CLCT

WILLIAM & MARY
LAW SCHOOL

Cybersecurity and Information Security Newsletter

Issue 12 | November 8, 2021

Table of Contents

Daniel Shin, research scientist with the Commonwealth Cyber Initiative (CCI) Coastal Virginia region, wants to hear from you! Submit any cybersecurity and information security news items or request related topics, via e-mail to dshin01@wm.edu.

This newsletter supports the mission of CCI. To learn more about CCI, including upcoming events, funded research, and news, please visit cyberinitiative.org.

Police arrest suspect for de-pixelating pornographic videos using AI technology

According to Mainichi Shimbun, a leading Japanese newspaper, Kyoto Prefectural Police arrested Masayuki Nakamoto (Nakamoto) in Takasago, Hyogo Prefecture, Japan, for violating Japan's Copyright Act and displaying AI-altered pornography. *West Japan man arrested for alleged sale of porn videos with pixelated images altered by AI*, available here. Nakamoto allegedly posted and sold on his website pornographic videos, where he used publicly available AI algorithms to render de-pixelated genitals from blurred-out areas of the films. *AI でモザイク除去 アダルトビデオ加工画像掲載の疑い 男逮捕 (Mosaic removal with AI Arrested man suspected of posting adult video processed images)*, archived version available here and here (note: the publisher removed the original article without explanation). According to Mainichi Shimbun, this may be the first arrest in Japan relating to an alleged crime involving AI algorithms used with pornographic media.

Due to the country's anti-obscenity law, Japan's pornography publishers pixelate genital areas. *Depixelating porn: Is it really possible to restore 8-bit genitals to their former glory?*, available here. Under Article 175 of the Criminal Code of Japan, an individual "who distributes, sells or displays in public an obscene document, drawing or other objects," may face imprisonment and a fine. *Penal Code (Ministry of Justice)*, available here. Although the criminal statute does not explicitly define the contours of what constitutes "obscene," the Supreme Court of Japan took the position that "public display of human genitals in any form," is obscene. *Obscenity decisions in the Japanese and United States Supreme Courts: cultural values in interpreting free speech*, available with a subscription here. To avoid violating anti-obscenity laws, Japanese pornography publishers use pixelation techniques in their media. Pixelation involves enlarging certain pixels beyond their original size to blur sections in an image. *Pixelation*, available here.

To overcome the pixelated censorship, mosaic removal machines (モザイク除去機), which allow users to edit censored video streams to attempt to de-pixelate images, have become available in Japan. *Depixelating porn: Is it really possible to restore 8-bit genitals to their former glory?*, *supra*. However, these machines require manual adjustments for each visual scene, and the output may not be realistic.

In the past few years, researchers have been utilizing easily accessible machine learning platforms to design and deploy various AI algorithms. Some of these algorithms utilize Generative Adversarial Neural Networks (GANs) to produce realistic but fake output datasets based on the input datasets. GANs have been integrated with major deep fake software, allowing individuals to produce face-swapped videos of celebrities in pornographic context.

Nakamoto is accused of using a publicly available deep learning algorithm, TecoGAN, to produce de-censored videos. *포르노 영상 모자이크 제거해주고 1 억 번 일본 40 대 남성 (Japanese man in his 40s earned 10 million yen by removing mosaic from porn video)*, available here. TecoGAN (also known as Temporally Coherent GAN) aims to produce super-resolution video based on a low-resolution video source. TecoGAN (GitHub), available here. This algorithm attempts to generate "fine details that persist over the course of long generated video sequences." *Id.* By using it, individuals can convert a low-resolution video that is *pixelated* to a detailed, high-resolution one. Nakamoto repurposed TecoGAN to convert pixelated images to detailed, reconstructed, and de-censored images.

After his arrest, Nakamoto admitted that he sold de-pixelated pornography online for financial gain. *Mosaic removal with AI Arrested man suspected of posting adult video processed images*, *supra*.

Analysis

Deep learning development has accelerated in the past few years as researchers are able to leverage both advanced techniques and vast computational resources to present various proof of concepts that demonstrate the next levels of AI-powered automation. As published deep learning projects become more accessible, however, individuals will use these technologies for purposes other than what they were designed for, including the creation of deep fake media.

Nakamoto's alleged acts demonstrate how deep learning models can be repurposed beyond the designer's intentions in ways that fulfill another user's needs. It is foreseeable that a deep learning algorithm primarily designed to fulfill a mundane automated task could be creatively repurposed to perform a different set of tasks, including those that raise ethical and legal concerns. Thus, lawmakers may be tempted to regulate the use of deep learning and other AI algorithms to prevent nefarious uses. However, the Kyoto police did not have to rely on any novel law to arrest Nakamoto. The current criminal laws of Japan were sufficient in this case to detain him for his illegal activity. AI technologies are just like any other tools, where users can misuse them for nefarious purposes. Proactive law enforcement may be a more effective measure for reducing criminal conduct involving AI technologies than passing specific legislation regulating AI algorithms.

