Economics, Law and Policy
ISSN 2576-2060 (Print) ISSN 2576-2052 (Online)
Vol. 9, No. 1, 2026
www.scholink.org/ojs/index.php/elp

Original Paper

Service Provider Liability for Damages Caused by Generative AI Hallucinations

Yuxin Yan[1]

[1] School of Public Administration, Southwest Jiaotong University, Chengdu, China

Received: April 25, 2026 Accepted: May 9, 2026 Online Published: May 11, 2026

doi:10.22158/elp.v9n1p354 URL: http://dx.doi.org/10.22158/elp.v9n1p354

Abstract

The concept of damages caused by generative AI hallucinations refers to situations in which the AI generates false information, thereby giving rise to fortuitous liability. As the controllers of the AI, service providers are the appropriate parties to bear such liability. Once their status as liable parties is established, it becomes necessary to further clarify the specific rules for determining liability: in principle, fault-based liability should apply, with the state of the art serving as the standard for assessing fault; regarding the proof of causation, presumptive rules should be adopted to allocate the burden of proof reasonably. The aforementioned liability framework strikes a balance between the dual values of technological innovation and the protection of individual rights.

Keywords

generative artificial, AI hallucinations, service providers, fault-based liability

1. Introduction

The rapid development and widespread application of generative AI technology, while bringing immense convenience to social life and production, have also given rise to new and non-negligible legal risks. Training data comes from many different sources. It is hard to tell what is real and what is fake. So generative AI may produce content that does not match objective facts. It may even make things up completely. This phenomenon is called AI hallucination. One thing needs to be clear. AI hallucinations are not caused by users on purpose. Users are not trying to trick the AI into generating false information. Instead, the AI passively generates false content during its normal operation. That false information can show up as fabricated facts. It can also show up as wrong citations or logical inconsistencies. How serious the harm is depends on the specific situation. At a minimum, it may mislead users. In the worst case, it may harm others. That harm could affect personality rights, property rights, or even the public interest. Given this background, service providers are the direct controllers of

generative AI. They face several pressing issues in legal practice and academic research. These issues include whether they should bear fortuitous liability for damages caused by AI hallucinations, the specific conditions under which such liability arises, and how to reasonably define the scope of that liability.

2. Reasons for Generative AI Hallucinations

The phenomenon of "hallucinations" in generative artificial intelligence is far more pronounced than in general AI; it refers to the generation of content that is inconsistent with objective facts, logically incoherent, or entirely fabricated. The causes of this phenomenon can be analysed across four dimensions: data, human oversight, technology, and objective limitations.

It is impossible to fully guarantee where the data comes from or whether it is authentic. Huge amounts of training data get collected from many different sources. These include books, magazines, online articles, and even encyclopaedias. Research companies do not guarantee the authenticity of these documents. On top of that, the sheer volume of data is a problem. Models tend to absorb and amplify errors, biases, or outdated information. This ultimately leads to factual hallucinations and value distortions. Now consider the people who manage these systems. They exert considerable influence over the content that gets generated. Manually annotated data helps refine reward models. It also helps improve the performance of large scale language models. But human evaluation is inherently subjective. It is also prone to structural cognitive biases. Different cultural contexts can lead to different judgments about the same fact. Those judgments can even be unfair. This results in a systematic bias in how hallucinatory content is produced. There are also technical limitations with the system itself. Generative AI does not understand context the way humans do. Instead, it relies on the probabilistic reconstruction of learned information. It has no built in fact checking mechanism. Add to that the problem of inadequate selection during decoding. Add also errors in the attention mechanism. Together, these can cause logical hallucinations. Examples include erroneous conclusions or internal contradictions. They can also cause referential hallucinations. That is when the AI invents non existent sources of information. Given current technical limits, hallucinations are somewhat inevitable. Even the most modern methods cannot completely rule them out. This is due to inherent problems within the system. These problems include statistical fitting biases and failure to delimit representation. This matters a lot when we later try to clarify responsibilities.

3. The Legitimacy of Service Provider Liability for Damages Caused by AI Hallucinations

Generative artificial intelligence can produce its own text. It can also generate images and videos. This kind of creativity looks very similar to human creativity. So people are now debating who should take responsibility for the hallucinations created by generative AI. We need a careful analysis. That analysis should look at whether granting legal personality to AI is even feasible. It should also examine whether holding service providers liable is a legitimate approach.

