Navigating the Unseen: Deepfakes in Indian Politics

SPRF.IN

01
24

| Neha Maria Benny

Issue Brief

TABLE OF CONTENTS

Cover Image credits: *Pexels / Markus Winkler*

If you have any suggestions, or would like to contribute, please write to us at contact@sprf.in

January 2024

ISSUE BRIEF

Navigating the Unseen: Deepfakes in Indian Politics

| Neha Maria Benny

WHAT IS DEEPFAKE TECHNOLOGY?

Deepfake technology uses a machine learning technique called generative adversarial networks (GAN) to modify or generate images and videos (Techslang, 2023). AI-driven software detects, learns and replicates subjects' movements and facial expressions from the source material, creating realistic deepfake content. Creators enhance authenticity by using a vast database of source images, which is why public figures like celebrities and politicians are more susceptible. One software generates the fake video, while the second detects forgery signs. The machine-learning models refine the deepfake until it becomes impossible for the other software to identify the manipulation. This process, called "unsupervised learning", makes it difficult for other software to detect deepfakes (Jain, 2023).

There are three primary categories of deepfake technology (Sawtell, 2023):

- Face-swapping deepfakes: This prevalent type involves substituting one person's face with another's in videos or images.
- Audio deepfakes: These replace an individual's voice in recordings with another person.
- Textual deepfakes: These generate text that appears to be authored by someone else.

Deepfakes can be used to create deceptive videos to spread misinformation, identity theft or financial fraud. They can also be used to develop misleading videos or audio recordings to access sensitive information or manipulate individuals into making payments. As this technology evolves, it poses increasing challenges. Detecting poorly generated deepfakes is currently feasible through human observation, noting nuances such as blinking or discrepancies in details (Groh, n.d.). However, distinguishing authentic videos from manipulated ones may become impossible with technological advances. Initiatives like the Deepfake Detection Challenge aim to develop AI-based countermeasures, but keeping pace with evolving technology remains crucial.

AI AND GLOBAL POLITICS

Beyond their use for creating morphed content for entertainment, deepfakes have also been employed to generate sexually explicit material. The technology also holds the potential for more nefarious applications, such as inciting political violence, sabotaging elections, disrupting diplomatic relations, and spreading misinformation (Ahmed, 2023). It can also be utilised to humiliate and blackmail individuals or attack organisations by presenting fabricated evidence against leaders and public figures.

In Argentina, the runoff election became a testing ground for AI in political campaigns, with both candidates and their supporters using the technology to manipulate images and videos. There was even “deepfake” video of a candidate discussing a market for human organs, blurring the line between entertainment and disinformation (Herrera & Nicas, 2023).

In the lead-up to Slovakia’s October 2023 election, viral deepfake audio recordings surfaced, allegedly featuring Michal Šimečka, leader of the pro-Western Progressive Slovakia party, discussing election rigging and a twofold increase in beer prices. The prevalence of AI in Argentina’s election highlights the technology’s potential impact on democratic processes globally, raising concerns about deception and confusion among voters (Meaker, 2023).

Internationally, legislative initiatives are being implemented to address the challenges posed by deepfake technology. The UK’s National Cyber Security Centre has recognised the dangers of deepfakes by listing the rise of AI and the evolving geopolitical landscape as significant risks to electoral processes. The UK’s Online Safety Act explicitly addresses the sharing of sexual deepfakes. In the United States, the 2020 Identifying Outputs of Generative Adversarial Networks (IOGAN) Act aims to establish metrics and standards for detecting deepfakes, complemented by state-level legislation. The proposed DEEPFAKES Accountability Act of 2023 criminalises the failure to identify malicious deepfakes, particularly those related to foreign interference in elections and criminal incitement.

China, through the Cyberspace Administration of China (CAC) office, has introduced comprehensive provisions covering the entire deepfake life cycle, including creation, communication, and consumption. Legislation mandates the disclosure of deepfake technology use in the media. The European Union mandates tech companies like Google, Meta, and X to counter deepfakes under guidelines, with the Digital Services Act and proposed EU AI Act addressing platform monitoring.

