

全国首例生成式AI“幻觉”引发侵权之诉

杭州互联网法院:为生成式人工智能服务厘清责任边界

□ 本报记者 宋朵云 本报通讯员 肖莧 吴永安

导读

生成式人工智能技术以其强大的内容生成能力,正深刻改变着信息的获取与交互方式。然而,AI并非全知全能,其生成的答案可能包含看似合理实则错误的“幻觉”信息。当AI生成的不准确信息误导他人时,是否构成侵权?近日,杭州互联网法院审结全国首例因生成式人工智能模型“幻觉”引发的侵权纠纷案,直面这一技术应用前沿的法律挑战,对AI生成内容的性质、服务提供者的责任边界以及侵权构成要件等核心问题进行了深入阐释,为类似纠纷处理提供了重要的裁判指引。



扫码观看视频

AI“承诺”生成内容有误将赔偿

某生成式人工智能应用程序(以下简称某生成式人工智能应用),是某科技公司基于自研大语言模型,开

发并运营的文本生成、信息查询类通用型智能对话应用程序。2025年3月,梁某在同意用户协议后,注册并开始使用

某生成式人工智能应用。2025年6月29日,梁某在某生成式人工智能应用中输入提示词询问某高校报考的相关信息时,某生成式人工智能应用生成了关于该高校某校区的不准确信息。梁某发现后,在对话中进行了纠正和指责,某生成式人工智能应用仍继续回复称该高校确实存在这一校区,并生成了对争议问题的解决方案,提出若生成内容有误将向用户提供10万元赔偿。之后,梁某将从该高校官网查询到的招生信息提供给某生成式人工智能应用,此时某生成式人工智能应用承认其生成了不准确信息,并建议梁某到杭州互联网法院起诉索赔。

梁某认为,某生成式人工智能应用生成不准确信息对其构成误导,使其遭受侵害,并承诺对其进行赔偿,遂起诉要求某科技公司赔偿损失9999元。

某科技公司辩称,对话内容由人工智能模型生成,不成立意思表示;某科技公司已充分履行了注意义务,无过错;原告未产生实际损失,某科技公司不构成侵权。



赵媛媛 柯燕婷 作

AI可否独立作出意思表示

生成式人工智能生成的“承诺”信息是否构成人工智能独立、自主的意思表示?是否可视为被告某科技公司的意思表示?法院认为,人工智能不具有民事主体资格,不能作出意思表示。案涉AI自行生成的“赔偿承

诺”亦不能视为服务提供者(被告)的意思表示,理由在于:第一,人工智能不具有民事主体资格,不能作为意思表示的传达人、代理人或代表人;第二,被告并无通过该AI模型这一工具来设定或传达其意思表示的行为;第三,一

般的社会观念、交易习惯等尚不足以使原告对该随机生成的“承诺”产生合理信赖;第四,无证据表明被告曾作出愿受人工智能生成内容约束的外在表示。因此,该“承诺”不产生意思表示的法律效力。

AI侵权应适用何种归责原则

法院认为,生成式人工智能服务依据《生成式人工智能服务管理暂行办法》属于“服务”范畴,而非产品质量法意义上的“产品”。本案应适用《中华人民共和国民法典》第一千一百六十五条第一款的一般过错责任原则,而非产品责任

的无过错责任原则。主要基于四点考量:其一,从概念与构成要件上,该服务缺乏具体、特定的用途与合理可行的质标准;其二,其生成的信息内容本身通常不具备民法侵权责任编所指的高度危险性,通常情况下不宜对信息内容本身采用无过错

责任原则;其三,生成式人工智能服务提供者缺乏对生成信息内容足够的预见和控制能力,对生成信息内容不宜适用产品责任;其四,从政策导向看,适用无过错责任原则可能会不当加重服务提供者的责任,限制人工智能产业的发展。

