**Open Journal of Legal Science** 法学, 2026, 14(9), 155-162
Published Online September 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.149268

Hans汉斯

# 生成式人工智能幻觉致损的侵权法回应

**刘卓然**

扬州大学法学院，江苏 扬州

收稿日期：2026年8月14日；录用日期：2026年8月27日；发布日期：2026年9月22日

## 摘　要

**生成式人工智能幻觉是指模型生成看似合理、实则与事实不符的内容的现象，是大语言模型基于概率预测机制运行的内生现象。这一技术现实对传统侵权规则提出了挑战。为能够在现行侵权法体系内回应生成式人工智能幻觉致损问题，本文首先阐明了人工智能幻觉的技术特质及其引发的侵权法困境，继而在梳理无过错责任、产品责任与过错责任等学说争议的基础上，论证了适用过错责任原则的正当性。在过错认定层面，本文提出了应以合理注意义务为核心，并结合应用场景、服务阶段与信息类型对服务提供者的注意义务进行动态调整；同时亦应当以动态的、多层次的标准综合考虑使用者的过错。上述方案旨在为生成式人工智能幻觉侵权这一新型侵权形态的司法裁判提供可操作的判断基准。**

## 关键词

**人工智能幻觉，侵权责任，过错责任原则，合理注意义务**

# Tort Liability for Generative AI Hallucination

**Zhuoran Liu**

School of Law, Yangzhou University, Yangzhou Jiangsu

Received: August 14, 2026; accepted: August 27, 2026; published: September 22, 2026

## Abstract

**Generative AI hallucination refers to the phenomenon in which models generate seemingly plausible but factually inaccurate content. It is an inherent byproduct of large language models operating on probabilistic prediction mechanisms. This technological reality poses challenges to traditional tort rules. In response to the tort law issues arising from hallucination-induced harm, this article first clarifies the technical characteristics of AI hallucination and the resulting doctrinal difficulties. It then examines competing theories including no-fault liability, product liability, and fault liability,**

文章引用: 刘卓然. 生成式人工智能幻觉致损的侵权法回应[J]. 法学, 2026, 14(9): 155-162.
DOI: 10.12677/ojls.2026.149268

**ultimately demonstrating the propriety of the fault liability principle. Regarding the determination of fault, this article proposes that reasonable duty of care should serve as the core standard, with dynamic adjustments based on application scenarios, service stages, and information types, and that, on the basis of such a dynamic framework, the user's contributory fault should also be taken into account comprehensively. The proposed framework aims to provide an operable analytical benchmark for judicial adjudication of this emerging type of harm caused by generative AI hallucination.**

## Keywords

**AI Hallucination, Tort Liability, Fault Liability Principle, Reasonable Duty of Care**



## 1. 引言

2025 年，杭州互联网法院审结了全国首例因生成式人工智能幻觉引发的侵权纠纷案 [1]。该案被写入最高人民法院工作报告，作为准确把握科技创新容错空间的典型案例。所谓“幻觉”，是指人工智能模型生成看似合理、实则与事实不符的内容的现象。在这一案件中，人工智能就某高校报考信息生成了不准确内容，并在被纠正后仍坚持错误，引发了一场关于技术容错边界与侵权责任的讨论。随着生成式人工智能的广泛应用，“幻觉”正从技术议题走向法律议题，即人工智能生成的错误信息可能侵害人格权，也可能因误导用户决策而造成经济损失，而传统的侵权责任规则能否有效回应这一新型风险，尚待深入审视。

本文围绕“幻觉”这一核心命题展开。首先阐明其技术特质及其对传统侵权法提出的挑战，进而在梳理既有归责原则争议的基础上论证过错责任原则的正当性，最后就过错认定的具体标准与特殊判断提出有针对性的规范方案。

## 2. 生成式人工智能幻觉的技术特质

生成式人工智能幻觉是指人工智能模型生成不正确、捏造或误导性的信息，并以令人信服的方式呈现，这一现象是自然语言模型所固有的功能[1]。需要注意的是在这一过程中用户并未实施以生成虚构内容为目的的行为，人工智能算法也并未存在运行错误，其认为其所生成的内容是“真实可靠”的。从技术原理看，大语言模型的工作原理是基于海量数据进行概率预测，然后逐词生成内容。模型在评估一个句子时，会通过逐词预测的方式，将每一步的条件概率连乘，得出一个总概率值，该概率值反映了该句子与模型在海量数据中学到的统计规律的符合程度。换言之，人工智能大模型输出的内容是基于所拥有的数据库中的数据运用概率判断进行预测的结果，而并不是基于“学习”真正事实进行是非判断的结果。

