MEDIANAMA

Deep Fakes And Democracy

January 2024

On January 17, MediaNama conducted a discussion on Deep Fakes and Democracy. The discussion focused broadly on the **implications of deep fakes for the upcoming general elections, the technical and policy solutions to tackling deep fakes, positive use cases of deep fakes, and how deep fakes threaten user verification**.

Our objective was to identify:

- Is there a **legitimate use case for deep fakes in elections,** or should they be outlawed altogether?
- What are the capabilities of political parties to generate deep fakes? What are their dissemination networks like?
- What are the **challenges of attributing deep fakes or even fake news to a bad actor**?
- How can the **spread of deep fakes be curbed** on end-to-end encrypted platforms like WhatsApp?
- Is there a need for a consumer app-level deep fake detection?
- Has there been enough deep fake activity in the recent elections? What is the current environment of deep fake like?
- What are the challenges with detecting deep fakes once **something is uploaded on social media**?
- Is watermarking an effective solution to curbing deep fakes?
- How does safe harbor play out in case of deep fakes? Should platforms lose their immunity if deep fake content is posted by users on their service?

Our speakers for the discussion included:

- Rakesh Maheshwari (Former Sr. Director and Group Coordinator, MeitY)
- Saikat Datta (CEO and Co-founder of Deepstrat)
- Jency Jacob (Managing Editor, Boom Fact Check)
- Gautham Koorma (Researcher, UC Berkley School of Information)
- Tarunima Prabhakar (Co-founder of Tattle Civic Technologies)
- Shivam Shankar Singh (Data Analyst and Political Campaign Consultant)

> We saw participation from companies and organizations like Samsung, HDFC Bank, Info Edge, Ministry of Electronics and IT, The Quantum Hub, Apollo 247, COAI, Thompson Reuters, Ikigai Law, Access Now, Truecaller, SLFC, Outlook, Meta, ShareChat, NDTV, EGaming Federation, Chase India, The Caravan, the Hindu, CCG NLUD, DataLEADS, University of Exeter, Center for Civil Society, SFLC, Google, Spotify, InShorts, Deloitte, Internet Society, The Internet Freedom Foundation, Logically.ai, Mozilla Foundation, Times Internet Limited, News Click, Mogambay India, The Asia Group, LT Mindtree, IndusLaw, Times Internet, CCAOI, Hasgeek, Citizen Digital Foundation, Dvara Research Junglee Games, among others.

MediaNama hosted this discussion with support from Google and Meta.
The following document captures and summarizes the key points raised during the discussion. You can also view a recording of the discussion on our YouTube channel.

Recommendations on how to combat Deep Fakes

Speakers at MediaNama's discussion on Deep Fakes and Democracy agreed that there isn't one simple solution to curbing the spread of deep fakes. Here are some of the key recommendations they made—

- **Companies should make AI-generated content using their services detectable:** The discussion highlighted the need for the government to work with major artificial intelligence (AI) companies to ensure that content generated using their services is detectable up to a certain extent. In cases where certain AI companies aren't cooperating with the government, it could consider regulation.
- **Self-regulation by AI companies is a good place to start:** Given that a certain group of companies dominate the AI-content generation market, the point to track misuse could be narrowed down. The discussion revealed that it would be effective for AI companies to have policies in place that regulate the creation of certain kinds of content. While it was suggested that AI companies should bring in self-regulatory policies, it was pointed out that free image-generation tools should have a certain degree of regulation. Further, it was said that courts should have the final say on whether social media platforms should lose their immunity for failing to effectively control the spread of deep fakes.
- **Social media platforms could shadow-ban a piece of content while it is under investigation:** Discussants pointed out that within 4-5 hours of a piece of content being uploaded, 50% of users have already seen it. In such a situation, the timeline for taking down such content as prescribed under IT Rules— which is 72 hours— might not be effective. As such, it was suggested that during this 72-hour period when the content is under investigation, platforms could prevent the algorithmic amplification of such content by shadow-banning it.
- **Banning deep fake use during elections:** It was suggested that the Election Commission of India could start with a complete ban on the use of deep fakes by political parties during the upcoming general elections. At a later stage, the commission could set out certain conditions in which deep fakes could be used. For

instance, allowing parties to create deep fake content of its party leaders, provided that it is labeled as such.

