Fading boundary between Information and Misinformation

Nikunj Saini

20312915028

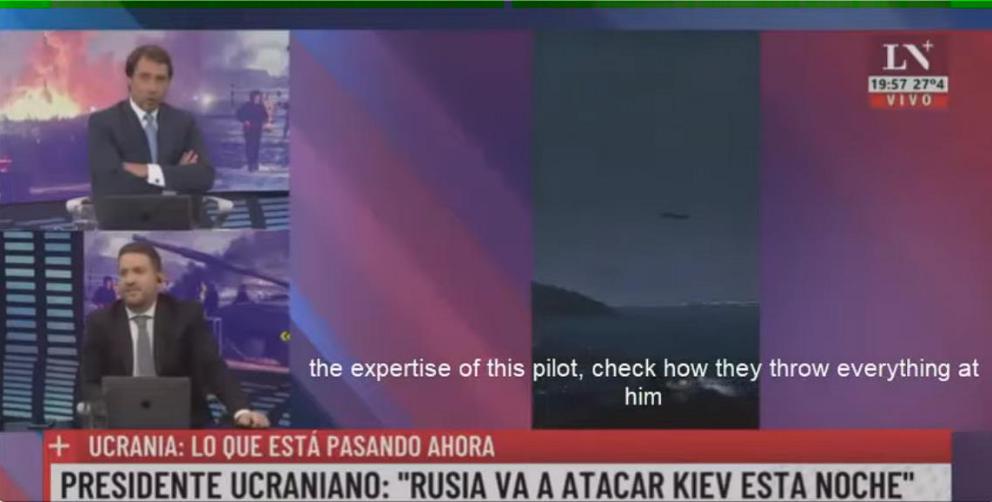
“False information wrapped in facts is expected to be believed by people”

~Anonymous

Well, the above quote has not been said by any prominent personality, yet it does reflect the mentality of people. People often fall for that one lie that is supported by facts or ‘meant to be obvious’ reasons. The flat earth controversy is an example of it. Eventually, it got busted by scientists, but for a long time, some people did believe in it. Well, that was an old example, and it did not impact the masses in an ugly way. But nowadays, the consequences created by such false information often lead to havoc in society.

Maybe we should start by discussing what is fake news. The unethical broadcast of incorrect or deceptive material appearing as actual news is known as fake news. It frequently seeks to advance agendas, confuse the public, or influence public opinion. It is vital for people to critically assess sources and double-check information before believing or spreading it in the digital era due to the rapid spread of false information through social media and online platforms. When such sort of misinformation is spread via mass media, then it does become harmful to society. Let us have a look at the example of Russia – Ukraine conflict. A prolonged and complex problem, the conflict between Russia and Ukraine has been defined by complicated historical circumstances, territorial conflicts, and geopolitical tensions. This dispute has gotten worse due in large part to the spread of false information and fake news.

Propaganda, fake news, and disinformation have all been employed as instruments to sway public opinion, rewrite history, and alter perceptions of the battle. There is now more division and hostility between Russians and Ukrainians because of both sides using the media, social networks, and internet platforms to spread misleading or erroneous information to further their own agendas were being covered by global media. The media instead of checking the validity and authenticity of the news, kept on showing clips of the game ‘Arma – 3’ on AIR and claiming it to be the live feed from battleground.