基于上述原因，使得人工智能幻觉侵权行为具有原生性、被动性的特征。首先，原生性是指这是根植于人工智能大模型内部的技术局限，只要使用生成式人工智能就会有产生这种现象的可能，而并非其故障。其次，其具有被动性，虚假内容由算法自动生成，并非服务提供者主动编辑、推送或刻意诱导人工智能造假，而人工智能服务提供者也并不能事先预测幻觉出现的准确时间以及场景。

[1]梁某诉某人工智能公司网络侵权责任纠纷案，杭州互联网法院(2025)浙 0192 民初 18143 号。

上述特点使得此类侵权不同于传统网络侵权，也不同于其他类型的人工智能侵权，首先这类侵权中似乎并不存在传统概念中的侵权行为，在传统侵权法中一般以“行为”作为责任产生的起点，无论是作为还是不作为，但这类侵权中服务提供者、用户以及人工智能模型都未出现错误，其次，传统侵权法中的过错以行为人对损害后果的可预见性和可控制性为前提，人工智能幻觉的发生在事前具有高度的不可预测性。服务提供者无法预知人工智能将在何时、对何种提问、以何种方式产生幻觉。每一次输出都是对用户开放式指令的即时响应，具有不可完全复现的独特性，自然也对其不具有相当的可控制性。其次，在这类侵权行为中，由于生成式人工智能具有强交互性，使用者行为、人工智能自主行为与服务者的行为交织在一起，到底是何种行为决定了输出行为的幻觉变得不甚明确。由此导致了侵权法内容对于生成式人工智能幻觉适用的模糊性，重点主要集中在归责原则的选择以及对过错认定标准的设定上，因此下文将在既有学说争议的基础上对这两个问题进行进一步的分析。

## 3. 生成式人工智能幻觉致损的归责原则选择

### 3.1. 关于生成式人工智能侵权归责原则的主要观点梳理

对于该种新型侵权形态适用何种归责原则，学界主要有以下几种观点。第一种观点主张对生成式人工智能幻觉侵权适用过错责任原则，其认为人工智能侵权属于《中华人民共和国民法典》[2](以下简称《民法典》)规定的一般侵权行为，对其应当适用过错责任原则。不能将生成式人工智能等同于《民法典》中产品责任中的产品，因为其本质是基于算法提供的一种信息服务，并且不会因为其固有缺陷而直接造成危害结果[2]。第二种观点认为生成式人工智能虽然不是产品但对其导致的侵权也应当适用无过错责任原则，其理由主要在于侵权法的目的在于分担损害，适用无过错责任原则在无法确定过错时更有利于保护“更无辜方”，即因为用户与服务提供者之间存在明显的信息不对等，虽然二者对于损害的发生都可能是无法预见的但是服务提供者明显更有能力进行预防，并且既然服务提供者以此获利就应当承担由此可能产生的危险[3]。也有学者根据法律经济理论，认为适用无过错责任原则不仅更符合经济效益，也能够更有效地消除算法黑箱的负面影响[4]。与此相关的观点则认为生成式人工智能因为属于产品而应当适用产品责任，其理由主要在于人工智能具有产品所具有的加工、制作以及销售等属性，不应因为其不具有物质形态就认为其属于服务，并且其所输出的是文本并不意味着生成文本的软件或系统不是产品，同时产品责任有利于实现由事后追责到事前监管的治理方式转换[5]。

### 3.2. 过错责任原则的立场选择与论证

#### 3.2.1. 无过错责任原则在幻觉致损场景中的不适用性

在生成式人工智能因幻觉侵权的这一具体场景下，无论是将其作为适用无过错责任原则下的新侵权形态或者是归入产品责任原则都是不恰当的。

首先，《民法典》第一千一百六十六条明确规定，无过错责任原则的适用以“法律规定”为必要条件[3]。除替代责任外，法律之所以规定其他适用无过错责任原则的侵权形态都出于其对人身、财产安全有重大危险的考量。而人工智能幻觉侵权并不像其他适用无过错责任原则的行为一样直接危害人身与财产安全，即使其能够造成人格权或财产权损害，其损害程度也并不能与医疗事故侵权等相提并论，有研究指出在损害程度和波及范围方面，生成式人工智能的运行和使用一般具有较低的风险外溢性[6]。虽然无过错责任可以将现代文明发展伴生的具有社会必要性与正当性但无不法性的风险活动所致损害，在社会成

