The Technological, Social, and Ethical Impact of Deepfakes

Paula V. Cardona

School of Communication, University of Miami

Abstract

Deepfake technology presents unprecedented challenges to media integrity in 2024. This research analyzes how artificial intelligence manipulates audio and visual content, examining recent cases from 2022-2024 where deepfakes influenced elections, spread misinformation, and enabled harassment. I investigated three key areas: technological advancement of deepfake creation tools, current detection methods, and emerging policy responses. My findings indicate deepfakes impact women disproportionately - with recent analysis showing 99% of deepfake adult videos feature women (KTimes, 2024). The research reveals critical gaps in current legislation and platform policies. Based on analysis of expert interviews and case studies, I recommend implementing mandatory digital watermarking, expanding platform liability, and strengthening user consent requirements for facial data collection. This study fills an important gap in understanding how media organizations should address deepfake content while maintaining public trust.

The Technological, Social, and Ethical Impact of Deepfakes

In January 2024, thousands of New Hampshire voters received a robocall from what sounded like President Biden telling them not to vote in the primary election. The voice urged, "Your vote makes a difference in November, not this Tuesday" (Han, 2024). This incident marked a significant escalation in the use of artificial intelligence to manipulate media for political purposes.

Deepfakes now challenge our fundamental ability to trust what we see and hear. As a media management student and future industry professional, I recognize how these technologies transform the media landscape. Through analysis of recent case studies, expert interviews, and emerging technological developments between 2022-2024, this study addresses three core questions:

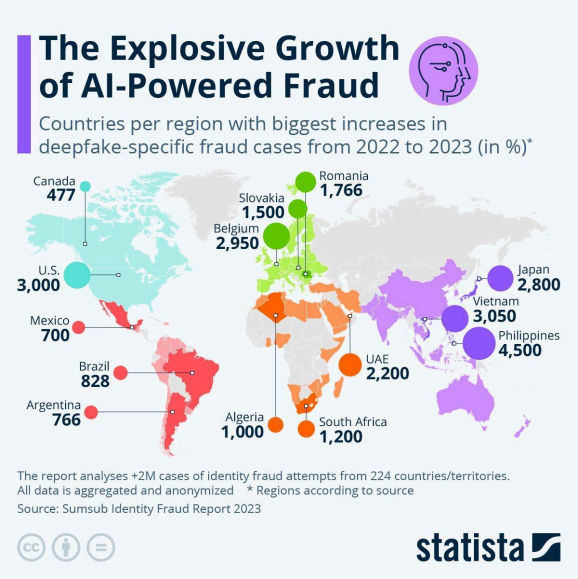
1. How do current deepfake technologies work, and what makes them different from previous forms of media manipulation?
2. What specific threats do deepfakes pose to media integrity, personal privacy, and democratic processes?
3. What solutions - technical, legal, and social - offer the most promise for addressing these threats?

The Sumsub Identity Fraud Report 2023 (Statista, 2024) reveals alarming growth in deepfake-specific fraud cases worldwide. Belgium experienced a 2,950% increase in cases, while the Philippines saw a 4,500% surge. The United States reported a 3,000% rise in deepfake fraud

incidents (see Figure 1). These statistics demonstrate the explosive growth of AI-powered deception across global markets.

Figure 1

The Explosive Growth of AI-Powered Fraud



The scope of deepfake proliferation extends beyond fraud. Recent data emphasizes the urgency of addressing deepfake technology. The amount of adult content created using deepfakes increased by 464% from 2022 to 2023, rising from 3,725 instances to 21,019 (KTimes, 2024). South Korea faces particular vulnerability, with 53% of individuals appearing in deepfake content being South Korean nationals. Eight out of the top ten targets worldwide are Korean singers (KTimes, 2024).

Deepfakes pose three distinct challenges. First, AI tools make deepfake creation increasingly accessible while detection technology struggles to keep pace. Second, the social impact disproportionately affects vulnerable populations - 99% of deepfake adult videos target women (KTimes, 2024). Third, limited legal frameworks exist for deepfake creation and distribution, leaving platforms and users without clear guidelines.