3.1 The Debate over Legal Personality for Artificial Intelligence

There is considerable controversy regarding the determination of liability for damages caused by AI hallucinations arising from the inherent flaws of generative AI technology. The negative view holds that neither weak nor strong artificial intelligence possesses the capacity to recognise rights and obligations, nor does it have the legal standing to enjoy rights or fulfil obligations. The theory of limited legal personality argues that AI entities possess a degree of independence, which could enable them to acquire a certain level of legal personality; however, as they are not natural living beings, they are subject to restrictions regarding the assumption of rights, obligations and liability. The theory represented by the affirmative view is the doctrine of legal fiction, which draws an analogy between AI and legal persons: legal persons are also non-living entities, and AI could similarly acquire legal personality through the means of legal fiction. Whether or not generative artificial intelligence is recognised as having legal personality has implications for the allocation of liability among service providers. If legal personality is affirmed for generative artificial intelligence, tort liability would be borne by the AI as property; if legal personality is denied, the service provider controlling the AI should bear corresponding liability under certain conditions.

3.2 Confirmation of the Service Provider as the Liable Party

3.2.1 Generative Artificial Intelligence Does Not Have Legal Personality

The author contends that generative artificial intelligence cannot be regarded as a subject of tort liability. Generative AI lacks the legal capacity to hold rights and obligations, nor does it possess moral agency. First, it is incapable of responding to rights and obligations and lacks personal dignity. Artificial intelligence operates solely according to algorithmic program mes designed by its controllers; rights and obligations ultimately bind those controllers. Imposing rights and obligations on artificial intelligence is futile, as such measures cannot achieve the purpose of legal norms. Second, requiring AI to bear liability is of no practical significance because AI lacks consciousness. Tort liability serves not only a remedial function but also a punitive one. Legal liability cannot punish AI, as financial and personal liability hold no meaning for it; AI will not show subjective remorse, nor will it experience any sense of punishment after a mistake. Finally, from a comparative law perspective, the EU's "Legislative Initiative on Liability for the Operation of AI Systems" states that AI lacks human cognition and legal capacity and is merely a tool serving humanity. The US "National Artificial Intelligence Initiative" also emphasises the instrumental status of AI. The view that generative AI does not possess legal personality remains the predominant stance adopted by most countries.

3.2.2 The Rationale for Treating Service Providers as Indirectly Liable Parties

Since generative artificial intelligence lacks legal personality, service providers should be regarded as one of the parties liable for its actions. This conclusion is mainly based on the theory of risk control. Generative AI relies on underlying data and algorithmic models. Those are provided mainly by technology providers. But service providers are the direct users of these models. They also engage in secondary development based on the original models. This gives them actual control over the model's

outputs. The theory of risk control says the following. The party that creates the source of risk has the ability to control that risk. That same party can also mitigate the risk. So that party should bear liability for any resulting harm. Service providers are in the best position to prevent or reduce hallucination related harm. They can do this through model adjustments and content filtering. Therefore holding them liable for fortuitous acts is consistent with the basic legal principles of risk allocation.

Another point relates to vicarious liability and the legal obligations of service providers. A service provider's omission may also count as a fortuitous act. The hallucinated content is directly generated by artificial intelligence. But the service provider might fail to take necessary measures. This failure happens when the provider discovers unlawful content. Or when the provider is reasonably expected to have discovered it. That missing action amounts to vicarious liability. China's Interim Measures for the Administration of Generative Artificial Intelligence Services makes this clear in Article 14. Service providers have a duty to take prompt remedial action once they discover unlawful content. This helps stop the harm from spreading. So service providers have two kinds of obligations. One is negative. They should refrain from actively creating false information. The other is positive. They must exercise due care. They must conduct inspections. They must take timely action. Now suppose hallucinations arise because of the AI itself. And suppose the service provider fails to meet these obligations. Then a legal causal link exists between that omission and the resulting harm. The service provider shall bear corresponding tort liability.

4. Service Provider Liability for Damages Caused by Generative AI Hallucinations

Service providers may be held liable for damages caused by generative AI hallucinations, and the elements constituting such liability require clarification.

