in legal protections for the public and criminalizing such acts. Deepfake child abuse material also poses an additional challenge for law enforcement as attempting to locate abused minors who don't exist can take up time and resources away from efforts to save those that do.

	Broad scope	**Targeted scope**
Enterprises	• KYC bypass	• Business email compromise • Employment scams
Individuals	• Fake advertisement • Romance scam* • Sextortion* • Child pornography	• Virtual kidnapping • Stranded traveler

Table 1. Deepfake attacks by victim type and breadth of scope

Deepfake technology has shifted from a niche experiment to a powerful weapon in the hands of cybercriminals, impacting both individuals and organizations. Table 1 shows how the different deepfake-enabled cybercrimes are divided in terms of breadth of scope and type of victim. It is worth pointing out that some broad attacks like romance scams and sextortion can also become targeted.

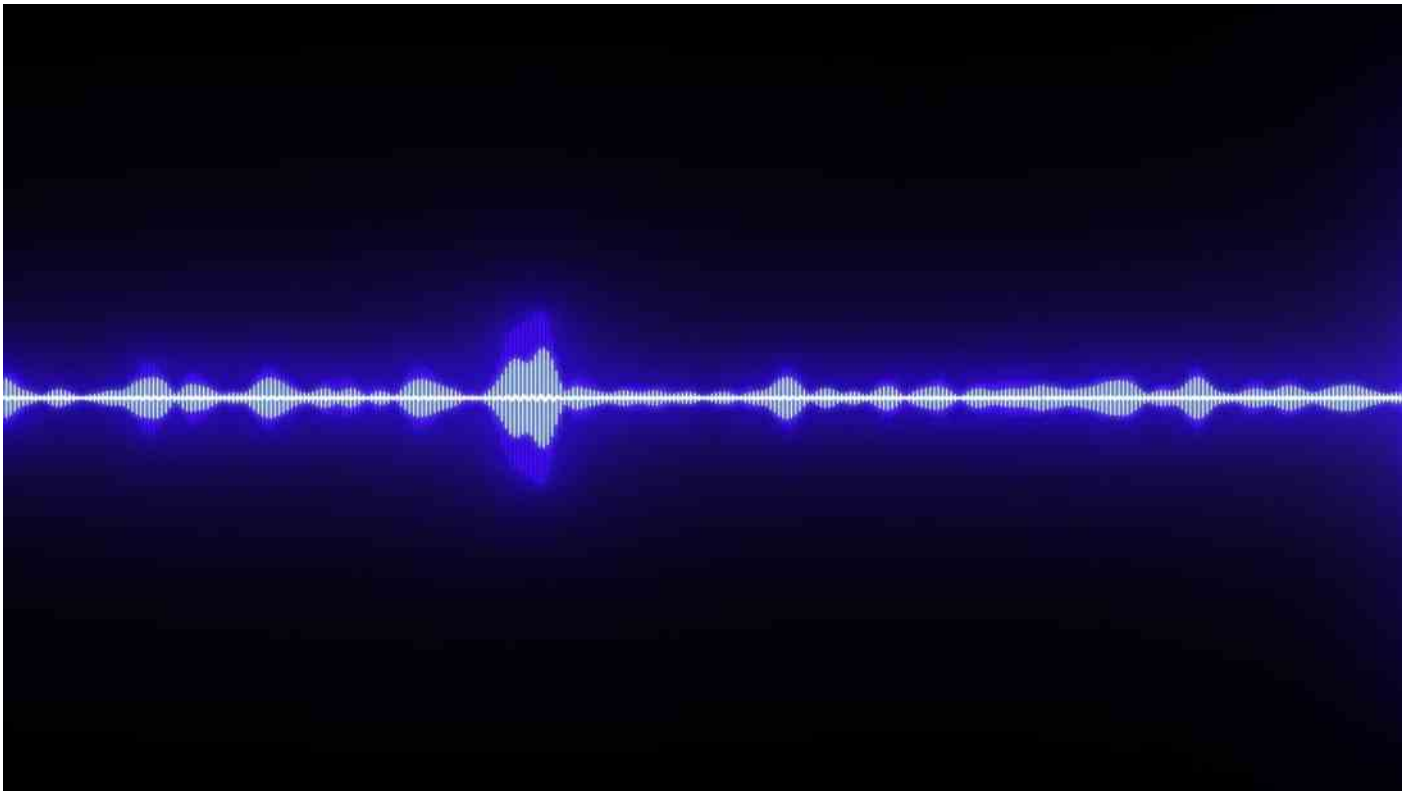
In our monitoring of the criminal underground, most of the criminal effort is directed towards targeting companies and organizations. However, it is just as important to mention consumer attacks as these deepfake-propelled targeted attacks such as sextortion and voice cloning can be easily modified to also target and compromise corporations.

