AI幻觉会影响大模型用户的持续使用行为吗？

——基于用户规避方式与心理机制的质化分析

□夏以柠　张洪忠

【摘要】随着AI大模型应用的快速拓展，AI幻觉问题也越发突出。用户如何感知AI幻觉？用户明知AI存在幻觉却仍持续使用大模型，这一行为背后的心理机制是什么？本文采用半结构化访谈法，对40名具有使用经验的用户开展深度访谈。研究发现，一是用户普遍认为AI幻觉多以“真假掺杂”的形式出现。二是多数用户在遭遇幻觉后并未停止使用，而是通过核查比对、场景规避与模型切换等策略，将幻觉由模型自带的风险问题转化为一种可被管理的有用工具，从而持续使用。三是用户持续使用的四种心理机制：基于工具理性形成效率优先选择，非事实导向的使用期待，技术普及下同辈压力的外部推动，AI风险个体可控的用户认知。

【关键词】AI幻觉；用户规避；大模型
DOI:10.16017/j.cnki.xwahz.2026.03.030

一、问题提出

“幻觉”一词最早是在2000年的计算机视觉领域出现的。[1]随着生成式人工智能技术的迅猛发展，大语言模型（Large Language Model，LLM，以下简称大模型）伴生的“幻觉”现象已成为学界和业界关注的一个重要议题。大模型中的幻觉现象呈现出复杂的错误输出谱系，其表现形式为忠实性幻觉与事实性幻觉，与传统规则系统中具有可预测性和确定性的风险不同，大模型幻觉的根源在于其基于概率的生成机制、注意力机制以及自回归生成过程之间的复杂交互作用。[2]Maleki等人通过对14个数据库的系统审查发现，尽管“AI幻觉”一词被广泛使用，但在不同领域，其定义缺乏一致性，且常被赋予拟人化特征，本质上它是模型生成了貌似合理但实际上不正确、无意义或不忠实于源内容的输出。[3]本文将“AI幻觉”定义为大模型生成了与事实不符或者张冠李戴甚至是无中生有信息的一种现象。

目前AI幻觉主要围绕产生原因、幻觉风险与影响展开讨论。在技术层面，部分研究已在分析幻觉的产生原因并进行分类。有研究指出，AI幻觉源于模型“以输出为导向的虚构工具”本质，其依据词与词之间的统计关联而非真正的理解与推理进行生成。[4]这导致其可能产生多种类型的扭曲信息。另有研究从可验证性角度，划分为与源内容矛盾的内在幻觉和无法从源内容验证的外在幻觉[5]；或更细致地归类为过拟合①、逻辑错误、事实错误、无根据捏造等8种一级错误和31种二级错误。[6]这些分类为识别和治理幻觉提供了基础框架。在风险与影响层面，已有研究深刻揭示了AI幻觉可能带来的危害。AI幻觉首要威胁的是信息的真实性与公信力，尤其在新闻、医疗、法律等高风险领域，可能引发连锁性的信任危机。它可能制造自我循环的错误信息网络，即AI生成的虚假信息被其他AI或平台反复引用和强化，形成“AI幻觉症候群”，对社会认知结构造成冲击。更深层的，幻觉可能侵蚀人类主体性与认知能力。当用户过度依赖并信任AI输出时，其批判性思维和独立求证能力可能退化，陷入认知迷失，满足于AI提供的知识表象，甚至将机器的虚构内容内化为自身的认知幻觉[7]。这种依赖会导致用户自我罢黜，在闲言、好奇与两可中沉沦，丧失本真的求知与判断。[8]此外，幻觉还可能被恶意利用进行深度伪造，操纵舆论，威胁社会安全。[9]