4.1 Fault-Based Liability as the Applicable Principle

4.1.1 Controversy over the Principle of Liability

There are two main schools of thought on liability principles for providers of generative AI services. One view says service providers should bear strict liability. Under this view, service providers should be subject to product liability. Their status is like that of a seller or a manufacturer. The reasoning behind this view has two parts. Companies mass produce products. Any infringement caused is likely to be on a large scale. Imposing strict liability works as a deterrent for manufacturers and sellers. Generative AI has a broad reach. Any incident that causes harm is highly likely to be a mass infringement. So product liability should apply. Another reason is that users find it hard to prove fault by the generative AI service provider. If we apply fault based liability, users must follow the rule that the one who asserts must prove. But users of generative AI or other injured parties have limited access to internal information. They cannot prove that the provider breached its duty of care. The other school of thought advocates for a different approach. It supports the general tort liability principle. That principle is fault based liability. The rationale supporting this view lies primarily in denying the product-like nature of generative AI and, taking into account the state's policy orientation of

encouraging technological innovation, supports the application of the fault-based liability principle.

4.1.2 The Legitimacy of Applying the Fault-Based Liability Principle

The author submits that the relevant rules on product liability cannot be directly applied to generative AI, particularly in cases where damage is caused by AI hallucinations. Article 2(2) of the Product Quality Law of the People's Republic of China defines a product as an item that has been processed or manufactured for the purpose of sale. Whilst most scientists who advocate for the introduction of a product liability regime tend to cite physical products as examples to support their arguments, there is, however, a fundamental difference: generative AI can deliver its functional value without the need for a physical medium. One of the main characteristics of a product is its high degree of uniformity, which means that liability risks are distributed through market mechanisms. Conversely, in the context of generative AI services, each user's input data and their use of it vary; it is precisely there that human characteristics lie. Consequently, the system of liability for defective products should not apply in this case. A more detailed analysis of the causes of hallucinations in generative AI shows that, under current technical conditions, certain types of hallucinations are inevitable. Imposing strict liability on service providers would not only be contrary to the principle of fairness but would also hinder the innovative development of generative AI technology. In summary, it can be argued that, in cases of damage caused by hallucinations in generative AI, the principle of liability for negligence should apply to service providers.

4.2 The Determination of Fault Is Based on the State of the Art

In traditional tort proceedings, the principle of fault generally applies, which means that the burden of proof lies with the claimant. The claimant must prove that the defendant has indeed committed a fault. However, the situation becomes more complicated when generative AI comes into play. This is because the mechanisms by which generative AI produces information are extremely complex and involve specific technical principles that are beyond the understanding of the average user. Added to this is the fact that the algorithms are effectively "black boxes". It is therefore difficult for users to prove whether the training data was flawed or whether the AI system itself failed to meet the prescribed standards. In these circumstances, it therefore seems not only unfair but also impractical to place the burden of proof on the user. Article 3 of the EU's Draft Directive on Rules on Non-Contractual Civil Liability for Artificial Intelligence provides that, in the case of high-risk AI systems, a court may order the service provider to provide evidence demonstrating that it has exercised due care; if such evidence is not provided or fails to meet the required standard of proof, it shall be deemed that due care has not been exercised. It is evident, therefore, that the EU adopts a presumption of fault, shifting the burden of proof regarding fault to the service provider.

Service providers must demonstrate that they are not at fault. This is their responsibility. They can demonstrate that they have exercised due diligence. They can also demonstrate that they have fulfilled their obligation to prevent harm. The criterion for assessing these obligations is the current state of the art. If technical measures have been taken but the harm is unavoidable, or if it is not possible to remove

the unlawful content, the service provider is deemed to have fulfilled their obligation. How is the current state of the art assessed? We can analyse this from different perspectives. Let us consider the time factor. The state of the art at the time the harm occurred serves as the criterion. We need to look at the state of the art within the sector. We also need to consider what technology is available to the service provider. This standard does not require the very latest technology. It requires technology that is generally available within the sector. Another point is about foreign technology. Service providers cannot simply exclude it. They cannot do so just because the technology is not available in their country. They must thoroughly examine feasibility. They must see if they can acquire the technology. They must see if they can implement it. They also have to comply with international standards. They must adapt to these standards as far as possible.

