Available Online: 20 November 2025 DOI: 10.54254/2753-7102/2025.29861

# Hallucinated harm: legal liability for AI-generated false content in mass media

*Kaiyan Shen*

University of California, Berkeley, USA

chloeshen1210@163.com

**Abstract.** As Large Language Models (LLMs) become increasingly embedded in news production, their tendency to generate “hallucinated” content—fabricated or misleading information presented as fact—raises serious legal concerns. This paper examines the implications of such content through the lens of both copyright and personality rights, focusing on civil defamation, privacy infringement, and unauthorized reproduction of third-party materials. Using a comparative doctrinal methodology, it analyzes regulatory and tort law frameworks in China and the United States, with particular attention to the heightened standards for public figure defamation in U.S. law and the broader scope of reputational protection under Chinese civil law. By bridging the intersection of copyright and personality rights, this study offers a novel perspective on the legal classification and liability of AI-generated news. It further proposes actionable compliance strategies for media organizations and generative AI providers, including content review mechanisms and attribution standards. Finally, the paper reflects on future governance trends—especially the tension between innovation and accountability—as jurisdictions worldwide grapple with the social and legal consequences of hallucinated media content.

**Keywords:** generative artificial intelligence, AI hallucinations, copyright law, personality rights, tort liability

## 1. Introduction

The rise of generative artificial intelligence (AIGC), especially large language models (LLMs) like ChatGPT and Claude, has fundamentally reshaped the way news is produced and disseminated. From automated headline generation to full-length reporting, AI is increasingly integrated into the workflows of major news organizations. The New York Times reportedly tested OpenAI’s models to summarize business content; China’s Xinhua News Agency has experimented with AI anchors and content templating; and the BBC has launched internal tools to automate climate reporting and election coverage. While these developments promise greater efficiency and scale, they also give rise to a troubling phenomenon: AI hallucinations. This term describes instances where LLMs generate factually incorrect or entirely fabricated content that appears syntactically plausible. In a media context, hallucinations can blur the line between truth and fiction, causing potential harm to individuals and institutions. This paper explores the legal risks posed by hallucinated content in AI-generated journalism, with a specific focus on copyright infringement, violations of personality rights (including defamation and privacy breaches), and broader civil liability for both developers and news organizations. Drawing on a comparative doctrinal analysis of Chinese and U.S. legal frameworks, the study investigates how each jurisdiction defines and regulates reputational harm, authorship, and the limits of intermediary liability. While U.S. courts rely heavily on doctrines such as actual malice to determine liability in defamation cases, China’s civil law tradition provides broader protection of reputation and privacy under the Civil Code. The article also examines how hallucinated outputs challenge the boundaries of copyright law, particularly in attributing authorship and determining infringement. By synthesizing doctrinal insights and real-world cases, this paper contributes to emerging scholarship at the intersection of AI, media law, and comparative regulation. It further offers practical recommendations for compliance and risk mitigation as generative AI becomes embedded in mainstream journalism.

## 2. Copyright ownership of AI-generated news content

Understanding the copyright status of AI-generated content helps clarify who is legally responsible when such content causes harm. If no human authorship exists, it becomes harder to assign liability for defamation, privacy breaches, or copyright violations in AI-assisted journalism. But when journalists significantly shape AI output, they may be both entitled to copyright and accountable for misuse. Copyright analysis, then, offers a useful framework for tracing authorship, assessing control, and allocating responsibility in disputes involving reputational or legal harm from AI-written news.

