

# The Dilemma of Applying Reputation Rights Norms in the Context of Information Authenticity and Its Institutional Optimization

-- An Empirical Study Based on Deep Synthesis and Text-to-Video Technologies

Fang Du

Law School, Beijing Normal University, Haidian District, Beijing 100091, China

---

**Abstract:** In the digital age, the authenticity of information has become increasingly challenging to verify due to the pervasive influence of generative artificial intelligence technologies, such as deep synthesis and text-to-video. This has led to a proliferation of reputation infringement incidents. In the face of the crisis of information authenticity brought about by new technologies, current regulations governing reputation rights reveal several critical shortcomings, including incomplete identification of infringing parties and inaccurate judgment of infringing acts. To address these problems, firstly, providers of true information should also be included within the scope of reputation infringers and be required to bear a portion of the duty of reasonable verification. Secondly, the scope of assessment for insult and defamation should be expanded to account for cases in which the difficulty in discerning the authenticity of information leads to an initial surge, followed by a decline, in the social evaluation of individuals. Lastly, it is crucial to focus on the interconnectedness of regulations concerning generative artificial intelligence, information dissemination, and reputation rights, ensuring that these areas are harmoniously aligned to effectively protect individual reputation in the digital era.

**Keywords:** Right to Reputation; Information Authenticity; Deep Synthesis Technology; Generative Artificial Intelligence; Technology Governance.

---

## 1. The Impact of Deep Synthesis and Text-to-Video Technology on the Right to Reputation in the Context of Authentic Information

The digital age has witnessed a shift towards fragmented and fluid information sources, accompanied by the emergence of free recombination, novel interactions, and nested dissemination of various types of information [1]. While the abundance of information brings convenience to life, the arbitrary combination or iteration of information poses significant challenges to the authenticity of such information. From simple video and audio splicing to deepfakes that blur the line between real and fake, and further to the independently creative text-to-video technology, such as the Sora model, not only has the public's ability to discern the authenticity of information been weakened, but even authoritative online news media and official websites have come under scrutiny. Incidents of infringement upon the right to portrait, reputation, and privacy have become increasingly frequent. In the early days of the internet, identifying false information was not difficult. For example, forged images struggled to replicate subtle changes in lighting, and forged videos had difficulty capturing the rhythm and pitch of sound. Before the advent of deepfake technology, forged videos were highly identifiable [2]. With the advancement of biometric technology, the use of human biological characteristics to identify, alter, and utilize individual identities has gradually matured. These characteristics can be expressed in specific codes and shared in various locations such as residential communities, libraries, and train stations through technological means [3]. This has been followed by deeper

applications of biometric information, namely deep synthesis technology represented by face swapping and text-to-video technology represented by the Sora model.

Deep synthesis technology operates in a manner akin to the functioning of brain ganglia, utilizing "neural networks" to conduct extensive sample learning. This process creates increasingly precise models, ultimately enabling the realistic imitation or alteration of audio, video, or images [4]. With the iteration of technology, the authenticity of deep-synthesized information has become increasingly difficult to discern, leading to a surge in reputation infringement incidents caused by face-swapping. As early as 2018, a deepfake video of Donald Trump speaking on climate change issues caused a significant uproar. Statements in the video such as "As you know, I have the guts to withdraw from the Paris Climate Agreement" nearly incited public outrage [5]. Indian journalist Rana Ayyub, after advocating for victims, herself became a victim of deepfake technology. This not only severely damaged her reputation but also caused significant psychological trauma, leading to her hospitalization due to palpitations and anxiety [6]. According to a survey of 200 people in the United States, almost everyone was negative about the deep falsification of pornographic videos as a damage to their mental health and reputation [7]. Text-to-video technology has evolved beyond mere imitation to possess "creative" capabilities. The Sora model, for instance, can generate a realistic image of a mammoth based solely on a 67-word description [8]. Currently, text-to-video technology is gradually evolving from collaborative creation to "independent creation." Some scholars believe that it may achieve "mind uploading" at the consciousness level once the technology meets necessary conditions such as high-precision simulation [9]. However, the so-called "inspiration" for these

