Deepfakes: Envisioning Prospects & Perils

Deepfakes: Altering Reality?

In a rare instance, the US President stated that "AIDS" had been eradicated across the world. While this was a fake video, its realism shocked many and deliberately damaged the US President's credibility. This is a classic instance of a "Deepfake", a new gamut of technologies that are aimed to fake or manipulate videos, or other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds, thereby making them appear real.

The Technology Behind Deepfakes

The whole phenomenon of Deepfakes centers around the Generative Adversarial Networks (GANs) approach. GANs was discovered by Ian Goodfellow and researchers at the University of Montreal in the year 2014[1].

Rather than a technology, GANs is an approach to "generative modeling using deep learning methods". "Generative" implies that the underlying property of GANs is to create or generate data of its own. For example, if GANs were to be fed with innumerable images, it can generate an image similar to but not the same as the input images, on its own. This can be extended to other forms of media as well. However, there is no mechanism to evaluate the authenticity and acceptability of the output "generated". For this purpose, the "Adversarial" aspects of GANs encompass a discriminative network* that validates the generated data against true data. Simply stated, the generative network and discriminative network are put against each other, creating adversaries against each other.

Essentially, GANs are a "two-player game where each player is trying their hardest to beat one another": the Discriminator's role is to keep a check on the generated values of the Generator; and the Generator's task is to push the discriminator into thinking that generated values are actually real.

The Powerful Possibilities of Altered Reality

While the negative implications of this nascent technology have been widely discussed, firms and businesses are yet to focus their efforts in realizing the unlimited opportunities and generative applications this technology has to offer. The positive impacts of Deepfakes are not limited to the media, entertainment and advertising industries alone, but cuts across several other domains.

Deepfakes offer two key benefits – Firstly, from the consumer's perspective, they offer enhanced, customization or personalization of media and user experiences. Secondly, from a firm's point of view, the generative aspects of Deepfakes can

be widely employed in automating time-consuming, human-centric repetitive tasks. Some of the potential applications of Deepfakes are outlined below:

- **Video Advertisements:** The advertising industry presents a range of opportunities for the application of Deepfakes. Below are two examples:
  - **Celebrity Endorsement:** In the fashion industry, GANs could help in blending the real world with virtual artefacts. For e.g. celebrities could agree to advertise for a fashion line, without doing a photoshoot. The use of GANs could even go to the extent of personalizing a website and its e-commerce arms for individual shoppers, thereby increase customer visits, and purchase through virtual celebrity endorsements.
  - **Product Placement:** A key area that remains unexplored is the usage of Deepfakes beyond human faces and sounds and extending it to include objects. Big advertising agencies and media companies are engaged in crafting out advertisements for business entities, especially consumer goods companies that have multiple product placements in their ads It is a tedious task for the designer to alter products in an advertisement manually for different geographies. Using Deepfakes, product placement and advertising campaigns could be personalized for geographies by morphing region-specific products onto the ad content.
- **Localization of Short Videos:** Deepfakes introduce a new way of localizing video content. For example, if the government launches a new campaign directed towards Indian farmers, in order to disseminate content free of ethnicity bias, Deepfakes can be deployed to alter the actors' ethnicity and customize content to different ethnic groups. Such an application of the technology can be applied across different media elements from face to even audio using technologies such as Native Dubbing[2], to ensure that a wider audience relates to the content.
- **Dialogue customization:** Deepfakes can be widely used in advertising campaigns, e-learning materials etc., that are directed towards different demographic groups and geographies by customizing the dialogue delivery to suit the audience's taste as well as for movie post-production. In fact, a group of scientists from Stanford University, the Max Planck

Institute for Informatics, Princeton University, and Adobe Research[3], have made a headway to solve this tricky problem. They have designed a tool to "edit talking-head video based on its transcript to produce a realistic output video in which the dialogue of the speaker has been modified".