No sweat deepfakes

In today's digital marketplace, creating deepfakes has become remarkably easy. Countless tools are now available for generating audio and video deepfakes, some even supporting real-time face swapping during video streaming. Mainstream AI-based image generators allow users to create hyper-realistic portraits of people who don't exist, or to fabricate convincing images portrayed as real individuals. More nefarious tools go further, such as generating nude images from photos of real people, or of fabricated persons.

Deepfake generation tools are often bundled into convenient all-in-one suites, but while some platforms market themselves as comprehensive, one-stop solutions for generating synthetic media, others focus on a single domain - like audio, video, or images - and often deliver more convincing results. These services will be the focus of the next subsections.

It should be noted that platforms for synthetic media generation, like any other technology, are not inherently good or malicious. While these tools are designed to generate content for legitimate purposes, there is always potential for misuse, which will also be discussed in this section.



Deepfake audio

The first thing we noticed in our analysis is that the market for AI-generated voice technology is extremely mature, with numerous services offering voice cloning and studio-grade voiceovers. These services have revolutionized this industry because they eliminate traditional barriers to producing high-quality audio, such as the need for recording equipment, a quiet environment, or even the speaker's physical presence. For example, Speechify, a subscription-based platform priced at $9.99 a month, enables anyone to create audiobooks. It relies on voice cloning and text-to-speech tools, allowing users to generate a narrated version of a book in their own voice, without ever stepping into a recording studio.

Meanwhile, other platforms like Play.ht offer APIs to integrate voice synthesis with conversational AI. This enables businesses to build natural-sounding, interactive AI agents that can speak to users in real-time, a feature that is particularly useful in sectors like customer support.

Although these services have many legitimate applications, their potential for misuse cannot be overlooked, and providers should be mindful of how their products are both presented and used. *Play.ht*, for instance, markets their lower-tier API plan, available for just $5 a month, under the label "Hacker Plan," a name that sounds catchy but could also raise concerns given the risks associated with the improper use of synthetic audio.

On the other hand, there are companies like Resemble AI that take a more cautious and responsible approach by providing custom voice generation, voice cloning, and text-to-speech features, with an

added layer of protection through neural watermarking and adherence to the Coalition for Content Provenance and Authenticity (C2PA) standard. This makes it possible to embed metadata in the audio clip, adding a digital certificate that helps verify that the content was generated using AI and trace its origin.

Despite these safeguards, deepfake audio remains a concern in that most voice synthesis services support multilingual output, a feature that, in the hands of bad actors, becomes a potent tool for deception. Combined with the ability to control pronunciation, intonation, and emotion, malicious users can craft persuasive and emotionally manipulative audio clips in different languages, which can be used across geographical borders and cultural boundaries.

Even more concerning is how easy it has become to generate audio deepfakes. Many services now offer one-shot voice generation, where just a few seconds of source material is enough to create a convincing replica of someone's voice.

