S&T Digital Forgeries Report

Technology Landscape Threat Assessment

January 24, 2023

Message from the Under Secretary

I am pleased to present the "S&T Digital Forgeries Report, Technology Landscape Assessment," which is the result of a collaborative effort between the Science and Technology Directorate (S&T), Department of Homeland Security (DHS) component agencies, and our industry partners. It reflects the growing attention that both DHS and Congress are placing on the threats posed by this rapidly evolving set of tools and technologies.

Our growing integration of smarter and more advanced technologies into our everyday lives brings with it new focus on how to attend to emerging risks associated with this rapid progress. Digital content forgery technologies are an example of new risks associated with emerging technologies, including artificial intelligence (AI) and machine learning (ML) techniques, to fabricate or manipulate audio, visual, or text content with the intent to mislead. The adversarial side of AI is a growing domain that is also now adding new approaches to the list of long-used software tools that create and manipulate information. These new methods are both increasing the quality of digital content forgeries, as well as reducing the amount of time and skill required to create the content. As Congress has noted, digital content forgery technologies could be misused to commit fraud, cause harm, harass, coerce, silence vulnerable groups or individuals, and/or violate civil rights.

Digital content forgery technologies enable adversaries to create or manipulate digital audio, visual, or textual content, to distort information, undermine security and authority, and ultimately erode trust in each other and in our government. Additionally, development and deployment of digital forgeries is inexpensive, putting common cybercriminals on par with nation state adversaries, exponentially increasing the threat we face daily.

This report is the first of five annual assessments, which establishes our path ahead and is consistent with the requirements set forth in Section 9004 (a) through (e) of the *National Defense Authorization Act (NDAA) for Fiscal Year 2021* (P.L. 116-283), concerning the state of digital content forgery technology.

This report is being provided to the following Members of Congress:

The Honorable Mark E. Green
Chairman
House Committee on Homeland Security

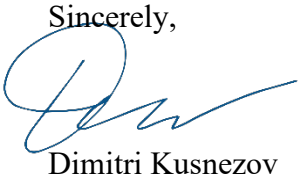
The Honorable Bennie G. Thompson
Ranking Member
House Committee on Homeland Security

The Honorable Gary C. Peters
Chairman
Senate Committee on Homeland Security and Government Affairs

The Honorable Rand Paul
Ranking Member
Senate Committee on Homeland Security and Government Affairs

Inquiries relating to this report may be directed to the DHS Office of Legislative Affairs at (202) 447-5890.

Sincerely,



Dimitri Kusnezov
Under Secretary for Science and Technology

Executive Summary

Images, audio, text, and video generated or modified by AI technologies and ML methods provide a foundation for synthetic media known as Digital Forgeries, evolving technologies often used for financial, social, or political gain. The ability to make counterfeit media indistinguishable from the original can violate personal civil rights and civil liberties, as well as foment inflammatory or contentious issues that sow division and potentially harm the integrity of the democratic process.

This inaugural assessment presents an overview of the digital forgery landscape by identifying current methods in which synthetic media continues to manifest harm. It provides an overview of tools and techniques currently used by bad actors to generate digital forgeries and deepfakes (a subset of digital forgeries created through ML).

The assessment also identifies various detection and authentication technologies and tools that currently exist or are in development as countermeasures, providing some insight into ways to challenge or rebut fake or altered content. For example, to help address the forgery and security problems posed by face morphing, the DHS S&T[1] Biometric and Identity Technology Center is performing and sponsoring limited research, development, testing, and evaluation with the National Science Foundation's (NSF) Center for Identification Technology Research (CITeR) and National Institute of Standards and Technology (NIST). DHS S&T is also performing work at NSF CITeR to identify methods to generate images that can be used to establish performance baselines for detection techniques. Additionally, NIST has several concurrent projects related to evaluating facial recognition algorithms that are applied to their large image databases.[2] These projects are co-funded by DHS S&T, DHS Office of Biometrics and Identity Management, and Federal Bureau of Investigation Criminal Justice Information Services, and executed under the NIST Face Recognition Vendor Test (FRVT) portfolio. In particular, the FRVT Morph project focuses on providing ongoing independent testing of prototype facial morph detection technologies.

During the next several years, future reports will present updated threat assessments that will continue to inform solutions. This 2022 report is an important first step in our effort to identify the everchanging threats in this space as we work with our federal, state, local, tribal, territorial, and private sector partners to assess and develop tools to prevent and mitigate future threats.

[1] DHS S&T activities pertaining to digital forgeries are done in accordance with federal privacy requirements and DHS privacy policy.

[2] Ngan, Mei, Patrick Grother, Kayee Hanaoka, and Jason Kuo. 2021. Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8292.

S&T Digital Forgeries Report

Technology Landscape Threat Assessment, January 24, 2023

Table of Contents

1. Legislative Language

The *National Defense Authorization Act (NDAA) for Fiscal Year 2021* (P.L. 116-283) includes the following requirements:

SEC. 9004. DEPARTMENT OF HOMELAND SECURITY REPORTS ON DIGITAL CONTENT FORGERY TECHNOLOGY.

(a) REPORTS REQUIRED. — Not later than one year after the date of enactment of this Act, and annually thereafter for 5 years, the Secretary of Homeland Security, acting through the Under Secretary for Science and Technology of the Department of Homeland Security, and with respect to paragraphs (6) and (7) of subsection (b), in consultation with the Director of National Intelligence, shall submit to Congress a report on the state of digital content forgery technology.

(b) CONTENTS. — Each report produced under subsection (a) shall include the following:

(1) An assessment of the underlying technologies used to create or propagate digital content forgeries, including the evolution of such technologies and patterns of dissemination of such technologies.

(2) A description of the types of digital content forgeries, including those used to commit fraud, cause harm, harass, coerce, or silence vulnerable groups or individuals, or violate civil rights recognized under Federal law.

(3) An assessment of how foreign governments, and the proxies and networks thereof, use, or could use, digital content forgeries to harm national security.

(4) An assessment of how non-governmental entities in the United States use, or could use, digital content forgeries.

(5) An assessment of the uses, applications, dangers, and benefits, including the impact on individuals, of deep learning or digital content forgery technologies used to generate realistic depictions of events that did not occur.

(6) An analysis of the methods used to determine whether content is created by digital content forgery technology, and an assessment of any effective heuristics used to make such a determination, as well as recommendations on how to identify and address suspect content and elements to provide warnings to users of such content.

(7) A description of the technological countermeasures that are, or could be, used to address concerns with digital content forgery technology.

(8) Any additional information the Secretary determines appropriate.

(c) CONSULTATION AND PUBLIC HEARINGS. — In producing each report required under subsection (a), the Secretary may —

(1) consult with any other agency of the Federal Government that the Secretary considers necessary; and

(2) conduct public hearings to gather, or otherwise allow interested parties an opportunity to present, information and advice relevant to the production of the report.

(d) FORM OF REPORT. — Each report required under subsection (a) shall be produced in unclassified form, but may contain a classified annex.

(e) APPLICABILITY OF OTHER LAWS. —

(1) FOIA. — Nothing in this section, or in a report produced under this section, may be construed to allow the disclosure of information or a record that is exempt from public disclosure under section 552 of title 5, United States Code (commonly known as the ‘‘Freedom of Information Act’’).

(2) PAPERWORK REDUCTION ACT. — Subchapter I of chapter 35 of title 44, United States Code (commonly known as the ‘‘Paperwork Reduction Act’’), shall not apply to this section.

(f) DIGITAL CONTENT FORGERY DEFINED. — In this section, the term ‘‘digital content forgery technology’’ means the use of emerging technologies, including artificial intelligence and machine learning techniques, to fabricate or manipulate audio, visual, or text content with the intent to mislead.

2. Background

The genesis for legislation on the subject of digital content forgery began with a bill introduced in the Senate (S. 3805) on December 21, 2018, with the short title, "Malicious Deep Fake Prohibition Act of 2018," to prohibit fraud in connection with audiovisual records. The bill defined "deep fake" as "an audiovisual record created or altered in a manner that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of an individual," with the term derived by the contraction of "deep learning" (i.e., neural networks manipulating media) and "fake."

Now more commonly referred to as "deepfake," its underlying artificial intelligence (AI) and machine learning (ML) techniques continue to evolve in sophistication and are increasingly difficult to detect. Furthermore, both open source and commercially available software for editing text, audio, and imagery provides an inexpensive suite of tools to the average consumer, contributing to the array of threats to privacy, property, and institutions. These less sophisticated methods, along with deepfake and other emerging threats supported by AI and ML, comprise the broader definition of "digital content forgery," used in the legislative language of Section 9004 of the *National Defense Authorization Act (NDAA) for Fiscal Year 2021* (P.L. 116-283), that provides guidance for this report.

3. Technology Landscape

A considerable suite of easy-to-use, low- or no-cost software applications are available to anyone interested in creating digital forgeries for research, entertainment, or any other purpose. In the 1990s, it became possible for anyone with a computer to alter an image to some degree of realism, depending on the skill of the editor and the nature of the source image, including putting the face or head of one person onto the body of another.

Today, numerous commercial and open-source applications exist for image, video, and text editing; face swapping; puppeteering; and generating deepfake video or audio. While there are many lawful and legitimate uses of these technologies, the proliferation of free and low-cost software technology tools and easily accessible public data, in the wrong hands, presents an ever-increasing threat.

3.1. Generation Tools and Techniques

Emerging digital content forgery technologies utilize AI and ML to create forgeries with such a level of accuracy that it becomes tremendously difficult to distinguish between real and manufactured data. Some representative examples follow, but this is not an exhaustive list. In the future, software will permit full-body deepfakes, real-time impersonations, and the whole removal of elements within videos. The most recent technology has the capability to produce levels of realism that run in near-real-time.