- **Movie Production:** While viewers mourned the tragic death of the actor Paul Walker before the completion of the film, Furious 7, computer-generated image (CGI) technology "put him back" into the film[4]. As on-demand streaming media providers and telecom companies are introducing a plethora of original movies and programs on their platforms, Deepfakes will play a crucial role during production by bypassing labor-centric and lengthy processes.
- **Radio Advertisements & Podcasts:** Using a text-to-speech deep learning system, radio ads and podcasts can be made more personalized to suit a consumer's preference. Such an application of a Deepfake was exhibited in a recently developed replica of popular podcaster Joe Rogan's voice that made his "talk" about how he was sponsoring a hockey team made of chimpanzees[5].
- **Synthetic Media:** Using Deepfakes, synthetic media (synthetic celebrities, videos and audios etc.) can be generated. Since there are no ethical guidelines associated with such media, a mere disclaimer to consumers of their nature usually suffices. Such an application can go a long way in managing scheduling, could result in cost savings, and guarantee greater customization.
- **Integration into AR/VR applications:** If Deepfakes and immersive experience technologies like AR/VR were to operate symbiotically, their impact increases exponentially, in that they reinforce each other's perceived level of realism. Creation of 'synthetic realities' has predominantly been restricted to games but here we are looking at scenarios where a prominent professor conducts a session, business executives holds remote meetings, with a feeling that the parties are in the same room, without incurring travel costs and seeing superimposed faces of choice in a movie. Another example of the application of AR/VR with Deepfakes is witnessed in retail companies trying to show the placement of different furniture and home décor artefacts with respect to the layout of the customers' actual rooms or houses. This brings customization, and helps the customer satisfy their

need to select in terms of color, size, design, style, positioning etc.

- **Deepfakes in the Recreation and Amusement Industry:** A museum in St. Petersburg uses Deepfake to entice their visitors with words from the "real" Salvador Dali[6]. There can be many such applications in museums, amusement parks etc., that can help improve customer experience and result in more footfall.

Imminent Threats posed by Deepfakes

While we discussed the powerful possibilities of Deepfakes in future applications, the applications of Deepfakes that have captured mass attention are those that have a negative connotation associated with them. A few such are highlighted below:

- **Misinformation & Deceptive Campaigns:** The application of Deepfakes has surpassed the mere publishing of fake news articles; they have also been used as a weapon for malpractices across online channels. Such activities are detrimental to companies' businesses, credit ratings and their overall reputation. Very recently, Deepfakes were used to compromise the organizational cybersecurity of a UK-based energy firm that fell prey to a hoax call, where the caller posed as the CEO of the parent company[7]. The call requested an immediate money transfer of €220,000 ($243,000) to an unknown Hungarian supplier. The adverse effects of this new type of identity theft may not only be felt by corporates and smaller organizations but could even trickle down to the simulation of someone's image and behavior and using them as spam callers to obtain personal information.
- **Political Manipulation:** One of the most widely publicized renditions of Deepfakes are the videos of Donald Trump claiming "AIDS is over" and the "satirical" interview of Democratic Congressional candidate, Alexandra Ocasio Cortez, where she sheepishly shakes her head when asked if she understood anything about politics[8]. The videos went viral across social media. Such videos can play a key role in swaying public opinion for or against political figures to the extent of creating uproar and causing threat to the national security of countries having numerous sensitive issues.
- **Creation of non-consensual pornographic videos:** This was one of the earliest malicious applications of Deepfakes.

It emerged anonymously on Reddit in 2017 and in the following year led to Discord[9], a digital distribution platform, shutting down a chat group that was spreading Deepfake pornographic videos of female celebrities without their consent. Lately, DeepNude was released that used neural networks to remove clothing from images of women[10] (needless to add, the app was shut down).

- **Privacy and Consent:** The era of Deepfakes has also ushered in several privacy and consent issues. For instance, in September 2019, a Chinese Deepfake app called Zao[11] was launched; it allowed users to morph their faces over that of celebrities.
- **Cyber Blackmail:** Deepfakes induced dangers for targeted individuals and celebrities, forcing them to pay over the cyber internet, might not be a rare phenomenon in the future. It could go hand in hand with similar incidents of Malwares, Spywares etc. in the recent past.

