

2026
年第
5期

新闻界

Journalism and Mass Communication

新闻界



2026

第 5 期

总第 398 期

ISSN 1007-2438



四川日报报业集团主管主办

中文核心期刊
CSSCI 来源期刊
中国人文社科综合评价 (AMI) 核心期刊
四川省品牌期刊



公益广告



川观新闻
—看四川，观天下—

新闻界

宣



欢迎订阅

《新闻界》杂志

思想影响传播 媒介推动文明

240元/年(12期)

《新闻界》近年荣誉

- 中文核心期刊
- 中文社会科学引文索引(CSSCI)来源期刊
- 中国人文社科综合评价(AMI)核心期刊
- 2025年中国新闻奖一等奖
- 2023年中国新闻奖三等奖
- 2021年中国新闻奖二等奖
- 四川省品牌期刊
- 多次荣获四川新闻奖一等奖



国际标准连续出版物号: ISSN 1007-2438

国内统一连续出版物号: CN 51-1046/G2

编辑部电话: (028)86968603 (028)86968876

发行部电话: (028)86969774

刊期: 月刊(每月10日出版)

投稿渠道: www.ixinwenjie.com

电子邮箱: xinwenjie66@163.com

目次

聚焦“十五五”·“人工智能+”

P₄ / 生成式人工智能以来：人机关系研究的三条线索与未来进路
文 | 陈昌凤 梁亦昆

前沿视点·人工智能时代的事实核查

P₁₆ / “概率事实”的核查：拆解 AI 大模型的新闻生产机制
文 | 白红义 万旭琪

P₂₈ / 再造可信生态：AI 幻觉风险下主流媒体事实核查的价值与权威重塑
文 | 杨奇光

理论研究

P₃₇ / 平台化生存：生活世界的“可见”与“遮蔽”
文 | 张秀丽 贺子阳 牛超杰

编委会主任
编委会副主任
总编辑
副总编辑
执行副主编
编辑
编委

谭江琦
徐文礼
段吉平
舒 炜
李韵奕
孙尚如 徐秋染
王炎龙 / 四川大学新闻传播与出版学院教授
张志安 / 复旦大学新闻学院教授
刘海龙 / 中国人民大学新闻学院教授
常 江 / 深圳大学传播学院特聘教授
曹 晋 / 复旦大学新闻学院教授

冯建三 / 台湾政治大学传播政治经济学教授
李天铎 / 昆山科技大学文化创意产业讲座教授
Christian Huchs / 希腊雅典大学政治经济学教授
Graham Murdock / 英国拉夫堡大学社会学系教授
Jlises Mejias / 纽约州立大学奥斯威戈分校传播政治经济学副教授

栏目主持
学术顾问
前沿视点栏目支持

常江：前沿视点 蒋晓丽：博士生新论
尹 鸿 李良荣 陈力丹 杜骏飞 郑保卫 童 兵 喻国明 蒋晓丽
全媒体技术与传播认知四川省重点实验室（四川日报报业集团）

媒介与社会

P₄₉ / 技术、身体与社会角色：中外用户对陪伴式 AI 的情感幻觉省思
文 | 肖珺 于林海 高天赐

P₆₂ / 对抗“算法父爱主义”：抖音短视频防成瘾机制下的心理逆反与过度使用警觉性变化研究
文 | 于茜瑶

博士生新论

P₇₇ / 高 / 低语境概念的矛盾性及转化：语境化的视角
文 | 王宵静

新闻一线

P₈₉ / 县级融媒体中心自我造血机能的结构困境与解题路向
——基于 Y 省的考察
文 | 朱天 冯楷 高敬

编辑出版

地址

电话

投稿渠道

电子邮箱

邮编

刊名题写

封面图片

国内统一连续出版物号

国际标准连续出版物号

发地

订阅

邮发代码

发行部电话

《新闻界》编辑部

四川省成都市红星路二段 70 号

(028) 86968603 (028) 86968876

Http:// www.ixinwenjie.com

xinwenjie66@163.com

610012

甘惜分

C 视觉影像数据库 | 韦洪涛

CN 51-1046/G2

ISSN 1007-2438

国内外公开发行

全国各地邮局 (所)

62-334

(028) 86969774

印刷

出版日期

定价

成都鑫达彩印务有限责任公司

2026 年 5 月 10 日

20 元 / 期

版权声明

凡本刊所刊登稿件,即视为作者授权本刊拥有网络传播、图书出版等再使用权,本刊有权决定是否以某种纸质出版或互联网出版形式进行二次出版。作者对此有异议者,须在来稿时特别注明;未经本刊允许,其他刊物或网络不得转载本刊作品;获本刊允许转载者,须注明转自《新闻界》。



生成式人工智能以来：人机关系研究的三条线索与未来进路

陈昌凤 梁亦昆

摘要 近三年来，生成式人工智能正在迅速基础设施化。围绕人机关系变迁的研究逐渐形成了三条主要线索：一是聚焦宏观层面的控制危机，关注“真实性”消解带来的认识论挑战与新型权力秩序的重构；二是侧重生产实践中的赋能悖论，探讨去技能化风险、职业边界游移以及智能繁荣背后的隐形劳动；三是转向情感维度的人机亲密纠缠，审视人机依恋机制及共生想象背后的本体论意涵。三条线索之间相互勾连，共同构成理解智能时代人机关系变迁的框架。未来研究应进一步回归生活世界，追踪人机关系在长时段内的演化轨迹，并在此基础上打破对用户的同质化预设，通过对异质性经验的探究在具身化与情境化的人机交往中寻求与人工智能共存方案。

关键词 智能传播；人机关系；生成式人工智能；人机传播

中图分类号 G206 **文献标识码** A

作者 陈昌凤，暨南大学国际传播研究院教授，广东广州 510632；清华大学新闻与传播学院教授，北京 100084；梁亦昆，暨南大学新闻与传播学院博士研究生，广东广州 510632

自 2022 年年底 ChatGPT 发布以来，生成式人工智能迅速进入公共视野，并逐步从公众讨论的热点话题嵌入到日常生活的各项实践之中，短短数年间便完成了从令人新奇的技术物到智能基础设施的转变。与以往以规则驱动或统计预测为主的人工智能型态不同，大语言模型凭借其特有的自然语言处理能力，首次使机器能够以近似对话者的身份参与人类的意义协商、叙事建构与日常交往，从而重塑了人机关系的基本形态。

生成式人工智能对社会生活的渗透引发了学界的高度关注。一方面，学者们十分警惕大模型在知识生产中的不透明性、概率性与幻觉风险；另一方面，随着人工智能在信息生产、创意劳动与经验决策中的广泛应用，围绕劳动替代、职业边界、人机协作与主体性让渡的讨论也在不断深化。与此同时，人机关系开始突破技术工具论框架，人们在与智能体的日常互动中，不仅开始发展出情感联结，也促使我们重新思考自我与他者之间的边界，进而引发关于共情、伦理与人类自我理解的复杂争论。尽管讨论的话题与关切的现象各有不同，但近年来围绕生成式人工智能研究共同指向一个核心问题：在机器日益具备“类人”交流能力与准主体性的背景下，人类应当如何重新界定自身的主体性边界？

在相关学术讨论不断涌现的语境下，本研究以生成式人工智能时代的人机关系变迁为核心焦点，对近年中外相关研究进行系统梳理，尝试在多维度的理论图景中把握人机关系演变的学理逻辑，从而为理解智能时代的社会形态提供更为清晰的认知坐标。总体而言，既有研究对人机关系的讨论，大致可提炼出三类不同的话语框架：第一类聚焦于宏观层面的控制危机与秩序重构，第二类框架侧重于生产实践中的赋能悖论与劳动博弈，第三类框架则转向情感维度的人机亲密纠缠与共生想象。三类话语在不同场景中交错并存，并促使研究议程在争议中不断分化与推进。

一、系统性焦虑：认识论危机与权力秩序重构

面对突如其来的技术变革，许多学者往往会本能地保持警惕。尤其当某项技术开始转变为不得不依赖的生存环境，学者讨论的重心就会由功能性想象移向规范性问题，由此激发的是关于一套技术失控风险与安全治理秩序讨论的问题域。研究者们把大模型视作一种不透明且概率化的知识生产装置，以及一种正在基础设施化、会重塑权力结构的社会技术系统，关注其给人类社会带来的多重影响，以及在寻求有效治理方式时所面临的结构性困境。

（一）认知挑战：真实性的消解

由大模型所引发的第一重危机在于其动摇了“真实”作为判断基准的经验性基础。传统的信息获取与验证逻辑，主要依赖于索引与指向关系——通过可追溯的线索建立起信息与事实之间的对应关系。而大模型的语言生成机制则截然不同，大模型所输出的文本是来自相关性的统计拼接和机遇概率的预测。因此，其虽能生成语法连贯、字面合理的文本，却缺乏对现实世界中因果结构的把握，也缺乏基于身体经验或事实对照的校准能力。这就导致信息的形式连贯性与内容的真实性之间出现系统性断裂。

在实践中，最直接的问题便是大模型“幻觉”现象的频发^[1]。在新闻生产领域，人工智能虽显著提升了内容生成效率，却因其输出的信

息中潜藏的虚假信息，导致事实核验成本上升。即使是面对已存在事实的检索概括，幻觉也难以避免。哈加尔（Nicholas Hagar）等在针对新闻报道式任务的实证研究中发现，大模型在基于真实文档进行查询时，输出的内容中仍可能有30%以上存在幻觉，模型会为来源添加无依据的特征描述，或将归因观点悄然转化为普遍陈述^[2]。当然，对真实性基础构成调整的并不限于幻觉。以文生视频为代表的新一代生成技术，能够制造出高度逼真的视听内容，使观看者产生身临其境的感知体验。在这种情境下，真实逐渐从对事物本体的判断转向对感官效果的认同，形成一种由生成式媒介主导的新型视听逻辑^[3]。在此意义上，真实性越来越像一种被媒介与技术条件所规定的“中介化真理”，数字域甚至可能生成相对独立于人类经验的表象世界，从而迫使人类重新面对“相合性”与“可证成性”的哲学追问^[4]。

而一旦模型生成内容大规模进入传播环境，认识论危机便会进一步外化为信息生态风险。源头数据污染、黑箱机制与低成本合成共同叠加，使辨识与溯源难度上升，公共讨论便可能陷入信息迷雾所引发的结构性不确定性之中^[5]。与此同时，通过训练-使用-再训练的递归回流机制，社会噪声与低质量内容被不断注入并沉淀于模型内部，推动风险从孤立的虚假信息扩散为系统性的知识退化，最终危及整体内容生态的质量^[6]。

（二）权力转移：新权力形态的诞生

如果说“幻觉”是一种显性的认知干扰，那么，更隐蔽的控制形态则来自大模型的基础设施化。伴随着技术的普及，生成式人工智能逐渐沉降为底层环境，像电与水一样成为日常生活默认要素，进而在不被察觉的情况下重塑人类的表达方式、互动规则与意义秩序^[7]。有研究指出，如若大模型成为了我们今天的“语言基础设置”，那输出内容的准确性其实并非最根本的困境。抛向整个社会的一个真正的问题是，人类是否还能在算法预设的语言框架之外，自主地组织经验、表达情感并建构自我^[8]？

大模型的基础设施化催生一种新型权



力形态，有学者借用社会学布拉滕（Stein Bråten）于上世纪提出的“模型权力”（Model Power）概念来捕捉这种微妙的支配关系。布拉滕认为，如果一方拥有一个能够解释现实、预测结果的强大模型，而另一方没有，那么即便是在看似民主、平等的对话中，弱者也会不由自主地被强者的逻辑带走，最终沦为强者的“思想复印机”。而大模型时代的AI与用户之间便很可能形成这种强弱关系，从而导致人类用户沦为模型的思想附庸^[9]。模型权力之所以难以识别和抵抗，是因其常以效率提升或操作便利的服务形式出现，使控制逻辑藏身于优化用户体验的表层之下。在社会实践层面，隐形的权力也催生着新的符号暴力。生成系统从训练数据中学习既有社会结构中的偏见模式，并将其转化为生成偏好^[10]。有研究者发现，在新闻形象的视觉生成中，算法对性别气质、职业姿态与权威符号的选择性强化，会把女性记者的能动性压缩为可被观看的装饰，从而把职业秩序以视觉形式重新写入公众想象^[11]。同样地，随着AI嵌入新闻生产的核心环节，传播权力由传统媒体机构转移至掌握算力、模型与平台接口的技术主体，新闻业对重要性与真实性的定义权也进一步旁落^[12]。

（三）治理困境：地缘差异与技术同步性的错位

在认知权威面临挑战与技术权力日渐扩张的背景下，“对大模型的治理”被迫成为重建控制权的关键路径。然而，关于大模型治理实践从一开始就深嵌于全球政治、经济格局之中。大模型因其在知识生成、产业赋能与制度塑造中的核心作用，事实上构成了通往数字未来的关键“全球网关”。但全球南方国家在接入这一网关时普遍面临连接质量差、算力获取成本高、本地适配能力弱等结构性限制，导致其在技术依赖中处于被动位置，反而可能被进一步边缘化^[13]。

与此同时，AI技术的发展与部署迅速被纳入地缘政治的话语框架。围绕特定模型或企业的争议，常被西方置于“安全化”叙事之下，

强调风险与威胁；而中国等国家则更多诉诸“技术主权”与开放合作的逻辑^[14]。围绕AI的叙事分歧，治理实践日益承载制度竞争与话语权争夺的意涵。因此，治理文本的分歧往往并不来自“是否要治理”，而来自“以什么优先级治理”。对于大模型的期待和定位，各区域各有不同权重，因此治理文本的叙事也各有不同^[15]。而当治理叙事趋于封闭时，规则就更容易被固化为阵营化工具，反而可能削弱跨域协作。

除此之外，研究者们的焦虑还与技术演进与治理响应之间的严重错配相关。大模型的发展速度远超制度演进节奏，很容易形成典型的科林格里奇困境：问题初现时缺乏干预共识，待后果显现时又已难以逆转。因此，有研究主张构建一种多方利益相关者合作的全球机制来对待大模型风险^[16]；也有研究强调，若仍停留在“技术控制论”，就很难触及价值对齐与人机共生的根本矛盾，必须打破主客二分，在技术人文主义视角下重新界定人、系统与社会之间的关系^[17]。而无论采取何种理念，治理的有效性最终取决于可操作的制度条件。费拉里（Fabian Ferrari）等人提出三项基础要求——工业可观察性、公共可审查性与技术可修改性，为将抽象原则转化为具体实践提供了关键支点^[18]。这三者共同构成一道门槛：唯有满足这些条件，治理才可能从话语宣示走向实质约束。

二、赋能的悖论：生产实践中的边界游移与劳动异化

围绕生成式人工智能的人机关系，第二类话语将注意力放在技术赋能背后的协作困境与劳动重组上。生成式AI逐渐以准主体性行动者的姿态进入写作、编辑、策划乃至管理等环节，职业合法性长期依赖的“人类独有技能”边界随之松动，职业分工也开始调整。人们曾经用创造性表达、逻辑推理、审美判断与战略决策来标记专业壁垒，如今模型在不少场景中提供相近的输出速度与组织能力，研究与公共讨论因而重新追问“何为劳动”“谁是作者”。生产现场一方面看见更便捷的生产力形态——模型

整合计算能力、语言模板、检索和生成流程，用户只需在界面上发出指令就能获得成稿、提纲或方案；另一方面，生产现场也同时遭遇另一组变化，去技能化趋势推动边界继续消蚀，隐形劳动随之增长，个体劳动者与弱势群体往往承担更多隐性的劳动。

（一）效率之殇：去技能化危机

在技术的赋能叙事中，传统的工具论观点倾向于将 AI 视为“增强智能”，认为通过自动化繁琐的检索与整理任务，技术能够释放人类的认知盈余用于复杂问题的思考和决策，从而催生出更高阶的创造力^[19]。然而，近年来研究者们逐步转向反思路径，开始梳理技术介入的组织后果与个体代价，并将“去技能化”作为一个需要系统解释的风险点纳入讨论。

在此背景下，有学者提出了“创造性伤痕”（Creative Scar）概念，指出人机协作在短期内能提升产出数量与形式质量，但会带来内容趋同与差异性衰减。也就是说，效率提升的代价是牺牲内容的异质性与多样性。并且这种人机协作模式具有强烈的“成瘾性”，一旦撤去 AI 的技术支持，长期依赖机器辅助的使用者独立创造力水平不仅无法维持原状，甚至会跌落至未接触 AI 前的基准线之下^[20]。效率的提升其实意味着把创造性劳动的关键部分转移到模型的概率空间中完成，它使人误以为自己在更高效地创作，实则降低了对不确定性的耐受与对原创路径的生成能力。

“修辞债务”的概念也因此出现，被用以反思人工智能在写作与教育中对人类思维能力的潜在侵蚀。麦克阿瑟（Marit MacArthur）指出，写作对人类而言最重要的意义其实在于帮助我们组织思维和训练语言逻辑。而如果我们这部分过程被持续外包给大模型，虽然短期内获得了即时完成的各类文本，但却可能在思考和认知训练上累积亏空^[21]。这笔债务终将以批判性思维能力的丧失与主体性思考的匮乏来偿还。某种程度上来说，AI 生成的文本越完美，人类主体性的亏空就越巨大。

不仅如此，人机互动过程本身也在重塑人

类的语言实践逻辑。有研究者指出，大模型作为“无理解的统计中介”，通过提示-生成-筛选-修订的循环，将传统“深思熟虑的创作”转变为“即时、可检索、机器可读”的协作过程，使得语言变得去个性化^[22]。人类在获得算法算力加持的同时，也在不知不觉中被算法逻辑所同化，不仅语言表达变得机械，对所产出内容的所有权感知也明显下降^[23]。因此，所谓“赋能”，在某种程度上演变为一场以主体性让渡为代价的效率交易。