3.1.1. Face Swap

One of the most common techniques for creating deceptive images is the *face swap,* which pre-dates AI-based methods. One major harmful use of face swap technology is seen in deepfake pornography, where the faces of unsuspecting victims are swapped with those in pornographic videos. Recent examples include well-known actors who were face-swapped in several pornographic videos without their consent, with one of the fake videos being labeled as "leaked" footage and gaining more than 1.5 million views.[3]

Most private individuals do not have the same level of resources as famous ones, socially or financially, nor do they command the same public attention to refute such damaging claims or videos.[4] Also, current laws do not adequately protect individuals from these crimes or offer them much recourse. Even though there exist some laws that prohibit the non-consensual distribution of intimate digital images that provide protection to victims, they are geared more towards protecting against the revelation of real images, rather than deepfakes, which are neither completely fake nor completely real.[5]

---

[3] Willen, Claudia. 2020. "Kristen Bell Says She Was 'Shocked' to Learn That Her Face Was Used in a Pornographic Deepfake Video." Insider. June 11, 2020. https://www.insider.com/kristen-bell-face-pornographic-deepfake-video-response-2020-6.

[4] Department of Homeland Security Office of Intelligence on behalf of the Office of the Director of National Intelligence. 2021. "Increasing Threat of Deep Fake Identities."
https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

[5] Gieseke, Anne. 2020. "'The New Weapon of Choice': Law's Current Inability to Properly Address Deepfake Pornography." Vanderbilt Law Review 73 (5): 1479–1515.

With current technology advancements, especially with that of AI, face swaps are now more realistic and convincing. Adversaries can swap the face from one person onto another person's face and body using a variety of applications and tools, such as autoencoders or Deep Neural Network (DNN) technologies.[6]

Autoencoding requires training an AI to deconstruct and reconstruct images. As shown in *Figure 1*, during training, the networks for both Original Face A and Original Face B are sharing an encoder but are using separate decoders to reconstruct the original face without changes. Completely separating the networks during training would make them incompatible because their latent face would represent different features. By sharing the same encoder, each network is trained to determine shared meaningful features. During this training process, it is expected that the networks will begin to understand the concept of a "face."[7] The so-called latent image, or latent face in this example, captures the salient aspects of the images as a lower-dimensional abstraction of the image, that allows the image to be reconstructed from it.

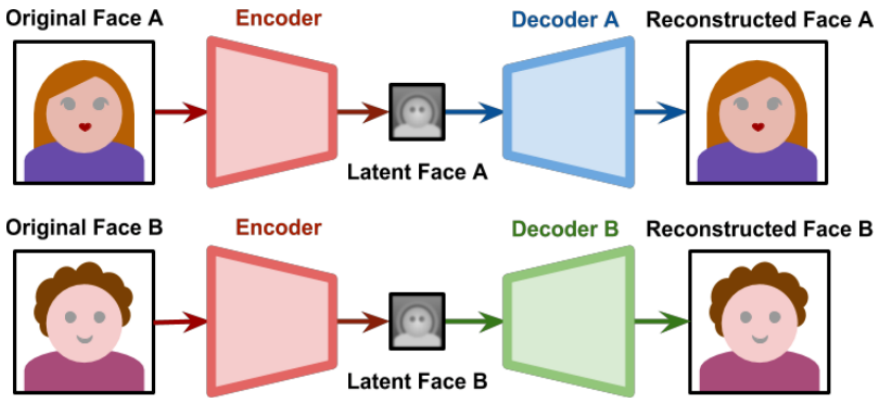


Figure 1:Autoencoding networks during training. Diagram is from Zucconi (2018)

Once training is complete, the networks will swap their latent faces using each other's decoder. As presented in *Figure 2,* Latent Face A will go through Decoder B and Latent Face B will go through Decoder A. Decoder A will then attempt to reconstruct Original Face A from the information it received from Latent Face B. Conversely, Decoder B will attempt the same with Latent Face A.[8] If the networks are trained sufficiently, the ending result should be a reconstruction of the original face with the swapped Latent A or B features imbedded.

[6] Ibid.

[7] Zucconi A. "An introduction to DeepFakes. Part 6: Understanding the technology behind DeepFakes." 2018 March 14. Available at: https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes.

[8] Ibid.

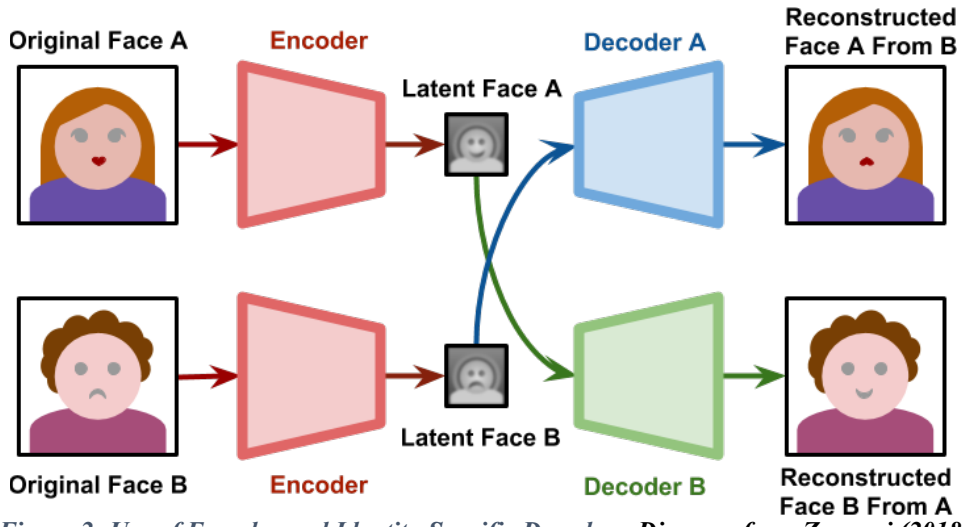


Figure 2: Use of Encoder and Identity-Specific Decoders. ***Diagram from Zucconi (2018)***

This type of technology is not limited to faces but is also used for any other type of image. The main concern is for the objects being swapped to have as many similarities as possible to ensure that meaningful features are easier to transfer.[9]

Face swap technology is also widely available. There are many applications that allow a user to swap faces but not all use the same technology. Such applications include FaceShifter,[10] FaceSwap,[11] DeepFace Lab[12], Reface,[13] and TikTok.[14] In addition, well known social applications, like Snapchat[15] and TikTok, require minimal knowledge and they have lower computational needs, which allows users to generate manipulations in real time.[16] It is anticipated that these technologies will continue to mature rapidly in the coming years.

3.1.2. Face Morphs and GANs

A face morph is a face image that is synthesized from two or more individuals such that the resulting image is a photo-realistic combination of identities and no longer a true representation of any single person. If morphed face images are inserted into databases or identity documents, they can confuse human judgement and cause errors in automated face recognition systems that assume face images only correspond to a single individual. Morphed face images increase false

[9] Ibid.

[10] Pascu, Luana. 2020. "Microsoft Research, Peking University Develop AI Deepfake Detector | Biometric Update." Www.biometricupdate.com. January 7, 2020. https://www.biometricupdate.com/202001/microsoft-research-peking-university-develop-ai-deepfake-detector.

[11] "Faceswap." 2020. Faceswap. October 21, 2020. https://faceswap.dev/.

[12] Perov, Ivan, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr Dpfks, et al. 2020. "DeepFaceLab: A Simple, Flexible and Extensible Face Swapping Framework." ArXiv:2005.05535 [Cs, Eess], May. https://arxiv.org/abs/2005.05535.

[13] Lomas, Natasha. 2021. "Reface Now Lets Users Face-Swap into Pics and GIFs They Upload." TechCrunch. May 17, 2021. https://techcrunch.com/2021/05/17/reface-now-lets-users-face-swap-into-pics-they-upload/.

[14] TikTok. 2022. "TikTok." Tiktok.com. TikTok. 2022. https://www.tiktok.com/.

[15] Quora. 2017. "The Inner Workings of Snapchat's Faceswap Technology." Forbes. Forbes. March 17, 2017. https://www.forbes.com/sites/quora/2017/03/17/the-inner-workings-of-snapchats-faceswap-technology/?sh=4b01e16564c7.

[16] Increasing Threat of Deepfake Identities. Page 9.

match error rates, creating situations where two or more people could present and use the same identity document, such as a passport. [17]

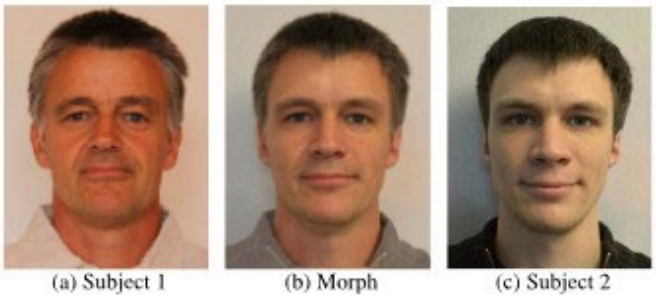


Figure 3: Example of a morphed face image (b) of subject 1 (a) and subject 2 (c). [18] *The Morph was manually created using FantaMorph software. (Scherhag et al))*

Face morphs are synthesized by first aligning the geometry of the facial images of two individuals and then blending the face pixels, usually in equal portions, to create a third image. The third morph image is no longer a true representation of either person, yet the image may contain enough content from either or both individuals to trigger recognition responses. When adversaries create a synthetic image with characteristics of two individuals to use as a reference image in a database, it is known as a *face morphing attack*. As seen in Figure 4, different instances of face images of both subjects contributing to a face morph are successfully matched against a database using a commercial-off-the-shelf (COTS) face recognition software with a default decision threshold of 0.5, resulting in an FMR (false match rate) of 0.1%.[19]

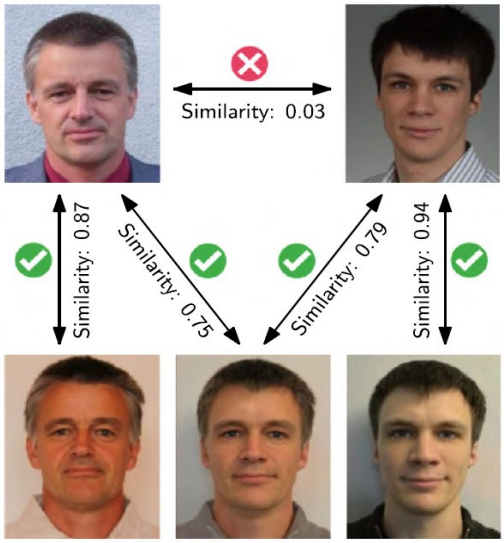


Figure 4. Example of a face morphing attack (Scherhag et al)

[17] Ferrara, Matteo, Annalisa Franco, and Davide Maltoni. 2014. "The Magic Passport." IEEE Xplore. IEEE. September 1, 2014. https://doi.org/10.1109/BTAS.2014.6996240.