Deepfakes: Evolution and Future Forecasts

A careful analysis of the opportunities and threats posed by Deepfakes reveals that this new technology can take three possible paths as shown in the Fig 1, with varied implications for each of the three paths.

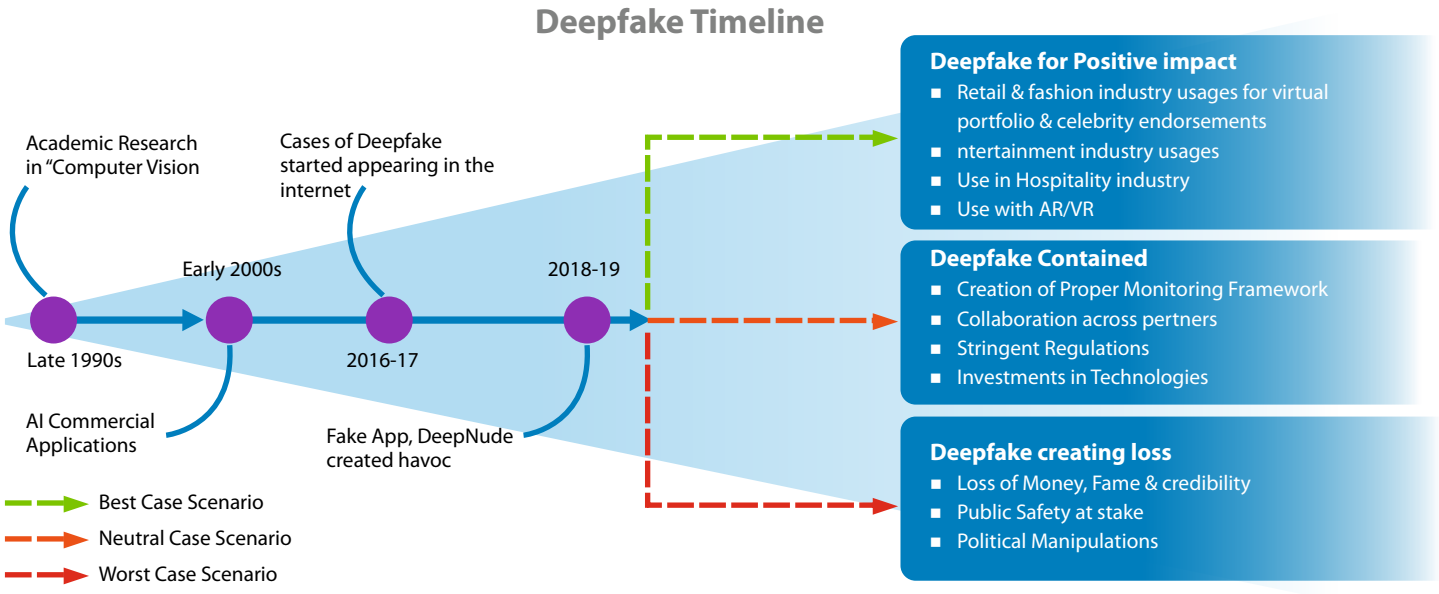

Fig 1: Hypothetical Deepfake timeline

The figure above reflects the probable journey of the evolution of various technologies and aspects related to Deepfakes in the future.

Combating Deepfake threats: The S-L-T approach

Even though Deepfakes are becoming quite convincing, it is still relatively easy to spot the differences between the "real" and the "fake". Down the line, the task will become more difficult as content that is more realistic starts to appear, thus sparking the need for better ways to identify Deepfakes. A framework has been envisaged (Fig 2), to help fight Deepfakes, ensure its proper monitoring and control, and leverage the immense potential of this technology.

