

Article

Not peer-reviewed version

The Socio-Political Implications of Deepfakes in Developing Countries

[Lailoma Noor](#)*, Itba Malahat*, Haya Noor*

Posted Date: 23 September 2024

doi: 10.20944/preprints202409.1654.v1

Keywords: deepfakes; artificial intelligence; misinformation; developing countries; political stability



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

The Socio-Political Implications of Deepfakes in Developing Countries

Lailoma Noor *, Itba Malahat * and Haya Noor *

Faculty of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan

* Correspondence: u2021266@giki.edu.pk (L.N.); u2021250@giki.edu.pk (I.M.); u2021225@giki.edu.pk (H.N.)

Abstract: Highly realistic media created through Artificial Intelligence and Deep Learning, commonly known as deepfakes, presents a serious risk to political stability and the integrity of information. We are entering an era where images and videos, which were once considered reliable evidence, are now being altered to spread false information and incite unrest. This threat is especially heightened in developing nations, where misinformation and manipulation of public sentiment could lead to the collapse of the fragile democracies. This study delves deep into the increased vulnerability of these regions to deepfakes due to slower technological advancement and widespread lack of awareness. This study provides a detailed analysis and sets the stage for future investigations, proposing a combined approach to counteract the consequences of deepfakes. This includes combining technical methods for detection of deepfakes with public awareness campaigns and legal frameworks. Reducing the harmful effects of deepfakes requires cooperation among a diverse group of experts.

Keywords: deepfakes; artificial intelligence; misinformation; developing countries; political stability

1. Introduction

The term Deepfake was originally used to describe manipulated media in which one person had been swapped with another. Since then, the scope has expanded to include any sort of realistic media content generated by Artificial Intelligence. Deepfake technology can be used to alter specific parts of existing media, or generate entirely new content. The most common forms of deepfakes are face swaps, lip syncing, voice cloning, and full-body deepfakes.

In recent years, deepfake technology has advanced rapidly, causing the content to appear indistinguishable from authentic content. Various advanced technologies can be used to generate deepfakes, such as:

- Generative Adversarial Networks (GANs): GANs consist of a generator and a discriminator. The generator creates content and the discriminator determines if the content received is authentic. The discriminator's feedback is used to improve the content created by the generator [1].
- Diffusion Models: Diffusion Models iteratively add noise to data and then reverse the process. They are often useful in producing high-quality images [2].
- Autoencoders: Autoencoders consist of two parts - an encoder that compresses the incoming data, and the decoder that restores the input data. It is particularly useful in learning fundamental features and traits from an image while largely ignoring extra features [3].
- Recurrent Neural Networks (RNNs): RNNs are similar to Artificial Neural Networks, except they also have additional layers that allow them to process sequences of data. They are most often used in voice cloning [4].

2. Literature Review

The phenomenon of deepfakes has garnered significant attention in recent years due to its potential to disrupt various societal structures. Chesney and Citron [5] discuss the geopolitical implications of deepfakes, highlighting their potential to undermine democratic processes. Korshunov and Marcel [4] focus on the threat deepfakes pose to face recognition systems and propose detection methods. Their work emphasizes the importance of developing robust detection methods to maintain the integrity of biometric systems.

Nguyen et al. [1] provide a comprehensive survey of deepfake detection methods, categorizing them into three main approaches: visual artifacts detection, physiological signal analysis, and deep learning-based detection. Their survey underscores the need for continuous research to keep pace with evolving deepfake techniques.

Ajder et al. [6] examine the landscape of deepfakes, identifying emerging threats and evaluating the effectiveness of existing detection tools. Their report highlights the gaps in current detection technologies and calls for more advanced solutions to address the growing threat.

The socio-political implications of deepfakes are profound, particularly in the context of misinformation and public trust. Chesney and Citron [5] explore the geopolitical ramifications of deepfakes, arguing that they represent a new front in the disinformation war. Their work highlights how deepfakes can undermine democratic processes by spreading false information and creating political instability.

Westerlund [7] provides a review of the societal impacts of deepfake technology, noting its potential to disrupt various sectors, including politics, entertainment, and media. He argues that the emergence of deepfakes necessitates new regulatory frameworks and ethical guidelines to mitigate their negative effects.

In the context of developing countries, the vulnerabilities are even more pronounced. Ajakaiye [8] discusses how deepfakes have been used in the Nigerian election to manipulate public opinion and influence the outcome. This example underscores the urgent need for awareness and countermeasures in regions with less media literacy and technological infrastructure.

Byman et al. [8] explore the international security implications of deepfakes, highlighting their potential to escalate conflicts and undermine international relations. Their research calls for coordinated international efforts to address the threat posed by deepfakes.