[18] Scherhag, U., Rathgeb, C., Merkle, J., Breihaupt R., Busch, C. (2019, February), Face Recognition Systems Under Morphing Attacks: A Survey. IEEE Access. Retrieved from: https://ieeexplore.ieee.org/document/8642312

[19] Ibid.

Another method for synthesizing face morphs is using generative adversarial networks (GANs), which can be used for many digital forgery technologies. GANs computing architecture and concepts were introduced in 2014 by Goodfellow et al.[20] GANs consist of two processes, a generator, and a discriminator, that are connected and pitted against each other. Conceptually, the generator can be described as a counterfeiter and the discriminator as a process that tries to detect and reject counterfeits, thus learning from each other. Once the GANs is sufficiently trained, the generator becomes able to defeat the discriminator and can synthesize convincing counterfeits. As the discriminator's counterfeit detection methods are improved, the generator works further to pass the new detection tests.

GANs can generate high-resolution, realistic faces, as commonly seen in video games.[21] The software tools can project a real face image into the GANs' latent space where encoded representations of the faces are combined or manipulated. The combined face encoding is then converted back to a new generated image. An interesting property of GANs-based morph generation methods is that they may include face recognition as part of the discriminator's process to check that resulting morphed faces not only look realistic but also maximize their similarity to the original images.

Advances to image editing tools have introduced AI-based, face-aware manipulations and special effects. While the alterations are not necessarily face morphs, the tools are popular, easy to use and can also confuse face recognition processing. Face-editing tools can, for example, modify face geometry, gender, and expression, or apply digital cosmetics, such as hair style, blush, and makeup. As altered face images are increasingly easy to produce, the provenance and integrity of face images used in identification documents and for identity verification is an important security consideration.

To help address the forgery and security challenge problems posed by face morphing, the Department of Homeland Security (DHS) Science and Technology Directorate (S&T)[22] Biometric and Identity Technology Center is performing and sponsoring limited research, development, testing, and evaluation with the National Science Foundation's Center for Identification Technology Research (CITeR) and National Institute of Standards and Technology (NIST). DHS S&T is performing work at NSF CITeR to identify methods to generate images that can be used to establish performance baselines for detection techniques.

DHS S&T is funding work at NIST to evaluate face-morph detection technologies and methods. The NIST face morph evaluation[23] reports on detection performance of multiple morph detectors

[20] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Ward-Farley, Sherjil Ozair, Aaron Courville, and Yushua Bengio. n.d. Review of Generative Adversarial Nets. In Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014), 2672–80.

[21] Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. "Analyzing and Improving the Image Quality of StyleGAN." IEEE Xplore. June 1, 2020. https://ieeexplore.ieee.org/document/9156570.

[22] DHS S&T activities pertaining to digital forgeries are done in accordance with federal privacy requirements and DHS privacy policy.

[23] Ngan, Mei, Patrick Grother, Kayee Hanaoka, and Jason Kuo. 2021. Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8292.

against multiple morphing methods along with providing analysis on the effectiveness of the morphs against leading face recognition systems.

Work sponsored through CITeR and evaluated by NIST has demonstrated incremental improvement on detection. However, the problem is technically difficult and, even with recent improvements, face morph detection performance remain below what is operationally needed.

3.1.3. Voice Synthesis and Voice Editing

Voice synthesis, or *deepfake voices*, are emerging as part of speech and voice technologies used in automated call centers, text-to-speech applications, video games, and dictation and audio production tasks. Advances in AI/ML software techniques over the past decade have produced convincingly natural synthetic voices, complete with appropriate inflections, emphasized words, and pauses. These fake voice scams, sometimes called *vishing*, can be used to impersonate others, potentially undermining remote identity verification processes that are critical to government services and benefit programs, and for access to financial services.

Voice synthesis with M/L techniques can produce convincing impersonations when trained against voice recordings of a target individual. Advances over approximately the last five years make it possible to mimic an individual's voice when sufficient pre-recorded samples are available for algorithm training. The amount of data and pre-recorded samples needed for training varies from several minutes up to 10 or 20 hours, depending on the application and desired fidelity.[24]

As reported in *The Wall Steet Journal,*[25] an AI-generated voice was used to impersonate a chief executive's voice to successfully request a fraudulent transfer of $243,000. The scam was possible because the CEO's subordinate was tricked by the quality of the impersonation and thus carried out the request.

Synthetic speech content can be generated from any arbitrary text or script and made to sound like the target speaker to, in effect, put words in their mouth. Short phrases and sentences are quite difficult for humans to distinguish as being synthetic. Longer audio and conversations are more prone to contain nuanced artifacts that humans could detect, especially if they are familiar with the speaker or have reason to question the audio's authenticity. The 2021 film *Roadrunner* used an AI-generated version of the late chef Anthony Bourdain's voice in three phrases that were indistinguishable from his real voice otherwise used throughout, raising questions on the ethics of not disclosing which passages were manufactured and not Bourdain's actual speech.[26]

Voice editing is another new application that has been enabled by advances in ML. Voice editing works by creating a text transcript from recorded speech, and then allowing the transcript

[24] Zhang, Maggie, and Rafael Valle. 2020. "Training Your Own Voice Font Using Flowtron." NVIDIA Technical Blog. October 3, 2020. https://developer.nvidia.com/blog/training-your-own-voice-font-using-flowtron/.

[25] Stupp, Catherine. 2019. "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case." WSJ. Wall Street Journal. August 30, 2019. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

[26] Rosner, Helen. 2021. "The Ethics of a Deepfake Anthony Bourdain Voice." The New Yorker. July 17, 2021. https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice.

to be edited and making corresponding edits to the audio. Usually, the voice edits are achieved with no discernable gap or artifacts. This technique can be used by an audio producer to clean up audio tracks, removing filler works and errors. But the techniques can just as easily add, delete, or replace words, negating statements that were affirmative, changing names, or otherwise altering the meaning and the message of the speech.

3.1.4. Puppeteering

Puppeteering techniques superimpose animated motion and speech from a source character onto another character. This involves digitally rendering 2D and 3D models in a virtual environment in real-time using computers. These renderings can include full-body and facial movements, or even smaller areas of manipulation, such as blinking and lip-syncing.

3.1.4.1. Digital Puppetry

In traditional computer graphic approaches, such as motion capture puppetry, motion is captured or specified at various key points (anchor points) from the source character's face and body. The motion is then applied to a set of corresponding key points on a target character such that the target character behaves and moves according to the source character. Motion capture hardware and computer vision techniques have automated much of the motion creation and transfer process, and additional modeling and ML techniques have improved facial animation that historically required dedicated efforts from computer graphic artists working with specialized production tools.

Even though puppeteering technology provides an improved and realistic experience to visual arts, such as filmmaking, video games, and theatre attractions, threat actors can also use this technology nefariously. By employing face morphs and GANs methods with motion capture puppeteering, users can target specific individuals and make them appear as if they are moving or acting in ways that did not actually occur.

Additionally, puppeteering technology is now easily available to the masses, especially with the advent of smart phones. Users can download applications that allow mapping of their or others' faces onto source materials by simply choosing a source image and using their own face to *puppeteer* the image. Other applications use deepfake technologies to make still-images move as if they are alive or a video. Some examples of easily accessible applications, which are not part of an exhaustive list, include Reface, Avatarify, and MyHeritage,[27] though there are numerous others available with similar or even more advanced capabilities.

3.1.4.2. Lip Sync

Another aspect in puppeteering are lip sync methods, which focus specifically on mouth and lip motions corresponding to speech. In addition to the transfer of pre-defined motion, mouth and lip movements are synthesized from sound models. These models enable lip synchronization

[27] Fowler, Geoffrey A. 2021. "Perspective | Anyone with an IPhone Can Now Make Deepfakes. We Aren't Ready for What Happens Next." Washington Post, March 25, 2021. https://www.washingtonpost.com/technology/2021/03/25/deepfake-video-apps/.

from arbitrary audio recordings, allowing a content creator to seemingly put words in someone's mouth. Related audio and voice applications are discussed further in section 4.

Examples of this technology are Wav2Lip [28] and Wombo,[29] which generate lip synchronized motion and can operate on still images, video, and even cartoons or sketched faces. Regarding Wav2Lip, the lip motion is generated from sound and not from an explicit phonetic speech model. Thus, instrumental music and non-speech audio will create lip motion even though it does not make sense. Wombo is curated to sync with the lyrics of songs, which can make user-selected characters appear as if they are singing themselves. This is not an exhaustive list of current lip sync technologies, which may have even more capabilities and features.