creations may stem from just a few photos or even a casually written text. The authenticity of the content post-creation is not only difficult to discern, but the hyper-realistic nature of the content can also influence the user's original judgment. Compared to deepfake technology, text-to-video technology poses a greater risk of deception when used to commit reputation infringement. On one hand, text-to-video technology can not only replace the facial features of the infringed party but can also generate new environments that adapt to those facial features. For instance, on April 27, 2024, China's self-developed text-to-video large model, Vidu, was unveiled for the first time at the annual meeting of the Zhongguancun Forum. This model can maintain high consistency of characters and scenes across time and space [10]. On the other hand, text-to-video technology allows creators to inflict damage on reputation rights simply by inputting basic text commands, without the need for illegally processing others' information. For instance, realistic videos can be generated using celebrity videos from the internet or photos posted by friends on social media, combined with fabricated text, making the process highly covert. Moreover, the new text-to-video model, EMO, only requires a voice audio and a single image to produce a video avatar with rich facial expressions and various head poses, presenting an external appearance that is convincing enough to "pass as real" [11].

Overall, deepfake technology represents the pinnacle of information modification techniques, while text-to-video technology exemplifies information creation techniques. Both have significantly advanced the development of information technology in the digital age but have also introduced concerns regarding the authenticity of information. At this stage, deep synthesis technology has evolved to a point where it is difficult for the untrained eye to detect, and text-to-video technology can even operate without the constraints of a "data source." The misuse of these technologies has exacerbated issues related to information authenticity, with reputation infringement being a particularly severe problem. Therefore, while vigorously developing technologies to identify information authenticity, it is imperative to promptly refine existing reputation rights regulations and related interpretations.

## **2. The Application Dilemma of Current Reputation Rights Norms in the Context of Information Authenticity**

Reputation refers to the social evaluation of a civil subject's character, prestige, talent, creditworthiness, and other attributes. Civil subjects enjoy the right to reputation, and no organization or individual may infringe upon another's right to reputation through means such as insult or defamation. However, with the continuous advancement of deep synthesis and text-to-video technologies, the challenges to the right to reputation posed by artificial intelligence are increasingly growing: Firstly, operators can generate content that infringes on reputation with simple instructions, making it difficult to establish infringement even if the video creator is traced. Secondly, due to the unresolved challenge of "distinguishing between true and false information," it is difficult to determine whether the actor has committed acts of insult or defamation. Lastly, the duty of reasonable verification can hardly achieve the purpose of protecting the right to

reputation, necessitating supplementation by other regulations.

### **2.1. The norms governing reputation infringement have failed to cover providers of true information**

During the early development of "deepfake" technology, although the training process relied on unsupervised learning, the collection of source data still required substantial manual effort to procure target materials such as videos, audio recordings, and images of specific individuals. At this stage, the creation of synthetic videos necessitated significant human intervention. If a party objectively engaged in unlawful acts of insult or defamation through such operations, their liability for reputation rights infringement could be readily established. However, complexities arise when a user merely inputs lawfully obtained and publicly available images or videos of the victim into a text-to-video AI system, which then autonomously generates content containing defamatory elements: While the user performed data input operations, the input materials themselves are authentic and publicly accessible, making it difficult to characterize such actions as unlawful under current tort law frameworks. For instance, in a scenario where a defendant provides AI software with learning capabilities, and a user customizes an AI persona closely resembling the plaintiff through experiential training and intentional parameter adjustments, such conduct may constitute infringement—even absent direct content creation by the user [12].

A more legally complex scenario arises when software autonomously commits reputation infringement after processing user data. For instance, the defendant's preventive risk control system automatically labeled the plaintiff—who had no violation history—with tags such as "This account may be abnormal" and "Avoid financial interactions", resulting in the plaintiff's friends mistakenly believing them to be a fraudster, thereby causing reputational harm [13].

### **2.2. The difficulty in information discernment makes it challenging to determine acts of insult and defamation**

Article 1024 of the Civil Code of the People's Republic of China (Hereinafter referred to as the Civil Code), stipulates that no one may infringe upon another's right to reputation through insult, defamation, or other means. Scholars have debated the nature of reputation damage: some argue that it manifests as an unreasonable reduction in the social evaluation of the civil subject [14]; others contend that without a detrimental effect on the social evaluation of the reputation rights holder, no defamation can be established [15]. However, in the current era of information authentication challenges, negative information is merely one manifestation of reduced social evaluation. In other words, even the dissemination of positive but false information can lead to a decline in social evaluation. For example, during a disaster, a celebrity who did not donate any funds might be falsely portrayed through deepfake technology as having donated millions, leading to public praise. Subsequently, when the truth is revealed—that no donation was made—the public may perceive the celebrity as exploiting the disaster for personal gain, resulting in widespread negative criticism and a significant decline in their social evaluation.

