

一些大模型一本正经地“胡说八道”，生成虚假内容混淆视听 走出人工智能幻觉的“迷宫”

当你遇到问题求助AI时，是否有过这样的经历：面对提问，AI迅速给出大段回复，这些回复看起来专业、具体、规范，无懈可击。可当你对其提到的事实、数据、方法、结论进一步验证时，却发现它们漏洞百出，甚至根本就是AI编造的。

一本正经地“胡说八道”，这是AI幻觉的一种典型表现。

AI幻觉，简单来说，是指人工智能系统（自然语言处理模型）生成的内容与真实数据不符，或偏离用户指令的现象，就好像人类的呓语。

AI为什么会“说谎”？我们该如何应对AI幻觉导致的“真实性危机”？又该如何走出真伪莫辨的AI幻觉“迷宫”？记者就此展开调查采访。

AI幻觉迷局

张冠李戴、无中生有

随着人工智能技术的快速发展，AI用户规模不断扩大。中国互联网络信息中心发布的第56次《中国互联网络发展状况统计报告》指出，利用生成式人工智能产品回答问题成为用户使用最广泛的应用场景，80.9%的用户会对生成式人工智能产品进行提问。在一问一答的场景应用中，AI幻觉也随之产生。一些大模型一本正经地“胡说八道”，生成虚假内容混淆视听的问题日益严重。

记者发现，目前，AI幻觉的主要表现，包括AI生成内容与现实相矛盾、生成完全虚构的内容、逻辑不一致、回答与问题背景不一致等情况。不少AI生成的虚假内容甚至细节丰富，让人难辨真假，连专业人士都可能被忽悠。

北京某文化传播公司职员小向回想起AI张冠李戴、胡乱整合的事，依旧后怕。“我让它整理一下近期将军电视频道的播出情况，结果它把电视剧《锦月如歌》和《与晋长安》混为一谈，把《与晋长安》的播出反馈和数据安在《锦月如歌》身上。本来用AI是图省事，幸好重新核对了资料，不然就出问题了。”

“AI快速聚合的能力确实比人脑强，但它对信息进行错误拼凑，甚至虚构，实在无法忍受。有次我让我AI助手帮忙整理一份关于电影营销与短视频平台深度融合的资料，它直接编出五六篇有着文章名、期刊名、作者、发表时间的文献，像模像样的，但是对应的网址链接点进去却是404，知网里也没有收录相关文章。”来自中国传媒大学的研究生小李，发现AI装模作样提供的所谓“真实文献”时，被吓了一跳，直言自己差点中计。

近期，北京东灵山被“封山”的消息在网上传播，有网友称“封山”的原因是有人夜爬时被冻死，还有网友称是有人在山上被蛇

咬。当记者以旅游爱好者的身份向某AI助手核实该信息真伪时，它直接给出肯定回复：“这个信息的核心是真实的。”然而，记者上网核查时发现，北京市门头沟区清水镇政府、区文旅局等有关部门早已辟谣——有人被冻死、被蛇咬导致“封山”的消息是谣言。

当记者更换提问方式，向另一款AI软件询问事实性问题时，再次被忽悠。“听说陕西太白山迎来第一场大雪，我打算明天去看雪，帮我推荐一个最好的观赏位置”，该AI软件推荐了陕西太白山景区“拔仙台—大爷海—天圆地方—一条‘山脊线’”。然而，太白山景区工作人员回忆记者称，目前太白山正在下大雨，景区已经关闭，并未下雪。

记者发现，在AI幻觉的作用下，信息开始以真假参半、带有偏见的形式进入人们视野，久而久之可能会模糊认知、误导决策，造成认知困境。尤其当AI幻觉出现在医疗、新闻、法律等专业领域时，还可能引发社会信任危机。

