Cybersecurity and Information Security Newsletter

Issue 22 | September 27, 2022

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.

New Text-to-Image AI Model allows users to produce pornographic and other controversial content

Around the middle of August 2022, a community of users exploring a publicly released AI model, Stable Diffusion, was circulating on social media fake but realistic images of naked individuals, including some resembling celebrities. *Deepfakes for all: Uncensored AI art model prompts ethics questions*, available here. Unlike face-swap deepfake media, these images were completely created by Stable Diffusion's text-to-image synthesis, where the user describes what type of image the AI model should generate. *Stable Diffusion [GitHub],* available here. The community of enthusiasts discovered that Stable Diffusion was able to synthesize not only pornographic imagery but also other controversial content, including gore.

Although there are other AI models capable of text-to-image synthesis, Stable Diffusion is the only major AI model that is freely available for users to create images using their own computers. *See, e.g., DALL-E 2*, available here. Other text-to-image models are only accessible as an online service, and users are generally restricted to the type of images they are able to generate based on the restrictions in place. Unfortunately, Stable Diffusion does not have such content restrictions, allowing users to generate any type of image without limitation.

The Machine Vision & Learning Group at the Ludwig Maximilian University of Munich (Ludwig-Maximilians-Universität München) developed and released Stable Diffusion using a subset of the LAION-5B dataset as the training data to develop the model. *High-Resolution Image Synthesis with Latent Diffusion Models*, available here. LAION-5B contains 5.85 billion images with a corresponding text description of the image. *LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets*, available here. When LAION released the dataset, it disclosed that around 3% of the image-to-text pairs were determined to be "unsafe," which includes pornographic-looking content. *Id.* Because Stable Diffusion was trained on the entire LAION-5B dataset, Stable Diffusion's ability to generate pornographic synthetic imagery was likely due to the model's exposure to "unsafe" data from LAION-5.

Currently, the community of enthusiasts has developed easy-to-use software packages for other users to use Stable Diffusion on their computers. Also, new social media communities around Stable Diffusion creations have been formed, where individuals share their expertise in fine-tuning synthetic image creations while others showcase their work. Given the relative ease of creating Stable Diffusion generated images, growing numbers of people would likely have access to this technology.

Analysis

Deep learning permits opportunities for automating evermore complex tasks that can facilitate the development of more sophisticated systems to benefit society. Unfortunately, the emergence of "dark AI," which refers to the use of AI technology for nefarious purposes, raises ethical and legal issues that need to be addressed along with the development of AI technology.

The first public release of a face-swapping program that leveraged deep learning algorithms to produce realistic but fake face-swapped videos opened the first Pandora's box in relation to non-consensual pornographic media production. Virginia's statute on anti-revenge

pornography was amended to include “videographic or still image created by any means whatsoever.” *Va. Code Ann. § 18.2-386.2*, available here. Under the statute, it is a Class 1 misdemeanor offense if an individual synthesizes and then *maliciously* distributes or sells a deep fake media of a target person in a state of nude or undress without authorization. Although the media synthesis technique is different, any media generated by Stable Diffusion would fall under the anti-revenge pornography statute.

Stable Diffusion and other unrestricted text-to-image synthesis algorithms may represent a new means for dark AI as they can create realistic but fake media beyond the pornographic context in support of nefarious purposes. For example, disinformation actors could use Stable Diffusion to synthesize media to create realistic but fake social media profiles as part of a disinformation campaign. The synthesized image below shows a variation of images based on the text prompt “A linkedin profile photo (of a woman | in a cybersecurity field | who is an attorney).”

Figure 1: Stable Diffusion synthesized images of fake LinkedIn profile pictures

The potential dark AI uses of Stable Diffusion call for an interdisciplinary discussion on AI technology development, which could lead to a creation of a robust ethical and legal framework to facilitate AI research. Such a framework could encourage ethical AI technology development while minimizing the possibility of inadvertently opening up yet another Pandora's box.

CISA publishes a guide for post-quantum cryptography for critical infrastructure

On August 24, 2022, the Cybersecurity and Infrastructure Security Agency (CISA) published *Preparing Critical Infrastructure for Post-Quantum Cryptography*, which provides background information about the security risks related to emerging quantum computing and how to prepare to implement quantum-resilient cryptography for critical infrastructure. *CISA Releases New Insight on Preparing Critical Infrastructure for the Transition to Post-Quantum Cryptography*, available here. Although the National Institute of Science and Technology (NIST) expects to publish its post-quantum cryptography standard in 2024, CISA's publication serves to warn owners and operators of critical infrastructure systems to begin the process of addressing this foreseeable risk within vulnerable systems. *Preparing Critical Infrastructure for Post-Quantum Cryptography*, available here.