While there is substantial research on the technological and detection aspects of deepfakes, there is a relative paucity of studies focusing on the socio-political impacts in developing countries. Existing literature often emphasizes the technological advancements and detection methods but overlooks the unique vulnerabilities of these regions. This paper aims to fill this gap by providing a detailed analysis of the socio-political consequences of deepfakes in developing countries and proposing comprehensive mitigation strategies.

3. A Distinct Threat in Vulnerable Contexts

While media manipulation is not new, traditional methods lacked the realism that deepfakes now provide. Its easy access to the public significantly increases societal uncertainty and distrust. Developing countries, with lower levels of media literacy and awareness, are particularly at risk. Malicious actors can use this technology to exploit weak trust in institutions and manipulate public opinion to influence elections and incite unrest [6].

Developing countries, due to fragile democracies and comparatively weaker institutions, are especially vulnerable to the disruptive potential of deepfakes. The absence of strong legal frameworks and technological infrastructure further amplifies this vulnerability. Thus, the potential for deepfakes to destabilize political systems and create turmoil is an urgent issue that requires immediate action [5].

4. The Corrosive Effect of Deepfakes

The latest surge in deepfakes has caused a society riddled with doubt and mistrust. Their potential to manipulate and fabricate content material disrupts set up systems and establishments, consisting of the judiciary, medicine, academia, and government. This segment delves into these detrimental effects, with a particular awareness on the vulnerability faced by developing countries:

4.1. Disinformation War

Deepfakes are increasingly being weaponized to spread misinformation. By manipulating political speeches and fabricating news announcements, even regular people can use deepfakes to propagate their personal agendas with just a few clicks. For example, a deepfake video of former U.S. President Barack Obama making offensive comments raised worries about using deepfakes to smear political

figures [5]. In Burkina Faso, fake videos of various people supporting the military junta spread on social media, worsening trust in already scarce reliable information sources [2].

4.2. Identity Theft and Financial Fraud

The ease with which deepfakes facilitate scams is particularly alarming. In 2019, a fake audio of a CEO led to a \$243,000 loss [6]. While this didn't happen in a developing country, these nations could be more vulnerable to such scams due to less awareness about the dangers of deepfakes.

4.3. Reputational Assassination

Deepfakes can damage a person's reputation by showing them in embarrassing situations. Additionally, fake pornographic videos have been used to defame people, causing severe mental distress and lasting shame [?]. In developing countries, victims are left especially vulnerable due to limited legal action.

4.4. Conflicts and Social Tensions

Deepfakes can be used to escalate existing conflicts between different religions, races, or ethnicities. As an example, fake videos displaying violence among religious groups could incite real-world violence. This is especially concerning in developing nations with histories of ethnic or religious conflict.

4.5. Silencing Dissent

Political figures can use deepfakes as a weapon to silence dissent. In India, a pornographic video created through deepfake technology was used to tarnish the reputation of a journalist critical of the BJP and ultimately silence her [?]. This raised concern over how easily deepfakes can be used to discredit opponents and suppress disagreements.

4.6. Electoral Interference

Deepfakes can control the result of elections through circulation of fake media. In Nigeria, for instance, fake videos showing Hollywood actors and public figures endorsing a candidate had been circulated on social media to boost his image [?]. Such use of deepfakes undermines democracy and can result in candidates being elected based on false information.

4.7. International Relations

Deepfakes can affect the outcome of wars, cause conflicts between allies, or weaken nations facing threats. For example, a fake video showing Ukrainian President Zelenskyy surrendering to Russia was used to attempt to get the Ukrainian troops to surrender [8].

4.8. Undermining the Judiciary

Deepfakes pose a significant threat to the justice system. If used as proof, they could cause wrongful convictions. That is especially concerning for developing nations with outdated legal systems and limited forensic resources.

4.9. Cybersecurity Threats

Deepfakes may be used in sophisticated phishing scams. Videos or audio recordings impersonating government officials or bank representatives can trick people into revealing personal records or monetary information. Developing nations with less robust cybersecurity infrastructure are at a greater risk of these attacks.

4.10. Cultural Appropriation and Exploitation

Deepfakes can misappropriate or exploit cultural heritage. Motion pictures providing respected ancient figures endorsing commercial merchandise may be disrespectful and undermine cultural identification.

4.11. Disinformation Cascades and Echo Chambers

Deepfakes can be easily shared and amplified on social media, leading to echo chambers where users are only exposed to data that confirms their present ideals. This exacerbates social and political polarization, hindering constructive talk. Developing countries with restrained access to diverse media sources can be more prone to these.

These examples illustrate the large and corrosive effect of deepfakes, especially in developing nations. As technology evolves, so should our strategies to mitigate the threats deepfakes pose to trust, truth, and the very fabric of society.

