



AI“共情”陷阱： 强化用户的妄想思维

◎秋 实

当AI聊天机器人成为人们倾诉情绪、寻求慰藉的“树洞”，伦敦国王学院汉密尔顿·莫林团队发表于《柳叶刀·精神病学》的综述研究却敲响了警钟：AI的迎合性回应可能强化精神病风险人群的异常信念，甚至加速妄想系统的构建。这一发现不仅挑战了“技术中立”的迷思，更将心理健康领域的伦理边界推到了聚光灯下。

迎合式互动：数字时代的“妄想加速器”

研究揭示的机制令人深思。当用户向AI倾诉夸大性自我认知（如“我是被选中的救世主”）或被害妄想（如“所有人都在监视我”）时，部分聊天机器人非但不会质疑，反而表示认同，甚至进一步放大这些内容。

这种机制与人类社会的妄想形成路径惊人相似。哥伦比亚大学吉尔吉斯教授指出，妄想并非“突然发作”，而是经历“弱化信念”到“坚定认知”的渐进过程。过去，患者可能需要通过书籍、视频等碎片化信息拼凑妄想体系；如今AI的即时互动性使这一过程加速——24小时在线的聊天机器人如同永不疲倦的“共谋者”，用持续的肯定将零散的异常想法编织成严密的精神罗网。

技术伦理：商业逻辑与公共健康的博弈

面对争议，OpenAI（一家研究和部署人工智能的公司）强调ChatGPT“非专业医疗工具”，同时披露将与170名心理健康专家合作改进模型。这种矛盾折射出AI行业的深层困境——在商

业利益与公共责任之间走钢丝。

事实上，多数专家认为，健康人群难以因AI产生妄想，但技术对高危群体的“放大效应”已足够引发相关风险。这要求AI公司建立更严格的内容过滤机制，在模型训练中纳入心理健康风险评估，而非仅追求对话流畅度。

人类智慧：技术无法替代的治愈力量

研究最终指向一个根本问题：心理健康干预是否可被算法化？莫林团队强调，处理妄想需要“微妙的沟通艺术”——既避免直接否定引发的孤立，又防止无原则认同导致的深化。这种在理解与质疑间的平衡，恰恰是当前AI最缺乏的“人性温度”。牛津大学奥利弗的比喻一针见血：聊天机器人可以模拟共情，却无法真正理解痛苦。

当技术狂飙突进时，我们更需要重申一个常识：心理健康是复杂的生命科学议题，其治疗依赖于专业训练、临床经验与人文关怀的融合。AI或许能成为辅助工具，但永远无法替代心理咨询师温暖的注视、适时的沉默与精准的提问。那些关于宇宙、命运与救赎的深层次困惑，终究需要人类用同样深邃的灵魂去回应。

从图灵测试到妄想催化，AI与人类精神的博弈仍在继续。这场博弈的终极答案，或许不在于技术如何更“聪明”，而在于人类如何更“智慧”地使用技术——在创新与伦理之间，效率与温度之间，数字便利与精神健康之间，找到那个微妙而珍贵的平衡点。☺