**University of Maryland, Baltimore County.**

Hinal Bhavesh Desai.

Ethical and Legal Issues in Data Science.

Artificial Intelligence (Paper 3)

Prof. Sunela Thomas

May 11, 2022.

**Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. AI is an interdisciplinary science with multiple approaches, but advancements in machine learning and deep learning are creating a paradigm shift in virtually every sector of the tech industry.** (Artificial Intelligence, n.d.)**. Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.** (Education, 2020)**. Artificial Intelligence (AI) is already the future of innovation and technological advances.**

**The term "deepfake" comes from the underlying technology "deep learning," which is a form of AI. Deep learning algorithms, which teach themselves how to solve problems when given large sets of data, are used to swap faces in the video and digital content to make realistic-looking fake media. There are several methods for creating deepfakes, but the most common relies on the use of deep neural networks involving autoencoders that employ a face-swapping technique. You first need a target video to use as the basis of the deepfake and then a collection of video clips of the person you want to insert in the target.** (Johnson, 2021)

**Deepfakes are one of the most unique phenomena of the last five years in the world of synthetic media. Synthetic media is AI-generated or AI-modified content. With traditional media, people relied on broadcasting networks (radio and TV) and social networks to create and distribute their content. This method of receiving information came with certain restrictions, which, of course, were significantly weakened by the emergence of social networks. With synthetic media, content creators can create content with a high level of quality that was previously only available to major studios with massive budgets. AI content is cheaper and easier to scale. However, this new trend raises several ethical concerns. They all boil down to one essential problem: how to distinguish AI-synthesized content from authentic content? In short, deepfakes are artificial intelligence-based images and sound synthesis techniques. They are used to join and overlay existing images, videos, and soundtracks onto original content.** (What Are Deepfakes: Synthetic Media Explained, 2021)**.**

**Deepfake works in 3 key steps:**

1. **It begins with inserting the original video or voice of the target character into a neural network. Autoencoder and GAN algorithms go to work by analyzing the subject's facial expressions and main features.**
2. **Combining an autoencoder with GAN allows the algorithm to generate fake images until it can no longer distinguish them from the original.**
3. **The video with the stunt double is then inserted into the neural network. After having analyzed the facial characteristics of the target subject, the network can easily generate a deepfake. The target subject's face is then overlaid onto the video of the stunt double.** (What Are Deepfakes: Synthetic Media Explained, 2021)

**One such use of deep fake technology was done by the advertising giant WPP. WPP will send unusual corporate training videos to tens of thousands of employees worldwide. A presenter will speak in the recipient’s language and address them by name while explaining some basic concepts in artificial intelligence. The videos themselves will be powerful demonstrations of what AI can do: The face, and the words it speaks, will be synthesized by software. But WPP’s CTO claims that the deepfake model used was not accurate and had various flaws like the prosody of the presenters’ delivery was off. The term deepfakes comes from the Reddit username of the person or persons who in 2017 released a series of pornographic clips modified using machine learning to include the faces of Hollywood actresses. This is ethically incorrect. It can cause mental trauma and depression to a few people. Defaming celebrities in such a manner is incorrect if we consider the Social Contract theory where human rights are to be given justice, Virtue Ethics where a person’s actions should be virtuous, and Utilitarianism where the costs weigh the benefits by a huge margin. Synthesia is one such organization that creates videos using deep fake technology. It lets you choose an avatar from its vast list and then lets you enter the script. This might have various uses but is also equally hazardous. Synthesia has celebrity avatars and one can use one such avatar to spread wrong messages to the masses which can lead to various implications. This is ethically incorrect, as the celebrity is unaware of what his/her avatar is being used for. Another organization that uses deepfake technology to create portfolios is Rosebud. They help small brands with limited resources produce more powerful portfolios of images, featuring more diverse faces. Though it has vast uses and is helpful but only if used in the right way. If this is misused, then it can have huge repercussions. It is unethical to use someone’s face without their consent for any kind of work.** (Simonite, 2020)

**Another such example is the use of deepfake technology by the MBN news channel in South Korea. The regular news anchor Kim Joo-Ha started to go through the day's headlines and the news was normal. Yet this particular bulletin was far from normal, as Kim Joo-Ha wasn't actually on the screen. Instead, she had been replaced by a "deepfake" version of herself - a computer-generated copy that aims to perfectly reflect her voice, gestures, and facial expressions. The viewers were already informed about the same and had mixed reviews. Some were impressed by what technology can do whereas some expressed their concern about how technology and robots will soon replace not only the newscasters but also the producers and directors. This concern is not all wrong. Bringing automation to the labor industry is helpful and can also save the lives of various laborers that do life-threatening jobs. But bringing automation to the entertainment industry might just eat up the jobs of various people. Various people might lose their jobs and hence lose the means of earning a livelihood. Some people might slip into a depression whereas some might even try committing suicide if they do not get any other means of earning. This is morally and ethically incorrect. If utilitarianism is considered, using deepfake for synthetic media purposes is not wrong because it has a lot of benefits over the few repercussions. Also if Kantianism is to be considered, it is not wrong to use deepfake as Kantianism states to keep human emotions aside and to view the benefits. But if Social Contract theory and Virtue ethics are to be considered, human rights and emotions are to be prioritized. Here, we are putting human emotions in hindsight which is ethically incorrect. One more major concern is who owns the videos that are created using deepfakes. The videos that are created using these technologies have no credibility and hence no ownership. There is still a debate going on whether the content creator, the celebrity used for deepfake, or the organization that was used for creating the content. Hence the content becomes vulnerable and might be ethically incorrect. Also, deepfake imposes a threat to society which states ‘we can no longer believe what we see’. This is a huge concern as it targets people’s trust. It might become very easy to deceive people using deepfake in this manner. This might be ethically incorrect in all means.** (Jr, 2021)