5. Ethical and Legal Considerations

Deepfakes raise significant ethical and legal questions. Who is accountable for harmful deepfakes? Legal systems must evolve to create clear regulations and penalties. Ethical guidelines should ensure consent, transparency, and harm minimization [5].

Accountability and Legal Measures

- **Challenges in Identifying Perpetrators:** Deepfakes can be created and distributed anonymously, complicating accountability.
- **Evolving Legal Frameworks:** Existing laws may be insufficient; new legal frameworks are necessary. Current legal provisions related to defamation, privacy, and intellectual property may not adequately address the unique challenges posed by deepfakes.
- **International Collaboration:** Global standards and cooperation are essential. The transnational nature of the internet requires coordinated international efforts to combat the spread of deepfakes.

6. Mitigation Strategies

Robust strategies are essential for combating the misuse of deepfakes. This section outlines key approaches to tackle this threat:

6.1. Legislative Frameworks and Policy Development

Governments must create clear legal frameworks defining deepfakes and assign liability for their malicious use. Penalties should deter misuse while upholding free speech.

6.2. Empowering Social Media Platforms

Social media platforms must take responsibility. They should empower users to flag suspected deepfakes and utilize AI tools for detection. This collaborative approach can quickly identify and remove harmful content.

6.3. Accessible Media Verification Tools

Developing easy-to-use deepfake detection apps is crucial. Tools like Sensity AI's Deepware Scanner help individuals verify media in real-time. These apps should be widely available to promote responsible information consumption.

6.4. Public Awareness Campaigns

Educating the public is vital. National task forces can lead campaigns to teach people how to spot deepfakes. Promoting media literacy helps users distinguish genuine content from manipulated media, reducing the impact of deepfakes.

6.5. Collaboration Among Tech Companies

Tech companies must collaborate, despite competition. Initiatives like the Deepfake Detection Challenge show the power of collective action. Standardizing detection tools and sharing deepfake databases are key steps.

6.6. Ethical Guidelines for Deepfake Technology

Establishing ethical guidelines is crucial. Industry organizations should set standards for consent, purpose, and harm potential to ensure responsible use of deepfakes.

By implementing these strategies, we can build a robust defense against deepfakes. Collaboration across legislation, platforms, tools, education, technology, and ethics is essential to navigate the deepfake landscape and uphold truth in the digital age.

7. Future Directions and Research Frontiers in Deepfake Mitigation

Combating deepfakes requires ongoing research and innovation. Here are key areas to focus on:

7.1. Improving Detection

Enhance the accuracy and speed of detection algorithms. Research should refine existing methods and explore new ones to keep up with evolving deepfake tactics [1].

7.2. Interdisciplinary Collaboration

The rise in the misuse of deepfakes is a complex issue that requires attention from a diverse group of experts. Computer scientists, legal experts, government officials, and social scientists must all work together to form comprehensive solutions to combat deepfakes [7].

7.3. Blockchain for Content Authentication

Blockchain technology offers a promising solution for verifying media authenticity. Its immutability and decentralization make it ideal for establishing trust in digital content.

7.4. Counter-Deepfake Techniques

Research should also focus on proactive measures. Developing tools that not only detect deepfakes but also generate authentic counter-narratives can help combat misinformation effectively.

Exploring these research avenues will help manage the issue of deepfakes and restore trust in digital media. Collaboration between experts is crucial for ensuring a responsible and ethical future for deepfake technology.

8. Conclusions

Deepfakes present a significant threat to the stability and integrity of information in developing countries. As this technology evolves, so must our strategies to combat its misuse. Through a combination of legislative action, technological advancements, public awareness, and international cooperation, we can mitigate the harmful effects of deepfakes. By addressing the ethical and legal implications and promoting interdisciplinary collaboration, we can work towards a future where digital content remains trustworthy and reliable.

1. Nguyen, T.; Nguyen, T. Deep Learning for Deepfakes Creation and Detection: A Survey. *arXiv preprint arXiv:1909.11573* **2019**.
2. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* **2020**, *33*, 6840–6851.
3. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *science* **2006**, *313*, 504–507.
4. Korshunov, P.; Marcel, S. DeepFakes: a New Threat to Face Recognition? Assessment and Detection. *arXiv preprint arXiv:1812.08685* **2018**.
5. Chesney, R.; Citron, D. Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs* **2019**.
6. Ajder, H.; Patrini, G.; Cavalli, F.; Cullen, L. The State of Deepfakes: Landscape, Threats, and Impact. *Deeptrace* **2019**.

7. Westerlund, M. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review* **2019**, 9, 40–53.
8. Byman, D.L.; Gao, C.; Meserole, C.; Subrahmanian, V. Deepfakes and International Conflict **2023**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.