（二）边界游移：职业边界的再协商

个体层面的去技能化并不会停留在个人经验，而会外溢到社会分工与职业秩序之中，表现为职业边界的松动与身份合法性的再谈判。新闻业之所以成为高频样本，是因为它同时具有强烈的公共性与高度流程化的生产链条，使其成为观察人机协作如何重塑职业边界的典型场域。

面对生成式智能的进入，从业者往往首先采取防御性策略，通过“边界工作”重申职业不可替代性。研究者对 ChatGPT 发布后元新闻话语的分析后发现，记者倾向于将生成式 AI 排除在新闻权威之外，或通过强调集体专业主义，或突出个体原创性，坚持人在新闻生产过程的控制权和主导权^[24]。然而，这种防御策略的效力极为有限。因为职业边界不是由一份详细的能力清单构成，而是由工具链、组织结构与平台规则共同生产的一种实践秩序。机器不必完全取代人，只要在关键节点改变定义权与评价机制，就足以让人的不可替代性变成一种需要不断被重新证明的脆弱主张。有研究引入“边界域”概念，用以描述媒体场域与技术场域之间形成的中间地带。认为与其固守封闭的职业领地，不如将视线投向新闻场域与技术场域之间的边界域，在新闻专业价值与技术算法逻辑的动态协商与混合杂交中重塑自身的合法性^[25]。

在无法逆转的智能浪潮面前，单纯的防御已显无力，学界开始探索更具建设性的共生策略。有研究提出了具有辩证色彩的“对抗式合作”策略。认为新闻业不应追求人机之间无缝的平



滑融合，而应正视并保留人机之间的认知冲突，在承认差异的前提下展开对抗式合作，才能有效实现人机协同，并破解智能革命带给新闻业的焦虑^[26]。在人机协同方面，研究者认为未来可以从行动者网络理论视角出发，根据人机协同内容生产的不同阶段、不同场景使用需求，合理发挥协同网络中的人类行动者和非人行动者的优势，重塑一个充满弹性张力的流动边界以应对技术发展的不确定性^[27]。

（三）繁荣背后：智能生产中的隐形劳动

把前两节视为劳动过程的可见面，那么在其背后，还有着许多看不见的结构性影响。所谓智能革命叙事的光鲜表象之下，是隐形劳动的制度化与不平等结构的深化。

一项关于儿童智能陪伴机器人的研究，将批判的矛头指向了AI生产过程中的性别化机制。研究发现，所谓的AI情感陪伴，其实源自一群被称为“智伴妈妈”的女性工程师的情感劳动。为了让机器人在互动中表现出“爱”与“关怀”，这些女性被要求将自己的母职经验、育儿话术与情感反应“喂养”给算法，从而生产出可被复制的互动脚本^[28]。在这里，劳动异化呈现出双重形态——女性经验被抽取为可售卖的“情感功能”，而用户关系又被技术化为可管理、可复制的照护流程。此外，在更基础的训练数据生产环节，数据标注员的劳动也呈现出被平台技术“幽灵化”的趋势^[29]。有研究者发现，在以人机交互为诉求的生产场景中，数据标注员需要将其个人的自然认知转化为与计算机程序相匹配的产生式认知，才能高效生产出符合机器逻辑的标准化数据^[30]。人作为劳动主体，在组织形态上被“幽灵化”，在精神结构上又遭遇“机械化”。

与人类的幽灵劳动相呼应的，是知识产权与数据占有层面的结构性矛盾。大模型训练依赖对全网文本与图像的抓取与挖掘，创作者的智力成果在未经许可或缺乏有效补偿的情况下被转化为“训练燃料”。相关研究将其概括为对智力成果的非表达性使用，指出现有著作权体系以人类作者为前提，难以稳定界定训练、

生成与再利用之间的权利边界，从而使侵权、补偿与责任链条长期处于不确定状态^[31]。跨国比较研究进一步表明，这种制度滞后会把新闻机构与内容创作者推入去制度化困境。创造性劳动的价值被平台与模型体系重新定价，而收益分配却缺乏对等机制，形成“有产出者却无获益者”的结果^[32]。

综合来看，大模型时代的生产实践呈现出三组互相勾连的张力。其一，个体在获得效率与规范性的同时，可能经历能力外包、表达规训与创造力伤痕；其二，职业群体在协作与抵抗之间重做边界工作，却不得不面对平台化与基础设施化导致的定义权转移；其三，产业繁荣建立在隐形劳动、性别化剥削与数据占有之上，使赋能叙事变得可疑。因此，不应止步于技术乐观主义的工具化想象，也不能仅以“不会取代人”作为问题终点。重要的是把人机协作理解为劳动过程的重新组织，并追问其中的主体性与责任如何被重新配置。

三、亲密纠缠：情感连接与人机关系的本体论转向

随着人工智能从生产工具延伸至日常生活领域，它在人类社会中的角色也开始发生质的转变，并逐渐进入情感与存在的维度^[33]。大模型时代的人机关系研究正经历从功能性互动向本体论共生的范式转型，学者们开始关注人机情感的复杂流动、依恋机制以及技术作为“他者”的存在状态。

（一）理论推进：超越“媒介等同”的新交往观

长期主导人机传播研究的“计算机作为社会行动者”（CASA）范式认为，人类会无意识地将社交规则应用于机器。生成式人工智能的出现促使学界对该理论进行修正与超越。有研究指出CASA范式存在人类中心主义倾向，过度关注机器如何触发人类的社交本能，而忽略了互动过程中的双向建构^[34]。通过实证研究发现，用户与生成式AI的交互实际上已经超越了传统意义上的媒介等同，“用户和对话式人工智能之间的人机关系既不是主体和工具间的使用关

系，也不是对人际与社会关系的直接复刻”，而是用户在意识到机器非人属性的前提下，主动寻求的一种能够填补现实人际交往缺陷的替代性方案^[35]。

德庞蒂（Iliana Depounti）和纳塔莱（Simone Natale）提出了“人工社会性”（Artificial Sociality）的概念，以界定这种新兴的交际动态。他们指出，生成式人工智能所展现的是一种旨在构建社会行为表象或幻觉的技术与实践^[36]。与CASA范式侧重于人类对媒介的单向、无意识反应不同，“人工社会性”视域强调了机器的“模拟”本质与人类的“投射”机制之间的双向互动：机器通过算法与语言策略构建出社会性的假象（Artifice），而人类则通过激活社会框架和意义投射参与其中。有研究引入技术现象学中的“它异关系”概念，生成式人工智能作为交往对象呈现出介于工具与主体之间的状态，互动中的主体性并不由单方决定，而在关系推进中逐步形成^[37]。

（二）情感纠缠：人机依恋的生成及后果

在人机交往的具体实践中，情感呈现出明显的动态性。研究发现，在与情感陪伴类人工智能交往的过程中，人机关系会在连接、疗愈、折损、中辍与抵抗等状态之间移动，使用者会在不同阶段调整投入强度与关系定位^[38]。外界常用“假性亲密”来概括人机情感，但用户体验却更复杂。从“情感现实主义”的视角进行的研究认为，与其纠结于机器情感的本体论真伪，不如关注用户在互动过程中构建的主观真实性。通过一种无关利益、无需负责且永远在线的纯粹情感体验，能够帮助个体对抗现实中的社交倦怠并重构自我认同^[39]。实际上，这种主观真实性的建立，是人们在特定生活条件下对关系功能的主动选择。尤其在现实人际支持薄弱或社交成本高昂的情境中，生成式人工智能所提供的低门槛回应机制成为情感调节的重要资源。以针对视障群体的研究为例，研究发现，语音助手因其持续可用性与非评判性特征，能够形成稳定且包容的“感官依恋”，并且其稳定性和包容性上往往优于现实人际关系^[40]。

当然，凡事都有两面性，人工智能的情感效能与不稳定性往往并存。大模型故障、逻辑混乱或回应偏离会带来强烈断裂感，此时便需要用户投入额外的解释、提示与情绪修补来维持互动连续^[41]。人机情感交往实践的涌现也自然进一步引向对学者们智能化情感生产的反思。当前主流的情感陪伴模式被概括为“爱的机器人化”，其以高效率与可预测性为核心特征^[42]。高度可控的情感输出虽然降低了交往摩擦，却也可能削弱人类关系中由不确定性、冲突与共同修复所孕育的情感深度。在追求稳定回应的同时，技术反馈可能无意中过滤掉了那些构成真实人际经验的重要张力。值得注意的是，并非所有用户都偏好完美无瑕的回应。有实验研究提供了一种反向视角，当聊天机器人适度展现脆弱性，如承认错误、表达困惑或暴露知识局限之时，反而更容易激发用户的共情意愿^[43]。这说明，相较于一个能够解决任何事的“万能机器”，用户在人机交往中所期待的反而可能是能够自觉暴露其局限性且能够真诚回应的对象。

（三）追问存在：人工智能的本体深思与人机共生

讨论人机关系的本质，往往需要先从技术工具论的视线里抽身，转向对交往得以成立的前提条件，以及其在本体论层面的意义结构所展开的审思。围绕人机关系的理解条件，既有研究首先在技术路径上区分了离身沟通与具身沟通两条路线，并指出更可能出现的图景是交织推进的演进形态，而非沿着单一方向完成替代^[44]。在这一框架中，对离身AI的批评强调，语言能力的强化并不会自动带来AI对文化语境与价值脉络的整体把握，而身体经验的缺位会在根本上限制AI情境理解与价值判断的生成空间^[45]。通过考察人工智能的哲学系谱，有研究认为，人类的智能本质上是具身的（Embodied）、情境性的，如果继续忽视身体，试图用数字模型复制人类心智，结果注定是徒劳无功^[46]。与此同时，将文本交互直接等同为完全离身的经验形态，也可能遮蔽用户侧的实践复杂性。研究发现，用户常以想象方式补全AI的身体，这类想象是参与在场



感与关系感生成的关键环节，而不是可有可无的附着物^[47]。因此，离身技术与离身经验之间并不存在简单的对应关系，文本界面之下仍可能展开一种经由想象而被具身化的交往过程。

与“具身/离身”的路径讨论并行，另一条重要的理论脉络聚焦于技术他者性的重新界定。

“镜子与他者”的框架颇具启发性。人工智能既如一面镜子，映照出人类认知结构与制度安排的固有边界；同时作为一个运作逻辑相异的“他者”，又迫使我们在相似与差异之间重新定位技术主体^[48]。进一步地，有研究引入海德格尔的生存论视角，深化了对人工智能这一“他者”的本质的探讨，将以 ChatGPT 为代表的生成式 AI 界定为“新常人”——一种基于概率计算的“知识渊博的认知平庸者”。这种新型存在者虽然在交互形式上表现为顺从的工具，但在知识生产的底层逻辑上却通过“平均化”消解了独特的决断，并作为隐匿的“算法大他者”对人类施加凝视^[49]。相较上述将 AI 视为镜像反射或支配性他者的路径，还有研究尝试以更具系统性的视野拓展讨论，提出“生态人工智能”的构想，主张跳出“斯普特尼克时刻”式的冷战对抗思维，在古今互鉴与东西共生中推动人工智能从工具理性的极致计算走向面向“人类命运共同体”的生态响应时代^[50]。

人工智能的发展正促使人机关系超越工具性范畴，迈向一种存在论意义上的共在结构。要理解人机共生的新现实，我们就需要超越将 AI 视为“类人媒介”的旧有范式，转而关注其作为社会性存在所参与的情感、伦理与身份的协同塑造过程。面对这一不可逆转的趋势，人类需要在拥抱技术便利的同时，持续反思主体性的边界，从而在人机共生中重塑情感连接与存在的意义。

四、交叉与张力：走向整合性的人机关系分析框架

前述三条研究线索在各自的问题域内形成了相对独立的论述逻辑，但若将它们并置审视，便会发现彼此之间存在相互渗透与内在张力。这种渗透与张力根植于生成式人工智能同时改变认知环境、劳动结构与情感生活这一基本事实。

三条线索所指向的，是同一个社会技术现实的三个截面，彼此之间的勾连关系构成了理解人机关系变迁的重要背景。

控制逻辑向生产实践的渗透是三条线索之间最为直接的关联。第一条线索所讨论的模型权力，在理论层面常以“基础设施化”与“认识论危机”的形式呈现；而它真正发挥作用的场所，恰恰是第二条线索所聚焦的生产现场。大模型以效率提升的界面形式进入工作流程，将控制逻辑包裹在优化用户体验的外壳之下。当记者习惯于以 AI 草拟初稿、编辑借助模型校对文本，提示词策略逐步固化为职业惯例，原本属于认知劳动的判断与取舍便悄然转移至模型的概率空间之中。去技能化因此从个体能力退化的问题，演变为权力重新分配的结构性问题。技术主体通过掌控模型能力的边界，间接获得了对职业实践的定义权。数据标注员将个人认知转化为机器可读的标准化输出，如“智伴妈妈”将母职经验喂养给算法，这些隐形的劳动形式之所以得以成立，也正是因为技术系统的基础设施化制造出一种需要持续填充的结构性空缺，而填充的代价是由最脆弱的劳动者承担的。

权力结构对情感纠缠的形塑是第一和第三条线索之间较为隐蔽的勾连。情感陪伴类产品对“永远在线”“无条件回应”等特质的着力塑造，是主动激活并强化特定依恋模式的策略性设计，用户在情感投入的过程中同步贡献着对话数据与使用时长。“爱的机器人化”所追求的高效率与可预测性，也可以理解为平台治理逻辑向情感领域延伸的一种表现，将关系简化为便于管理且能够复制的互动流程，在降低运营成本的同时也压缩了用户在关系中的主动性空间。当然，平台设计意图与用户实际体验之间的落差，仍有待更多实证研究加以细化，但情感实践与权力结构之间存在关联，已是这一领域研究难以绕开的议题。

生产实践与情感纠缠之间的潜在关联，目前在研究层面尚属探索性讨论。去技能化的忧虑主要来自生产效率与能力训练之间的张力，情感纠缠的研究则更多聚焦于陪伴与依恋的心理机

制，两条线索目前各自延伸，交汇处的经验证据仍较为有限。但有一点值得持续关注，当AI在工作情境中成为日常性的协作对象，职业使用与情感交往之间的边界如何演变，以及二者的相互渗透会在多大程度上影响用户对技术的整体依赖方式，都是有待后续研究填补的空白。

关联之外，三条线索之间也存在真实的张力。控制危机框架所呈现的人类主体，是面对技术权力时相对被动的承受者；情感纠缠的研究却不断发现，用户在与AI的交往中展现出相当程度的能动性，如主动寻求补偿性关系、选择性披露隐私、在系统的标准化回应中开拓私人化的意义空间。技术的结构性力量与个体的实践能动性之间，存在持续的摩擦与协商，二者在具体情境中的互动方式，远比任何单一框架所能描述的更为复杂。赋能悖论的讨论亦然。职业边界的游移既可能呈现为能力的萎缩与定义权的旁落，也可能在特定条件下为部分从业者开辟出重新协商身份与价值的空间。同一种技术介入，在不同的社会位置与制度环境中会激活截然不同的后果，这正是将异质性经验纳入分析的根本理由。

三条线索的相互渗透与内在张力共同表明，生成式人工智能时代的人机关系变迁，难以用任何单一框架加以统摄。认识论的危机、劳动的重组与情感的纠缠，在现实中相互缠绕，有时彼此强化，有时又相互扰动。更具解释力的路径，或许在于正视不同层面之间的摩擦与张力，将这种复杂性视为分析的起点，在具体情境之中追踪权力、劳动与情感如何交织运作，从而形成对人机共存现实更贴近的把握。

五、局限与未来：迈向情境化与具身化的人机关系研究

三条线索的并置与交叉，已经初步呈现出既有研究的覆盖范围，也使其共同的认识论倾向愈发清晰。现有研究大多将人机关系作为可从外部加以描述的既成状态，而非视其为在时间中持续生成、在情境中不断沉淀的动态过程。控制与秩序的框架擅长揭示技术的结构性后果，

却较少追问这些效果如何在日常实践中被感知与内化；赋能与劳动的框架精准捕捉了生产关系的重组逻辑，却较少关注个体在职业边界游移中如何逐步调整对自身能力的认知；情感与共生的框架触及了人机交往的主观体验，却仍多以横截面方式呈现情感形态，难以说明依恋如何生成、断裂又如何修复。技术进入日常之后发生的那些更难以察觉的变化，如嵌入时间节律的微观适应、场景转换之间的意义协商、由不同人群以差异方式展开的技术实践，极易在宏大叙事的牵引下退为背景，在理论叙事里失去应有的可见性。

局限的形成有其客观原因。生成式人工智能的爆发速度远超研究积累的节奏，在现象尚未沉淀之际，借助宏观框架把握变局几乎是研究者的必然选择。但随着相关研究逐渐从对新现象的命名走向对关系机制的解释，研究重心也需要相应下移，从对人机关系性质的宏观判断，转向对人机关系如何在具体生活中被生产、维持与改变的过程性追踪。沿着这一方向，以下四条研究进路值得重点推进。