AI服务提供者是否构成侵权

在案涉情形下,法院基于一般过错责任原则,对侵权构成的各项要件进行了逐一审查。

首先,关于侵权行为的问题。原告主张受到的侵害是因信息不准确,致其受误导错失报考机会,额外产生信息核实成本、维权成本等,即纯粹经济利益被侵害,而非人格权、物权等绝对权被侵害,因此,不能仅依据权益本身被侵害而认定行为的非法性或不法性,而须从被告是否违反注意义务进行判定。

其次,关于过错的认定。生成式人工智能技术还处于高速发展的过程中,其应用场景亦具有很强的泛在性,故服务提供者的注意义务处在一

个动态调整的框架之中。法院采用动态系统论,阐述了服务提供者应尽的三层注意义务:一是对法律禁止的“有毒”、有害、违法信息负有严格审查义务;二是需以显著方式向用户提示AI生成内容可能不准确固有局限性,包括明确的“功能局限”告知、保证提示方式的“显著性”、在重大利益的特定场景下进行正面即时的“警示提醒”,以防范用户产生不当信赖;三是应尽功能可靠性的基本注意义务,采取同行业通行技术措施提高生成内容准确性,比如检索增强生成技术措施等。经审查,被告已在应用程序欢迎页、用户协议及交互界面的显著位置呈现AI生成内容功能局限的提醒标识。结

合被告已采用检索增强生成等技术提升输出可靠性的事实,法院认定其已尽到合理注意义务,主观上不存在过错。

最后,关于损害结果与因果关系。原告主张其因信息误导而错失报考机会并产生额外成本,但未能就此实际损害的发生提供任何有效证据,依法难以认定损害存在。进一步分析因果关系,法院采用相当因果关系标准,认为案涉AI生成的不准确信息并未实质性介入或影响原告的报考决策过程,二者不存在因果关系。

综上所述,被告的案涉行为不具有过错,未构成对原告权益的损害,依法不应认定构成侵权。最终,法院驳回了原告的诉讼请求。原、被告均未上诉。判决现已生效。

法官心语

为AI发展铺平法治轨道

杭州互联网法院跨境贸易法庭庭长 肖莧

人工智能时代已经来临,应当坚持发展与安全并重、促进创新与权益保障相结合,既要鼓励生成式人工智能服务的创新发展,也不能忽视当事人合法权益的保护。当下存在一种观点,主张将产品责任的适用对象扩大至虚拟系统,对生成式人工智能侵权纠纷适用产品责任的无过错责任原则。然而,传统产品的物理形态、功能边界和风险范围是明确且相对固定的,生成式人工智能生成的内容是算法基于海量数据、对用户开放式指令的即时响应,其每一次输出都是不可完全复现的独特生成。对生成式人工智能侵权纠纷适用产品责任,缺乏法律依据和法理基础。同时,人工智能的研发和应用具有较高的公共价值属性,如果在技术发展初期对其苛以过重的严格责任,可能产生“寒蝉效应”,导致

某些创新技术无法落地应用,减缓甚至阻碍技术的创新和发展。对生成式人工智能服务侵权采取过错责任原则能够全面评价服务提供者的行为,有利于建构兼具弹性、适应性与演化能力的回应型规则治理体系:既设立红线、底线,严格禁止生成各类“有毒”、有害、违法信息,又激励生成式人工智能服务提供者采取合理措施进行技术创新,防止风险发生。通过对服务提供者有无过错的判断形成对相关行为的理解与共识,在动态回应的过程中不断寻找系统、合理的注意义务标准,不断演化形成新的标准与规范,最终实现技术创新和权益保障的平衡。

在当下的人工智能时代,模型通过在含有数万亿词元(Tokens)的庞大语料库上训练,学习了复杂的语言统计模

式和词汇共现规律,但并未真正习得“事实”,因此难免产生“幻觉”。需要特别提醒,社会公众在面对生成式人工智能高度流畅、自然的语言应答时,应提高警觉,清醒认识到大模型在当前技术条件下仅是人们生活、工作的“文本辅助生成器”和“信息查询的辅助工具”,尚不是可靠的“知识权威”,更不能作为“决策替代者”,绝不应轻信、盲从,而需多方验证,审慎决策。在人工智能时代,应通过案例宣传与科普教育,揭示AI“幻觉”的形成机制,使社会公众更加理性地认知AI的功能和局限,在全社会广泛培养人们对AI自动生成内容的基本怀疑精神与核查能力,既充分利用好AI的创造性潜力,同时又避免“幻觉”带来的风险。