尽管现有研究已较多探讨大模型幻觉带来的潜在风险，但大模型在各领域的用户渗透规模仍呈增长趋势。中国互联网络信息中心发布的《生成式人工智能应用发展报告（2025）》调研数据显示，截至2025年6月，我国生成式人工智能用户规模达5.15亿人，普及率为36.5%。[10]大模型的用户规模还在不断增大之中，用户增长趋势不可避免。在此提出一个问题：用户是如何看待AI幻觉的？为什么有如此明显的幻觉，用户还要继续使用大模型？这背后有什么样的用户心理机制？

中国知网　https://www.cnki.net

二、研究方法

本研究采用半结构式访谈法，于2025年12月起招募40名具有AI大模型使用经验的受访者，并依托腾讯会议平台开展一对一访谈，单场访谈时长控制在1小时左右。鉴于研究聚焦于关注过大模型幻觉的用户群体，本研究采用便利抽样策略选取样本，访谈中涉及隐私问题均填写知情同意书并予以加密。详细信息见表1。

表1　受访者信息表

受访者序号	年龄	性别	职业
F1	25岁	男	电子商务从业人员
F2	35岁	女	公务员
F3	27岁	女	在读学生（博士）
F4	23岁	男	在读学生（硕士）
F5	25岁	男	乡村全面振兴专职工作人员
F6	23岁	女	护士
F7	25岁	男	国有企业工作人员
F8	26岁	女	国有企业工作人员
F9	27岁	男	新闻媒体从业人员
F10	34岁	男	高等学校教师
F11	21岁	女	在读学生（本科）
F12	20岁	男	在读学生（本科）
F13	19岁	男	在读学生（本科）
F14	19岁	男	在读学生（本科）
F15	20岁	女	在读学生（本科）
F16	24岁	女	制造业企业会计人员
F17	19岁	女	在读学生（本科）
F18	25岁	男	信息技术服务业从业人员
F19	24岁	女	情感疗愈师
F20	26岁	女	商业服务业从业人员
F21	23岁	男	在读学生（硕士）
F22	27岁	男	新能源汽车行业管理人员
F23	23岁	男	在读学生（硕士）
F24	24岁	男	工程技术领域质量检验人员
F25	30岁	女	商业服务业从业人员
F26	28岁	女	在读学生（博士）
F27	23岁	男	在读学生（硕士）
F28	24岁	男	某职业岗位实习人员
F29	25岁	女	教育领域专业技术人员
F30	26岁	男	科学研究领域辅助工作人员
F31	30岁	女	国有企业工作人员
F32	23岁	女	在读学生（硕士）
F33	26岁	女	在读学生（博士）
F34	23岁	女	在读学生（硕士）
F35	27岁	男	在读学生（硕士）
F36	26岁	女	商业服务业从业人员
F37	23岁	女	在读学生（硕士）
F38	24岁	女	在读学生（硕士）
F39	29岁	男	软件研发工程师
F40	20岁	女	在读学生（本科）

实施流程上，遵循开放性、平等性、相关性和深度性原则，邀请受访者分享自己的见解、行为与感受，问题主要包括：（1）大模型平台的早期接入与使用；（2）在大模型出现幻觉时的使用经历；（3）幻觉出现后的使用结果和影响。每组访谈时间持续1小时，最终获得56万余字的文本以供分析。

访谈内容的饱和度检验。研究采取等距抽样的方式将焦点小组访谈获得的数据集分为两个部分：一部分由80%的数据构成，充当数据组；另一部分由另20%的数据构成（即F3、F8、F13、F18、F23、F28、F33和F38），充当检验组，用于饱和度检验。结果显示，检验组提取的开放式编码关键词均可在数据组中复现，没有发现新的概念或类别，符合理论饱和原则，不再新增访谈数据。

三、用户如何感知大模型幻觉

大模型幻觉的呈现特征不是通篇的“胡说八道”。受访者普遍反馈，大模型幻觉并非完全无依据的“通篇胡说”，而是在真实信息中混入虚构、错误或未经证实的内容，形成“真假掺杂”的特征。