（一）从事件到过程：引入历时性视角

控制危机框架所揭示的认识论风险与权力重构，乍看之下像是技术部署后随即显现的结构性后果，但这些效应的真正发生机制，恰恰依赖于时间的积累。模型权力之所以难以被厘清，很大程度上是由于它藏身于一次次日常使用的惯例化之中，不会在某个时刻突然降临。现有研究在方法选择上常借助实验任务与横截面调查。这类研究便于控制特定变量，也能够在规定时间内收集可比较的数据，容易抓住人机互动的“起始面貌”，如初次接触时的惊艳感、效率的即时提升、特定伦理情境下的态度摆动。但是，这类视角往往在无意间把人机关系拆解成一组组可被截取的“互动回合”，互动被呈现为一次次可复现的反应链条，时间在其中只充当背景，难以进入解释的中心。关于采纳意愿、新奇感、满意度的结论，常常停留在关系尚未确定之前，也就难以说明关系如何被日常化，又如何在日常中进入倦怠、习惯与重新调适。



其实真正的问题在于，一些研究开始便将人机互动视为可重复发生的标准化事件，默认同一用户在第一次与第一百次使用之间的交互逻辑基本不变。但人机关系实际上恰恰是不可逆的积累性过程。信任一旦建立便不会回到原点，去技能化一旦发生便会在认知结构中留下痕迹。一切关系都是在时间中沉淀出来的，而非在某一时刻一次性完成的。因此，如果说人机关系是一种伴随交往时间里生长出来的关系机制，那么许多关键环节就需要沿着时间轴展开。用户需要通过持续验证与误差调适，不断建构或削弱对机器的信任。随着使用习惯趋于稳定，个体的身体节奏与工作程序随之发生系统性调整。从工具性依赖转向情感依恋的转变，往往嵌入在用户反复求助与系统反复确认的交互循环中。生成式人工智能的技术属性显著加剧了关系建构对时间过程的依赖。开发者通过频繁的版本更新重新划定系统的能力范围，用户感知到的关系连贯性也会随模型版本的更迭产生波动^[51]。人机关系的稳定性于是被放置在一个持续变动的地基上，用户的适应与系统的变迁交织在一起，单次互动越来越难承担解释重量。

未来的研究需要把用户的长期使用或者交往行为从经验描述拆解为可追踪、可编码的过程性证据，从关系历程甚至是关系史的角度展开研究。值得被纳入观察的线索包括：新奇感退潮之后，互动如何进入常态维持的节律，哪些时刻触发重新学习或重新退出；提示词策略如何从试探性尝试转向可复用的套路，再转向更隐性的分工安排，如把智能系统定位为校对者、草拟者、陪练者等；系统侧的技术变迁也应被制度化地纳入时间变量，版本更新、记忆机制调整、对齐策略变化如何与用户的主体性适应形成结构性耦合。时间维度被拉长之后，主体性在日复一日的微调中被重塑的轨迹才会显现，研究也才能从对关系状态的描述，推进至对关系生成与演变过程的解释。

（二）从悬置到嵌入：回归生活世界

部分研究倾向于采用技术中心主义叙事或抽象思辨，把人机互动从具体生活语境中抽离，

交往仿佛发生在低噪音、无干扰的纯净空间里。在宏大概念的遮蔽下，具体的人容易消失，现实中的人机关系被压缩成心理效应或技术效用的叠加。赋能悖论框架揭示了去技能化、隐形劳动与职业边界游移等生产端的结构性变化，但相关讨论大多发生在剥离了具体生活语境的组织或职业场域之内。研究者会默认人机互动是一种可以从周遭社会关系中剥离出来、单独加以测量的行为单元。但在经验层面，互动总是嵌入情境之内，并与既有的社会关系网络缠绕在一起。它可能发生在赶稿的深夜，也可能发生在家庭争执后的喘息间隙。使用者带入互动的，不只是一个待解决的任务，还有当下的情绪状态、既有的关系负担，以及对这段交往可能被他人知晓的顾虑。理解人机关系，需要把这些“噪音”视为结构条件，而非需要被清除的干扰项。

更重要的是，现有研究大多继承了一种将技术使用行为与社会生活过程相剥离的研究传统。在生活世界中，AI作为非行动者的进入方式是具体的，它会挤入家庭琐事的缝隙，成为情绪的临时支点；也会在个体最脆弱的时刻悄然在场，承接那些难以向真实他人开口的表达。在这些场景中，人机互动的意义并不由技术功能单独决定，而是在具体的社会情境、关系处境与个体经验的交织中被生产出来的。正因如此，“人们在压力下如何使用”“使用之后的代价由谁承担”“使用的痕迹在关系网络中如何留存”之类的问题，在脱离情境的研究框架内几乎无从回答。

因此，后续研究应当致力于将人机互动归置于日常生活的具体脉络之中，关注技术如何以微观方式渗透进个体的生命体验。尤其值得关注的是那些非工具性的、甚至看似无意义的互动瞬间。用户在完成任务之外与AI闲聊、倾诉、试探边界的时刻，往往正是意义被私人化赋予的关键节点。对这类互动的捕捉，有助于揭示生成式人工智能作为生活基础设施，如何在潜移默化中重构了主体的感知方式与存在状态。这些经验若不进入研究视野，人机关系便很难从悬浮的哲学论辩落到地上。

（三）从同质到异质：打破普遍性预设

现有研究常以高学历、主流语言环境的用户作为默认样本，边缘群体在高度依赖生成式系统时遭遇的特殊脆弱性，因此难以在普遍化的论断中获得应有的呈现。类似的问题也存在于情感纠缠的研究之中，当依恋机制与关系感知被作为普遍性规律加以描述，性别、阶层、文化背景对人机情感实践的差异性塑造便被一并抹平。在宏观论述的层面，人类常被视作无差异的同质化个体，大模型也被当作无差别的机器对象来对待，由此遮蔽了相当复杂的经验面向。同样，技术系统内部也绝非铁板一块。不同平台的治理逻辑、干预强度、个性化能力、数据政策，会框定关系的边界与可持续性，影响着用户能做什么不能做什么，以及关系能否延续。将“大模型”作为单一的分析对象，就如同把性质各异的社会制度归并为同一类型，遮蔽的恰恰是制度差异中最有解释力的部分。

因此，后续研究亟需将“差异”制度化地纳入分析视野。在主体维度，研究应打破标准化用户的迷思，系统性地追问不同社会位置的用户在进入人机关系时面临的结构性差异。深入关注边缘群体、数字弱势群体乃至个体等非主流文化语境下的人机实践，揭示平台权力如何与既有的社会不平等产生共振，并在技术互动中重新分配资源与风险。比较研究的路径在此尤为重要，跨群体、跨文化的系统性比较，能够将一些在单一情境研究中被视为理所当然的前提条件暴露出来，使其成为需要解释的变量而非隐含的假设。在技术维度，应当将大模型视为具有具体架构与制度嵌入性的社会技术系统。长期记忆机制、价值观对齐策略、可解释性设计与申诉通道等要素的异同，不仅改变着交往的稳定条件，也预埋了不同的脆弱点。唯有充分关照对象差异与系统异质性，研究才能触及技术落地过程中复杂而真实的影响，进而在经验描述与规范建构之间架起更可靠的桥梁。

（四）从离身到具身：拓展多感官经验

现有人机关系研究在讨论交互形态时，多数研究聚焦聊天对话场景，以文字界面为主，

用户通过键盘输入指令，系统以文本形式回应，交互在语言符号的往返中完成。当然，在很大程度上是因为大语言模型的核心能力确实首先体现在语言处理上，早期用户与生成式 AI 的接触也主要经由文字界面发生。然而，当语音对话、视频对话、AI 绘图、视频生成乃至具身机器人的应用场景迅速扩展，人机关系的研究边界也应当随之扩展。人与 AI 的交互越来越频繁地经由声音、图像、空间感知等多种感官通道展开，身体在人机关系中的参与方式也随之变得更加多元而复杂。

以语言交互为默认场景的研究框架，在认识论层面预设了一种“离身”的交往模式，交往的主体是能够抽象处理符号的认知主体，而非拥有具体身体、处于特定感官处境中的经验主体。尽管这使研究得以在相对可控的条件下处理语言层面的意义生产，但也无意中遮蔽了身体感官在人机关系中的构成性作用。事实上，感官信息是意义生成的基本条件。声音的音色与节奏影响着用户对 AI “人格”或者说性格的感知；图像的视觉质感改变着内容的可信度判断；语音交互中的停顿、语调与响应速度，在无意间塑造着用户对关系亲密程度的体验。这些感官维度所产生的效应，无法被简单地还原为语言内容的分析，也不会因为研究者的忽视而在实际的人机关系中消失。随着 AI 系统向多模态方向持续演进，离身预设与现实交互经验之间的裂缝将愈发显著。

将具身维度系统性地纳入人机关系的分析框架，将是未来研究需要考虑的。在研究对象上，语音交互、图像生成与空间感知等多模态场景应当获得与文字交互同等的学术关注，尤其是那些感官介入程度更深的使用情境。在研究方法上，生理信号采集、录像民族志等新兴具身性研究方法能够捕捉那些难以通过自我报告呈现的感官层面的反应，与访谈、问卷等传统方法形成有效互补。在理论资源上，现象学传统中关于身体图式、感知意向性与知觉经验的论述，以及具身认知研究积累的概念工具，为分析感官在人机关系中的构成性作用提供了可资借鉴



的理论支撑。

当前,随着人工智能的不断发展,技术愈发向着某种主体的角色迈进,而人类也同时感到更尖锐的“主体性危机”。一方面,智能技术为我们提供新的认知盈余与情感补偿,人们在效率红利与陪伴慰藉中获得切实收益;另一方面,大模型幻觉带来的不确定性、隐形劳动累积出的负担、去技能化引发的焦虑,也持续冲击着个体对自身能力边界与存在意义的把握。人机关系已从“人类使用工具”转变为“人机之间的纠缠”。智能技术在持续渗透的过程中,形塑着人的观念与行为,人也在通过具体的抵抗、协商与驯化重塑着技术的社会面貌。

人机关系的未来不会由某一方单方面书写,人类需要在一次次具体互动中协商边界、确认意义,未来图景才会逐步成形。这就要求研究视线从宏大理论下沉至经验层面,去捕捉个体如何在反复试错、持续修补与高频互动的微观过程中,在组织、平台与制度的多重牵引下,生长出应对风险与不确定性的具体关系机制。

在生成式人工智能突飞猛进的世代,我们要探讨的不是技术系统本身,而是一个正在被不断改写的“人”的定义。研究者其实不必执着于机器能在多大程度上替代人,而应去追问人类如何在与人工智能的长期互动中形成新的适应性智慧,最终更好地实现与人工智能的共存共生。

(责任编辑:李韵奕)

参考文献

- [1] 张铮,刘晨旭.大模型幻觉:人机传播中的认知风险与共治可能[J].苏州大学学报(哲学社会科学版),2024,45(5):171-180.
- [2] Hagar N, Agustianto W, Diakopoulos N. Not Wrong, But Untrue: LLM Overconfidence in Document-Based Queries [EB/OL]. (2025-09-29)[2025-12-23] arXiv preprint arXiv:2509.25498, 2025. <https://arxiv.org/abs/2509.25498>.
- [3] 徐帅,陈昌凤.生成的真实感:Sora的拟真特征与视听逻辑转向[J].现代传播(中国传媒大学学报),2024,46(12):76-86+96.
- [4] 曾白凌.“被中介的真理”:Sora对媒介相合性的追问[J].现代传播(中国传媒大学学报),2024,46(5):1-10.
- [5] 张梦晗,沈文乾.错位的升级与结构性失序:AI幻觉主导的信息迷雾风险与分类治理可能[J].传媒观察,2025(10):53-63.
- [6] 胡翼青,蒋艺瑶.大语言模型神话的破灭:论机器学习中的

递归与偶然[J].新闻记者,2025(8):3-13+24.

- [7] 蓝江.提丰的幽灵与代码秩序的建构——人工智能大模型时代的物体间性理论[J].新闻记者,2025(10):3-14+34.
- [8] 胡翼青,胡欣阅.作为语言基础设施的ChatGPT[J].新闻记者,2023(6):21-27.
- [9] Brandtzaeg P B, Skjuve M, Følstad A. Understanding model power in social AI[J]. *AI & SOCIETY*, 2025, 40(4): 2839-2849.
- [10] Bender E M, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? [C]//Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021: 610-623.
- [11] 王彦博,任昊炯,姚俊臣.符号权力的“智能再生产”:中文AI大模型生成记者图像分析[J].新闻界,2025(9):74-86.
- [12] 陈昌凤,王金鹏.人工智能成为新闻基础设施:技术动因、权力转移与信任重塑[J].新闻界,2025(4):4-13+52.
- [13] 黄阳坤,朱泓宇,陈昌凤.迈向“全球网关”:接入视角下的大语言模型基础设施化研究[J].国际新闻界,2025,47(6):137-158.
- [14] 郑恩,赵懋,李振康.媒介如何建构AI技术的地缘政治想象——基于DeepSeek-R1国际争议的报道框架的研究[J].新闻与写作,2025(5):85-100.
- [15] 李耘耕,姚欣桐.中美欧人工智能治理文本中的技术政治[J].现代传播(中国传媒大学学报),2025,47(4):38-50.
- [16] 方兴东,宋珂扬.全球AI治理的科林格里奇困境:多维挑战与协同治理路径[J].未来传播,2025,32(3):19-28+116-117.
- [17] 林爱璐,常云帆.人工智能大模型价值对齐的人文主义思考[J].新闻界,2024(8):24-33.
- [18] Ferrari F, Van Dijck J, Van den Bosch A. Observe, inspect, modify: Three conditions for generative AI governance[J]. *New Media & Society*, 2025, 27(5): 2788-2806.
- [19] Jarrahi M H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making[J]. *Business Horizons*, 2018, 61(4): 577-586.
- [20] Zhou Y Y, Liu Q H, Huang J H, Li G Q. Creative scar without generative AI: Individual creativity fails to sustain while homogeneity keeps climbing[J]. *Technology in Society*, 2025: 103087.
- [21] MacArthur M. Large language models and the problem of rhetorical debt[J]. *AI & SOCIETY*, 2025: 1-14.
- [22] Nieto W J. Large Language Models, and the Humanization of Nature through Artificial Communication[J]. *P&D-Philosophy & Digitality*, 2025, 2(1): 17-32.
- [23] Kosmyna N, Hauptmann E, Yuan Y T, et al. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task [EB/OL]. (2025-06-10) [2025-12-25]. arXiv preprint arXiv:2506.08872, 2025. <https://arxiv.org/abs/2506.08872>.
- [24] Van Dalen A. Revisiting the algorithms behind the headlines. How journalists respond to professional competition of generative AI[J]. *Journalism Practice*, 2024: 1-18.
- [25] 徐天博,严康.新闻“边界域”:生成式人工智能对媒介场域的再造[J].传媒观察,2024(6):58-65.

- [26] 王辰瑶,宋沁语.“大模型焦虑”何解:新闻业与大模型的“对抗-合作”式关系创新[J].新闻与写作,2025(12):18-28.
- [27] 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.
- [28] 朱琳.用AI生产爱:“智伴妈妈”情感劳动的性别化机制[J].新闻与写作,2025(10):72-84.
- [29] 束开荣.平台外包人:论幽灵工作的可见性——以人工智能数据标注员为个案[J].福建师范大学学报(哲学社会科学版),2023(3):122-133+171.
- [30] 贾文娟,颜文茜.认知劳动与数据标注中的劳动控制——以N人工智能公司为例[J].社会学研究,2022,37(5):42-64+227.
- [31] 朱鸿军,李辛扬.ChatGPT语境下版权正当性的再反思[J].国际新闻界,2024,46(9):92-110.
- [32] Kuai J. Unravelling copyright dilemma of AI-generated news and its implications for the institution of journalism: The cases of US, EU, and China[J].*New Media & Society*,2024,26(9):5150-5168.
- [33] 赵云泽,黄子婷.突破边界:人机情感交互中的伦理风险及其治理[J].东南学术,2025(5):59-72.
- [34] 徐琦.超越“计算机为社会行动者”范式:智媒时代人机传播理论创新的突破点[J].现代传播(中国传媒大学学报),2023,45(6):160-168.
- [35] 姜泽玮.超越媒体等同:人机传播用户隐私披露中的补偿性交流[J].新闻界,2025(11):70-80.
- [36] Depounti I, Natale S. Decoding Artificial Sociality: Technologies, Dynamics, Implications[J].*New Media & Society*,2025,27(10):5457-5470.
- [37] 别君华.AI是传播主体吗?人机传播的兴起及超越传播本体论的限度[J].全球传媒学刊,2024,11(3):57-73.
- [38] 闫佳琦,孙萍.流动的情感:理解智能陪伴中的人机关系[J].西南民族大学学报(人文社会科学版),2024,45(10):145-154.
- [39] 曾一果,王可心.心“真”是归处:一种情感真实主义的人机亲密关系阐释路径[J].现代传播(中国传媒大学学报),2024,46(9):123-131.
- [40] 王敏,刘鑫,吴丹.感官依恋、情感补偿与妥协关系:视障者在线亲密关系研究[J].新闻大学,2024(9):1-16+117.
- [41] 宋思茹,洪杰文,常静宜.“在断裂中纠缠”:人机情感的不稳定性考察[J].新闻界,2025(5):61-72+96.
- [42] Lin B. The AI chatbot always flirts with me, should I flirt back: From the McDonaldization of Friendship to the Robotization of Love[J].*Social Media+ Society*,2024,10(4):20563051241296229.
- [43] 周敏,赵秀丽,陈飞扬.聊天机器人的脆弱性表达是否会提高人机共情?——一项基于自主开发聊天机器人的实证分析[J].新闻与传播研究,2025,32(11):54-76+127.
- [44] 邓建国.“延展的心灵”和“心灵的延展”:人机传播研究的具身AI转向[J].新闻大学,2024(3):77-90+119-120.
- [45] 邓建国,韩志瑞.人工智能有文化吗?论人机交流中的文化之困及其化解[J].新闻与写作,2024(10):25-35.
- [46] 胡泳,王梦瑶.神经网络中的“幽灵”:笛卡尔主义与人工智能[J].新闻大学,2025(5):13-28+117-118.
- [47] Carolus A, Wienrich C. “Imagine this smart speaker to have a body”: An analysis of the external appearances and the characteristics that people associate with voice assistants[J].*Frontiers in Computer Science*,2022,4:981435.
- [48] 彭兰.“镜子”与“他者”:智能机器与人类关系之考辨[J].新闻大学,2024(3):18-32+117-118.
- [49] 刘海龙,连晓东.新常态统治的来临:ChatGPT与传播研究[J].新闻记者,2023(6):11-20.
- [50] 邓建国.新新媒介、互鉴共生:DeepSeek与人类“生态人工智能”时代的到来[J].新闻大学,2025(5):29-43+118.
- [51] 宋思茹,洪杰文,常静宜.“在断裂中纠缠”:人机情感的不稳定性考察[J].新闻界,2025(5):61-72+96.

