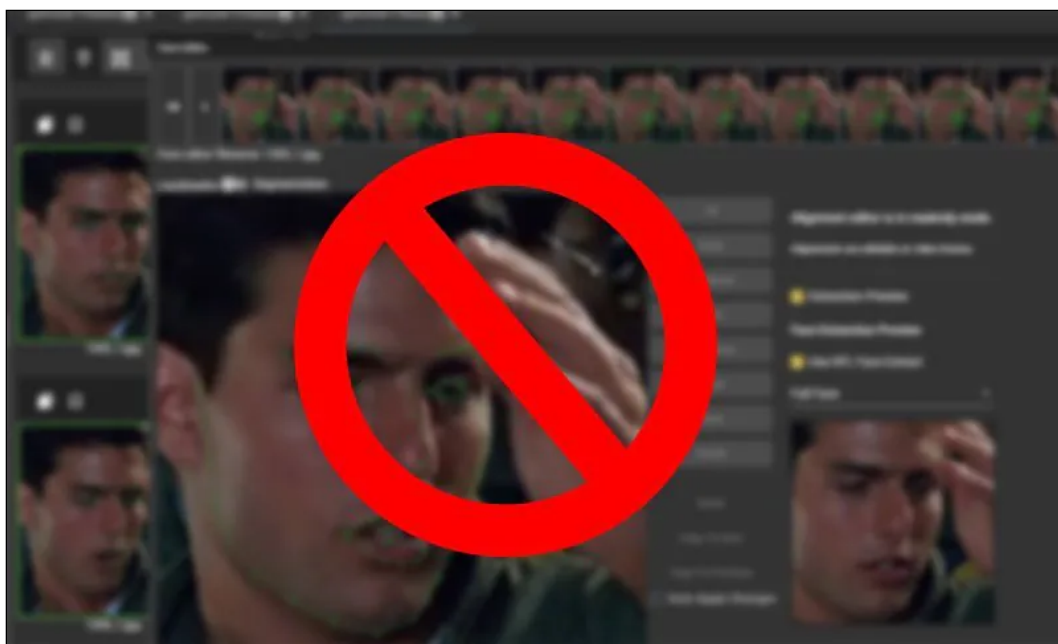


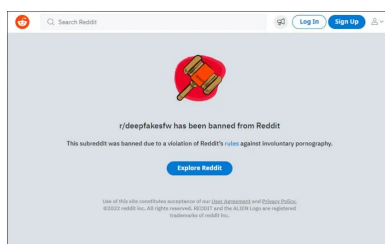
Reddit Bans 'SFW' Deepfake Community

Originally published at <https://www.unite.ai/reddit-bans-sfw-deepfake-community/>



Over the weekend, apparently sometime around Sunday, Reddit banned the nominally 'SFW' (Safe For Work) deepfakes community titled r/deepfakesfw. The subreddit was one of the early responses to the social media giant's prompt [deletion](#) of the original, AI-porn-ridden r/deepfakes sub in 2018.

The (relatively) boiler-plate legend that now greets (archive snapshot taken Monday, June 13) anyone visiting the sub explains that it '*was banned due to a violation of Reddit's [rules](#) against involuntary pornography*'.



The new front page for r/deepfakesfw.

The r/deepfakesfw sub was not frequently archived by popular conservation platforms, but the most [recent WayBack Machine snapshot](#), taken around ten days ago (on 3rd June 2022), indicates that the sub had 3,095 readers at that time.

That's actually a higher number of subscribers than [r/DeepFakesSFW](#) (note the extra 's'), which currently has [2,827 readers](#) (archive snapshot taken Monday, June 13, 2022). The most popular 'legitimate' SFW deepfakes community appears currently to be [r/SFWdeepfakes](#), which at the time of writing has [16,636 readers](#) (snapshot taken Monday, June 13, 2022).

Don't Even Ask

What appears to have doomed r/deepfakesfw was not the posting but the *requesting* of deepfake pornography. Reddit's rules on Sexually Explicit Media state:

'[Images] or video of another person posted for the specific purpose of faking explicit content or soliciting "lookalike" pornography (e.g. "deepfakes" or "bubble porn") is also against the Rule.'

Based on the last available RSS feeds from the now-banned SFW sub, it has been the requesting rather than the posting of NSFW deepfake content that has finally triggered the ban:



Some of the final, clearly NSFW posts from the now-dead r/deepfakesfw subreddit. Source: Inoreader.

Though the posts that these RSS posts point to are no longer available, clicking the requesters' profile links (blurred out in image above) reveals that many of the posters request banned deepfake content across more than one forum, and are either asking for or offering some of the most extreme content of this type, including incest porn and various shades of involuntary pornographic themes.

This suggests that the remaining 'SFW' Reddit subs survive by the speed, diligence and fast response times of their moderators, since it seems that Reddit is unwilling to become a kind of 'classified ad' forum for people to connect and transact illegitimate deepfakes via DMs (in cases where this cannot be explicitly requested in a post).

