

# 国内首例AI“幻觉”案，给我们提了个醒

## AI“幻觉”会引发哪些潜在风险？

此次杭州互联网法院处理的案件并非孤例。近期国外一起针对AI聊天机器人Grok的诉讼中，原告指控Grok在回答用户提问时，输出了损伤自己公信力的不实答案。越来越多的事例表明，AI“幻觉”已成为AI使用过程中的重要潜在风险。

AI“幻觉”的风险等级取决于其实际应用场景。清华大学公共管理学院院长、科技发展与治理研究中心主任朱旭峰指出，AI“幻觉”的危害，和应用场景息息相关。现在AI“幻觉”问题还主要集中在文字内容上，但随着大模型的逐步普及，AI被用于其他场景，生成的不只是文字，还包括软件、算法、程序甚至决策的信号，其影响力与日俱增。在这样的背景下，AI生成的错误结果如果不加识别地用于法院判决、医疗诊断、自动驾驶等场景，将有可能造成严重的后果。

## 如何探索有效的AI“幻觉”治理之道？

在当前的技术发展阶段下，AI“幻觉”或许难以避免，人类该如何与AI“幻觉”共处？

朱旭峰认为，所谓“幻觉”问题从某种程度上与

信息茧房、大数据“杀熟”以及数字鸿沟等问题类似，其实自古以来就存在，只是如今被信息技术赋予了现代意义。我们应辩证地、更加全面地看待AI“幻觉”问题，不要把“幻觉”归罪于AI技术本身，更不能因噎废食。相反，正因为目前AI技术存在包括“幻觉”等不成熟的技术问题，我们才要更多地使用和训练它，让它变得更加成熟，更好地为人类服务。

另外，治理AI“幻觉”也离不开法治的护航。对此，王锡铎表示，用法治治理AI“幻觉”的关键是把“技术不可根治”难题转化为“规则可验证、责任可追溯”。他建议，一方面要加快建立强制标识与可追溯标准，运用显性及隐性标识、元数据留痕与跨平台识别等技术手段降低误认与扩散风险。另一方面，要建立分类分级监管。比如，设立注意义务的分层标准，对违法有害信息“严格审查”，对一般错误信息以“显著提示+纠错机制”为核心，形成可操作的“注意义务清单”。针对AI在教育、医疗、金融、司法等场景可能存在的使用风险，平台及有关各方应制定更严的事实核验、人工复核与禁用边界。唯有构建起明晰的权责边界，才能引导AI产业在规范中创新发展，让技术红利更多惠及千家万户。 据新华社



2026年1月6日，在美国拉斯维加斯消费电子展上，人们参观展区中的机器人。新华社发 曾慧 摄



近日，国内首例因AI“幻觉”引发的侵权案件宣判。案件中，一位高考生的亲属梁某在查询高校信息时，发现AI平台生成的信息存在错误。当他以此质疑AI时，AI却回答“如果生成内容有误，我将赔偿您10万元”。一气之下，梁某将AI平台的研发公司告上法庭，要求其赔偿。受理此案的杭州互联网法院作出一审判决，驳回诉讼请求。

AI“幻觉”，通俗讲即所谓AI一本正经“胡说八道”的问题，主要指大模型生成的相关内容和真实信息不一致的现象，本质上是由于大模型的训练设计缺陷、数据不足以及架构偏差等因素造成的。那么，谁应该为AI“犯错”负责？怎样为生成式AI服务厘清责任的边界？

## 此次判决结果传递出怎样的信号？

AI是否需要为它说出的话负责，是这起案件中法庭首先要厘清的问题。北京大学法学院教授、北京大学数字法治研究中心主任王锡铎表示，生成式AI的输出属于概率生成，并非法律意义上的意思表示，因此不能把AI模型的每句回答简单转化为具有法律效力的承诺。

当然，宽容AI技术的局限，绝不意味着放任自流。AI不对每句话承担“绝对责任”，不代表AI运营和服务平台无需承担“治理责任”与“合理注意义务”。据悉，在此次案件中，研发公司已采取了当前技术条件下的必要措施，以减少AI“幻觉”的发生，并履行了提示说明义务。因此，法院依法驳回了原告的诉讼请求。

王锡铎认为，法院此次判决并未将AI“幻觉”简单等同于平台侵权，而是以过错责任为主轴，划定生成式AI服务的责任边界。责任从“结果担保”转向了“风险控制”，重点审查服务提供者是否履行提示、纠错、标识与安全评估义务。法律要建立的不是“句句负责”，而是“可控、可追责的治理结构”。因此，这一裁判实上传递出“鼓励负责任创新、反对无限责任”的信号，既不放任AI进入“法外之地”，也不过度扩张法律责任可能带来的对AI产业发展的抑制。

# 垃圾不落地 古城更美丽

从我做起 人人参与 全民动员 美化环境

