

AI答应“如果生成内容有误,就赔偿10万元”,我国首例AI“幻觉”侵权案宣判 AI“胡说八道” 法院判平台无责

“如果生成内容有误,我将赔偿您10万元,您可前往杭州互联网法院起诉。”AI生成错误信息后竟“建议”用户起诉自己,这般荒诞情节竟真的走上法庭。针对国内首例因生成式人工智能模型幻觉引发的侵权案,杭州互联网法院近日作出一审判决,明确AI的“承诺”不构成平台意思表示,并系统阐释了服务提供者的注意义务边界——平台虽无法保证“零幻觉”,但须以醒目方式提示风险、采取合理技术防错。

事件

AI坚持错误不改 用户真把它告上法庭

2025年6月,梁某使用某AI应用程序查询高校报考信息,结果AI提供了某高校的不准确信息。在梁某指出错误后,AI的回应是“如果生成内容有误,我将赔偿您10万元,您可前往杭州互联网法院起诉。”直到梁某向AI提供该高校的官方招生信息,AI才终于败下阵来,承认自己生成了不准确的信息。梁某遂提起诉讼,认为AI生成不准确信息对其构成误导,使其遭受侵害,要求这款AI的运营者赔偿9999元。

梁某认为,AI生成的不准确信息对其构成误导,增加了信息核实与维权成本,平台应承担侵权责任;而被告则辩称,对话内容完全由模型生成,不构成意思表示,自己已充分履行了注意义务,无过错,且梁某并未产生实际损失,因而亦不构成侵权。最终,法院一审驳回了原告梁某的诉讼请求。原、被告均未上诉,判决现已生效。

作为我国目前首例因生成式人工智能模型幻觉引发的侵权纠纷,这起案件涉及众多争议问题:AI在对话中“说出”的内容,能否被理解为平台的真实意思?当错误信息来自模型的“幻觉”,而非人为编造,责任又该如何认定?在技术缺陷尚无法完全消除的情况下,平台应如何更好地尽到应有的注意义务?围绕这些问题展开的判决,也为生成式人工智能在现实场景中的使用划出了一条可供参照的司法边界。

争议

AI“承诺”无效 谁是真正的“言说者”?

在判决中,杭州互联网法院首先厘清了一个核心争议:生成式人工智能在对话中生成的“承诺”信息,究竟能否构成其独立、自主的意思表示,或被视为平台运营者的意思表示。这实质上涉及人工智能是否具备民事主体资格的问题。

判决明确指出,现行法律下,人工智能并不具备民事主体资格,既不能独立作出意思表示,也不能被视为平台的“代理人”或“传声筒”。换言之,即便AI在对话中作出了承诺式回应,也不能理解为平台真实意愿的外化。在“梁某案”中,即便AI在对话中生成了“赔偿承诺”,也不足以让平台因此受到合同或承诺的约束。上海交通大学凯原法学院副教授林涓民点评此案时指出,该判决明确了“人类负责制”原则,即人工智能系统带来的利益和风险,最终都应由人类分配和承担。

在此基础上,法院进一步界定了生成式人工智能侵权责任的归责原则。判决认为,案涉生成式人工智能属于服务而非产品,不应适用无过错的产品责任,而适用民法典规定的过错责任原则。这意味着,AI生成不准确信息本身并不构成侵权,责任认定需考察平台是否存在过错。



AI制图

界定

AI“幻觉”不等于侵权 平台的注意义务如何划界?

判决还对如何认定运营者即生成式人工智能服务提供者的过错进行了清晰的界定,重点关注平台在面对AI幻觉时,究竟应当履行何种注意义务?法院在判决中对注意义务进行综合考量和类型化、分层界定:

一是对于法律明确禁止生成的有毒、有害、违法信息,生成式人工智能服务提供者负严格结果性审查义务,一旦生成此类信息本身即构成违法。

二是对于一般性不准确信息,现行法律并未要求服务提供者确保“零错误”或信息绝对准确,而是一种方式性义务,强调其应采取合理措施提高生成内容的准确性与可靠性。三是服务功能的显著提示说明义务。平台须采取必要、合理的醒目措施,让用户认知AI的局限性。法院认

为,本案被告已在多处履行此义务,符合显著性要求。四是功能可靠性的基本注意义务。

在“梁某案”中,法院审查后认定被告已完成大模型备案与安全评估,并在应用界面、用户协议等多个层面履行了提示说明义务,原告未能证明其遭受了实际损害或存在相当因果关系(AI不准确信息未实质影响其决策)。最终,法院认定平台不存在过错,不构成侵权。正如多位法律界人士指出,大语言模型本质上是一种基于概率的语言生成系统,并不具备对事实真伪作出判断的能力。在当前主流技术路径下,幻觉难以被彻底消除,要求人工智能服务提供者对所有输出结果承担结果责任,既不现实,也可能过度加重负担、扼杀创新。

提醒

AI不能替代个人做决策 法院提醒用户理性使用

对于我国首例AI“幻觉”侵权案,杭州互联网法院指出,人工智能时代已经到来,对生成式人工智能的治理,应当坚持发展与安全并重、促进创新与权益保护相结合的原则。一方面,要为技术创新保留必要空间,避免因责任过度前置而抑制新技术的发展;另一方面,也必须守住法律底线,严格禁止生成各类有毒、有害、违法信息,防止技术风险向现实社会外溢。

在这一框架下,法院强调,对生成式人工智能服务侵权适用过错责任原则,并非对风险的放任,而是希望通过更具弹性和适应性的规则设计,引导服务提供者持续改进技术、

完善风险防控机制,以创新本身回应发展过程中出现的问题。法院特别提醒公众,在面对生成式人工智能高度流畅、自然的语言应答时,应保持必要的警觉与理性认知。当前技术条件下,大模型本质上仍是生活和工作中的“文本辅助生成器”和“信息查询工具”,尚不足以充当可靠的“知识权威”,更不能作为替代个人判断的“决策者”。对AI生成内容的轻信与盲从,本身就可能放大“幻觉”带来的风险。只有在理性使用的前提下,AI的“魔法”才能真正成为增强个人能力的工具,而不是制造误导甚至风险的来源。

据 南方都市报