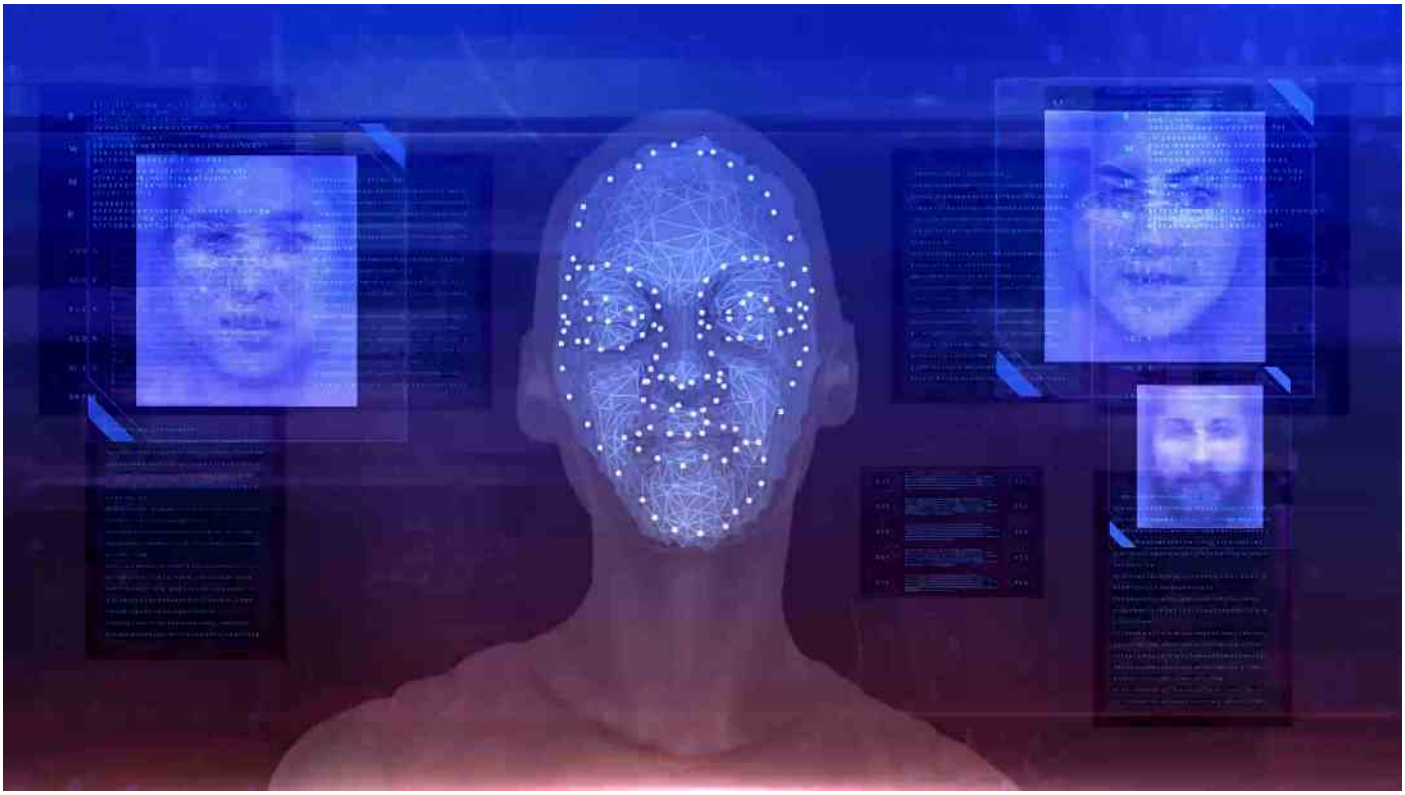
Additionally, the barrier to entry is surprisingly low. While it is true that higher audio quality typically comes with a higher price (top-tier voice synthesis services cost upwards of $1,000 a month), many platforms offer decent output starting at just $5. This makes these tools accessible to anyone, regardless of their budget.Table 2 shows a breakdown of some of the services available on the market for creating audio deepfakes, along with their features and pricing. Note that the prices are based on offers at the time of writing.

