警惕DEEPSEEK-R1的法律幻觉

2025-08-02 翟冠慧

直观地看，对咨询者所提出的法律问题， DEEPSEEK-R1（“R1”）会按照探寻法律依据、归纳司法实践、确定应对策略、提出处理建议和注意事项的套路给出言之凿凿的答案。

这一程式化的流程，让使用者容易产生一种错觉：R1具备拆解和分析复杂法律问题的能力；它能够像人脑一样进行信息整合和逻辑分析；它通过快速检索、筛选信息，在断断数十秒钟的时间内即能完成对复杂法律关系的梳理进而给出建议和解决方案，简直可以完美契合法律行业的需要。

正是因为R1 展现出了非凡的“思考”能力和流畅对话的能力，它被国内外各大平台、各类垂直应用广泛接入。在R1等大模型的助推下，中国法律行业自2025 年起，也全面进入了AI 赋能时代。以北大法宝、威科先行等为代表的法律检索工具平台，通过整合AI 技术，显著提升了法律检索、文书生成等任务的效率。

但是，必须提出的是，R1的推理泛化能力的代价之一，便是幻觉（HALLUCINATION）问题突出。在复杂推理与专业领域，**R1的中高风险领域是法律建议、医疗诊断、金融投资建议。**

在AI 评估组织Vectara 今年3 月4日公布的大语言模型（LLM）幻觉排行榜上，R1的幻觉率达到了14.3%，远高于其V2.5 版本的2.4%。而Google Gemini-2.0-Flash-001、Google Gemini-2.0-Pro-Exp、OpenAI-o3-mini-high-reasoning的幻觉率都低于1%。GPT-4o 的幻觉率为1.8%。阿里Qwen2.5-7B-Instruct 的幻觉率则为2.8%。14.3%的幻觉率意味着R1平均每7次生成就会有1次幻觉。

有论者将R1的法律“幻觉”现象划分为两类，直接引述如下：

一是内容性“幻觉”，即生成虚构、失实的法律主张。例如编造不存在的《民法典》条款（如声称“第1321 条规范虚拟财产继承”），或误述最高人民法院已废止的司法解释效力（如将1998 年《关于审理盗窃案件具体应用法律若干问题的解释》作为现行依据）。此类错误直接违反法律规范的确定性原则，属于“硬性幻觉”。

二是关联性“幻觉”，即法律依据与主张间的逻辑断裂或效力误植。典型表现有：援引普通案例支撑“应当参照”类裁判规则（如用基层法院判决论证高级法院的法律适用）；混淆部门规章与行政法规的效力层级（如将《网络交易监督管理办法》等同于《电子商务法》适用）；或误用域外法理分析中国本土法律问题（如以美国“长臂管辖”原则解释我国民事诉讼管辖规则）。此类错误触及法律适用的正当性基础，构成“软性幻觉”。

可见，R1等大模型的AI幻觉涉及编造法律条文、引用虚假案例，还涉及对判例效力的误判，对法律事实的误解等。此类错误如不加甄别直接加以采用，会冲击以严谨为生命的法律实践，并且还可能会造成严重的法律后果。

特别需要指出的是，相关研究还表明，“幻觉”问题具有不可根除性。因此，在使用R1等AI工具时，**尤其是针对所谓的“硬性幻觉”问题，要进行人工复检，谨慎使用。**

地址：中国上海市徐汇区漕溪北路18号上海实业大厦38层
电话：+86 136 3641 6891
邮箱：admin@xinbenlaw.com
传真：+86 21 5466 3852