## 2.1. Human authorship as a legal precondition

Copyright protection is fundamentally premised on the presence of human authorship. Most jurisdictions—including the United States, European Union, and China—maintain that only works created through original intellectual effort by natural persons qualify for copyright. AI systems, by contrast, do not possess consciousness, intention, or creative judgment. They operate based on statistical prediction and algorithmic execution, guided by prompts and training data defined by their programmers. Consequently, where there is no substantial human creative input, the output of an AI system is generally ineligible for copyright protection. In the United States, the U.S. Copyright Office has repeatedly affirmed this principle. Under its current policy, applicants must disclose which portions of a work were generated by AI and demonstrate a "meaningful human authorship" contribution to register the work [1]. The recently released draft Copyright Registration Guidance explicitly requires creators to identify the nature and extent of human involvement in any AI-assisted submission. Similarly, under EU law—grounded in the InfoSoc Directive and the Berne Convention—authorship remains tied to human creators [2, 3]. Machines, as non-human agents, cannot hold or originate copyright under the current doctrine. These frameworks align in viewing originality and authorship as legal concepts that require individual, human creative input.

## 2.2. Human-AI collaboration: conditional protection

However, when humans engage with AI systems in a way that reflects their personal judgment or expressive intent, some copyright protection may be available. If the output reflects sufficient human creativity—even if the AI was a tool in the process—courts or copyright offices may find valid authorship. This could include scenarios where the human user:

- Crafts detailed and expressive prompts that shape AI output in a unique direction
- Edits, curates, or reframes the generated material with editorial discretion
- Combines various generated components into a distinctive, integrated composition

In such cases, copyright may vest in the human contributor or, in institutional settings, their employer. However, determining whether the level of human involvement reaches the threshold of authorship is a fact-sensitive inquiry that must be assessed case by case [4].

## 2.3. Commercial use and legal risk management

When AI-generated material lacks sufficient human authorship, it is typically deemed to fall into the public domain, meaning it cannot be exclusively owned or controlled and may be freely reproduced, reused, or modified by anyone. That said, this legal status does not eliminate risk. Several practical dangers remain:

- Unintentional infringement may occur if AI replicates copyrighted materials embedded in its training data
- Platform-imposed restrictions, such as those found in OpenAI or Stability AI's terms of use, may limit how generated content is distributed or monetized

Other legal liabilities, including violations of trademark rights, likeness or image rights, or unauthorized use of personal data, may still attach depending on the content involved [5].

## 2.4. Comparative analysis of legal trends in mainland China and the U.S.

While both China and the United States deny copyright to AI-generated content that lacks human authorship, their legal frameworks reflect distinct underlying philosophies. In the U.S., courts have reinforced a long-standing precedent that copyright protection requires human creativity, as recently confirmed in Thaler v. Perlmutter, where a fully machine-generated image was deemed ineligible for copyright [6]. The U.S. Copyright Office further mandates disclosure of any AI-assisted elements in registration applications [7]. In China, the Copyright Law does not expressly address AI authorship, but government agencies have been reluctant to grant protection to purely machine-generated works. Instead, regulators emphasize the need for demonstrable human creativity in the production process [8]. Though Chinese courts have yet to issue definitive rulings, ongoing policy consultations suggest a converging view: human involvement remains the bedrock of legal protection, even in a rapidly evolving technological landscape [9].

## 3. Civil liability for infringement by AI-generated content

Civil liability for AI-generated content arises when false, harmful, or privacy-violating material is published and causes damage to individuals. If human users or platforms negligently deploy AI without proper oversight, they may be held responsible under tort or personality rights law. Liability depends on human involvement, foreseeability of harm, and regulatory compliance.

### 3.1. Personality rights infringement under Chinese law

China's Civil Code provides robust and detailed protections for personality rights, including the rights to reputation and privacy, codified in Articles 990 to 1039 [10]. These provisions emphasize that individual dignity and autonomy should be safeguarded, particularly in the digital era, where exposure through AI-generated media content can result in reputational or emotional harm. The emergence of generative AI in journalism and media production has introduced new risks, particularly through the phenomenon of AI hallucinations—the generation of false or misleading content presented as fact. When such outputs refer to real individuals and falsely associate them with crimes, misconduct, or private circumstances, the resulting harm may constitute a violation of their legally protected personality rights [11]. According to Article 1024, the right to reputation prohibits defamation and wrongful exposure that diminishes one's social standing, while Article 1032 affirms an individual's right to control personal information, private space, and private life [10]. When a journalist, knowingly or negligently, uses AI tools to produce defamatory or privacy-invading content and subsequently publishes it, they may bear primary civil liability for the resulting infringement. Platforms, too, may face joint liability if they fail to moderate or respond to harmful content. Under Article 1195, internet service providers that are aware of infringing acts and fail to take appropriate action can be held jointly liable with users [12]. In rare but growing cases, plaintiffs argue that AI developers themselves should bear responsibility—particularly when systemic flaws or insufficient safeguards result in foreseeable harm. Although the Civil Code does not yet expressly define such "algorithmic negligence," courts may analogize from product liability and tort law principles found in Articles 1165–1176 [13]. Ultimately, as AI continues to reshape media production, Chinese civil law offers a flexible but evolving framework for protecting individual dignity in the face of synthetic, potentially harmful outputs.