- **Mandating digital fingerprinting of deep fakes:** It was suggested that digital fingerprinting (adding provenance information) of AI-generated content at the source of creation is necessary. AI companies must prevent users from creating content without digital fingerprints and social media platforms could consider penalizing content that does not have a fingerprint. While concern was raised about the potential to fake such digital fingerprints, discussants argued that the solution could be similar to the relatively spoof-free nature of file formats like ".jpg".
- **Social media platforms can takedown election-related deep fakes within 2 hours:** It was pointed out that social media platforms have signed a voluntary code of ethics that they observe during general elections. Under this code, platforms take down fake news content within two hours of getting a lawful order asking for such a takedown. Discussants argued that the same could be implemented in case of election-related deep fakes as well.
- **Increasing funding for deep fake detection:** Speakers said that the Silicon Valley has been rapidly working on coming up with deep learning techniques that have the right product market fit. They suggested that since these deep learning applications have now become well known, some of the development funds need to be dedicated to countermeasures for the harms they pose. This includes funding deep fake detection and provenance-based techniques.
- **Media literacy could help dampen the impact of deep fakes:** It was suggested that users should practice attention-conservation, and evaluate whether something is even worth their attention before they choose to consume a piece of content. Discussants also expressed that journalists should shoulder the responsibility of media literacy, instead of it being left solely to educational institutions.
- **The Election Commission could form a team to tackle deep fakes:** Speakers argued that the ECI needs a broad coalition of people— psychologists, sociologists, political science experts, and technologically capable individuals— to effectively tackle deep fakes.
- **AI companies shouldn't be able to monetize disinformation:** Discussants suggested that AI companies should alter their terms of use to ensure that they don't monetize disinformation and misinformation spread using deep fakes.

Executive Summary

Ahead of elections in multiple different jurisdictions, the proliferation of deep fakes and their potential to spread misinformation has emerged as a major concern. Speakers at MediaNama's discussion pointed out that the current measures to curb deep fakes like watermarking or content provenance have limitations, with bad actors being capable of removing watermarks and faking provenance. They highlighted that one of the reasons for the limited success of detection techniques is that once the adversaries learn about the detection techniques, they improve their generation capabilities.

Another challenge to detection is the circulation of deep fakes on social media platforms. It was pointed out that an audio deep fake detection algorithm, which may have 90 percent accuracy in lab settings, becomes less effective when asked to detect a deep fake circulating on social media because platforms transcode the audio and video, thereby changing its properties, making it difficult to compare with the original version.

Discussing the responsibility of social media platforms in curbing deep fakes, it was found that platforms have two main tools in their arsenal—content takedowns and shadow banning. Speakers called attention to the importance of shadow banning before any final decision is made about a piece of content, suggesting that this would prevent the algorithmic amplification of the deep fake. Discussants pointed out that while social media platforms are capable of curbing child sexual abuse material (CSAM) to a certain extent, curbing deep fakes becomes a challenge because they cannot create databases to cross-reference the deep fake, which is how they currently take down CSAM. When considering the time-sensitive nature of curbing misinformation, it was highlighted that platforms have adhered to a two-hour takedown timeline in previous elections as a part of their voluntary code of ethics that social media platforms have agreed to follow during general elections.

When considering the impact of deep fakes on the upcoming elections, discussants pointed out that this technology would make it easier for political parties to flood people's social media feeds with their desired narrative. Consequently, this flood of deep fakes would also flood any reporting system that the Election Commission of India (ECI) would create, overwhelming it and making it impossible for the ECI to act against any of the complaints it receives. It was also mentioned that the flood of misinformation created by deep fakes is similar to fake news that was spread in past elections, and will be spread through sources that cannot be tied to political parties, making it hard for the ECI to hold any party responsible. Discussants expressed that international actors could also make attempts to influence election results using deep fake-generated misinformation.