 

Fig: (Bateman, 2020)

Deepfake also has the ability to cause financial harms. Deepfakes and synthetic media do not pose a serious threat to the stability of the global financial system or national markets in mature, healthy economies. But they could cause varying degrees of harm to individually targeted people, businesses, government regulators; emerging markets; and developed countries experiencing financial crises. Financial threats from synthetic media appear more diverse than political threats but may in some ways be easier to combat. Some financial harm scenarios resemble classic political disinformation scenarios that seek to sway mass opinion. Other financial scenarios involve the direct targeting of private entities through point-to-point communication. There are 3 malicious techniques to deceive people: deepfake vishing, fabricated private remarks, and synthetic social botnets. Deepfake voice phishing (vishing) uses cloned voices to impersonate trusted individuals over the phone, exploiting victims’ professional or personal relationships. Fabricated private remarks are deepfake clips that falsely depict public figures making damaging comments behind the scenes, challenging victims to refute them. Synthetic social botnets are fake social media accounts made from AI-generated photographs and text, improving upon the stealth and effectiveness of today’s social bots. This poses a greater ethical concern. Here people’s earnings are at risk. The damage that can be caused due to such unethical things has no measure. You can create a fake credit card account (identity theft), you can pressurize people to pay money in the name of the government (imposter scam), etc. Hence here if we apply Kantianism, Utilitarianism, Social Contract Theory, and Virtue ethics, the deed cannot be justified and is morally incorrect and is vulnerable. (Bateman, 2020)

**Imagine a few days before an election, a video of a candidate is released showing them using hate speech, racial slurs, and epithets that undercut their image as being pro-minorities. Imagine a teenager seeing an explicit video of themselves on social media. Imagine a CEO on the road to raise money to take their company public when an audio clip of them stating their fears and anxieties about the product is sent to the investors. These are a few scenarios of the malicious use of AI-generated synthetic media known as deepfakes. Deepfakes can harm individuals, businesses, society, and democracy, and can accelerate the already declining trust in the media. Such ab erosion of trust will promote a culture of factual relativism, unraveling the increasingly strained fabric of democracy and civil society. Additionally, deepfakes can enable the least democratic and authoritarian leaders to thrive as they can leverage the ‘liar’s dividend’, where any inconvenient truth is quickly discounted as ‘fake news’. Creating a false narrative using deepfakes is dangerous and can cause harm, intentional and unintentional, to individuals and society at large. Deepfakes could worsen the global post-truth crisis as they are not just fake but are so realistic that they betray our most innate senses of sight and sound. Deepfakes created to intimidate, humiliate, or blackmail an individual are unambiguously unethical. Pornographic deepfakes can threaten, intimidate, and inflict psychological harm on an individual. It reduces women to sexual objects and torments them, causing emotional distress, reputation harm, abuse, and in some cases even financial or employment loss. There are moral and ethical concerns about how deepfakes can be used to misrepresent political leaders’ reputations posthumously to achieve political and policy motives. Although there are some legal protections to using the voice and face of a deceased person for commercial gain, if the heirs have the legal right to use these features, they can use them for their commercial benefit.** (Jaiman, 2020)

**In conclusion, I would just like to add, deepfakes make it possible to fabricate media, often without consent, and cause psychological harm, political instability, and business disruption. The weaponization of deepfakes can have a massive impact on the economy, personal freedom, and national security. The ethical implications of deepfake are enormous. Deepfake threat models, harm frameworks, ethical AI principles, and commonsense regulations must be developed through partnerships and civil society oversight to promote awareness and encourage advancement and innovation.** (Jaiman, 2020)

# References

*Artificial Intelligence*. (n.d.). Retrieved from builtIn: https://builtin.com/artificial-intelligence

Bateman, J. (2020, July 8). *Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios*. Retrieved from Carnegie Endowment: https://carnegieendowment.org/2020/07/08/deepfakes-and-synthetic-media-in-financial-system-assessing-threat-scenarios-pub-82237

Education, I. C. (2020, June 3). *Artificial Intelligence (AI)*. Retrieved from IBM: https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Jaiman, A. (2020, August 27). *Debating the ethics of deepfakes*. Retrieved from Observer Research Foundation: https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/

Johnson, D. (2021, January 22). *What is a deepfake? Everything you need to know about the AI-powered fake media*. Retrieved from BusinessInsider: https://www.businessinsider.com/what-is-deepfake

Jr, B. D. (2021, March 8). *Deepfake is the future of content creation*. Retrieved from BBC News: https://www.bbc.com/news/business-56278411

Simonite, T. (2020, July 7). *Deepfakes Are Becoming the Hot New Corporate Training Tool*. Retrieved from Wired: https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/

*What Are Deepfakes: Synthetic Media Explained*. (2021, June 15). Retrieved from Respeecher: https://www.respeecher.com/blog/what-are-deepfakes-synthetic-media-explained