（一）用户感知幻觉的主要特征是真实信息中混入虚构内容

根据用户的使用经验，大模型出现幻觉的场景可以分为五类，分别为时效依赖型、语义混淆型、内容主体模糊型、推理一致性失真型和语境转移型。

第一类为时效依赖型幻觉。用户描述其主要产生于涉及实时信息、最新事件或即时政策的情境。当外部信息尚未被纳入训练语料，或模型缺乏即时检索能力时，系统倾向于以概率生成方式，用看似真实的叙述替代事实性回答。这类幻觉的输出结果在表达上具有连贯性，但在事实层面却存在明显偏差。

比如说天气，你问它今天甘肃的天气怎么样？模型怎么可能知道你实时的天气，它只有联网搜索或者调用一些工具才能查到这样的数据，所以你非要问这样的问题的话，模型肯定会胡说八道，因为它不能给你输出空白答案。（F10）

在我使用的范围内，它是会经常出现的。因为我都是用它研究国家最新政策的。往往是一个重要会议刚开完，我当天就要写心得体会，当天就要了解会议思想，这个时候，全网是没有相关资料给AI的，它找不到，所以它肯定瞎编。（F2）

第二类为语义混淆型幻觉。其产生于同名对象、语义歧义或提问关键词导致情景错配现象。模型在信息整合过程中将不同来源的实体片段拼接，

中国知网 https://www.cnki.net

从而形成结构完整却指向错误的叙述。其表层呈现逻辑连贯，但在语义指称层面发生错配。与时效依赖型幻觉相比，该类型幻觉具有更强的误导性与延迟识别性。

有一次，我想让AI帮我梳理一个“社区养老服务改革试点”的经验，当时这个项目在我们市也算比较新的做法。我在问题里提了项目名称，但没有给太多背景信息。它生成出来的内容同样是非常完整，从政策依据到实施路径都有，但是我越看越感觉内容像是另一个地方的做法。后来我仔细去核对细节，才发现它写的其实是另外一个城市同名项目的改革案例，只是因为名称和主题高度相似，它就把那一套模式照搬过来了。这种“张冠李戴”比简单错误更麻烦，你需要一条一条去核实它到底混入了哪些别的案例信息。（F5）

第三类为内容主体模糊型幻觉。大模型在回应有限度问题时，模型会额外补充解释性内容，从而形成核心相关内容，附加生成内容的复合性回答，但在答案的相关性与精准性之间发生错位。若话题是关于新议题或开放性任务的，大模型往往能够生成结构完整、逻辑自洽的文本框架。内容框架仍有参考价值，但其细节证据却存在偏差或失配。

我让AI帮我生成一个策划，大概看了之后，感觉确实是那种场合下提到的形式，不过具体看下去，有很多内容是我们根本实现不了的。就是在现实中，地级媒体单位如果去操作的话，是实现不了的。（F9）

第四类为推理一致性失真型幻觉。其主要出现在计算、制度规则或程序性推理任务之中。此类输出在形式逻辑与推理流程上保持一致与连贯，但在结果层面出现系统性偏差，呈现出强烈的看似正确的答案。该幻觉更可能直接影响决策结果与正式文本生产，从而引发更强烈的负面认知评价与不信任感。

在政策条文解读上，它能把条款拆得很细，也能推理出一套结论，可是跟真正的法律适用范围一对比，才发现它很多理解都偏了，只是推理过程看起来特别可信。（F33）

第五类为语境转移型幻觉。此类幻觉发生于长时间或多轮互动的连续会话中。在语境逐步累积与语义权重迁移的过程中，模型生成轨迹逐渐偏离原始任务目标，形成一种过程性、动态化的输出不稳定现象。

