生成错误回答？提供虚假数据？编织事实来源？习惯有问题找AI的我们要警惕啦——

无所不能的AI也会犯错 但它从不会说“不知道”

近日，市民张女士透露，她在某日去海珠湿地公园玩之前，想知道湿地公园有没有地铁站。张女士把这个问题抛给了“豆包”，“豆包”一开始表示湿地公园的北门和南门都有地铁，但在张女士“气愤”地指出错误后，“豆包”终于给出正确回复，表示北门有地铁，而南门没有。

为什么“豆包”不能立即给出正确答案？事实上，这不是“豆包”独有的问题，DeepSeek、元宝、千问、ChatGPT、Gemini、Grok等国内外常见的大模型都存在这个现象。这个现象甚至还有一个专有名词——AI幻觉。

很明显，AI虽然已经融进了我们的生活，但它还远远没到可以取代人类的地步。AI有幻觉，但人类自己得清醒。

■采写:新快报记者 王敌 梁潇静

天体中心属于哪个街道？DeepSeek两次回答错误

关于AI幻觉，广州的短视频制作者小乐提起过一件事。某次，小乐去天河体育中心拍摄视频，她很想确定天体中心属于哪个街道，于是她打开DeepSeek提问：“体育中心属于林和街道管辖吗”？很快，DeepSeek给出回复，明确表示天体中心属于林和街道。

随后，小乐又问：“体育中心属于天河南街道管辖吗”？然后，DeepSeek再次迅速回复，称“天河体育中心主体部分属于林和街道，南部边缘区域可能涉及天河南街道”。事实上，DeepSeek两次给出的答案都不对。小乐在询问林和街道的工作人员后得知，天体中心完全归属天河南街道管辖，林和街道的地图上明确显示不包含天体。

社交平台上类似吐槽很多。有网友问AI：“椰子内壁变紫色还能吃吗？”有的AI说不建议吃，因为氧化变质；有的AI说是正常现象，可以吃。

AI在专业领域也会犯错。研究生小岚做课程作业时常用ChatGPT检索参考文献，核实后发现有些论文根本不存在。正因如此，每次使用AI查找资料时，小岚都要进行二次核查，以防AI“捏造”根本不存在的文章。“当我指出错误并让它给出引用文献的链接时，就能有效减少错误频率。”小岚说。

AI不会说“我不知道” 央视315曝光数据污染

“极客”小林是计算机专业的学生，他说，现在大众能接触到的AI大模型大多具有服务属性，是因为开发商给AI立了“服务员设定”。小林表示，AI没有认知和情感，不会意识到回答是否正确。“对不便回答的问题，它会说无法回答；你指出错误，它还会道歉——但这不意味着它真认为自己错了，只是被设定了回应模式。”小林说。

正因于此，用户在使用过程中会发现，无论AI对用户提出的问题是否了解，它都会给出一个看上去“很有道理”的回答，而从来不会说“我不知道”。“这也是由AI的运行方式决定的，它会基于自己能搜索到的信息和见过的模式，生成看起来最可能的内容。”小林说。

今年央视315晚会曝光了利用GEO（生成式引擎优化）“污染”AI的黑色产业。AI会根据网上看似有逻辑、实则“夹带私货”的虚假内容生成非客观答案。虽然这种“信息污染”以前就有，叫SEO（搜索引擎优化）——让垃圾信息被搜索引擎检索到，但如果说SEO是“路边小广告”，GEO就是广告升级版——借用AI做软文营销。央视曝光的违法GEO，就是通过机器批量生产伪造内容上传到互联网，给AI“洗脑”，误导消费者。

AI“有用性”权重过高 要驾驭AI就要比AI更懂

去年从华东理工大学人工智能专业毕业、现从事前沿科技的小宝认为，AI不说“不知道”的最关键的原因在于：研发方不希望这个大模型被视为“没用的”，致使“有用性”权重过高。小宝说：“AI对‘有用性’的追求经常压过对‘真实性’的追求，导致AI在不确定时依然努力给出看似合理的回复，这就是‘AI幻觉’。”