3.1.4.3. Puppet-Master Forgeries

Building from both motion-capture and lip-sync puppetry methods, along with voice synthesis, some of the most complex and realistic deepfakes are known as a puppet-master deepfakes. In this type of puppetry, a video or recording of a source character is used to puppet an unmodified target video or image by overlaying them together, which can include just the head or entire body movements. Facial expressions from the source character are also mimicked, such as eyes, head, and mouth gestures. This type of technology is significant because it not only swaps faces but can also swap a subject's entire body.[30] Additionally, in many puppet-master forgeries, voices are overlayed to match expected lip and facial movements.

An example of an advanced puppetry deepfake is shown through the MIT's Center for Advanced Virtuality, *In the Event of a Moon Disaster* project,[31] where they used both a voice actor and a video repository of U.S. President Richard M. Nixon's speech patterns. Using deep learning, they superimposed the voice actor's speech with AI generated lip-syncing and sound to an original video of President Nixon to create a hyper-realistic rendition of him delivering a fabricated 1969 contingency speech for an Apollo 11 disaster. Due to having both manipulated video and audio, this type of synthetic media is considered a complete deepfake and is one of the most difficult types to detect visually, audibly, and with technology.

3.1.5. Text and Language Generation

Language and synthetic text generation is a central component to natural language processing (NLP). A long-standing goal of NLP has been to create machine-generated language that responds contextually and appropriately to human prompts. Today's AI-generated chatbots are commonly used as automated virtual assistants and pop-up chat windows on many websites. In

[28] K, Prajwal, Rudrabha Mukhopadhyay, Vinay Namboodiri, and C Jawahar. 2020. "A Lip Sync Expert Is All You Need for Speech to Lib Generation in the Wild." In.

[29] Asarch, Steven. 2021. "Wombo.ai Lets Users Make Silly Deepfake Videos of Their Friends or Celebrities Singing Songs." Insider. March 11, 2021. https://www.insider.com/wombo-ai-womboai-download-transforms-photo-a-singing-deepfake-face-2021-3.

[30] Masood, Momina, Marriam Nawaz, Khalid Mahmood Malik, Ali Javed, and Aun Irtaza. 2021. "Deepfakes Generation and Detection: State-of-The-Art, Open Challenges, Countermeasures, and Way Forward." NASA ADS. February 1, 2021. https://ui.adsabs.harvard.edu/abs/2021arXiv210300484M/abstract.

[31] "Tackling the Misinformation Epidemic with 'in Event of Moon Disaster.'" 2020. MIT News | Massachusetts Institute of Technology. July 20, 2020. https://news.mit.edu/2020/mit-tackles-misinformation-in-event-of-moon-disaster-0720.

news, entertainment, and social media applications, NLP tasks may include reading comprehension, text summarization, question answering, recommendations, sentiment analysis, completion tasks, and language translation. On social media platforms, machine NLP capabilities are inextricably integrated into the platforms; automated content selection and information flows co-exist and combine with the likes and dislikes of its human users, forming a human-machine ecosystem that generates, propagates, and amplifies content according to social networks and algorithm decisions.

AI-based NLP tools excel at the summarization and personalized filtering of news and information at a massive scale. However, the same NLP tools and content distribution mechanisms work equally well for generating and spreading misleading or malign content. In part, this is because platforms consider and favor user engagement and screen time. Platforms also generally lack processes to distinguish content that is authentic and factual from what is misleading or harmful.

Detecting computer-generated synthetic text in the wild is challenging due to the enormous scale and the nature of active and reactive dynamics between humans, platforms, and information content. Moreover, a large portion of synthetic text is ambiguously benign summarizations and translations. Thus, detection of synthetic text is frequently reframed as the problem of detecting misinformation and malicious content, regardless of how it is produced and disseminated. For example, a threat actor could possibly saturate search engines with forged stories created by text generators to fool users into believing an event is true.

3.1.6. Text to Image and Text to Video Generation

In 2021 and 2022, a number of organizations introduced transformative new capabilities to generate images and video from text descriptions. Tools, such as DALL-E 2,[32] Make-A-Video,[33] and Imagen[34] process text descriptions and use a variety of AI/ML techniques to render photo realistic images and video, as well as artistically stylized imagery. These tools can combine different subjects, concepts, scenes, attributes, and styles to creates images and video that are highly relevant to the input text descriptions.

Generative AI research is a rapidly evolving branch of AI/ML, some of which appears to rely on large, mostly un-curated, web-scraped datasets to permit rapid development and iteration. Recognizing the potential for misuse, some research teams have implemented various safeguards, such as limiting distribution of source code, as well as filtering input text descriptions and removing offensive language and imagery from training sets.

3.1.7. Counterfeiting Identity Documents

Digital imaging, editing, and printing technologies play a major role in the fabrication of forged identity documents, such as bank cards, passports, driver's licenses, and identity cards.

---

[32] OpenAI. n.d. "DALL·E 2." OpenAI. https://openai.com/dall-e-2/.

[33] Facebook. 2022. "Introducing Make-A-Video: An AI System That Generates Videos from Text." Ai.facebook.com. Meta AI. September 29, 2022. https://ai.facebook.com/blog/generative-ai-text-to-video/.

[34] Saharia, Chitwan, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, et al. n.d. "Imagen: Text-To-Image Diffusion Models." Imagen.research.google. https://imagen.research.google/.

Counterfeit operations advertise the sale of fake documents on the internet. Most facilities, organizations, and institutions, both public and private, rely upon identification (ID) cards to determine risk, admissibility, and eligibility of individuals based on their identity. The sharp rise in pandemic relief benefits has also created new targets of opportunity for fraudsters, and some fraud may have included the use of fake identity documents populated with stolen personal information.

Historically, faked documents, such as driver's licenses, were produced using simpler methods, like switching pictures (i.e., photo substitution) or text and printing a new copy. Conversely, current-day producers of fake licenses can create high-quality products, produced to match the look and feel of genuinely issued documents, most complete with machine scannable barcodes. Many web sites that offer fake licenses provide third party quality reviews and user-submitted reports to suggest that their product is usable in other identity-verification situations beyond getting into bars. Customization is also an option to buyers; when creating their fake license, they can choose their desired age, state, photo, and other identity information they want printed on the final product. These sites typically accept payment in the form of cryptocurrency or gift cards.

Counterfeiters claim they can replicate specific security features, such as raised text, holograms, micro-printing, and ultraviolet features. Investigative news reporting from across the country,[35, 36, 37] present examples of high-quality forgeries obtained online for $100 to $200. Most reports attribute the high-quality fakes as coming from China.

Government agencies, such as U.S. Customs and Border Protection (CBP), are dealing with the issue of counterfeit documents. For example, since at least 2017, CBP has consistently seized thousands of U.S. state counterfeit driver's licenses on a monthly basis. The counterfeit stream did not dissipate during the COVID-19 Pandemic and the quality of simulations has steadily risen. To date, the quality of a prototypical counterfeit U.S. state licenses can effectively simulate a host of security features, including multiple laser images, laser perforation, tactile data, and other surface plate features. DHS has observed a rise in frequency of simulations with redrawn background artwork printed using the genuine security process of offset lithography. In 2022, Transportation Security Administration's (TSA's) review of a popular open-source website that administers fake identification showed there are at least 19 vendors offering fake licenses to individuals across the United States, as well as the United Kingdom with the approximate cost, delivery times (ranging from 1-3 weeks) and accepted payment methods. Most of the payment methods are harder to trace, such as digital currency, prepaid cards, and cash.[38]

[35] ABC15 Arizona. 2011. "Fake IDs so Good They Fool Even the Pros." Video. YouTube. https://www.youtube.com/watch?v=C2YXNn7Z9U0.

[36] TheDaily. 2012. "Scary-Real Fake IDs Could Fool Police." Video. YouTube. https://www.youtube.com/watch?v=A_ydiur0iUY.

[37] CBS Chicago. 2014. "2 Investigators: Fake Driver's Licenses Flooding in from China." Www.youtube.com. YouTube. September 22, 2014. https://www.youtube.com/watch?v=oF8hLsdesZ8.

[38] In 2022, TSA's Systems Risk Analysis Division in coordination with Homeland Security Investigations Forensic Lab (HSI-FL) reviewed the site https://fakeidvendors.com/page/vvl as part of an open-source analysis on the influence fake ID reviews have on illicit purchases of counterfeit State driver's licenses.

To attempt to mitigate fraud, the financial sector and some government agencies have turned to commercial face recognition and document authentication tools, techniques, and services to strengthen identity verification as part of establishing accounts. Remote identity verification services offer the promise of an online process and experience that attempts to replicate in-person identity verification transactions, such as going to a bank or visiting a department of motor vehicles. Most remote identity verification implementations require the user to present live images of their driver's license or ID card and their face selfie that are authenticated and verified against reference data from the original issuance. The adoption of these technologies has grown due to significant user convenience and reduced costs to government and private sector service providers who would otherwise need to have staff and physical facilities to allow a user to assert their identity. While these capabilities have already achieved widespread adoption, there is little independent or objective data characterizing the performance and fairness of the technologies, as well as the degree to which they may reduce fraud at scale.

3.2. Detection and Authentication Technologies

Detection tools exist and are improving for some use cases, but they still do not generalize well and work reliably against all forms of forged or altered content, a key example being against compressed video or printed photographs. Since most deepfakes that have the potential to cause harm spread virally on social media, it is concerning that most detection algorithms suffer degraded performance on lower quality videos that have been compressed and re-encoded through online platforms. Many techniques are a continuation of digital forensics research and now utilize more AI-based approaches, convolutional neural networks, and GANs. However, as these techniques enable improved detection capabilities, they also enable more capable digital content forgery technologies.