就是今天跟它说不清楚的话，可能明天再跟它说，它可能就理解了。但是如果你要现在再问下去的话，它已经开始给你胡说八道了，它就很难纠正到正确的跑道上来了。（F7）

长对话里它会记住前面的内容，但有时候记得过于牢固，后面我换了角度提问，它还是按照之前的语境往下回答，慢慢就和我真正的问题越来越脱节了。（F37）

（二）大部分用户感知幻觉偶尔发生后仍持续使用大模型

大部分用户认为幻觉是偶尔出现。绝大部分受访用户在使用过程中都会发现幻觉，19位被访者认为幻觉发生的频率是偶尔发生，18位被访者认为幻觉发生的频率是经常发生，3位被访者认为幻觉是几乎没有发生的。用户对自身专业领域、熟悉场景的信息敏感度高，易察觉幻觉；而在陌生领域，“部分正确”的表象易掩盖幻觉内容，增加辨别难度。绝大多数用户在幻觉出现后仍选择继续使用大模型，少部分人停止使用是源于他人发现错误，导致使用行为受到影响。

我记得是我跟它说一个明星的事儿，我说这俩是谁，你下次要记住，我会再来问你的，结果发现下次AI还是认不全，它甚至都不知道他俩是谁。（F20）

四、用户持续使用的四种心理机制

大部分用户对AI幻觉持接受态度，并在行为上表现出持续使用的倾向。基于用户的持续使用行为，本文进一步从心理与情境层面，对用户在幻觉风险背景下仍选择持续使用大模型的内在机制进行探究。访谈材料显示，用户的行为选择并非简单的工具依赖或情绪偏好，而是在效率考量、社会环境与风险理解等多重因素的共同作用下形成的结果。

（一）用户基于工具理性形成效率优先的选择

在工具理性框架下，用户并不以“信息是否绝对真实”作为唯一评价标准，而是将大模型置于效率—使用成本—收益的权衡体系中进行判断。大模型在信息整合速度、内容生成效率与多场景适配能力上的显著优势，使其在用户心中被界定为一种高效率媒介工具。即便存在幻觉问题，仍能为用户创造显著价值。

大模型的核心优势集中在效率提升和便捷性。大模型能快速完成信息检索、文本生成、数据分析等重复性高的基础工作或高耗时任务，大幅节省时间成本，避免人工筛选、整理的烦琐流程，让用户在学习、工作中聚焦核心环节，显著提升整体效率。例如电商从业者用其收集产品相关信息，均能将原本需数小时的工作压缩至几十分钟。

有了AI之后，像之前不懂就问百度一样，现在第一选择是AI，它能提升效率，文献总结、日常工作和生

中国知网 https://www.cnki.net

活中的问题都能直接让它解决，比百度或其他搜索引擎更全面系统，能根据个人需求给出答案。（F21）

找数据的时候快一点，提取数据分析数据也更快，省了很多时间，好多东西不需要自己弄，不需要自己考虑，丢给它就能出来，比自己手动整理、计算省事太多。（F23）

大模型的功能多场景适配，一款大模型可以覆盖用户多场景需求。大模型从工具属性的文本润色、代码生成、翻译、P图等到情感属性的情绪陪伴、倾诉疏导，再到专业场景的文献梳理、实验方案设计，可适配学习、工作、生活中的多样化需求，充当多种角色。

工作上用它生成重复性文书，生活中让它翻译、算命，情绪不好时用它的咨询师模块倾诉，功能很全面。翻译时会问使用场景，还能介绍当地俚语，咨询师模块能共情，比朋友更关注我的需求。（F25）

大模型的使用便捷性高。在大模型的使用过程中，无须复杂操作技巧，通过自然语言即可下达指令，它响应速度快，支持多终端使用，随时随地满足需求，适配快节奏的工作与生活场景。