The discussion also covered policy solutions to deep fake misinformation, wherein it was pointed out that the spread of deep fakes generated outside the nation-state should be curbed by autonomous bodies like the ECI. Self-regulation was discussed as a measure against deep fakes with speakers arguing that it could only be applied on a case-by-case basis.

Discussants said that fact-checkers can struggle to call out deep fakes that are furthering a political agenda under the guise of satire. In certain cases, a deep fake intended as satire could be clipped out and spread as fake news as well. It was mentioned that previous attempts to debunk satire, which was being mistaken for real events, had landed fact-checkers in legal trouble. This makes fact checkers cautious of looking into content that has been marked as satire, even if said content is pushing a political narrative.

The discussion established that deep fakes would pose a risk to practices like video know-your-customer (KYC) verification. Speakers pointed out that there have been instances of job candidates who have hidden their identities in interviews using deep fakes. It was pointed out that if biometrics were used as a means of identification, they would have to constantly evolve to keep pace with advancements in deep fake technology.

Key Points from the Discussion:

I. Technical solutions to curbing deep fakes

It’s an arms race between detectors and adversaries:

Social media platforms and researchers are publishing papers on detection but it remains a hard problem because once the adversaries know of the detection algorithms, they make improvements upon their generation capabilities.

There is a need for AI companies to create a level of traceability for the content generated through them:

The government needs to work with platforms that provide AI content-generation services to establish a certain level of traceability from the source. Further, the government should regulate those content-generation platforms that it cannot work with.

All watermarks can be broken:

> ***“This is some work coming out of Soheil Feizi's lab at the University of Maryland. They have looked at every watermarking technique that exists out there, and they've broken every one of them. I think we should be using watermarking, but keeping in***

> ***mind its downside, that it can be easily broken by a sophisticated adversary. Sometimes it's not even a very sophisticated adversary."*** — *Gautham Koorma, UC Berkley*

Why hashing used for CSAM detection won't work for deepfakes:

> ***"The biggest problem with deepfakes is that [something new is] generated every time, and the content can change quite a bit, as opposed to the classical photographs that we have. These techniques are based on the idea that there's a database of hashes that we can cross-check with really quickly. And so it's really hard for a company to maintain a database of hashes of every possible content that can be generated."*** — *Gautham Koorma, UC Berkley*

Mandate digital fingerprinting to curb deep fakes:

Content provenance or fingerprinting of deep fakes at source would be a necessary step to curbing deep fakes. Companies must prevent users from removing the fingerprint and social media platforms should penalize content that doesn't have a fingerprint.

Content labels leave users responsible for verifying information:

Elon Musk had also talked about getting detailed labeling for content to maintain a public record of everything that happens around the content but this puts the onus on users to review content history and make up their mind.

Solving the CSAM problem is also going to be much harder now:

> ***"Before the advent of generative AI, CSAM was usually photographs, disturbing photographs of explicit material involving children that could be collected by law enforcement agencies. The scale at which they were being generated was limited. […] This is not true for deepfakes because now what we're seeing is the likeness of a child being taken and used to generate explicit material, which can be generated at scale and volumes that are unprecedented."*** — *Gautham Koorma, UC Berkley*

Detecting audio deepfakes has 90 percent accuracy in lab settings, but drops significantly when the content is "in the wild":

> ***"For example, when you upload an audio clip to Facebook or when you send it on WhatsApp, each of these platforms do something called transcoding, essentially changing the bit rate, changing some properties of the media. And once that happens, we see that the accuracy of detection drops a lot, sometimes higher than 10***

> ***percent. So, it's really hard to detect these things once they've gone on social media."*** *— Gautham Koorma, UC Berkley*

Even if a detection algorithm has 90 percent accuracy, it is not enough:

> ***"An algorithm that has 90% accuracy, you're saying that 10 out of 100 cases, you're getting it wrong. That is huge at the scale of social media, where you have billions of posts, versus if you're doing it for something in a court case, it's still a reasonable tolerance. So, when you're using these algorithms on social media for automated detection of deepfakes you want to get 99.999% accuracy."*** *— Gautham Koorma, UC Berkley*

Silicon Valley needs to dedicate funds to deep fake detection:

> ***"I think in the past five years, what we've seen in Silicon Valley and across the world is kind of rapid generation in deep learning techniques and them finding what we call product market fit. I think they have found product market fit. Now they're scaling and we know these applications will exist. So, we need to take a little bit of money out of that and start spending it more on the countermeasures for the harm, which is basically detection and start developing detection and provenance-based techniques. There's very limited money going into this space right now. And I'd like to see more of venture capital, government money, whatever sort going on to not just the generation part, but also the detection part."*** *— Gautham Koorma, UC Berkley*

II. How deep fakes could affect upcoming elections

The deep fake problem will be similar to fake news:

> ***"If you look at a lot of the fake news that was generated during past elections, a lot of it did not come from the political parties directly. Initially, it came through circuits that are not directly affiliated to a political party and not directly affiliated to a politician. These might be different Facebook pages that a party supports somehow, but you could never link them directly to a political party. The funding, the people, all of them are essentially de-linked. They might be paid through some business houses, they might be paid in cash. I see the deepfake problem in a similar light as the fake news problem, because it will be very difficult to attribute a deepfake to a political party very directly." -*** *Shivam Shankar Singh, campaign consultant*

Political parties already have extensive distribution networks for spreading deep fakes:

> ***"There are two sides to any kind of content. There is the content and the narrative that a political party is trying to push, and the second one is a distribution channel. What political parties have focused on for the longest time is creating a distribution channel, because you could have the best content in the world, but if you don't have a mechanism to deliver it to the voters, then it's pretty worthless. So, as of right now, the distribution channel exists, and political parties have the capacity to push any content that they want to millions and millions of voters across the country. This might be through their own Facebook pages, this might be through, it's something like Twitter and a group of influencers, and it's through a lot of the WhatsApp groups that have been created, and a lot of Telegram channels that have been created. So, there is capacity for it to wreak-havoc essentially as soon as the content comes out"-*** *Shivam Shankar Singh, campaign consultant*

Political parties are experimenting with both positive and negative deep fake use cases:

> ***"Say, if you talk about something like promises during an election campaign, if they want to put the name of a particular political constituency or a particular region or a particular state and dub it into that language, that capability, a lot of companies are pitching right now, and the political parties are experimenting with it themselves. The exact same tool and the exact same technology can be used for something nefarious, as in you could end up creating a deepfake of someone that you're not authorized to create a deepfake of. It might be an opposition leader. It might be someone that has not authorized it. It might be a celebrity endorsing a certain political candidate." -*** *Shivam Shankar Singh, campaign consultant*

Parties are currently reliant on existing AI models:

> ***"No one's making their own, say, GAN [generative adversarial network] or their own model or running any kind of, say, open-source technology on their own servers right now. That hasn't happened yet. Because the platforms really like, one, it's pretty cheap to just do it on the larger companies' trained models. The second one, I think OpenAI just mentioned day before yesterday that they wouldn't allow, say, GPTs to be used on creating deepfakes for election purposes specifically. So, things like that are happening now. So, you'll see political parties spend more resources on that side. Otherwise, generation's very cheap. Distribution's already there. All that you need is, okay, this is a message that we want to spread and we'll use deepfake technology to do it. Political will is what hasn't come about yet, but it most likely will during the 2024 elections."*** *- Shivam Shankar Singh, campaign consultant*

Dear Reader,

MediaNama’s work of covering the key policy themes that are shaping the future of the Indian Internet is made possible by support from its subscribers. If developments in technology policy are key to you or your organization, I would urge you to subscribe to MediaNama to support us.

