Proceeding of ICSPHS 2026 Symposium:Ethics, Law, and Psychology: Interdisciplinary Reconstruction of Humanities in Medicine
DOI: 10.54254/2753-7048/2026.ZJU30919

Legal Regulation of AI-Generated Video Behaviors

Shujie Zeng

CAS Department of Music, New York University, New York, USA
sz3826@nyu.edu

Abstract. While AI-generated video technology provides convenience to the public, it may also be used by malicious individuals to cause varying degrees of social harm, making legal regulation necessary. This paper first focuses on introducing several common current AI-generated video technologies and their principles. Subsequently, it elaborates on the potential harms of AI-generated videos from four dimensions: personal rights and interests, fraud risks, political chaos, and social stability. Finally, it explores the legal regulation paths for AI-generated video technology from three perspectives—regulatory legislation, regulatory measures, and the responsibilities of various subjects—to prevent the abuse of AI-generated video technology.

Keywords: AI-generated video technology, video forgery, harms, legal regulation

1. Introduction

The recent launched AI-generated video technology Sora 2 had again gained attention on the Internet. It is a further proof of AI having the capability to blur the line between what's real and fake.

However, it has often occurred that even when AI-generated video technology is immature and the results are flawed, people struggle to identify them, or even refuse to do so. They are more inclined to believe the videos are real. This phenomenon carries enormous risks: once more advanced AI technology is used in the future to create videos that infringe on others' rights and interests, the harm will be immense.

Therefore, preventing the potential harms of AI-generated videos and regulating their use is an urgent issue. However, current global legal regulations on AI remain extremely inadequate. While China and Europe have begun to pay attention to this problem and promote relevant legislation, the regulatory framework is far from perfect.

Based on the basic principles and harms of current AI-generated videos, this paper discusses the existing regulation of AI-generated videos, and finally proposes specific paths to improve the regulatory system for AI-generated videos.

2. Principles of AI-generated videos

Today's AI already possesses some of the capabilities required for "intelligent machines" as defined by Roger C. Schank [1]. Among these, communication ability—on which current AI generation technology relies most heavily—has achieved significant development. AI video generation

DOI: 10.54254/2753-7048/2026.ZJU30919

technology has multiple branches, most of which rely on text processing capabilities to convert user input into generated content. Additionally, AI video generation technology embodies the "creativity" defined by Schank.

According to the classification method proposed by YuFeng Huang [2], AI generation technology can be divided into five types: Automatic Story Generation, Image Generation, Deepfake, Speech Generation, and Motion Control. This paper focuses on two of these video generation technologies: Image Generation and Deepfakes. Under the category of Image Generation, there are two sub-types: text-generated images and text-generated videos.

2.1. Text-generated images

Text-generated images use user-provided text as a "prompt" to generate images. AI employs natural language processing models to analyze and process the input prompt. It then retrieves images from existing databases that are relevant to the prompt, synthesizes these images to construct a visual representation, and finally generates high-quality, diverse images through independent models. More advanced AI models not only process prompts and generates images but also retain memory of both the user's prompts and the generated images through advanced Long Short-Term Memory (LSTM) technology. In other words, modern AI can further refine generated images based on user requests without having to recreate them from scratch. For example, if a user requests a photo of a sunset at the beach and the AI generates an image featuring a sunset sky and a crowded beach, the LSTM-enabled AI can remove the people from the beach while preserving the rest of the scene, rather than generating an entirely new image. Beyond memory systems, current AI also features advanced text processing capabilities: users can input not only short prompts but also full scripts. The AI extracts keywords to search for relevant 3D scenes in databases and then uses these scenes to render realistic environments.

