Written Testimony of Zohaib Ahmed, CEO and Founder of Resemble AI

For a hearing on "Oversight of AI: Election Deepfakes" Before the Judiciary Committee

Subcommittee on Privacy, Technology, and the Law
United States Senate
April 16th, 2024

Introduction

Chairman Blumenthal, Ranking Member Hawley, and Members of the Committee, thank you for the opportunity to discuss the oversight of AI as it relates to understanding the impact this technology can have on elections with you. I'm Zohaib Ahmed, CEO and Founder of Resemble AI.

Resemble AI is a research and development lab focused on the creation of Generative Voice AI models and is being used by some of the largest media, gaming, financial and telecom companies as well as content creators.

Our team has spent the last five years developing and researching AI voice technology and we are uniquely positioned to understand both the remarkable potential and possible risks associated with the rapid advancement of voice synthesis and cloning capabilities. It is with this balanced perspective that Resemble AI has created innovative solutions to address the emerging challenges posed by unauthorized or unethical uses of voice cloning technology.

Given our tenure in responsible voice cloning technology, we started building speaker identification, watermarking, and deepfake technology to enable safer deployment of voice AI products. Over the last nine months, we have opened up the research on speaker identification, watermarking and deepfake detection to a broader audience to help ensure that Generative AI is safely and responsibly deployed around the world.

I'd like to address the critical role of responsible technology practices in shaping the future of AI. We have always maintained a commitment to ethical standards, including transparency, privacy, and security; and these principles guide our product development and user engagement.

In this written testimony, I will share some of the technologies that we have developed since Resemble AI was founded in 2019, especially around watermarking and deepfake detection, and my recommendations around 1) transparency and disclosure, 2) safeguards and mitigation, and 3) integrity verification.

Ethics and Consent

At Resemble AI, we've always held ourselves to an exceptional standard of ethics. To uphold our mission to help the private and public sector use generative AI for good, we developed the following guardrails early on to maintain our ethical standards, including:

- a dedicated team that is responsible for ensuring that our voice cloning technology is used ethically and responsibly
- policies and guidelines that we require our customers to adhere to, including a consent line that must be recorded in the same voice that is being used to generate new audio
- policies and guidelines are designed to ensure that our technology is used in a way that is respectful of people's privacy and data rights and that it is not used in a way that might cause harm
- technological improvements that detect AI-generated media including our neural watermarker and AI-speech detector

At Resemble AI, we require users to provide a recording of a consent clip in the voice they are attempting to clone. If the voice in this clip does not match the other clips, the user is blocked from creating the AI voice.

When recording, Resemble AI requires the user to say an array of particular sentences in your own voice. Misuse of this can be easily detected by our algorithm. Once the voice is created, the user owns all rights to that voice. We do not use that voice data to train other models, nor do we resell the voice data to third party companies.

For customized solutions, we work with companies through a rigorous process to make sure that the voice they are cloning is usable by them and have the proper consents in place with voice actors.

Materials used through integration of Resemble and related metadata must be produced by the publisher itself, correctly licensed from the third–party rights holder, used as allowed by the rights holder, or legally used in any other way.

In our terms of service, we state that you can not use AI voices built by Resemble for:
- claiming to be any person, company, administration, or entity without explicit authorization to make this statement and/or impersonating to gain illegal information or privileges;
- propagating hate speech;
- discrimination, libel, terrorism, or violent activities;
- spreading unattributed content or misrepresenting sources.
- exploiting or manipulating children;
- making unsolicited phone calls, vast communications, postings, or messages;
- deceiving or deliberately misleading people;

Watermarking
After creating open source speaker identification with Resemblyzer several years ago, in 2023 we took it several steps further by introducing the Neural Speech Watermarker, an "invisible watermark" that tackles the malicious use of AI generated voices. With a deep neural network watermarker, the data is embedded in an imperceptible and difficult-to-detect way, acting as an "invisible watermark."