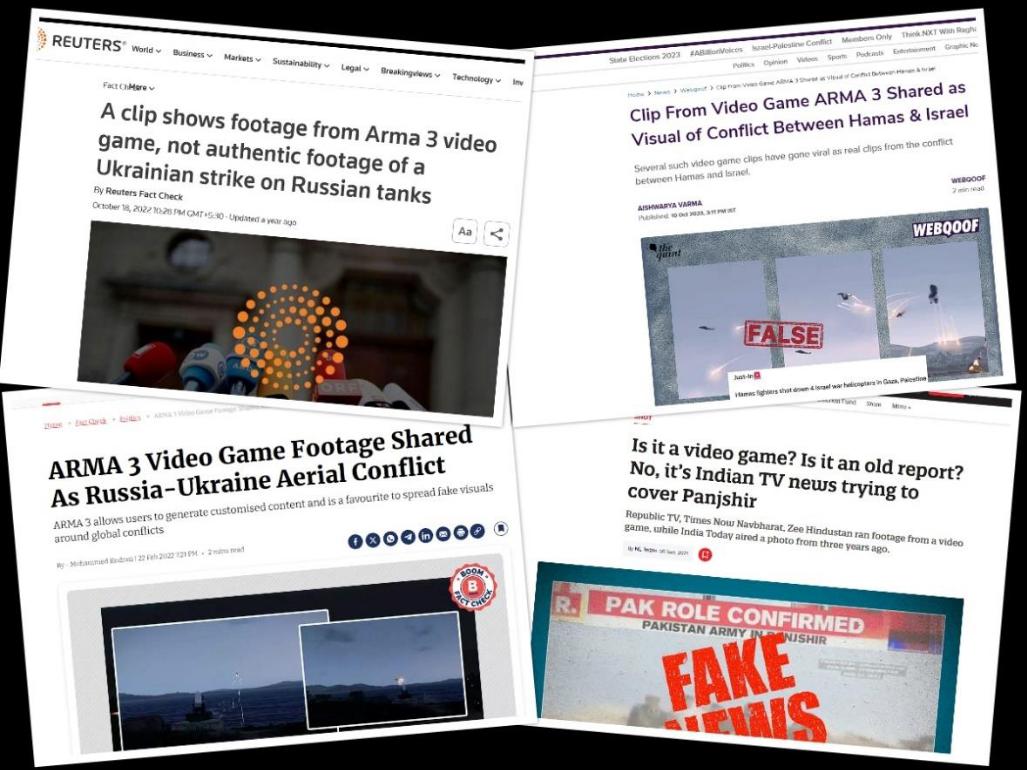
*Figure 1- Argentine news channel shows ArmA 3 video as Ukrainian footage*

Indian media was also seen broadcasting these clips with different headings and in the context of the India – Pakistan conflict.



*Figure 2- Indian news channel shows ArmA 3 video*

Later some fact-checking organisations such as Reuters, The Quint, BOOM and Newslaundary recognised the clip from the game and raised flags against mainstream media spreading such fake news to such a great extent.



*Figure 3 - Articles from Fact checkers raising flags*

Such kind of fake news was being spread by the media houses due to negligence and carelessness. But sometimes, fake news is also spread from social media which can cause social unrest on a very large scale. Let us consider the case study of Manipur violence. The recent violence in Manipur, India, has been a source of concern, originating from several underlying causes. Tensions have arisen in the area, resulting in fights over resources and territory as well as ethnic confrontations. Conflicts between various communities have occasionally gotten out of hand, which has tragically resulted in violent confrontations, evictions, and fatalities. Fake news has been fanning the flames in Manipur by stoking fear, promoting disinformation, and escalating animosity amongst various communities. On social media, aggressive words, photoshopped photographs, and false narratives have increased tensions and further divided communities.



*Figure 4- Fake News being spread in the context of Manipur violence*

Artificial Intelligence – Generative AI taking deep fakes to new levels

Till now, we have talked about fake news that was created by humans. They can still be checked for their authenticity. But what if something that has never happened is created by and spread out in the public? To know about this, we need to know about generative AI.

Artificial intelligence that mimics human creativity by producing new content, such as literature, music, or graphics, is known as generative AI. This technology creates new material based on patterns it has learned from massive volumes of data by using neural networks. Over time, generative AI has advanced significantly, moving from simple models to more complex and adaptable systems. Autoencoders and Variational Autoencoders (VAEs) were employed in the initial stages. The primary goals of these early models were to generate new data from latent representations and to learn data representations. Specifically, VAEs made it possible to generate fresh, credible data points. After that, Ian Goodfellow proposed the idea of Generative Adversarial Networks in 2014. By combining the generator and discriminator networks in a manner like a game, GANs created a new paradigm in generative modelling. To promote competition and raise the quality of the generator's output, the discriminator looks for differences between created and real samples, while the generator produces samples.

The invention of Progressive GANs, StyleGAN, and BigGAN was primarily responsible for advancements in GAN technology. Through the gradual addition of layers, progressive GANs generated high-quality images by beginning with low resolution and gradually increasing complexity. Better control over generated images was made possible by StyleGAN and StyleGAN2, which allowed for the adjustment of some properties (such as facial emotions in faces) without sacrificing realism. Subsequently, BigGAN, which was composed of large-scale models, succeeded in producing images of superior quality through the process of scaling up the model size and training data.