This scenario is not a simple case of insult or defamation,

nor does it involve the direct dissemination of negative statements. Instead, it leverages deep synthesis and text-to-video technologies to first elevate the victim to a position of exaggerated praise, only to later "debunk" or criticize them, thereby causing a decline in their social evaluation. Due to the difficulty in quickly verifying information authenticity, the dissemination of such false information can have a more severe and lasting impact on the victim's reputation than direct insults or defamation.

As early as the beginning of the 21st century, scholars have argued that both positive and negative evaluations of others must be based on facts and evidence, with as little emotional involvement as possible [16]. In recent years, some scholars have determined whether an individual has committed defamation or insult by examining factual statements and value judgments [17], with a focus on the truthfulness of the content expressed by the actor. Under this criterion, the elevation of social evaluation due to false information may also result in reputational infringement. If the actor states objective facts, even if it leads to a decrease in the social evaluation of the victim, it does not constitute reputational infringement [18]. Accordingly, if the increase and decrease in the victim's social evaluation are considered as two separate parts, it may lead to contradictions in the evaluation of reputational infringement; if the increase followed by a decrease in the victim's social evaluation is considered as a whole, the boundary between false and true information may become difficult to define.

### **3. Analysis of Changes in the Subject and Conduct of Reputational Infringement in the Context of Information Authenticity**

In the current era of widespread internet technology, ordinary netizens can generate false information with simple operations. As long as the video content is sufficiently eye-catching, even self-media users without a fan base can quickly gain traffic. The influence of videos incorporating deepfake technology is even greater, such as the Barack Obama face-swapping incident, which often garners millions or even hundreds of millions of views. Some argue that the concept of the "actor" in reputational rights norms can no longer be simply interpreted as news media at this stage; self-media and internet users have also exacerbated the complexity and variability of information dissemination models [19]. If the concept of the "actor" in reputational rights norms is expanded to include ordinary netizens, should they also be obligated to perform reasonable review duties? Moreover, text-to-video technology only requires providing publicly available, authentic images of the victim to generate false videos. If a false video generated using real and legal information constitutes reputational infringement, who should be held responsible—the provider of the authentic information or the manager of the generative artificial intelligence?

#### **3.1. An Analysis of the Expansion of Reputation Tort Concepts in the Context of Information Reality**

This paper posits that real information providers utilizing deep synthesis or generative video technologies, irrespective of being individuals or organizations, should be mandated to undertake reasonable content-level auditing obligations.

Furthermore, in instances where reputation infringement arises from false videos generated based on authentic and public content, the party responsible for the infringement remains the original information provider.

#### **3.1.1. Expansion of the Scope of Actors Based on Reasonable Review Obligations**

The subjects bound by reasonable review obligations are generally news media. From the perspective of the Civil Code system, Article 999 stipulates that when news reporting involves public interest and fair use, the protection of the reporting entity takes precedence over the protection of personality rights [20]. Although Article 1025 of the Civil Code states that actors who conduct news reporting or public opinion supervision for the public interest and thereby affect the reputation of others shall not bear civil liability, it still does not provide an explanation for the concept of "actors". Moreover, terms such as "actors" and "news reporting" are inherently related to Article 999 of the Civil Code, making it difficult to broadly interpret "actors" as individuals or organizations using the internet. Some scholars believe that the provisions on defamation by self-media (including individuals and institutions) should apply Articles 1194-1197 of the Civil Code [21]. This view may be over-interpreted: firstly, the scope of self-media is too broad; secondly, from the legislative system perspective, these provisions mainly restrict network service providers, with fewer restrictions on network technology users; thirdly, these provisions do not point to clear infringements, and it is better to directly apply the general provisions of the tort liability section rather than using them to explain defamation. Therefore, although the Civil Code provides detailed provisions on the right to reputation, there are no clauses directly applicable to individuals or organizations other than news media. If ordinary netizens, as self-media users, spread false information, they can only be processed according to the natural person's defamation process. Self-media users neither have the right to public interest defense nor the obligation to verify content. Some scholars, from the victim's perspective, argue that it is unreasonable to apply the general fault liability principle to all news reporting that infringes on the right to reputation. They suggest dividing victims into ordinary citizens and state public officials, applying the no-fault liability principle to the former and the fault liability principle supplemented by legitimate legal reasoning to the latter [22]. This view is also inadvisable, firstly, China's right to reputation norms are from the perspective of the infringing subject, and rashly distinguishing the infringed cannot achieve the self-consistency of the normative system; secondly, the classification of ordinary citizens and state public officials cannot exhaust the special circumstances of the infringed, such as celebrities with a duty of moderate tolerance and minors protected by sensitive information norms. It is evident that neither the scope of the infringing subject nor the infringed should be excessively expanded.