Service	Capabilities	Key Features	Pricing
Eleven Labs	Voice cloning, Voice changer, Text-to-speech	Multilingual support, One-shot-generation, Scalable API integration	Free plan; Paid plans from $5 to $1320/month
FakeYou	Voice changer, Text-to-speech	Zero-shot voice cloning, Zero-shot voice conversion, Community-generated voices, Strong character/meme focus	Free plan; Paid plans from $7 to $25/month
Google Cloud Text-to-Speech	Text-to-speech	High-fidelity voices, Multilingual support with 50+ languages and variants	Free tier allows 4M characters/month; Paid plans from $4 to $160 per 1M characters
IBM Watson Text-to-Speech	Text-to-speech	Real-time speech synthesis, Customized pronunciation, Expressiveness control, Controllable speech attributes	Free tier allows10K characters a month; Paid plans from $2 per 100K characters

Lovo AI (Genny)	Voice cloning, Text-to-speech	Pre-set voices, Multilingual support, Emotional tone control, Video editor integration	Paid plans from $24 to $149/month
Murf.ai	Voice changer, Voice cloning, Voice dubbing, Text-to-speech	Multilingual support, Pre-set of 200+ voices	Free trial; Paid plans from $19 to $199/month
Play.ht	Voice changer, Text-to-speech	Multilingual support, Pronunciation and inflection control via SSML, Mobile-friendly	Free trial; Paid plans from $5 (hacker) to $999 (growth) per month
Replica Studios	Voice changer, Text-to-speech	Multilingual support, Extensive voice library, Customizable pitch and tone	Free plan; Paid plans from $10/month
Resemble AI	Voice cloning, Voice synthesis from text prompts, Text-to-speech	Multilingual support, Emotion and tone control, Real-time output, Neural watermarking and adherence to C2PA standard	Free trial; Paid plans from $5 to $699/month
Speechify	Text-to-speech with voice generator, Voice dubbing, Voice cloning	Audiobook creation service, Optimized for reading/listening, Mobile-friendly	Text-to-speech is $11.58/month; Audiobook service is $9.99/month

Table 2. Some of the services available on the market for creating audio deepfakes, along with their features and pricing.

Given the vast amount of audio clips publicly available on social media or in video content and how accessible the creation of deepfake audio has become, the threat of voice theft is no longer something we can afford to underestimate.



Video deepfakes

AI-driven video generation platforms are rapidly gaining popularity as they make it easier to produce professional-looking video content. But that same accessibility also opens the door to misuse: with minimal effort, technical skill, or resources, anyone can now create convincing and potentially harmful synthetic videos.

Synthetic video platforms are a game-changer particularly for content creators. They enable anyone to produce engaging, high-quality videos, even individuals who are extremely shy or have limited experience presenting on camera.

For instance, Argil AI allows users to generate videos directly from text by generating a clone of the user from a short sample video, learning how they move and speak, eliminating the need to appear on camera entirely. While this is incredibly useful for content creation, a monthly subscription of only $39 could also allow for a malicious user to easily generate convincing fake videos of someone else, with full control over their body language. This can easily be done to create videos of business executives giving false directives or engaging in unethical behavior that could seriously damage public trust and corporate security.

Another example is AI Studios, a suite designed to help content creators generate material for platforms like YouTube, which offers services such as document-to-video generation, avatar creation, voice cloning,

and AI dubbing. What stands out is the accessible pricing; both platforms operate on a freemium model: a limited free plan and tiered paid subscriptions start at $19.99 and $24.00 per month, respectively.

Synthetic video platforms are also remarkably easy to use: such as Reface.ai which offers a mobile app that allows users to swap faces in GIFs and short videos, which can then be uploaded to social media directly from the app. These tools are intuitive, making synthetic video creation more accessible than ever. However, there are open-source alternatives like DeepFaceLab and Deepfake_tf that offer a more powerful, but also more complex approach to synthetic video generation that swap faces in video clips using deep learning frameworks, often leveraging neural networks like GANs and CNNs to boost realism and editing capabilities. While free, these open-source tools also require a certain level of expertise and powerful hardware. But this doesn't eliminate the risk of misuse: it simply limits the potential for exploitation to bad actors with more advanced skills.

