Deep Fakes
Understanding the illicit economy for synthetic media

Author: Rob Volkert
Contributions: Michael Eller, Dev Badlu, and Kyle Ehmke (Threat Connect)

Are deep fakes actually being weaponized?

In this paper we will examine the illicit ecosystem for deep fakes, their technology evolution and migration paths from surface web to deep and dark sites, and uncover some of the actors creating and disseminating these videos. Nisos undertook research into deep fake technology (superimposing video footage of a face onto a source head and body) to determine if we could find the existence of a deep fake illicit digital underground economy or actors offering these services. Our research for this white paper focused specifically on the commoditization of deep fakes: whether deep fake videos or technology is being sold for illicit purposes on the surface, deep, or dark web.

Our research found that indeed deep fakes are being commoditized but primarily on the surface web and in the open. Most deep fake commoditization appears to be focused on research and basic cloud automation for satire or parody purposes. The underpinnings of the illicit ecosystem is nonconsensual face swapping pornography where commoditization happens primarily via ad revenue and subscription fees. While we expected to find a community, we did not find evidence of a marketplace (selling as a service) for e-crime or disinformation purposes. We assess the lack of an underground economy for uses other than satire, parody, or pornography is due to the resource and technological barrier to entry, lack of convincing quality in the videos. In short, the dark market is not yet lucrative enough.

As deep fakes become easier and quicker to create, social media platforms (currently the primary vehicle for deep fake dissemination) and users will simply not be able to keep up with the volume of synthetic manipulated videos or images. Instead, this will require a partnership between on-platform (social and traditional media), off-platform (cyber security and watch dog groups), and public sector entities (governments and private firms). We see this partnership nexus as having maximum impact in the following three areas:

- **Public accountability and information sharing:** Which platforms are being targeted and why?
- **Threat actors and intent:** Who are the actors actually creating and deploying deep fakes?
- **Action:** What can be done to stop illicit use of deep fakes?

What are deep fakes?

A deep fake superimposes existing video footage of a face onto a source head and body using generative adversarial networks, or GANs, which is generative modeling using deep learning or AI methods.[1]

This technology is readily accessible to the public, primarily via GitHub, online messaging platforms like Discord, research labs, and deep web interest forums. Most deep fakes involve facial manipulation; for example, a deep fake could appear to be a real person's recorded face and voice but the words they appear to be speaking were never really uttered by them. Most deep fakes involve four types of facial manipulation: face synthesis, face swapping, facial attributes, and facial expression.[2] We also identified several other emerging technologies to 'enhance' deep fakes in our research, including synthetic audio (most deep fakes use dubbed-in voice actors) and "talking head" technology which instead of generating new mirrored audio uses a text-based editing approach where dialogue is modified with no jump cuts (see graphic on evolution of illicit deep fakes).[3]

[1] https://interestingengineering.com/generative-adversarial-networks-the-tech-behind-deepfake-and-faceapp

[2] https://arxiv.org/pdf/2001.00179.pdf

[3] https://www.ohadf.com/projects/text-based-editing/

Deep Fakes Economy

Commoditization

Deep fakes are currently being commoditized primarily via open source repositories (GitHub), service platforms (websites automating creation with GUIs), and marketplaces sellers (Fiverr and NSFW pornography forums). Deep fake vendors engaged in fee-based video production are operating almost exclusively on the surface web (defined as public, open, and searchable by web crawlers). The primary illicit use of deep fake video creation is nonconsensual face swapping pornography on both the surface and dark web, confirming prior research conducted by Deeptracelabs in 2019.[4] Our research into over 200 deep and dark web sites did not find strong evidence of an economy (marketplaces for selling as a service) specifically for e-crime or disinformation purposes.

Revenue

The types of profits currently being generated through licit or illicit deep fake technology have three basic characteristics: deep fake video creation for a fee (service portals and marketplaces like Fiverr), ad revenue or subscription fees on websites (pornography sites), and donation solicitations for continued research (via PayPal, Bitcoin, or Patreon). We envision the *illicit* commodification and underground economy for this synthetic manipulation as evolving in several phases.