2.2. Text-generated videos

The principle of text-generated videos is very similar to that of text-generated images, but it imposes higher requirements on the generation model—since the core task is to directly convert text into video. Video generation models create a matching text-video corpus using public video resources, enabling them to generate diverse videos that meet user requirements. Due to the greater complexity of video generation, AI can typically only produce short videos based on simple descriptions. Many AI-generated videos circulating online are created by editing multiple short AI-generated clips and hiding the edit points to form a longer video. A recent example is the promotional video for Megadeth's new album, which was compiled from multiple AI-generated short clips.

There are other methods for text-generated videos. Many developers build on text-generated images and apply temporal modules [3], training models using both image and video data. After training, the model first generates a low-resolution video to establish the overall structure and movement, then gradually optimizes the image quality. Meanwhile, the temporal module ensures the correct temporal dynamics between consecutive frames. Additionally, model developers may use techniques such as Super-resolution to further enhance video quality.

2.3. Deepfake

Deepfake combines "Deep Learning" and "Fake." It uses AI algorithms and Deep Neural Networks (DNNs), which are trained through deep learning technology. Deepfake technology can

DOI: 10.54254/2753-7048/2026.ZJU30919

automatically fuse and replace elements in a video—for example, swapping one person's face with another's while preserving facial expressions. Furthermore, it can superimpose specified images, audio, or video onto another video [4]. According to Tolosana [5], Deepfake technology is classified into four types: Entire Face Synthesis, Identity Swap, Attribute Manipulation, and Expression Swap.

Entire Face Synthesis focuses on generating images, aiming to create entirely new, non-existent human faces. Developers input a large number of human faces from public databases, and the AI analyzes these faces through deep learning to learn facial features and generate realistic-looking faces.

Identity Swap focuses more on video generation, replacing the face of a person in a specified video with another's. Through deep learning of databases, the algorithm captures the facial features of the original video and replaces them with the target face, while aligning the original and new videos to ensure accurate results.

Attribute Manipulation and Expression Swap operate on similar principles. Attribute Manipulation focuses on altering facial features such as hair color, skin tone, gender, or adding accessories like glasses. Expression Swap, like Identity Swap, replaces one person's expression with another's—but it changes the expression of a person in a video while preserving their facial features (e.g., turning a smile into a sad expression on the same person).

3. Harms of AI-generated videos

As demonstrated, current AI-generated video technology already possesses sophisticated and powerful capabilities. While Alexey Turchin [6] argues that truly dangerous AI capable of triggering global crises will not emerge for 10 to 20 years, AI-generated video—as an already highly advanced tool—is inevitably vulnerable to abuse by malicious individuals to create harmful content. According to Anushi Jayalath [7], one category of AI-related harms is "Security and Malicious Use." As Davis Morris [8] notes, humans are the most harmful element in AI technology. This section explores the harms of AI-generated videos from four perspectives.

3.1. Infringement of personal rights and interests

In the digital age, people commonly share videos and photos online. In the context of AI-generated videos, this means almost anyone can become a victim. One of the most notorious examples of personal rights infringement involves using Deepfake technology to replace faces in pornographic videos with those of others. The most common victims of Deepfake porn are celebrities, but ordinary people are also targeted occasionally. For powerless individuals, such videos can severely damage their careers, lives, reputations, and mental health. Moreover, due to inadequate AI regulation, even seeking police assistance often fails to identify the perpetrator.

In 2022, Irish politician Cara Hunter was celebrating her grandmother's birthday when her phone was flooded with hateful messages from strangers. Many claimed to have watched her "little video."

Hunter found the video in question and discovered it was a pornographic clip featuring a woman who looked remarkably like her. She immediately contacted the police and lawyers for help, but received the same response: their current technology could not identify and trace down the person who created the video.

The spread of this fake video drastically changed Hunter's life. She constantly received online sexual harassment from around the world, and eventually, some people even approached her in person to request sexual services. Even some of her friends and family began to believe the video

DOI: 10.54254/2753-7048/2026.ZJU30919

was real. Explaining such a situation to elderly relatives unfamiliar with modern AI technology proved extremely difficult.