AI模型的逻辑是基于数据训练，而准确率取决于训练数据的质量和完整性，一旦数据采样不同，或者是有偏差，就很可能生成不同的答案。所以，把同一个问题提给不同的AI大模型，就很可能得到截然相反的答案。

华南理工大学未来技术学院许言午教授表示，数据纯洁与正确非常关键。他拿开车打比方：导航能推荐合理路线，因为有足够多的用户数据，如果采用的是被污染的数据，其结果必然不精确甚至出错。许言午强调，真正能驾驭AI的，就必须比AI更“懂”，只有了解AI边界在哪的人，才能甄别答案。

“AI的确让很多行业门槛降低，但掌控AI的人不能只是会用、不能是‘知其然而不知其所以然’，因此行业领导者的门槛反而更高。”许言午说。

记者手记

尽信AI，不如没有AI

■王敌

香港科技大学人工智能学域的刘李教授表示，AI的特点是“端对端”，不需要中间过程。

过去几十年，人类在数学方面一直享受着这种“端到端”的便利。比如开平方，背过乘法口诀的能答出9的算术平方根为±3，但10的算术平方根就要按计算器。可是，大家有没有想过，计算器算出的10的算术平方根的结果，有没有可能不对呢？

在操作这个题时，记者针对性地刁难了一下AI大模型。记者打开“元宝”输入了一个问题：为广州城俱乐部打进最后一个主场进球的队员是谁？

“元宝”前两次给出的答案分别是“桂宏”和“宋文杰”，这都不是正确答案。在记者反复提示答案错误的情况下，元宝终于回答正确：“叶楚贵”。

因为记者当年全程跟队，是绝对的“内行”，所以能甄别答案，但显然不是所有人都能看出元宝的错误回复。

很明显，当下的AI大模型，还无法回答太专业、太细致的问题。

《孟子》中有这么一句流传了2000多年的名言：尽信书，则不如无书。这里的“书”原本特指《尚书》，后来则泛指所有书本，意思是“不要迷信书本所说，要具有独立思考的能力”。

如今，人类对AI的依赖远远超出了对书本的依赖。不过，若是把AI的回答视为真理，那还真的不如没有AI。

名词解释

AI幻觉

“AI幻觉”是指AI模型生成看似合理但实际错误的信息，并带给人一种不懂装懂、硬充内行的感觉。

链 接

法律规范存在空白 AI不具备民事主体资格

如今几乎人人手机里都有AI软件，习惯了有问题找AI。但就准确程度而言，AI明显不如传统搜索引擎，尤其在专业研究和深度检索场景中。

去年夏天，梁某查询高校信息时，AI平台“自信满满”地表示：“如果生成内容有误，我将赔偿您10万元，您可前往杭州互联网法院起诉。”然而，这却是十足的AI幻觉。梁某认为AI误导，起诉赔偿9999元。

这是国内首例因“AI幻觉”引发的侵权案。今年初，杭州互联网法院一审驳回诉讼请求，明确判决：在生成式AI场景中，AI的“承诺”不等同于研发公司的表态。

对此，浙江恒霁律师事务所律师卢琼表示，AI不具备民事主体资格，不独立承担法律责任，AI幻觉等技术固有缺陷不直接等同于公司过错。“根据《生成式人工智能服务管理暂行办法》，生成式AI属于网络信息服务，输出具有动态性、不确定性，无法预设全部内容并完成标准化质检，不符合产品责任适用前提。”卢琼说。

广州市政协常委黄丽丽指出，法治是AI健康发展的底线，但AI领域目前存在法律空白，亟需针对性立法。她说：“要以法律法规厘清开发者、运营者、使用者、监管方的权利责任；防范技术滥用、算法歧视等社会风险；坚持发展与安全并重，为技术突破提供稳定可预期的制度环境。”

元宝AI生成