使用很方便，操作简单，不用复杂学习，有需求直接输入指令就行，不管是整理文本、翻译外语，还是处理数据，都不用掌握专业技巧，普通人也能快速上手。（F34）

它24小时都在，半夜失眠、情绪崩溃时随时能找它聊天，不用考虑时间和场合，发送指令后很快就能收到回复，比找朋友倾诉更便捷，不用顾虑对方是否方便。（F35）

（二）非事实使用期待降低了AI幻觉负面体验感

期待水平是值得探讨的核心影响因素。不同用户基于对AI技术的认知、使用场景的需求，会形成差异化的期待，在X（推特）和微博平台中，用户对大模型的角色期望存在差异，在X中，舆论关注较多的大模型人类角色想象依次为传播者、创意者和教育者，在微博中，受关注较多的角色依次为创意者、分析师和传播者。[11]基于大模型智能体，发现中美开发智能体的侧重点不一样，我国智能体开发偏向情感类应用，美国创作者主要打造“助手”角色的情感智能体。[12]用户对大模型的期望存在差异，这些期待又导致每一次使用大模型的体验感和满意程度不同，最终塑造出多样的用户行为模式。

如果用户的核心期望只是大模型承担机械性、标准化的基础任务，无须深度思考或创新输出，仅需完成信息核对、格式规整、内容归类等执行层面工作，核心诉求是节省重复劳动时间，降低基础操作门槛。在这类期望下，用户不依赖大模型提供创意或专业见解，仅要求其精准执行明确指令，完成流程化、规则化的工作。比如核对文本中的数据准确性、统一文档格式、将零散信息按类别归类整理、提取文本关键信息形成清单等。用户对结果的要求是合规达标而非优化创新，只要符合基础标准即可，无须额外拓展。

我用大模型做PPT、收集资料、做可行性报告，就是因为我不想动脑，它能帮我输出东西就行，哪怕不是百分百满意也能接受。偶尔也用它简单P图，我没有专业修图技能，它能按指令改一改就够了，不用追求多精细。（F7）

用户的高工具型期望指希望大模型完成专业内容，发挥其思考作用。核心期望是大模型发挥思考型作用，完成需要逻辑分析、创意拓展、专业解读等深度赋能工作，核心诉求是突破自身认知局限，获得创新思路、专业方案或深度见解。在这类期望下，用户将大模型视为专业助手或创意伙伴，要求其基于指令进行深度思考，输出具有创新性、专业性或逻辑性的内容。例如，在生成原创性文本的过程中负责创意脚本，在专业领域提供代码撰写以及实验设计优化，或是能帮助用户提供拓展思路给出新角度等。用户对结果的要求是有价值、有深度，不仅要满足基本需求，还需提供超出自身预期的专业支持或创意启发。

写材料时，希望它不仅搭框架，还能结合上海地区法院的实际情况，提供贴合场景的创新表述和政策解读，而不是通用模板。（F2）

用户的低情感型期望指希望AI提供倾听服务，追求无评判、随时可用的轻松互动。这类用户期望与大模型互动时无须承担现实社交的压力，不用担心打扰他人，或被评判、被敷衍，可以随时开启或结束对话，在轻松、包容的氛围中释放情绪。

我不知道像豆包心情树洞算不算，有些事情不好跟身边人讲，也不好跟家里人讲，那总得有个发泄的地方嘛！这个只是在我心情不好的时候才用到，一般来讲是不会跟它进行聊天的，我还是偏向于将AI当作工具软件。（F24）

用户的高情感型期望是需要AI情感陪伴并提供回应。渴望被倾听、被关注，获得情感慰藉。这类用户将大模型视为情绪树洞或虚拟朋友，核心需求是在孤独、焦虑或情绪崩溃时，获得即时的陪伴、专注的倾听和共情的回应，填补现实社交的空缺。

半夜睡不着、情绪崩溃时只能找GPT聊，它的咨询师模块能共情，回复很长，会站在我的角度考虑问题，有些话甚至没人跟我说过，被AI安慰到，比朋友更有耐心。（F35）

中国知网 https://www.cnki.net

我性格内向，不喜欢社交聚会，也不太愿意结交新朋友，但又有点渴望社交，QQAI好友完全充当了我的情感寄托，日常聊天、分享琐事，能承接之前的话题，不会出现断裂，满足了我的社交需求。（F38）