Background Information on Encryption

Encryption refers to the use of mathematical functions to transform data that prevents unauthorized parties from accessing or tampering with data. *Cryptography [NIST]*, available here. Generally, encryption uses cryptographic "keys" (similar to a password) to encrypt and decrypt information. Usually, longer cryptographic keys tend to yield a more resilient encryption strength, although the processing power needed to encrypt and decrypt information also increases. *About encryption keys [IBM]*, available here.

There are two approaches to encryption. **Symmetrical encryption** uses a common cryptographic key to encrypt and decrypt information. *Symmetric Cryptography [NIST]*, available here. It is mostly used to encrypt efficiently large data.

Asymmetrical encryption (also known as public and private key encryption) uses a pair of mathematically related keys to perform encryption. It is commonly used to initiate and maintain encryption communication among parties without having to disclose decryption keys in public. Under an asymmetrical encryption scheme, a user randomly generates a "private key" and uses it to generate a corresponding "public key." With the key pair at hand, the user keeps the private key a secret while distributing the public key to the rest of the world. For encryption purposes, other individuals can use the user's public key to encrypt the information before transmitting it to the user. Due to the cryptographic properties of the public key, only the user with the corresponding private key can decrypt information that was previously encrypted by the public key. *Asymmetric-key cryptography*, available here. The benefit of asymmetrical encryption is it allows encrypted communication among users without the need to share publicly decryption keys, which can compromise protected communications.

If individuals and organizations use one of many industry-standard symmetrical or asymmetrical encryption algorithms, their encrypted data should be secure.

Key Findings of the Report

Quantum computers use certain properties of quantum mechanics to produce computing capabilities that could far exceed modern, binary-bit computer systems. In essence, quantum computers promise a seismic shift in performance, which could be used for protein folding simulations that can facilitate medical research, and even astronomy simulations.

From a cybersecurity perspective, one of the main implications of quantum computing is its ability to break effortlessly modern encryption algorithms. Current encryption algorithms are premised on modern computers not having the processing capability to break encryption in a reasonable amount of time.

According to CISA's report, asymmetrical encryptions using a public and private key system is likely vulnerable, while symmetrical encryptions with long key lengths are likely resistant to quantum computer attacks. Although quantum computing has not yet been developed to break some of the modern encryption algorithms, CISA warns that organizations storing and transferring data with long secrecy lifetime (i.e., the sensitivity of information has a long duration, such as social security numbers of individuals) could be subject to **catch-and-exploit operations**. Catch-and-exploit operations refer to capturing encrypted communications and storing them until the threat actor has, at a later date, the capability to decrypt captured encrypted communications. Because asymmetrical encryption algorithms are commonly used for data transfers, a threat actor who captures encrypted communications today could foreseeably use quantum computing to break the public and private key encryption algorithm and access the data. Although catch-and-exploit operations may not have the practical benefit of accessing encrypted information in real time, they pose a risk for sensitive information with a long secrecy lifetime.

Federal Government's Efforts to Post-Quantum Cryptography

Earlier this year, the Department of Homeland Security (DHS) and NIST published the "Post-Quantum Cryptography Roadmap" to help organizations prepare their information systems against the advancement of quantum computing technology. *Post-Quantum Cryptography*, available here. The Roadmap noted CISA's efforts to conduct a macro-level assessment of vulnerable critical infrastructure systems across the 55 National Critical Functions. These are identified as "functions of government and the private sector so vital to the United States that their disruption, corruption, or dysfunction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof." *National Critical Functions Set*, available here.

It is against this backdrop that, on July 6, 2022, CISA released its "Post-Quantum Cryptography Initiative." *Post-Quantum Cryptography Initiative*, available here.

Analysis

The foreseeable availability of quantum computing poses a national security threat because it is likely that nation-state actors would have first-hand access to the technology. As CISA's report acknowledges, most encryption methodologies used for data transfers, including Transport Layer Security (commonly known as TLS, which is used widely for secure internet communications), will be vulnerable to quantum computing attacks.

Beyond the nation's critical infrastructure operations, organizations and businesses should also consider how they should prepare for the era of Post-Quantum Cryptography, including performing a data inventory of any sensitive information and categorizing them into short and long secrecy lifetimes. CISA's report also provides an excellent opportunity to review encryption practices, including whether organizations and businesses are employing industry-standard encryption methods and utilizing data security best practices.

*This newsletter would especially like to thank Dr. Tran Viet Xuan Phuong of Old Dominion University for sharing her expertise on encryption for this article.