Vidnoz Face Swap is another platform that supports AI voice cloning, text-to-speech, and video generation with personalized avatars. But the services this company offers do not stop there, the platform also allows users to generate questionable content, such as kissing videos created from real photos or simulated footage involving children, all with a few clicks.

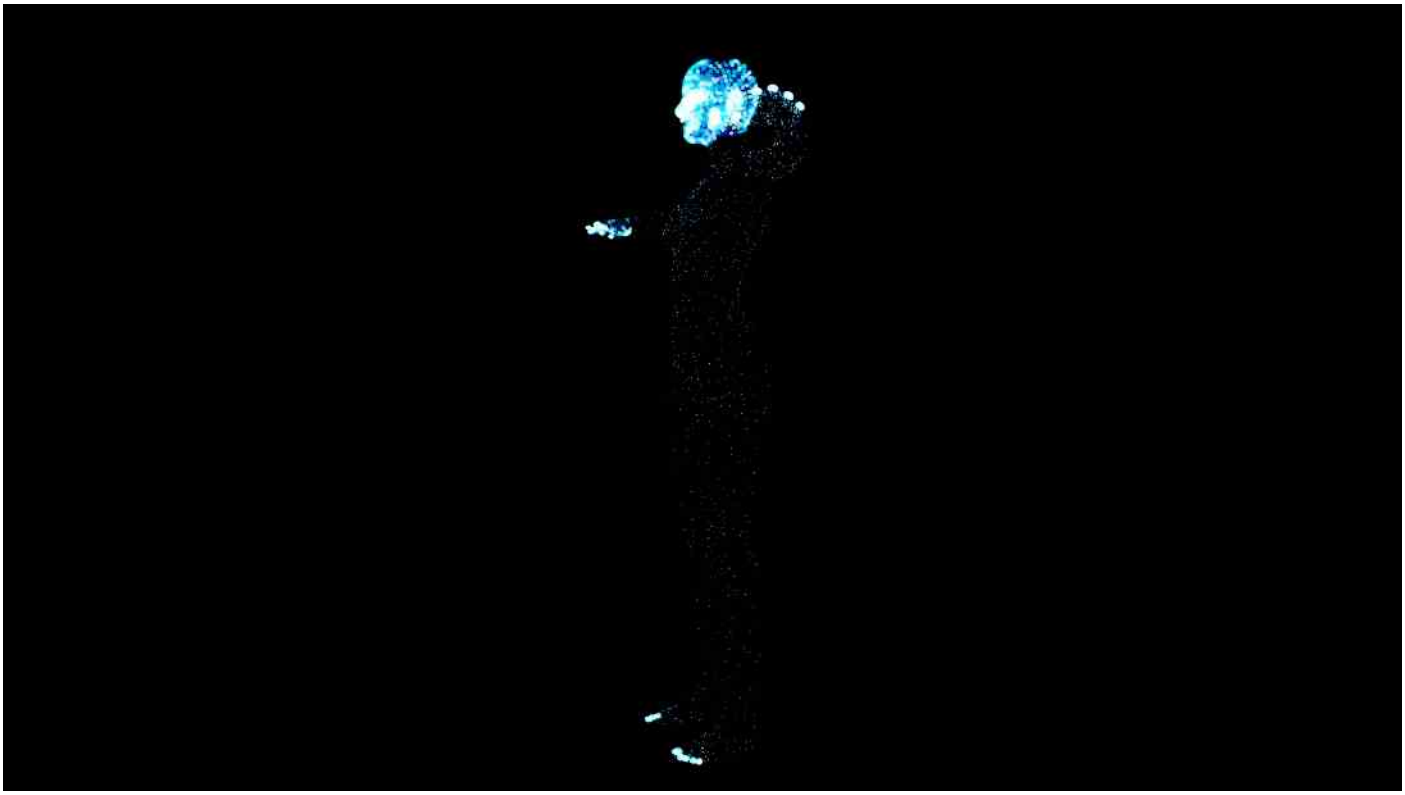
Another significant barrier to synthetic video creation is also being removed. Powerful systems were once a prerequisite for generating high-quality synthetic media. Now, with the advent of cloud-based services using GPU clusters to synthesize video, this obstacle has largely disappeared. One example is Deepfakes Web, which eliminates the need for dedicated computing resources as it allows users without high-end hardware and with limited skills and resources to produce realistic synthetic video content.