However, generative AI also presents ethical questions and obstacles, despite its amazing potential. The topic of bias included in the data used to train these algorithms is still a hot topic of debate. Prejudiced datasets can produce AI results that reinforce or even mirror existing prejudices, resulting in the spread of false information or discrimination. Moreover, trust and accountability are hampered by the generated content's legitimacy. Differentiating AI-generated content from human-created content gets harder and harder as these systems get more advanced. This raises the question of whether content producers and platforms should identify or verify AI-generated material to preserve user confidence and transparency.

With only a single input, users may generate everything from poetry to exotic pictures thanks to generative AI technology, which went viral last year. By 2030, the generative AI market is projected to reach a valuation of over $100 billion, with an average annual growth rate exceeding 35 percent. The past year has seen significant advancements in generative AI, making the task of spotting deepfakes more difficult.

How Generative AI is causing problems?

According to a study by the international consulting firm KPMG, there are more than 100,000 AI-based models available for producing or identifying deepfakes. simply roughly 3,000 of them are "able to identify deepfakes properly," with most of them simply being able to distinguish between the more blatant fakes. Sensity.ai, a visual threat intelligence firm based in the Netherlands, claims that since 2018, the quantity of bogus films on the internet has increased dramatically, nearly tripling every six months. As of December 2020, the company claims to have found 85,047 fraudulent videos on the internet.

The emergence of incredibly lifelike AI-generated deepfakes has raised several issues and concerns in a variety of fields. Deepfakes can change audio and video content to create fake stories, spread false information, or harm people's reputations. The accuracy of information published online is seriously threatened by this, which could increase the distrust of the media and institutions. It gets harder to distinguish altered content from true content as deepfakes get more realistic. Institutions are also impacted by this decline in trust, which breeds doubt and uncertainty regarding the veracity of information. Deepfakes can be used to fabricate events, speeches, or interviews with important personalities to sway public opinion and possibly spark political or social turmoil.

Let us have a look at the case when an AI-generated image that purported to show an explosion near a structure in the Pentagon complex went viral on social media on Monday. On Twitter, the picture of a towering, dark grey cloud of smoke immediately became viral, even getting shared by verified accounts. Where it started is still unknown. The photograph was a hoax, as certified by the US Department of Defence. Also, as reported by CNN, the stock market appeared to briefly decline because of its virality.



*Figure 5- Viral image of explosion in Pentagon Complex. Later it was found to be AI-generated.*

One of the verified pages that shared the images was OSINTdefender, a Twitter handle with over 336,000 followers that provides news about foreign military engagements. The person in charge of the page expressed regret for disseminating false information and stated that the episode demonstrated how dangerous it may become in the future and how readily these kinds of photos could be exploited to alter the information space.

Another story of the deepfake audio that was used to clone the boss's voice and steal $243,000 from a multinational company's UK unit made news in 2019. New Kite Data Labs, a US-based data analytics company, recently alleged that SpeechOcean, a Beijing-based private artificial intelligence company with clients connected to the People's Liberation Army of China, has been gathering voice samples from India, mostly from the sensitive border regions of Jammu

and Kashmir and Punjab. It appears that locals are paid to record pre-written lines of speech or conversations, which are then uploaded to servers located in China.

People's faces and voices can be placed over explicit or compromised content, which can violate their privacy and damage their reputation. Deepfake technology compromises security and integrity since it can be used for phishing attacks or to tamper with audio or video evidence in court. Deepfakes are made and shared, which brings up difficult moral issues with consent, privacy rights, and responsible usage of AI-generated media.



*Figure 6- Miles Fisher impersonating Tom Cruise with the help of AI*

Although deepfakes can be used fraudulently, Miles Fisher, a popular figure on TikTok, demonstrated that there is a lighter side to the phenomenon by brilliantly impersonating Tom Cruise. In his videos, Fisher emphasises the significance of realising the possibilities of this cutting-edge technology, stressing that it may be used for artistic expression in addition to being a tool for deception.

How to stop Deepfakes?

Well till now we have talked about what all problems are caused by deepfakes. But the problem is how to deal with it, well there are many steps that are in the pipeline to be executed to deal with them. To address the issues raised by deepfakes, several actions have been made with the goal of lessening their detrimental effects and improving methods for detection and prevention.

To identify deepfakes, IT businesses, academic institutions, and research centres are investing in the development of advanced tools and algorithms. These instruments analyse anomalies, objects, and inconsistencies in audio and video clips that point to manipulation. To detect any

manipulations, researchers are using algorithms that examine patterns, discrepancies, or distortions in audio, pictures, and videos. To improve their detecting powers, these algorithms learn from large datasets of both authentic and altered content. Some solutions make it more difficult for deepfakes to remain undetected by using cryptographic techniques or blockchain technology to establish digital signatures or watermarks that confirm the validity of content.