To deploy safe neural speech in the wild, Resemble AI introduced the PerTh Watermarker, a deep neural network watermarker. It uses machine learning models to both embed packets of data into the speech content that we generate, and recover said data at a later point. Because the data is imperceptible, while being tightly coupled to the speech information, it is both difficult to remove, and provides a way to verify if a given clip was generated by Resemble. Importantly, this “watermarking” technique is also tolerant of various audio manipulations like speeding up, slowing down, converting to compressed formats like MP3, etc.

Our model is called PerTh, which is a combination of Perceptual and Threshold. It was designed around the concept of exploiting the way we perceive audio to find sounds that are inaudible, and then encoding data into these regions. Further care is taken to ensure we can extract the embedded data from any part of the audio (aside from silence), and that the data is encoded into frequencies most common for speech. This ensures our data payload is difficult to corrupt with common audio manipulations.

We use Psychoacoustics (the study of human sound perception) to find the best way to encode the data. One psychoacoustic phenomenon is that we have varying sensitivity to different frequencies — this means we can embed more information into frequencies we are less sensitive to. Another, more interesting effect called “auditory masking” also exists, in which quiet sounds nearby in frequency and time to a louder sound are not perceived. As a result, the masking sound (the louder sound) produces a “blanket” in amplitude-frequency-time space that drowns out other sounds beneath it.

Further research of our watermark supports the traceability of data. Our watermark has proven that it can persist through model training, therefore the output of a Generative AI model that uses watermarked data can be traced back to the source.

Deepfake Detection

Resemble AI's Generative Product, encompassing Voice Cloning combined with text-to-speech or speech-to-speech technologies, is extensively utilized by enterprise customers for conducting red-team exercises. These exercises are critical for security testing, where the technology's ability to replicate voices accurately is used in simulated attack scenarios to assess and improve organizational security measures. By creating realistic voice interactions, the technology aids in identifying potential vulnerabilities, thereby enabling companies to enhance their security protocols and safeguard against actual voice-based security threats.

In July 2023, we launched Resemble Detect, our advanced deepfake detection tool, which utilizes a cutting-edge neural network to differentiate between real and fake audio files, addressing the increasing concern of deepfakes in various sectors. Resemble Detect is designed to work in real-time, providing up to 98% accuracy in exposing deepfake audio, thereby safeguarding the authenticity of voice content.

For telecom providers, Resemble Detect can uncover social engineering fraud attempts using cloned voices in phone calls. These can be flagged for enhanced screening to protect customers against pretexting scams.

To further support our enterprise customers, many of whom serve as a gateway to consumer safety, we launched our Deepfake Detection Dashboard to help security teams monitor deepfake activity in a central place.

We acknowledge that consumer education and awareness is another critical piece to addressing the risk of misinformation. Starting last fall, we also began publishing weekly Deepfake Incident Reports to bring transparency to the use of deepfakes using our Detect technology. We make these available on our blog and offer our reporting to journalists to accelerate the identification of deepfakes in mainstream media channels.

We also joined the FTC's Voice Cloning Challenge and signed Canada's voluntary AI code of conduct, underscoring our commitment to accountability, safety, fairness, transparency, human oversight, and robustness in AI systems.

Our free, real-time Deepfake Detector tool is available for anyone to combat fraud and promote the responsible use of generative voice technologies. With this tool, you can quickly verify the authenticity of widely circulated audio content, making it a valuable asset for journalists, content creators, and the general public who are often on the frontline of combating misinformation.

Our solution also helps enforce the recent FCC ruling against robocalls by providing a means to verify the authenticity of audio content. It can serve as a line of defense for businesses and individuals who wish to ensure that the voices they are hearing in calls are genuine and not AI-generated deepfakes intended to deceive.

And just last week, we added real-time deepfake detection for meetings, starting with Google Meet, that leverages our Detect technology. By integrating into Google Meet, our Detect solution is now available on a platform that people are already familiar with and it provides them with a seamless detection experience that can give them a piece of mind they are protected from deepfakes and ease AI fraud concerns.

Policy Recommendations

I believe your legislative framework is a comprehensive approach and positive step towards effectively regulating AI. Given the rapid advancements in generative models and the surge in synthetic voice technology, especially during this crucial election period, I would like to propose additional recommendations for your thoughtful consideration as you continue to refine legislation in this area:

1. **Transparency and Disclosure**

First and foremost, we support the proposed legislation that requires clear labeling on AI-generated content in the election process. Similarly to how disclaimers appear at the end of

political ads, consumers should be made aware that they are interacting with an AI model or AI-generated content.