However, with the continuous advancement of generative artificial intelligence technology today, self-media users represented by ordinary netizens can also impact others' reputations through realistic false content. If such cases are handled according to general defamation laws, the rights of the infringed parties may not be effectively protected. Based on this, providers of authentic information who utilize deep synthesis or text-to-video technologies could be regulated as a special category of subjects and included within the "reasonable review obligations" stipulated in Articles 1024

and 1025 of the Civil Code. The reasons for not establishing a separate clause are primarily as follows:

Firstly, expanding the concept of actors within the framework of Articles 1024 and 1025 does not disrupt the existing system of reputation rights norms and does not affect the Civil Code's regulation of news media. Some scholars argue that media infringement has its unique aspects, and the optimal approach is to regulate it as a special type of conduct separately [23]. Under this premise, if authentic information providers were to be listed separately, it would necessitate restating provisions such as "not fabricating or distorting facts," "fulfilling reasonable verification obligations," and "not using insulting language to damage others' reputations." This could easily lead to content repetition. Secondly, generative artificial intelligence technology is not yet widespread, and its impact remains incomparable to that of news media, making it more reasonable to treat it as a supplement to existing regulations. Lastly, imposing the same reasonable verification obligations on individuals as on news media exceeds the capacity of individuals. Therefore, when interpreting the reasonable verification obligations of individuals, the standards could be appropriately lowered, meaning that as long as there is a clear indication of the use of generative AI and the content does not involve insults or defamation, liability can be exempted.

### **3.1.2. Identification of Actors in Reputation Infringement Involving Authentic Information**

Under the perspective of traditional network technology, actors who input a large number of instructions to fabricate false content to insult or defame others meet the constitutive elements of reputation infringement and can be identified as committing reputation infringement. However, with the widespread application and refinement of technologies like Sora, users only need to input authentic and publicly available images or text instructions to generate a large number of videos, which contain many fictional events, activities, or social scenarios. These contents are sufficient to mislead viewers of the videos. Faced with false videos generated from authentic information, applying the constitutive elements theory to determine whether the actor constitutes reputation infringement encounters obstacles: on one hand, the content provided by the actor to the generative artificial intelligence is authentic and legal, making it difficult to deem the act illegal even if instructions like "publish" or "share" are added; on the other hand, although there is a causal relationship between the act of the authentic information provider using legal information to create false videos and the damage to reputation rights, it is also difficult to determine reputation infringement due to the objective authenticity of the information source.

This article posits that when judging the illegality of false videos generated from authentic information, the constitutive elements of reputation infringement should be analyzed according to a certain logic, with the core being the actor's subjective intent. If the actor subjectively intends to insult or defame others, then consideration should be given to whether the information and instructions provided to the generative artificial intelligence can achieve this purpose. Specifically, it manifests as: the actor, knowing that the generative artificial intelligence will generate videos detrimental to others' reputation rights after obtaining this information and these instructions, still actively promotes or allows their generation. This is similar to the "actual malice" principle in the United States, where the actor knowingly uses false statements to

harm others. Once the subjective purpose aligns with the objective behavior, the actor's behavior can be deemed illegal, ultimately constituting reputation infringement. Additionally, the standard for subjective intent can also refer to Article 1 of the "Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in Handling Criminal Cases of Defamation Using Information Networks," which stipulates "fabricating facts to defame others."

In summary, when determining whether videos produced or published by generative artificial intelligence constitute reputation infringement, the subjective intent of the authentic information provider should be considered first, followed by other constitutive elements. Of course, if the authentic information provider bears a reasonable review obligation, then failing to prominently label or review the content of the video may constitute subjective intent.

