Technology Blog

Approaches to Address AI-enabled Voice Cloning

By: FTC's Office of Technology | April 8, 2024

Today, the FTC announced four winners of the Voice Cloning Challenge, which was launched to address the present and emerging harms of artificial intelligence, or "AI"-enabled voice cloning technologies. The FTC received submissions from a wide range of individuals, teams, and organizations. The winners are outlined in the press release and on the Challenge website.

The agency called for ideas that can be implemented during at least one of the following three intervention points to address the risks of fraud, scams, and misuse of biometric data: (1) upstream prevention or authentication; (2) real-time detection or monitoring; and (3) post-use evaluation of existing content. Solutions were evaluated for their administrability and feasibility to execute, ability to go upstream and reduce consumer burden, and resilience to an evolving technical landscape.

Below, we highlight several potential methods at each stage of intervention. While some of the methods may not differ across the three intervention points, the implementation at each stage does. To that end, we discuss some challenges and benefits of implementing approaches to address the risks of voice cloning—and highlight that there is no single solution.

Leveraging solutions to provide upstream prevention or authentication

Prevention or authentication refers to techniques that limit the application and misuse of voice cloning software by unauthorized users. These interventions are important because they aim to address the core source of potential harm with regards to voice cloning technologies by focusing on ways to systematically authenticate which content is real versus fake before it reaches a consumer.

One commonly discussed approach for prevention and authentication is watermarking, which often refers to a "broad array of techniques" for embedding an identifying mark into a piece of media to track its origin to help prevent the misuse of cloned audio clips.[1] For instance, when embedded by AI developers, such techniques can be used to allow someone to know if a piece of content was generated by AI.

While some implementations of watermarking can help detect and deter harm, there are also limitations. For example, fraudsters may be able to remove a watermark signaling content is inauthentic or distort the audio in order to make their audio clip seem authentic.[2] Invisible or visible watermarks can be altered or removed, potentially rendering them unhelpful for differentiating between real and synthetic content.[3] If, for example, a bank or hospital incorrectly marks a voice as authentic when it is actually cloned, such false positives could be particularly harmful, especially if watermarking is implemented in sensitive settings, such as authentication for a financial or healthcare accounts.[4]

Another option to help address harm further upstream may be establishing clear ways of validating humans who use communication systems – such as phones or videoconferencing platforms. For example, if a person's doctor's office is calling, they could have confidence in the call through a verification scheme that guarantees the call is originating from the correct device—giving assurance to the patient that the voice on the other end of the call is indeed their doctor's. Upstream actors, including handset developers, telecommunication providers, video conferencing providers, and messaging platforms will need to work to ensure that such authentication mechanisms are implemented for verified users.

Applying solutions to detect solutions in real-time

Real-time detection or monitoring includes methods to detect cloned voices or the use of voice cloning technology at the time during which a specific event occurs. Studies reveal a spectrum of efficacy for voice cloning detection solutions.[5] The effectiveness of such solutions is especially important when considering the types of AI-enabled voice cloning scams – such as fraudulent extortion scams – that the technology can enable.

Real-time solutions should also be feasible to deploy. For example, a solution that can be implemented via a Software Development Kit (SDK), would allow for more developers to incorporate the solution. In addition, solutions that are less resource intensive could be more widely implemented.[6] For example, detection software that is able to run directly on consumer devices – like phones and

laptops, as opposed to software that requires complex hardware – could give consumers real-time feedback to know if the voice on the other end of the line is real.

Using solutions to evaluate existing content

In an ideal world, upstream actors would have prevented nefarious voice clones by authenticating the voice or detecting it in real-time. However, we already live in a world where troves of AI-generated content exist and are accessible to anyone – and voice cloning tools are more readily available every day.

The post-use evaluation of existing content includes methods to check if already-created audio clips such as voice mail messages and audio direct messages contain cloned voices. One potential way to evaluate existing audio clips is to develop algorithms that detect inconsistencies in voice cloned clips.[7] For example, some voice cloning evaluation tools look for audio signals, like abnormal soundwaves in an audio file, to detect whether a clip is real or fake.[8]

As voice cloning technology develops, fraudsters will continue to try to avoid detection, and those developing detection technology will continue to fight it. As voice cloning technology becomes more sophisticated and these signals become harder to pick up, current evaluation schemes may need to be adapted in order to be effective in detecting cloned voices.

Looking forward: Preventing and deterring AI-enabled voice cloning scams and fraud

To select the winners, the judges and FTC analyzed the concepts' strengths, limitations and workarounds to ensure that the response was robust enough to combat the evolving risks of voice cloning. While there are many exciting ideas with great potential, there's still no silver bullet to prevent the harms posed by voice cloning. The Challenge recognizes people who are pushing science and proposing different options so that society has a varied landscape of solutions; but more innovation and testing is always needed. The top solutions can help highlight possible ways to prevent fraudulent uses of AI-enabled voice cloning technologies today, as well as show a path to how concepts like these can move forward to implementation across the three stages of prevention and authentication, real-time detection, and post-use evaluation of existing content.

Any solution to voice cloning-enabled scams will require engagement with key partners – including the FCC, DOJ, and states, as well as the Industry Traceback Group and other industry stakeholders, to maximize the tools and abilities each have. It is also necessary to incentivize actors to comply with and implement new technologies and policies.

As seen with the agency's recent history on robocalls, the FTC can be more effective by going upstream. A strong approach to AI-enabled voice cloning ensures that AI companies releasing tools that have the potential for misuse may be held liable for assisting and facilitating illegal activity in certain circumstances if they do not implement guardrails to prevent it. Further, voice service providers – telephone and VoIP companies – need to continue making progress against illegal calls.

We stand at an inflection point. The FTC has made clear that it is prepared to use all tools to hold bad actors accountable. That includes sponsoring this Challenge, as well as law enforcement under Section 5 of the FTC Act and the Telemarketing Sales Rule. In addition, the Commission has recently enacted a new Impersonation Rule, which will give the agency additional tools to deter and halt deceptive voice cloning practices. The Commission also is considering the adoption of other rules, such as the pending Consumer Reviews and Testimonials Rule, which similarly would expand the agency's toolkit. There is no AI exemption from the laws on the books and the FTC remains committed to protecting consumers from the misuse of "AI"-enabled voice cloning technologies.

---

[1] https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/

[2] https://arxiv.org/pdf/2306.01953.pdf

[3] https://www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/

[4] https://www.banking.senate.gov/imo/media/doc/voice_cloning_financial_scams.pdf

[5] https://arxiv.org/pdf/2307.07683.pdf; https://arxiv.org/pdf/2005.13770.pdf

[6] https://arxiv.org/pdf/2308.14970.pdf

[7] https://arxiv.org/pdf/2402.18085v1.pdf

[8] https://arxiv.org/pdf/2307.07683.pdf; https://arxiv.org/pdf/2402.18085v1.pdf

Tags: Technology | Artificial Intelligence | Office of Technology

More from the Technology Blog

Technology Blog

AI and the Risk of Consumer Harm

Staff in the Office of Technology and the Division of Advertising Practices | January 3, 2025

Technology Blog

Lenses of security: Preventing and mitigating digital security risks through data management, software development, and product design for humans

Staff in the Office of Technology & the Division of Privacy and Identity Protection | December 13, 2024

Technology Blog

Solving the Traveling Salesman Problem? Not quite, but here are more research questions from the Office of Technology

Staff at the Office of Technology | December 4, 2024

Technology Blog

Unpacking Real Time Bidding through FTC's case on Mobilewalla

Staff in the Office of Technology & Division of Privacy and Identity Protection | December 3, 2024