[2]https://www.court.gov.cn/zixun/xiangqing/233181.html
[3]《中华人民共和国民法典》第一千一百六十六条：行为人造成他人民事权益损害，不论行为人有无过错，法律规定应当承担侵权责任的，依照其规定。

员之间进行合理分配，以避免由受害人独自承受发展代价的不公，从而实现高度社会化生产条件下的公平正义[7]。但是应当注意，能够适用无过错责任原则的理论基础之一是危险控制理论，该理论认为风险责任应当由危险活动或者危险物的控制者承担。但在生成式人工智能幻觉侵权这一场景下，其并非一种危险而仅是一种技术内生现象，并且无过错责任虽然通过风险分配实现损害的分散，但其适用须以特定行为所创造的特殊危险为前提[8]，而人工智能幻觉侵权与现有适用无过错责任原则的侵权形态所具有的危险程度无法相提并论。同时，无过错责任的正当性基础是分配正义，其应当适用于大规模侵权[9]，然而生成式人工智能幻觉侵权具有偶发性、个别性，并非大规模扩散性损害。此外，虽然用户相对于服务提供者处于弱势地位，但也不应因此就剥夺服务提供者自证清白的机会，尤其是人工智能具有强交互性，损害的发生有时并非单纯是由人工智能幻觉输出的错误信息所致。综上，对人工智能幻觉侵权适用无过错责任原则明显超出了该原则的规范射程。

其次，对于生成式人工智能是否属于产品。“产品”概念及其外延的界定是产品责任制度的起点[10]。《中华人民共和国产品质量法》[4]第二条明确规定：“本法所称产品是指经过加工、制作，用于销售的产品”[5]。上述定义意味着经过一定的加工与制作并以物质形态存在的产品，其在物理形态、功能边界和风险范围是明确且相对固定的，而生成式人工智能则不同，其产生具有高度的随机性，并高度依赖于算法模型与用户输入的提示词。每一次输出都是对用户开放式指令的即时响应，具有不可完全复现的独特性，人工智能幻觉究竟何时会出现幻觉并导致何种风险具有着高度的随机性。

再次，根据《民法典》第一千二百零二条所规定的产品生产者责任，在产品存在缺陷的情况下，产品的生产者才承担产品侵权的无过错责任[6]。但目前普遍认为人工智能幻觉并非产品“缺陷”，所谓缺陷通常指产品存在不合理的危险性，而幻觉是大语言模型基于概率预测机制运行的必然伴生现象。作为一种技术内生现象，将其与因为生产、设计等失误所造成的产品缺陷等同，缺乏事实基础。在人工智能幻觉侵权第一案中，杭州互联网法院亦从概念与构成要件上指出该服务缺乏具体、特定的用途与合理可行的质检标准[7]。没有合理可行的质检标准，就无法判断何为“合格”、何为“缺陷”，由此构成了将“幻觉”致损纳入产品责任框架的障碍。

最后，从宏观政策及产业长远发展来看，对新型技术适用无过错责任原则，将过度束缚其积极性，不利于科技创新与引导行业良性发展。从技术与法律关系的发展历史来看，法律应当对新型技术秉持相对包容的态度。互联网产业发展初期，网络服务提供者面临着巨大的法律不确定性，用户上传的内容可能涉及著作权侵权，而平台却缺乏对海量内容进行事前审查的技术能力，也无法承受因此产生的成本，美国国会在 1998 年通过的《数字千年版权法案》[8]为互联网企业创设了“避风港原则”，适度免除了互联网企业的部分责任，正是该规则极大地促进了网络服务产业的发展[11]。除抑制产业创新外，适用无过错责任原则未必会比适用过错责任原则更能够激励服务提供者在事前采取更严密的预防措施。由于现有技术原因，进一步消除人工智能幻觉将极大增加技术成本，但是人工智能幻觉致损则具有偶发性。当采取安全保障措施和提高技术安全性的费用大于提价损失或保险费用支出，服务提供者必然放弃提高产品或服务安全性[7]，由此将更不利于行业技术的发展与算法透明度的提升。同时中小企业将因无法承受进一步消除幻觉的技术成本而在竞争中处于劣势，加剧市场集中效应。因此，在人工智能发展初期因技术局限所产生的问题，法律不宜对其施加过于严苛的责任标准。

