

Task 2: Ethics in AI

1. Title of the Case / Scenario

Ethical Concerns of Deepfake Videos in the Age of Generative AI

2. Description of the Scenario

Advanced AI models like Sora, Gemini, and ChatGPT are advancing rapidly. As a result of their advancements, it is becoming harder to tell if the content is real or AI-generated. These systems are able to produce highly realistic images and videos that appear to be real. Hence, those who are unfamiliar with AI may fall for these fabricated videos and be convinced that they are real and thus, cause confusion, misinformation, and harm. An individual's face or voice is used in a deepfake video without the individual's consent raises serious concerns about ethics. Their misuse could lead to damage of reputation, misinformation, emotional and social harm.

3. Ethical Concerns and Stakeholders Involved

Some people abuse deepfakes because their intention is to treat someone in an unethical manner. Deepfakes can be used to slander or defame people, impersonate people, or even greatly influence politics. In essence, deepfakes can cause serious harm to those targeted as well as society in general.

Stakeholders involved in deepfake technology:

- People who have had their faces/voices misused and put onto a deepfake
- Unknowing general public who might be fooled by a deepfake
- AI researchers/companies who developed generative technologies
- Governments, media organizations, etc. who are responsible for regulating what content can be distributed

Ethical values being violated: honesty, transparency, privacy, etc.

4. Role of Bias or Lack of Transparency

- Lack of transparency plays the largest role in this issue. Most deepfake creation algorithms do not mark their creations as artificially generated. As a result, viewers or listeners may never be able to tell whether what they're seeing and hearing is real.

- Additionally, bias does play a role in deepfake creation. For example, some faces, voices, or scenes are disproportionately targeted by deepfakes.
 - Organizations rarely disseminate deepfakes of random people. Instead, they are likely to target certain celebrities, genders, etc for whatever reason.
 - Overall, the lack of marking/accountability with deepfake creation is what makes this technology so dangerous. People can no longer trust whether digital media is authentic.
-

5. Possible AI Governance or Technical Solutions

We can fix these problems by setting up clear rules and using technical safeguards.

- All AI-created videos and images should always have a label on them that says “AI-Generated Content” in plain text or a watermark.
- All created files’ metadata need to have a special digital signature that shows if it was made by AI, and no one can change or remove it.
- Platforms need to use special programs that can always catch and warn users if they see deepfakes.
- Tech companies and governments must work together to make laws and systems for using generative AI in a good way.

The government’s rules and oversight will keep AI in the right lane and will only be used as a tool for creativity and innovation, not for manipulation.

6. Personal Viewpoint

I believe that deepfake technology is the most dangerous aspect of AI. Generative AI can be used creatively in film, education, or art. But it can be even more dangerous if people use it in the wrong way. It can ruin people’s trust in digital media. I think there should be strict rules to make what you see transparent and AI-generated content must always be labeled in some way so that the public is always aware of what they are looking at. If used ethically and with global cooperation, AI can remain a force for good and won’t cross the line to become unethical.