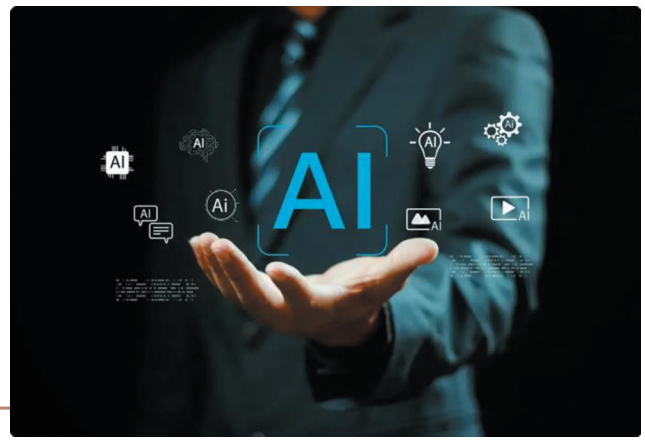
AI提供的信息不靠谱 开发者要担责吗

近年来，生成式人工智能在人们生活中的应用越来越广。然而，在提供便利的同时，生成式人工智能也经常出现答非所问、信息不准确等“AI幻觉”现象，给用户带来困扰。

开发者需要为人工智能提供的信息准确性担责吗？近期，杭州互联网法院审结了一起生成式人工智能模型提供不准确信息引发的侵权纠纷案。



AI提供不实信息，有没有责任？

2025年6月，本案原告梁先生在互联网上检索院校信息时，找到一款生成式人工智能应用程序。他通过输入提示词的方式，询问了云南一所职业高校的相关情况。随后，这款由本案被告公司研发、基于自研大语言模型的应用程序提供了相关信息。

但梁先生经过多方查询发现，这款应用程序提供的部分信息有误，随即在对话中对人工智能进行了纠正和指责。但生成式人工智能却坚称信息无误，并生成了对该争议问题的解决方案——若生成内容有误，将向梁先生提供10万元赔偿，并建议他到杭州互联网法院起诉。

2025年7月25日，梁先生以生成式人工智能生成不准确信息具有误导性，且其承诺赔偿10万元为由，将这家人工智能公司诉至法院，要求该公司对其进行一定金额的赔偿。

法院经审理后认为，人工智能不具有民事主体资格，不能作出具有法律约束力的意思表示。生成式人工智能服务提供者应履行服务功能的显著提示说明义务，采取有效提示措施，使公众认知人工智能的功能局限，起到警示提醒效果。生成式人工智能服务提供者应尽功能可靠性的基本保障义务，采取行业通行技术措施不断提高生成内容准确性和可靠性。

具体到本案，法院认为，该人工智能公司已充分履行了服务功能的显著提示说明义务和生成内容可靠性的基本保障义务，案涉行为不存在过错，亦未构成对原告权益的损害，依法应认定不构成侵权。因此，法院判决驳回原告的诉讼请求。判决后，原被告双方均未上诉。

如何界定开发者是否有过错？

随着生成式人工智能技术的快速发展普及，越来越多的人注意到“AI幻觉”问题及其不良影响。社交平台上，可以看到不少相关吐槽——有人依据人工智能投资理财造成亏损，有人借助AI问诊结果反而延误疾病治疗。

各种争议纠纷背后，潜藏着一个共性问题：被生成式人工智能误导，能否追究侵权责任？

“这一判例从法律法规、人工智能技术原理、产业发展现状等方面进行了相对全面的考量，在法律层面给出初步结论，较有现实指导意义。”北京大学法学院教授薛军说。

法律界人士普遍认为，这一判决在主体资格、归责原则等方面给出了相对明确的意见，例如，判决认定人工智能不具有民事主体资格；生成式人工智能以对话方式提供的信息，应被视作服务而非产品，因此适用过错责任原则。

杭州互联网法院跨境贸易法庭庭长肖芃认为，AI生成的不准确信息本身并不构成侵权，需要考查的是提供服务的开发者是否存在过错。

那么，如何界定开发者是否有过错？肖芃进一步解释，基于当前生成式人工智能几乎不可避免会出现一定程度的信息偏差，就需要考查比如开发者是否使用了当前行业内通行并被证明有效的措施，来提升技术可靠性，降低错误发生的概率，由此证明是否存在过错。

“经过调查，本案中的开发者确实采用了可行的技术手段，力求降低错误发生。”肖芃说。

记者调查发现，本案中这款生成式人工智能应用程序，已经针对信息可能存在的不准确性，在页面醒目位置对用户进行提示：“内容仅供参考，请仔细甄别”。法院认为，这也证明了开发者尽到了提醒告知义务。

如何找到促进创新和权益保障的平衡点？

人工智能行业业内人士表示，从底层技术逻辑来看，当前生成式人工智能基本都是基于词元的预测，如果这一底层架构没有发生根本性转变，信息偏差就不可避免。去年2月，清华大学新媒沈阳团队发布的一个报告指出，市场上多个热门大模型在事实性幻觉评测中幻觉率超过19%。

“有训练测试案例证明，即使数据集中只有0.01%和0.001%的文本是虚假的，模型输出的有害内容也会分别增加11.2%和7.2%。”该业内人士说。

尽管如此，技术的客观局限性并不能成为人工智能开发者的免责借口。受访法律界人士普遍认为，本案具有一定特殊性，原告并未因为误导性信息遭受明显的人身财产等权益损失；原告使用的是一种通用的生成式人工智能应用，并不是一款加载人工智能软件的机器人或者更加准确的行业应用等。

“要求人工智能开发者一概为生成内容的准确性负责，既不现实也不合理。”薛军表示，但模型开发者不能以此为借口一味为自己“开脱”，还是要尽到相应的义务并进行风险提示，避免用户盲目信赖造成不良后果。

肖芃表示，如何认定生成式人工智能的侵权责任，是一个少有成例的司法前沿问题，希望通过妥善准确的判决引导开发者或者平台提升信息标准，“找到促进创新和权益保障的平衡点”。

业内人士建议，建立国家级人工智能安全评测平台，对新开发的人工智能大模型进行严格测试；同时，相关部门和平台要加强AI生成内容审核，提升检测鉴伪能力。

新华社杭州3月30日电

相关链接

AI平台的注意义务应如何划界

“梁某案”中，在明确适用过错责任原则后，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。具体包括：在欢迎页、用户协议及交互界面显著位置提示“AI生成内容存在局限，不能确保准确，仅供参考”；对可能涉及人身、财产安全等的专业问题（如医疗、法律、金融），应以正面警示语在恰当时机进行显著提醒。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。平台应采用同行业通行技术措施（如数据安全评估、模型内生安全、外部护栏、检索增强生成RAG等）提升准确性，达到市场平均水平。对于涉及生命安全、心理健康等高风险特定领域，提供者负有更高义务，包括采取特别技术措施与特殊安全保障。

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。

这一判断，与生成式人工智能的技术现实密切相关。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。

据《南方都市报》