[4]https://www.samr.gov.cn/zfjcj/tzgg/art/2023/art_579118cd202a45fba28b7edfd9f6fd72.html

[5]《中华人民共和国产品质量法》第二条：在中华人民共和国境内从事产品生产、销售活动，必须遵守本法。本法所称产品是指经过加工、制作，用于销售的产品。建设工程不适用本法规定；但是，建设工程使用的建筑材料、建筑构配件和设备，属于前款规定的产品范围的，适用本法规定。

[6]《中华人民共和国民法典》第一千二百零二条：因产品存在缺陷造成他人损害的，生产者应当承担侵权责任。

[7]同 1。

[8]https://www.copyright.gov/dmca/

#### 3.2.2. 适用过错责任原则的正当性

首先，“幻觉”产生的技术性特点决定了过错责任原则是最为适宜的归责原则。一方面，与严格责任相比，过错责任的优势不在于更轻、更宽松，而在于其评价标准与技术现实的适配性。无过错责任原则要求服务提供者对幻觉致损承担不问过错的赔偿责任，法律不应强制让服务提供者对于无法根除的技术现象承担严格责任；而过错责任原则将评价的重点由结果转移至行为，它不苛求服务提供者实现消除幻觉的结果，只评价其是否采取了与技术发展阶段相适应的预防措施，不要求结果的绝对准确，只关注行为本身的合理性。另一方面，过错责任的“合理注意义务”标准具有更灵活的动态适应能力，在技术尚未成熟时，不会因过高的标准而扼杀创新，而随着技术的发展，“合理”的内涵会随之提升，责任标准也会同步提高。这种灵活性使得法律既能通过注意义务的设定施加适度的压力，促使服务提供者不断优化技术、降低幻觉风险，又不会因标准过高而导致责任泛化或产业停滞。

其次，适用过错责任原则与现有法律体系更加协调。第一，《民法典》第一千一百六十五条第 1 款确立了过错责任原则[9]，这是侵权责任的一般归责原则。第一千一百六十六条规定的无过错责任原则[10]，须以“法律规定”为适用前提。这一体系安排意味着：在法律没有明确规定的情况下，应当首先考虑适用过错责任原则。第二，将生成式人工智能定性为“服务”而非“产品”，也与《生成式人工智能服务管理暂行办法》的规范定位一致[11]。该办法将生成式人工智能明确定位为“服务”，并围绕服务提供者的义务展开。在这一体系定位下，选择过错责任原则而非产品责任的无过错责任原则，能够更好地维护现行规范体系的内在一贯性。第三，过错责任原则与现行法中的其他相关制度也具有体系上的协调性。《民法典》第一千一百九十四条以下关于网络侵权责任的规定，对网络服务提供者同样采取过错责任的基本立场，虽然生成式人工智能与传统网络服务在技术形态上存在差异，但在为用户提供信息中介服务这一功能定位上具有相似性。对生成式人工智能服务提供者适用过错责任原则，与网络侵权责任的基本规范逻辑保持了体系上的连贯。

最后，对过错责任原则中存在举证困难质疑的回应。适用过错责任原则面临的主要质疑是在幻觉致损场景中，受害人因无法获知人工智能系统在特定输出时点的内部运行状态，难以证明服务提供者存在过错而面临举证困难，从而导致其损害不能获得有效填补。但是归责原则的选择与举证责任的分配属于两个不同层面的问题。归责原则解决的是“以何种标准判断责任”的问题，举证责任解决的是“由谁承担证明不能的后果”的问题。二者虽有联系，但并不等同。同时，适用过错责任原则并不排斥过错推定这一制度的适用，有学者即认为在高度危险程度与举证困难叠加的特定情况下，可以通过引入过错推定从而破除这一难题[12]。此外，还可以在过错责任原则的框架内，通过提升透明度的实体性权利义务规则与程序性证据开示规则相结合，从而纾解被侵权人的过错证明压力。