After the Rise of Generative Artificial Intelligence: Three Research Threads and Future Directions in Human-Machine Relationship Studies

Chen Changfeng, Liang Yikun

Abstract: Over the past three years, generative artificial intelligence has been rapidly becoming infrastructural. Research on the transformation of human-machine relationships has gradually coalesced around three main threads. The first focuses on the macro-level crisis of control, examining the epistemological challenges posed by the erosion of "authenticity" and the restructuring of emergent power orders. The second centers on the paradox of empowerment in productive practice, exploring the risks of deskilling, the shifting boundaries of professional roles, and the invisible labor underlying the appearance of intelligent prosperity. The third pivots to the affective dimension of human-machine intimacy and entanglement, scrutinizing the mechanisms of human-machine attachment and the ontological implications embedded within imaginaries of symbiosis. These three threads are deeply intertwined, collectively constituting a framework for understanding the transformation of human-machine relationships in the age of intelligence. Future research should return more fully to the lifeworld, tracing the evolutionary trajectories of human-machine relationships over extended timeframes, while simultaneously dismantling homogenizing assumptions about

(下转第27页)



“概率事实”的核查： 拆解 AI 大模型的新闻生产机制¹⁾

白红义 万旭琪

摘要 AI 大模型以概率方式生产“事实”，挑战了事实核查赖以成立的前提。以往对事实核查的理解经历了“声明可验证”的专业实践与“应对信息失序的生态转向”两个阶段，均以输出文本为审查终点，遮蔽了事实背后的概率性生产机制。本文认为，在 AI 大模型的生产逻辑下，事实并非被发现与查验的，而是被数据、模型与社会反馈机制操作化生产的；核查对象需要从“虚假新闻”这一结果，转向“生产虚假新闻的系统”及参与生态。受人机交互研究“拆解 AI”视角启发，本文提出“概率事实的核查”框架，将核查视野追溯至输出、数据与生态系统三个维度，并以三个经验案例揭示各维度的实践张力与权力不对等。概率事实的核查，不仅是理解人机共生时代“何为可信事实”的一个关键命题，也为数字新闻业的媒介韧性建设提供了可操作的方法路径、可探索的经验议题。

关键词 事实核查；AI 大模型；人机交互；拆解 AI；新闻生产

中图分类号 G210 **文献标识码** A

作者 白红义，复旦大学新闻学院教授，复旦大学全球传播全媒体研究院研究员，上海 200433；万旭琪，复旦大学新闻学院博士研究生，上海 200433

一、问题的提出

事实核查是新闻业为捍卫公共知识可信度，抵抗并适应信息挑战而开展的策略实践。从信息员的核实惯例到独立核查运动兴起、国际事实核查网络（IFCN）建立，随着实践边界拓展，其功能已远不止于甄别与纠正错误声明，逐渐演变为抵御虚假信息操纵、赢得公共信任的媒介韧性机制。然而，生成式人工智能（GAI）的规模化应用使韧性机制陷入紧张。内容生产体量呈指数扩张，与此相对的是信源归属、生成主体愈发难以识别。

这场危机最集中地爆发在 AI 大模型的幻觉输出与训练数据“投毒”问题上。2023 年美国科技媒体 CNET 的 AI 写作项目在流畅专业的语体中夹杂了大量事实谬误，成功绕过传统编辑审核流程，误导了公众。2026 年中央广播电视总台“3·15”晚会曝光了 AI 大模型被“投毒”的黑色产业链，即一些 GEO 服务商通过系统性“投喂”虚假信息生成数以万计的看似合理的虚构新闻报道。这些案例凸显了一个事实，即便我们明确地知道错误可能在内容生

1) 基金项目：复旦大学全球传播全媒体研究院重点课题“AIGC 影响下的数字新闻业”。

成之际便已造就，传统的核查方法却只能望“模型黑箱”而却步，滞留在事后补救的被动局面。

新闻业的回应迅速而直接，形成了两条“以AI之道还治AI”的路径：第一种是透明标签法，披露人工智能系统参与了内容的制作或传播，编辑会在文章下方添加免责声明；第二种则是针对大模型内容特征的自动化核查，利用AI来分析潜在的错误信息、检索证据，让人类参与决策过程。然而，这两种路径都将AI的输出文本作为审查的终点，忽略了文本背后那个生产“真伪”的概率性机制。

这就引出了一个被遮蔽的问题：当我们讨论事实核查时，我们默认真实是可被验证的。但在GAI语境下，又是什么在生产“真实”？这些症结并非没有先见之明者。批判AI研究者卢克·芒恩（Luke Munn）等人在新近研究中做出诊断，真实不是被发现与查验的，而是被数据采集、模型架构、话语社会反馈机制所操作化生产的^[1]。如果接受这一论断，我们或许可以拥有新的理论视野：事实核查的对象必须从“虚假新闻”这一结果转向“生产虚假新闻的系统”及其参与生态。

为此，本文借鉴人机交互（HCI）研究中“拆解”（Unmaking）视角，提出一种面向AI大模型事实生产机制的系统性核查框架，以期为人机共生时代建立可信、韧性的信息生产秩序探寻可行之道。这一框架的适用边界以AI大模型参与内容生产或传播为前提，试图弥补既有核查认识及方法的不足，进而将核查的视野从输出与分发向前追溯至塑造AI“真实观”的三个层面——社会生态、训练数据以及生成输出。

二、事实核查的认识转向

事实核查的认识论经历了两次关键转向。它最初旨在抵御政治意识形态、维护新闻业自主性，但随着GAI带来信息生态的剧变，其实践范畴与理论认识再度构成一个值得反思的议题。

（一）作为“声明可验证”的实践：专业准则的默会前提

事实核查的传统根植于一个默会前提：存

在一种可供验证的客观真实，核查工作便是将人类言说者的“声明”（claim）与客观证据进行比对。这一专业准则预设源于20世纪中叶《纽约客》《明镜》等杂志内部纠错程序^[2]，目的是消除不真实的信息，而非引起人们的注意。进入21世纪，以FactCheck.org、PolitiFact和Fact Checker三家精英核查机构为代表的政治事实核查运动兴起，将内部程序转为公开问责实践。一套稳固的认识框架逐渐确立：核查对象是明确的声明，真相藏于可查证的事实中，证据与声明遵循可比对、可客观检验等裁决原则。

然而，当事实核查从一种技术性工序走向广阔的社会政治场域时，上述看似稳固的前提开始出现裂缝。格雷夫斯在其里程碑式研究《谁在判定真相》（*Deciding What's True, 2016*）中揭示了事实核查的实践认识论（practical epistemology）本质。通过对核查员群体的长期民族志追踪，他指出“事实可能是主观的”，核查结论的可信度来自核查者在专家网络中的位置、对证据标准的情境性判断以及对新闻叙事惯例的熟练运用^[3]。围绕“什么算是可核查的声明”，学界早已展开类似争论。批评者如约瑟夫·乌辛斯基（Joseph E. Uscinski）与赖登·巴特勒（Ryden W. Butler）认为，核查本身的选择性、复杂因果的判断隐含价值倾向，“当白纸黑字的事实与政客说法发生冲突时，事实核查者只能看到‘谎言’”^[4]。支持者米歇尔·阿玛津（Michelle Amazeen）以跨机构核查结论具有高度一致性作为反证，论证事实核查者之间存在可操作的方法共识^[5]。这展现出“声明可被验证”的脆弱性。上述争论共同揭示出，在GAI出现前，事实生产的“人为可操作性”，即核查标准与过程的社会建构性早已被学界所关注。

（二）应对信息失序的生态转向：自动化、算法审计与社会技术协同

面对社交平台汹涌而来的信息洪流，传统新闻业缓慢、审慎的核查流程不堪重负，此前的认识论反思在近十年中让位于一次务实的“生态转向”。其核心关切从“核查什么”转向以社会技术和解决问题为导向的“如何有效核查”^[6]，



致力于拓展核查的工具、基础设施与参与网络，并发展出数字行动网络、自动化核查与算法问责三条路径。简言之，前两者是向外构建生态，后者则是向内审查机制。

将事实核查视为一种社会技术生态系统，得益于其日益呈现出的数字行动网络特征^[7]。由软件开发者、数字取证专家、OSINT情报员与社交媒体平台等多方形成的协作网络，揭示了人类与技术行动者间的相互塑造关系，平台方为核查行动提供了必要的合法性、社会技术资源（如识别潜在错误信息的策略）与权威程序（分发事实核查新闻的算法系统）^[8]。

自动化事实核查的兴起是生态转型中最具代表性的项目，但也揭示了不同学科间的认知分歧。拉沙·卡夫塔拉泽（Lasha Kavtaradze）的系统性综述揭示了来自计算立场与新闻立场的认知竞争：计算机科学倾向于将自动核理解为声明检测、证据检索、立场分类等可模块化处理的工程问题，新闻学则坚持核查是以人为本，并涉及认知权威、专业判断和社会语境的实践，不应被简化为算法流程^[9]。这两种理解始终未能真正整合，使得自动化核查的实践落地远比理论设想困难。格雷夫斯的观察至今有效，最成熟的自动化系统不过是将新声明与既有核查数据库进行语义匹配，其真正自动化的程度极为有限^[10]，更多是为人类核查者赋能的技术手段。

在核查实践的生态扩张过程中，算法问责这条相对独立的研究脉络也在平行发展。以尼古拉斯·迪亚科普洛斯（Nicholas Diakopoulos）为代表的学者主张，平台算法也应成为新闻业监督的对象，这种核查行动可以通过逆向工程法实现^[11]。其可行性已得到实证研究的支持，例如，卢卡斯·格雷夫斯（Lucas Graves）与C·W·安德森（C. W. Anderson）的研究便指出，借助结构化新闻的数据工具、新的数据标准以及建立通用标签系统，专业事实核查机

构同样能参与到对搜索引擎与推荐系统的算法审计中^[12]。尽管算法问责与事实核查鲜有交集，但它为后者带来了一次重要的认识进步，算法系统作为认知基础设施也可以通过筛选来源、分配知识权重来生产事实^[13]，核查对象应从人类生产者扩展至塑造信息呈现方式的算法本身。

至此，科学技术研究与批判新闻研究传统对于“核查实践的社会建构性”已有充分积累。但既有研究的关切重点，始终是核查实践本身的人类主观性与争议性，即谁来核查、如何判断、谁的标准被采信，而非事实生产机制本身。AI大模型让问题前置，当“事实”在被任何人核查之前，就已经在模型训练数据的概率分布中被统计性地塑造，那么，核查实践的认识论转向已然不够，必须同时引入对生产机制的“技术认识论”审查。

三、当传统核查遭遇“概率事实”

GAI的出现让事实生产的主体从人类行动者变为人机混合系统。尽管人机传播理论早已预示了人类会将机器视为拟人化的社会行动者^[14]，或因交互性将机器视为信息来源^[15]。但停留在外部会将AI的运作机制简化为机械的刺激-反应序列，使机器和人类在认识论上都变得机械化。本文主张，真正的挑战在于，AI大模型以一种前所未有的方式生产“事实”，人为验证或干预AI大模型每一项决策本质上是增加了记者的任务负累。核查须深入AI大模型的技术内核及概率生产机制，以审视其系统性偏差、不确定性校准以及模型解释风格。

在AI开发中，“真实”被操作化为可供测量的基准，即“真实值”（ground truth）^①。它被定义为一组源于客观观察、未经中介的事实，模型的准确性便取决于输出与该基准的吻合度。然而，即便是AI领域内部也承认，这种泾渭分明的关系具有欺骗性，不存在原始数据，数据处理本身就是一种定义现实、界定问题并构想

① 在机器学习和人工智能领域，真实值（ground truth，简称GT）代表我们所拥有的“真相”或“事实”，指在训练和评估AI大模型时使用的已知正确答案或标签。它是模型学习的基础，也是评估模型性能的标准。

解决方案的过程^[16]。将现实中的事实转化为数据点的过程,充满了人为选择与偏见。数据选择、人工标注等环节的微小差异都会造成巨大的下游影响^[17]。从这个意义上讲,“系统的设计者拥有决定训练集所定义的世界真相的权力”^[18]。

这种权力结构给事实核查带来一系列棘手的问题。

其一,模型会复制常见的误解,即“常见词元偏差”(common token bias),当一个说法在数据集里出现的频率更高时,其被重复的可能性就越大,这可能导致流行的小道消息比罕见的真相更被模型采信^[19]。其二,它系统性地复制并固化了训练数据中内嵌的文化价值偏见。这些问题无法通过简单增加数据量解决,更大规模的模型可能反而最不真实,这也印证了大型语言模型本质上是“随机鹦鹉”的判断。

本文采用“概率事实”(Probabilistic Fact)说法定义这类由AI大模型生成的特殊陈述。其内容表面上流畅连贯,本质上却是庞大训练数据统计规律的再现,既缺乏可验证的信源,也不携带任何明确的生产意图。“概率”具有双重隐喻。技术层面,它指向大语言模型通过计算词元排序概率来生成文本的机制;认识论层面,则指向其生成内容的可靠性因技术机制而无法担保。简单来说,技术层面的概率性,是认识层面不稳定的结构根源。

需要澄清的边界是,“概率事实”与人类生产的不确定信息存在本质差异。第一,不可归因性。人类信息生产的错误原则上可追溯至具体的判断者与情境,存在可问责的主体;AI的概率性输出则混合、重组了信源,无法精准还原为某条数据,问责链在技术上断裂。第二,系统规模性。人类的失误是个体的、偶发的;AI的概率偏差是嵌入模型权重的系统性偏向,会随着用户交互被大规模复制放大。第三,技术不透明性。即便面对同一错误,人类可被追问动机,而模型的开发者自身也无法完整解释为何在某次生成中输出了某个特定词元。这三点差异,使概率事实构成一个需要被独立分析的认识论对象。

概率事实的出现,使AI大模型成为一个“不稳定传播主体”。其概率机制决定了它必然是一个系统性的、不可靠的信息生产者;但其成功且拟人的语言模仿,又诱使用户将其误认为可靠信源。这导致此前研究所积累的两大方法论路径在各自展开时陷入失灵。在以逆向工程为核心的算法问责路径中,基于提示词的输入-输出不稳定,记者无法通过重复提问、控制实验来理解生产机制。在自动化核查路径中,传统将声明检测拆分成模块化任务来处理的做法,也因概率事实并非离散声明而失效。其中,事实性内容与修辞性表达相互交织,难以被现有技术有效切割。

当“核查什么”与“如何核查”的问题无法脱离AI的生产逻辑独立回答时,两个曾经平行发展的研究传统,如今不得不面临历史性的对话时刻。事实核查转向对事实生产系统的过程审计。这一新议程要求我们将视野投向两个被传统核查忽视的关键领域,其一是溯源核查,即审计构成模型“真实”基础的训练数据来源;其二是“商业与监管生态的价值嵌入审查”,即审视何种价值判断与商业逻辑在开发中被编码进概率机制,从而系统性地影响其事实的生成与呈现。这些问题超越了传统内容核查范畴,要求新闻记者主动成为“系统审计师”,这也是新闻业在人工智能时代继续履行社会责任的必然选择。

四、拆解AI:“概率事实”的系统核查框架

为有效应对人工智能的挑战,新闻业的分析视野需要从“AI辅助报道”的应用层面,拓展至大模型开发、训练的全生命周期及其嵌入的复杂社会技术生态。AI大模型并非中立的技术工具,而是由数据管理、训练范式、开发者文化与商业模式共同塑造的产物,是特定利益与决策汇集的结果。然而,当前新闻业对其底层档案、逻辑和局限性的理解仍然有限。这一认知困境主要源于模型的固有不透明性,无论是闭源模型的数据、模型参数“黑箱”,还是开源模型被开发公司严格保密的投资结构或因



技术晦涩难以理解的对齐流程。

尽管批判 AI 的研究已致力于打开黑箱，但其工作多聚焦于特定偏见，其技术审计路径（如参数调整、模型卡）未能为事实核查者提供一套与新闻知识体系融贯的认识方法论。为此，本研究从人机交互（HCI）领域引入一个新兴视角——“拆解”（Unmaking）。该视角倡导一种启发式路径，能将对 AI 大模型生成内容的事实核查，置于其所处的更广阔社会文化背景中。

借鉴芒恩等人的框架，我们将“拆解”分为三个相互关联的层次：“拆解输出”聚焦于模型生成结果，通过提示、反思和迭代揭示其逻辑^[20]；“拆解数据”审视模型所使用的信息来源（图像和文本）；“拆解生态系统”则分析支持模型生产的价值观与组织结构。接下来，我们将在新闻事实核查领域发展这一框架，其适用边界以 AI 大模型参与内容生产或传播为前提。