S (Societal) – L (Legal) – T (Technological) Framework to Tackle Deepfakes

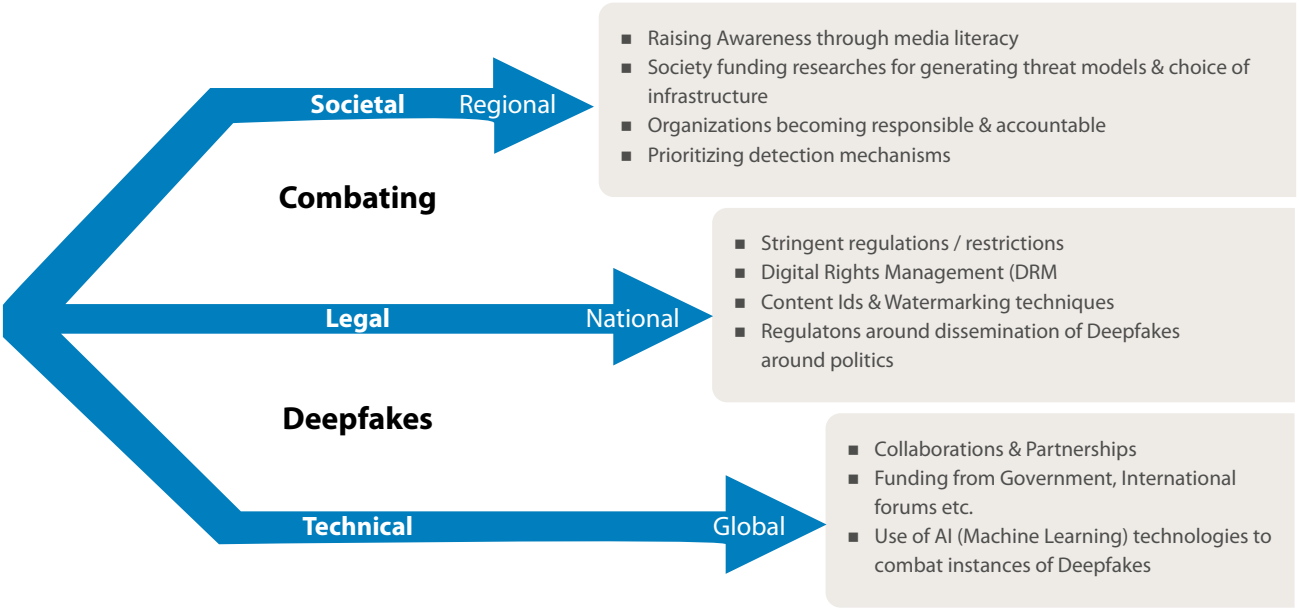

Fig 2: Fighting Deepfakes with an S-L-T framework

The Socio-Legal-Technical (SLT) framework is a three-pronged approach to mitigate the risks associated with Deepfakes.

Societal Approach

A regional approach that cannot be generalized for a nation or for the whole world, tasking societies to raise awareness along with providing proper monitoring mechanisms to subdue the cases of Deepfakes that may arise. The ambiguity of satire versus genuine content is a conundrum that tricks many. This

approach for reducing Deepfakes seeks to reduce such an ambiguity by improving media literacy amongst viewers or consumers, enabling them to identify the differences between a real piece of media and a Deepfake. Local governments and societies are also expected to act and fund research to combat potential damages. This approach would lead to the formation of global threat models and relative mechanisms to combat threats and provide background information on the choice of infrastructure.

Legal Approach

This nationwide approach would require that developed and developing countries delve upon upcoming threat scenarios and come up with stringent regulations and restrictions regarding Deepfakes. Regulations around the dissemination of Deepfakes for political propaganda and non-consensual adult content have already been rolled out in the US in states such as Texas, Virginia and lately, California. The adoption of Digital Rights Management (DRM) will protect certain video files and grant partial permission for their use. Also, introducing content IDs and watermarking of original videos will help keep a check on any future manipulations.

Technical Approach

The technical approach is more global or universal in nature and would require collaborations across various parties, international companies, multinational enterprises and more specifically deep engagements with technology companies. AI has been leveraged to create the Deepfakes in the first place, therefore, using the same AI technology to detect instances of Deepfakes has been a buzzing area of research. Facebook and Microsoft have also partnered with universities for Deepfake detection and Amazon is a recent addition to this "army" against Deepfakes[12] Organizations should also be prudent enough to instill desired responsibility and accountability with respect to the creation of platforms and tools and thereby prioritize the methods of detection systems in collaboration with other stakeholders.