4.3 The Burden of Proof Regarding Causation Is Subject to the Rules of Presumption

Depending on the stage at which the damage occurs, it can be categorised as damage arising during the generation stage or damage arising during the removal stage. The generation stage refers to the stage at which the AI generates content and outputs information; the removal stage refers to the stage at which the service provider, having become aware of the existence of hallucinated content, fails to take appropriate measures. Regarding the causal relationship of AI hallucinations during the generation stage, users need only demonstrate that they have reported or complained about the issue, and that the service provider knew or ought to have known of it but failed to take measures to prevent the damage from spreading, in order to establish the causal relationship. However, when damage caused by AI hallucinations occurs, it is difficult to establish the causal relationship during the generation stage.

We divide the assessment of causation into two parts. One is factual causation. The other is legal causation. Factual causation is not hard to figure out. AI hallucinations are produced by generative AI. The service provider is the controller. So the provider naturally bears factual causation. Legal causation is harder to determine. In China, the prevailing view uses the theory of sufficient causation. This theory looks at an ordinary person's objective experience. It asks whether a causal relationship exists between an act and its result. The process behind generative AI's results is complex. Users find it hard to pinpoint specific infringing acts by the service provider. To address this difficulty, we should ease the user's burden of proof. One way is to adopt a presumption of causation. This method is not the same as reversing the burden of proof. The injured party still needs to provide prima facie evidence of causation. That evidence helps establish the presumption. Then the service provider can rebut it. The burden of proof stays with the injured party. However, the standard of proof is lowered. This approach balances the interests of service providers and users better.

4.4 Grounds for Exemption and Limitation of Liability for Service Providers

Service providers should bear some liability. But they may claim exemption or mitigation under specific circumstances. One situation is malicious user input. Or leading questions. Or authorised tampering with generated content. If damage from AI hallucinations comes from these, the service provider may be exempt. Or its liability may be mitigated. Another situation involves the state of the

art. The service provider may have taken measures. These measures could include filtering, prompts, and corrections. These measures are in line with the current state of the art. Yet hallucinations remain unavoidable. In that case, the provider is deemed not at fault. It shall not be liable for compensation. There is also the context of secondary development of open source models. A service provider might just integrate the base model. It makes no substantial modifications. It has also fulfilled its duty to provide prominent warnings. Then the provider may enjoy limited liability under the safe harbour principle. Clarifying these grounds for exemption and limitation of liability helps incentivise service providers to actively adopt preventive measures, whilst avoiding the imposition of unreasonable absolute liability upon them.

5. Conclusion

The technical limitations of generative AI make AI hallucinations difficult to avoid, and the harm caused by such hallucinations may infringe upon the legitimate rights and interests of users or third parties. As generative AI lacks legal personality, service providers, as its controllers, should bear liability for damages caused by such hallucinations. To balance technological innovation with the protection of rights and interests, the liability principle for service providers should be based on fault, with the standard for determining fault being the current state of the art. Given the difficulty in proving causation, presumption rules should be applied to allocate the burden of proof reasonably.

Looking ahead, relevant legislation and judicial practice should, building upon the existing liability framework, further refine the criteria for determining the "state of the art", improve the rules of evidence concerning damages caused by AI hallucinations, and strike a dynamic balance between encouraging technological innovation and safeguarding individual rights. In doing so, a regulatory framework for AI tort liability that is both forward-looking and practical should be gradually established.

References

Choi, B. H. (2024). AI malpractice. *DePaul Law Review*, *73*(2), 301-362.

Crootof, R. (2020). Tort liability and the internet of things. In W. Barfield, & J. M. Williams (Eds.), *The Oxford handbook of ethics of AI* (pp. 543-558). Oxford University Press.

DiMatteo, L. A., Poncibò, C., & Cannarsa, M. (2022). AI and legal personhood. In L. A. DiMatteo, C. Poncibò, & M. Cannarsa (Eds.), *The Cambridge handbook of artificial intelligence: Global perspectives on law and ethics* (pp. 288-303). Cambridge University Press. https://doi.org/10.1017/9781009072168

Fernández Llorca, D., Charisi, V., Hamon, R., Sánchez, I., & Gómez, E. (2023). Liability regimes in the age of AI: A use-case driven analysis of the burden of proof. *Journal of Artificial Intelligence Research*, *76*, 1302-1348. https://doi.org/10.1613/jair.1.14565

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework* (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology.