Synthetic Media & Deepfakes

Manipulated imagery has been around for over 150 years but it has reached a new level with "deepfakes." The term "deepfake" originated in 2017 to describe audio and videos manipulated with the assistance of artificial intelligence (AI) to resemble a real person even when the person portrayed did not say or do what is depicted in the content. Deepfakes are a subset of "synthetic media" which include audio, images, text and video content created with the assistance of AI. There continue to be conversations about how to differentiate synthetic audiovisual content from deepfake audiovisual content.

Key Takeaways

I. Understanding the nuances between synthetic media and deepfakes is crucial. Considerations include deception, intent, harm, and consent. Like many topics CNTI covers, the definitional clarity of these terms remains a work in progress and is important when considering policy.

II. New labeling and disclosure tools (e.g., watermarking) are making methods for addressing deepfakes more accessible globally. These tools assist in asserting provenance, or displaying the authenticity/origin of content. However, these approaches are likely not silver bullets as research finds provenance is not fully understood by the public.

III. Most countries do not currently have any existing policies that specifically target deepfakes. It is important that current and future efforts to do so are compatible with laws protecting freedom of speech and expression.

Key Research

I. Examining the impact of provenance-enabled media on trust and accuracy perceptions
Proceedings of the ACM on Human-Computer Interaction (2023)
Summary: The use of provenance information can lower the public's trust in deceptive or synthetic content, though incomplete provenance information can decrease trust in accurate media content as well.

II. Deepfake detection with and without content warnings Royal Society Open Science (2023)
Summary: Individuals are unlikely to recognize content that is, in fact, a deepfake even with the inclusion of a content warning.

III. Deepfakes and scientific knowledge dissemination Scientific Reports (2023)
Summary: The authors find both older individuals and those more trusting of information sources are more susceptible to believing deepfake content related to climate change is authentic.

IV. Deepfakes and the epistemic apocalypse Synthese (2023)
Summary: A philosophical overview of how deepfakes are similar to other forms of manipulated media witnessed throughout history.

V. The new political ad machine: Policy frameworks for political ads in the age of AI
UNC Center on Technology Policy (2023)
Summary: This report examines how AI-generated and manipulated media may influence politics and offers ten policy recommendations.

Synthetic Media & Deepfakes

Legislative Activity

Country	Legislation
China	The Deep Synthesis Provisions update prohibited creating deepfake content without the explicit consent of the individual(s) being altered.
European Union	The AI Act presents a classification scheme for AI usage based on risk level and requires disclosure of content that has been "artificially generated or manipulated."
South Korea	Concerns about how deepfakes may influence South Korea's 2024 parliamentary elections have led to banning this type of content during the 90 days prior to the elections.
United Kingdom	The Online Safety Act, passed in October 2023, stipulates criminal offenses for transmitting and sharing deepfake adult content, along with numerous other safety provisions. The act has a broad focus on online safety but does not include "deepfake" or "synthetic" in defining this type of content of interest, instead using terms like "automated tools."
United States	U.S. legislation related to manipulated media has been proposed at both the state and national level, with several bills at the national level specifically considering deepfake content. The DEEPFAKES Accountability Act would require disclosure when producing synthetic deepfake content. Related bills under consideration include the Protect Elections from Deceptive AI Act, the Advisory for AI-Generated Content Act, and the AI Labeling Act.