To take it one step further, we propose the creation of a public database where all AI-generated election content is registered, allowing voters to easily access information about the origin and nature of the content they encounter.

Furthermore, we believe that voter education initiatives are crucial in promoting transparency. Resemble AI recommends the development of public awareness campaigns that inform voters about the existence and potential impact of AI-generated content in elections. These campaigns should provide clear examples of how AI can be used in election-related communications, both positively and negatively, and equip voters with the tools to critically evaluate the information they receive. Language translation is an example of how voice cloning technology can be used responsibly in an election year.

2. **Safeguards and Mitigation**

To adequately safeguard against misinformation, particularly during critical events like elections, collaboration is key. We see additional opportunities for bipartisan support, as well as private and public sector partnerships:

At Resemble AI, we know firsthand that deepfake detection technology is a powerful tool, capable of providing crucial context and labeling to identify potentially misleading or AI-generated content. By distributing and seamlessly integrating this type of technology into widely-used platforms, as seen in today's phone spam filters, we can protect more Americans regardless of their own resources or knowledge in this area.

We've also seen the effectiveness of red team exercises, that have a sole purpose to test the robustness of AI systems by simulating real-world cyber threats and attacks.The red team can create a deepfake audio, video, or image and distribute it within the organization. The blue team, who is working toward improving an organization's security, is then tasked with detecting the deepfake and responding appropriately. This will help assess the effectiveness of the organization's detection and response systems.

Moreover, we suggest the creation of a national task force composed of experts from the private and public sectors, including representatives from AI companies, government agencies, academic institutions, and civil society organizations. This task force would be responsible for developing best practices and standards for the use of AI in election-related content, as well as coordinating efforts to detect and counter AI-generated misinformation.

To further strengthen security measures, Resemble AI supports the implementation of strict penalties for the malicious use of AI in election-related content. This could include fines, legal action, and the revocation of broadcasting licenses for media outlets that knowingly distribute AI-generated misinformation. By establishing clear consequences for the misuse of AI, we can

deter bad actors and create a stronger incentive for compliance with transparency and labeling requirements.

3. **Integrity Verification**

We believe that AI watermarking technology is a readily available solution to verify the integrity of audio content in this current election year, ensuring that voters can distinguish between voices created with consent and unauthorized deepfakes.

We strongly advocate for the implementation of AI watermarking technology, such as our Neural Speech Watermarker, to verify the integrity of audio content in elections. We propose that all election-related audio content, including political advertisements, campaign messages, and public statements by candidates, be watermarked using this technology. The watermark would serve as a tamper-evident seal, allowing voters and election officials to easily verify the authenticity of the content.

To facilitate the widespread adoption of AI watermarking, we recommend the establishment of a certification program for AI watermarking technologies. This program would set standards for the effectiveness and reliability of watermarking solutions, ensuring that only trusted and vetted technologies are used in the election process. Additionally, we propose the creation of a centralized database where watermarked election-related audio content is registered and can be easily accessed by the public for verification purposes.

Furthermore, we recommend the development of user-friendly tools and platforms that allow voters to easily check the authenticity of election-related audio content. This could include a mobile app or website where users can upload suspicious audio files and receive an instant verification of their authenticity. By empowering voters to take an active role in verifying the integrity of the content they encounter, we can create a more resilient and trustworthy election ecosystem.

To ensure the effectiveness of AI watermarking in preventing deepfakes, we propose regular security audits and stress tests of watermarking technologies. These assessments would be conducted by independent security experts and would simulate real-world scenarios to identify potential vulnerabilities and areas for improvement. By continuously evaluating and strengthening the integrity verification measures in place, we can stay ahead of the evolving tactics of malicious actors and maintain the highest standards of security.

Conclusion

Thank you for the opportunity to provide insight into voice cloning technology and preventative measures that can be taken now to ensure the integrity of this year’s election. We are always willing to help facilitate partnership between the private and public sectors to ensure today’s innovation is used responsibly.