Thanks,
Nikhil Pahwa,
Editor and Publisher

The quality of the deep fake doesn’t matter, political parties just want to flood people’s timelines:

Political parties have been capable of spreading misinformation even before deep fakes, by taking pictures from older events or from different countries to form their desired narratives. With AI they are capable of creating fake images on a larger scale. It does not matter how sophisticated these images appear, the sheer quantity of them pushes the targetted user in a certain direction.

Political parties can make it hard for the election commission to take down deep fakes:

> ***“So, what happened even during the fake news reporting stuff that they said during 2019, they had a mechanism in place to report it to say the platforms and get content taken down. But when a political party really wanted to spread a certain piece, they actually flooded the election commission stream itself. So, the election commission was so inundated with complaints and reports and this and that, that they themselves were able to forward very few of them to the platforms. That's why the platform to election commission link worked pretty well. That's not where the bottleneck was. The bottleneck was at the election commission itself. And how many people can the election commission really deploy for something like this? It could be thousands and thousands of pieces of content generated every day with something like deep fake technology itself. Generative AI, the entire point is that the cost of generation is like pretty negligible. In such a circumstance, the election commission just stands no chance.”*** *— Shivam Shankar Singh, campaign consultant.*

Identifying deep fake audio remains a challenge:

> ***“Some of the stories that we looked at during the recent state elections in Madhya Pradesh, Chhattisgarh, and Rajasthan, were the clips of Kaun Banega Crorepati, where Amitabh Bachchan is asking some questions and some answers are being given, which is targeting one of the political leaders. In fact, we couldn't even come to a proper conclusion there whether they were just dubbed audio or whether it was AI-generated cloned audio.” —****Jency Jacob, Boom Live Fact Check*

The use of deep fakes in state elections was for uplifting candidates:

> ***"During the Telangana elections, we saw some use of deep fakes, both from the BRS [Bharat Rashtra Samithi] and from the Indian National Congress, but those are mostly trying to create a positive image of the leaders who are participating in the elections to try and reinforce them as popular leaders or sensitive leaders, so on and so forth," —*** *Saikat Datta, Deepstrat*

Foreign state actors could use deep fakes to affect elections:

Inimical powers could make attempts to influence the results of elections through deep fakes by spreading messaging that aligns with their desired outcome. There was evidence of meddling by foreign state actors in the Taiwanese elections. However, ultimately, the deep fakes did not end up impacting the elections significantly.

III. Policy solutions to control deep fake technology:

Self-regulation by artificial intelligence companies is a good place to start:

Even though self-regulation has its limits, it is a good place to start. Given that AI generation capabilities are concentrated, the points at which you have to track misuse have narrowed. As such, big companies putting restrictions on the generation of certain kinds of content is a good step. There isn't a need for specific regulation for deep fakes, they already get captured under existing self regulation policies, what is important is to enforce these regulations effectively.

Self-regulation is a case-by-case solution:

> ***"So different strokes for different folks. So, we'll have to break it up into, there is the content generation part. So, those who are creating tools which can be used because it's mostly for free, like Image Creator is now free. And so therefore, there should be certain degree of regulation for them. Then the platforms which could be misused, then another set of regulations for them. And then look at also how we can deploy certain degrees of traceability, which again is challenging, but we've seen it work to an extent in CSAM. So, can we use some of those techniques that we could employ here?"*** *- Saikat Datta, Deepstrat*

Curbing cross-border flow of deep fakes should be left to autonomous bodies:

Tackling deep fakes cannot be left to the government because the political parties in power have their own biases. To curb the spread of deep fakes coming in from foreign state actors, there is a need for greater investment in autonomous bodies like the Election Commission of India. These autonomous bodies need a broad coalition of people including psychologists, sociologists, political science experts, and technologically capable people.

