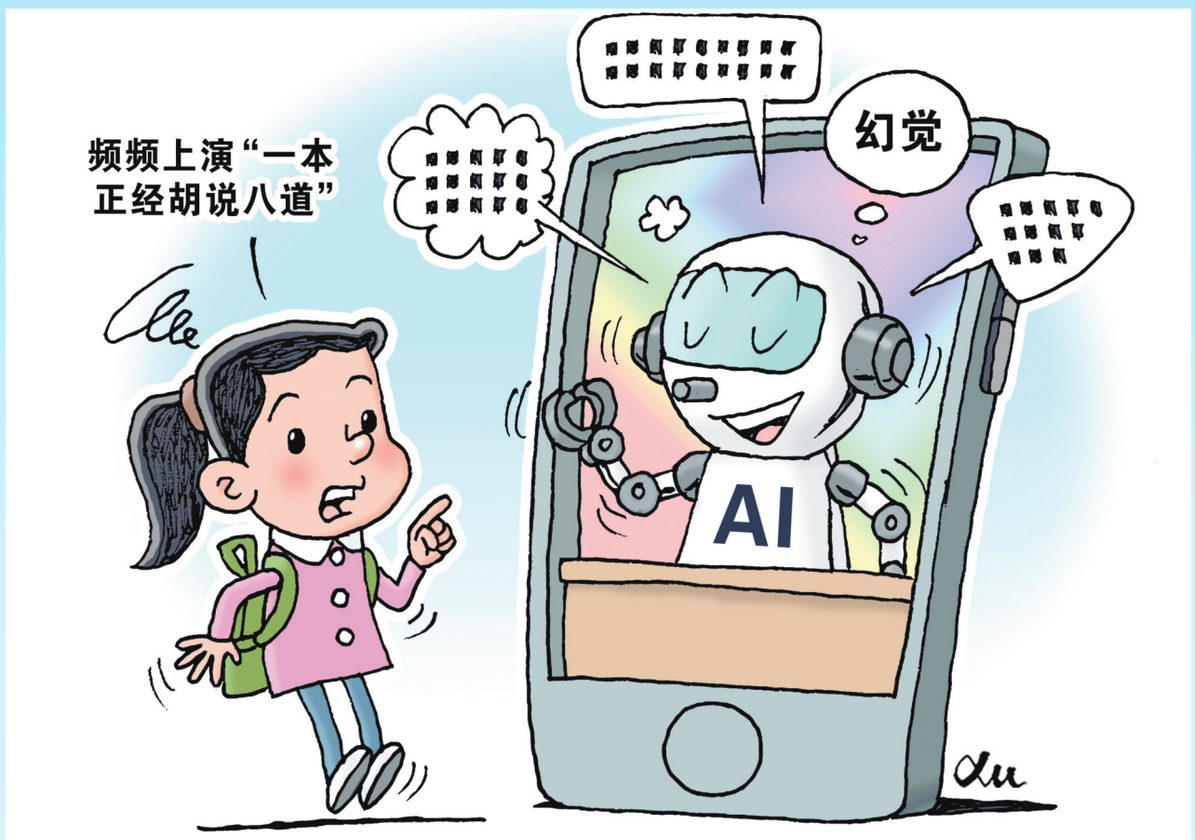
2025年9月25日 星期四
责任编辑：拉珍 编辑：次旦央吉 校对：赵耀铁
拉萨日报

当AI“一本正经胡说八道”……

新华社“新华视点”记者 颜之宏 胡林果

当前，AI正赋能千行百业，为人们的工作、学习、生活带来极大便利。与此同时，不少人发现，用AI搜索数据，给出的内容查无实据；用AI辅助诊疗，出现误判干扰正常治疗……AI频频上演“一本正经胡说八道”。社交平台上，AI幻觉引发热议。



频频出现的AI幻觉。 （新华社发）

AI好用但不时像是“中邪”了

用AI检索海量信息、让AI辅助查看三维病灶、打造AI互动课堂……如今，AI已深度融入现代生活，“人工智能+”产品赋能各行各业，从多个维度提供便利。

作为AI深度使用者，“95后”女生瑞希坦言，AI好用，但不时像“中邪”了一样胡说八道。“我让AI推荐10本高分小说，结果一多半都是它编的。反复确认后，它承认虚构了答案。”

现实生活中，不少人遇到相似情况。业内人士表示，这是由于AI幻觉导致。“AI可以快速给出答案，但生成内容可能与可验证事实不符，即凭空捏造；或生成内容与上下文缺乏关联，即‘答非所问’。”一名主流人工智能厂商技术人员说。

记者使用一款AI软件，让其给出某行业未来市场规模及信源，AI迅速回答称某投资机构预测2028年该行业的市场规模将达到5万亿美元，并提供相关链接，但链接页面找不到上述信息。记者看到，页面内容虽然包含该投资机构名称和5万亿美元表述，但预测数据并非该机构作出，且不存在2028年时间节点。

社交平台上，AI幻觉相关话题浏览量达数百万，网友吐槽涉及金融、法律、医疗、学术等多个领域。

第三方咨询公司麦可思研究院近期发布的2025年高校师生AI应用及素养研究显示，四千余名受访高校师生中，近八成遇到过AI幻觉。今年2月，清华大学新媒沈阳团队发布的报告指出，市场上多个热门大模型在事实性幻觉评测中幻觉率超过19%。

AI幻觉已经影响了人们的生活与工作。

近期，一名国外男子被诊断出溴中毒。他此前询问AI，过量食用食盐不利于身体健康，有无食盐替代品，AI回答称可以用溴化钠代替。但溴化钠存在一定毒性，需要严格遵医嘱服用。该男子用溴化钠代替食盐三个月后出现精神错乱等症状。

这几年，美国多起案件中的律师因在法律文件中使用AI生成的虚假信息，被法院警告或处分。

AI幻觉为什么会发生？

受访专家认为，AI幻觉的背后存在多重因素。

——数据污染。AI“养成”过程中，数据“投喂”是关键环节。研究显示，当训练数据中仅有0.01%的虚假文本时，模型输出的有害内容会增加11.2%；即使是0.001%的虚假文本，其有害输出也会相应上升7.2%。

奇安信集团行业安全研究中心主任裴智勇解释说，人工智能大模型需要海量数据，训练数据来自开源网络，难免会错误学习一些虚假、谬误数据，还有一些不法分子会恶意进行“数据投毒”。

“如果把AI比作一个学生，数据污染就像是给学生看了错误的教科书，自然会导致‘胡说八道’。”暨南大学网络空间安全学院教授翁健说。

——AI本身“认知边界模糊”。翁健认为，人类智能的一个重要特征是“元认知”能力——知道自己懂什么、不懂什么，而当前AI技术架构缺乏这种自我认知机制。

翁健解释称，AI可以博览群书，但并不一定理解书里的内容，只是根据统计规律把最有可能的词语组合在一起，在准确评估自身输出的可信度方面尚存盲点。

——人为调校和干预。在中国通信学会数据安全专业委员会副主任委员左晓栋看来，相较于事实真相，AI更在意自己的回答是否契合用户需求，从而导致AI有时为了“讨好”用户而编造答案。

“针对不同需求，AI的训练、打分方式也不同。”一位从事大模型训练的技术人员说，当面对写作等创意性需求时，偏理性的事实严谨在打分系统中占比相对较低，偏感性的词语优美、富有感情色彩等占比更高。“所以可能会出现一篇辞藻华丽但词不达意的文章，里面内容甚至与事实相悖。”

多方合力减少AI幻觉

第55次《中国互联网络发展状况统计报告》显示，截至去年12月，有2.49亿人使用过生成式人工智能产品，占整体人口的17.7%。受访专家表示，应通过多方合力应对AI幻觉带来的风险挑战。

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，训练语料管理不严、未落实内容标识要求、利用AI制作发布谣言等均为整治重点。

“可靠、可信、高质量的数据对降低AI幻觉非常重要，应优化人工智能的训练语料，用‘好数据’生成‘优质内容’。”左晓栋认为，可以加快推动线下数据电子化，增加“投喂”的数据量；同时探索建立具有权威性的公共数据共享平台，“各大厂商也应加强优质数据筛选，提升训练准确性”。

多家主流人工智能厂商已经采取措施，从技术层面减少AI幻觉发生。

豆包升级深度思考功能，由先搜后想变为边想边搜，思考过程中可以基于推理多次调用工具、搜索信息，回复质量明显提升；通义千问在20多个通用任务上应用强化学习，增强通用能力的同时纠正不良行为；元宝持续扩充引入各领域的权威信源，在回答时交叉校验相关信息，提高生成内容的可靠性。

翁健建议，建立国家级人工智能安全评测平台，就像生物医药新药上市前要做临床试验一样，大模型也应该经过严格测试；同时，相关平台加强AI生成内容审核，提升检测鉴伪能力。

“AI可能‘欺骗’用户，公众应客观认识人工智能的局限性。”左晓栋等专家提示，可以通过改进使用方式，如给出更加明确的提示词、限定范围等避免AI幻觉。“无论是工作、学习还是生活，现阶段的人工智能还不能全面替代人类的认知和创造能力，大家在使用AI时要保持怀疑态度和批判思维，不过度依赖AI给出的回答，多渠道验证核查。”

（新华社广州9月24日电）

外卖平台服务“新国标”向社会征求意见
聚焦外卖平台治理之痛

新华社记者 赵文君

市场监管总局组织起草的《外卖平台服务管理基本要求（征求意见稿）》，24日正式面向社会公开征求意见。

这意味着外卖平台服务管理将有“新国标”，引入标准化方法规范外卖平台服务管理行为，提升平台服务质量。

今年以来，外卖平台企业为争夺即时零售流量入口，反复发起“百亿补贴”“大额神券”等外卖大额补贴活动，“外卖大战”一定程度上加剧了餐饮市场“内卷”。

市场监管总局会同有关部门多次对饿了么、京东、美团等外卖平台进行集中约谈，直指当前外卖行业竞争中存在的突出问题，明确要求平台企业严格遵守法律法规，落实主体责任，维护市场公平竞争秩序。

在此背景下，征求意见稿聚焦当前外卖平台服务管理中存在的突出矛盾和关键短板，针对社会广泛关注的“内卷式”竞争、平台收费不透明不合理、“幽灵外卖”、配送员权益保障不足等热点问题形成标准条款。

记者从市场监管总局了解到，全国平台经济治理标准化技术委员会于今年8月底成立，这一标准正是由全国平台经济治理标准化技术委员会归口管理，主要起草单位有中国标准化研究院、中国网络安全审查认证和市场监管大数据中心等。主要外卖平台均参与了标准研制全过程，对标准内容已有充分了解和前期准备。

征求意见稿通过优化平台收费规则与促销规则、明确与落实平台对商户的管理责任、保障配送员权益，促进外卖行业规范发展，维护市场公平竞争秩序，保障多元主体权益。

针对平台收费规则不透明、不合理问题，征求意见稿明确了收费管理的具体要求，提出平台应规范收费项目、公示收费信息、进行合规审核等。

针对平台“内卷式”竞争问题，征求意见稿明确了对价格促销管理的要求，包括确保促销活动公开透明、促销规则公平合理等，有效维护市场竞争秩序。

如何防范外卖食品安全风险？征求意见稿从入驻条件、信息审核、运营管理等方面入手，对外卖平台对商户的管理提出系统要求，明确平台应建立规范的准入机制，严格审核商户资质与经营信息，提供必要的技术支持与运营指导，加强日常经营行为管理，促进商户合规经营，提升平台整体服务质量和食品安全水平。

如何保障外卖小哥权益？征求意见稿聚焦配送员社会保障、劳动报酬、工作时间、职业发展和关怀等方面，要求平台规范用工关系，保障配送员合理收入与劳动时间，加强业务培训与职业保障，推动建立公平、合理、可持续的配送员权益保障机制。

值得注意的是，征求意见稿对平台在服务管理方面提出了总体要求，包括平台应依法经营、保障各方合法权益、维护公平竞争秩序、落实社会责任、推动多方共治等方面，进一步督促外卖平台履行责任。

相比于同类国际标准，这一标准有何新意和亮点？从征求意见稿看，更侧重于对主体行为的规范，聚焦外卖行业中平台、商户、配送员等多元主体间的行为协同，填补现有标准在主体关系协调层面的空白，推动多元主体利益诉求协调一致，进一步促进外卖行业的规范治理。

中国人民大学食品安全治理协同创新中心研究员孙娟娟表示，这一标准通过整合食品安全、公平竞争、劳工权益等不同维度的监管要求和合规管理来优化外卖行业生态，实现不同利益相关者的共生共赢。

从可操作性上看，这一标准立足外卖平台服务管理的现实需求，紧密结合外卖平台的服务管理实际，系统梳理服务管理流程中可能遇到的关键问题，对相关问题进行深入分析，能够为外卖平台企业从聚焦单向度合规到整合多维度合规的管理优化，提供可参照适用的服务管理模式与程序指引，有效规范外卖行业的服务行为，提升整体服务质量与效率。

业内表示，为外卖平台服务研制“新国标”有助于回应社会关切，推动行业尽快形成公平、有序、可持续的格局，提升消费者信任度与行业整体形象。

（新华社北京9月24日电）

AI时代电影行业的变与不变

新华社记者 陈怡 孙一然

22日，一场名为“未来影像”的AI电影国际峰会在第30届釜山国际电影节期间举办，引发业内人士关于AI对电影行业影响的热议和深思。如大家所见，AI正以前所未有的速度和广度渗透电影工业，带来一场堪比工业革命的深刻变革；但不少电影人坚持，电影创作的核心本质无法被取代。

釜山国际电影节官方产业交易平台“亚洲内容与电影市场”在今年推出全新单元InnoAsia，将AI等前沿科技纳入视野，探索AI如何重塑电影表达。此次AI电影国际峰会由字节跳动旗下AI创作与内容平台即梦AI、云和AI服务平台火山引擎和上海电影股份有限公司等机构共同举办，此前他们联合发起的“未来影像计划”全球AI影像作品征集活动的最终获奖作品，也被带到InnoAsia平台展映。

此次“未来影像计划”获奖短片的创作均使用即梦AI等生成式人工智能创作平台。依托字节跳动自研的Seedance视频生成模型，即梦AI能够帮助创作者突破技术门槛，实现“创意即产出”。火山引擎则为创作过程提供了底层技术支持，其云端算力集群能够保障Seedance模型的高效运算。

即梦AI产品经理林捷表示：“国内有着非常繁荣的短视频生态，（在技术层）会比较紧密地去向创作者收集各种各样的反馈，创作者也会反过来帮助我们去快速迭代模型。”

不论是短视频还是大电影，AI技术正以“全链条渗透”方式融入影视产业生态。火山引擎旗下大模型服务平台火山方舟模型策略产品高级总监向凯接受新华社记者采访时表示，目前AI在影视里的应用虽然还处于初级阶段，但几乎覆盖了电影制作的全部流程。AI开拓了创作者们的自由度，助其提升效率。

参加峰会的博纳影业集团影视制作副总经理曲吉小江告诉记者，“AI技术逐步成熟，整个业界从初期的怀疑排斥到现在慢慢地接受、尊重并使用，AI生成影视可以说快迎来要爆发的‘奇点’”。

曲吉小江在当天的峰会演讲中说，AI正重构电影制片模式，将原本难以把控的人物、造型、光线等等变得“精准可控”，符合电影工业化指标。她表示，AI不仅仅是一种工具，更像是人类的“外脑”，赋予了创作者“超能力”。与历次电影所面临的技术变革不一样，AI将彻底改变人们的思维方式。

上海温哥华电影学院电影制作系高级讲师奥黛·阿瓦迪亚也认为，利用AI技术将是影视行业的主流趋势。“尽管好莱坞对于AI评价褒贬不一，但不能否认的是，越来越多的制片人会使用AI。AI能节省预算，也能缩短工期，有了AI，可以尝试更多可能。”

然而，在AI给影视行业带来的巨大变化中，也有声音认为AI无法改变电影创作的核心。此次“未来影像计划”获奖短片导演——韩国影视创作者姜信圭坦言，AI的出现对于影视行业带来的改变如同工业革命，但AI影视不可能完全取代传统创作，因为演员身上独特的“气息”是AI无法复制的。

姜信圭认为，AI当然有其强大的表现力和优势，但演员的呼吸、表演中传递的信息，源自其走过的人生，这是机器不可能拥有的。他相信，AI最终会成为电影制作中特效技术的一部分，与传统电影并行发展。姜信圭表示，未来他的电影创作将是“双轨”并行模式：一方面，继续用传统方式拍摄在现实条件下能实现的电影；另一方面，利用AI去尝试那些原本想都不敢想的故事。

（新华社韩国釜山9月23日电）

观赏“新月吐蛾眉”

9月26日5时26分将迎来年度最小蛾眉月。天文科普专家表示，如果天气晴好，9月25日和26日傍晚日落后至月落前，感兴趣的公众可赏这轮“新月吐蛾眉”。

（新华社发）