### 3.2. Copyright infringement from AI-generated journalism

One growing legal risk in AI-assisted journalism is the potential for copyright infringement. Generative models may produce content—such as text, images, or audio—that closely mimics or directly reproduces portions of existing works without authorization. Since most models are trained on large, publicly available datasets that may include copyrighted material, output may inadvertently "remix" protected content [14]. Journalists who publish such material, even unintentionally, may still be held liable for unauthorized use of another's work under China's Copyright Law (Articles 10 and 48), which recognizes both literal and substantial similarity as infringement thresholds [15]. Civil remedies may include takedown orders, monetary damages, and public apologies. When used for commercial purposes—such as monetized articles or sponsored media content—the penalties can be significantly increased. This places a growing compliance burden on content creators to vet AI outputs and ensure originality before publication.

### 3.3. False light and public disclosure under U.S. law

Under U.S. tort law, individuals harmed by AI-generated falsehoods may seek redress through defamation, false light, or public disclosure of private facts. For defamation, plaintiffs must prove a false statement was published, caused reputational harm, and was made with fault—negligence for private figures, "actual malice" for public ones [16]. This becomes complex with AI: who is the "speaker"? The end-user, the platform, or the model's developer? False light offers an alternative where a statement isn't strictly defamatory but is misleading and offensive, such as placing someone at a protest they never attended [17]. Public disclosure applies if AI output reveals sensitive personal details unrelated to public interest [18]. However, First Amendment protections pose major obstacles, especially in matters of public concern. Courts often hesitate to expand liability where speech rights are at stake. Moreover, Section 230 of the Communications Decency Act shields platforms from liability for third-party content unless they co-create or materially contribute to its illegality [19]. Still, ongoing debate surrounds whether algorithmic amplification of harmful AI content might someday pierce that immunity.

## 4. Criminal liability from AI-generated content

While most legal risks from AI-assisted journalism fall under civil law, criminal liability may arise when AI-generated content causes serious reputational or public harm. In China, Article 246 of the Criminal Law penalizes defamation that "seriously undermines social order," particularly if false information is deliberately fabricated and widely disseminated [20]. If a journalist knowingly uses AI to publish defamatory materials—such as fabricating accusations of crime or corruption—the conduct could be prosecuted as criminal defamation. Courts have emphasized intent and actual harm to public order as thresholds for criminalization [21]. Further, criminal negligence may apply in cases where developers or platforms fail to prevent foreseeable harms caused by flawed algorithms. Under Article 15 of the Criminal Law, negligence arises from a failure to anticipate a risk that ought to have been foreseen. If news agencies or AI vendors deploy systems with inadequate safeguards—e.g., unfiltered hallucination-prone outputs—they may be liable where the harm reaches a criminal threshold [22]. In the United States, criminal liability for AI misuse is narrower but evolving. Although defamation is generally a civil tort in the U.S., criminal charges may arise under statutes involving fraud, identity theft, or digital impersonation. For instance, political deepfakes—false AI-generated voice or video used to manipulate elections—could trigger criminal sanctions under state deception or cybercrime laws [23]. Prosecutors may also invoke conspiracy or aiding-and-abetting doctrines where multiple actors coordinate the malicious use of AI tools for harm [24]. Thus, while AI itself lacks intent, those who deploy it recklessly or maliciously—especially in media contexts—may face criminal exposure depending on the content, intent, and resulting harm. As AI adoption expands in journalism, both Chinese and American legal systems are beginning to apply existing doctrines to this novel domain.