Additionally, although use of multiple different approaches can help improve detection, they also currently require expert interpretation, making them inappropriate for general field use at this time and better suited for use in forensic laboratories. This may also present a challenge in explaining results in legal proceedings.

3.2.1. Deepfake Detection Challenges

This section lists challenge datasets and research competitions that seek to measure and advance capabilities for detecting forgeries. Technology contests, such as these are commonly used to encourage researchers, companies, and enthusiasts in finding solutions to difficult computer science problems.

3.2.1.1. Defense Advanced Research Projects Agency Semantic Forensics (DARPA SemaFor)

The Semantic Forensics (SemaFor) program seeks to develop innovative semantic technologies for analyzing media. Detection techniques that rely on statistical fingerprints can often be fooled with limited additional resources (algorithm development, data, or compute). However, existing automated media generation and manipulation algorithms are heavily reliant on purely data driven approaches and are prone to making semantic errors. For example, GANs-generated faces may have semantic inconsistencies, such as mismatched earrings. These semantic failures

provide an opportunity for defenders. A comprehensive suite of semantic inconsistency detectors would dramatically increase the burden on media falsifiers, requiring the creators of falsified media to get every semantic detail correct, while defenders only need to find one, or a very few, inconsistencies. These technologies include semantic detection algorithms, which will determine if multi-modal media assets have been generated or manipulated. Attribution algorithms will infer if multi-modal media originates from a particular organization or individual. Characterization algorithms will reason about whether multi-modal media was generated or manipulated for malicious purposes. These SemaFor technologies will help detect, attribute, and characterize adversary disinformation campaigns.[39]

3.2.1.2. Facebook's Kaggle Deepfake Detection Challenge

From late 2019 through 2020, Facebook, Microsoft, the Partnership on AI coalition, and academics from seven universities launched one of the most comprehensive AI contests to date, the *Deepfake Detection Challenge*.[40] One of the biggest obstacles with deepfake detection is the need for a large sample of examples for AI to train from. Facebook's challenge[41] is unique due to the massive, state-of-the-art public dataset of over 100,000 total video clips, considered the largest publicly available dataset at the time.[42] The dataset, consisting of both original videos and corresponding deepfakes of varying quality, were specifically recorded for use in machine-learning tasks and were altered using various deepfake generation models to replicate real-world videos encountered online. Due to such a large and scoped dataset, along with over two-thousand teams and 35,000 models entering the contest, this challenge helped create and pave the way for the latest wave of deepfake detection innovation.

3.2.1.3. NIST Open Media Forensics Challenge

NIST's Open Media Forensics Challenge (OpenMFC) is a media forensics evaluation to facilitate development of systems that can automatically detect and locate manipulations in imagery (i.e., images and videos). The OpenMFC focuses on three major tasks:[43,44]

> **Manipulation Detection (MD):** For the MD task, the objective is to detect if imagery has been manipulated, and if so, to spatially localize the edits. Manipulation in this context is defined as deliberate modifications of media (e.g., splicing and cloning etc.), as well as localization, which is encouraged but not required for OpenMFC. The MD task includes three subtasks, namely, image manipulation detection (IMD), image splice manipulation detection (ISMD), and video manipulation detection (VMD). The IMD

[39] Corvey, William. "Semantic Forensics (SemaFor)." Darpa.mil, Defense Advanced Research Projects Agency, 2022, www.darpa.mil/program/semantic-forensics.

[40] Finley, Klint. 2019. "Facebook, Microsoft Back Contest to Better Detect Deepfakes." Wired. September 5, 2019. https://www.wired.com/story/facebook-microsoft-contest-better-detect-deepfakes/.

[41] "Creating a Data Set and a Challenge for Deepfakes." 2019. Facebook.com. 2019. https://ai.facebook.com/blog/deepfake-detection-challenge/.

[42] Dolhansky, Brian, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. 2020. "The DeepFake Detection Challenge Dataset." ArXiv:2006.07397 [Cs], June. https://arxiv.org/abs/2006.07397.

[43] National Institute of Standards and Technology. n.d. "Open Media Forensics Challenge." Mfc.nist.gov. NIST. https://mfc.nist.gov/.

[44] Guan, Haiying, Yooyoung Lee, and Lukas Diduch. 2022. "Open Media Forensics Challenge 2022 Evaluation Plan." Https://Mfc.nist.gov. NIST. https://mig.nist.gov/MFC/Web/EvalPlan/OpenMFC2022EvaluationPlan.pdf.

subtask will detect if a probe image has been manipulated. The ISMD subtask will detect if a probe image has been spliced. The VMD subtask will detect if a probe video has been manipulated.

Deepfakes Detection (DD): Using GAN or Deepfake based techniques, the DD task will detect if imagery has been manipulated. The DD task includes two subtasks, namely, image deepfakes detection (IDD) and video deepfakes detection (VDD). The IDD subtask will detect if a probe image has been manipulated based on GAN models, while the VDD subtask will detect if a probe video has been manipulated based on Deepfake models. For each DD trial, which consists of a single probe, the DD system must render a confidence score with higher numbers indicating the probe image is more likely to have been manipulated using GAN-based techniques.

Steganography Detection (StegD): The StegD task will detect if a probe is a stego image, which contains the hidden message either in pixel values or in optimally selected coefficients. For each StegD trial, which consists of a single probe image, the StegD system must render a confidence score with higher numbers indicating the probe image is more likely to be a stego image.[45]

The primary goal of these challenges is to engage the public research community and work on the latest media forensics topics. By making participation free and accessible worldwide, NIST is providing an opportunity to organizations and individuals to test their systems against diverse datasets collected under controlled environments.

3.2.1.4. NIST FRVT Morph

NIST has several concurrent projects related to evaluating facial recognition algorithms that are applied to their large image databases.[46] These projects are co-funded by DHS and FBI and executed under the NIST Face Recognition Vendor Test (FRVT) portfolio. In particular, the FRVT Morph project focuses on providing ongoing independent testing of prototype facial morph detection technologies.

Initiated in 2018 and still ongoing, FRVT Morph is open for participation worldwide and free of charge. NIST tests submissions against datasets created using different morphing methods to evaluate algorithm performance over a broad range of morphing techniques. Testing is also conducted with a tiered approached, where algorithms are evaluated against lower quality datasets created with easily accessible and intuitive tools, as well as against higher quality datasets like ones from academic research and commercial-grade tools. If a submitted algorithm shows promising results, NIST could curate a larger dataset to further evaluate the technology.[47]

---

[45] Ibid.

[46] "Face Challenges." 2019. NIST. U.S. Department of Commerce, National Institute of Standards and Technology. March 27, 2019. https://www.nist.gov/programs-projects/face-challenges.

[47] Ngan, Mei, Patrick Grother, Kayee Hanaoka, and Jason Kuo. 2022. Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection. NISTIR 8292 Draft Supplement. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology.

Throughout this project, NIST tests morph detection performance and face recognition accuracy on morphs. Submitted algorithms are measured on attack presentation classification error rate (APCER), or morph miss rate, when a bona fide classification error rate (BPCER), or false detection rate, is set to a detection of .1 and .01, respectively. The algorithms are tested for low BPCER rates due to the assumption that most transactions in the real world are conducted on bona fide (nonmorph) photos of individuals. Even if an algorithm has a higher APCER rate, a low BPCER rate is preferable and can still provide operational gains compared to not having any detection capabilities. NIST has worked on three types of morph experiments: single-image morph detection, two-image differential morph detection, and morph-resistant face recognition testing. Algorithms are tested on both digital and print-and-scanned imagery.[48]

3.2.1.5. FVC-onGoing

Created in 2009 by the University of Bologna's Biometric System Laboratory, the Fingerprint Verification Competition-onGoing (FVC-onGoing) project is a continuation of four previous FVC competition successes. Initially, the goal of this FVC project was to track the advances of fingerprint recognition technologies by using defined benchmarks. Unlike previous FVC initiatives, FVC-onGoing was built as a continuous competition, always open to new participants without constricted timeframes, with the benefit of an evolving public repository of evaluated metrics and results.[49] Since then, FVC-onGoing continues to host multiple biometric-based competitions and benchmarks. Even though there is still an emphasis on finger-print recognition algorithms, recent benchmark areas now include several deepfake detection and face detection challenges. Each facial detection challenge, along with their current number of evaluated and published algorithms,[50] are noted below:

> **Single-Image Morph Attack Detection**[51] – This benchmark area focuses on face morphing detection on a single image. The goal is to analyze a suspected morphed image and produce a score representing the probability that the image was manipulated. Applicants are provided several different levels of algorithm testing benchmarks that range in complexity, from testing algorithms with a simple dataset to testing with high-resolution printed and scanned face images. As of August 2022, at least 59 algorithms were evaluated with 12 published.
>
> **Differential Morph Attack Detection**[52] – This benchmark area focuses on face morphing detection between two images. Algorithms submitted are required to compare a suspected morphed image against a bona fide (non-morphed) image and produce a score representing the probability that the suspected morphed image was manipulated.

---

[48] Ibid.

[49] "Background." n.d. FVC-Ongoing. Biometric System Laboratory. Accessed August 31, 2022. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/Background.aspx. *See also*,: Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti and A. Mayoue, "Fingerprint and On-Line Signature Verification Competitions at ICB 2009", in proceedings International Conference on Biometrics (ICB), Alghero, Italy, pp.725-732, June 2009

[50] "FVC-OnGoing." n.d. Biolab.csr.unibo.it. Accessed September 28, 2022. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/PublishedAlgs.aspx.