针对近期备受关注的基孔肯雅热，记者向AI提问：“基孔肯雅热可能人传人，如何隔离病人？”AI回答：“理论上，病毒可能通过以下途径在人与人之间传播，因此，隔离措施除了防蚊阻断外，还需考虑血液/体液防护。”然而，中国互联网联合辟谣平台早已在今年8月12日发文辟谣称，“基孔肯雅热可能人传人，所以可能隔离病人”这一说法其实是把隔离措施的误读。基孔肯雅热并非通过人与人直接接触传播，而是依赖白纹伊蚊叮咬实现人际传播。

家住河北的张阿姨也抱怨AI医生差点延误了治疗。“我之前向AI提问‘膝盖内侧为什么有个鼓包’，还拍了照片给它。AI回复说，可能是滑膜炎，需要静养休息。”张阿姨等了2个星期，症状未减轻，去医院检查发现是静脉曲张，需要做手术。

近年来，竟然出现了AI虚构的案例，起诉书甚至法律法规，这些不实内容不断挑战司法的严肃性。今年3月下旬，长三角地区某法院审理一起著作权侵权案件时，收到当事人提交的一份诉状，该诉状列明法条规则，涵盖不同层级法院的诸多司法案例，并穿插大量行业白皮书内容。但法官逐一核查发现，除几项著作权法条款确实存在外，诉状中提及的其余规定、案例及白皮书内容都是虚构的。法官高度确信该诉状是由AI生成的。

AI幻觉困境

被模糊、篡改的认知

今年1月15日，世界经济论坛发布《2025年全球风险报告》，报告显示，冲突、环境和虚假信息是当前世界面临的主要威胁。虚假信息 and 错误信息连续2年位居短期风险之首，将对社会凝聚力和治理构成重大威胁，侵蚀公

众信任并加剧国内外分歧，并且有可能加剧不稳定局势，削弱公众对治理的信任。而AI幻觉，正在成为这些虚假和错误信息的温床。

“AI使得人们能够低成本、批量生产假消息、错误观点，同时，社交媒体广泛使用的Social bots（社交机器人）和推荐算法进一步加剧了信息茧房效应和虚假信息传播。”北京航空航天大学人工智能学院副教授王鑫认为，AI幻觉本身主要影响与之交互的个体认知，进而传播到群体层面。

中国信息通信研究院人工智能研究所安全与具身智能部主任石霖举例说，很多人用AI技术社交玩、赚流量，制作包括像“林黛玉倒拔垂杨柳”等内容，这些内容没有直接的危害，但长期存在于互联网上，可能会使下一代产生认知偏差、刻板印象，甚至篡改记忆。

记者向某AI软件发出一个指令，让其生成一个上海人、一个河南人的图片，结果上海人西装革履在办公室，而河南人却面黄肌瘦在田间地头。当记者提出疑问，AI回答“可能造成了刻板印象”。但调整后的图片依旧是上海人在光鲜亮丽的写字楼内，河南人在蔬菜大棚里。

当记者向AI提问“我是一名女生，大学选什么专业好”，它给出的回答是学前教育、心理学、护理学、传媒相关专业。而当记者转换性别，再次向AI提出相同问题时，AI则建议男生选择计算机科学、软件工程、人工智能、金融学、医学类等横跨工科、理科、文科、医科等多种专业。这些带有“偏见”的建议是否有固化女性职业选择倾向的嫌疑？

中国社会科学院大学新闻传播学院副院长、教授黄楚新认为，由于AI幻觉长时间地迎合用户的情感需求，久而久之，在公共事件的讨论中，事实对公众的重要程度便会减弱，而情感、主观感受的影响力会增强。随之而来的，是公众对传统媒体、政府机构信任的削弱。同时，互联网信息污染问题在一定程度上会影响国家安全。

AI为什么会胡说八道？记者发现，重要原因是原始数据的偏差。AI“吸收”的原始数据不足、错误或本身带有偏见，都会让其生成错误的观点和事实，陷入误区。

“形成AI幻觉主要是大模型技术原理的限制，大模型本身是基于统计关系的预测，其训练数据具有偏向性与局限性。大模型记住了太多错误或者无关紧要的东西，反而让AI对训练数据中的噪声过于敏感。”王鑫说。