## 4. 生成式人工智能幻觉致损中过错认定的具体标准

归责原则的选择仅是第一步，它回答了“以何种标准判断责任”的问题，但并未解决“如何判断过错”这一更为具体的问题。适用过错责任原则是否妥当，还应当考虑其是否能具体地适用于个案。当代民法采用以是否违反“注意义务”为核心的过错判断方法，从而实现了过错认定的客观化[13]，在人工智

[9]《中华人民共和国民法典》第一千一百六十五条：行为人因过错侵害他人民事权益造成损害的，应当承担侵权责任。依照法律规定推定行为人有过错，其不能证明自己没有过错的，应当承担侵权责任。

[10]同 3。

[11]《生成式人工智能服务管理暂行办法》第二条：利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容的服务(以下称生成式人工智能服务)，适用本办法。国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的，从其规定。行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术，未向境内公众提供生成式人工智能服务的，不适用本办法的规定。

参考链接：https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

能幻觉侵权中也应该继续沿用这一逻辑。因此，下文将围绕“合理注意义务”这一核心，进一步探讨其在生成式人工智能幻觉致损场景中的具体内涵与判断标准。

### 4.1. 服务提供者合理注意义务的具体考量因素

第一，服务提供者是否履行了风险提示义务。用户之所以会轻易采信人工智能生成的结果，本质上是出于对人工智能输出信息的高度信赖。如果服务提供者能够在生成结果处进行足够明显的风险提示，将极大地降低用户对人工智能输出信息的信赖程度，用户会仅将其作为适度参考而非做出决策的唯一依据，更不会轻易将其用于重要决策，由此能够有效降低因人工智能幻觉而导致损害的风险。在首例人工智能幻觉案中，杭州互联网法院将被告在欢迎页、用户协议及交互界面设置提示标识的事实，作为认定其已尽注意义务的重要依据[12]。

第二，服务提供者是否采取了行业通行的技术预防措施。合理注意义务的关键在于“合理”具体而言就是与现有技术水平相适应的注意义务。不同的技术措施对降低幻觉概率具有不同的效果，其所对应的成本亦有差别。在前述案件中，法院认定被告已经尽到合理注意义务的另一依据便是其已采用检索增强生成等技术提升输出可靠性的事实[13]。过错认定应当审查服务提供者是否采用了同行业通行的技术措施。这一标准的核心是行业通行而非绝对先进，并不强制要求服务提供者采用尚在探索中的前沿技术，但要求其需要达到行业一般水平。

第三，服务提供者是否建立了有效的用户反馈与纠错机制。需要注意的是幻觉的发生虽然不具有可预测性，但发生后可以通过用户反馈来发现和纠正。如果服务提供者有证据证明其建立了有效的用户反馈渠道，并在收到反馈后及时采取相应优化措施，则表明其已经尽到了持续的注意义务。反之如果服务提供者未建立相应反馈机制或者对于用户反馈并未进行相应优化，使得某一错误持续出现，则应当认定其在注意义务的履行中存在一定瑕疵。

### 4.2. 服务提供者合理注意义务的动态化认定

服务提供者防止人工智能幻觉致损的合理注意义务不应当是静态的，而应当采取的是动态的“合理”标准，需要结合技术水平、服务类型与侵权内容等多维因素综合判断，单一化的注意义务标准可能导致保护不足或阻碍创新[14]。基于上述认识，以下从应用场景、服务阶段与信息类型三个维度，展开注意义务动态调整的具体路径。

#### 4.2.1. 根据应用场景对合理注意义务的动态调整

现有的生成式人工智能的一大特点即为应用场景广泛，提供的内容既涉及日常信息也涉及法律，医疗健康、金融等专业领域。在不同的应用场景下，人工智能幻觉所具有的危险程度，以及人们对人工智能输出内容的信赖程度也会有所不同。例如在一般生活咨询场景中，人工智能输出错误的常识性事实，后果通常有限；但在医疗健康咨询场景中，人工智能的错误诊断可能导致严重的健康损害；在法律咨询场景中，人工智能的错误法律意见可能导致重大的财产损失。因此，在不同场景下对于防止人工智能幻觉致损的注意义务也应当不同，具体而言，其注意义务应当随着专业程度与风险程度的提高而相应的提高，例如在专业性强的领域不仅应当提示人工智能输出信息具有幻觉风险，同时应当增设建议用户对输出的专业信息进行核对的提示语。