There is also a particular class of video generation services that can become especially dangerous when used for malicious purposes: those that allow real-time manipulation of video streams. Some platforms, like Avatarify, enable real-time face swapping during video calls. Others, like Synthesia and HeyGen let users create interactive AI clones from real footage and voice recordings, which can then join video conferences. In the wrong hands, these platforms could become powerful tools of deception, enabling malicious actors to infiltrate corporate meetings by impersonating trusted individuals or creating entirely fabricated personas. These services are available at every budget level. While some platforms operate on freemium models, others cost several hundred dollars per month. The higher price may act as a barrier of entry for some malicious users, but it also tends to reflect better performance, which can make deepfakes even harder to detect.

For criminals, even the most expensive services are only a fraction of the potential return on investment from a successful attack, making them well worth the cost.Table 3 lists some services that offer AI-generated video content creation. Note that the prices specified are based on offers at the time of writing.

Service	Capabilities	Key Features	Pricing
Argil AI	Text-to-video synthesis using a short sample clip	Full control over body language and camera angles	Free plan; Paid plans from $39/month
Avatarify	Real-time face swapping in video calls	Open source, Integrates with video conferencing apps like Zoom	Free
Deepfake_tf	Synthetic video generation, face swapping in videos	Open-source, TensorFlow-based, Customizable training models	Free
DeepFaceLab	Synthetic video generation, Face swapping in videos	Open-source, TensorFlow-based, GPU-accelerated	Free
Deepfakes Web	Face swapping in video	Cloud-based service, Digital watermarks, “Imperfect by design” policy	$10/video
DeepSwap	Face swapping in images and videos	High-quality output, Multi-face video swap, Handles obstructed faces	Paid plans from $19.99/month
FaceMagic	Face swapping in video	User-friendly mobile app, Cloud-based service, Watermark removed with paid plan	Free plan; Paid plans from $9.99/month
HeyGen	AI avatar generation	Specialized in unscripted conversation	Free plan; Paid plans from $29/month
Reface	Face swapping in GIFs and short videos	User-friendly mobile app, Social media integration	Free with in-app purchases
Synthesia	AI avatar generation	Multi-language support, Can be used in video conferencing apps	Free plan; Paid plans from $23/month

Table 3. Some services for AI-generated video and video streaming manipulation.



AI-generated images

AI-powered image manipulation has become so advanced and accessible that it is no longer confined to skilled actors or niche communities. There are various deepfake image-generating services that offer face-swapping capabilities for images as well as for video clips and streams.

DeepSwap provides this type of service, but it stands out for the quality it delivers, particularly when it comes to still images. Their platform produces convincing results even in what should be challenging scenarios involving various facial expressions and emotions, as well as different camera angles and lighting conditions. The effectiveness of their service could be misused by a malicious user to place someone in a situation they have never been in, or to superimpose a new face onto a digital document, making it a powerful tool for disinformation and deceit.

Although less relevant in enterprise settings, a particularly alarming form of AI-generated images that are worthy of discussion are “deep nudes.” Several nudifying services have recently emerged, allowing users to digitally undress individuals in photos. Nudify.Online and XPicture.ai are two examples: both platforms rely on proprietary engines built specifically for this purpose, with the latter even allowing users to upload pictures directly from Instagram. UndressAI.Tools and Nude Fusion go even further: the former enables the creation of pornographic images from real photos, while the latter allows users to swap faces onto a set of available pornographic scenes.