If so, then the not-yet-banned r/DeepFakesSFW may be living on borrowed time; its two current top posts when entering the root subreddit URL with the 'old' Reddit theme enabled have been specifically marked 'NSFW' (a tag that should presumably not even be available in this particular community).



Depending on circumstances, some rule-breaking posts currently appear in the r/DeepFakesSFW sub. Source: <https://www.reddit.com/r/DeepFakesSFW/>

You can see these two NSFW posts topping out today's r/DeepFakesSFW forum archived [here](#) (snapshot taken Monday, June 13, 2022).

There is actually no NSFW 'flair' available when making a new post at r/DeepFakesSFW, which has been operational for four years, according to its sidebar information:



Available 'flairs' when posting to r/DeepFakesSFW.

Therefore the NSFW tags at r/DeepFakesSFW seem to have either been applied by moderators at the forum or by higher-level admins – or even algorithms, possibly without any intervention or available opposition by local moderators. I reached out to the sole listed moderator of r/DeepFakesSFW about this, and will post any available updates.

(EDIT, Tuesday, June 14, 2022 – A moderator at r/DeepFakesSFW has responded, stating 'It seems to happen when the user's account is NSFW'. Presumably the only way to slow down the likelihood of such requests in SFW deepfake subreddits would therefore be to ban users who have 'earned' this classification.)

Most of the content at this subreddit, and at the larger r/SFWdeepfakes is concerned with using deepfake technology to 'recast' actors from popular or classic movies, or otherwise to have some more innocuous AI-driven fun with popular media and deepfake technology, and do indeed qualify as 'SFW'.

Even though many such posts skirt perilously close to copyright-driven bans, the continuing survival of 'IP-appropriated' non-porn deepfake video posts by major YouTube deepfakers such as Ctrl-Shift-Face suggest a broad studio and rights-holder tolerance towards such activity, at least for the time being.

The 'Innocent' Appearance of the Deepfake Process and Ecostructure

Two weeks ago it [came to light](#) (and much [discussion](#)) that Google had suddenly banned deepfake-creation software from its popular Colab cloud processing environments. Even to date, Colab is apparently only 'triggered' by attempts to run [DeepFaceLab](#) (DFL), the most popular deepfakes software in the world (and the one most [closely associated with AI porn](#)), rather than less-frequented implementations such as [FaceSwap](#), which is better-reputed, even though it performs exactly the same tasks as DFL.

To an extent, this approach may have been the only way that Colab could have implemented *any* restriction on the production of deepfake pornography, since the actual deepfake training process uses nothing but open source software and highly-cropped face images (usually at a mere 512x512px, often less), with almost zero possibility for image recognition algorithms to intercept any actual pornographic content.

By the time the deepfake model is trained, the deepfaker no longer needs the enormous processing power of Colab or an expensive local GPU. Most actual *conversions* (i.e. the process of imposing a face into a video, pornographic or otherwise), can be accomplished locally, on low-powered laptop GPUs – or even on a CPU (with a slightly longer processing time). By the time anything pornographic is happening in the deepfake process, Colab has long since been cut out of the picture.

Therefore, the constituent elements that pass between creators of deepfake porn rarely represent more than a potential minor copyright infringement, in the form of the passing around of ‘face sets’, which deepfake creators have extracted from YouTube videos, Blu-rays, televised interviews, social media, and various other sources.

Is a ‘Scorched Earth’ Policy Coming for Deepfake Communities?

The question that haunts the denizens of deepfake Discord communities such as [DeepFaceLab](#) (including [DeepFaceLive](#)), [FaceSwap](#), [Machine Video Editor](#) (MVE, a GUI-based [environment for DFL](#), with many added features), and new audio-based deepfake Discord servers such as [Audio Deepfakes](#) is *Will current portal and platform owners soon simply ban deepfake communities on principle, rather than bear the expense of constant moderation against a tide of NSFW adherents and creators?*

Bryan Lyon, one of the developers of FaceSwap, and owner of the project’s associated Discord server, commented[†]:

‘We’ve managed to stay on Discord, despite several purges, by staying absolutely squeaky clean. We don’t have any hidden channels or a secret second server with illicit content. There was a purge of the popular DeepFaceLab server a couple of years back. Everyone on it got banned. They had invite-only roles with access to channels with non-consensual pornography.’

Given that Google, itself one of the world’s largest technological hosting resources, has blanket-banned deepfake training on its servers and GPUs, it’s even possible that the Microsoft-owned GitHub may decide that [facilitating these technologies](#)* will eventually not be worth the moral ambiguity or investment of oversight.

At this time, the process of legitimizing deepfake technologies so that their use transits away from homespun porn towards movie and TV production, and non-infringing/criminal uses, is too nascent to produce clear policy guidelines for platforms and community host services.

* *Archive of <https://github.com/search?q=deepfake> from Monday, June 13, 2022.*

† *In private DMs on Discord.*

First published 13th June 2022. Edited 3:08 PM UTC for image amend.

Amended Tuesday, June 14, 2022 6:24am UTC, to add Reddit moderator’s response regarding NSFW status of some posters. Also image amend on RSS feeds illustration.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG’s Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More