## **3.2. Analysis of the Criteria for Judging Insults and Defamation in the Context of Authentic Information**

The primary manifestations of deepfake and text-to-video technologies infringing on reputation rights are: placing an individual's information in a fabricated, potentially insulting context, or altering the individual's expressions in a video, thereby inciting public insult or defamation against the individual [24]. These infringing methods can be divided into two categories: first, the actor actively creates an environment that insults or defames others, thereby lowering the social evaluation of the civil subject's character, reputation, talent, credibility, etc.; second, the actor releases false videos that mislead the public about the victim, leading to insults or defamation and ultimately causing a decline in the victim's social evaluation. The difference between the two scenarios lies in the fact that the first constitutes direct harm, where the victim's reputation rights are damaged as soon as the video is released. Of course, insulting or defamatory behavior can only be recognized as infringing on reputation rights if it has caused a certain impact. The impact here mainly refers to the social evaluation of the victim, with the sign of impact being the presence of a third party or disclosure to a third party [25]. Since false information is released on the internet, these conditions are already met, and thus the victim's reputation rights are directly harmed. The second scenario constitutes indirect harm, where the false video released by the actor does not contain content that insults or defames the victim, but the public, in conjunction with the victim's existing social evaluation, draws a negative conclusion, subsequently making negative evaluations that lead to a decline in the victim's social evaluation. The first scenario conforms to the general standards for identifying reputation infringement and is not unique. The second scenario presents a situation not yet addressed by reputation rights norms, where although the actor releases a false video, the content of the false video is a positive evaluation of the victim. As the public may not promptly discern the truth, such false videos can quickly gain traction, elevating the victim's social evaluation. However, once the video is debunked, the victim's social evaluation can plummet sharply, and some victims may even suffer online violence unknowingly, significantly impacting their lives.

### **3.2.1. Adopting Dynamic Continuous Evaluation as the Criterion for Assessing Insults and Defamation**

There are two reasons why reputation rights norms are

difficult to apply to situations where social evaluation first rises and then falls: first, the false information released by the actor triggers positive evaluations. From the perspective of action and result, false information does not directly infringe on others' reputation rights. Second, if the issues reflected in the critical articles written and published by the actor are substantially true, or the content of literary and artistic works is based on real people and events, it does not constitute reputation infringement. This article argues that although the information ultimately released by the actor is true, it may still constitute insult or defamation: on one hand, there is continuity between the decline and rise in the victim's social evaluation, which is the result of dynamic development and should be evaluated together. Specifically, the subject who commits the insult or defamation should be directly related to the decline in the victim's social evaluation and damage to their reputation. This relevance is dynamic because the formation of an individual's social image is inseparable from the participation and evaluation of others during the flow of information [26]. Therefore, participants in the process of changes in the victim's social evaluation may constitute insult or defamation. On the other hand, the judgment of dynamic relevance should not focus on the authenticity of the information but on the purpose of the behavior. That is, as long as the actor has the purpose of insulting or defaming, they may constitute reputation infringement. The act of insult or defamation is necessarily carried out with a specific purpose, such as ruining someone's reputation or bringing oneself into the public eye, which excludes groups that make wrong judgments due to being misled by information. Additionally, the consequences of insulting or defamatory behavior should also be dynamically evaluated, meaning that while paying attention to changes in the victim's social evaluation, one should also examine the damage to the reputation rights of the victim's associates caused by the false video, or the secondary harm brought about by this video.

### **3.2.2. Taking Unfair Commentary as the Core of Evaluating Insults and Defamation**

The criterion for determining acts of insult and defamation is unfair commentary, not positive or negative evaluations. In real life, if internet users make fair comments based on facts and law, it does not constitute reputation infringement. The situation where the victim's social evaluation first rises and then falls differs from fair commentary: although the social evaluation ultimately decreases due to the public's genuine comments, this process is not always in a state of fair commentary but transitions from unfair to fair commentary. For reputation, both fair and unfair commentary can either damage or enhance someone's reputation. For reputation rights, evaluations that constitute reputation infringement must be brought about by unfair commentary; otherwise, they would not carry the characteristics of insult or defamation. Therefore, to determine whether a situation where the victim's social evaluation first rises and then falls constitutes insult or defamation, attention should be paid to whether the reason for the decline in social evaluation involves unfair commentary, rather than the positive or negative nature of the social evaluation.