This research examines three critical challenges:

1. Technical Evolution: AI tools make deepfake creation increasingly accessible while detection technology lags behind.

2. Social Impact: Deepfakes threaten personal privacy, reputation, and democratic discourse.

3. Regulatory Gaps: Limited legal frameworks leave platforms and users vulnerable to deepfake exploitation.

Through systematic analysis of recent cases, expert interviews, and emerging technologies, this study provides media organizations with strategies to:

- Verify content authenticity
- Protect vulnerable populations
- Maintain public trust
- Combat disinformation

As you read this research, consider:

- How does deepfake technology affect your trust in media content?
- What responsibilities do media organizations hold in verifying content authenticity?
- How should society balance innovation with protection from harm?

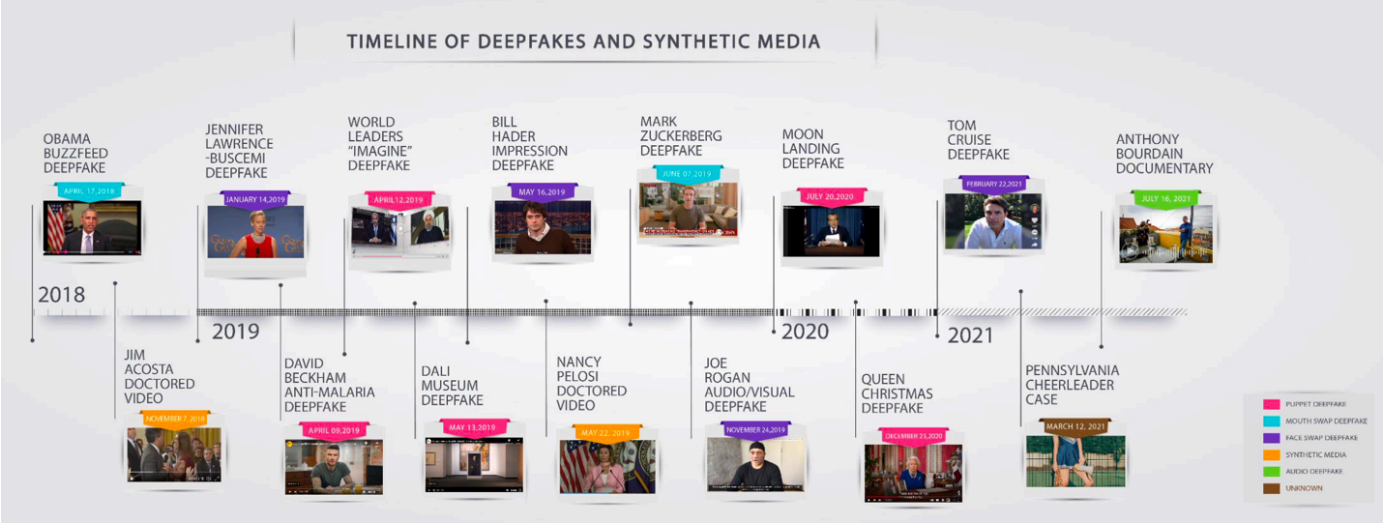
Technological Aspects of Deepfakes

Evolution of Deepfake Technology

The first widely known deepfakes appeared in 2017 when a Reddit user uploaded manipulated videos featuring celebrity faces grafted onto other actors' bodies (Paris & Donovan, 2019). Since then, deepfake technology has evolved rapidly.

Figure 2

Deepfakes and Synthetic Media



Note: The timeline "provides a listing of some of the most well-known and representative examples of deepfakes, as well as some "cheap fakes" and one example of an instance in which deepfakes were initially implicated, but never proven to have been used" (Brooks et al., 2019).

Deepfakes rely on three primary AI technologies:

1. Generative Adversarial Networks (GANs)

- Two AI networks work against each other
- One network creates fake content
- Second network tries to detect the fake
- Process continues until fakes become undetectable (Byman et al., 2023)

2. Deep Neural Networks

- Process vast amounts of facial data
- Learn patterns in human expressions
- Generate realistic facial movements (Heikkilä, 2024)

3. Voice Synthesis Models

- Analyze voice patterns and intonation
- Clone voices with minimal source material
- Create realistic audio from text (Bond, 2024)

Recent advances have dramatically reduced the technical barriers to deepfake creation. The Biden robocall incident in New Hampshire demonstrated how AI voice synthesis now requires only short audio samples to create convincing fakes (Han, 2024).