[51] "Benchmark Area: Single-Image Morph Attack Detection." n.d. FVC-Ongoing. Biometric System Laboratory. Accessed August 31, 2022. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/BenchmarkAreas/BenchmarkAreaSMAD.aspx.

[52] "Benchmark Area: Differential Morph Attack Detection." n.d. FVC-OnGoing. Biometric System Laboratories. Accessed August 31, 2022. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/BenchmarkAreas/BenchmarkAreaDMAD.aspx.

Like the Single-Image Morph Attack Detention benchmark area, the sub-benchmarks range in complexity, from a simple dataset to datasets with high-resolution printed and scanned face images. As of August 2022, at least 118 algorithms were evaluated with 23 published.

Face Image ISO Compliance Verification[53] – For this benchmark area, algorithms are required to check the compliance of face images to International Organization for Standardization (ISO) standards, which include identifying shadows, blurriness, pixilation, and other undesired characterizations. There are two sub-benchmarks for applicants: the first is a simple dataset to test algorithm compliancy with the testing protocol, and the second is a large dataset with high-resolution face images related to all pre-defined ISO requirements. As of August 2022, at least 1,223 algorithms were evaluated with nine published.

By allowing both large organizations and the public to submit and test their algorithms against these data sets, the FVC-onGoing project provides a platform for continuous innovation and fine-tuning of future detection technologies. Given that the project is including deepfake and facial detection benchmarks into their biometrics-testing portfolio, one can expect that their interest and dedication to these types of efforts will continue. Regarding subjects that are not explicitly labeled as being related to deepfakes, such as the Face Image ISO Compliance Verification benchmark area, such research and testing could possibly expand into or help influence future deepfake identity fraud detection technologies.

3.2.2. Examples of Mitigation Technologies

This section lists some well-known detection technologies applicable to some digital content forgeries, along with other mitigation approaches. There are multiple companies and software programs that have developed comprehensive mitigation technologies. This is not an exhaustive list and is only a small sample to illustrate the current state-of-the-art technologies.

3.2.2.1. Reality Defender

By using a public dataset from Face Forensic++ and testing on the Deepfake Detection Challenge, Microsoft created Video Authenticator. This detection tool can analyze still photos and videos to provide the percentage chance, or confidence score, that the media is artificially manipulated. For videos, the tool can detect this percentage frame by frame in real-time. In 2020, Microsoft partnered with the AI Foundation's Reality Defender 2020 initiative to make Video Authenticator available to organizations involved in democratic processes.[54] This partnership developed into today's Reality Defender application, which is used by various government agencies and organizations.

[53] "Benchmark Area: Face Image ISO Compliance Verification." n.d. FVC-OnGoing. Biometric System Laboratory. Accessed August 31, 2022. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/BenchmarkAreas/BenchmarkAreaFICV.aspx.

[54] Burt, Tom. 2020. "New Steps to Combat Disinformation." Microsoft on the Issues. Microsoft. September 1, 2020. https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/.

Reality Defender[55] is a deepfake scanning tool with actionable results. It provides continuous security developed by the world's top AI research teams. Reality Defender provides deepfake detection by users scanning assets via a web application or secure application programming interface. The Reality Defender portal provides real-time scanning across press, social media, and websites related to any given company or organization and its customers and employees.

3.2.2.2. University of Buffalo's DeepFake-o-meter

DeepFake-o-meter[56] is an open-source online platform developed by the University of Buffalo Media Forensics Lab (UB MDFL) to detect third-party deepfake algorithms. It provides a convenient service to analyze deepfake media with multiple state-of-the-art detection algorithms, with secure and private delivery of the analysis result. For developers of deepfake detection algorithms, it provides an application program interface architecture to wrap individual algorithms and run on a remote machine. For researchers of digital media forensics, it is an evaluation and benchmarking platform to compare performance of multiple algorithms on the same input.[57]

3.2.2.3. iProov

iProov provides biometric capabilities that assesses the "liveness," or presence, of users. This in turn may reduce fraud risk against machine-driven attacks from deepfakes and other emerging technological threats.[58]

iProov's technology is being used by organizations around the world to authenticate user identities. Their customer base includes the government, public and financial sectors. Regarding governmental use, they specialize in supporting national digital identity projects, visas and immigration, and border patrol control activities.[59] An example is seen through their awarded contract from S&T's Silicon Valley Innovation Program (SVIP).[60]

3.2.2.4. V7 Fake Profile Detector

V7 Fake Profile Detector[61] is a Google Chrome extension that examines social media profile photos to determine if they were created using StyleGan, a popular GAN-based method used to

[55] "Reality Defender." n.d. Reality Defender. https://www.realitydefender.ai/.

[56] Li, Yuezun, Cong Zhang, Pu Sun, Lipeng Ke, Yan Ju, and Siwei Lyu. 2020. "DeepFake-o-meter." DeepFake-o-meter. UB Media Forensic Lab. 2020. http://zinc.cse.buffalo.edu/ubmdfl/deep-o-meter/.

[57] Li, Yuezun, Cong Zhang, Pu Sun, Yan Ju, Honggang Qi, and Siwei Lyu. 2021. "DeepFake-o-meter: An Open Platform for DeepFake Detection." 2021. https://cse.buffalo.edu/~siweilyu/papers/sadfe21.pdf.

[58] "Genuine Presence Assurance | IProov." n.d. Www.iproov.com. https://www.iproov.com/iproov-system/technology/genuine-presence-assurance.

[59] "Biometrics for Government and Public Sector | IProov." n.d. Www.iproov.com. Accessed September 28, 2022. https://www.iproov.com/what-we-do/industries/government-and-public-sector

[60] S&T Public Affairs. 2020. "News Release: S&T Award for Genuine Presence Detection and Anti-Spoofing | Homeland Security." Www.dhs.gov. November 6, 2020. https://www.dhs.gov/science-and-technology/news/2020/11/06/news-release-st-award-genuine-presence-detection-and-anti-spoofing.

[61] V7Labs. 2022. "V7 Releases Deep Fake Detector for Chrome." Www.v7labs.com. V7Labs. April 6, 2022. https://www.v7labs.com/news/v7-releases-deep-fake-detector-for-chrome.

create artificial photorealistic images of human faces. The tools only detects images created using StyleGAN and does not generalize to other methods, face swaps, or deepfakes.

3.2.3. Examples of Potential Mitigation Policies

There are also potential policy changes that could reduce vulnerabilities associated with digital content forgeries. For instance, a group of activists in Germany used face morphing technology to create and authenticate fake passports,[62] likely helping spur the German Government to change its policy of accepting self-submitted passport photos. In 2020, Germany passed a law requiring that only photos taken under the supervision of German authorities are permitted.[63] The danger in such use cases is that not only does an individual have a high probability of passing facial recognition algorithms, but human visual inspection has a low probability of recognizing an imposter.[64]

Moving the other direction, a growing number of government and other organizations accept self-submitted physical or remotely submitted digital photographs as part of an identity document issuance and application process or to verify identity to receive some benefit. A key example is Passport Renewal, a process during which citizens mail two physical photographs of themselves to the Department of State. While this process contains some other identity-verification safeguards, it lacks an onsite human interaction and may be susceptible to some forms of identity attacks, like morphed photographs capable of defeating human reviewers and biometric technologies.

Policies restricting the provision of a State or Federal identity documents (requiring physical in-person, or at least real-time interactive provision) would add burden to these processes, yet they mitigate the susceptibility to digital content forgery attack opportunities. In addition, (or if necessary, as an alternative), government and other organizations could increase the level of collateral identity verification safeguards as part of issuing, or renewing, an identity-based benefit (especially identity-verifying benefits, such as a passport or a driver's license).

Policy mitigations are a necessary and concurrent component to the technological solutions, of which the readiness level for detecting morphs and altered face images on physical artifacts is critically low. As a result, increasing funding and/or supporting technology rallies or sprints to research strategies to detect morphed photographic evidence is equally critical.

---

[62] Thelen, Raphael, and Judith Horchert. 2018. "Biometrie Im Reisepass: Peng! Kollektiv Schmuggelt Fotomontage in Ausweis." Der Spiegel, September 22, 2018, sec. Netzwelt. https://www.spiegel.de/netzwelt/netzpolitik/biometrie-im-reisepass-peng-kollektiv-schmuggelt-fotomontage-in-ausweis-a-1229418.html.

[63] https://www.reuters.com/article/us-germany-tech-morphing/germany-bans-digital-doppelganger-passport-photos-idUSKBN23A1YM and https://digit.site36.net/2020/01/10/laws-against-morphing/

[64] Scherhag et al., Face Recognition System Under Morphing Attacks: A Survey, IEEE Open Access Journal vol. 7 (2019) at 23021, Table 1 (describing the expected morph quality from various morph software); see also Pikoulis et al., Face Morphing, a modern threat to border security: recent advances and open challenges (2021) https://www.mdpi.com/2076-3417/11/7/3207

3.3. Human Capabilities and Human-Algorithm Teaming

Most humans cannot detect morphed and altered face images reliably, especially as the morphed images increase in quality and have less obvious imperfections from the morphing process, such as outlines of another person's hair.[65] NIST describes a high-quality tier III morph as having no visible artifacts to human observation, or in other words, all pre-existing defects are removed using digital image editing software to make the image look authentic.[66]

Research on human face morph detection has been limited to relatively small experiments, usually with unrealistically high occurrences of morphs compared to the suspected low occurrence rates in operational data and workflows. In one experiment, 100 participants reviewed 108 face images and were asked if they thought the image was a morph or not. Participants were also asked to indicate their confidence in their decision on a 6-point scale where 1 was uncertain (guessing) and 8 was absolute certainty. The results of the experiment showed an average human accuracy of 54.1 percent. Moreover, accuracy was similar across all confidence levels, suggesting that high confidence does not correlate to correct judgements. The work also concluded that training did not necessarily improve human detection capabilities for the task. Nonetheless, training has positive effects on awareness, staff involvement, and collaboration.