### **3.2.3. Valuing the Initial and Final States of Social Evaluation**

The impact of negative social evaluations can vary greatly, and not every instance requires accountability. For example, it is difficult to determine that an actor has committed an insult based solely on a few curse words. The same applies to

the internet: it is unreasonable to deny the damage to reputation caused by false videos based solely on a few praises, and there is no basis to conclude that false videos constitute reputation infringement based solely on fluctuations in social evaluation. Acts of insult and defamation must have a certain impact or cause serious consequences to be evaluated as infringing on reputation rights [27]. When a victim's social evaluation decreases due to genuine comments, the reference should be their original social evaluation and the final social evaluation after changes. The enhancement of their reputation during the event's fermentation period can only serve as a reference and not as the final evaluation result.

Overall, whether the situation where an actor creates a false video causing someone's social evaluation to first rise and then fall constitutes insult or defamation cannot be simply determined by whether the evaluations received are positive or negative. It is necessary to analyze the entire reputation infringement incident, identify the essence and final impact of the decline in social evaluation, and then make a comprehensive judgment based on conditions such as "unfair commentary + a certain impact."

## **4. Suggestions for Improving Reputation Rights Norms in the Context of Authentic Information**

Article 18 of China's 2021 "Regulations on the Administration of Public Account Information Services for Internet Users" emphasizes the importance of information authenticity and identifies the dissemination of false information, damage to others' reputation, and misleading the public as illegal. Unfortunately, the regulation does not delve deeper into the protection of reputation rights in the digital age. With the rapid improvement of internet governance capabilities, refining current reputation rights norms is of significant importance for fostering a healthy public opinion space, promoting the reasonable use of generative artificial intelligence technologies, and safeguarding the personality rights of internet users.

### **4.1. Suggestions for Improving the Norms Regarding Actors in Reputation Rights Regulations**

This article suggests that the norms related to actors should be improved from both the legal provisions and legal interpretation perspectives:

Firstly, the refinement of specific provisions in the Civil Code. Authentic information providers should be included within the scope of actors in reputation rights regulations as a special category. Specifically: authentic information providers have the obligation to label the content they provide and bear reasonable verification obligations. This provision could be added to the Civil Code by inserting a new clause between paragraphs 2 and 3 of Article 1025, stating: "Authentic information providers who use generative artificial intelligence to create information without prominent labeling or who fail to fulfill verification obligations shall bear responsibility." The specific requirements for prominent labeling can refer to Article 17 of the "Chinese Internet Information Service Deep Synthesis Management Regulations"; the main body of authentic information providers should, based on the previous explanation, refer to individuals or organizations other than news media that

provide authentic and legal information; the tool part, "generative artificial intelligence," should be interpreted according to Article 23 of the "Chinese Internet Information Service Deep Synthesis Management Regulations," which includes a series of network information technologies, including text-to-video models.

Secondly, the improvement of the interpretation of reasonable verification obligations. As mentioned above, it is unreasonable to expect ordinary netizens to verify content generated by artificial intelligence according to the standards of news media. For ordinary netizens and self-media users among authentic information providers, overly stringent verification obligations should not be imposed. It is somewhat reasonable for them to use eye-catching, moderately exaggerated expressions to attract traffic or showcase their views. This article believes that the scope of reasonable verification obligations for authentic information providers should be limited to not adding false information directly or indirectly related to others in videos; and not creating content that insults, defames others, or leads viewers to insult or defame others. This approach neither overly restricts the creative ideas of video producers nor violates the provisions of Article 1026 of the Civil Code regarding verification capabilities and costs.

## **4.2. Suggestions for Improving the Criteria for Insult and Defamation Determining**

Based on the previous discussion, the creation of fake videos that initially elevate and then subsequently lower an individual's social standing can constitute insult and defamation. However, from the victim's perspective, the theory of constitutive elements still requires refinement. The reason lies in the fact that although the victim's social reputation is diminished, they gain widespread attention. In other words, if the victim is aware of or consents to the creation of a fake video that elevates their social standing, they should bear the risk of a subsequent decline in social reputation when the truth is revealed. Similarly, if the victim derives certain benefits from the elevated social standing, it can be presumed that the victim was aware of or consented to the release of the fake video. In these cases, the infringer may argue that no reputational harm has been committed.