Applications of Deepfake Technology

The applications of deepfake technology seem to be able to be categorized under two categories: Commercial Applications and Harmful Applications.

Commercial Applications

Entertainment and media companies embrace deepfake technology for legitimate business purposes. Movie studios use deepfakes to recreate deceased actors or make actors appear younger, as seen in recent Star Wars films (Rosner, 2021). Marketing campaigns leverage the technology for personalized content - David Beckham appeared to speak nine languages in a 2023 anti-malaria campaign through AI voice synthesis. Corporate training has evolved through companies like Synthesia, which creates AI-generated training videos in multiple languages from a single recording. Educational applications include historical reenactments and language learning tools, offering new ways to engage students with interactive content (Heikkilä, 2024).

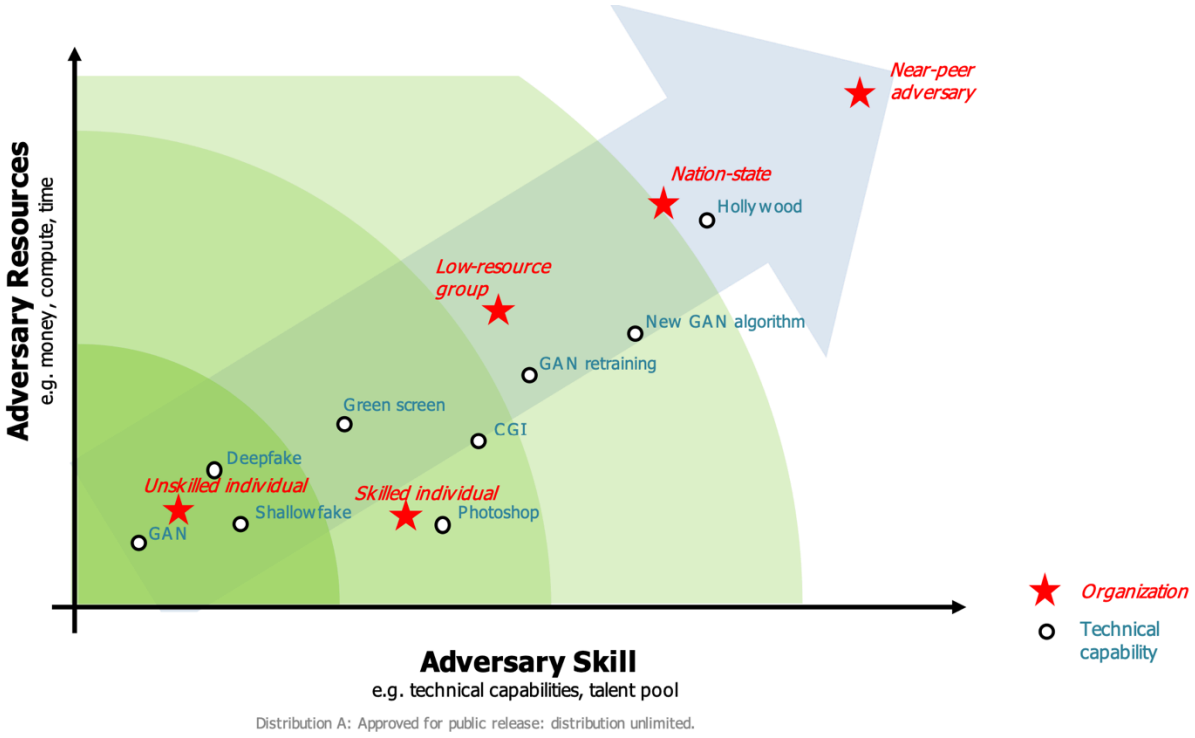
Harmful Applications

The accessibility of deepfake technology enables significant potential for abuse. Political manipulation through deepfakes threatens electoral integrity, as demonstrated by the 2024 New Hampshire robocall incident impersonating President Biden (Han, 2024). Financial fraud cases involving deepfakes increased dramatically - Belgium saw a 2,950% rise in cases from 2022 to 2023 (Statista, 2024). Personal exploitation remains a primary concern, with 99% of malicious deepfakes targeting women through non-consensual pornography (KTimes, 2024). Identity theft

has evolved to include voice cloning for financial scams, while social media platforms struggle to combat the spread of synthetic media used for disinformation campaigns (Bond, 2024).

