Written Testimony of Rijul Gupta

Founder and CEO, Deep Media, Inc.

Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law

Hearing on "Oversight of AI: Election Deepfakes"

April 16, 2024

Introduction

Chairman Blumenthal, Ranking Member Hawley, and distinguished members of the Subcommittee, thank you for the opportunity to testify today on the critical issue of AI-generated Deepfakes and their potential impact on U.S. elections. My name is Rijul Gupta, and I am the Founder and CEO of Deep Media, Inc., a leading Deepfake Detection and AI Security company at the forefront of combating the threats posed by unethical AI and Deepfake misinformation.

As an expert in the field with over 15 years of experience developing cutting-edge AI algorithms, I am **deeply** concerned about the rapid proliferation of Deepfakes and the significant risks they pose to our democracy. Today, I will provide an overview of Deepfake technology, discuss its potential impact on elections, and propose solutions to address this growing threat.

Defining Deepfakes

Deepfakes are images, audio, or videos that have been generated or manipulated by AI in a manner that may harm or mislead the public. It is important to note that Deepfakes do not include AI-generated text, such as the output of language models like ChatGPT. While

AI-generated text is a related technology, it poses distinct challenges and should be addressed separately.

The human brain is uniquely susceptible to manipulation by AI-generated images, audio, and video in ways that it cannot be influenced by text alone. Deepfake images, audio, and video present a distortion of truth, they present themselves as reality, and in doing so **undermine** truth in a powerful and dangerous way. Deepfakes represent a form of "Counterfeit Truth" that can distort our perception of reality and undermine trust in the very information we consume.

Technology Overview

Modern Deepfake generators employ a combination of advanced AI techniques, including utilizing Transformers, Generative Adversarial Networks (GANs), and Diffusion Models. These models are trained on vast datasets containing **millions** images, audios, and videos including millions of unique human identities and can process hundreds of thousands of samples in just a few hours, enabling them to create highly realistic imitations of human motion and appearance.

In addition to these core AI technologies, Deepfakes often incorporate traditional post-production techniques such as Photoshop, After Effects, and sound design to further enhance their realism, making them nearly indistinguishable to the human eye.

While "Shallowfakes" – simple Deepfake manipulations like filters or image generation – may seem less threatening, they pose a unique and significant risk due to their scale and accessibility. Shallowfakes are highly effective at deceiving viewers on low-resolution phone screens or in

areas with poor internet connectivity, and can be created by anyone with minimal to no training. In concert with higher quality Deepfakes, which remain particularly dangerous because of their near-perfect quality, it becomes increasingly difficult to distinguish manipulated from genuine content. Both Shallowfakes and Deepfakes present unique challenges, and the gap between them is rapidly narrowing as the technology advances, making it crucial to address the **full spectrum** of AI-generated media.

Proliferation of Deepfakes on Social Media

The rapid advancement of Deepfake technology has led to an **exponential increase** in the realism, affordability, and prevalence of Deepfakes on social media platforms. Research and development efforts in the AI industry are continually making Deepfakes more realistic and cheaper to produce, with these advancements typically taking 6-12 months to reach end-users.

Given the substantial investments in AI research over the past six months, we anticipate major problems arising from Deepfakes in the next 3-6 months. Election seasons worldwide have already sparked a significant increase in Deepfake creation and interest, particularly from malicious actors. Disturbingly, many pornographic or unsafe Deepfake models are freely available to **anyone** with an internet connection.

Impact of Deepfakes on US and Global Elections

While the United States is not unique in facing the challenges posed by Deepfakes, their potential impact on our elections is particularly concerning. Deepfakes can and are used for voter intimidation and confusion, as exemplified by the Biden robocall incident. They can also be

employed for virtual political assassination, such as the circulation of fake images depicting Donald Trump's arrest. Deepfakes can redirect narratives, as seen in the fabricated Hillary Clinton endorsement of Ron DeSantis, or create the illusion of groundswell support, like the manipulated images of Trump with Black supporters or Eric Adams speaking Spanish.

However, the most alarming aspect of Deepfakes is their ability to provide bad actors with plausible deniability, allowing them to dismiss genuine content as fake. This erosion of public trust strikes at the very core of our social fabric and the foundations of our democracy. The human brain, wired to believe what it sees and hears, is particularly vulnerable to the deception of Deepfakes. As these technologies become increasingly sophisticated, they threaten to undermine the shared sense of reality that underpins our society, creating a climate of uncertainty and skepticism where citizens are left questioning the veracity of every piece of information they encounter. In a world where the **very nature of truth** is called into question, the foundations of our democratic institutions, which rely on an informed and engaged citizenry, are at risk.

Deepfake Solutions

On the technological front, while adding clear metadata and watermarking manipulated images have served as important first steps in combating Deepfakes, they alone are not sufficient. These techniques can be easily circumvented by bad actors, and as Deepfake technology continues to evolve, more robust and comprehensive solutions will be necessary to effectively detect and mitigate the threat posed by these malicious AI-generated media.

The development of advanced Deepfake detection platforms on the other hand have proven critical, and show great promise for detecting even the highest quality Deepfakes. Just as the models used to create Deepfakes continue to evolve, so too must our detection efforts. Investigative journalists, such as Donnie O'Sullivan at CNN, Geoff Fowler and the Washington Post, and Amanda Florian at Forbes, play a vital role in uncovering and exposing Deepfakes.

Government entities, including the military and intelligence community, must also actively engage in Deepfake detection and mitigation. Independent community organizations like Witness, the DARPA AI Force, the Content Authenticity Initiative, and Deep Media-led coalitions are essential in coordinating research, sharing best practices, and advocating for effective policies.

Concluding Remarks

As we navigate the challenges posed by Deepfakes, it is important to recognize that the growth of Generative AI and the development of appropriate policies and regulations **need not be in conflict**. Our Generative AI companies represent an economic boon in the global competition with China and other near-peer nations and play a crucial role in protecting the US from rapidly proliferating external AI-related threats.

However, Deepfakes represent a clear market failure – an abuse of a public good that creates negative externalities and erodes trust in the information era. By internalizing these negative

outcomes through smart regulation and industry collaboration, we can accelerate the growth of the Generative AI market while ensuring its safety and integrity.

I **believe** in the positive potential of AI, and I am optimistic about the future. The fact that we are holding this hearing today is a testament to our collective commitment to addressing the challenges posed by Deepfakes. By working together to implement effective solutions, we can harness the power of AI to enrich our lives, improve our political discourse, and build a brighter future for all.

Thank you for your attention and the opportunity to contribute to this critical discussion.