Future Forward: Reinventing viewer perceptions

Although media manipulation applications and techniques have been in existence almost since the inception of media itself, the introduction of this GAN-backed technology has added an entirely new flavor to the era of malicious and cybersecurity attacks.

However, at a time where Deepfakes are gaining a lot of traction for their negative consequences, the application of the SLT framework serves as a roadmap not only restricted to combating malicious Deepfakes, but rather, streamlining their positive applications. This, thus, reduces the risk factor and paves the way for an era of Deepfakes being used extensively to aid creative and graphic designers, media production houses, advertising agencies and the like. In short, it involves reinventing the very perception Deepfakes have made in the audiences' mind and replacing it with the promising applications they bring to the table. With that said, Deepfakes have just made their appearance, and recent events suggest that they are likely here to stay.

References

[1] *Research Paper: "Generative Adversarial Nets": Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley,Sherjil Ozair, Aaron Courville, Yoshua Bengio (2014)*

* *Discriminative network helps in keeping the generated values from the generator under check, so that the generator is not being able to create false values against the real ones. This is the major underlying concept defining GANs.*

[2] *Deepfake technology: a "Localization" Boon for Sales Trainers?*

[3] *Research Paper: "Text-based Editing of Talking-head Video" (July 2019)*

[4] *Furious 7 used 350 CGI shots of Paul Walker*

[5] *RealTalk: This Speech Synthesis Model Our Engineers Built Recreates a Human Voice Perfectly*

[6] *Museum creates Deepfake Salvador Dalí to greet visitors*

[7] *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case – The Wall Street Journal*

[8] *A million Facebook users watched a video that blurs the line between bad satire and 'fake news'*

[9] *Discord just shut down a chat group dedicated to sharing porn videos edited with AI to include images of celebrities*

[10] *Github is banning copies of 'deepfakes' porn app DeepNude*

[11] *Another convincing deepfake app goes viral prompting immediate privacy backlash – The Verge*

[12] *Amazon joins Facebook and Microsoft to fight Deepfakes*

Other References

1. *FT 2020 Trend Report for Entertainment, Media & Technology (3rd Annual Edition)*

2. *The State of Deepfakes: Reality Under Attack (DeepTrace 2018 Report)*

About The Authors

Iaphi Tariang,

Innovation Evangelist, Media, Entertainment & Advertising – Corporate R&I, TCS

Iaphi Tariang is an Innovation Evangelist for the recently established Media, Entertainment and Advertising (ME&A) Research Area. She has experience in marketing, communications and market research across various domains such as Energy, IT, Magazine, Media and Education Management. She holds an MBA degree from Indian Institute of Technology, Madras, and has completed Grade 5 Piano from Trinity College, London.

Swayambhu Dutta

Research Analyst, Marketing Transformation – Research team, TCS

Swayambhu Dutta is a research analyst in the Energy & Resources (E&R) domain looking after the North America, ANZ, & MEA geographies. He has around seven years of experience spanning across the E&R and telecom verticals. Currently, he works as a business analyst, enabling strategic decisions and deals support for the company. Swayambhu holds an MBA gold medalist from XIM, Bhubaneswar. He is a telecom engineer with a deep interest in oil and gas industry value chain technologies.

Chayan Bandyopadhyay

Research Analyst, Marketing Transformation – Research team, TCS

Chayan Bandyopadhyay is a research analyst in the Banking, Financial Services and Insurance (BFSI) vertical focusing on major markets such as North America and Europe. Chayan also provides research and advisory services to the Research and Innovation (R&I) team at TCS. He has around nine years of cross-industry experience across verticals such as Information Technology, Telecom, Banking and Energy. He is keenly interested in cutting-edge technologies and their effects in various industries. In his current role, he enables strategic decisions within the company through his research and advisory. Chayan is a Gold medalist during his MBA and completed B. Tech. in Electronics and Communications Engineering.

Contact

Visit the Research and Innovation page on www.tcs.com

Email: innovation.info@tcs.com

Blog: #Research and Innovation

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 02 I 20