Senate Judiciary Subcommittee on

Privacy, Technology, and the Law

Hearing on

"Oversight of AI: Election Deepfakes"

April 16, 2024

Questions for the Record

Senator Amy

Klobuchar

Rijul Gupta, CEO, Deep Media

Can you expand on your testimony about the risks that AI-generated Deepfakes pose to our elections, including how watermarking alone is not sufficient?

Deep Media Response:

Thank you for the opportunity to expand upon Rijul's testimony regarding the risks that AI-generated Deepfakes pose to our elections and the importance of a multi-layered defense strategy.

Watermarking is a valuable and powerful tool, but it can be circumvented in several ways. Deepfake algorithms can be trained to identify and remove watermarks, rendering them ineffective. Additionally, if a watermark is not embedded robustly enough, it may be lost during compression or other transformations that the media undergoes when shared online. Furthermore, watermarks are only effective if they are consistently applied to authentic media, which requires widespread adoption and standardization. For these reasons, Deep Media believes only a robust Defense in Depth approach can overcome the shortcomings of individual mitigation solutions.

Another powerful tool for detecting AI-generated Deepfakes is metadata analysis. While useful, it is not foolproof because metadata can be easily manipulated or stripped from an image or video file. Deepfake creators can intentionally alter or remove metadata to hide their tracks, making it appear as though the manipulated content is authentic. Moreover, the absence of metadata does not necessarily indicate a Deepfake, as some genuine media may lack metadata due to how it was captured or processed.

Lastly, cryptographic hashing provides another piece of the puzzle by creating unique fingerprints for media verification, but it is not a one-size-fits-all solution. While hashing can

detect changes to a file, it cannot determine the nature or intent of those changes. A manipulated Deepfake and an authentic video that has been slightly compressed or resized may produce different hashes, making it difficult to distinguish between the two. Moreover, hashing is only effective if there is a trusted database of hashes for authentic media to compare against, which requires significant coordination and maintenance.

Recent discoveries confirming China and Iran's Deepfake influence targeting the US 2020 election highlight the importance of proactive measures required for our democratic integrity. From misleading campaign videos, fake endorsements, and disinformation campaigns to fraudulent speeches, manipulated debate footage, and viral fake news, Deepfakes can manipulate public opinion, damage candidate reputations, and sow confusion among voters.

Additionally, Deepfakes can sabotage voter trust, cast doubt on electoral integrity, and even lead to identity theft or targeted manipulation of swing voters. As these scenarios illustrate, the impact of Deepfakes on elections cannot be underestimated, necessitating ongoing vigilance and the implementation of effective measures to detect and combat their influence. Identifying key characteristics of Deepfakes is critical to developing and strengthening our defensive posture.

This is why we believe that a comprehensive, Defense in Depth strategy for Deepfake Detection is necessary. By combining these various techniques in concert with each other, we believe we can create a more resilient defense that addresses the limitations of each individual technique.

By employing these strategies in a cohesive manner, we can create a robust, adaptable defense against the ever-evolving threat of Deepfakes to our electoral process. It is only through this multi-layered approach that we can effectively safeguard the integrity of our elections from the corrosive influence of AI-generated disinformation.