Evolution

We envision three phases and general timeframes for deep fake technology and illicit commodification migration. The initial phase was defined by nonregulation and niche hobbyists (from the original creation in late 2017 to mid 2018). The second phase, which we are currently in, is witnessing semi-regulation and the emergence of an initial underground economy. The next phase is likely to see the illicit community go largely underground and deep fakes sold as a service for e-crime and nation-state level activities.

[4] https://deeptracelabs.com/mapping-the-deepfake-landscape/

Evolution of the Illicit Deep Fake Economy

Late 2017 - Mid 2018

Phase 1: Unregulated

- **Technology Originates:** Reddit user "deepfakes"
- **Community Grows:** Subreddit r/deepfakes; Gfycat; Pornhub; Twitter
- **Software Applications:** FakeApp
- **Uses:** Celebrity Pornography; Revenge Porn
- **Methodology:** Exclusively Face Swap
- **Actors:** Hobbyists

Mid 2018 - Mid 2020

Phase 2: Semi-Regulated

An Underground Economy Emerges

- **Technology Shared:** New training models; GitHub; Discord
- **Community Migrates:** After bans on Reddit and other platforms, moves to Voat, Telegram, and Deep Web Forums
- **Profits Emerge:** Surface Web Video-for-Fee sites; Ads and Subscriptions on Pornography Sites; Donations via PayPal and BitCoin
- **Uses:** Nonconsensual Pornography; E-Crime and Election Cases Emerging
- **Methodology:** Primarily Face Swap; Beginning of Synthetic Audio; Talking Head
- **Actors:** Hobbyists; Labs; Tech Enthusiasts

Late 2020 - ???

Phase 3: Underground

Sold as Service, E-Crime, Nation-States

- **Technology Refined:** No longer just celebrities / high-profile figures; synthetic audio mimicking technology spreads
- **Seller Community Expands:** Popularity grows and sellers emerge on the deep and dark webs; smartphone applications
- **Illicit Profits and Uses Grow:** E-Crime (fraud, blackmail, social engineering); Disinformation (mixed with real videos for confusion, cast additional doubt on truth / fact); Pornography
- **Methodology:** Refined synthetic audio; Better pairing with video; Body movement authenticity (not just face mimicking)
- **Actors:** Criminals; Cyber Community; Nation-State Actors

Icons by ivisual from Noun Project

NISOS

A screenshot from the popular deep fake video, "Home Stallone."

Phase 1: Unregulated

Deep fake technology emerged in late 2017 when a Redditor named "deepfakes" created a pornographic video using celebrity face swapping. Soon after, a subreddit was dedicated to deep fakes, growing to 15,000 subscribers, and a user-friendly application called FakeApp was created that allowed anyone to recreate videos with their own datasets.[5] Most of the individuals creating deep fakes tended to be either "hobbyists," researchers, or in the entertainment industry, and the line blurred between illicit and legal use.

Additional interest forums emerged on other social media platforms and pornographic sites. After technology media outlets spotlighted these videos, social media platforms and some adult sites reacted to privacy and potential criminal implications of these nonconsensual videos and took action in an attempt to stop the deep fake activity. In early February 2018, Reddit shut down the deep fakes subreddit and in the same week Discord, Gfycat, Twitter, and Pornhub all denounced involuntary pornography and banned deep fake videos from their platforms.[6] Deep fake activity soon migrated to other online platforms and information sharing sites.

[5] https://www.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley

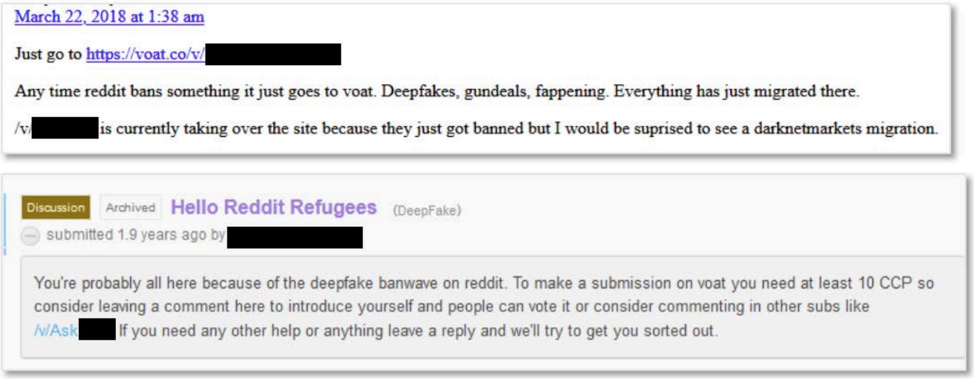
[6] https://www.vice.com/en_us/article/neqb98/reddit-shuts-down-deepfakes

