当如何应对模型『幻觉』挑战？AI生成的诉状呈上法庭

N南方都市报

3月下旬的一次庭审中，身为法官的高铭（化名）收到案件当事人提交的一份特殊诉状。初看之下，这份诉状的内容写得头头是道。后经核查，高铭高度确信该诉状是由AI生成，“特别能体现AI幻觉的特征”。高铭表示，当律师提交了AI生成的有差错诉状，法官可以向律协发送司法建议函；但如果是案件当事人犯下此类错误，除了无奈一笑外别无他法。

AI幻觉
提及的白皮书均属虚构 案件序号都写错

这是长三角地区某法院在审的一起著作权侵权案件。高铭介绍，当事人出示的诉状表面上条理非常清晰且规范，将事实理由拆分成多个部分依次阐释。每部分内容里，先是列明法条规则，再摆上诸多司法案例，并穿插大量行业白皮书。

“哪怕是专业律师代理，也很少能把证据说得这么全面。”高铭提到，诉状中罗列了七八个案例，涵盖最高法、高院、基层法院等不同层级法院的判决书。

但在对援引的法条、案例和白皮书逐一核对之后，高铭发现，除了几项著作权法的法条确实存在外，其余提及的规定、案例和白皮书均属虚构。

法院的判决书编号由法院代码、案件序号等元素组成。上述案件诉状中给出的案例文书，其法院代码准确无误，但案件序号却全是诸如“12345”这样的连续数字。

案件当事人并未承认使用了AI来生成诉状，高铭分析背后有两种可能性：一种是当事人谎称有律师撰写，实则自己用AI生成；另一种则是当事人寻求过在线法律咨询服务，咨询平台不负责任地反馈了一份AI生成的文书。

“人一般不会这么理直气壮地编造出头头是道的内容。”高铭表示，把假的当作真实的加以引用，高度符合AI幻觉的特征。

律师引用AI生成虚构案例 两年间美国发生七起

域外法院同样遇到过AI生成错误诉状的情形。据路透社2月中旬报道，美国知名人身伤害领域律师事务所Morgan&Morgan向其1000多名律师发送了一封紧急邮件，警告AI可能编造虚假的判例法，若律师在法庭文件中使用这些虚构内容，可能会被解雇。

这份警告邮件事出有因——该律所两名律师在一起针对沃尔玛的诉讼中，引用了AI生成的虚构案例。而在过去两年间，此类事件在美国至少发生了七起。

发展趋势
有平台开通AI写文书模块 用户输入信息生成法律文书

“有的律师可能会误以为，AI现在已经很强大，输出后就不用审核了，这显然是不正确的。”常成是法律AI公司华宇元典旗下“元典问达”产品的运营负责人，据其介绍，为避免用户过分信赖AI生成结果，“元典问达”会在页面上提醒用户，AI生成的内容仅供参考。

当前的法律AI工具市场上，除了通用型AI聊天产品可被用于法律咨询，像华宇元典这样的公司也在推进生成式AI在法律垂直领域的应用。“元典问达”平台开通AI写文书模块，可以根据用户输入的信息，生成民事起诉状、律师函、法律咨询意见书等多种类型的法律文书。

根据记者的实际体验，若要生成一份涉及民间借贷案由的民事起诉状，用户首先需要在“元典问达”平台输入案件情况概要。随后，系统会自动分析并提取案件的关键要素，包括原告与被告的身份信息以及具体的诉讼请求等内容。在用户确认这些初步信息后，AI将自动生成一份格式规范的诉状，把案情扩充讲述得更详细，同时援引了涉及民间借贷案件的法律法规作为支撑。如果不满意，用户可在此基础上进一步修改再提交重新生成。

当事人利用AI 参与诉讼过程难以避免

法律AI产品不仅面向专业律师，更覆盖有法律服务需求的普通民众。深圳市律协2023年7月下发的一份文件显示，2022年全国法院受理各类案件3372万多件，律师代理的诉讼案件数量只有824万多件，占比仅约为24.4%。常成说，基于这样的司法现状，即便AI存在出错的可能性，但对于未聘请专业律师的当事人而言，他们依然会利用AI工具参与到诉讼过程中，这一趋势难以避免。利用AI工具撰写诉状，可能很多情况下要好于当事人自行撰写的质量。

但常成提醒，法律AI产品仍然只能作为辅助提效的工具，使用者自己要担负起责任，自己要检查和最终审核AI生成的文本。

规范纠错
法律AI公司精调训练 限制底层基座大模型随意发挥

一些法律AI公司通过精调训练的手段，限制底层基座大模型的随意发挥。

武汉百智诚远科技有限公司（下称“百智诚远”）推出了一款“起诉状辅助生成系统”，目前主要面向G端（政务端）市场。该公司产品总监孙莉莉介绍，由于通用大模型难以直接落地到具体行业场景，公域部署的诉状生成系统在DeepSeek和阿里的通义法睿等大模型基座模型的基础上，必须再灌注一些法律行业专用数据展开进一步的精调。百智诚远正在和一些G端客户合作开展诉状生成系统的内测，用户与AI交互后，AI便可生成符合规范格式的诉状草稿。

孙莉莉表示，在处理如法条引用和案件金额计算等高度专业化的任务时，这种叠加式的训练方法显得尤为重要。例如，在劳动争议案件中，精调后的法律AI模型能够根据当事人所在地的工资标准，自动计算出相应的赔偿额度。如果交由通用大模型执行相同任务，幻觉问题可能比较严重。

不过，受限于训练数据量的丰富度，不同难易程度案件的诉状生成效果有差异。孙莉莉以知识产权案件举例说，对于外观设计专利侵权的案件，AI生成的结果质量较高；而对于其他更复杂的知产侵权案件类型，AI撰写的诉状可能只有部分可用，无法达到专业律师的成文水平。

律师提交AI生成差错诉状 法官可向律协发送司法建议函

面对AI逐步渗入诉讼场景，如何规范其正确使用，已然成为不同司法辖区面临的共同课题。

为此，2024年7月，美国律师协会发布的第一份关于律师使用生成式AI的正式意见指出，使用AI可能导致律师引用不存在的案例或不准确的分析，该意见警告律师应审查所有输出内容的准确性。

同年9月，新加坡最高法院、家事法庭、州法院同时发布一份《生成式人工智能工具使用指南》，要求包括案件当事人、代理律师在内的案件相关人员在使用AI时，必须确保生成内容的准确性与真实性、没有创造虚假案例，更不应用此途径来生成或编造证据。若发现有不当使用情况，法庭可以对相关使用者进行惩戒。

目前，中国尚无类似的专门指引。高铭表示，当律师提交了AI生成的有差错诉状，法官可以向律协发送司法建议函；但如果是案件当事人犯下此类错误，除了无奈一笑外别无他法。