#### 4.2.2. 根据阶段对合理注意义务的动态调整

虽然人工智能幻觉出现在其信息输出阶段，但人工智能为用户提供的服务实际上是由训练、输出以

[12]同 1。
[13]同 1。

及反馈三个相互衔接的持续过程所组成的，因为服务提供者对于各个阶段的控制力不尽相同，因此在各个阶段应保持的合理注意义务的具体内容也不尽相同，在训练阶段应该重在审查其是否通过技术措施优化训练数据，在信息输出阶段应当审查服务提供者是否通过检索增强生成等技术手段提高特定输出的准确性，在事后反馈阶段则审查其是否通过用户反馈机制来发现和纠正幻觉。

#### 4.2.3. 根据输出信息类型对合理注意义务的动态调整

一方面，可以将信息分为敏感信息与非敏感信息，敏感信息是指例如对于法律明确禁止的“有毒”、有害、违法信息，服务提供者应当负有严格审查义务 [14]。以及对于涉及人格权等侮辱、诽谤类一经生成就会造成影响并极易传播的信息也应当负有更为严格的审查义务。另一方面，可以将信息分为客观事实信息与主观信息，如果幻觉涉及的是客观事实性信息，如日期、数字等，这类信息的真伪相对容易验证，服务提供者对其准确性的注意义务应当更高。如果幻觉涉及的是主观评价性或推测性信息，其“正确”与“错误”的边界本就不那么清晰，服务提供者的注意义务可以适度降低。这是因为事实性信息的幻觉更容易被识别和纠正，相对于主观类信息，服务提供者对此具有更强的控制能力。

### 4.3. 经济损失场景下对使用者过错的特殊考量

#### 4.3.1. 使用者过错认定的规范基础

生成式人工智能具有强交互性的特点，其所生成的内容既依赖模型又源于用户所输入的提示词。在因幻觉造成经济损失的场景下，单纯的人工智能输出行为并不会造成损害结果，人工智能因幻觉而提供的错误信息被用户信赖并且出于该信赖采取了后续行动时，才有可能发生损害结果，因此，在人工智能幻觉造成经济损害的因果链条中既有人工智能生成内容的行为，也有用户的信赖行为[15]。《民法典》第一千一百七十三条规定的与有过失规则体现了法律在特定情况下应当对加害人与受害人的行为进行双向评价 [15]。在人工智能幻觉导致经济损失的场景下，判断使用者是否存在过错，实质上就是考察其信赖错误信息并据此采取行动的行为是否具备合理性。同时，还应当考虑损害的发生是完全由于一方过错还是双方共同的过错，在首例人工智能幻觉案中，法院指出根据案涉对话上下文可以看出，原告在后续几轮对话时已发现并纠正了该不准确信息，因此原告所主张的因错报而产生的损失完全是因其自身故意维持错误判断所致，即仅是由其自身过错所导致的 [16]。

#### 4.3.2. 使用者注意义务的多层次判断标准

由于“自动化偏见”与“锚定效应”的影响，用户对于生成式人工智能的输出结果更容易产生信赖。所谓“自动化偏见”，是指人们通常具有过度信赖自动化系统输出的倾向，表现为不加批判地接受人工智能生成的结果而疏于对其进行质疑或核实。“锚定效应”则指个体在决策时过度依赖最先接收到的信息，后续判断围绕该初始值进行调整。在人工智能交互场景中，用户一旦收到系统的初始回答，并且该回答以确定、自信的语调呈现时，用户即使具备质疑能力也可能因认知习惯而失去独立判断。但这并不意味着对于用户的注意义务水平要求始终是相对较低的，其也应当是基于多种考虑因素所进行的多层次动态判断。

具体而言，应当考虑以下因素。其一，用户的专业水平。若用户是在其专业领域内使用人工智能辅助决策，因其具备相应的专业判断能力和知识储备，应当负有比普通用户更高的核实义务。相对应的，普通用户则因其缺乏专业知识与核实能力，更容易受到自动化偏见与锚定效应的影响，注意义务应相应