This case illustrates how easily an ordinary person's reputation and life can be destroyed by fake videos—especially when the victim is not a government official, head of state, or wealthy individual.

3.2. Fraud risks

Fraud risks associated with AI-generated videos primarily involve creating fake videos of company executives, celebrities, or other prominent figures to defraud money. For example, generating a fake video of a company CEO instructing a senior manager to transfer a large sum of money.

A recent example occurred in 2023, when a video circulated showing American billionaire Warren Buffett announcing that he would give away expensive Bitcoin for free on his website—a site that was actually a fraudulent cryptocurrency platform. The video was later confirmed to be an AI-generated Deepfake.

3.3. Political chaos

AI-generated videos can also be used to create fake clips of political figures. These videos leverage extensive public speech footage of political figures to produce false content. In addition to damaging the personal reputation of political figures, such videos may spread false national policies or provoke diplomatic conflicts. While such videos are often easy to debunk, the negative public opinion they generate in a short period can cause significant political chaos. Governments are forced to urgently respond to and refute these fake videos, which in itself constitutes a form of political disruption.

In 2022, during the Russia-Ukraine War, a video went viral on the internet, gaining 120,000 views on Twitter in a very short period of time. The video showed Ukrainian President Volodymyr Zelensky sitting solemnly and telling Ukrainian citizens to lay down their weapons and surrender to Russia immediately. If Ukrainian citizens had believed the video was real, the consequences of the war would have been unimaginable. Zelensky quickly posted a video on his social media accounts, informing his people that the clip was fake and that Russia was the one who needed to surrender.

If such disinformation is not debunked promptly, its interference in national internal affairs and diplomacy will undoubtedly be enormous.

3.4. Undermining social stability

AI-generated videos undermine social stability primarily by manipulating public opinion. The victims are often public figures: perpetrators use AI to generate videos of public figures making politically incorrect or incendiary remarks—such as racist comments or other forms of discriminatory speech.

In 2019, two artists, Bill Posters and Daniel Howe, used AI to modify a 2017 video of Mark Zuckerberg speaking. They hired a voice actor to dub the clip, creating a 21-second video in which Zuckerberg appeared to discuss "using data to control humanity."

The video was initially part of the artists' installation art exhibition, but when it was uploaded to Instagram, it quickly went viral—attracting massive views and even coverage by mainstream media. Fortunately, the video was soon debunked. While it was not removed from Instagram, Facebook officially restricted its reach and added a Deepfake label to inform users that it was AI-generated. To

DOI: 10.54254/2753-7048/2026.ZJU30919

some extent, the creators exploited widespread public resentment toward capitalists, aligning with popular disdain for the wealthy to depict Zuckerberg as "evil" in the video. Although the video was debunked promptly, the consequences could have been catastrophic if someone had used an AI-generated video to incite public anger without timely refutation.

4. Legal regulation paths for AI-generated videos

Given that AI-generated videos already have the ability to cause significant harm to individuals and society, regulating AI-generated video behaviors—especially through legal means—is essential. This section discusses the legal regulation paths for AI-generated videos from three perspectives: legislation, supervision, and liability.

4.1. Improving regulatory legislation on AI-generated videos

Currently, the European Union (EU) is undoubtedly at the forefront of AI regulatory legislation. According to the content of the EU AI regulation compiled by Matt O'Shaughnessy [9], the EU AI Act adopts a vertical regulatory approach: it classifies AI algorithms into four risk categories to determine the intensity of regulation. The highest risk category is "unacceptable risk," and AI falling into this category is directly banned in the EU. The second category is "high risk," which requires AI algorithms to comply with specified modifications both before and after entering the market. The third and fourth categories are "limited risk" and "minimal risk": limited-risk AI requires transparency management, while minimal-risk AI only requires developers to conduct self-regulation to a certain extent.

Drawing on the EU's experience, China's legislation on regulating AI-generated videos should focus on two key points: first, determining regulatory measures based on the risk level of AI-generated videos, i.e., adopting a classification and grading principle.