AI IN INDIAN POLITICS

India first faced the risk of AI intervention in elections during the Delhi Assembly polls in 2020. Users came across videos featuring the then state BJP chief, Manoj Tiwari, criticising Chief Minister Arvind Kejriwal’s policies in various languages. Later, it was ascertained to be a deepfake video.

The recent assembly elections witnessed the proliferation of viral political videos, subsequently revealed as deepfakes. Amitabh Bachchan taking digs at Madhya Pradesh Chief Minister Shivraj Singh Chouhan on Kaun Banega Crorepati (Balkrishna & Triwedi, 2023), or Telangana minister Malla Reddy making false claims that voters would not get jobs under K Chandrashekar Rao’s re-election were such instances (Valaboju, 2023). The BRS has filed a complaint against the Congress in Telangana to the Election Commission (EC), accusing them of employing ‘deepfake’ technology (PTI, 2023). These underscore the issue of online authenticity and the tangible threat deepfakes pose to democratic processes, influencing public opinion.

THE POTENTIAL IMPACT OF UNREGULATED DEEP FAKES

Social media platforms in India are already inundated with deepfake content, making the country the sixth most vulnerable to harmful, explicit deepfake content (Balobanov, 2023). Political campaigns are increasingly exploring AI solutions, such as The Indian Deepfaker, offering personalised political ad campaigns that utilise lip-synching, voice cloning, and AI technology to connect with voters on a more personalised level (Saha & Tiwari, 2023).

Even legitimate domestic political entities may resort to disinformation to achieve their goals. Two main categories of deepfake videos may be utilised in election campaigns: those generating positive sentiments for endorsed candidates and those spreading misinformation about opponents (Sherif, 2023). These deepfake videos can then be disseminated within the voter base of the concerned constituency through WhatsApp groups. Misinformation videos are primarily designed for closed messaging platforms that play a key role in elections, like WhatsApp and Telegram, with different content moderation guardrails compared to platforms like Instagram and Facebook (Goel, 2018). Given their unprecedented realism, speed, scale, and capacity to personalise disinformation (Diakopoulos & Johnson, 2021), deepfakes exacerbate the broader issue of fake news on social media. Groups specifically formed for propagating political content, including deepfake-enabled misinformation, are referred to as “scratch groups’’ within electoral circles. These scratch groups are categorised based on the susceptibility of their members, with the age group between 18 and 25 being the most preferred. The intent is to circulate the deepfake videos in numerous WhatsApp groups, making their removal from public-facing social media platforms only a secondary concern (Sherif, 2023).

As AI progresses, the ability to discern fabricated audio and video content featuring individuals uttering statements they never made or engaging in actions they never performed is nearing a point of virtual impossibility. Consequently, the notion of “seeing is believing” will erode, compelling individuals to make judgments without reliable evidence about the authenticity of depicted events or statements. Even if they do not deceive individuals, they may sow uncertainty which, in turn, reduces trust in news on social media. It may also make them less likely to behave collaboratively and responsibly towards other users when they share news themselves (Galston, 2020).

Political candidates may exploit this technological advancement by dismissing genuine yet inconvenient representations as fake, known as the “liar’s dividend.” For instance, leaked recordings of Palanivel Thiagarajan, a notable Indian official, in April 2023 stirred controversy as they depicted him criticising fellow party members. Although Thiagarajan claimed a machine generated the audio, independent researchers verified the authenticity of at least one recording (Christopher, 2023). This scepticism towards authenticity can make meaningful debate difficult as citizens struggle to reconcile their tendency to believe visual content with the need to be vigilant against manipulative deepfakes. Voters are then likely to remain entrenched within their partisan echo chambers, trusting only those politicians and media sources aligned with their political beliefs, challenging evidence-based persuasion across partisan and ideological divides (Vaccari & Chadwick, 2020).

In the global context, deepfakes can also be used to legitimise wars (Galston, 2020), falsify orders (Allyn, 2022), sow confusion, divide military ranks, and undermine popular support. These applications, limited only by the creativity of those creating deepfakes, pose significant risks to international relations and the stability of nations. The potential for deepfakes in conflict scenarios may amplify the effectiveness of disinformation efforts, leading to greater confusion, distrust, and manipulation.