While human perception by itself is limited for detecting high quality morphs and other digital forgeries, a better understanding of how humans interact and team with computer algorithms may help improve detection rates and strengthen the integrity of workflows that rely on authentic documents and media. DHS S&T Biometric and Identity Technology Center-sponsored research has shown humans tend to adhere to recommendations made by face recognition algorithms.[67] This cognitive effect and other nuances of human-algorithm teaming are less studied for content forgery problems but will be foundational in understanding how to best integrate solutions to detect forgeries and prevent fraud. Effective human-algorithm teaming is especially important for digital forgery detection applications where human capabilities by themselves are limited; detection capability comes from building human experience with computer tools and techniques combined with a deep understanding of media sources in question.

[65] Nightingale, et al. "Perceptual and Computational Detection of Face Morphing."

[66] "FRVT MORPH." Nist.gov, NIST, pages.nist.gov/frvt/html/frvt_morph.html.

[67] Howard, John J., Laura R. Rabbitt, and Yevgeniy B. Sirotin. 2020. "Human-Algorithm Teaming in Face Recognition: How Algorithm Outcomes Cognitively Bias Human Decision-Making." Edited by Paola Iannello. PLOS ONE 15 (8). https://doi.org/10.1371/journal.pone.0237855.

4. Risks associated with Digital Content Forgeries and Synthetic Media

Since early photography, people have manipulated images for various purposes, as well as for creative enhancement and as a form of expression. Inherently, the technologies used to modify images, videos, and other digital content are not harmful in and of themselves, possessing many positive and Constitutionally protected use cases. Examples are seen in the entertainment industry with puppeteering-based animation, in the service industry with generative text customer support, and even in education with AI allowing the creation of complex anatomy and machinery simulations. Additionally, the manipulation of media for entertainment, comedy, and parody are frequently cited as pre-existing uses, and exceptions that require interpretation and protection under provisions for free speech and fair use.

Unfortunately, when various technologies or methods are used to create realistic renditions of objects and people for nefarious ends, such as identity theft and fraud, they can cause wide-spread harm. Deepfakes are powerful tools that those with harmful agendas can use for exploitation.[68] These tools are easily available and are becoming increasingly difficult to detect as technology continues to progress.

4.1. Deceitful and Harmful Uses

The following sections provide insight on some of the harmful ways digital forgery technology is used, such as with identity fraud and for legal proceedings.

4.1.1. Identity Fraud

Identity fraud refers to crimes in which someone wrongfully obtains and uses another person's personal data, often for financial reasons.[69] There are three effects that result from intentionally manipulating identity-based digital media, depending on the context for how and where the media is used – fake identity, impersonated identity, and de-identified identity. Fake identities and impersonations are harmful forms of fraud when used to circumvent security controls or seek benefits or entitlements that one otherwise would not be eligible to receive.

4.1.1.1. Fake Identity

While some users might simply filter photos for creative or aesthetic purposes, bad actors could use photo editing applications to create fake identities from photo IDs or passport photos. A fake identity is an identity that is fabricated and not known to belong to anyone. Fake and fabricated identities are not a new capability, but its use has broadened as work, communication, news, and entertainment are increasingly digitized. The wide availability of easy-to-use, low or no cost tools to generate high-quality synthetic faces and voices is a more recent development. Although these are not part of an exhaustive list of available photo-editing tools, FaceApp,[70] Oldify, and

[68] Office, U. S. Government Accountability. 2020. "Science & Tech Spotlight: Deepfakes." Www.gao.gov, no. GAO-20-379SP (February). https://www.gao.gov/products/GAO-20-379SP.

[69] US Department of Justice. 2017. "Identity Theft." Justice.gov. February 7, 2017. https://www.justice.gov/criminal-fraud/identity-theft/identity-theft-and-identity-fraud.

[70] "FaceApp - AI Face Editor." n.d. Faceapp.com. https://www.faceapp.com/.

Faceapp![71] are some examples of smartphone-based photo-editing applications that offer users AI filters, backgrounds, and other effects to alter their photos.

In 2019, researchers significantly improved the image quality of generated faces,[72] demonstrating that computer generated faces can be nearly indistinguishable from photos of real human faces on the website https://thispersondoesnotexist.com. Other tools and applications can generate fake names, addresses, and assign fabricated data to other attributes, such as gender, age, email, user ID, occupation, hobbies, and professional interests. Representative examples of these applications include Faker[73]and Fake Person Generator.[74] In a more recent example, security researchers have examined the proliferation of fake identities on professional networking sites by generating fake names, resumes, and profile photos to target corporations and recruiters.[75]

4.1.1.2. Impersonated Identity

An impersonated identity is an instance in which someone intentionally presents themselves as someone else. Impersonated identity is a major factor in financial fraud, fraudulent loans, and unauthorized use of credit. Credit card and payment fraud in the United States totaled an estimated $9.62 billion in 2019 according to the Nilson Report.[76] This estimate does not include fraudulent applications for government benefits and claims. Historically, fraud in unemployment insurance has been largely attributed to issues of eligibility, misrepresentation of seeking work, the terms of work separation, and unreported income. However, the Department of Labor recently stated identity theft as a primary factor in recent fraud targeting unemployment benefits during the pandemic.[77] The amount of improper unemployment payments related to the pandemic is estimated to be in the billions of dollars.[78]

In the cyber domain, the interception, theft, and unauthorized use of personally identifiable information (PII) and identity credentials are enduring problems that enable cybercrime. Phishing, business scams, and predatory romance scams frequently involve online impersonation and use of false accounts. Targets typically receive an email that appears to originate from a government agency or a reputable company, such as a financial institution, that encourages the recipient to click on a link. This can redirect to a fake website made to look official that prompts

[71] Bell, Karissa. 2019. “FaceApp Clones Are Also Going Viral, You Should Still Be Careful.” Mashable. July 18, 2019. https://mashable.com/article/faceapp-alternatives-viral.

[72] Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. “Analyzing and Improving the Image Quality of StyleGAN.” IEEE Xplore. June 1, 2020. https://ieeexplore.ieee.org/document/9156570.

[73] Faraglia, Daniele. 2020. “Joke2k/Faker.” GitHub. December 10, 2020. https://github.com/joke2k/faker.

[74] “Fake Name Generator | Fake Person Generator.” n.d. Www.fakepersongenerator.com. Accessed August 23, 2022. https://www.fakepersongenerator.com/fake-name-generator.

[75] Krebs, Brian. 2022. “Glut of Fake LinkedIn Profiles Pits HR against the Bots – Krebs on Security.” Krebs on Security. October 5, 2022. https://krebsonsecurity.com/2022/10/glut-of-fake-linkedin-profiles-pits-hr-against-the-bots/.

[76] Robertson, David. 2020. Nilson Report. HSN Consultants, Inc.

[77] Costa, Thomas, Mary Hannah Padilla, Seth J. Bagdoyan, Lawrance L. Evans, and Carol C. Harris. 2022. Unemployment Insurance: Transformation Needed to Address Program Design, Infrastructure, and Integrty Risks. Washington DC: United States

[78] Romm, Tony, and Yeganeh Torbati. 2022. "A magnet for rip-off artists: Fraud siphoned billions from pandemic unemployment benefits." The Washington Post. May 15. https://www.washingtonpost.com/us-policy/2022/05/15/unemployment-pandemic-fraud-identity-theft/.

the targeted individual to enter their private information, which could be used to create a false identity.[79]

In the physical domain, impersonated or misrepresented identity presents enforcement challenges to travel security and immigration processing. Multiple aliases can stem from cultural naming conventions, so that responding to or using different names in different situations may not be purposefully deceitful. However, individuals wishing to avoid government or official scrutiny may employ different names and identities. This latter use is problematic, such as during a traffic stop with a police officer or at an airport security check.

A Known and Suspected Terrorist could use an impersonated identity to access a TSA Pre-Check lane. TSA's risk-based approach to its counterterrorism mission relies on effective identity verification. TSA uses identity information to do the following: ensure that individuals on the No Fly List are prevented from boarding an aircraft when flying within, to, from, and over the United States; route individuals who represent an increased risk to enhanced screening; and enable TSA's Pre-Check program, which offers expedited screening to vetted individuals. Additionally, TSA uses identity verification to mitigate the threats from transportation sector insiders. These workers are required to undergo background checks and recurrent vetting to ensure their eligibility to access sensitive systems, equipment, and locations. Additionally, the tactics to compromise and impersonate multiple identities to commit financial crimes could be used by adversaries to conceal their identities when planning an attack against the U.S.

4.1.1.3. De-Identified or Anonymous Identity

De-identified or anonymous identity situations refer to the intentional removal and minimization of information such that recognition or identity verification is untenable – that is, the removal of PII. With some online accounts, such as social media, the default settings provide the least amount of privacy. People avoiding recognition could simply be exercising a desire for privacy and security for their personal data. With the internet making personal information relatively simple to find, people may purposely limit their identifying information online to protect their privacy. From banking data to medical information or even school registrations, personal data is often submitted electronically. Removing identifying information mitigates the individuals' privacy risks, while allowing the data to be used in comparative assessments, policy research, or scientific studies in which the individual's data is not needed.[80] For example, one user could create multiple social media accounts on a single platform because they are not required to provide their PII. Similarly, a bad actor could use a de-identified persona to create multiple accounts for nefarious purposes.