Many of these platforms include ethical disclaimers or usage advisories on their websites, but it should be said that such warnings do little to stop users from exploiting them for malicious purposes. Anyone can pull a photo from a person's social media profile and uploading it to one of these platforms to be used as a weapon for extortion, humiliation, or harassment. Consequences can go far beyond reputational damage and can undermine a person's sense of identity and beliefs, seriously affect their well-being and cause devastating psychological impact.

Most concerning is how accessible these services are: many platforms offer free trials or free limited plans, and even the paid versions are relatively affordable, with subscriptions ranging from just $9.99 to $22 per month. As these tools continue to evolve and become even more accessible, the potential for personal misuse and harm will naturally grow alongside them. When anyone can weaponize a picture with just a few clicks, the risks become impossible to ignore. Table 4 lists some examples of these services, their features, and their respective pricing. Note that the prices are based on offers at the time of writing.

Service	Capabilities	Key Features	Pricing
Nudify.Online	Photo nudifying	User-friendly interface, Drag-and-drop upload	Free
Nude Fusion	Face swapping in NSFW scenes	Image and video presets for face swapping	Free trial; Paid plans from $22/month
PornWorks AI	Text-to-image generation, Text-to-video generation, Photo nudifying, Face swapping in NSFW images and videos	High-quality images, Private storage	Paid plans from $2.99 to $14.99 /month
UndressAI.Tools	NSFW image generation from real photos	NSFW mode presets	Free plan; Gem-based pricing starting at $19.99/bundle
XPicture.ai	Photo nudifying	Instagram photo import,Age and ethnicity presets	Free plan; Paid plans from $10.90/month

Table 4. A list of some applications that allow for swapping images onto nude bodies and explicit content generation services.

Deepfakes in underground discussions

The criminal business plans involving deepfake images, audio, video, and video streaming develop and improve scams and other criminal endeavors for a relatively small investment in skill and money. This

makes them an ideal cybercrime enabler. Synthetic media generation services are easy to use and there are many tools already available on the internet that produce high-quality fakes at very affordable prices.

On a survey of the criminal underground in 2024, we found some criminal deepfake creation tools that are no longer in use. This could be attributed to the wide availability of general-purpose deepfake creation software that makes criminal-specific tools obsolete.In our observation of the criminal underground in 2025, malicious actors are already discussing how to use general–purpose and widely available deepfake creation software as their toolkit of choice. We have seen deepfake creation tutorials and playbooks that teach other criminals how to take advantage of open- and close-source tools on the internet to accomplish criminal objectives.

As early as November 2024, we observed an actor on a criminal forum sharing a playbook to create video deepfakes on Deep-Live-Cam, a free open-source tool that uses ffmpeg as the multimedia library and Visual Studio to compile the sources. The tutorial is technical and very detailed, so that any reader can follow the steps and create deepfake videos for free. It also includes the option to use video streaming driver Open Broadcaster Software (OBS) to change the webcam to superimpose a deepfake on top of the criminal's face.

In early December 2024, a criminal on an underground forum was offering a tutorial and playbook that provided services and tools to facilitate successful KYC verification on cryptocurrency exchanges, dating and financial applications, and online casinos. The cybercriminal's tutorial allegedly included:

- A detailed manual to bypass KYC checks and identity verification services using a web browser or a dedicated application. This included configuration files, text tutorials, training materials and videos.
- A licensed version of the *VcamPro* aka *Vcam* virtual Android camera application.
- A fake iPhone Camera application for iOS versions 15 and 16.
- A private tool to make deepfake videos, which only supports Windows operating systems and requires a GeForce RTX 2070 graphics card or more advanced technology.
- Services to configure OBS, OBS Studio for video recording and live streaming.
- Services and advice to bypass KYC verification for accounts that customers provided.