Figure 3

Adversarial Landscape



[Figure 3: Evolution of synthetic media and visual manipulation technologies mapped against required skill level and resources. Note how deepfakes have moved from high to low barriers of entry. Source: Brooks et al. (2019)]

This visualization demonstrates how Generative Adversarial Networks (GANs) have transformed deepfake creation from a highly technical process to one requiring minimal skills

and resources. The democratization of this technology explains the rapid increase in both commercial applications and harmful uses (Brooks et al., 2019).

The technology's dual-use nature presents significant challenges. While companies like Synthesia create legitimate business applications, their technology faces potential misuse. In 2023, their avatars appeared in pro-China misinformation campaigns despite company safeguards (Heikkilä, 2024).

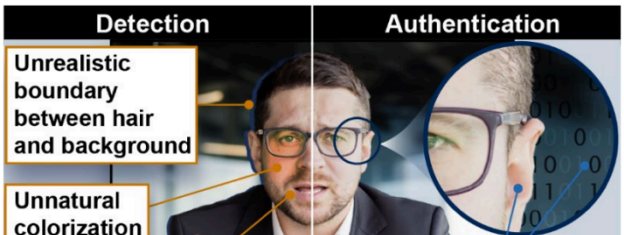
Challenges in Detection and Authentication

Current detection methods confront rapidly evolving deepfake technology. Technical analysis focuses on three key areas: facial movement patterns, audio inconsistencies, and metadata examination. Reality Defender and similar companies now offer real-time detection tools for video calls, yet these solutions struggle to keep pace with advancing AI capabilities (Rogers, 2024).

Platform responses include automated content flags, user reporting systems, and human content moderation. Meta recently implemented detection systems for AI-generated content, requiring clear labels for synthetic media (Bond, 2024). However, these methods face significant limitations. Deepfake technology improves faster than detection tools, leading to high false positive rates. The verification process demands extensive computational resources, making real-time screening challenging (Paris & Donovan, 2019).

Figure 4

Examples of Deepfake Detection and Authentication



Emerging solutions offer promising developments in three areas. Digital watermarking leads platform innovation, with Meta implementing invisible markers to track content origin and enable automated detection (Bond, 2024). Blockchain authentication creates permanent records of original content, verifying integrity and tracking modifications throughout the content lifecycle (Bothwell, 2024). AI-based detection employs sophisticated machine learning models to identify synthetic content patterns, offering real-time screening capabilities with continuous improvement through training (Byman et al., 2023).

The technological arms race between deepfake creation and detection shapes the future of digital media authenticity. As deepfake tools become more sophisticated, detection methods must evolve continuously. The challenge extends beyond technical solutions - media organizations need comprehensive strategies combining automated detection, human oversight, and clear verification protocols.