AI REGULATION IN INDIA

India currently lacks specific legislation addressing deepfakes and AI-related crimes. Some of the existing provisions in various laws can provide civil and criminal remedies to an extent (Shankar, 2023). The Information Technology Act, 2000, Section 66E applies to deepfake crimes involving the capture, publication, or transmission of a person's images in mass media, violating privacy and carrying a penalty of up to three years imprisonment or a fine of ₹2 lakhs. Section 66D of the same Act punishes individuals using communication devices or computer resources with malicious intent, leading to impersonation or cheating, with a penalty of up to three years imprisonment and/or a fine of ₹1 lakh.

Sections 67, 67A, and 67B of the IT Act can also be used to prosecute individuals for publishing or transmitting obscene or sexually explicit deepfakes. The IT Rules prohibit hosting content that impersonates another person, requiring prompt takedowns of artificially morphed images on social media platforms; failure to comply risks losing 'safe harbour' protection.

The Indian Penal Code, 1860, includes Sections 509 (insulting the modesty of a woman), 499 (criminal defamation), and 153 (a) and (b) (spreading hate on communal lines) for addressing cyber crimes related to deepfakes. The Copyright Act of 1957 can be invoked for unauthorised use of copyrighted material in deepfakes under Section 51.

IDENTIFYING GAPS

The current IT Rules only address instances after the illegal content has been uploaded and harm has already occurred. The identification of deepfakes poses a significant challenge, as their high quality means that even forensic labs may take days to establish their authenticity, allowing false narratives to spread and cause damage. The existing laws concentrate on online takedowns through censorship or criminal prosecution. Under this framework, the burden of filing complaints rests entirely on victims, and the local institutional response may not be equipped to deal with it satisfactorily. There needs to be a shift toward preventive measures, such as empowering users to recognize morphed images. As technology progresses, vigilance and continuous development of countermeasures remain essential in combating the potentially damaging impact of manipulated media.

Legal experts suggest a need for comprehensive legislation designed for emerging technologies like AI. Current laws may not be adequately equipped to address the challenges posed by technologies such as deepfakes. This inadequacy necessitates a regulatory framework based on market studies, a comprehensive understanding of generative AI, and the different types of harm caused by it (Bhaumik, 2023).

RECOMMENDATIONS FOR POLICY AND REGULATION

India, facing a growing deepfake challenge, requires a robust regulatory and policy framework. Defining deepfakes is a crucial starting point, consolidating provisions from existing laws encompassing AI regulation, data protection, copyright, and disinformation action plans. Users engaging with technology for content creation, publishing, or communication should obtain consent, verify identities, report deepfakes, and provide watermark disclaimers. The policy must incorporate public awareness initiatives, fund research and development for detection technologies, and establish mechanisms for effective coordination among ministries and social media intermediaries.

Regulating deepfakes and manipulated media in the political sphere is crucial to ensuring an informed electorate and maintaining the integrity of the electoral process. Regulations should cover synthetic images, audio, and video, focusing on limiting the dissemination of fabricated events or statements, particularly in paid campaign ads (Cortés et al., 2023). Explicit exemptions for parody, news media, and other protected speech should be incorporated into new rules. Transparency measures, such as the labelling of manipulated content, are defensible in court; however, policymakers should also consider outright bans on deceptive audio and visual material, especially those misleading about voting details. Regulations should primarily target creators and disseminators of deceptive media, with potential considerations for platform regulations.

Enforcement poses a significant challenge, considering the anonymity and adaptability of malicious users. Balancing enforcement with safeguards against overreach is crucial, addressing potential human rights violations and upholding the right to privacy and personal data protection. A comprehensive policy approach is indispensable to navigating the deepfake landscape's complexities.

CONCLUSION

The rise of AI-generated content, like deepfakes, poses complex challenges for democracy and truthful information. The use of such technology in political campaigns, as seen in Argentina, highlights the pressing need for clear rules and awareness efforts. Beyond elections, the impact of manipulated media extends to conflicts, disinformation, and impediments to social unity. Finding the right balance between transparency and potential misuse is crucial as India navigates the regulation of deepfakes. The evolution of deepfake accessibility, from major political entities to local politicians, requires adaptable laws. The intersection of AI, politics, and media calls for proactive steps to uphold the core principles of democracy and informed public engagement.

