

AI“幻觉”误导他人是否构成侵权

全国首例纠纷案宣判

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。

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。

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检标准;其二,其生成的信息内容本身通常不具备民法典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之中,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及时的“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地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。

(中国普法)