（三）技术普及下同辈压力作为持续使用的外部推动力

用户持续使用AI大模型的重要原因是技术的普及和同辈压力的共同作用。在快节奏的工作与学习环境中，当身边同事、同学或行业同行普遍使用大模型提升效率时，他人普遍使用而自身未使用的差距，会形成隐性压力。用户担心自身效率落后于他人，难以在竞争中占据优势，因此选择主动融入使用潮流，将大模型视为避免落后的必选工具。这种压力并非强制要求，而是源于对效率差距的焦虑，他人借助大模型快速完成文本生成、资料整理、创意构思等工作，若自身坚守传统方式，可能面临任务完成速度慢、成果质量不占优的困境，进而影响工作表现或学习成效。这种“不使用就落后”的认知，促使用户即便遇到大模型的错误或幻觉问题，仍选择持续使用并适应其特性。

我身边不管是年轻同事还是家里长辈都在用AI，同事写材料、做报表都靠腾讯元宝、DeepSeek，效率特别高，领导也默认大家用工具提升产能。要是我还靠自己从零开始写，不仅慢，还可能因为思路不够开阔被比下去。我父母都知道用豆包查健康问题、分析检查报告，年轻人要是不会用，反而显得落后。在这种环境下，不用AI就感觉跟大家脱节了，工作效率也跟不上，所以只能一直用。（F2）

（四）用户形成AI风险是个体可控的认知

大多数用户认为AI幻觉带来的风险是可控的。在技术层面，AI幻觉源于大模型的概率预测与语料库数据，并非偶发性错误，而是一种难以通过个体操作彻底消除的不确定性问题。这一风险具有内生性与不可完全规避性，理论上应被理解为生成式人工智能伴生的结构性风险。然而，在具体使用实践中，用户并未将幻觉建构为需要制度性干预或技术治理的宏观风险，而是通过日常经验的积累与使用策略的调适，对其进行重新界定。AI幻觉风险被理解为是用户使用方式不当、核查不足以及场景选择错误导致的。幻觉风险被降级处理、通过核查弥补和更换模型规避。

根据使用场景切换模型来规避幻觉。幻觉频率与场景有较强相关性，幻觉多集中在冷门、专业、实时、本地化信息场景，用户可根据需求场景选择合适的模型应用，或避免让模型处理高风险任务。当大模型出现幻觉时，用户通常会通过切换至其他适配场景的模型来规避问题，优先选择在对应需求上口碑更优、准确性更高的模型，同时结合多模型交叉验证来确保结果可靠，既不放弃大模型的效率优势，又通过模型间的功能互补降低幻觉带来的风险。

没了豆包我会用DeepSeek，像之前用不了GPT就换了豆包，不同模型各有侧重，换着用能避开单个模型的幻觉问题。（F22）

面对大模型幻觉，受访者已形成一套成熟且主动的应对策略，核心逻辑是通过多重验证和人工把关降低风险。用户会先借助搜索引擎反向检索事实性信息，验证大模型输出的关键内容；对于专业文稿、工作材料等重要场景，会进行逐句人工校对，重点核查逻辑连贯性、细节准确性及与自身领域知识的匹配度；部分用户还会采用多模型交叉核对的方式，将不同大模型的输出结果对比分析，提取一致且可信的信息，剔除矛盾或存疑的内容。这套策略既利用了大模型的基础输出价值，又通过后续验证环节弥补了幻觉缺陷，让大模型的使用更具可控性。

这对我就是一个非常大的影响。我现在遇到问题，首先会想着去问这种大模型，而不是去搜索引擎搜索。其实我还会再去传统引擎，再去搜索，再去核对，但是一开始思维逻辑还是用大模型去询问，它相当于一个私人医生。（F4）