Governments, IT firms, universities, and non-governmental organisations can work together to share expertise, pool resources, and launch collaborative projects to create detection techniques that are more accurate. Diverse stakeholders are brought together by initiatives such as the Partnership on AI to tackle issues linked to AI, such as deepfakes. The development of detection technology is aided by projects such as the DARPA Deepfake Detection Challenge and collaborations between academic institutions and business entities.

Governments are thinking about enacting or revising laws and rules that particularly concern the production, use, and distribution of deepfakes. By guaranteeing that those who abuse technology will face consequences from the law, these regulations seek to make people or organisations responsible for the destructive use of modified content.

Conclusion

Misinformation, fake news, and deepfake technologies pose a complex problem that affects governance, technology, and society. False information has spread and become more intense over time, impacting people's views on a local, national, and international level. Examples of this include historical events such as the flat earth controversy, contemporary conflicts, and the rise of generative AI-driven deepfakes. Although generative AI has significantly improved the creation of creative content, it has also created new ethical problems and obstacles. Transparency, accountability, and ethical standards are essential in the development and application of AI because of the possibility of bias in training data and the challenge of telling AI-generated content from real human-created content.

The goal of projects concentrating on the creation of algorithms and detection tools is to locate and stop the spread of deepfakes. Cooperation aims to pool resources and expertise for more efficient solutions across governments, IT businesses, and research institutes. To counter the exploitation of deepfake technology and hold persons accountable for its harmful usage, legislative measures are being developed.

To wrap things up, the spread of false information, the development of generative artificial intelligence, and the appearance of realistic deepfakes provide intricate problems that call for diverse approaches. To protect against the harmful effects of disinformation and guarantee the responsible development and use of AI-driven technologies in an increasingly digital world, a coordinated effort involving technological advancements, ethical guidelines, education, legislation, and international collaboration is essential.

**References**

1. [How AI-Powered Fake News is Set to Devastate India | Akash Banerjee & Manjul](https://www.youtube.com/watch?v=p_6Aune8dbQ)
2. <https://www.newslaundry.com/2021/09/06/is-it-a-video-game-is-it-an-old-report-no-its-indian-tv-news-trying-to-cover-panjshir>
3. <https://www.reuters.com/article/factcheck-arma-3-ukraine-idUSL1N31J1L5/>
4. <https://www.thequint.com/news/webqoof/video-game-arma-3-clip-viral-amid-israel-and-palestine>
5. <https://www.boomlive.in/fact-check/politics/arma-3-video-game-footage-shared-as-russia-ukraine-aerial-conflict-16874>
6. <https://www.reuters.com/world/india/why-is-indias-manipur-state-grip-ethnic-violence-2023-07-21/>
7. [https://thewire.in/security/manipur-violence-reasons-failure-and-solution-q-and-a#:~:text=Fake%20news%20about%20a%20Meitei,of%20them%20was%20reportedl](https://thewire.in/security/manipur-violence-reasons-failure-and-solution-q-and-a#:~:text=Fake%20news%20about%20a%20Meitei,of%20them%20was%20reportedly%20raped) [y%20raped.](https://thewire.in/security/manipur-violence-reasons-failure-and-solution-q-and-a#:~:text=Fake%20news%20about%20a%20Meitei,of%20them%20was%20reportedly%20raped)
8. <https://www.boomlive.in/fact-check/akshay-kumar-old-video-on-bengaluru-molestation-shared-as-recent-reaction-on-manipur-incident-22573>
9. <https://www.boomlive.in/fact-check/politics/fake-news-viral-video-manipur-police-poll-violence-2022-manipur-violence-kukis-meitei-factcheck-22646>
10. <https://www.theweek.in/news/india/2023/05/05/manipur-riots-army-warns-against-fake-videos-congress-blames-bjp-for-politics-of-hate.html>
11. <https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image-explosion>
12. <https://timesofindia.indiatimes.com/entertainment/english/hollywood/news/rashmika-mandanna-deepfake-horror-scarlett-johansson-tom-hanks-and-other-celebs-whose-images-were-morphed/photostory/105028778.cms?picid=105028809>
13. <https://www.indiatoday.in/magazine/cover-story/story/20231204-deepfakes-clear-and-present-danger-2466673-2023-11-24>