专家点评

司法实践为AI治理提供法理支撑

清华大学法学院教授、博士生导师 程啸

本案是我国目前第一起因生成式人工智能模型“幻觉”引发的侵权纠纷。审理本案的杭州互联网法院在充分权衡民事权益保护与鼓励促进人工智能技术发展基础上,依法对上述争议问题作出了正确的裁判,具有极为重要的理论意义和实践价值。

第一,所谓人工智能可否独立作出或代替服务提供者作出意思表示的问题,其实质就是人工智能是否属于独立的民事主体的问题。理论界曾有人提出应当承认并赋予人工智能以法律人格,将其作为独立的民事主体的观点。然而,这种观点无论是在现实中,还是依据当前法律的规定,都是不能成立的。本案判决依据我国民法典关于民事主体以及意思表示的规定,正确地指出:人工智能既不是生物学意义上的人,也并未被我国现行法律赋予民事主体资格,故而不是民事主体,不具有民事权利能力、行为能力和责任能力。同时,考虑到社会一般观念、交易习惯以及人工智能服务提供者已经在用户协议中明确表明生成内容并非自身意思表示等因素,法院在判决中也否定了人工智能可以代替服务提供者作出意思表示,即生成式人工智能不能作为服务提供者意思表示的传达人、代理人或代表人。

第二,生成式人工智能侵权责任的归责原则问题是目前理论界争议很大的一个问题。理论界有人主张适用过错责任原则,有人则将生成式人工智能视为产品,认为应适用产品责任的无过错责任原则。对此,法院依据我国产品质量法、《生成式人工智能服务管理暂行办法》等法律法规中关于产品、生成式人工智能服务的规定,并考虑到生成式人工智能所生成的信息内容不具有民法侵权责任编所指的高度危险性、国家促进人工智能发展的产业政策等因素,正确认定了案涉生成式人工智能属于服务,而非产品,依法不适用产品责任,而应适用民法典第一千一百六十五条第一款的一般过错责任原则。这一认定对于今后处理人工智能侵权纠纷时,法院准确适用侵权责任编的归

责原则具有重要的启示。

第三,在生成式人工智能侵权责任适用过错责任原则前提下,如何科学合理认定生成式人工智能服务提供者的过错,也是本案判决呈现的一大亮点。在判决中,法院依据我国现行法对生成式人工智能服务的规定,区分了不同类型的注意义务并提出了相应的认定标准。首先,服务提供者对于确保生成式人工智能不生成法律明确禁止生成的各类“有毒”、有害、违法信息负有的是结果性义务,即一旦生成违法信息就应当认定违法,服务提供者存在过错。其次,生成式人工智能服务提供者应当依法履行对服务功能的显著提示说明义务,即采取必要、合理的显著提示措施,使得用户能够了解、认知服务功能的局限性,并在特定情形下起到警示提醒的作用。再次,就所生成内容的准确性而言,服务

提供者负有的不是结果性义务,不能因为生成内容不准确就当然认定服务提供者存在过错,而是一种方式性义务,即应当在综合考虑现有的技术措施水平、生成内容对当事人合法权益的影响、当事人是否开启联网搜索功能选项等多种因素的基础上,确定服务提供者合理的注意义务。

应当说,本案判决就服务提供者注意义务所做的上述类型化区分及提供的相应判断标准,既充分考虑了民事主体合法权益的保护与社会公共利益的维护,也注意到生成式人工智能的自主性特点、目前技术措施下模型“幻觉”尚无法全部消除等客观现实,充分发挥了民法典侵权责任编的风险治理功能,对于未来我国法院审理类似的案件具有极大的参考价值。