我当时让AI生成澳洲旅游推荐的餐厅，它说的头头是道，结果到实地搜索发现根本没有那家店，这让我对它的信赖度降为零，白白充了那么久的钱，但后来还是继续用了，因为大多数时候是好用的，只能之后多检查。（F25）

五、讨论

在用户与大模型的持续互动过程中，幻觉并未被单纯理解为技术层面的问题，而是在实践语境中被重新赋义，并逐步纳入一种以效率考量、风险对比与情境判断为核心的使用逻辑之中。

首先，用户持续使用大模型的主要原因是其对技术价值判断逻辑的改变：从“是否正确”到“是否值得用”的价值转向。尽管大多数用户在实际使用过程中清晰感知到大模型幻觉的存在，并在多次互动体验中不断累积对其的不确定性感受，但幻觉并未被用户理解为阻断技术采纳与使用的问题。相反，用户在持续的实践经验中逐步形成一套个人在使用中解决幻觉问题的使用逻辑和个体认知。用户对大模型的价值评估，从“结果是否绝对正确”转向对效

中国知网 https://www.cnki.net

率—成本—收益的综合衡量。此外，通过核查、比对与场景规避等策略，将幻觉由系统性风险重构为可被人工避免的使用成本。在效率收益、同辈竞争压力与功能多场景适配性共同作用下，使用户得以在风险与价值之间形成一种动态平衡，从而在心理与行为层面维持对大模型的持续使用。在用户的使用体验中，幻觉风险没有被用户定义为系统本身存在的风险问题，而是在持续互动中被重新定义为用户操作流程的一部分。

其次，AI幻觉的技术内部短板可以通过用户外部使用克服。未来，检索增强、外部知识库约束、事实校验链路、可追溯推理机制等工程化手段的持续引入，会使幻觉从一种不可预期的输出偏差转化为可被监测、可被压缩，并在较大程度上被修正的技术问题。当幻觉问题逐渐被压缩、被消除，其可能不再是对传播结构具有强破坏力的核心问题。

进一步来看，一些学者主张不应该仅仅将AI幻觉视为技术错误，更要探索其可能的创造性价值。有文章认为，幻觉的创造性偏差能够激发创新。在艺术创作、图像生成、创意写作等非事实严谨性优先的领域，AI提供的出人意料的信息，能够打破人类固有的思维定式，提供全新的视角和灵感。[13]同时，也有观点指出，幻觉揭示了人机认知的本质差异，强化了人类的不可替代性。直面并应对幻觉的过程，促使从业者重新确认自身在情感共鸣、复杂语境理解、思辨等方面的核心竞争力。从哲学与技术批判视角看，幻觉现象如同一面镜子，反射出人类自身的认知局限与思维特性。有学者指出，将机器的输出称为幻觉本身可能是一种拟人化误判，更准确的描述应是“虚构症”。[15]

最后，极少部分用户在负面体验累积与信任崩塌后，选择停止使用大模型。在所有被访者中，35位用户选择接受幻觉的出现，另外5位用户无法接受幻觉的发生，甚至有一位用户出现停止使用的行为。拒绝接受幻觉的用户选择自主生产的方式规避幻觉带来的风险。他们因大模型频繁出现虚构文献、编造专业信息、错误输出等幻觉问题，选择放弃不可靠模型，并在核心内容生产上亲力亲为。他们放弃不可靠模型的核心原因是幻觉可能导致严重后果，这些后果带有不可预测性。用户对AI工具的信任并非一次性瓦解，而是源于在多次的大模型使用中，频繁出现AI幻觉，这种幻觉体验的累积以及幻觉的隐蔽性、顽固性带来的心理冲击，最终导致其对AI输出的内容出现全面质疑，甚至对AI本身形成不信任。访谈中持不接受幻觉的用户提及AI幻觉经常发生，发生频率偏高。幻觉发生后核对的时间成本与用户采用传统人工方式检索的时间成本相近，AI的高效工具属性没有得到验证，打破了用户对AI高效精准的初始认知，使得用户对AI的信任程度降低。