## 5. Comparative regulatory approaches and doctrinal analysis

### 5.1. China: personality rights and platform liability

As generative AI becomes integrated into media and journalism, hallucinated outputs—fabricated or misleading information presented as fact—have triggered significant concerns around reputation and privacy rights. China and the United States offer starkly different responses rooted in their legal traditions. Under Chinese law, personality rights are protected as a distinct legal category in the Civil Code, with Articles 1024–1039 governing reputation and privacy. If AI-generated content falsely links an individual to criminal behavior or moral misconduct, the publisher can be held liable—even if the content came from an automated system [25]. This reflects a dignity-centered approach, where objective harm to social evaluation is enough to constitute infringement. Notably, Chinese courts have begun recognizing that AI systems, due to their lack of human reasoning and moral judgment, may misinterpret or hallucinate insults, sarcasm, or socially sensitive concepts. For example, terms rooted in internet slang or cultural nuance may be inappropriately applied to public figures, resulting in reputational harm. In such cases, courts tend to examine whether the user or publisher should have foreseen the harm, and whether sufficient human review was conducted before dissemination [26]. China also imposes platform liability under Article 1195, which holds that internet service providers may be jointly liable if they fail to act upon known infringing content. If a platform hosts AI-generated news or commentary that damages someone's reputation, and refuses to remove it after notice, it may bear legal consequences—regardless of whether it created the content itself [27]. This expands liability along the AI content chain: from individual user to platform to potentially the model provider.

### 5.2. United States: safe harbor and the section 230 debate

By contrast, in the United States, courts emphasize free speech protections under the First Amendment, especially for matters of public concern. The defamation doctrine, as shaped by New York Times Co. v. Sullivan, requires public figures to prove actual malice—that the publisher knew the statement was false or acted with reckless disregard for the truth [28]. This standard makes it particularly difficult for plaintiffs to prevail in suits involving hallucinated AI content unless human actors clearly manipulated or endorsed the false information. Moreover, Section 230 of the Communications Decency Act offers strong immunity to platforms for third-party content. Courts have generally interpreted this to include AI-generated material hosted by platforms, unless the platform itself materially contributes to the illegality [29]. Thus, while Chinese law emphasizes remedial responsibility and preventive obligations, U.S. law favors speech freedom and platform neutrality, even when AI amplifies harm.

### 5.3. Governance priorities and liability pathways: a comparative table

**Table 1.** Legal dimensions of AI output regulation: China and the United States

| Dimension | China | United States |
|---|---|---|
| Regulatory Orientatio n | Prioritizes personal dignity and social order. AI output is subject to the same legal standards as human speech. | Prioritizes freedom of speech and press autonomy, especially in matters of public concern. |
| Liability Attribution | Legal responsibility can flow from AI user → media platform → AI developer. Emphasis on the duty of care and foreseeability of harm. | Liability usually stops at the user; platforms are largely protected under Section 230, and developers are rarely liable. |
| Remedies for Victims | Include injunctive relief, apology, deletion of content, and damages. Courts may order platforms to act upon notification. | Civil suits are possible, but harder to win. Courts often deny relief unless actual malice or clear negligence is shown. |
| Emerging Legal Trends | Draft policies suggest exploring algorithmic transparency, platform content moderation obligations, and traceable AI output logs [30]. | Legal scholars debate amending Section 230 or creating "algorithmic accountability" standards for generative AI platforms [31]. |

In short, China adopts a victim-centered model focused on preventing and correcting reputational or privacy harm caused by AI-generated content as shown in Table 1. Even when platforms are not the original publishers, they may face joint liability if they fail to remove unlawful hallucinated content. Chinese courts emphasize the predictability and social impact of such outputs, and routinely impose remedies even without proof of malice. The United States, by contrast, applies a speech-centered model, prioritizing the preservation of expressive freedom. Victims of AI hallucinations face greater legal hurdles—especially public figures, who must show "actual malice" to succeed in defamation claims. Platforms remain largely shielded under Section 230, even when hosting harmful AI content, unless they are found to have co-developed it. These divergent approaches reflect deep philosophical differences: China's emphasis on social harmony and moral accountability, versus America's emphasis on open discourse and media independence. As generative AI becomes more embedded in news ecosystems, these legal frameworks will likely continue to evolve in distinct directions.