De-identified identity information is designed to create confusion, ambiguity, and mistaken identity. In some situations, de-identified and anonymous identities are associated with malign

[79] "Phishing Attack Prevention: How to Identify & Avoid Phishing Scams." 2019. Occ.gov. April 6, 2019. https://occ.gov/topics/consumers-and-communities/consumer-protection/fraud-resources/phishing-attack-prevention.html.

[80] Rights (OCR), O. for C. (2012, September 7). Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. HHS.gov. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#:~:text=Section%20164.514%20%28a%29%20of%20the%20HIPAA%20Privacy%20Rule

actors and unsolved crimes. Removing identifying information makes it more difficult, although not impossible, to detect the source of the criminal activity.

4.1.1.4. First Party Fraud

First-Party Fraud is a hybrid type of risk that includes elements of both credit and fraud risk. An individual opens a line of credit with a true or false identity, maxes out the credit line, and allows the account to default. Specifically, first party fraud involves an individual who makes a promise of future repayment in exchange for goods and services without the intent to repay.

4.1.2. Digital Media as Legal Evidence

With the number of deepfakes circulating the internet increasing alongside public knowledge of deepfake technology, the ability to separate fact from fiction will also become increasingly difficult, especially due to how realistic these forgeries are. Providing media, such as photos, videos, and audio recordings, is a common method to corroborate one's story during legal proceedings. With the proliferation of technologies to create or modify digital media, the ability of a witness to correctly identify the authenticity of evidence may become more difficult and complicated. Additionally, the assumption that technological experts can distinguish between real and synthetic media is eroded due to the chance that small, yet important, details are missed. Despite technological advances in detection technologies and tools, the proliferation and sophistication of content editing tools will increasingly hinder and erode the integrity of the court,[81] making using media as evidence less reliable.

One example relates to a term coined by Professors Chesney and Citron, the *Liar's Dividend*, which occurs when increased public knowledge of digital editing and deepfake tools makes it easier for liars to avoid accountability for things that are true.[82] The difficulty in determining the integrity and provenance of digital content can make media and recordings no longer universally trusted as truthful, factual events, and thus a liar can cast doubt on true images or recordings, claiming they are false and never happened.

For instance, if an incriminating or damaging video, audio recording, or picture of someone is found, the affected individual could simply deny the authenticity of the media, claiming that it was manipulated or created without their knowledge. Conversely, bad actors can use deepfakes to frame innocent people, and even if the innocent individual tries to denounce the falsities, the public might still believe the forged media. Furthermore, defense teams could legally challenge evidentiary standards for face images, video, and audio voices in court, making it possible for them to claim evidence as unreliable, manipulated, edited, or tampered.

---

[81] LaMonaga, John P. (2020) "A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes," American University Law Review: Vol. 69 : Iss. 6 , Article 5.

Available at: https://digitalcommons.wcl.american.edu/aulr/vol69/iss6/5

[82] Citron, Danielle, and Robert Chesney. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." 107 California Law Review 1753, December. https://scholarship.law.bu.edu/faculty_scholarship/640.

5. Conclusion

Technologies to create or manipulate audio, visual, or text content are not new. Recent advances in AI/ML have enabled a new generation of technologies that have substantially increased the quality of the content and reduced the amount of skill and time required, resulting in the proliferation of extremely realistic digital content. The ever-growing availability of rich mixed media data combined with readily accessible free or commercially available AI/ML capabilities portends a massive increase in fabricated and manipulated digital content. Many of these technologies can be used for legitimate and lawful purposes, including creative expression, entertainment, comedy, and parody, and could be protected under provisions for free speech and fair use.

Unfortunately, these technologies may also be used to create digital content forgeries with the intent to mislead and cause wide-spread harm. Digital content forgery technologies and the resulting content, like deepfakes, are powerful tools that those with harmful agendas can use for exploitation.[83] The core technologies can be subverted by criminal actors employing a range of sophistication from COTS software to specialized tools, all with the intent to misrepresent legitimate textual, audio, and visual content.

The actors who would seek to profit from the use of these technologies are diverse. Criminals are likely to develop creative ways to identify new opportunities to commit fraud. Foreign actors may conduct disinformation campaigns, create false narratives, and recycle existing media to mislead the public and other targeted audiences. Foreign intelligence agencies have long benefited from the creation of false identities, and these technologies can increase the depth and detail of their legends. Violent extremists may circulate disinformation and misinformation on social media to alter public perspectives, influence global affairs, spread falsehoods, and incite violence.

Assessing the broad spectrum of criminal actors and technologies that can be used to threaten our institutions and way of life is ongoing and demands detection and mitigation solutions that are forward looking. Detection and mitigation methods and technologies continue to evolve in an effort to improve detection.

Over the next installments of this assessment, which may include classified annexes, S&T will expand its content to capture the evolving threat landscape, as well as review analytical methods to assess threats, summarize performance of detection and mitigation methods, and examine countermeasures to manage these threats.

[83] Office, U. S. Government Accountability. 2020. “Science & Tech Spotlight: Deepfakes.” Www.gao.gov, no. GAO-20-379SP (February). https://www.gao.gov/products/GAO-20-379SP.

6. Appendix: Acronyms

AI	Artificial Intelligence
APCER	Attack presentation classification error rate
BERT	Bidirectional Encoder Representations from Transformers
CBP	U.S. Customs and Border Protection
BPCER	Bona fide classification error rate
CISA	Cybersecurity and Infrastructure Security Agency
CITeR	National Science Foundation's Center for Identification Technology Research
DD	Deepfakes detection
DNN	Deep Neural Network technology
DHS	Department of Homeland Security
FMR	False Match Rate
GPT-3	Generative Pre-trained Transformer 3
IMD	Image manipulation detection
ISMD	Image splice manipulation detection
ISO	International Organization for Standardization
MD	Manipulation Detection
ML	Machine Learning
NLP	Natural Language Processing
NIST	National Institute of Standards and Technology
OpenMFC	Open Media Forensics Challenge
S&T	Science and Technology Directorate
StegD	Steganography Detection
TSA	Transportation Security Administration
VMD	Video manipulation detection

7. Appendix: Definitions

Attack presentation classification error rate (APCER): is the proportion of presentation attacks incorrectly classified as bona fide by an identity system.

Artificial Intelligence: refers to automated, machine-based technologies with at least some capacity for self-governance that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments

Autoencoder: is an unsupervised neural network that learns how to efficiently compress and encode data. It then learns how to reconstruct the data from the compressed/reduced encoded representation to a representation that is as close to the original input as possible.

Bona fide classification error rate (BPCER): is the proportion of bona fide presentation incorrectly classified as a presentation attack by an identity system (based on ISO 30107-3).

Chatbot: is a computer program that uses artificial intelligence and natural language processing to simulate and process human conversation (written or spoken) with a user. Chatbots are employed through text messages, telephones, websites, mobile applications, and other similar modes of communication.

Commercial off the shelf (COTS): a software and/or hardware product that is commercially ready-made and available for sale, lease, or license to the public.

Computer vision: is a field of artificial intelligence that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, seeking to replicate human vision, observation, and understanding.

Cryptocurrency: is a form of digital currency where transactions and records are verified and maintained by a decentralized system using cryptography (a special system of encrypting and decrypting information).

Dark web: is a part of the World Wide Web that is encrypted and only accessible by means of special software that allows uses and website operators to remain anonymous or untraceable.

Deepfake: a video, photo, or audio recording that seems real but has been manipulated with artificial intelligence technologies.

Deep learning: is a machine learning technique, comprised of a neural network with three or more layers, that enables computers to solve complex problems. This technique attempts to simulate the behavior of the human brain and how the human brain gains knowledge, allowing it to "learn" from large amounts of data. Deep learning allows automation of predictive analytics.

Deep Neural Network (DNN): is a category of machine learning algorithms and is a more complex neural network that has two or more layers.

Face morphing attack: an attack used by threat actors to create a synthetic image with characteristics of two individuals to use as a reference image in a database. The reference image could be used to bypass recognition processes.

Generative Adversarial Network (GAN): is a deep neural network framework used for unsupervised machine learning. If given a large set of data to train from, a GAN can create a new unique set of data that is almost indistinguishable from the original. It does this by contesting two neural networks against each other, in which they learn from each loss and/or gain.

Human-algorithm teaming: regarding face recognition, it refers to when humans "team" with algorithms, by using the algorithm results to assist in making identity decisions.

Interface architecture: refers to the user-centered principles, guidelines and designs used when creating software and machines. The "interface" is a point of human-machine interaction, specifically referring to the areas that users interact with directly.

Machine Learning (ML): is a branch of artificial intelligence that focuses on the use of data and algorithms to give computers the capability to learn without being explicitly programmed.

Motion capture: is a technique to record all or part of a person's movement so that it may be converted into the action of a computer-generated 3D figure on screen.

Natural Language Processing (NLP): refers to a branch of artificial intelligence concerned with giving computers the ability to understand, analyze, and manipulate the human language.

Neural networks: are a subset of machine learning that are the foundation of deep learning. They are computing systems with interconnected nodes that are inspired by the human brain, employing pattern recognition and the passage of input through various layers of simulated neural connections. Layers refer to an input layer and an output layer, which can have layers in between. Each layer performs specific types of sorting and ordering.

Rich media: is a digital advertising term for an ad that includes advanced features like video, audio, or other dynamic visuals and interactive elements that encourage viewers to engage with the content.

Semantic technology: is a set of methods that provide advanced means for categorizing and processing information, helping machines understand data.

Steganography: is the practice of concealing information within another message, image or physical object.

Stego (steganography image): is an image that has been embedded with a secret message via a steganography algorithm. The hidden message is either in pixel values or in optimally selected coefficients.

Synthetic (media): is term for the artificial production, manipulation, and modification of data and media by automated methods, especially artificial intelligence algorithms.