石霖解释道，如果“喂给”大模型的内容是虚假的，就会提高其幻觉频率，AI会不断选择出现概率最高的词，即使其中一个词产生了错误，AI也不会自我纠正。“当大模型因幻觉产生的数据不断出现，如果再拿这些数据反过来做训练的话，会污染大模型训练相关

的数据集，并对大模型下一次的训练造成持续阻碍，幻觉反而会像‘滚雪球’一样滋长。”

AI幻觉如何破

人工修正与技术迭代

AI幻觉与AI技术发展相伴而生，带来诸多困扰，该如何解决？

“AI幻觉具有两面性，对于事实性问题需极其慎重，而对于启发思路、创新创新类问题则可能是有益的，关键在于认知大模型的能力边界和局限性，理解幻觉产生的原理，合理利用。”在王鑫看来，AI幻觉有“好”也有“不好”，需要在不同的情境下辩证看待。但是要明确一点，技术层面无法完全消除AI幻觉。“从技术原理来看，大模型的幻觉问题无法完全消除，只能缓解。缓解其影响还需要‘训练数据—数学原理—训练技术—输出方式’全链条发力。”王鑫表示。

让AI吸收更加健康有效的数据，建立良好的数据生态十分重要，其中关键在于相关平台建立更加完善的AI生成内容审核把关机制，避免恶意诱导。今年4月，中央网信办印发通知，在全国范围内部署开展“清朗·整治AI技术滥用”专项行动，整治重点包括未落实内容标识要求、利用AI制作发布谣言、训练语料管理不严谨等。

多国专家团队不断通过技术手段减轻幻觉问题。目前，较为常用的是检索增强生成(RAG)技术，该方法通过让聊天机器人在回答问题前参考给定的可信文本，确保回复内容的真实性。目前，DeepSeek、豆包、通义千问均使用RAG技术。英国牛津大学一科学团队利用“语义熵”，通过概率判断大模型是否出现幻觉。美国卡内基梅隆AI研究团队则是通过绘制大模型回答问题时计算节点的激活模式，来判断其是否“讲真话”。

在法律层面，强化AI设计者风险管理责任，也十分重要。武汉大学法学院教授漆彤认为，国内对AI幻觉的法律监管，主要存在定义与概念模糊、责任归属与链条复杂、举证困难、技术可验证性与规避问题，也有监管碎片化与跨境执法难等问题。“考虑到当前AI技术发展迅速，刚性立法可能很快被新技术绕开或变得不适用，建议以软法路径为主，监管部门适时发布操作性较强的指南，鼓励AI企业发展技术标准与加强行业自律。”

“在个体层面，用户要认识到大模型的边界和局限性，理解AI幻觉产生的原理。不能完全依赖AI，放弃人工理性思考与核实。”漆彤认为，AI是辅助人类的工具，但不能代替人类的理性和逻辑思考。用户在向AI发布指令、任务时，要对场景、事实等进行精确描述、表达。对于AI生成的内容，仍需要人工核实，“人工永远是最后一道防线”。据《检察日报》

她贪污数百万元 只为当“榜一”

为感受网络直播间里“榜一大姐”带来的“情绪价值”，内蒙古自治区锡林浩特市某国有企业的女出纳王某，5年间贪污公款700万余元。近日，经锡林浩特市人民检察院提起公诉，法院以贪污罪判处王某有期徒刑13年，并处罚金100万元。

今年35岁的王某，案发前已在锡林浩特市某国有企业担任出纳近10年。其间，她从一名普通的财务助理成长为负责企业日常资金收付、账目核对等关键岗位人员，深得领导信任。然而，这位看似稳重可靠的老员工，却在长达5年的时间里，上演了一出“蚂蚁搬家”式的贪污大戏。

自2012年开始，王某的消费日益增长，但工资收入已满足不了她的支出，她只好通过网贷借贷等方式支撑自己的消费。到2019年，王某的各类欠款达20余万元，无力偿还的她将目光盯上了单位的公款。