Social Impacts of Deepfakes

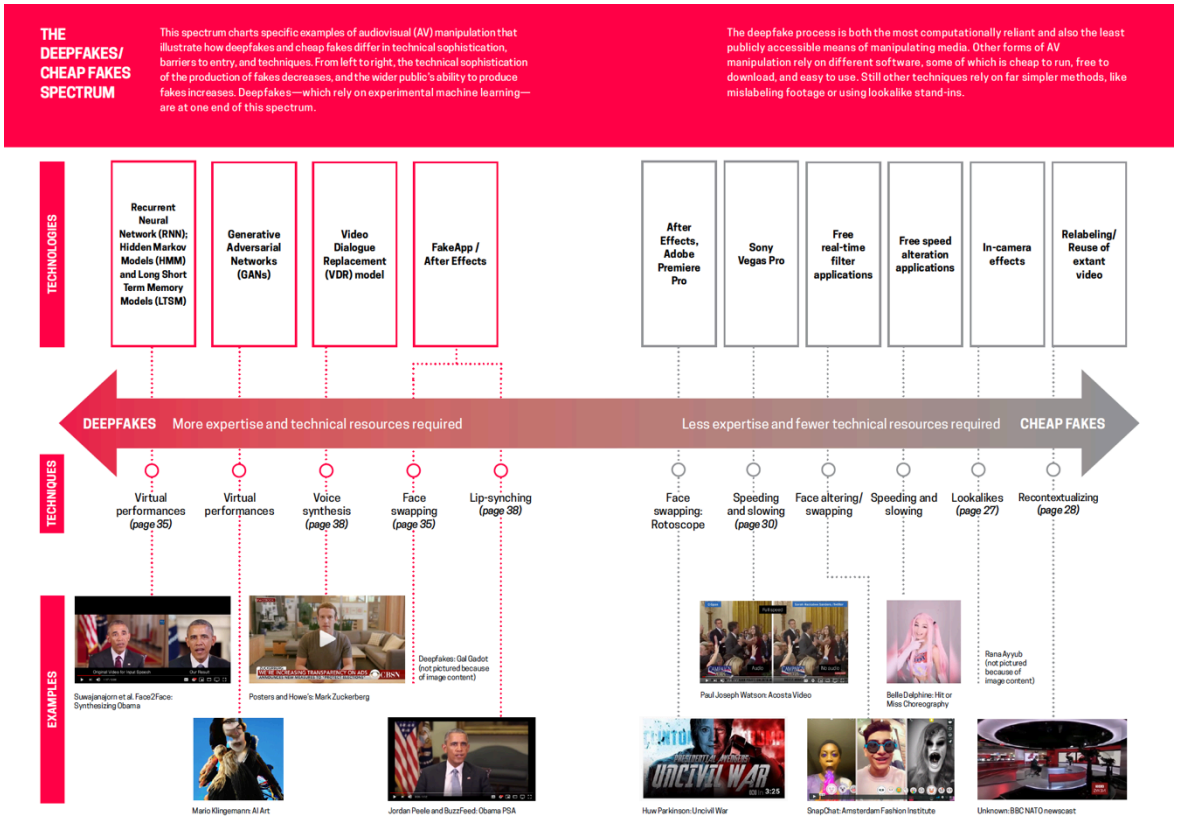
Political Manipulation and Misinformation

The January 2024 New Hampshire primary incident exemplifies how deepfakes threaten electoral integrity. AI-generated robocalls impersonating President Biden attempted to suppress voter turnout by providing false voting instructions (Han, 2024). This event marks a significant escalation in the sophistication of election interference tactics.

Disinformation campaigns increasingly leverage deepfake technology across multiple platforms. In Slovakia's 2023 election, manipulated audio of a candidate discussing vote rigging circulated days before the election, potentially influencing the outcome in favor of a pro-Russian candidate (Heikkilä, 2024). These incidents, highlighted in the spectrum below (Figure 5), demonstrate how rapidly deepfake content can spread and impact voter behavior.

Figure 5

The Deepfakes / Cheap Fakes Spectrum



The implications extend beyond individual elections. A Meta Oversight Board review revealed significant loopholes in the platform's "manipulated media" policies, highlighting the

challenge of maintaining public trust in digital information (Bond, 2024). Democratic institutions face unprecedented challenges in preserving information integrity during electoral processes.

Cultural and Social Influence

Entertainment industry adoption of deepfake technology creates complex social dynamics, presenting both opportunities and ethical challenges. Val Kilmer, who lost his voice to throat cancer in 2015, regained his ability to "speak" through Sonantic's deepfake technology. Similarly, Snoop Dogg's commercial appearances were efficiently modified using deepfakes to create localized versions without requiring reshoots (Lalla et al., 2022). However, these advancements raise concerns about actors' rights and industry equity. The technology could potentially create a "caste of deepfake actors who never achieve celebrity status or the ability to monetize their name and likeness" while established celebrities' likenesses become increasingly valuable digital assets (Lalla et al., 2022). The use of AI to recreate Anthony Bourdain's voice in a documentary sparked ethical debates about posthumous consent and authenticity in media (Rosner, 2021). While such applications demonstrate creative possibilities, they also raise questions about content authenticity and viewer trust.