Phase 2:

Semi-Regulated

An underground economy emerges

After Reddit shut down the r/deepfakes subreddit in 2018, several landing zones emerged on both the encrypted messaging chat Telegram and the deep web forum Voat. These platforms offered continuity for deep fake discussion and image and video sharing (high-quality image repositories, especially for celebrities, are frequently shared and required for celebrity deep fake pornography). Voat, a content-hosting and discussion forum site that claims to protect privacy and free speech but has also become a hub for the alt-right, picked up where the subreddit left off and posted links to deep fake software downloads, tutorials, and technology discussions. This forum also continues to offer the original FakeApp technology that has since been removed from most social media platforms. Additionally, a simple search for "deepfake" on Telegram reveals numerous channels for deep fake interest and technology discussions. We noted most GitHub users sharing deep fake technology maintain Telegram and Discord channels for interacting with other researchers as well as "customers." For example, we posted numerous questions and contacted administrators on Discord forums when conducting our own deep fake video creation efforts. While Discord may have banned the actual posting of deep fake videos, discussion and information sharing around the technology itself continues to generate large user communities.



March 22, 2018 at 1:38 am

Just go to https://voat.co/v/[redacted]

Any time reddit bans something it just goes to voat. Deepfakes, gundeals, fappening. Everything has just migrated there.

/v/[redacted] is currently taking over the site because they just got banned but I would be suprised to see a darknetmarkets migration.

Hello Reddit Refugees (DeepFake)

submitted 1.9 years ago by [redacted]

You're probably all here because of the deepfake banwave on reddit. To make a submission on voat you need at least 10 CCP so consider leaving a comment here to introduce yourself and people can vote it or consider commenting in other subs like /v/Ask[redacted] If you need any other help or anything leave a reply and we'll try to get you sorted out.

Users on the dark web posting about deep fake migrating from Reddit to less-regulated forums such as Voat (top); and a Voat welcome note for Reddit users posted at same time r/deepfakes was shut down (bottom).

Open Source Repositories

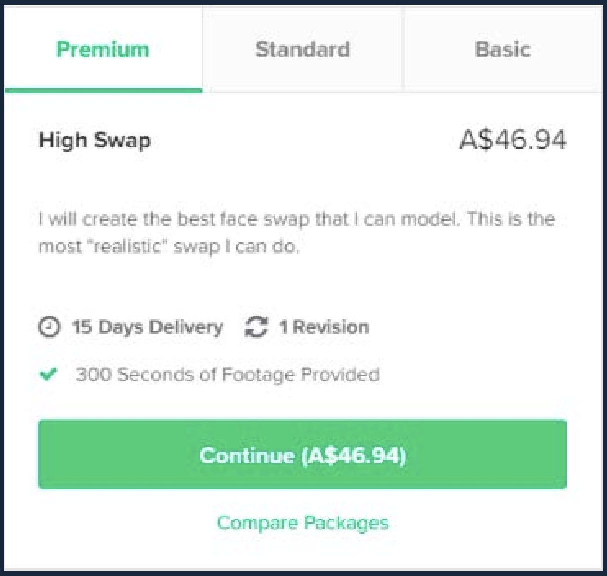
The foundation for deep fakes commoditization continues to be open source repositories like GitHub where various software and training models are shared without a fee (although many users request donations for continued research). Along with GitHub, we found interest forums or personal sites from researchers or labs that are using the technology from GitHub or likely the same users posting the code and models on the repository site. In one case, we tied a GitHub project promoting a particular face swapping technology being “forked” to a deep web forum that was promoting the same technology and also operating a celebrity pornography site. This site and forum was likely another migration area for users after the Reddit takedown, based on the earliest post timeframes, and has an entire thread dedicated to guides and tutorials on extracting and training models for matching celebrity faces with pornographic videos.

Sellers Emerge

In this second phase, we also discovered several service platforms and marketplace sellers on the surface web actually offering deep fake video creation for a fee. These service platforms appear to the user as websites purporting to automate the process of creating deepfakes through a graphical user interface (GUI). It is unclear what elements of the process are automated, or whether the website's owner or employees are manually operating open source software with the training data uploaded through the GUI.