In January 2025, another actor offered on-demand service to create deepfake images, specifically geared towards defeating KYC checks. This person shared the toolkit they used: they mentioned the open-source tools DeepFaceLive and DeepFaceLab, along with the voice cloning tool AI Voice Changer. The actor also mentioned that he had been able to bypass KYC checks from two different cryptocurrency exchanges.

We have generally observed that KYC bypass seems to be the biggest application for deepfake creation, likely because it allows cybercriminals to open anonymous cryptocurrency exchange accounts that they can use to launder money.

We have also seen in the last few months how law enforcement agencies have arrested criminal gangs that specialize in deepfake-enabled scams. In October 2024, the Hong Kong police arrested 27 individuals who focused exclusively on deepfake-enabled romance scams. Similarly, in 2025, the Spanish police arrested six people that operated a global investment scam that was powered by ads featuring deepfakes of national celebrities. Deepfake-enabled scams are quickly becoming very fashionable in the criminal world, and this seems to be accelerating.

This should prompt the companies developing these platforms to implement appropriate safeguards: user identity verification, justification of use when generating content with someone else's likeness, real-time prompt filtering, and adding fingerprinting information to AI-generated media are just some of the ways to safeguard against potential misuse and abuse. These platforms should also block output that violate their own policies and keep comprehensive server-side logs, ensuring a trail of evidence in case of abuse.

While there are safeguards in some platforms such as watermarked generated content, these are often automatically removed in paid plans. In practice, for a fee, users gain the ability to generate synthetic media with no embedded signatures. This takes away one of the few reliable mechanisms we can use to distinguish real from fake, which is particularly important when AI-generated content is used to exploit or harm people.

Conclusion

Deepfake technology is now more accessible and affordable than ever, and it's only going to get cheaper and more advanced as time goes on. Unfortunately, the risks tied to this technology are growing just as quickly: from corporate espionage to identity theft, financial fraud, and personal extortion, enterprises and individuals are both at risk.

More alarmingly, cybercriminals are increasingly relying on legitimate, inexpensive, and easy-to-use services. Malicious actors not only know how to use these tools but also share tips and tricks on underground forums to refine their methods, making it even easier for more bad actors to exploit this technology for nefarious purposes.

While we have made recommendations for providers to secure their tools from abuse, it's crucial to remember that deepfake technology is already widespread; when even one provider fails to respond appropriately, the consequences can be severe and irreversible.

Ideally, as AI-generated media proliferates, society will become more aware of its potential for misuse and the associated dangers. But this growing awareness brings with it a deeper, more insidious threat: the erosion of trust. As deepfakes become more widespread, they blur the lines between fact and fiction and thus undermine our ability to trust what we see and hear, ultimately eroding the foundation of how we acquire knowledge about the world.

This means we must accept that all content could be fake by default. The recent mindset of “trust but verify” might not be enough any longer. In other words, we are moving away from a world where we can trust what we hear and see and towards a “zero-trust” reality. It is a world where content must prove it is authentic and is assumed suspicious otherwise.

As we can see, simply being aware that content might be AI-generated will no longer be enough. As trust dwindles, people will begin to dismiss genuine content and legitimate communication. To stay protected, we must adopt tools for detecting deepfakes.

Trend Vision One™ AI Security has a deepfake detection technology that uses a variety of advanced methods to spot AI-generated content. Going beyond techniques like image noise analysis and color detection, the platform also analyses user behavioral elements to provide a much stronger approach to detecting and stopping deepfakes. Upon detection, Trend immediately alerts enterprise security teams, enabling them to learn, educate, and take proactive measures to prevent future attacks. It can help verify if a party on a live video conversation is using deepfake technology, alerting users that the person or persons with whom they are conversing may not be who they appear to be. This capability is also available for consumers today in Trend Micro Deepfake Inspector.

While these solutions are not foolproof, they can raise the bar for criminals, making deepfake attacks less likely to succeed and, over time, help restore trust in digital content. While no one is fully immune, not everyone has to become a victim.