Second, ensuring algorithm transparency. Transparency is an extremely important concept—among the 84 AI regulatory documents compiled by Anna Jobin [10], transparency is the most frequently mentioned term. For harmful AI-generated videos, the key is to trace the creators. Legislation should hold offenders accountable, which requires AI development companies to implement traceability measures for videos and make relevant information transparent. This way, government agencies can obtain sufficient information and evidence to pursue liability in judicial proceedings. Transparency is also crucial in this regulatory approach: Miriam C. Buiten [11] proposes that by making AI algorithm models transparent, it is possible to examine factors (such as training data) that lead AI to generate harmful content at various levels.

Third, protecting victims. The most effective method is to design AI systems to prevent the generation of harmful videos during the development phase. Philipp Hacker [12] suggests intentionally training and testing AI during the training phase to ensure it does not generate harmful content. Transparency also plays a key role in this approach.

Fourth, establishing an informed consent rule. For regulating harmful content, China's regulatory laws should clearly stipulate that AI must obtain personal consent before using personal information (such as photos or voiceprints) for content creation.

4.2. Strengthening supervision of AI-generated videos

Supervision of AI-generated videos should focus on the following aspects: first, establishing an international regulatory body to oversee AI globally. Patricia Gomes Rêgo de Almeida [13] proposes

DOI: 10.54254/2753-7048/2026.ZJU30919

in her paper that the regulatory body should develop review processes, and companies whose AI products and services pass the review will receive a certificate confirming compliance.

Second, reviewing AI-generated video algorithms. Governments around the world should reach international agreements to set unified ethical standards for AI algorithms. Under these standards, AI generation algorithms should be prohibited from producing videos that infringe on human rights, portrait rights, privacy rights, intellectual property rights, or other personal rights and interests. Based on these unified standards, a series of review mechanisms should be established to enable third-party regulatory bodies to conduct more effective reviews and supervision of AI algorithms in accordance with specific processes. On a national level, countries should also adhere to their own bottom lines and requirements—similar to the EU AI Act's AI risk standards and China's AI Law, which requires reviews of individual algorithms. Countries can additionally set standards to determine whether individual AI algorithms are permitted for release within their borders, whether specific functions need to be restricted before release, or whether algorithms need to be adjusted based on national conditions.

Third, cracking down on the creation of illegal AI-generated videos. Transparency is crucial for combating such illegal activities: governments need to formulate relevant laws to ensure that AI-generated video companies can trace users who create illegal content in accordance with the law. Simultaneously, internet real-name authentication should be promoted to ensure that users who generate illegal videos can be held accountable in the real world. Such illegal acts may constitute crimes such as defamation, fabricating and intentionally spreading false information, inciting subversion of state power, or inciting secession. Those who commit serious offenses should be held criminally responsible in accordance with the above charges.

4.3. Strengthening the responsibilities of all subjects

The regulatory requirements mentioned above primarily involve government and regulatory bodies. However, AI regulation is not solely the responsibility of government and regulatory agencies—AI development companies and online platforms also need to participate. Roger Clarke [14] notes that the most effective regulatory model is "Co-Regulation," which involves legislative bodies establishing a regulatory framework while other institutions develop detailed regulatory rules. The ultimate goal of this model is to formulate enforceable guidelines for the market to follow, thereby protecting the public. This process requires the joint efforts and compliance of legislative bodies, regulatory agencies, and development companies.

In addition, all online video and social platforms need to label AI-generated content. It is necessary to review content and use AI video detection systems (such as the one proposed by Peng Zhou [15]) to identify AI-generated videos. Once detected, platforms should use watermarks or additional labels to ensure users are aware that the content is AI-generated. Individual users also bear responsibilities for AI regulation: similar to the social norms governing online speech, users of AI-generated videos should understand the potential harms of the technology and refrain from creating videos that infringe on others' rights and interests. Additionally, when encountering AI-generated videos that violate community rules or laws, users have a responsibility to report them to the platform for removal. Platforms should penalize users who violate community norms.