The process is fairly simple; the user uploads the video for editing, then uploads a photo (or another video) of a person, waits for the video to render and downloads the modified video. Deep fake marketplace sellers advertise on sites like Fiverr and NSFW message board sites like Voat or 4Chan, and utilize the same video creation process as service GUI platforms (and likely utilize the same open source face swapping software).

The creation time frame varied by sophistication of the output: swapping a photo onto a video for basic modification could take a matter of minutes, whereas merging two videos for higher quality and authentic output could take hours. Most of these sites run the entire platform on the cloud and do not require any user downloads; we assess they are using basic face swap technology likely taken from open source repositories due to the similarity with these models from the screenshots and example videos on their sites.



High Swap A$46.94

I will create the best face swap that I can model. This is the most "realistic" swap I can do.

15 Days Delivery 1 Revision

✓ 300 Seconds of Footage Provided

A premium quality face swap video from one seller on Fiverr costs $47.

Who are the deep fake creators?

For privacy reasons we are not identifying the names of the individuals or companies associated with deep fake commoditization, but we will share a few characteristics for each. We note the geographic dispersement of these deep fake creators, confirming this technology is a global phenomenon.

Japan

We identified one online service portal seller running an automated deep fake video creation website as likely based in Japan and a professor or associate professor at a Japanese university specializing in neural network research and AI technology. In this case, we assess the website is likely for showcasing technology and generating additional income, as well as possibly for determining overall industry demand for deep fake videos.

Russia

We attributed another entity sharing information and training models on GitHub (asking for donations) as Russian in origin. This individual included a training manual in the Russian language as well as one of the donation outlets via Yandex.Money, a popular Russian online payment method. This individual has a GitHub page for a particular face swapping technology and is possibly also the owner of or associated with the deep fake topic forum (mentioned above) associated with a celebrity pornography site. This user had links to the forum via the GitHub page and asked for PayPal donations on both sites. This user communicated on two Telegram channels (one of our migration areas mentioned above) and reminds participants to “don’t forget to hide your phone number.”

China

The typical deep fake purveyor profile is more likely the hobbyist and researcher. One user likely located in China and marketing to a primarily Asian audience writes about deep fake and faceswap technology and is consistently reworking existing code to improve capabilities in that area, especially the DeepFaceLab project. The user has a running GitHub account where they list another personal site blog they publish articles to. The GitHub activity, as well as the writer's frequent publishing on multiple blogging platforms, suggest this person is a programming enthusiast specifically interested in emerging deep fake technology and other related technologies like GANs. Like many hobbyists, this user is a talented programmer actively involved in the open-source community and regularly working with others to improve the code powering these emerging technological advances.

Phase 3: Underground

Sold as Service, E-Crime, Nation-States

We foresee an increasing growth in deep fake deployment for criminal or nation-state activities as most likely in the next phase of illicit deep fake commodification. We anticipate that as deep fake videos reach higher quality and "believability," coupled with increasing technology proliferation, the videos will be used for more criminal purposes. For example, instead of using an email for a social engineering cyber attack, deep fake synthetic audio or video could be used in real time or as evidence of a company executive's decision.

Any emerging technology is merely an extension of a malicious actor's toolkit, but the criminal still has to use (and be highly effective with) social engineering tactics to induce someone into taking an action. Criminals also learn from each other, so as more high-profile criminal success stories gain notoriety, we anticipate more illicit actors trying them and learning from others who have paved the way.

Disinformation

Our company's numerous investigations into domestic and foreign disinformation campaigns targeting various electoral contests across the globe has not yet identified any successful illicit use of deep fake videos in recent election cycles. So-called cheap fakes or shallow fakes – typically defined as manipulated or misleadingly-edited videos, often taken out of context – have been much prevalent online in recent years.

Malicious actors spreading political disinformation on social media platforms have also recently used fake, life-like pictures of non-existent individuals produced by GAN technology on a free website for their sock puppet accounts.[7] The aforementioned methods are far cheaper and easier to produce compared to existing deep fake technology.

Deep fakes can easily be exposed as lies as we assess most disinformation campaigns will continue to attempt to spin or manipulate existing content and videos, creating doubtt and deceiving their target audiences about the truth of factual events.