The broader impact on social trust manifests in multiple ways:

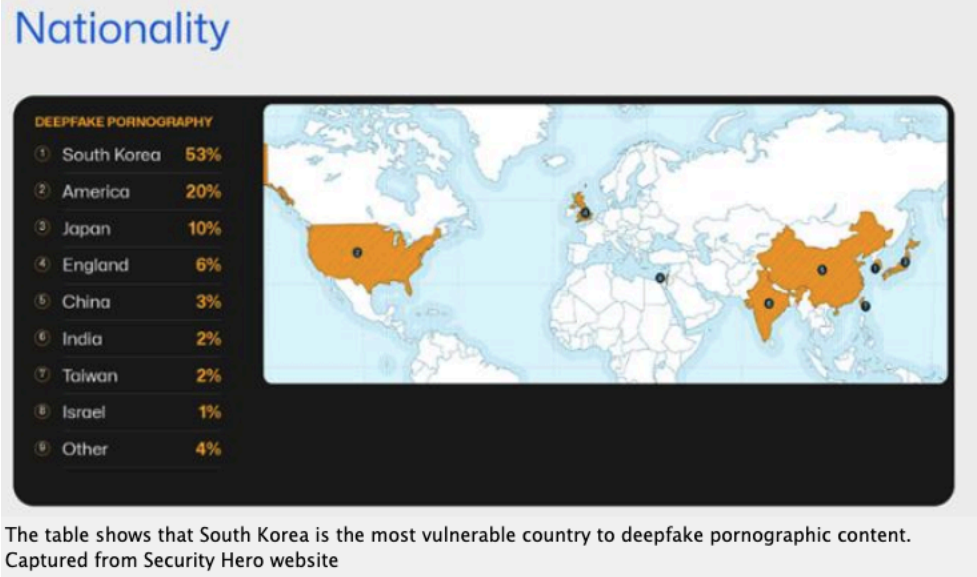
- News organizations struggle to maintain credibility
- Public skepticism of video evidence increases
- Historical documentation faces new challenges
- Entertainment and reality blur increasingly

Privacy Violations and Non-consensual Deepfakes

South Korea's experience highlights the devastating impact of non-consensual deepfakes. Eight out of ten most frequently targeted individuals worldwide are Korean singers, demonstrating how deepfake technology disproportionately affects public figures in certain regions (KTimes, 2024).

Figure 6

DeepFake Targeting by Nationality



The table shows that South Korea is the most vulnerable country to deepfake pornographic content. Captured from Security Hero website

The personal impact extends beyond public figures. Recent analysis reveals:

- 99% of deepfake adult videos target women
- Deepfake production increased 464% from 2022 to 2023
- Victims face significant challenges in content removal (KTimes, 2024)

Individual cases demonstrate the profound personal toll. Noelle Martin, a law graduate and activist, discovered her images manipulated into pornographic content at age 17, affecting her employment prospects and mental health (Paris & Donovan, 2019).

Deepfakes in Geopolitical Conflicts

Nation-state actors increasingly deploy deepfakes for strategic advantage. In March 2022, a manipulated video of Ukrainian President Zelenskyy calling for surrender circulated during Russia's invasion, marking the first high-profile use of deepfakes in armed conflict (Byman et al., 2023).

The military implications include:

1. Falsified command orders
2. Disrupted communication chains
3. Compromised intelligence gathering
4. Psychological warfare tactics

The Northwestern Security and AI Lab's research demonstrates how deepfakes could potentially destabilize military operations through fabricated communications (Northwestern University, 2023). This technology transforms modern information warfare capabilities.

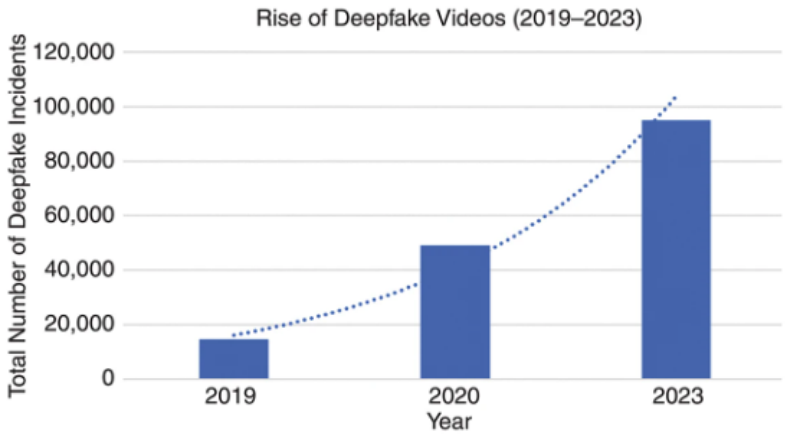
Threats to Democratic Societies

Deepfakes pose fundamental challenges to democratic discourse. The technology enables electoral manipulation and institutional trust erosion. Electoral manipulation through deepfakes manifests in multiple forms that threaten democratic processes. The 2024 New Hampshire

primary incident previously referenced, demonstrated sophisticated voter suppression tactics, instructing Democrats not to vote. This incident exemplifies how deepfakes enable both candidate impersonation and targeted voter suppression (Han, 2024). As show in Figure 7, research indicates a 550% increase in deepfakes online since 2019, highlighting the technology's growing sophistication and widespread deployment in disinformation campaigns (Kharvi, 2024).