5. Conclusion

At a time when AI-generated videos are already capable of causing significant harm, there is sufficient theoretical support to formulate laws and establish regulatory bodies to regulate AI-

DOI: 10.54254/2753-7048/2026.ZJU30919

generated videos and even broader AI models. What is needed now is to collectively realize the vision of reasonable AI regulation both domestically and internationally—creating common guidelines, ensuring that countries and companies comply with these rules, formulating corresponding laws for regulation, and holding criminals accountable for using AI-generated video technology to harm others. However, AI regulation must also avoid endangering the trade secrets of development companies, as well as the economic and technological development potential of AI. Transparency—one of the most important aspects of AI regulation—requires the most rigorous oversight, but the formulation and implementation of specific regulations must be carefully planned to ensure that economic and technological development, as well as human rights, are not sacrificed for the sake of regulation.

References

[1] Roger C. Schank, What is AI Anyway? , Cambridge University Press, AI Magazine Volume 8 Number 4, 1987
[2] YuFeng Huang, ShiJuan Lv, Kuo-Kun Tseng, Pin-Jen Tseng, Xin Xie & Regina, Fang-Ying Lin, Recent Advances in Artificial Intelligence for Video Production System, Informa UK Limited, Enterprise Information Systems, 17 Aug 2023
[3] Hessen Bougueff, Mamadou Keita, Wassim Hamidouche, Abdelmalik Taleb-Ahmed, Helena Liz López, Alejandro Martín3, David Camacho3, Abdenour Hadid, Advances in AI-Generated Images and Videos, International Journal of Interactive Multimedia and Artificial Intelligence, 28 November 2024
[4] Marie-Helen Maras, Alex Alexandrou, Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos, SAGE Publications, The International Journal of Evidence & Proof, 28 October 2018
[5] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, Javier Ortega-Garcia, Deepfake and Beyond, Elsevier BV, An International Journal on Multi-Sensor, Multi-Source Information Fusion, December 2020
[6] Alexey Turchin, Assessing the future plausibility of catastrophically dangerous AI, Elsevier, Futures: for the interdisciplinary study of futures, visioning, anticipation and foresight, March 2019
[7] Anushi Jayalath, Nimesha Rodrigo, D A Gunasekera, What are the potential dangers of Artificial Intelligence?, ResearchGate, October 2023
[8] Davis Morris, Magical thinking and the test of humanity: we have seen the danger of AI and it is us, Springer Nature, AI & Society 14 Sep 2023
[9] Matt O'Shaughnessy, Matt Sheehan, Lessons From the World's Two Experiments in AI Governance, Carnegie Endowment for International Peace (CEIP), 14 Feb 2023
[10] Anna Jobin, Marcello lenca, Effy Vayena, Global landscape of AI ethics guidelines, Springer Nature, Nature Machine Intelligence, 02 September 2019
[11] Philipp Hacker, AI Regulation in Europe: From the AI Act to Future Regulatory Challenges, feoma Ajunwa & Jeremias Adams-Prassl (eds), Oxford Handbook of Algorithmic Governance and the Law, Oxford University Press, 6 Oct 2023
[12] Miriam C BUITEN, Towards Intelligent Regulation of Artificial Intelligence, Cambridge University Press, 29 April 2019
[13] Patricia Gomes Rêgo de Almeida, Carlos Denner dos Santos, Josivania Silva Farias, Artificial Intelligence Regulation: a framework for governance, Springer Nature, 21 April 2021
[14] Roger Clarke, Regulatory alternatives for AI, Elsevier, Aug 2019
[15] Peng Zhou, Xintong Han, Vlad I. Morariu, Larry S. Davis, Learning Rich Featuers for Image Manipulation Detection, IEEE Xplore, 16 December 2018