BIBLIOGRAPHY

Ahmed, N. (2023, November 9). Why has the government issued a directive on deepfake? | Explained. The Hindu. https://www.thehindu.com/sci-tech/technology/why-has-the-government-issued-a-directive-on-deepfake-explained/article67516589.ece

Allyn, B. (2022, March 16). A deepfake video showing Volodymyr Zelenskyy surrendering worries experts. NPR. https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Balkrishna, & Triwedi, J. (2023, November 15). How 'KBC' became a template for belting out non-stop election spins. India Today. https://www.indiatoday.in/elections/story/kbc-fake-video-amitabh-bachchan-election-madhya-pradesh-2463407-2023-11-15

Balobanov, K. (2023). 2023 State of deepfakes: Realities, threats, and impact. Home Security Heroes. https://www.homesecurityheroes.com/state-of-deepfakes/#key-findings

Bhaumik, A. (2023, December 4). Regulating deepfakes and generative AI in India. The Hindu. https://www.thehindu.com/news/national/regulating-deepfakes-generative-ai-in-india-explained/article67591640.ece

Christopher, N. (2023, July 5). Indian politician blames AI for alleged leaked audio. Rest of World. https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/

Cortés, E., Norden, L., Frase, H., & Hoffmann, M. (2023, December 5). Regulating AI deepfakes and synthetic media in the political arena. Brennan Center for Justice. https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena

Galston, W. (2020, January 8). Is seeing still believing? The deepfake challenge to truth in politics | Brookings. Brookings Institution. https://www.brookings.edu/articles/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/

Goel, V. (2018, May 14). In India, Facebook's WhatsApp plays central role in elections (Published 2018). The New York Times. https://www.nytimes.com/2018/05/14/technology/whatsapp-india-elections.html

Groh, M. (n.d.). Overview ‹ Detect DeepFakes: How to counteract misinformation created by AI — MIT Media Lab. MIT Media Lab. https://www.media.mit.edu/projects/detect-fakes/overview/

Herrera, L. C., & Nicas, J. (2023, November 16). Is Argentina the first A.I. election? The New York Times. https://www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html

Jain, S. (2023, December 4). Supervised and unsupervised learning. GeeksforGeeks. https://www.geeksforgeeks.org/supervised-unsupervised-learning/

Meaker, M. (2023, October 3). Slovakia's election deepfakes Show AI is a danger to democracy. Wired UK. https://www.wired.co.uk/article/slovakia-election-deepfakes

PTI. (2023, November 30). BRS complains to EC on Congress' alleged use of 'deepfake' technology in Telangana poll campaign | India News - Times of India. The Times of India. https://timesofindia.indiatimes.com/india/brs-complains-to-ec-on-congress-alleged-use-of-deepfake-technology-in-telangana-poll-campaign/articleshow/105621669.cms?from=mdr

Saha, B., & Tiwari, S. (2023, November 17). How deepfakes could impact Indian elections. India Today. https://www.indiatoday.in/elections/story/how-deepfakes-could-impact-indian-elections-2464241-2023-11-17

Sawtell, J. (2023, January 23). What is a deepfake? (definition, how to spot one). Built In. Retrieved January 4, 2024, from https://builtin.com/machine-learning/deepfake