Figure 7

Deepfake Incidents (2019-2023)



Note: The number of deepfake incidents is expected to increase in the coming years.

False campaign messaging extends beyond direct impersonation - manipulators create synthetic videos showing candidates in compromising situations or making inflammatory statements they never made. These tactics significantly degrade institutional trust. Meta's recent investigation revealed how synthetic content undermines public confidence in legitimate news sources and official communications. When voters cannot distinguish between real and fake content, they increasingly question all media, including authentic campaign messages and

official government communications (Bond, 2024). This erosion of trust creates a "liar's dividend," where bad actors can dismiss authentic evidence of misconduct as fake, further degrading public discourse and democratic accountability (Rosner, 2021).

Mitigation strategies must address multiple fronts:

Technical Solutions

- Advanced detection tools
- Content authentication systems
- Platform monitoring capabilities (Bothwell, 2024)

Policy Responses

- Updated election laws
- Platform regulations
- International cooperation (Byman et al., 2023)

The European Union leads in regulatory response through the AI Act, providing a potential model for other democratic societies (Heikkilä, 2024). However, technological advancement continues to outpace policy development.

These threats to democratic society require coordinated responses from:

- Government agencies
- Technology companies
- Media organizations

- Civil society groups

Ethical Considerations

Moral Responsibility of Creators and Distributors

III. Ethical Considerations (3 pgs)

A. **Moral Responsibility of Creators and Distributors**

- The ethical dilemmas in creating and distributing deepfakes.
- The role of media platforms in curbing harmful deepfake content.
- Example and explanation (Statista Chart)

B. **Challenges to Trust and Truth**

- How deepfakes erode the public's trust in information, media, and even government institutions.
- Ethical implications for news organizations, social media platforms, and society at large.

C. **Ethics of Detection and Prevention**

- Balancing individual privacy and freedom of expression with societal harm prevention.
- Ethical challenges in developing and deploying detection technologies.

IV. Legal and Regulatory Challenges (4 pgs)

A. **Existing Legal Frameworks**

- Review of laws addressing deepfakes in various countries (U.S., EU, etc.).
- Legal gaps in prosecuting deepfake creators and distributors.
- Example and explanation (Statista Chart)

B. **Policy Recommendations**

- Suggestions for developing new legal frameworks to address deepfake creation and dissemination.
- International cooperation for establishing global standards and regulations.

C. **Public Awareness and Media Literacy**

- The role of education in helping the public recognize and understand deepfakes.
- Strategies for fostering digital literacy and critical thinking in content consumption.

D. **Codes of Conduct and Government Responsibility**

- Proposals for government action to manage the use of deepfakes in national and international contexts.
- Case studies of governmental responses to deepfake-related crises.

5. Conclusion (3-4 pages)

- **Summary** of the key technological, social, and ethical challenges presented by deepfakes. Include your own perspectives.
- **Analysis and Opinion**: Discuss the broader implications of deepfakes on society and media integrity.
    - Weighing the benefits and risks of deepfake technology.
    - Possible future developments in deepfake creation and detection.
- **Recommendations for the Future**:
    - How advancements in detection technologies can address these challenges.
    - Strengthening legal frameworks and promoting corporate and societal responsibility.
    - Importance of ongoing public education and awareness campaigns.

References

Al-Khazraji, S. H., Saleh, H. H., Khalid, A. I., & Mishkhal, I. A. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23, 429-441.