（一）输出核查：识别不确定性地带

输出核查最为直接与普遍，但面临新的困惑。如何有效识别 AI 大模型生成文本中真实与虚构无缝交织的“不确定性地带”？AI 大模型随意编织训练数据中的冲突知识，并以高度自信的语态输出，形成一种修辞流畅性与事实可靠性脱节的“合成真实”（synthetic truth）。这使得核查问题从判断“对不对”转变为诘问“哪些论述有据可循，哪些没有”。

这要求记者从裁判的角色，转变为不确定性的诠释者。这与“以谦逊姿态约束真实声明”这一认识规范不谋而合^[21]。核查报道需在真与假之外容纳第三个范畴，即认识上不可判断、来源上无法确认的内容。标注这种不确定性并非降低标准，而是一种更诚实的知识实践。

这项任务的复杂性还源于 AI 大模型在核查流程中扮演的双重角色，它既是被审视对象，也是辅助核查的助手。这引出一个普遍质疑：

借助不稳定的概率事实生产者进行核查，是否会陷入“以错检错”的循环？答案在于对 AI 大模型两种功能的区分。当批评 AI 大模型时，我们指的是它不可靠的知识声明功能；当使用 AI 大模型辅助核查时，我们借助的是它高

效的计算操作功能（如海量筛选证据、语义检索）。换言之，这种做法并不是把 AI 大模型当作可信的裁判，而是将其作为扩展记者认知能力的计算工具，其输出结论必须经由人类的批判性诠释才可作为判断依据。这正是“人在回路”（human-in-the-loop）框架的意义所在——“机器完成计算，人类保留判别权威”。这就要求记者对 AI 辅助工具保持批判距离，清醒地认识到工具的结论只是判断的起点。来自密西西比大学的一项实验为记者们提供了警示，当 AI 辅助工具的检测结果与他们的初始判断一致时，记者容易陷入认知陷阱，下意识地依赖工具从而放弃对其他验证路径的探索^[22]。经过重塑的核查认知，最终需要落实到一套具体的操作路径上。

首要的便是将提示词设计本身视为一种核查行为。这种“提示词即探究”（prompting as inquiry）的思路，目的是通过与模型的反复互动来揭示其知识边界、内在偏见。对核查者而言，意味着需要对同一问题从不同角度进行多轮追问，来观察输出的一致性；设计对抗性提示，主动试探并触发模型的知识盲区。输出越不稳定，文本的不确定性区间就越大。

自动化核查在此阶段依然适用，但必须在“人在回路”（human-in-the-loop）框架下，构建人机协作的多智能交叉验证流程。自动化工具可以高效处理命题筛查、初步信息检索等大规模、重复性任务，从而将记者的宝贵认知资源解放出来，使他们能专注于情境化理解、批判性审视与综合判断等更高层次的工作。尤其当工具的结论与其他核查路径（如专家访谈、原始文档比对）产生矛盾时，矛盾点本身往往就是最有价值的调查线索。

（二）数据核查：批判信息来源

超越输出层面的表象，数据核查旨在深入模型肌理，对其训练数据进行“拆解”。训练数据从根本上塑造了模型的认知倾向与局限，其内在偏见是 AI 幻觉的主要根源。正如莉萨·吉特爾曼（Lisa Gitelman）提出的开创性观点，“没有原始数据这种东西”^[23]，训练数据是经过筛

选、策划与清洗的语料集合。这一洞见引导我们必须追问：这些数据经历了怎样的处理流程？谁在塑造、策划或审查它们？这些行动者的动机是什么？当我们将这些因素纳入考量，事实核查者便需要具备系统设计师的思维。为此，信息来源批判（source criticism）提供了一套行之有效的框架^[24]。斯蒂恩·斯滕森（Steen Steensen）等人区分了来源（生产材料的行动者，如个人或机构）与来源材料（其产物，如文件、陈述）。AI大模型的训练数据集就是“来源材料”，而生产这些材料的人类写作者、发布平台与各类机构则是“来源”。更进一步，那些决定数据取舍与过滤规则的模型开发者，我们称之为“元来源”（meta-source）。

记者的核心任务是核查预训练数据（pre-training data）的来源构成。记者无需审查模型权重，可通过分析公开证据（如技术报告、数据卡）追溯数据来源的结构特征。例如，当报告披露了数据总量但未详述各语料来源的构成比例时，记者便可追问：语料覆盖了哪些历史时段？多语言语料是否均衡代表了不同社群？

数据维度的核查观要求记者承认并揭示来源的倾向性。斯滕森认为，所有来源及来源材料都内嵌着倾向，包括来源者的认知视野、材料的生成情境、更广泛的社会文化语境，以及技术本身的可供性。记者的溯源不能止步于“这段内容从何处复制”，必须深入知识传统与文化语境的层面。例如，在分析有关国际冲突的伤亡数据时，需辨别其主要信源是倾向于官方叙事的新华社，还是代表不同立场与视角的当地媒体。

还有一种常见的反面案例是，记者在报道中简单陈述“经AI检测，该内容可能由AI生成”，却不说明使用了何种检测工具、该工具的已知误判率，以及是否进行了二次人工验证，这种做法恰恰剥夺了读者评估结论可靠性的能力，因此，诠释透明度（interpretive transparency）构成了另一项核心规范。核查记者不仅要评估来源倾向，更有责任在最终的核查报道中向公众清晰地展示自己的分析过程与判断依据。其具体操作路径包括：

一是追踪训练数据的关系网络。来源与来源材料并非孤立存在，它们总是嵌入于一个由其他来源构成的复杂网络中。记者需要追问的不仅是“语料包含哪些来源”，更是“这些来源之间存在何种关联”。例如，美联社与OpenAI签署的语料授权协议，意味着其新闻框架与写作风格将以一种有别于未授权媒体的方式影响模型^[25]。值得注意的是，数字原始资料通常包含嵌入在链接中的直接关联，这就需借鉴斯托勒·格鲁特（Ståle Grut）提出的“网络阅读”概念，并将其从跨网站阅读延伸为一种跨文本的制度性调查^[26]，通过整合技术文档、法庭记录、商业协议与劳工证词等多元信息，将这些节点连接起来，才能看清一批训练数据的真实来源图谱。

二是识别数据中的沉默部分。并非所有掌握相关信息的行动者都能成为模型的“来源”，某些群体或机构可能因技术或经济因素等无法或不愿发声。这种“缺席”在数据层面表现为特定知识权重的分布偏低。例如，当记者发现“气候变化对特定小岛屿国家的影响”这类议题在多个模型中均呈现信息缺失或失真时，其任务便是调查导致这类数据被遗漏或边缘化的结构性原因。

核查报道作为新一轮来源批判的对象，其自身就是带有特定倾向与局限的来源材料。这一提醒对容易陷入误区的AI核查报道尤为必要。一是因技术门槛而过度依赖开发机构的自述，使报道实际上成为企业公关的传声筒；二是因问责冲动而将“训练数据存在偏向”的发现，简化为“该模型不可信”的结论。自我审视能以坦诚的认知边界替代虚假的客观性。

（三）生态系统核查：记者作为系统分析者

生态系统核查将视野扩展至塑造模型的商业力量、监管框架以及利益相关者社群。如同理解传统记者的专业实践需考查商业主义的逻辑^[27]，或理解新闻机构的数字创新需分析国家政策的激励一样^[28]，理解AI模型也必须将其置于所嵌入的社会生态中。在用户输入第一个提示词之前，生态系统就可以通过施加各种利益相关者认为可取的特定目标，强有力地塑造模型的技术形态与内容产出^[29]，从而预先完成对“什么是可信事实”



的筛选与塑造。

因此，记者的角色演变为系统分析者，需要具备一种批判性的模型素养（model literacy），即能对模型的生产机制提出问题。例如，谁的数据训练了这个模型？谁的劳动（如数据标注员）定义了何为“有用”或“无害”的输出？模型的对齐过程优先考虑了谁的价值观？

具体实践中，记者需将模型与商业、监管的结构性关联作为核心分析对象。首先，审视模型的商业激励与其输出倾向的关联。正如商业主义话语将媒体机构实现商业价值的一切行为定义为“明智的”，催生了服务企业的舆情调查机构、涨粉吸流量的编辑原则^[30]，AI 的商业模式同样会通过“定义何为有价值的生产方式”来影响其对齐机制，让新闻生产的导向从告知变为服务、从公共价值变为用户黏性。例如，以用户订阅为主要收入来源的模型，其输出可能更倾向于维持用户的正向情感体验，而非呈现争议性或令人不悦的真相。当记者凭借经验识别出某段内容带有特定模型的风格时（如“这段话有很强的 DeepSeek 风格”），需要进一步探究的是，这种风格倾向究竟是展现当前技术局限，还是被商业激励所固化的设计。

其次，国家监管与产业政策的塑造作用同样不容忽视。尤其在国内，大模型的迭代周期与监管政策的出台往往紧密相连。例如，追踪分析每次新规发布后，特定模型在敏感议题上的输出边界是否发生了整体变化，这些变化是否与当时的重大社会事件存在时间强相关性。

此外，新闻业自身也需要将长期形成的核查范式与问责机制纳入“自我核查”的范畴。一个来自核查者的反身性悖论是，记者往往将自己视为“跟随者而非创新的发起者”，导致在面对技术变革时常采取“表面顺从而非自下而上的创造性应对”。当前新闻机构引入 AI 的模式便是一个例证。AI 常被定位为提升效率的工具，记者则成为其输出的“审核员”，但审核的标准、深度与责任归属往往模糊不清。新闻编辑室在拥抱 AI 技术的同时，必须进行深刻的内部反思与核查机制建设。

面向生态系统的核查周期更长，方法上更接近深度调查和机构报道。

第一步，开发机构与商业调查。审查模型开发机构的资本结构与商业合作方、追踪对齐标准的制定文件（如公开的标注指令、安全政策）。由于企业透明度有限，这项工作常需要综合分析新闻稿、财报、创始人访谈、技术论坛帖子乃至在线社群讨论等多种信息源，进行类似数字民族志的探查。

第二步，叙事来源的网络追踪。针对 AIGC 虚假信息，从孤立声明转向对整个叙事网络的分析，关注其协调性、背后账号网络的联动模式。

第三步，使用者动机与影响分析。调查谁在积极传播特定的 AIGC，他们在何种语境下传播，以及这些传播行为旨在达成何种政治或商业目的。这一环节将核查的闭环从内容生产延伸至其社会影响，完整地呈现了 AIGC 在当代信息生态中的角色与作用。

五、实践中的“拆解”

为将理论框架付诸实践，核查过程需被细致地“拆解”。鉴于尚无任何单一案例能完整覆盖输出、数据与生态系统这三个维度，本研究将采用分层剖析的方法，以澎湃明查的实践探讨输出核查，以《华盛顿邮报》的调查审视数据核查，并以斯坦福互联网观察站的报告分析生态核查，最终构建一个多层次的核查行动图景。

（一）澎湃明查“以 AI 辨 AI”的输出核查实践

输出核查面临的挑战，是如何在 AI 流畅、自信的文本表面之下，识别那些来源无法确认、真实值无法标注的不确定性地带。澎湃明查·AI 实验室进行了两轮探索。第一阶段的探索始于 2023 年末，早期测试对象 BingChat、文心一言与 Perplexity AI 均出现幻觉与误判。在一个特定测试中，它们对 ChatGPT 现场生成的且被检测器判定为人类书写的虚假推文进行核查，合计得分仅为 14 分，未能及格^[31]。这揭示了 AI 作为核查工具与新的不确定性风险源的双重角色。

到了 2025 年，第二阶段测试的对象更新

为 Claude Sonnet 4、ChatGPT-5、豆包与 DeepSeek，结果呈现出显著飞跃，模型平均分高达 19.125/20^[32]。明查转向探究模型“如何作答”的推理方式。如豆包倾向于引用权威来源并补充传播动机分析，ChatGPT-5 能识别并保留虚假陈述中的真实元素，而 DeepSeek 则直接援引包括明查自身在内的专业核查报告，并提供识别此类信息的通用建议。这一转变标志着明查方法论的成熟，开始深入理解“模型如何思考”。

与文本测试并行，明查针对 AI 生成图像建立了一套包含四个观察点（手部细节、配饰一致性、图中文字、背景失真）的检查框架，结合 EXIF 元数据分析、反向图像搜索构建了多层交叉核验体系。对于深度伪造视频，则采用 Deepware 等工具分析与人工判读结合的方式。这些实践背后贯穿着“以 AI 辨 AI”的对抗性思路：通过“变着法子提问”触发模型的不稳定性，才能找到幻觉的边界。例如，在“大模型读图”实验中，Grok 对“桦加沙台风前被‘五花大绑’的雕塑”AI 图片给出截然相反的结论，ChatGPT 在改变提问方式后撤回判断^[33]。这些不稳定性正是识别合成真实的裂缝所在。

基于以上探索，明查创新性引入“0.5 分”评分，专门用于标记模型在无法确定时“提示用户自行甄别”的回答。这一设计体现出可信度源于解释过程的透明度，而非结论的绝对确定性。有研究指出，这种透明度在新闻实践中具有双重性，核查工具与证据链的公开属于规范性透明，而选题决策等内部流程则遵循已有的策略性透明。二者共存揭示了“诠释透明度”在实践中存在的特殊张力，而非二元对立^[34]。这也成为推动明查将其方法制度化的原因。

（二）《华盛顿邮报》对谷歌 C4 数据集元来源的核查实践

若说输出核查揭示的是概率事实的症状，数据核查则试图追溯其病因。到底是什么样的训练数据结构，使得特定内容以较高概率被模型生成成为“事实”？《华盛顿邮报》对谷歌 C4 数据集的核查，是迄今最具代表性的尝试，它展示了记者如何以系统设计者的思维，进入训

练数据的来源关系网络，揭示概率事实的数据生成机制。该项目首次拆解了谷歌的 C4 数据集（Colossal Clean Crawled Corpus），而 C4 作为 Common Crawl 网络爬取数据的关键快照，被广泛用于训练如 Google T5 和 Meta LLaMA 等主流模型^[35]。这项核查的价值在于，它展示了核查记者如何用系统设计者的思维来追溯海量数据集的来源。

核查的第一项工作是识别关于训练数据的数据来源。C4 数据集的内容构成由两套权力关系共同决定，其一是 Common Crawl 的爬取机制，它优先抓取最重要、最权威的网站，但并不回避受版权保护的内容；其二是谷歌在此基础上施加的过滤层，用以定义什么是干净的语料。正是这第二层机制使谷歌成为所有训练语料的“元来源”，它不生产任何内容，却通过制定过滤规则，对所有下游来源材料拥有底层塑造权。

紧接着，核查团队对来源关系网络进行全局追踪。团队最初获得的只是排名前 1000 的域名，但这些域名仅占词元总量的 8%。发现这一局限后，报道团队进一步协商，从 Allen Institute for AI 获取了全部 1570 万个域名，随后与网络分析公司 Similarweb 合作，将约 1000 万个可识别域名归类为新闻、科技、医疗等行业类别。这种跨越技术文档与商业工具的核查逻辑意味着，记者需要将多个异质来源连接起来，才能看清一批训练数据的真实图谱。此外，这张图谱中最值得关注的，是那些沉默的结构。谷歌的过滤机制在清除色情内容的同时，系统性地过滤了非性相关的 LGBTQ+ 内容。

核查团队还披露了一种利益共生关系。Common Crawl 由亚马逊赞助，而《华盛顿邮报》的所有者正是亚马逊创始人贝佐斯。这一自我审视，是对“核查报道本身也是来源材料”这一反身性命题的实践，代表着记者没有回避这个结构性关联，而是将其写入报道，以坦诚的认知边界替代了虚假的客观。

（三）斯坦福互联网观察站（SIO）的生态核查与自我消逝

生态系统核查的操作对象是制度力量与行



为网络，它们塑造着 AIGC 以何种轨迹被传播与被接受的。然而，当前还没有任何核查机构完整地针对某一 AI 大模型的生态链条开展过系统报道。本文将斯坦福互联网观察站（SIO）的案例作为“方法论参照”，虽然 SIO “追踪叙事网络中的非协调行为”这一核查思路诞生于大模型普及之前，但在 AIGC 时代具有直接的延伸价值。当 AI 能够以极低成本炮制高度拟真的内容时，真正难以应对的不只是单条概率事实，而是那些被有组织部署与放大的整个叙事传播网络。

为此，SIO 在 2019 年至 2024 年间的实践正是生态核查可行的操作路径。但遗憾的是，其自身遭受政治反噬而解体，这一命运揭露出生态核查所面临的权力结构风险。在 SIO 于 2022 年与 Graphika 联合发布的报告《沉默的声音》（Unheard Voice）中，其通过分析 Twitter 和 Meta 等平台移除账号后提供的数据集，识别出了一个跨平台、持续近五年推广亲西方叙事的操作网络^[36]。分析的关键证据包括集中的账号注册时间、异常的内容发布节奏、一致的跨平台叙事等行为指标。这在 AIGC 时代尤为关键，尽管 AI 生成的内容在事实上难以直接证伪，但其背后协调一致的传播行为痕迹依然可以被追踪和识别。

同时，SIO 对平台私有数据的依赖构成生态系统核查的关键门槛。其研究能力高度依赖于平台在完成内容封禁后，授权其访问相关数据集。这种“先处置、后研究”的合作模式，构成了个体记者或小型机构难以复制的优势。尽管如此，记者仍可以参照 SIO 在个体层面做简化版本，比如追踪一个关键词或叙事框架，而非全量数据。