[7] https://abcnews.go.com/US/facebooks-latest-takedown-twist-ai-generated-profile-pictures/story?id=67925292

The Future

While we do not anticipate widespread deep fake use in disinformation campaigns in the near term (to include the 2020 election cycle), as deep fakes become easier to create or purchase and the quality significantly increases, coupled with synthetic audio enhancements, we do anticipate increasing deployment of these in information warfare. As previously mentioned, this technology is simply another tool in the vast toolkit of information operations. If an operation required the use of a completely fake doctored video for maximum impact, and it is worth the money and resources, it will likely be used. Or think of a scenario where a video, actually a deep fake, is 'accidentally' left on a hard drive or thumb drive by a nation-state entity as part of a highly covert information operation.

Again, if it's the most impactful tool necessary to accomplish a mission, it might be worth the resources.

What is happening on the dark web?

We examined deep and dark web content from over 200 marketplaces, forums, and communications channels (primarily English-language but with global user bases – also the term "deepfake" is most commonly referred to in English and does not appear to have a close equivalent term in other languages). Our research found a significant lack of sellers on deep and dark web underground forums and therefore we assess there is a lack of large-scale demand for deep fake video creation at this time. We did identify at least one entity on the Dread dark web forum advertising deep fakes for a fee, in addition to creating deep "nudes" for a lower cost, however that was the exception.

The vast majority of dark web discussions center around technology sharing and instructing users to learn how to create deep fakes themselves, as well as references to locations where software or tutorials can be obtained. We found discussions most frequently on dark web forums (as opposed to marketplaces) in mostly English and Russian language, and almost all centered around instructing users to download the software and teach themselves. We also found one dark web forum post claiming neural voice cloning software is not publicly available and asking whether someone had software to clone a person's voice that would allow for synthetic audio mirroring – there was no response to the question.

Dark web marketplaces are precisely that – a market to sell a good or service – and the lack of presence on these indicates the demand has not yet materialized. It is possible deep fake video creation is being sold entirely on private and encrypted channels, but that is not conducive to high and recurring profits and there is almost no ability to market to a wider audience.

Conclusion

Nisos research points to a combination of factors likely explaining the lack of a larger deep fake video illicit economy. The technology and resource barrier to entry is still high, particularly when using non-celebrity individuals that lack high quality videos with footage captured from different angles. Second, the quality of deep fake videos is not yet high enough for convincing use in either disinformation or e-crime activities. Third, and as result, the underground market is not lucrative enough yet for a large number of video production sellers.

As deep fakes become easier and quicker to create, however, social media platforms (currently the primary vehicle for deep fake dissemination) and users will simply not be able to keep up with the volume of synthetic manipulated videos or images. Instead, this will require a partnership between on-platform, off-platform, and public sector entities.

On-platform entities (social media and traditional media outlets) can provide the kind of internal telemetry data that off-platform organizations (cyber security, non-profit watchdogs, investigative journalists) can use as start points for additional investigations. These off-platform “middle brokers” can help attribute the identities and locations of the deep fake creators and proliferators, and identify the evolutions in technology as deep fake software and training models improve and proliferate.

These on-and-off platform entities can then share this information with the public sector and private legal entities where we see maximum impact in the following three areas:

- **Public accountability and information sharing:** Partnering with multiple entities ensures cross-platform analysis that could answer questions like which social media or traditional media outlets are being targeted and why? This information could then be shared with both the domestic public and international government and non-government partners.
- **Threat actors and intent:** Attributing the origin points of this technology is critical and would shed light on the actors actually creating and deploying deep fakes. This would answer questions such as is it all nation-state level influence operations? How prevalent is criminal use for digital crimes like identity theft or fraud, what is are marketing companies and social influencers doing to commoditize this technology?
- **Action:** Finally, what can be done to stop illicit use of deep fakes? Bringing the right partners together would generate mitigation options and strategy, such as potential new legislation, increasing funding for detection capabilities, criminal prosecution, innovative authentication procedures, and increased public awareness of synthetic manipulation.

NISOS™

Responsive strategies and intelligent defenses for protecting your enterprise.

Nisos.com | **info@nisos.com**
603 King Street, Suite 300 Alexandria, VA 22314