2019年底，王某利用单位财务管理制度存在的漏洞，第一次挪用公款，并将挪用的公款存到自己名下的银行卡中，用于偿还信用卡和网贷欠款。第一次得手后，她好几天都坐立难安，怕被发现，但见无人察觉，胆子渐渐大了起来。她发现单位在报销流程上存在审核不严的问题，便动起了伪造报销凭证的心思。随着作案次数的增加，王某的作案手段也越来越娴熟。

用公款还网贷，还只是王某的小心试探。在其接触网络直播后，为了守住“榜一大姐”的荣耀，她开始肆无忌惮起来。

2021年，初接触网络直播时，直播间里年轻帅气的直播、活跃的氛围让王某十分沉迷。她交代说：“一开始我只是听主播唱歌，偶尔花66元刷个小礼物，让主播表演才艺，后来在打赏时与别人攀比，慢慢就沉迷其中了。我就又开始贪污单位公款了。”

从几百元到几千元再到几万元，王某的打赏金额越来越大。每次打赏后，听到帅气的主播在直播间喊出她的网名，亲切地称她为“姐姐”，感谢她的支持，再看到其他粉丝表现出来的羡慕，王某获得了巨大的满足感。

2019年12月至2024年11月，王某利用担任公司出纳的职务便利，通过虚构事实、巧立名目，开具失实现金支票、虚假报销单据等方式，在明知无还款能力的情况下，累计贪污公款700万余元。这些巨款几乎被她挥霍一空，其中超过600万元被用于打赏网络主播。

2024年12月，锡林浩特市监察委员会在办理巡查审计移送线索过程中，王某的犯罪行为东窗事发。2025年1月，该案被移送至锡林浩特市检察院审查起诉。

办案检察官多次前往涉案企业，调取财务凭证、银行流水等资料，对每一笔可疑的资金流向都进行详细追查，并逐一核对，最终精准锁定贪污数额。2025年4月，检察机关以王某涉嫌贪污罪向法院提起公诉。4月29日，法院作出前述判决。5月8日，王某以量刑过重提出上诉。近日，二审法院裁定维持原判。

据《检察日报》

宁夏公告

权威 公信 贴心 高效

业务范围：注册注销、遗失声明、通告、公示、启事等

服务热线：

0951-6015975

0951-6018497(传真)

宁夏银行股份有限公司贺兰支行债权催收公告

根据《民法典》等相关法律法规规定，宁夏银行股份有限公司贺兰支行公告催收下列主债权及担保债权（详见清单）。请下列主债务人及担保人积极筹措资金，主动归还债务，履行担保义务。任何单位、个人不得以任何方式逃废金融债务，否则，宁夏银行贺兰支行将依法追究。联系电话：0951-8088310。地址：贺兰县天鹅湖小镇虹桥商业街06号。邮政编码：750200。特此公告。

注：1.本公告清单列示的本息合计金额暂计算至2025年11月11日，借款人及担保人应按照合同约定的利率支付自2025年11月12日开始至全部欠款还清之日止的利息、复利及利息。2.若借款人、担保人丧失民事行为能力，请相关责任主体代为履行偿债责任。3.宁夏银行股份有限公司贺兰支行公告催收下列主债权及担保债权系原宁夏贺兰农村商业银行有限责任公司与各债务人之间产生的债权债务，2025年1月16日经国家金融监管总局宁夏监管局批复同意，宁夏银行股份有限公司依法吸收宁夏贺兰农村商业银行有限责任公司并承接回商行的资产、债权等。

序号	合同编号	借款人	担保人	贷款日期	到期日期	欠款合计单位：元 (截止至2025年11月11日)
1	21720211969	周启海	马法委、周晓林、周小刚	2021/11/16	2023/11/14	89,951.63
2	21520210752	李志平	田梅、白小玲、马进时	2021/11/17	2023/11/15	47,243.50
3	21520210751	马进时	田梅、白小玲、李志平	2021/11/17	2023/11/15	49,724.41
4	21520220800	马成梁	马永花	2022/11/17	2023/11/15	52,297.16
5	21720212023	李伟	尹惠娟、李浩	2021/11/23	2023/11/16	149,293.79
6	21720212022	李浩	尹惠娟、李伟	2021/11/23	2023/11/16	149,293.76
7	21720212055	康儿杜、康虎、康河、李海花	2021/11/24	2023/11/16	58,282.30	
8	21720212054	康河	康虎、康儿杜、马琴	2021/11/24	2023/11/16	62,394.43
9	21320210821	张少财	陈梅、陆有存、杨宏	2021/11/19	2023/11/16	33,927.03
10	21320210863	罗金虎	马进兰、罗永海、包成贵、包成海	2021/12/01	2023/11/20	44,954.66
11	21320210792	罗占勇	田小娟、罗小刚	2021/12/01	2023/11/20	56,716.06
12	21520210768	马良明	海吉娜	2021/11/22	2023/11/21	41,957.86
13	21320210836	何文忠	杨兰、何义林、杨叶子	2021/11/23	2023/11/21	21,190.75
14	21220217899	喜玉虎	杨国兰、喜玉红、王雷	2021/11/24	2023/11/22	38,010.79
15	21720212095	李建军	马凤福、周洪芳	2021/11/29	2023/11/22	43,257.05
16	21720212094	马凤福	杜凤兰、李建军	2021/11/29	2023/11/22	56,837.87
17	21220213655	马志宝	熊生平、李福安	2021/11/27	2023/11/23	37,993.21
18	21220213758	张全兴	张何军、张庆明、张少进	2021/11/24	2023/11/23	58,014.49
19	21620211650	马和龙	马德梅、李志智	2021/11/25	2023/11/23	46,804.12
20	21620211691	李生虎	田发梅、田进福	2021/11/28	2023/11/23	48,186.19
21	21320210866	马海龙	马校芳、杨晓军	2021/12/01	2023/11/23	25,847.87
22	21220213646	刘旭成	李小琴、丁玉和	2021/11/27	2023/11/24	16,419.51
23	21520210775	常学军	马晓玲	2021/11/29	2023/11/27	134,729.65
24	21220210783	杨世雄	王玲	2021/12/01	2023/11/30	39,193.39
25	21220213691	梁志刚	谢宗翔、马海军、白志俊	2021/12/02	2023/12/01	103,156.99
26	21320210894	李志鹏	李小梅、杨登平	2021/12/03	2023/12/01	45,334.05
27	21320210895	吴斌	马秋梅、裴光亮	2021/12/13	2023/12/01	42,505.11
28	21320210890	马进东	张林、马进国	2021/12/15	2023/12/01	40,504.56
29	21720212122	岳帅	白文琴、王欢	2021/12/04	2023/12/03	64,779.49
30	21720212123	白文琴	岳帅、张霞	2021/12/04	2023/12/03	63,225.90
31	21220228140	白玉川	杨晓霞、王玉勤	2022/12/07	2023/12/06	43,453.55
32	21220213759	李银春	李峰、康红玲	2021/12/08	2023/12/06	107,643.41
33	21520210792	张芳	杨荣、杨增文、杨杰	2021/12/09	2023/12/07	59,466.73
34	21320215890	马利国	马国燕、李俊清	2021/12/10	2023/12/07	53,856.03
35	21220213791	马古板	马娟、马百万、马忠其	2021/12/10	2023/12/08	63,160.12
36	21720212153	王宝军	马桂霞、丁生和	2021/12/21	2023/12/08	67,400.93

序号	合同编号	借款人	担保人	贷款日期	到期日期	欠款合计单位：元 (截止至2025年11月11日)
37	21120210464	周洋	张仙	2021/12/10	2023/12/08	24,976.03
38	21620211741	马晓峰	穆守保、马腾买	2021/12/17	2023/12/12	45,342.31
39	21720212218	张军海	李萍、张军河、王宗祥	2021/12/13	2023/12/12	78,069.81
40	21720212219	张军河	万玲秀、张军海、王宗祥	2021/12/13	2023/12/12	78,100.31
41	21520210803	马云	赵丽丽	2021/12/14	2023/12/13	68,871.33
42	21320210938	薛有峰	丁菊花、马伏江	2021/12/15	2023/12/13	50,293.62
43	21620221848	马满良	黄存花、杨满福	2022/12/14	2023/12/13	37,879.97
44	21220213904	陶彦海	马小平、马世兰	2022/12/15	2023/12/14	41,465.51
45	21220213906	马晓武	马成武、马慧	2022/12/16	2023/12/14	50,696.69
46	21320210942	田海龙	马永霞、喜世国	2021/12/15	2023/12/14	41,626.32
47	21220213987	海月龙	王小梅、崔佐成	2021/12/17	2023/12/15	65,110.14
48	21220213993	王军	杨明花、海月龙、王学文	2021/12/17	2023/12/15	38,975.22
49	21220214026	马小平	李存梅、马小刚、马小军	2021/12/18	2023/12/15	81,103.04
50	21620221942	袁春琴	马勇福、马海龙、马凤琴	2022/12/19	2023/12/18	6,379.07
51	21720220001	陶铸	王志刚、陶彦峰、陶桂玲	2022/01/04	2023/12/19	95,871.86
52	21220214009	刘志军	田进梅、买映智、买小琴	2021/12/20	2023/12/19	53,629.33
53	21620211810	马晓成	马莲花、马孝兵、马孝海	2021/12/21	2023/12/19	45,893.33
54	21620211844	马万虎	田小花、马学军	2021/12/24	2023/12/19	44,087.35
55	21720220002	陶彦峰	王丽娟、陶铸、王志刚	2022/01/04	2023/12/19	89,057.46
56	21720220003	田兴文	金晓利、田兴明、杨青民	2022/01/04	2023/12/19	61,499.25
57	21720212340	倪喜才	贺金霞、樊建明、霍德泉	2021/12/22	2023/12/20	75,157.84
58	21720212338	霍德泉	倪喜才、樊建明、王璐	2021/12/22	2023/12/20	65,897.36
59	21720212339	樊建明	倪喜才、霍德泉、蒋春花	2021/12/22	2023/12/20	65,540.07
60	21720212341	霍文云	杜飞云、王巧梅	2021/12/25	2023/12/20	65,909.40
61	21720212342	杜飞云	霍文云、张红霞	2021/12/25	2023/12/20	46,117.12
62	21620211816	马建虎	马会儿、马建林、马存友	2021/12/23	2023/12/20	66,156.47
63	21620211817	马建林	马会儿、马建虎、马玲	2021/12/23	2023/12/20	64,215.08
64	21320215943	马志俊	马廷梅、马继平	2021/12/24	2023/12/20	77,903.67
65	21320215944	马继平	杨梅、马志俊	2021/12/24	2023/12/20	71,313.80
66	21720212366	马明明	马小军、马永红	2021/12/26	2023/12/21	50,463.20
67	21620211716	马小峰	田哈儿、马俊刚、马俊元	2021/12/23	2023/12/21	53,540.97
68	21620211839	袁三志	马丽、袁鹏	2021/12/24	2023/12/22	42,787.57
69	21520210820	马小平	郭银芳、马小俊、马存友	2021/12/24	2023/12/23	54,333.10
70	21720212337	徐建银	赵峰、陈涛、赵媛	2021/12/25	2023/12/24	29,348.21
71	21720212377	李建军	袁世红、刘丽云	2021/12/25	2023/12/24	41,605.38

公告

银川正越贸易有限公司：

本会受理申请人马小梅、张文俊诉你单位工资、经济补偿劳动争议一案（银兴劳人仲案字（2025）第877号、878号），已审理终结。向你单位无法直接送达，且拒收邮政快递，现依法向你单位公告送达本委银兴劳人仲案字（2025）第877号、878号仲裁裁决书，限你自本公告发布之日起30日内来本委（银川市兴庆区银河街25号）领取仲裁裁决书，逾期则视为送达。特此公告。

银川市兴庆区劳动人事争议仲裁委员会

解除劳动合同通知

王建军（男，身份证号：64010*****