### **4.2.1. Incorporating the Theory of Informed Consent into the Criteria for Insult and Defamation Determining**

According to Article 7, Paragraph 3 of the Interim Measures for the Management of Generative Artificial Intelligence Services in China, providers of generative artificial intelligence services must comply with the following regulations: when personal information is involved, they must obtain the individual's consent or meet other circumstances stipulated by laws and administrative regulations. By slightly modifying this provision, the existing criteria for insult and defamation determining can be improved. Specifically, there are two improvement options:

Option One: Amend Article 1024 of the Civil Code by adding the phrase "except with the informed consent of the individual" at the end. The advantage of this approach is that the addition is made after the provisions on insult and defamation, creating a clear connection with the determination of such acts and avoiding the risk of over-interpretation. However, this option has drawbacks, as it detaches from the regulatory framework for generative artificial intelligence, allowing any infringer to claim

exemption from liability based on the theory of informed consent, which is unreasonable.

Option Two: Amend Article 1025 of the Civil Code by adding a new paragraph after Paragraph 3, stating: "Using generative artificial intelligence technology to insult or defame others without their informed consent." The advantage of this amendment is that it explicitly regulates generative artificial intelligence and connects with the provisions on insult in Paragraph 3. If the regulations on the behavior of the infringer can also be improved, this would form a preliminary specialized regulation on generative artificial intelligence within the Civil Code.

### **4.2.2. Using the Victim's Gain as a Criterion for Insult and Defamation Evaluating**

Here, "gain" refers to the attention, rewards, or other benefits the victim receives during the process of their social reputation rising and then falling. Given the difficulty in promptly identifying deepfakes and text-to-video technologies, using the victim's gain as an evaluation criterion can more efficiently determine whether the perpetrator has committed insult or defamation. Specifically:

**Active Pursuit of Gain:** If the victim, based on the reputational damage incident, actively seeks attention, solicits rewards, or obtains other benefits on self-media or other online platforms, it can be determined that they have gained from the situation, and the perpetrator does not constitute insult or defamation.

**Passive Allowance of Reputational Infringement:** If the victim, knowing or should have known that their social reputation has been elevated or diminished by a video, does not actively seek help from internet platforms or relevant authorities to request the removal of the video, elimination of the impact, or restoration of their reputation, but instead allows the content to continue spreading, resulting in a certain impact or serious consequences, the perpetrator cannot be directly deemed to have committed insult or defamation. In this case, the victim also bears some fault, and a comprehensive judgment should be made based on factors such as the extent of the victim's passivity, the ultimate impact of the behavior, whether the perpetrator took steps to mitigate the damage, and whether the victim gained from the situation.

Additionally, the theory of informed consent can also be considered when evaluating whether the perpetrator has gained from the situation. For example, if the perpetrator damages the victim's reputation by releasing false information, but the victim believes that such "negative attention" could generate traffic and is willing to sacrifice their reputation for financial gain, this should be taken into account [28].

## **5. Conclusion**

Generative AI advancements, like text-to-video models, are transforming the digital age by evolving AI from a tool to a "creator." However, technology is a double-edged sword. While generative AI technologies benefit society, they are also susceptible to misuse by infringers. As a result, a plethora of fake images and videos, indistinguishable to the naked eye, have quietly emerged on the internet, spreading alongside genuine information across various media platforms and causing severe reputational harm.

Nevertheless, the Civil Code's protection of reputation rights remains confined to traditional news media and natural persons, and has yet to extend to areas such as self-media and artificial intelligence. Against this backdrop, establishing a

modern reputation rights protection system—rooted in the norms of the Civil Code, centered on generative AI regulations, and safeguarded by information authenticity standards—is not only an inevitable requirement for the rational use of internet technologies but also an urgent necessity for protecting the legitimate rights and interests of netizens.