U.S. House Members introduce The Justice Against Malicious Algorithms Act

On October 14, 2021, U.S. House Committee on Energy and Commerce Chairman Frank Pallone, Jr. introduced *The Justice Against Malicious Algorithms Act* ("Algorithms Act") that seeks to remove liability protection for websites and online platforms that *knowingly or recklessly* provide information that "materially contribut[e] to a physical or severe emotional injury to any person." *H.R. ______ The Justice Against Malicious Algorithms Act*, available here. Representative Pallone introduced the legislation to hold social media platforms accountable for using "personalized algorithms that promote extremism, disinformation, and harmful content." *E&C Leaders Announce Legislation To Reform Section 230*, available here. The introduction of this bill comes right after Frances Haugen, a former Facebook employee, accused the social media network of "knowingly amplifying harmful content and abusing the immunity of Section 230 well beyond congressional intent." *Id.*

Section 230 of the Communications Decency Act, commonly known only as "Section 230," "specifies that service providers and users may not 'be treated as the publisher or speaker of any information provided by another information content provider.'" *Liability for Content Hosts:*

An Overview of the Communication Decency Act's Section 230 (Congressional Research Service) (citing 47 U.S.C. § 230(c)(1), available here. Courts have interpreted this section to mean that "Section 230(c)(1) immunity may apply in any suit in which the plaintiff seeks to hold the provider liable 'as the publisher' of another's information." *Id.*

Thus, a website, which mainly serves content posted by its users, is protected by legal immunity that is generally not extended to traditional content publishers. Such a website is provided immunity from civil liability if it engages in self-regulation or makes good faith efforts to edit its online platform by screening, restricting, or blocking access to illegal content posted by users. *Ending Immunity Of Internet-Facilitated Commercial Sexual Exploitation Through Amending The Communications Decency Act*, available with a subscription here.

The implications from Section 230 are that websites, including social media, are generally immune from civil liability arising out of user-posted content on platforms. Importantly, Section 230 *only protects* online intermediaries that host or republish others' content, while original authors of such content remain liable for their speech. For example, if a person publishes a video that falsely accuses another of adultery, the online video platform is immune from liability for making the libelous video available, while the video's author could face a potential libel lawsuit. *See, e.g., Va. Code Ann § 18.2-417 (2020)*, available here.

Representative Pallone's newly introduced bill attempts to circumscribe the broad immunity enjoyed by online intermediaries. It removes immunity in cases where online platforms knowingly or recklessly recommend content that materially contributes to a physical or severe emotional injury to anyone. *H.R. _____ The Justice Against Malicious Algorithms Act*, *supra*. Given that most platforms use algorithms to personally recommend content to users, the Algorithms Act discourages the use of "personalized algorithms that promote extremism, disinformation, and harmful content." *E&C Leaders Announce Legislation To Reform Section 230*, *supra*.

Analysis

In his press release, Representative Pallone remarked, "[s]ocial media platforms like Facebook continue to actively amplify content that endangers our families, promotes conspiracy theories, and incites extremism to generate more clicks and ad dollars." *E&C Leaders Announce Legislation To Reform Section 230*, *supra*. Similarly, during her testimony in front of the U.S. Senate Committee on Commerce, Science, and Transportation, Ms. Haugen emphasized that Facebook "chooses profit over safety every day," and that "Congress can change the rules Facebook plays by and stop the harm it is causing." *Statement of Frances Haugen (United States Senate Committee on Commerce, Science and Transportation)*, available here. Both government representatives and certain civil society groups have been raising the issue of holding Facebook and other social media platforms accountable for recommending pernicious content that could lead to violence or other kinds of harm. *See, e.g., Facebook Says It Supports Internet Regulation. Here's an Ambitious Proposal That Might Actually Make a Difference*, available here.

The Algorithms Act attempts to hold online media platforms accountable by removing certain immunities granted by Section 230. Representative Mike Doyle, one of the co-sponsors of the bill, mentioned that the "era of self-regulation is ending . . . ." *E&C Leaders Announce Legislation To Reform Section 230*, *supra*. But this bill does not impose any governmental regulations on the methodology (e.g., algorithms) used to recommend content to users. In fact, the Algorithms Act fails to expand the government's role in monitoring the content recommendation algorithm landscape across social media websites.

This bill was intended to respond to the rise of online disinformation and extremism propagated via content recommendation algorithms in social media networks. *Id.* However, it neglects to make any references to disinformation (false information *intended* to mislead). Whether certain algorithms recommended disinformation has no bearing on lifting Section 230 immunity. Regardless of the factual nature of the content, the Algorithms Act only focuses on whether the content recommendation materially contributed to a physical or severe emotional injury to any person to determine whether Section 230 immunity should be lifted against an online platform.

From a litigation perspective, the bill does not define key legal terms that are critical for potential plaintiffs (the party initiating a lawsuit) to prevail. For instance, the Algorithms Act fails to define what constitutes "recklessly making a personal recommendation to users." Consequently, potential plaintiffs must overcome the burden of defining this legal element and, second, demonstrating to the court how this element applies to their case. By not defining key legal terms, the Algorithm Act also creates ambiguity for online platforms on how to manage their recommendation algorithms to conform with the law.

Also, the Algorithms Act exempts recommendation methodologies where a content recommendation "was made directly in response to a user-specified search." *H.R. _____ The Justice Against Malicious Algorithms Act*, *supra*. It is unclear whether these exemptions include content recommendation systems that utilize direct user input to provide an ongoing stream of recommended content. For example, certain social media networks, such as LinkedIn, allow users to specify user-interested topics to determine recommended contents on the user's timeline. Under the bill, it is indeterminate the extent to which a user's input may be considered as a "user-specified search." If the Algorithms Act passes, online media platforms may pivot towards the user-specified model for their recommended content streams to place their platforms under the exemptions category.

The Algorithms Act fails to directly regulate the content control mechanisms utilized by online media platforms. Furthermore, the ambiguous legal landscape it offers makes it difficult for potential plaintiffs to efficiently present their claims against these platforms. As it stands, the bill presents both a high burden for potential plaintiffs to present their legal claims and a relatively low one for platforms that want to use AI content recommendation tools. Unless lawmakers clarify key concepts written in the bill, it may have limited effect in holding social media accountable for utilizing "malicious algorithms." In essence, it is unclear whether the Algorithms Act changes the legal landscape enough for social media companies to alter their content recommendation algorithms to avoid civil liability.