## 6. AI-specific regulation prospects

China and the United States are both accelerating efforts to regulate AI, but their approaches remain distinct. China has adopted a top-down regulatory model, exemplified by the Generative AI Measures, which require algorithm registration, content labeling, and risk assessments [32]. These rules reflect China's emphasis on social stability and proactive oversight. In contrast, the U.S. relies more on sectoral and reactive mechanisms. While there is no comprehensive federal AI law, recent proposals—including the NO FAKES Act—seek to address deepfakes and AI-generated impersonation in entertainment and political speech [33]. States like California and New York are also experimenting with legislation on algorithmic accountability. Given these diverging legal cultures, a unified international framework is unlikely in the near term. A practical path forward may lie in a hybrid model combining tort liability, criminal sanctions for malicious misuse, and administrative supervision to balance innovation with harm prevention.

## 7. Potential compliance for AIGC in journalism

Writing that an AI-generated story is from a human journalist and releasing it without supervision can result in a negligent publication or invasion of privacy claim, or even intentional defamation. China's Civil Code places an obligation of care on people who post defamatory information, and criminal sanctions may follow if the publication was intended to cause serious harm. In the U.S., while criminal defamation is rare, civil claims may arise from reckless or negligent publication. Courts scrutinize whether journalists acted with actual malice, particularly regarding public figures. News organizations, too, may face vicarious or substitute liability where editorial control, economic benefit, or failure to screen content is demonstrated. To mitigate legal exposure, media outlets should:

- Require human editorial review of all AI-generated content
- Mandate disclosure of AI involvement in published materials
- Audit outputs for legal and factual accuracy
- Train journalists on responsible AIGC use
- Vet third-party tools for compliance with IP and privacy standards

In balancing speed with integrity, media institutions must prioritize due diligence when adopting generative tools into newsroom workflows.

## 8. Conclusion

As generative AI tools like ChatGPT and Claude become deeply integrated into journalism, they bring both opportunity and risk. While enhancing efficiency and scalability, these technologies introduce legal complexities that current frameworks in both China and the United States are still struggling to address. This paper has examined the copyright, personality rights, and civil liability implications of AI-generated journalism—particularly in cases of hallucinated or defamatory content. The comparative analysis reveals that China adopts a dignity-centered model emphasizing platform responsibility and reputational safeguards, whereas the U.S. remains rooted in free speech doctrines and intermediary immunity. These differences are not merely doctrinal—they reflect distinct cultural values and governance priorities. As AI journalism advances, both legal systems must find new tools to allocate liability, ensure compliance, and protect individual rights without stifling innovation. Moving forward, a hybrid regulatory model that combines tort remedies, criminal enforcement, and administrative oversight appears most promising. Meanwhile, media organizations must proactively implement internal safeguards—from editorial review and AI-disclosure protocols to risk audits and training—to reduce legal exposure. Ultimately, the integration of AIGC into journalism demands not only legal reform but also ethical and institutional transformation.

## References

[1] U.S. Copyright Office. (2024, February). Copyright registration guidance: Works containing material generated by artificial intelligence (Draft policy statement). https: //www.copyright.gov/policy/artificial-intelligence/

[2] European Parliament. (2001, June 22). Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society. *Official Journal of the European Union*, L 167, 10. https: //eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32001L0029

[3] World Intellectual Property Organization. (1971, July 24). Berne Convention for the Protection of Literary and Artistic Works (Paris Act), Art. 2. https: //www.wipo.int/treaties/en/ip/berne/

[4] U.S. Copyright Office. (2023, February 21). Zarya of the Dawn decision letter. https: //www.copyright.gov/docs/zarya-of-the-dawn.pdf

[5] Samuelson, P. (2023). Can copyright law handle generative AI? *Columbia Journal of Law & the Arts*, 46(2), 195–221.

[6] U.S. District Court for the District of Columbia. (2023, August 18).Thaler v. Perlmutter, 1: 22-cv-01564.

[7] U.S. Copyright Office. (2024, February). Copyright registration guidance: Works containing material generated by artificial intelligence (Draft policy statement). https: //www.copyright.gov/policy/artificial-intelligence/

[8] National Copyright Administration of China. (2023, November). Copyright protection and AI-generated content: Expert consultation report. http: //www.ncac.gov.cn

[9] Li, Y. (2024). Rethinking copyright authorship in the age of generative AI. *Peking University Law Journal*, 35(1), 88–105.

[10] Standing Committee of the National People’s Congress. (2020, May 28). Civil Code of the People's Republic of China (Book IV: Personality Rights, Arts. 990–1039).

[11] Ji, H., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 1.

[12] Standing Committee of the National People’s Congress. (2020, May 28). Civil Code of the People's Republic of China (Book VII: Tort Liability, Art. 1195).

[13] Wang, Q. (2023). Product liability and algorithmic harm in civil law: A Chinese perspective. *Tsinghua China Law Review*, 15(1), 67–89.

[14] Gervais, D. (2020). Is copyright law ready for artificial intelligence? *Houston Law Review,* 57(3), 871–908.

[15] Standing Committee of the National People’s Congress. (2020, November 11). Copyright Law of the People's Republic of China (Arts. 10, 48, as amended).

[16] Smolla, R. A. (2023). Defamation law (2nd ed., pp. 5–18). Thomson Reuters.

[17] Post, R. C. (1990). The constitutional concept of public discourse: Outrageous opinion, democratic deliberation, and *Hustler Magazine v.* Falwell. *Harvard Law Review*, 103(3), 601–686.

[18] Prosser, W. L. (1960). Privacy. *California Law Review*, 48(3), 383–423.

[19] Communications Decency Act, 47 U.S.C. § 230. (1996).

[20] Standing Committee of the National People’s Congress. (2017, November 4). Criminal Law of the People's Republic of China (Art. 246, as amended).

[21] He, W. (2022). Criminal defamation and its judicial limits. *Peking University Law Journal,* 20(2), 74–92.

[22] Zhang, W. (2023). Algorithmic harm and negligence in Chinese criminal law. *Tsinghua Law Review*, 18(1), 41–60.

[23] Citron, D. K. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*, 98(5), 147–155.

[24] Kerr, O. S. (2019). Aiding, abetting, and the expansion of criminal liability in cybercrime. *Harvard Journal of Law & Technology*, 33(1), 1–37.

[25] Standing Committee of the National People’s Congress. (2020, May 28). Civil Code of the People's Republic of China (Arts. 1024–1039).

[26] Zhou, Y. (2023). Personality rights and AI: Reconstructing defamation liability in the age of algorithms. *Peking University Law Journal*, 34(2), 45–68.

[27] Zhang, Y. (2024). Platform liability and AI-generated content under the Chinese Civil Code. *Tsinghua China Law Review*, 15(1), 102–117.

[28] New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

[29] Communications Decency Act, 47 U.S.C. § 230. (1996).

[30] Cyberspace Administration of China. (2023, July). Administrative measures for generative AI services (Draft for comments). https: //www.cac.gov.cn

[31] Citron, D. K., & Wittes, B. (2018). The internet will not break: Denying bad Samaritans Section 230 immunity. *Fordham Law Review,* 86(2), 401–424.

[32] Cyberspace Administration of China. (2023, August). Interim measures for the management of generative artificial intelligence services. https: //www.cac.gov.cn

[33] U.S. Congress. (2023, October). No Fakes Act (Discussion draft). Senate Judiciary Subcommittee on Intellectual Property. https: //www.congress.gov