最具反思意义的是，SIO 自身的终结成为生态系统压力最深

刻的反身性案例。自 2023 年起，SIO 因其在选举虚假信息方面的研究，遭受了来自保守派议员和法律团体的持续政治攻击，被指控为“政府审查的共谋”，斯坦福大学最终选择退却。随着核心研究人员离职和研究方向的终止，SIO 实质解体^[37]。这一遗憾的结局揭示了生态系统核查背后的风险——旨在揭示政权虚假言论的核查机构，其自身也无法脱离于同一套权力生态的博弈与反噬，而那些本应被核查的政治、法律与资本力量，最终也直接作用于核查者自身。SIO 案例提示，对 AI 大模型的生态核查，从一开始就不只是技术问题或方法论问题，而是一场在现实权力与资本结构博弈中开展的新闻行动；核查者面对被核查对象显得脆弱且力量羸弱，这种不对等的约束是实践中必须提前纳入考量的条件。

六、结论与讨论

本文关注人工智能新闻生产中的“概率事实”问题及其对传统核查路径的挑战。为此，研究主张将 AI 大模型开发与数据训练的全生命周期，乃至所嵌入的广阔社会文化生态纳入核查范畴。受“拆解 AI”视角启发，本文提出“概率事实的核查”这一面向 AI 大模型情境的系统性框架，以剖析新闻生产中由模型引发的偏差、危害和风险，从输出、数据与生态系统三个维度为核查实践提供新的认识与操作路径（表 1）。在此基础上，本文就框架内部的张力与权力不对等问题，提出两点看法。

表 1 概率事实的核查框架

	输出核查 (识别不确定性地带)	数据核查 (批判信息来源)	生态系统核查 (记者作为系统分析者)
核查对象	一种介于真实与虚假之间的“合成真实”。	预训练语料的来源（语言社群、历史时段、被排除或纳入的知识传统）。	塑造模型“可信事实”的制度力量，如商业激励、国家监管与利益相关者；核查机构自身。
核查认识	记者作为不确定性的诠释者，比如在核查报告中标注不确定的部分。	承认所有来源具有倾向性；诠释透明度，向公众展示自己的分析过程与判断依据。	具备批判性模型素养，追问谁的价值观被编码进了对齐机制、标注标准与内容筛选规则。
操作路径	提示词即探究，在迭代追问中试探模型能力边界；人机协作的交叉核查。	追踪训练数据的关系网络；识别数据中的沉默部分；对核查报道本身进行来源批判。	开发机构与资本调查；叙事来源的网络追踪；使用者动机与影响分析。

首先，三个核查维度在优先级、资源投入上存在实践张力。输出核查最为直接，是当前主要路径。但这意味着记者常处在信息链末端。引入数据与生态系统核查，旨在形成完整的认识框架：训练数据决定了何种知识被概率性地塑造成“事实与真相”；而商业与制度生态，则决定了何种与新闻专业性冲突的“有用性”（比如流量新闻导向、娱乐化新闻导向）被设为模型对齐的目标。新闻业在核查机制建设上需充分考虑这些关联。

其次，三个核查维度不具有同等的权力效应，其深度与难度与记者能力、技术可及性及权力关系高度相关。输出核查是记者“一人配备多个AI”就可以独立开展的工作。数据核查的可及性取决于开发者是否披露训练集信息，并受制于商业激励与监管压力。生态系统核查则需要记者具备制度调查的能力，其可行性受制于数据获取权限、信源关系以及机构支持。这种权力上的不对等让后两种核查路径在现实中相互掣肘。

当我们接受概率事实的核查思路时，它将为数字新闻业抛出一系列新的思考方向。以往研究对“事实核查作为媒介体制韧性”持有共识^[38]，并将AI核查视为建立人机信任、迈向系统信任新闻业的途径^[39]。倘若以此为指引，AI大模型时代的核查韧性与信任建立至少还需要回答——当核查对象的真实来源无处追溯时，韧性何处生长？面对概率事实，只有新闻生态系统中的专业媒体、平台媒体、自媒体与社群媒体四类新闻行动者^[40]参与的核查韧性机制真的足够吗？当涉及核查议题时，模型开发者（包括算法开发者、产品、数据运营者在内）是否应被纳入新闻行动者的范畴？如果是的话，又将如何与专业核查记者分配代理权？因此，未来还可以在实证层面继续发展以下议题，比如核查机构如何在实践层面应对来自输出、数据与生态的资源约束；当事实核查者进入模型系统内部与开发者展开协作时，在对大模型核查之前，如何让人类核查者在理念上做出价值对齐等。

最后，我们还必须正视全球化与本土特殊

性。当前以澎湃明查为代表的本土化探索尚在起步，虽已开启AI辅助核查的实验，但在其他方面仍有较大发展空间。相比之下，西方对人工智能训练数据的揭秘展现出纵深探索趋势，也为本土实践提供了重要参照。比较不仅有助于揭示文化制度偏向，也为理解AI时代事实核查的全球多样性，提供一条富有价值的研究路径。

（责任编辑：徐秋染）

参考文献

- [1] Munn L, Magee L, Arora V. Truth machines: synthesizing veracity in AI language models[J]. *AI & Society*, 2024(39): 2759-2773.
- [2] Sivek S C, Bloyd-Peshkin S. Where Do Facts Matter?[J]. *Journalism Practice*, 2018, 12(4): 400-421.
- [3] Graves L. *Deciding what's true: Fact-checking journalism and the new ecology of news*[M]. New York: Columbia University Press, 2016:1-2.
- [4] Uscinski J E, Butler R W. The Epistemology of Fact Checking[J]. *Critical Review*, 2013, 25(2): 162-180.
- [5] Amazeen M A. Revisiting the Epistemology of Fact-Checking[J]. *Critical Review*, 2015, 27(1): 1-22.
- [6] Westlund O, Belair-Gagnon V, Graves L, Larsen R, Steensen S. What Is the Problem with Misinformation? Fact-checking as a Sociotechnical and Problem-Solving Practice[J]. *Journalism Studies*, 2024, 25(8): 898-918.
- [7] Weikmann T, Lecheler S. Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network[J]. *Digital Journalism*, 2024, 12(10): 1505-1522.
- [8] Bélair-Gagnon V, Larsen R, Graves L, Westlund O. Knowledge Work in Platform Fact-Checking Partnerships[J]. *International Journal of Communication*, 2023, 17(0): 1169-1189.
- [9] Kavtaradze L. Dominant Disciplinary and Thematic Approaches to Automated Fact-Checking: A Scoping Review and Reflection[J]. *Digital Journalism*, 2025, 13(9): 1552-1577.
- [10] Graves L. Understanding the Promise and Limits of Automated Fact-Checking[EB/OL]. Reuters Institute for the Study of Journalism, University of Oxford, 2018. <https://ora.ox.ac.uk/objects/uuid:f321ff43-05f0-4430-b978-f5f517b73b9b/files/smc87ps17m>.
- [11] Diakopoulos N. Algorithmic accountability: Journalistic investigation of computational power structures[J]. *Digital journalism*, 2015, 3(3): 398-415.
- [12] Graves L, Anderson C W. Discipline and promote: Building infrastructure and managing algorithms in a “structured journalism” project by professional fact-checking groups[J]. *New media & society*, 2020, 22(2): 342-360.



- [13] 万旭琪. 新闻游渗: 数据流如何更新新闻认识论? [J]. 新闻大学, 2025(9):32-47+117-118.
- [14] Nass C, Moon Y. Machines and mindlessness: Social responses to computers[J]. *Journal of Social Issues*, 2000, 56(1): 81-103.
- [15] Sundar S S, Chen J. From CASA to TIME: Machine as a source of media effects[C]//Guzman A L.ed. *The SAGE handbook of human-machine communication*. SAGE Publications, 2023.
- [16] Kozyrkov C. What Is “Ground Truth” in AI? (A Warning.)[EB/OL]. Medium, (2022)[2026-03-30]. <https://towardsdatascience.com/in-ai-the-objective-is-subjective-4614795d179b>.
- [17] Jaton F. *The constitution of algorithms: ground-truthing, programming, formulating*[M]. Cambridge: MIT Press, 2021.
- [18] Crawford K. Excavating “Ground Truth” in AI: epistemologies and politics in training data[EB/OL]. UC Berkeley, (2022)[2026-03-30]. https://www.youtube.com/watch?v=89NNrQULm_Q.
- [19] Zhao T Z, Wallace E, Feng S, Klein D, Singh S. Calibrate before use: improving few-shot performance of language models[J]. *arXiv:2102.09690*, 2021.
- [20] Munn L, Khan A H, Hettiachchi D, Sabie S, Ghahremanlou L, Saghir S, Lambourne N, Magee L. Unmaking AI: Engaging Critically and Creatively with Generative AI[C]//Proceedings of the 36th Australasian Conference on Human-Computer Interaction (OzCHI '24). *Association for Computing Machinery, New York, NY, USA*, 2025: 892-897.
- [21][24] Steensen S, Belair-Gagnon V, Graves L, Kalsnes B, Westlund O. Journalism and Source Criticism. Revised Approaches to Assessing Truth-Claims[J]. *Journalism Studies*, 2022, 23(16): 2119-2137.
- [22] Sohrawardi S J, Wu Y K, Hickerson A, Wright M. Dungeons & Deepfakes: Using scenario-based role-play to study journalists' behavior towards using AI-based verification tools for video content[C]//Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). *Association for Computing Machinery, New York, NY, USA*, 2024: Article 776, 1-17.
- [23] Gitelman L ed. *Raw data is an oxymoron*[M]. Cambridge, MA: MIT press, 2013.
- [25] 财联社. 扛着“AI 语料版权战”大旗的美国上市媒体签下首份 AI 授权协议 [EB/OL].(2025-05-29). <https://mp.weixin.qq.com/s/d6njOwUIVczi6v8KJ0d-4g>.
- [26] Grut, S. *Digital Kildekritikk*[M]. Oslo: Gyldendal akademisk, 2021:27.
- [27][30] 李艳红, 陈鹏. “商业主义”统合与“专业主义”离场: 数字化背景下中国新闻业转型的话语形构及其构成作用 [J]. 国际新闻界, 2016, 38(9):135-153.
- [28] Fang K, Repnikova M. The state-preneurship model of digital journalism innovation: Cases from China[J]. *The International Journal of Press/Politics*, 2022, 27(2): 497-517.
- [29] Vertesi J, Ribes D, Forlano L, Loukissas Y, Cohn M L. Engaging, designing, and making digital systems[C]. Felt U ed. *The handbook of science and technology studies*. 2016: 169-194.
- [31] 明查·实验室. AI 造假对决 AI 核假, 谁占上风? [EB/OL]. 澎湃新闻, (2023-06-17)[2026-03-29]. https://www.thepaper.cn/newsDetail_forward_23507478.
- [32] 明查·实验室. “AI 核查员”上线, 四大模型谁最靠谱? [EB/OL].(2025-11-18)[2026-03-30]. 澎湃新闻, https://www.thepaper.cn/newsDetail_forward_31966879.
- [33] 明查·实验室. 这个“六指小男孩”大模型知道 TA 是 Ai 造的吗? [EB/OL].(2025-11-19)[2026-03-30]. 澎湃新闻, https://www.thepaper.cn/newsDetail_forward_31983647.
- [34] Zhang H. Fact-checking in China: normative and strategic transparency of Chinese journalists in fact-checking reports[J]. *Asian Journal of Communication*, 2025, 35(2): 81-99.
- [35] The Washington Post. Inside the secret list of websites that make AI like ChatGPT sound smart[EB/OL].(2023-04-19)[2026-04-10]. <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>.
- [36] Graphika and Stanford Internet Observatory. Unheard Voice: evaluating five years of pro-Western covert influence operations[EB/OL]. Stanford Digital Repository, 2022. <https://purl.stanford.edu/nj914nx9540>.
- [37] The Washington Post. How The Washington Post uncovered the sources that make AI chatbots sound so smart [EB/OL].(2024-06-14)[2026-04-01]. <https://www.washingtonpost.com/technology/2024/06/14/stanford-internet-observatory-disinformation-research-lawsuits-politics/>.
- [38] 刘于思, 闫文捷. 虚假信息应对韧性的生成条件——基于 38 个数字媒介体制的比较分析 [J]. 新闻记者, 2025(11):97-112.
- [39] 白红义, 万旭琪. 迈向系统信任的数字新闻业: 一个机器自主性的视角 [J]. 新闻界, 2025(6):4-16.
- [40] 吴涛, 张志安. 新新闻生态系统中的事实核查及对新闻真实观的影响 [J]. 新闻与写作, 2022(7):37-45.

Fact-Checking Probabilistic Facts: Unmaking the News Production System of Large Language Models

Bai Hongyi, Wan Xuqi

Abstract: Large Language Models (LLMs) produce “facts” probabilistically, challenging the epistemic foundations

of traditional fact-checking. The understanding of fact-checking has evolved from a professional practice of verifiability to an ecological turn against information disorder. Yet both paradigms treat the final textual output as the endpoint of scrutiny, thereby obscuring the probabilistic mechanisms of production. This article argues that, under the production logic of large language models, facts are not discovered but are operationally produced through a confluence of data, models, and social feedback. Consequently, fact-checking must shift from “false news” as outcome to the systems that produce false news and their participatory ecosystems. Drawing on the “Unmaking AI” perspective from Human-Computer Interaction (HCI) research, this study proposes a framework for checking “probabilistic facts” that extends the verification lens across three dimensions, which are output, data, and ecosystem. Through three empirical cases, it illuminates the practical tensions and power asymmetries within each dimension. This approach not only reframes the foundational question of what constitutes a credible fact in an era of human-AI symbiosis but also offers actionable pathways and new empirical agendas for building media resilience in digital journalism.

Keywords: fact-checking; large language models; human-computer interaction (HCI); Unmaking AI; news production

Authors: Bai Hongyi, School of Journalism, Institute for Global Communication and Integrated Media, Fudan University; Wan Xuqi, School of Journalism, Fudan University.

(上接第 15 页)

users. By investigating heterogeneous experiences within embodied and contextualized human-machine interactions, scholars may chart pathways toward coexistence with artificial intelligence.

Keywords: intelligent communication; human-machine relationships; generative artificial intelligence; human-machine communication

Authors: Chen Changfeng, Institute of International Communication, Jinan University; School of Journalism and Communication, Tsinghua University, Beijing. Liang Yikun, School of Journalism and Communication, Jinan University.



再造可信生态：AI 幻觉风险下主流媒体事实核查的价值与权威重塑¹⁾

杨奇光

摘要 传统事实核查多侧重于对虚假信息的事后纠错和纠偏，然而 AI 幻觉带来的“后真相递归”导致失真信息被重复引用、再生产，进而形成自我循环的信息污染。作为关涉意识形态安全和政治信任的主流媒体，其事实核查的开展不仅要宣告事实真伪层面的核查结论，还要通过真相的供给介入塑造社会现实的过程，并在系统性的信息治理中发挥能动价值作用。基于数字新闻认识论的理论主张，主流媒体的事实核查可以被界定为一套以新闻真实性准则为核心、以 AI 幻觉风险应对为旨归的系统性、专业化的真相确证实践，并于此过程中，主流媒体在主体可信性、信息可信性和系统可信性的层面重塑新闻业权威，彰显信息生态治理的专业价值。

关键词 事实核查；AI 幻觉；主流媒体；可信性；新闻权威

中图分类号 G210 **文献标识码** A

作者 杨奇光，中国人民大学新闻与社会发展研究中心研究员，中国人民大学新闻学院副教授，北京 100872

引言：从文本失真到幻觉迷雾

当前生成式人工智能（GAI）已泛在性地介入人类日常社会生活，互联网传播格局的变化体现为，从 Web2.0 时代“人人都有麦克风”的个性化信息传播，向人工智能时代“人人都能生成内容”的智能传播范式转向。在新闻生产领域，AI 大模型的介入可以模拟新闻报道、政策解读甚至专家评论，但产出的内容往往缺乏准确的事实来源，也难以经得起严密逻辑推敲，这也给事实核查实践带来新的挑战。与谣言传播、虚假新闻等多集中于文本层面的信息失真形态不同，“AI 幻觉（AI hallucination）”引发的信息失序已超越单一文本的真伪判定范畴。通俗而言，AI 幻觉是指大语言模型生成看似合理，但实则存在事实错误或逻辑谬误等问题，尽管这些内容在表面上体现为语法完整、逻辑连贯以及文体逼真等特点^[1]。

大语言模型生成的内容通过拟制近似专业的新闻语态并在社交媒体平台上扩散，后经由用户的再加工和模型再吸纳的循环后，最终演变为难辨真伪、

1) 基金项目：国家社科基金青年项目“智能新闻生产中人工智能幻觉的风险识别与协同治理研究”（25CXW074）。

极具迷惑性的虚假信息迷雾。人工智能生成内容（AIGC）的信息源头通常难以锁定和溯及，这导致传统事实核查所依托的事后纠错和纠偏以及人工求证等主要方式面临速度滞后等局限。与此同时，AIGC在新闻认识论层面还引发了何种信息是真实信息、何种主体能够判断真伪以及如何通过事实核查确证真相等一系列复杂性认知问题。

既往关于事实核查的研究较多关注核查的实践经验 and 效果评估，但是对于事实核查的主体价值以及事实核查的理论供给方面研究不足。党的二十届三中全会站在党和国家事业发展的战略全局，作出了“推进主流媒体系统性变革”的重大部署，“十五五”规划中也将“主流媒体系统性变革”作为大力繁荣文化事业的重要抓手。主流媒体作为内嵌于社会系统结构的沟通网络，关系着主流意识形态安全和政治信任^[2]。在AI幻觉引发信息迷雾的风险面前，传统分散式的事实核查模式已无法充分应对，相较于商业媒体、自媒体和第三方机构，主流媒体因其关键的意识形态属性、公共服务使命和组织新闻实践的专业能力，应勇于挺进包括事实核查等在内的信息生态治理主战场，充分发挥应对AI幻觉这一新型系统性风险的牵头者和引领者角色。但是，主流媒体如何发挥信息鉴别的专业主体性作用较少受到学术研究的关注，现有研究也未能从主流媒体系统性变革的战略高度探究主流媒体如何发挥专业示范效应进而引领AI时代的信息治理。

