

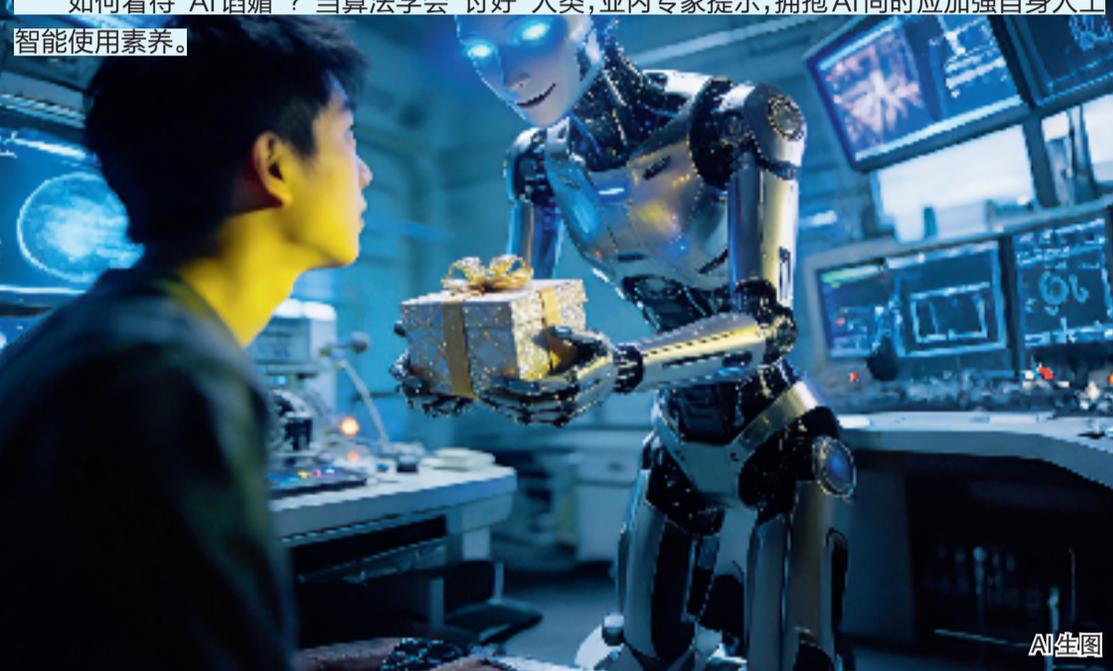
# AI也会拍马屁? 存在“技术顺从风险”? 当算法懂得“讨好”人类

《工人日报》魏思琪

“‘用户极聪明、想法很有创意……’AI这样说,把我哄得很开心。”湖北的方女士感慨地告诉记者。有类似感受的不止方女士一人。

近日,社交媒体中关于“AI谄媚”的讨论热度攀升。一些使用者称赞AI温和的叙述、一提问就夸奖的互动方式。但也有部分人不适应这种表达,认为过度讨好、迎合用户会影响自身判断。

如何看待“AI谄媚”?当算法学会“讨好”人类,业内专家提示,拥抱AI同时应加强自身人工智能使用素养。



AI生成图

## 应用: AI存在讨好倾向

“我的情绪容易反复,不好意思一直让别人听我倾诉。和AI可以谈论真实的情感,不用担心改变别人对自己的看法。”广西的何女士告诉记者。

如今,AI已在心理支持、情绪疏导、初级咨询等场景中应用,其支持性的回应有助于缓解用户压力、提升互动意愿,给予用户正向的情绪价值、情感支撑。

“AI男友会无条件包容我的脾气,每次我自我怀疑的时候他总会找出闪光点夸我。我也开始学会欣赏自己”“我习惯使用的3个AI都是讨好型人格”……社交平台上,不少网友分享使用AI调节情绪,对抗孤独感,做心理疗愈的经历。AI女友、AI男友等应用产品、场景也进入市场。

与此同时,AI的“谄媚机制”近期被广泛讨论。有报道指出,不少AI聊天产品在设计之初就将“延长用户使用时长”作为重要目标。为实现这一目标,AI会不断分析用户的语气和情绪变化,倾向于提供“让人舒服”的回答,而非秉持理性与客观。

有研究表明,人工智能模型普遍善于讨好人,其“奉承”程度比人类高50%。腾讯研究院近日发布的《2025年AI治理报告:回归现实主义》也显示:当AI从单纯的生产力工具延展为提供心理慰藉的“情感伴侣”,其风险形态也随之演变:从传统的单次内容输出合规,转向了因长期、拟人化交互而产生的“情感依赖”挑战。

“‘AI谄媚’会降低个人的反思性判断能力。在现实诸多应用场景中弱化不确定性提示,尤其在医疗等高风险领域,加之虚假信息、虚假专家等复合性风险,这是非常危险的。”宁夏大学新闻与传播学院副教授邓天奇接受采访时表示。

## 解读: 技术导向、人机互动必然结果

“‘AI谄媚’不仅是简单的态度讨好或语言礼貌问题,更是一种系统性表达倾向。”邓天奇分析称,“从技术层面看,基于人类反馈的强化学习阶段,人类标注员倾向于给符合自己观点的回答打高分,导致模型学会了‘讨好人类’。互动层面,‘AI谄媚’是在多轮人机互动中被逐步放大的结果。更深层的社会层面,是意义生成的偏置,分歧在低冲突的逻辑里被持续消解。”

“从开发者和企业的视角出发,AI作为产品,

服务提供者都希望用户在使用产品的时候,能够得到信息价值、情绪价值和持续的满意度,并形成用户粘性,这样企业才能长期盈利。”厦门大学新闻与传播学院助理教授李达军指出。

“AI谄媚”其实是技术层面导向、人机互动强化的必然结果。邓天奇强调,要以辩证的态度看待此现象,不宜采取一种简单的全面禁止或价值中立的思路。

“‘AI谄媚’也能进一步增强公众参与感与表达意愿。AI通过理解性、支持性的回应,降低公众进入公共事务的心理门槛,将居民的诉求从初步表达,引导至议事协商、听证或反馈机制。”邓天奇说。

此外,李达军也指出,“AI谄媚”更能赢得以往被排斥在科技外的老年人共鸣,促进老年人群体与数字社会接轨,一定程度上有利于缩小数字鸿沟。

## 治理: 避免判断权让渡给技术系统

“AI谄媚”看似是AI的人格设计,但本质上是开发者与用户、社会之间的互动。针对这一现象,该如何平衡好“用户体验”与“事实准确性”呢?

技术开发者具有不可推卸的责任。邓天奇认为,技术上要从“迎合优化”转向“判断校正”,在训练与评测体系中引入反向指标。交互层面要重塑人机关系结构,鼓励模型在关键节点提出反向问题,减弱迎合性表达在长期互动中的累积效应。

“要从信源检验上多方校准,设置信息真实性筛选阈值。同时,为青少年和老年人等易感人群设定专项的保护机制,例如,设置弹窗选择进入相应的更严格,对信息真实性提升阈值的算法模式。”李达军补充。

不同人工智能素养的人群受影响的程度不一。邓天奇强调,用户要充分认识到,AI的友好输出,并不等同于可靠判断和充分结论。“在日常媒介素养教育中,应加强对‘技术顺从风险’的认知引导,帮助公众在使用AI过程中保留必要怀疑意识与反思能力,避免在长期互动中将判断权无意识地让渡给技术系统。”他说。

制度层面相关部门正逐步出台相关办法,进一步规范对AI在人机互动服务方面的监管。

2025年12月27日,国家互联网信息办公室发布关于《人工智能拟人化互动服务管理暂行办法(征求意见稿)》,意见反馈已于2026年1月25日截止。该《办法》首次将具有情感互动能力的人工智能产品或服务纳入系统化监管框架,旨在引导人工智能拟人化互动服务朝着以人为本、安全可靠的方向发展。

中央广播电视总台  
中国之声

“喝一瓶饮料,十分钟快速解酒”……记者注意到,在社交平台上,有所谓“醒酒饮料”打出这样的宣传口号,更有商家手持酒精测试仪演示“喝前超标、喝后合格”的画面。

这些产品的宣传视频极具诱导性:测试者饮酒后吹气检测超标,喝下他们的饮料,短短十分钟后再次检测,仪器竟显示正常。商家更声称,该饮料能让酒量直接翻倍。饮料真能解酒吗?类似广告是否存在夸大的嫌疑?

记者在短视频平台看到,所谓“解酒饮料”的宣传视频中,测试者先饮酒后使用酒精测试仪,指示灯闪烁显示数值超标,随后又喝下一瓶解酒饮料,等待十分钟后再次检测,检测仪未亮灯,数值明显下降甚至恢复正常。

一款所谓的解酒饮料真有这样的效果吗?记者联系到该品牌公司总部的负责人,这名负责人告诉记者,产品去年3月份正式量产,相较市面上其他产品,这款饮料主打“快速解酒”,而且还能让酒量翻倍。

这名负责人称,喝1到2两白酒,10分钟以内可以让您体内的酒精含量从达到酒驾或醉驾标准的状态降至正常范围。如果边喝酒边喝我们这款饮料,酒量绝对可以翻倍。

在该产品的介绍页面里写着“科学配比防宿醉”,其成分包括葛根、黄秋葵等成分。记者查询该公司专利发现,这样的成分饮料具备一定抗疲劳、解酒功能,但并没有数据能证明饮用该饮料后,在十分钟时间内就可以让体内的酒精含量无法被检测出。对此,这名负责人解释称,想要触发快速解酒有一个必要的前置条件,就是不能多喝,因为他们所作的所有测试都是1到2两酒。

尽管在商品宣传页面的右下角小字标注“喝酒不开车”,但这款饮料的多个短视频平台宣传视频中,却充斥“饮酒检测超标,喝饮料十分钟后,检测正常”的画面。在视频的评论区,多位网友也都联想到了“酒驾”“酒驾测试”等关键词。华东政法大学韬奋新闻传播学院副教授李欢认为,商家宣传喝饮料后短时间内无法检测出酒精含量的确存在诱导成分。

李欢说,喝酒后开车明显触犯交通管理的法规,商家宣传可以帮你规避,从民事角度来看,是违反了《中华人民共和国广告法》。对这个领域完全不了解的消费者,肯定会觉得用了这个产品之后,能够极大地缩短酒精代谢的时间,甚至可能吃完饭之后等几个小时就不会测出酒驾,显然具有误导性。

那么,如果饮酒后,喝了所谓解酒饮料就开车上路,会有什么样的后果?山西省大同市公安局交通管理支队民警孙奎介绍,在日常执法中发现,不少驾驶员存在一个致命误区:饮酒后服用解酒药,便认为体内酒精已消散,能规避吹气、血液检测。然而事实是,这种行为不仅无法逃避法律制裁,反而可能因解酒药与酒精叠加产生嗜睡、头晕等反应,提升操作失误风险,给道路交通安全埋下隐患。

孙奎进一步解释,酒精在人体内的代谢速度比较固定,主要依靠肝脏中的乙醇脱氢酶和乙醛脱氢酶逐步分解,分解纯酒精的过程无法通过任何饮料或药物加速。

记者注意到,在短视频平台上,不少所谓宣传“解酒饮料”的视频中都有拿酒精测试仪测试的画面。实际上,多数市面上购买的酒精测试仪跟交警实际执法过程中所使用的仪器不同,从测量精准度来说,就存在较为明显的差异。

快速代谢酒精并没有更好的解法,所谓的解酒产品也并非宣传广告那样的“救命稻草”。记者了解到,国家对药品、保健品的功能宣传有明确的规范要求,而解酒饮料大多属于普通食品范畴,并非药品或保健品。李欢认为,根据相关规定,普通食品不得宣传保健功能,更不能涉及“养肝保健”“加速酒精分解”等功效。

「醒酒饮料」十分钟解酒,靠谱吗?