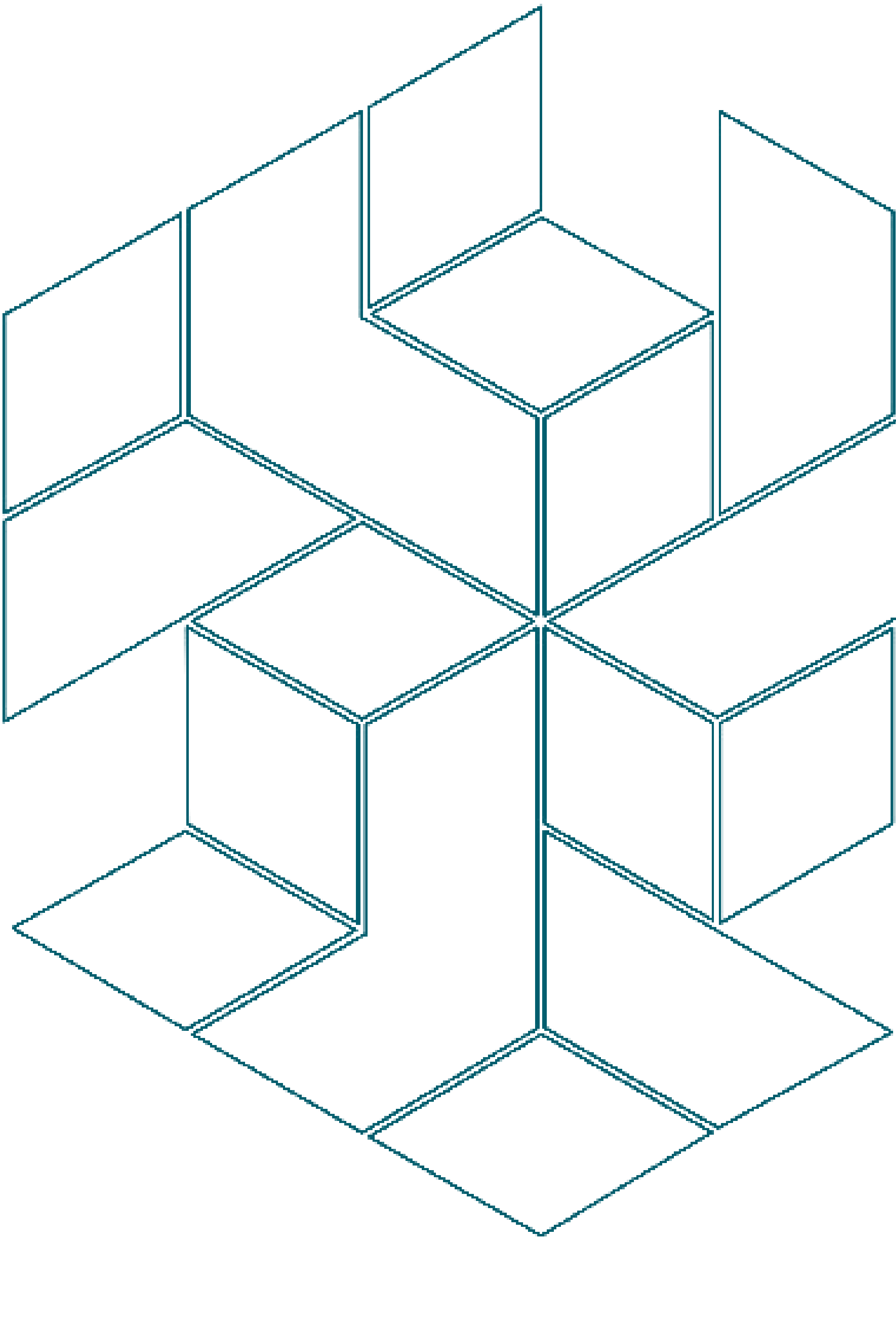
Shankar, V. (2023, July 16). Deepfakes call for stronger laws - The Hindu BusinessLine. The Hindu Business Line. https://www.thehindubusinessline.com/business-laws/deepfakes-call-for-stronger-laws/article67077019.ece

Sherif, A. H. (2023, November 23). Deepfake elections: How Indian politicians are using AI-manipulated media to malign opponents. Outlook Business. https://business.outlookindia.com/technology/deepfake-elections-how-indian-politicians-are-using-ai-manipulated-media-to-malign-opponents

Techslang. (2023, September 22). Deepfake technology: What is it and how does it work? — Techslang. Techslang. https://www.techslang.com/what-is-deepfake-technology/

Valaboju, G. (2023, November 5). AI, deepfake videos wreak havoc on political campaigns. Deccan Chronicle. https://www.deccanchronicle.com/nation/current-affairs/061123/ai-deepfake-videos-wreak-havoc-on-political-campaigns.html

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1). https://doi.org/10.1177/2056305120903408



SPRF.IN