本文基于数字新闻认识论的理论视角，主张在既往事实核查实践经验的基础上，进一步结合中国主流媒体事实核查的主体类型和实践价值等维度，重新锚定主流媒体在事实核查中所应扮演的主体角色，同时倡导主流媒体在生成式AI介入信息生态的背景下发挥重要的规范性作用，进而在主流媒体系统性变革进程中重塑自身的可信权威。

一、事实核查的路径转向与可能困境

（一）事实核查：从新闻样式到认识基础

设施

自现代新闻业诞生以来，挖掘真相始终是新闻从业者的终极追求，也成为新闻业承载公共信息传播使命的重要基础，新闻从业者群体因此也被描述为“真相寻求者（truth-seekers）”^[3]。事实核查是新闻业求证事实、宣告真相的重要实践活动，其延续着调查新闻的历史传统，不同类型的事实核查机构则是锁定可疑信息、开展科学调查并公开做出事实判断的主体力量。过去十余年来，事实核查成为一类全新的、独立于传统新闻观念的新闻样式，全球事实核查运动发展迅猛，并已在全球信息治理领域形成了显著的影响力与行业话语权^[4]。数字信息生态中的失真问题并非简单的内容真伪之争，而是平台结构、传播动机与认知偏差共同导致的信息污染，因此，事实核查的重要性更加凸显，其也被视为新闻业在信息失序日益加剧背景下，为重申新闻业坚守真实性原则而发展出的制度创新^{[5][6]}。

与此同时，事实核查面临着理论建构和认知维度的诘问，既往对于西方事实核查的研究指出，事实核查在内容筛选与价值评估环节不可避免地涉及主观判断，其客观主义的核心诉求与既有的真理体制之间形成了内生性张力^[7]。此外，事实核查在纠正公众错误认知、遏制不实信息与恶意虚假信息传播方面的实际效果究竟如何，学界对此仍存争议^[8]。面对智能传播环境中对于真相的认知难题，有学者提出将“事实核查作为认识基础设施”^[9]的命题，旨在强调在算法深度介入的社会中，事实核查活动已经转变为塑造公共知识与公民信任的底层机制。上述理论观点一方面延续了新闻认识论的基本主张，回应了新闻工作者如何确证自身所知内容的来源与边界，以及其所获取的核查结论如何能够被阐释和检验的问题；另一方面，这一理念主张还将事实核查纳入了知识建构的范畴，即事实核查如何能够成为一种独立的知识形态。

在具体实践层面，2018年建立的中国互联网联合辟谣平台构筑了全国网络辟谣联动机制，2025年中央网信办已部署开展“清朗·整治AI



技术滥用”专项行动。与此同时，我国主流媒体也在积极应对智能传播环境下的信息安全问题。例如，作为国家级通讯社，新华社正在开展“新华智信生态建设行动”，通过依托媒体融合生产技术与系统国家重点实验室探索筑牢信息生态安全防线^[10]。尽管事实核查在促进新闻业价值彰显过程中起到了积极的作用，但是对于事实核查的质疑以及实践发展过程中事实核查所面临的新挑战仍有待持续性研究。

（二）AI 幻觉与事实核查困境

在 GAI 技术介入的智能传播环境背景下，人类作为造假主体依然可能会蓄意制造和传播虚假信息，同时，大语言模型内嵌有算法系统内生性失真的底层技术属性，这二者的叠加效应，使得事实核查面临着更为复杂的局面。面对多模态、跨平台、隐蔽化的 GAI 引发的虚假信息，仅凭传统人工核查的方式已难以应对。如前文所述，GAI 所引发的信息风险并非表现为孤立的文本失真，在大模型幻觉问题难以解决的前提下，智能生成内容可能会引发谎言驱逐真相的“递归效应”^[11]。研究表明，相较于纯文字类信息来说，当某一虚假信息以“图+文”形式呈现时，用户对此类信息的信任度会更高^[12]。表面来看，这意味着 AIGC 给事实核查带来了更大挑战，但是从深层次剖析原因便会发现这一现象更加具有讽刺意味——人类“眼见为实”式的认识真相的方式，正在被 AIGC 所制造的多模态信息环境所解构，并在幻觉信息迷雾中走向迷失。

如果将传统新闻业开展的事实核查和智能传播环境下的事实核查相比较可以发现，二者在核查对象、责任归依及纠偏方式等方面存在显著差异。传统新闻业开展事实核查的主体主要是职业新闻从业者和新闻媒体机构，主体类型相对来说较为单一且明确。事实核查的对象也主要是依赖可核验和可溯源的事实，通常涉及新闻文本中的数据、言论和背景信息等。在责任归属判定上，传统事实核查主要以信源是否权威为依据。例如，如果核查发现了某类事实性偏差，那么可以将责任归属于新闻文本的

创作者或可能的信源提供方。在纠偏方式上，既往事实核查结果以媒体机构公开声明为主，且纠偏环节的开展多先于新闻报道正式刊发前。面对 AI 幻觉可能引发的风险隐患，智能传播环境下事实核查的困境和挑战包括但不限于以下：

其一，事实核查的对象不仅包括新闻报道中的存疑性事实，还拓展至社交媒体平台上的碎片化虚假信息、误导性叙事、圈层化传播的阴谋论，甚至还包括为了制造对立而编造的各类裹挟着偏见、情绪化的信息。当各类经由 GAI 制造的失实信息可以被批量化、低成本地制造出来，事实核查若仍以逐条甄别、逐篇回应为应对方式，势必会陷入被动状态。

其二，在责任归属的判定上，事实核查不仅需要追究内容原创者的主体责任，还要考虑平台算法设计责任，以及关键传播节点上“舆论意见领袖”的责任。在纠偏过程中，事实核查不仅需要澄清错误事实，还要披露导致不实信息的技术原因、传播链条等，此外，还需建立以算法筛查、跨模态比对、传播溯源和编辑复核为支撑的人机协同机制^[13]。

其三，传统意义上的假新闻常常是以新闻形式出现的违背客观事实、故意虚构或歪曲的信息，但是 AI 幻觉的棘手之处在于其高度拟真的证据外观。当人们因某一信息形式上的规范性而将其误读为一类可靠信息时，事实核查就已经陷入被动局面，因为这意味着，事实核查既要宣告事实本身之伪，还要告知用户其依据信息外观形式做出判断是不可靠的，后者因直接挑战用户自我认知，极易导致说服失效并加剧核查困境。

当前研究已关注到如何防范 AIGC 风险的策略，现有方案主要包括两类。一是为大语言模型建立“知识边界”，通过自监督机制识别已知与未知信息；二是在内容生成过程中引入外部知识库进行实时检索与事实核查，打破已有知识的静态局限^[14]。此类研究形成的结论多为依赖技术路径、面向通用语境的策略，但是对于以新闻业为主体，特别是以主流媒体为主体的事实核查实践的适配性仍显不足。

二、新闻认识论视角下主流媒体事实核查的价值重估

从新闻认识论（特别是数字新闻认识论）的理论视野来看，对于主流媒体事实核查的价值重估可以在媒介社会学范式下的新闻认识论传统中找到学理依据^[15]。主流媒体事实核查可以被视为（智能）技术驱动背景下的社会建构产物，当大语言模型、平台和用户都在竞相争夺关于事实真相的定义权时，主流媒体更要发挥制度性优势，借助事实核查这一专业实践，重新建立起自身和公众关于事实核查主体资格、核查判定标准以及核查权威的认识^[16]。

一方面，从专业实践价值来看，主流媒体事实核查可以被界定为一套以新闻真实性准则为核心、以应对 AI 幻觉风险为旨归的系统性、专业化的真相确证实践，同时也可被视为智能传播环境中保障公共信息真实性的媒介化治理实践。另一方面，从社会建构的意义来看，主流媒体事实核查所遵循的规则体系与操作标准，是掌握主流话语权的专业机构与行动主体围绕事实的定义权、核查的解释权展开制度化协商的产物，其价值在于为数字公共领域确立统一的事实基准，进而在凝聚社会共识、守护主流价值的维度发挥积极作用。

（一）防范系统性安全风险的真相供给价值

事实核查所面临的困境与挑战多与技术诱因直接相关，通常而言，AI 幻觉通常表现为两种形式：一是大语言模型生成违背客观事实与外部世界运行规律的内容，二是大模型对提示词、上下文及自身能力边界做出越界性误判^[17]。这意味着 AI 幻觉的产生并非信息生产和流通过程中某一局部环节的问题，其特殊性在于更易引发关于真实性的系统性风险。尤其在突发事件、公共卫生和战争冲突等场景中，一旦公众舆论的首轮认知被错误信息占领，那么即便后续核查给出了正确结论，也常常会丧失最佳舆论引导和观点矫正的时机，公众也会因为频繁受到虚假信息的冲击而加剧对智能技术失控的集体焦虑感^[18]。

AI 幻觉风险下的信息真实性危机除了会引

发公众对智能技术的焦虑外，还在具体司法纠纷中暴露出当前事实核查体系的深层制度性困境。例如，2026 年我国最高法、最高检工作报告中就曾提及了一起因 AI 幻觉引发的侵权纠纷案。在该案件中，梁某在使用某生成式 AI 查询高校报考信息时，AI 生成不实内容并自行作出 10 万元错误赔偿承诺，梁某遂起诉 AI 所属科技公司索赔。杭州互联网法院审理认定为，AI 无民事主体资格，科技公司已尽提示义务无过错故无需担责^[19]。这一事件进一步引发了如何在 AI 时代锁定事实核查主体的问题，在上述案例中，如果要求科技公司对 AI 生成的每条内容都进行逐一核查似并不现实，但是如若放任未经核查的信息在互联网平台上广泛传播，那么将制造更加混沌且更具危险性的信息生态环境。

从认识论的视角来看，事实核查需要实现从“程序客观”到“实质客观”的转变^[20]，主流媒体在系统性变革进程中更需肩负起核查的重任——不仅要宣告核查结论（事实层面的真伪），还要通过真相的供给介入塑造社会现实的过程，并在系统性的信息治理中发挥能动价值作用。具体而言，主流媒体应率先建立对 GAI 技术的正确认识。主流媒体对待 AI 最可取的方式，不是否认技术潜能，也不是轻率交出判断主权，而是在承认新闻专业实践与通用大模型之间存在张力关系的前提下，主动重组新闻生产链条，守住事实裁定与公共解释的最后关口^[21]。如此，事实核查才可能真正成为主流媒体在智能时代重建认知权力的重要支点。

对于我国主流媒体而言，越是在真假交错、情绪感染的议题中，越需要把核查能力前置到报道策划、内容生产和传播分发环节，而不是等到谣言成势之后再匆忙应对。从公众的认知角度来看，在 AI 幻觉引发的各类信息乱象案例中，普通民众既可能因“有图有视频”而误信错误内容，也可能陷入普遍怀疑而质疑一切事实核查的努力，不论是前者还是后者都会极大削弱新闻业通过发挥连接人与人、人与社会而维系公共理性讨论的价值^[22]。

（二）新闻信任危机下主流媒体制度优势



的价值延伸

AI幻觉对于新闻真实性的威胁，不仅发生于本源信息失实层面，还会动摇新闻业在再现真实与形塑信任真实层面的根基。特别是在各类AIGC叠加出现的复杂信息生态中，新闻记者和受众对新闻是否可信的评判标准也存在较大分歧^[23]。同样回归到新闻认识论的视角来看，新闻信任危机的实质是，新闻业在事实锚定、议程设置、接口控制与信任生产四个维度上的功能失调与权力位移^[24]。因此，主流媒体事实核查的重要专业性价值在于不只是一是要帮助判定某一条信息的内容真假，还要在真实性被污染的信息环境中，重建公众对于信任的共识基础。

在推进主流媒体系统性变革的政策导向下，各级主流媒体通过入驻头部社交平台、开设垂类官方账号、搭建传播矩阵等一系列转型举措，主动抢夺互联网流量的主要接口。此类实践表明，主流媒体将大众传播时代依托制度性权威与长期专业实践积淀而获得的信任与公信力资本，转化为适配算法分发规则的网络传播影响力，以此巩固和延伸自身议程设置能力和算法把关的能力。此外，既往事实核查的从业者可能因频繁暴露在包括仇恨言论、有害信息等在内的信息环境中从而引发职业倦怠，加之薪酬回报有限，这都加剧了该职业的不稳定性。反观主流媒体的组织架构则相对稳定，由主流媒体发起的事实核查实践也将体现更强的有组织色彩，这有利于事实核查的从业者更好地投入本职工作中。

对于主流媒体事实核查的价值重估，实际上也是对新闻生产逻辑再理顺和再组织。传统新闻生产实践中，对于真实性原则的坚守往往内嵌于新闻采访、写作、编辑与审校的流程之中，事实核查很少需要在前台凸显。然而，在AI幻觉引发信息迷雾的境况下，对于真实性的坚守越来越需要被明确展示和主动说明，从而最终形成真实性维护的专业机制。主流媒体肩负着自媒体和平台公司所难以秉持的专业职责，通过事实核查的开展能够更好地将主流媒体的影响力转化为公众可感知、可理解的新闻品质，从而也更有利于进一步修复新闻信任危机。

（三）以牵头者角色引领信息治理的规范价值

新闻认识论强调新闻活动在公共生活中应体现主体性作用，对于事实核查来说，其核心主体长期被默认为是唯一具有合法行动力的人类主体。然而，在数字新闻业的语境之下，事实核查的主体逐步表现为由人类、技术、机器等多元行动者在动态互动关系中共同构成的复合性新主体^[25]。事实核查的实践开展当然离不开人类核查者的专业判断，但同时事实核查中还关联着人机混合的验证系统及跨组织网络等形态，因此传统分散化的事实核查工作具有不稳定性。一项针对加拿大新闻编辑部的研究也指出，当GAI技术进入新闻生产环节之后，现有关于事实核查伦理规范的更新速度明显滞后于技术变迁的速度，新闻从业者往往仍依赖一种头脑中熟悉的规则或经验性直觉来作出抉择^[26]。

当前，全球范围内的事实核查实践已经体现出明确的合作特点，若要充分发挥跨平台事实核查的信息治理效能，那么，主流媒体则需要更明确地发挥牵头和组织作用，并尽早探索确立能够被行业共享的核查规范和操作指南。在事实核查的协作网络中，互联网平台多与第三方人工智能科技公司达成合作，借助其技术优势与专业核查能力，应对实时更新的海量信息带来的辟谣挑战。对于主流媒体来说，其事实核查的效能不在于对单一新闻事件的辟谣效果，而在于是否能够通过发挥引领作用，及早发现高风险议题、识别潜在的信息失真痕迹，并把事实核查的结果嵌入更广泛的舆论引导与社会沟通过程之中。例如，主流媒体通过与舆情监管部门联动可以将事实核查与议题监测、舆情风险预警相结合，那么，事实核查将成为阻断不实信息演化为舆情危机的前置性预防环节。这亦表明，主流媒体引领事实核查的价值已经超出传统新闻机构的纠错价值，与国家信息治理、社会风险防控以及公共安全维护紧密相连。

三、基于事实核查的权威再造与可信生态

主流媒体权威再造这一命题的提出，指向

“新闻”概念正逐渐从一种标准信息文本（产品）演变为动态性的“信息关系”^[27]。数字新闻认识论的理论主张也明确将新闻工作的边界模糊化和泛化现象作为重新认识新闻权威以及新闻业价值的前置因素，因此，主流媒体与事实核查的关系界定及主流媒体的权威地位已不再是经典专业主义话语所设定的概念范畴，由此更加呼唤一类新的阐释框架^[28]。新闻权威的塑造关系到新闻职业共同体如何理解自身，新闻权威并非新闻机构天然拥有的声望，权威的实现基础在于新闻机构及其生产的新闻产品能够拥有被社会公众承认为阅读或观看的资格^[29]。当 GAI 技术介入新闻产品生产和流通的过程，原本由新闻专业共同体持有的编辑权威与知识生产权力，会部分地被不透明的技术系统所消解，新闻权威的根基也需要持续加固。

不同于技术平台和第三方机构，主流媒体通过牵头引领应对 AI 幻觉的事实核查实践，可通过主体可信性、信息可信性和系统可信性重塑主流媒体权威、重彰专业价值。其中，主体可信性指向主流媒体作为事实核查机构本身的可信度；信息可信性指向主流媒体核查结论的可信度，包括其证据的充分性、论证的严密性和结论的准确性；系统可信性指向主流媒体在整个信息生态系统中的可信角色，包括组织引领内容生产、传播、核查等环节的规范性和可靠性。以上三个层面的可信性相互促进、相互强化，主体可信性的提升会增强信息可信性，信息可信性的提升又会进一步巩固主体可信性，两者共同推动系统可信性的提升和整体信息生态的可信度。

（一）主体可信性：以可问责的专业自治重建信任基础

对于主流媒体而言，权威的维护并不是要贸然排斥 AI 技术工具，而是需要建立规范的 AI 使用规则、适用于新闻生产的人工监督模式以及清晰的责任归属依据。这意味着涉及 AI 参与新闻采写、整理、翻译、可视化等环节，都应尽可能地建立更为细致的 AI 使用规范，通过主动披露 AI 的使用，让公众能够更好辨识哪些内容是由大模型辅助生成的，哪些事实性判断和价值性