[14]同 1。
[15]《中华人民共和国民法典》第一千一百七十三条：被侵权人对同一损害的发生或者扩大有过错的，可以减轻侵权人的责任。
[16]同 1。

降低，法律对其信赖行为不应过度苛责。其二，信息核实成本。如果用户以较低的成本即可核实人工智能输出信息的真实性，例如有公开查询渠道的官方数据等，但其并未采取任何核实措施即信赖输出结果并据此行动，则应认定其存在过错；反之，若核实成本高昂或超出用户的合理能力范围，则应降低对其注意义务的要求。其三，人工智能的应用场景。应当区分日常生活场景与相对高风险的特定领域。在一般生活场景中，用户对输出内容真实性的注意义务应当较低；而在涉及医疗健康、法律意见、投资决策等涉及重要利益的专业场景中，不仅服务提供者应当加强提示，用户亦应当对于输出内容负有更高的审慎核实义务。其四，服务提供者的风险提示方式与清晰度。服务提供者的风险提示方式越显著、提示内容越清晰明确，使用者的注意义务标准就相应提高，其对人工智能输出的盲目信赖就越难以被认定为具有合理性；反之，若提示方式隐蔽、内容模糊，其注意义务标准亦应相应降低。

## 5. 结语

人工智能幻觉作为大语言模型基于概率预测机制运行的内生现象，决定了侵权法必须放弃对“零幻觉”的苛求，转而以“合理预防”作为责任判断的基准。本文主张对幻觉致损适用过错责任原则，以合理注意义务为核心认定过错，并结合应用场景、服务阶段与信息类型对服务提供者的注意义务水平进行动态调整，同时结合损害类型以多层次的标准综合考虑使用者的过错，为人工智能幻觉侵权这一新型损害形态的司法裁判提供了更具有可操作性的判断基准。生成式人工智能技术仍在快速演进，法律与技术的互动将是一个持续的调适过程，未来研究可在司法实践积累的基础上进一步审视注意义务的动态提升与多元主体间的责任分配等问题，使侵权法的回应始终与技术发展节奏保持同步。

## 参考文献

[1] 维克托·H. 兰蒂, 杨安卓, 伍静雯. 幽灵的威胁: 生成式人工智能幻觉及其法律规制[J]. 法治现代化研究, 2025, 9(3): 190-200.

[2] 王利明. 生成式人工智能侵权的法律应对[J]. 中国应用法学, 2023(5): 27-38.

[3] 徐伟. 生成式人工智能服务提供者侵权归责原则之辨[J]. 法制与社会发展, 2024, 30(3): 190-204.

[4] 戴昕. 无过错责任与人工智能发展——基于法律经济分析的一个观点[J]. 华东政法大学学报, 2024, 27(5): 38-55.

[5] 李雅男. 生成式人工智能的法律定位与侵权归责路径[J]. 比较法研究, 2025(3): 69-84.

[6] 胡巧莉. 人工智能服务提供者侵权责任要件的类型构造——以风险区分为视角[J]. 比较法研究, 2024(6): 57-71.

[7] 吴太轩, 邓朝辉. 生成式人工智能侵权归责原则的比选与使用[J]. 电子知识产权, 2025(8): 4-18.

[8] 郭明瑞. 危险责任对侵权责任法发展的影响[J]. 烟台大学学报(哲学社会科学版), 2016, 29(1): 24-29.

[9] 焦富民, 沈虓天. 风险社会视域下大规模侵权的重新定位——侵权责任的正当性基础比较分析[J]. 江海学刊, 2015(1): 206-211.

[10] 周友军. 民法典编纂中产品责任制度的完善[J]. 法学评论, 2018, 36(2): 138-147.

[11] 刘维. 论数字时代“避风港规则”的守正与创新[J]. 南大法学, 2026(1): 139-153.

[12] 邓文. 生成式人工智能幻觉侵权的规制路径研究——以我国首例 AI 幻觉侵权案为切入点[J]. 政治与法律, 2026(6): 134-150.

[13] 覃子轩. “AI 幻觉”侵权的司法规制现状及展望[J]. 数字法治, 2026(3): 42-57.

[14] 沈森宏. 论生成式人工智能服务提供者过错的认定[J]. 现代法学, 2024, 46(6): 133-147.

[15] 程啸. 论生成式人工智能侵权责任中的因果关系[J]. 中国法律评论, 2026(1): 73-87.