## References

- [1] Pan Ji, From Reflecting Reality to Symbolic Framing: Assessing Information Credibility in the Networked Environment, *Modern Communication* (Journal of Communication University of China), no. 11 (2018): 86.
- [2] Wang Lusheng, On the Integrated Regulation of “Deepfake” Intelligent Technology, *Oriental Law*, no. 6 (2019): 60.
- [3] Jiao Yanling, The Definition of Personal Biometric Information, *Journal of Chongqing University (Social Science Edition)*, no. 3 (2023): 201.
- [4] Cai Shilin, The Technological Logic and Legal Reform of “Deepfakes”, *Political Science and Law Review*, no. 3 (2020): 132.
- [5] Oscar Schwartz, Technology can make it look as if anyone has said or done anything. Is it the next wave of (mis) information warfare? (Mon 12 Nov 2018), Available: <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>.
- [6] India Today Web Desk, I was vomiting: Journalist Rana Ayyub reveals horrifying account of deepfake porn plot,(Nov 21, 2018), Available: <https://www.indiatoday.in/trending-news/story/journalist-rana-ayyub-deepfake-porn-1393423-2018-11-21>.
- [7] Kugler, M. B., Pace, Carly. (2021). Deepfake privacy: attitudes and regulation. *Northwestern University Law Review*, 116(3), 611.
- [8] Digitaltrends Web, OpenAI’s new AI-made videos are blowing people’s minds, (February 15, 2024), Available: <https://www.digitaltrends.com/computing/openai-sora-text-to-video-generator/>.
- [9] Shi Anbin, Zheng En, Entering “Fusion Reality”: The Reshaping of the News Media Landscape by Text-to-Video Technology, *Media Observation*, no. 4 (2024): 31.
- [10] Xinhua News Agency, Directly Generating 16-Second HD Video: China’s Self-Developed Video Large Model Released in Beijing, available: <http://www.xinhuanet.com/20240427/96747214b15b4138a2709f9f65cf503d/c.html>.
- [11] EMO, Emote Portrait Alive - Generating Expressive Portrait Videos with Audio 2 Video Diffusion Model under Weak Conditions, available: <https://humanaigc.github.io/emote-portrait-alive/>.
- [12] Beijing Internet Court Civil Judgment No. 9526 (2020).
- [13] Beijing Internet Court Civil Judgment No. 11223 (2023).
- [14] Sun Lianbin, Exploration of the Paradigm of Tort Liability Objects in the Era of Self-Media, *People’s Forum·Academic Frontiers*, 2020, pp. 122-123.
- [15] Wang Yuting, Qin Qianhong, Public Interest Defense in Defamation Cases: Special Exemptions for News Reporting and Public Opinion Supervision, *Hebei Law Science*, 2022, pp. 180.
- [16] Zhu Hongbing, On the Right to Reputation of Citizens, *Journal of South-Central University for Nationalities (Humanities and Social Sciences Edition)*, No. 2, 2003, p. 225.
- [17] Wang Zejian, Characteristics and Interpretation of Chinese Civil Law, *Legal Application*, No. 13, 2020, p. 12.
- [18] Zhu Xiaofeng, On the Elements and Interest Balancing in the Determination of Personality Rights Infringement Liability, *Law Review*, No. 1, 2024, p. 109.
- [19] Wang Liming, The Right to Reply in the Era of Self-Media, *Oriental Law*, No. 4, 2023, p. 71.
- [20] Liu Wenjie, Exploration and Innovation of the Civil Code in Defenses against News Infringement, *Journalist*, No. 9, 2020, p. 64.
- [21] Xuan Gang, Yan Haiyan, Chaos and Regulation: An Empirical Analysis of Defamation in Self-Media Communication, *Hubei Social Sciences*, No. 7, 2022, pp. 123, 137.
- [22] Yin Jun, Research on the Liability Principles of Japanese News Reporting Infringing on the Right to Reputation, *Journalist*, No. 11, 2015, p. 89.
- [23] Wang Ranran, Wen Xueguo, Defenses Against Infringement of Personal Information Disclosure in the Self-Media Era—With Reference to the “News Value Test” in the United States, *Henan Social Sciences*, No. 11, 2020, p. 15.
- [24] Deng Jianpeng, Zhao Zhisong, Risks and Three-Dimensional Regulation of Text-to-Video Artificial Intelligence: A Perspective from Sora, *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)*, No. 6, 2024, p. 94.
- [25] Wang Liming, Cheng Xiao, Zhu Hu, Interpretation of the Personality Rights Section of the Civil Code of the People’s Republic of China, *China Legal Publishing House*, 2020, pp. 388-390.
- [26] Lu Qing, Identity Construction and Legal Protection in the Digital Age: Reflections Centered on Personal Information Protection, *Legal Studies*, No. 5, 2021, p. 13.
- [27] Yang Lixin, Controversial Issues in Civil and Commercial Law Theory—Compensation for Moral Damages, *China Renmin University Press*, 2004, p. 286.
- [28] Luo Xiang, Zhang Huimin, Homogenized Evaluation and Selective Application of Insult/Defamation Crimes in Cyberspace Governance, *Jianghuai Tribune* No. 5 (2023), p. 124.