AI companies should not be able to monetize disinformation:

> ***"I think we have to recognize that platforms are creating these generative AI tools with a vested interest of creating more content that gets more eyeballs and more advertising revenue for them. Keep that premise in mind and kind of start critically looking at platforms and how their terms of use and policies should evolve. So, they're not able to monetize based on disinformation and misinformation. " -*** *Gautham Koorma, UC Berkeley*

IV. Can social media platforms curb the spread of deep fakes?

Takedowns are the only solution platforms have at their disposal:

> ***"I'm not aware of any platform that is right now at source using detection. Variety of challenges, the computational complexity associated with, when you're uploading a video, every time you have to analyze it using many models, plus even if you do that, the accuracy is relatively not at the level that they would want to productionize. The approach I've seen here generally around deep fakes is kind of the takedown or like after post-fact some forensic expert or fact checker or someone finds out that it's a fake and makes a claim that it's a fake. And I don't have this information off the top of my head, but I'm not sure if there's been takedowns, but I know that there have been articles published post-fact of the thing going online that, oh, this was a fake. I mean, it's also worth noting that typically the half-life of a social media post is considered to be a few hours, like within four to five hours, 50% of the viewers have seen that post. So, that's another thing to keep in mind that makes this problem a little more harder for platforms to kind of do detection"*** *— Gautham Koorma, UC Berkley*

Platforms need to control algorithmic amplification of content:

> ***"I think this debate around free speech and regulation, we need to kind of think carefully about it because platforms have been living under the protection of this safe harbor, section 230 out here and free speech. But what they are doing is algorithmic amplification of content that drives engagement. And if their algorithms see that certain salacious content or misinformation or disinformation is what is driving engagement and getting more eyeballs, I don't know if there exists really good checks and balances to make sure if these platforms can control the algorithmic amplification of those contents."*** *— Gautham Koorma, UC Berkley.*

Fake news content is taken down within two hours during elections:

> ***"I think most specifically to the elections bit, if you look at it, our takedown times are supposed to be two hours, and I think all the platforms have adhered to this. So, if deepfakes are found during the MCC, the model code of conduct time, we take down content within two hours, and that's part of the voluntary code that we've all signed up to. And more often than not, I think all platforms that are part of that generally end up taking down content far faster than even two hours when we get a lawful order."*** *— Berges Malu, ShareChat*

Shadow banning of content on social media would be more effective than content takedown:

> ***"Perhaps it could also be that instead of banning a certain content, while it is under investigation, they should perhaps contain the amplification of that content until the investigation is complete. Because if you have like a 36 hour timeline or a 24 hour timeline or a 12 hour timeline at some point in time that requires action quickly in terms of removal, perhaps a better way would be to then control amplification rather than removal"*** *— Nikhil Pahwa, MediaNama.*

Courts should take the final call on whether a platform's immunity needs to be taken away:

> ***"If the platform continues to be in clear violation of the whole intent of the rules and the act, then of course, the platform or the action in respect of, I'll say, withdrawal of immunity should be initiated and that too, it will be taken, I mean, the final call will be taken by the courts and not by the government."*** *— Rakesh Maheshwari, ex-MeitY*

V. Should Deep fake satire be fact-checked?

Labels on satirical deep fakes can be removed:

Fact-checkers have seen that although a deep fake is intended for satirical purposes and labeled as such, it is clipped out and mixed with other content and spread as fake news. A lot of social media handles proclaim that they are producing satire but are actually political shills.

Fact-checkers can end up in trouble for checking satire:

> ***"There was this one quote of Raghuram Rajan, which went viral sometime last year, where his quote was being, someone had created a quote, which he had not given***

> ***and had published a story. And then that became viral on various other platforms, where people actually started believing that it is true, that it's Raghuram Rajan who said it. And in fact, we wrote a story fact-checking that, and we got a legal notice from the one who created that, saying that we defamed them. And we had to then give a reply and also give a legal reply to the person who sent it to us. So, not just the fact that the one who has created that satire, whether that person is crossing the lines of defamation versus misinformation, but the fact that even fact-checkers, if we attempt to fact-check that, even we face the danger of probably being accused of defaming those who intended it to be satire."*** *— Jency Jacob, Boom Fact Check*

VI. Is Media Literacy the solution to deep fakes?

There is a need for more direct forms of media literacy:

> ***"To say that media literacy efforts do not help, I think would be pushing it too far. But at the same time, it has to be more direct. It has to be more unique methods of doing it. It could be little clips like those ads that are going about or more awareness that there is misinformation on the internet and that people will then figure out that even watching a video where the voice sounds a little funny may actually be a deep fake or synthetic media or whatever else you want to call it. So, there is a level of requirement of educating people of how the internet actually functions. And that would actually push this discussion further on fighting this kind of bad content on the internet."*** *— Berges Malu, ShareChat*

Attention conservation is an important part of media literacy:

> ***"It's a newer idea of media literacy that's come about because of the internet and it's this idea of attention conservation, right? So in the past, people would say you have to be critical thinkers, you have to evaluate different sources, see what is right, what is wrong. But I think what's happened over the last four years is that the wisdom is that, or where the research is coming is that if you have read a piece of content on the internet, the battle by the creator is like, they've already half won it. So, where people are moving towards in media literacy is this idea of attention conservation, which is that you stop, you evaluate whether it's even worth it for you to open that piece of media content, whether it's worth it for you to spend time on that website and you conserve your attention. And I feel like that is something that is even relevant for deep fakes, right? Which is why, because we are now in this space where there's so much content on the internet and our attention is saturated, it is precious. And so let's all evaluate how do you, or think about why, where we spend it.*** *" —* *Tarunima Prabhakar, Tattle Civic Technologies.*

Journalism should take charge of spreading media literacy:

> ***“..my limited point is we can't ignore the need for media literacy, which is a very boring term, I agree. And the problem is media literacy is being left to people in universities, though I think the challenge now is to make it mainstream. So, it should no longer be an education issue. It should be a journalism issue or an opportunity. —*** *Venkatesh HR, Boom FactCheck*

VII How deep fakes threaten identity verification:

There is a need to start thinking about the impact of deep fakes on verification:

> ***“..this [using deepfakes for exploiting verification systems] is still not exactly technically possible as of now, not that I'm aware of. But because identity and verification, et cetera, from a regulatory perspective, is so critical, like video KYC allows you to function, carry, create your UPI handle, so on and so forth. From a cyber crimes perspective, from a fraud perspective, from a verification perspective, [there are] major implications that we have to start wondering and thinking about."*** *- Saikat Datta, CEO, DeepStrat*

Video KYC would be challenged in a big way by deep fakes:

> ***“..this [using deepfakes for exploiting verification systems] is still not exactly technically possible as of now, not that I'm aware of. But because identity and verification, et cetera, from a regulatory perspective, is so critical, like video KYC allows you to function, carry, create your UPI handle, so on and so forth. From a cyber crimes perspective, from a fraud perspective, from a verification perspective, [there are] major implications that we have to start wondering and thinking about."*** *- Saikat Datta, CEO, DeepStrat*

Biometrics could be a solution to the identity verification question:

> ***“Technicalities aside, biometric systems need to evolve. Biometrics are a good way to deal with a lot of these things, but once we put them in place, it doesn't mean that they're gonna work forever. There are gonna be people who are gonna break it and these things continuously need to evolve. And there is research that is showing promising results for biometrics.”*** *— Gautham Koorma, UC Berkley*

Real-time deep fake verification is an emerging threat to identity verification:

> ***“...so, out here in the US, we're seeing a lot of cases where people are kind of coming to interviews, deepfaked as another person in Zoom calls and kind of getting the job. And yeah, and then once the interview is done, a whole different person who is not, well, it's the same face, cause his face is deepfaked. And it might not be convincing, but it will be convincing two years from now. So, we constantly need to talk about biometrics and authentication. And I think what we need to keep in mind is as these***

> ***tools evolve and as the adversaries have more sophisticated methods, these biometric systems also need to evolve to keep pace with them.”*** *— Gautham Koorma, UC Berkley*

Reference Material

- Rajeev Chandrasekhar Tells Online Platforms To Modify Terms And Conditions And Inform Users Not To Publish Deep Fakes [Read]
- Microsoft To Help Combat Deepfakes In The Run Up To 2024 Elections [Read]
- YouTube Announces New Measures For AI-Generated Content, Users Can Ask For Takedown Of Deepfakes [Read]
- Report: IT Ministry May Bring Amendments To IT Rules, 2021, To Regulate AI, Deep Fakes [Read]

Additional Reading

Understanding the threat posed by deep fakes:

- Deepfakes Can Have A Huge Impact On International Relations, But Can It Lead To Nuclear War? [Read]
- To Deepfake Is Human, to Detect Is Divine [Read]
- Supreme Court Justice Hima Kohli Flags Concerns With Deep Fakes: Report [Read]
- US Agencies Publish Cybersecurity Report on Deepfake Threats [Read]
- As Deepfakes Flourish, Countries Struggle With Response [Read]

Global discussions around elections and deep fakes:

- Big Tech must deal with disinformation or face fines, says EU [Read]
- AI-Generated False Information May “Radically” Impact Elections In Several Countries: World Economic Forum Report [Read]
- How deepfakes affect elections in the UK, USA, India, South Korea, Argentina [Read]
- California Bans Political Deepfakes During Elections And Non-Consensual Pornographic Deepfakes [Read]
- US Senator Writes A Letter To Tech Companies Urging For Disclosures On AI-Generated Content [Read]
- U.S. lawmakers question Meta and X over AI-generated political deep fakes ahead of 2024 election [Read]
- EU Plans New Laws For AI Music Labeling And Deep Fake Regulation [Read]

Indian Government's response to deep fakes:

- Take Down Deepfakes Within 24 Hours, IT Ministry Tells Social Media Platforms: Report [Read]
- Platforms That Do Not Meet Deep Fake Takedown Obligations Will Lose Safe Harbour, Rajeev Chandrasekhar Says [Read]
- Work On Regulations For Deep Fakes To Start Immediately, Says Ashwini Vaishnaw: Report [Read]
- Indian Government To Discuss Deepfake Regulations With Social Media Giants [Read]
- Are Deep Fakes Different From Misinformation? Rajeev Chandrasekhar And Priyank Kharge Debate [Read]
- IT Ministry Issues Advisory To "All Intermediaries" On Combating Deep Fakes Through IT Rules, 2021 [Read]
- FIR Filed Against Owner Of Gaming Platform Hosting Sachin Tendulkar Deep Fake: Report [Read]

Tech companies' responses to deep fakes:

- Google Releases Large Dataset Of Deepfakes For Researchers [Read]
- Facebook, Microsoft Announce Challenge To Detect Deepfakes [Read]
- YouTube Announces New Measures For AI-Generated Content, Users Can Ask For Takedown Of Deepfakes [Read]
- YouTube Cracks Down On AI-Generated Content Featuring Victims Of Crime [Read]
- McAfee Reveals Its Deepfake Audio Detection Technology "Project Mockingbird" [Read]
- Election Integrity In The AI Era: Open AI Lists New Measures In Preparation Of 2024 [Read]

MediaNama is the premier source of information & analysis on Digital Policy in India. We focus on key issues like privacy, data governance, fake news, misinformation, cybersecurity, cyber diplomacy, digital payments policy, Net Neutrality, intermediary liability, website blocking, internet shutdowns, data localisation, e-commerce policy, IoT, content regulation and censorship, among others.

Our mission is to help build a digital ecosystem which is open, fair, competitive, and global.

On MediaNama.com, our reportage attracts a readership of policy professionals, government officials, Members of Parliament, startup founders and business leaders, as well as investors with an eye on policy developments shaping the future of the Internet in India.

MediaNama's events enable meaningful conversations for a high quality and curated audience to discuss opportunities, challenges and issues that they care about, in a manner that helps in building capacity. The fact that they are among peers with similar depth of understanding, and come at the same issues from a different point of view, ensures that our audience keeps coming back for more.

Nikhil Pahwa | Founder & Editor | nikhil@medianama.com

MediaNama.com | Twitter | Facebook | YouTube | LinkedIn
Subscribe to Telegram | Newsletter