Bond, S. (2024, February 6). Meta will start labeling AI-generated images on Instagram and Facebook. NPR. https://www.npr.org/2024/02/06/1228844753/meta-ai-generated-images-instagram-facebook

Bond, S. (2024, February 8). AI fakes raise election risks as lawmakers and tech companies scramble to catch up. NPR. https://www.npr.org/2024/02/08/1229641751/ai-deepfakes-election-risks-lawmakers-tec

-companies-artificial-intelligence

Bond, S. (2023, April 27). AI-generated deepfakes are moving fast. Policymakers can't keep up. NPR. https://www.npr.org/2023/04/27/1172089560/ai-generated-deepfakes-are-moving-fast-policymakers-cant-keep-up

Bothwell, B. (2024, March 11). Science & Tech Spotlight: Combating deepfakes. The U.S. Government Accountability Office (GAO). https://www.gao.gov/products/gao-24-107292

Brooks, T., G., P., Heatley, J., J., J., Kim, S., M., S., Parks, S., Reardon, M., Rohrbacher, H., Sahin, B., S., S., S., J., T., O., & V., R. (2019). Increasing Threat of Deepfake Identities. US Department of Homeland Security. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

Byman, D. L., Gao, C., Meserole, C., & Subrahmanian, V. S. (2023, January). Deepfakes and international conflict. Brookings. https://www.brookings.edu/articles/deepfakes-and-international-conflict/

Han, J. (2024, January 22). New Hampshire is investigating a robocall that was made to sound like Biden. NPR. https://www.npr.org/2024/01/22/1225872838/new-hampshire-biden-robocall-primary

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. Cyberpsychology, behavior, and social networking, 24(3), 149-152.

Heikkilä, M. (2024, April 25). An AI startup made a hyperrealistic deepfake of me that's so good it's scary. MIT Technology Review. https://www.technologyreview.com/2024/04/25/1091772/new-generative-ai-avatar-deepfake-synthesia/

Heikkilä, M. (2024, September 24). What the US can learn from the role of AI in other elections. MIT Technology Review.https://www.technologyreview.com/2024/09/24/1104347/what-the-us-can-learn-from-the-role-of-ai-in-other-elections/

James, S. B. (2019, July 31). The business case for protecting the technology behind "deepfakes." S&P Global. https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/th-business-case-for-protecting-the-technology-behind-deepfakes-53069518

KTimes. (2024, September 4). Deepfake disaster rooted in teaching kids technology without ethics. The Korea Times. https://www.koreatimes.co.kr/www/nation/2024/10/113_381480.html

KTimes. (2024, October). South Korea most targeted by deepfake crimes worldwide. The Korea Times. https://www.koreatimes.co.kr/www/nation/2024/10/113_381479.html

Lalla, V., Mitrani, A., & Harned, Z. (2022, June). Artificial intelligence: deepfakes in the

entertainment industry. WIPO Magazine. https://www.wipo.int/wipo_magazine/en/2022/02/article_0003.html

Meskys, Edvinas, et al. (December 2, 2019). Regulating Deep Fakes: Legal and Ethical Considerations. Journal of Intellectual Property Law & Practice, 15(1), 24–31. https://ssrn.com/abstract=3497144

Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes: The manipulation of audio and visual evidence. Data & Society Research Institute. https://datasociety.net/library/deepfakes-and-cheap-fakes/

Rogers, R. (2024, October 16). Startup can identify deepfake video in real time. Wired. https://arstechnica.com/security/2024/10/startup-can-catch-identify-deepfake-video-in-realtime/

Rosner, H. (2021, July 17). The ethics of a deepfake Anthony Bourdain voice. The New Yorker. https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice

Satariano, A., & Mozur, P. (2023, February 7). The people onscreen are fake. The disinformation is real. The New York Times. https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html

Statista. (2024). The explosive growth of AI-powered fraud

[Infographic]. https://www.statista.com/chart/31901/countries-per-region-with-biggest-increases-in-deepfake-specific-fraud-cases/

Swenson, A. (2024, July 29). A parody ad shared by Elon Musk clones Kamala Harris' voice, raising concerns about AI in politics. AP News. https://apnews.com/article/ai-kamala-harris-elon-musk-election-2024-voice-clone-parody-abc857f7ad66cdadb458a0ea77d63c55

The Rise of Artificial Intelligence and Deepfakes. Northwestern: Buffett Institute for Global Affairs. (2023, July). https://buffett.northwestern.edu/documents/buffett-brief_the-rise-of-ai-and-deepfake-technology.pdf

Villasenor, J. (2022, March 9). Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth. Brookings. https://www.brookings.edu/articles/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/