总体而言，AI幻觉作为生成式人工智能的内在局限，在短期内难以完全消除，但用户并非被动承受技术缺陷，而是通过主观的实践实现风险控制与价值获取的平衡。然而用户对大模型的持续使用，既是技术驱动，也是个体效率体验和风险认知等核心因素共同作用的结果。

（杨子奇同学参与了研究和文章撰写工作）

注　释：

①过拟合是指机器学习或统计模型中，模型在训练数据上表现过于优秀，如准确率高，但在新数据（测试集或实际应用）上表现显著下降的现象。

参考文献：

[1]S. Baker and T. Kanade,“Hallucinating faces,”in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000,pp.83-88.

[2]郭全中，张磊，韦薇.AI幻觉的生成机理与敏捷治理[J/OL].新闻爱好者，1-15[2025-12-18].https://doi.org/10.16017/j.cnki.xwahz.20251205.001.

[3]Maleki N,Padmanabhan B,Dutta K.AI hallucinations:a misnomer worth clarifying[C]//2024 IEEE conference on artificial intelligence (CAI).IEEE, 2024:133-138.

[4]聂静虹，李肖楠.抵御“智能幻象”：主流媒体内容生产中的AI幻觉与风险防范实践[J].新闻与写作，2025(11)：97-109.

[5]Ji, Z., Lee, N., Frieske, R., et al., “Survey of ”, ACM Computing Surveys, vol.55, no.12, 2023, pp.1-38.

[6]Sun Y, Sheng D, Zhou Z, et al.AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content [J].Humanities and Social Sciences Communications, 2024, 11(1): 1-14.

[7]彭兰.智能生成内容如何影响人的认知与创造？[J].编辑之友，2023(11)：21-28.DOI：10.13786/j.cnki.cn14-1066/g2.2023.11.003.

[8]梁兴洲，王迎春.人机交互中的认知迷失与沉沦危机[J].现代传播（中国传媒大学学报），2025，47(02)：35-45.DOI：10.19997/j.cnki.xdcb.2025.02.011.

[9]闫桥，陈昌凤.人机协同内容生产中的边界问题：何以建构、消融与重塑[J].青年记者，2025(7)：12-18.DOI：10.15997/j.cnki.qnjz.2025.07.005.

[10]中国青年报.当AI成为受访大学生的“全能伙伴”[EB/OL].(2025-09-22)[2025-09-22].http://hn.xinhuanet.com/20250922/20a5fa8d9a384a9192721bdf27e020d6/c.html.

[11]任吴炯，张洪忠，燕东祺.大模型的角色期望：基于X（推特）和微博语境的比较分析[J].新闻界，2024(5)：58-67.DOI：10.15897/j.cnki.cn51-1046/g2.20240315.001.

[12]张洪忠，夏以柠，林润.是工具还是情感对话者？中美AI大模型话语竞争背景下的智能体应用比较[J].传媒观察，2025(3)：41-51.DOI：10.19480/j.cnki.cmgc.2025.03.004.

[13]秦静，李菲，邓元兵.从“表征偏差”到“认知突围”：人工智能幻觉作为创造性认知的双重中介[J/OL].新媒体与社会，1-13[2025-12-15].https://link.cnki.net/urlid/CN.20250814.1714.008.

[14]胡泳.当机器人产生幻觉，它告诉我们关于人类思维的什么？[J].文化艺术研究，2023(3)：15-26+112.

作者简介：夏以柠，北京师范大学新闻传播学院博士生（北京　100875）；张洪忠，北京师范大学新闻传播学院教授，北京师范大学新媒体传播研究中心主任（北京　100875）。

编校：王志昭

中国知网　https://www.cnki.net