判断是由编辑所主张的，以及如何追责信息失实的问题。由此，主流媒体通过建立起可问责的专业自治体系，从而让主体可信性重新确立。

事实上，主体可信性之所以能够被重塑，是因为主流媒体在事实核查中表现出的专业能力，这些能力包括但不限于现场信息采集的能力、信源交叉核验的能力、符合情境的解释能力和公道的道德评估能力等。在常规新闻报道中，主流媒体的新闻叙事通过情感的能动确立了报道主体的可信性，即新闻工作者在具体新闻采访过程中能够对被采访对象的生活经历、人际关系和现实处境有较强的情感感知，而非简单地与被采访对象共情或将事件细节煽情化处理以拓展传播效果。

不同于常规新闻报道，事实核查则更加强调核查主体的可信和权威。当前中西方主流媒体的事实核查实践，也表现出主体可信性之于新闻权威重塑的重要意义。例如，新华网于 2025 年 12 月上线了“AI 探真”智能体，该智能体以全链条谣言治理为核心目标，整合新华社、中国互联网联合辟谣平台等权威资源，这也是主流媒体运用 AI 技术对抗 AI 生成式谣言、破解事实核查技术困境的新一类型实践^[30]。作为英国主流媒体的 BBC 也推出了名为“Verify 事实核查”的服务平台，其开展事实核查的目标不只是完成对某条假消息的驳斥，更在于持续训练公众理解证据、识别来源并做出判断的能力，最终将事实核查打造为一类媒体品牌^[31]。中西方主流媒体均在积极探索事实核查的创新手段，西方媒体的事实核查往往局限于媒体自身的业务范围，难以形成治理合力，中国主流媒体依托自身的公信力和组织协调能力，更有利于牵头构建统一、高效的智能事实核查平台。

（二）信息可信性：基于可信语料库构建可验证的事实

宣告真相是事实核查的重要任务，传统事实核查以过程的透明性原则为标榜。尽管透明性原则并不能保证事实核查实践完全免责，但如果连透明性都无法保证，那么事实核查便难以建立持久的专业说服力^[32]。对于主流媒体的权威重塑来说，透明性原则依然在指导事实核查实践中



体现着必要性，公众对权威的认识也取决于新闻媒体是否愿意把求证过程展示出来，换言之，用户越来越看重的并不是简单的结论正确，而是结论是否有迹可循、是否能够复核。

基于信息可信性，主流媒体在实践中可通过牵头开发组建数据基础设施、构建可信语料库等方式将事实核查由单次的求证提升为数据化和结构化的真相宣告行动。主流媒体的从业者需要使用时新的、不同类型的数字系统以满足事实核查的需求，开发有效的、可解释的、用于假新闻检测的机器学习算法^[33]，包括原图检索、视频拆帧、地理定位、数据比对、版本溯源等方法，都应成为主流媒体事实核查的常规工具，以提升智能新闻产品的可验证性。以“澎湃明查”的事实核查为例，一方面是媒体借助智能工具对海量信息进行自动筛查，另一方面也明确保留了人工事实核查权。研究表明，团队信源透明、过程透明与结论透明并非一种附加性质的修辞，其与用户互动和事实核查的传播效果存在显著正相关性^[34]。“澎湃明查”的可信语料库多来自国务院官网、国家统计局以及权威学术数据库等多源官方数据，这有助于提升对 AI 虚假信息的核查效率与精准度。此外，一些 AI 大模型也正在测试包括“引证”等在内的功能，所谓“引证”主要是指，当用户和 AI 对话中如果涉及新闻时事或政策制度等内容，大模型会内置对所回答内容事实的二次复核机制，如果用户点击相关按钮，则大模型会再次对外部信源进行核查，从而提升回答的可信度。

（三）系统可信性：面向协作网络的智能信息生态治理

前文提及 AI 幻觉可能引发递归式的信息失真循环，例如，当大模型生成中的失实内容经由搜索引擎、社交平台和群聊转述，便会以新的形式重新进入大语言模型检索和调用回路，在后续问答中该失实内容会持续被复制、改写和放大。因而，智能信息环境中不实信息的传播失控，不仅表现为不实信息数量层面的增多，其更严峻的风险还在于带来持续性的信息真实性污染。如此一来，公众面对的便不再是一个固定的谣言

对象，而是不断迁移甚至相互背书的失真版本。从新闻传播专业角度来看，对于递归式失真循环的认知和评价不应停留于 AI 技术是善是恶的道德判断，当务之急是拷问我们的公共信息传播体系是否建立了足以约束、校正并向公众说明原理的稳定作用机制，这也就解释了为什么事实核查必须从传统的单点纠错转向全链路的信息治理。

主流媒体的事实核查需要在平台、技术公司、研究机构、政府部门与公众之间，以协作的方式构建真实性保障机制，共同构建起 AI 幻觉风险应急响应通道，并更好地探索虚假信息治理的闭环模式。在现实治理实践中，近年来《生成式人工智能服务管理暂行办法》《互联网信息服务深度合成管理规定》等政策文件的出台已对 AI 生成内容的使用提供了操作准则，由国家互联网信息办公室、工业和信息化部、公安部和国家广播电视总局联合印发的《人工智能生成合成内容标识办法》也明确强调内容生产者应对包含 AI 生成的内容做出标注等^[35]。但是，现有规章制度仍只能在信息生成和传播的前序阶段起到防控作用，主流媒体机构在参与法律法规制定方面的贡献度还不足。从主流媒体的事实核查流程来看，在前端应通过流程规范、来源标注和 AI 使用披露等方式，减少失实信息进入新闻生产领域的可能；在中端则可通过技术监测和编辑复核识别出高风险内容；在后端通过公开核查、解释说明与可信语料库的沉淀维系长期的事实可追溯性，必要时还应联动平台方采取限流、下架、标记、阻断等技术化处置手段。

四、结论与讨论

AI 幻觉影响下的事实核查亟待突破传统基于事实层面的内容核查和纠偏模式，当事实核查从内容把关机制转变为复杂信息生态的治理模式，其主体已不再是单一媒体机构，而是拓展至包括主流媒体、专业机构、自媒体以及普通网民在内的多元行动者，其目标也更加体现出对公共信息秩序的维护。智能传播信息生态中，AI 幻觉风险对于新闻业的危害性远大于一般性的生

成式文本应用场景，主流媒体肩负着及时识别合成内容的潜在风险和引领专业事实核查实践的重任，新闻真实性原则仍然是维护新闻业权威和记者职业信仰与身份认同的根基。

不可否认的是，一方面，AI幻觉引发的风险使得事实核查任务更为艰巨，但另一方面，以事实核查为抓手，主流媒体可以更容易地找到证明自身为何比平台热帖、自媒体推断和大模型生成的内容更值得信赖的实践契机。为化解既往事实核查实践中主观意图可能引发的核查结果偏差，同时更好地弥合事实核查理念中客观性和主观性的罅隙，人工智能时代的事实核查应建立更为严谨和统一的操作标准，主流媒体事实核查也应发挥搭建链路更透明、操作更标准、核查更系统的专业内控机制优势，从而最大限度压缩核查流程中的从业者个人的主观裁量空间，进而从制度设计层面为事实核查结论的信度、效度与程序公正性提供保障。

从数字新闻认识论的理论视野来看，主流媒体事实核查的价值内涵已超越传统意义上的辟谣和求证，主流媒体的主体性引领作用，以及其在信息生成溯源、传播链路监测和公众认知引导等关键环节的价值作用需要重新评估。主流媒体事实核查因此也应被理解为在AI幻觉引发潜在风险下重建社会认知秩序的新闻业专业实践，主流媒体是将事实核查再组织为专业规划的关键主体，事实核查则是抵抗信息失序与认知权力外移的重要的制度选择。AI时代的事实核查，不只是新闻业对技术冲击的被动回应，更关乎一个社会能否在高度自动化、平台化和情绪化的传播环境中坚守事实基础。因此，本文把事实核查视为主流媒体在重建制度合法性和重塑专业权威的关键环节，而不仅仅将事实核查作为一类栏目或实践样态。这种对于主流媒体权威的重塑并不是对旧式媒体中心主义的简单回归，也不是对技术扩张的消极防守，而是一种建立在可问责主体、可验证信息与可协同系统之上的新型制度主张。具体而言，主流媒体通过事实核查重建权威的可能路径可以分为三个层面：一是主体可信性，通过发挥

主体作用在事实核查中巩固专业权威；二是信息可信性，通过构建可信语料库构建可验证的事实；三是系统可信性，以主流媒体引领、基于行业协同为网络空间虚假信息治理提供实践新范式。

未来研究还可以进一步通过提炼实践中的经验做法，总结出有效防范AI幻觉信息安全风险和推动人工智能规范发展的中国主流媒体方案。当然，研究还应考虑主流媒体的层级差异、资源占有等客观因素对事实核查效果的影响。例如，中央级媒体、头部平台与大型技术企业在数据、算力、人才与工具接入上拥有显著资源，而地方媒体、行业媒体和基层传播机构在核查人才、技术设备、法务支持和跨平台协作上往往相对薄弱。若不考虑这种能力落差，仅从规范层面要求所有媒体同步升级事实核查体系，容易把必要的制度期待变成难以兑现的责任负担。此外，需要承认的是，用户自身对于不确定信息的事实核查习惯尚未充分形成，公众对核查新闻的理解常常仍较多停留在辟谣的层面。同时也应注意，主流媒体事实核查并不意味着将所有社会意见的表达都纳入规训式的核查制度当中，主流媒体在强化事实核查时也需要注意维护讨论空间的多样性，从而更好地健全网络生态治理的长效机制。

（责任编辑：徐秋染）

参考文献

- [1] van Dalen A. Revisiting the algorithms behind the headlines. How journalists respond to professional competition of generative AI[J]. *Journalism Practice*, 2026, 20(3): 897-914.
- [2] 张涛甫. 人工智能推动舆论生态转型及其治理进路[J]. 学术月刊, 2024(2): 149-157.
- [3] Boesman J, Meijer I C. Nothing but the facts? [J]. *Journalism Practice*, 2018, 12(8): 997-1007.
- [4] Lauer L, Graves L. How to grow a transnational field: A network analysis of the global fact-checking movement[J]. *New Media and Society*, 2025, 27(6): 3505-3523.
- [5] Kim Y, Lee J. Balancing artificial intelligence and human expertise: Ideal fact-checking strategies for hard and soft news[J]. *Journalism & Mass Communication Quarterly*, 2026, 103(1): 134-160.
- [6] Graves L. *Deciding what's true: The rise of political fact-checking in American journalism*[M]. 2016, New York:



Columbia University Press.

[7] Suomalainen K, Nykänen N, Seeck H, et al. Fact-checking in journalism: An epistemological framework[J]. *Journalism Studies*, 2025(26): 1129-1149.

[8] Tulin M, et al. Beyond belief correction: Effects of the truth sandwich on perceptions of fact-checkers and verification intentions[J]. *Journalism Practice*, 2025, 19(11): 2576-2595.

[9] Shin D, et al. Fact-checking as epistemic infrastructure: From journalistic gatekeeping to platform governance[J]. *Emerging Media*, 2025(3): 183-202.

[10] 新华社国家重点实验室发布“新华智信”全链解决方案[EB/OL]. (2025-5-11)[2026-05-19]. <https://www.xinhuanet.com/tech/20260511/1915e1e925b04755967ba94977b3839d/c.html>.

[11] 杜骏飞. 奇幻社会的来临——DeepSeek 幻觉与后真相递归[J]. 探索与争鸣, 2025(3): 11-14.

[12] Yang G, Zhao L, Liang Y, et al. Shadows of deception: Unveiling the dominance of source versus recipient in the perceived credibility of misinformation[J]. *Journalism Practice*, 2025. https://www.tandfonline.com/doi/epdf/10.1080/17512786.2025.2464192?src=getftr&utm_source=clarivate&getftr_integrator=clarivate. DOI:10.1080/17512786.2025.2464192.

[13] 林爱珺, 姜泽宇. AIGC 时代多模态虚假信息的识别与人机协同核查[J]. 新媒体与社会, 2025(2): 88-104.

[14] Waldo J, Boussard S. GPTs and hallucination: Why do large language models hallucinate? [J]. *Queue*, 2024, 22(4): 19-33.

[15] 常江. 什么是数字新闻: 认识论的视角[J]. 山西大学学报(哲学社会科学版), 2022(3): 68-74.

[16] 常江, 王鸿坤. 集成多维社会: 主流媒体系统性变革的议题清单及价值锚定[J]. 传媒观察, 2025(8): 13-21.

[17] Huang L, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. *ACM Transactions on Information Systems*, 2025(43): 1-55.

[18] Dodds T, Zamith R, Lewis S C. The AI turn in journalism: Disruption, adaptation, and democratic futures[J]. *Journalism*, 2026, 27(3): 530-544.

[19] 人民日报. 守护合法权益 传递法治温度[N]. 2026-3-18(11).

[20] Cazzamatta R. Redefining objectivity: Exploring types of evidence by fact-checkers in four European countries[J]. *European Journal of Communication*, 2025(40): 144-164.

[21] 王辰瑶, 宋沁语. “大模型焦虑”何解: 新闻业与大模型的“对抗-合作”式关系创新[J]. 新闻与写作, 2025(12): 18-28.

[22] 黄文森. 从泛连接到部落化: 共识在新闻流通中的形成与演变[J]. 中国出版, 2024(24): 16-22.

[23] Enrique N, et al. The threat of misinformation on journalism's epistemology: Exploring the gap between journalist's and audience's expectations when facing fake content[J]. *Digital Journalism*, 2025, 13(3): 478-499.

[24] 王金鹏, 陈昌凤. 智能时代新闻的认知权力位移——基于主流媒体系统性变革的分析性框架[J]. 苏州大学学报(哲学社会科学版), 2025, 46(6): 172-183.

[25] 常江, 王茜. 机器人记者会思考吗——智能体新闻的“拟人”与“非人”?[J]. 现代视听, 2025(3): 5-9.

[26] Misri A, Blanchett N, Lindgren A. “There's a rule book in my head”: Journalism ethics meet A.I. in the newsroom[J]. *Digital Journalism*, 2025(4): 1-19.

[27] 常江. 媒介基础设施化、新闻弥散与倦怠社会的纾解[J]. 南京社会科学, 2025(1): 104-113.

[28] 徐笛. 数字时代, 谁是记者: 一种分层理解的框架[J]. 新闻界, 2021(6): 13-20.

[29] Amigo L, Porlezza C. “Journalism will always need journalists.” The perceived impact of AI on journalism authority in Switzerland[J]. *Journalism Practice*, 2025(19): 2266-2284.

[30] 新华网“AI探真”智能体正式上线[EB/OL]. (2025-12-03)[2026-03-29]. https://www.gov.cn/zhengce/zhengceku/202503/content_7014286.htm.

[31] Cushion S. The truth about fact-checking[J]. *British Journalism Review*, 2025(36): 27-33.

[32] Dierickx L, Lindén C. Transparency and fact-checking in open societies[J]. *Journalism*, 2024(27): 202-220.

[33] Lebernegg N, Eberl J, Tolochko P, Boomgaarden H. Do you speak disinformation? Computational detection of deceptive news-like content using linguistic and stylistic features[J]. *Digital Journalism*, 2024(13): 1373-1398.

[34] 潘文静, 陈嘉翌, 季江睿. 事实核查新闻的中国经验: 澎湃明查的透明度实践与传播效果[J]. 新闻记者, 2025(11): 65-79.

[35] 关于印发《人工智能生成合成内容标识办法》的通知[EB/OL]. (2025-03-07)[2026-03-29]. https://www.gov.cn/zhengce/zhengceku/202503/content_7014286.htm.

(下转第 61 页)

平台化生存： 生活世界的“可见”与“遮蔽”¹⁾

张秀丽 贺子阳 牛超杰

摘要 平台化生存已成为当代社会个体的基本存在样态。本文尝试将平台概念化为一种“启蒙性基础设施”——它并非中立的技术系统，而是一种内嵌着技术赋权与系统控制双重潜能的社会—技术装置。平台通过数据化、算法化与自动化，系统性地将生活世界纳入“算法视域”，塑造着“可见”的秩序，同时也进行着结构性的“遮蔽”。这种双重效应揭示了平台启蒙的深刻悖论：它在赋予个体前所未有的“看见”权力的同时，却可能以算法前置、资本驱动和景观替代的方式，悬置康德意义上的“敢于求知”的批判性反思。本文进一步提出“可见中的遮蔽”与“遮蔽中的可见”这一辩证分析框架，前者揭示平台如何以极致的“可见”掩盖社会现实的复杂性、算法的运作逻辑及个体生存的本真状态；后者则展现被技术中介所“遮蔽”的领域如何创造出新的感知维度与连接可能，为意义的重建与韧性实践提供空间。通过剖析这一辩证关系，本文揭示了平台化生存既形成结构性制约又催生抵抗性实践的复杂样态，为理解数字时代生活世界的转型提供了新的理论视角。

关键词 平台化生存；生活世界；启蒙性基础设施；可见；遮蔽

中图分类号 G206 **文献标识码** A

作者 张秀丽，西安外国语大学新闻与传播学院教授，陕西西安 071000；贺子阳，中央民族大学新闻与传播学院博士研究生，北京 100081；牛超杰，华南理工大学新闻与传播学院博士研究生，广东广州 510000

随着天猫提出“发现品质生活”、支付宝强调“生活好，支付宝”的品牌定位不难看出，数字商业平台深度渗透日常场景，已然成为现代社会生活中不可或缺的重要载体；抖音倡导“记录美好生活”，新浪微博致力于“随时随地发现新鲜事”，二者不断推动个体情感与公共议题高度可视化的趋势，并拓展了信息流动与社会连接的维度；微信将自身定位为“一种生活方式”，小红书致力打造“你的生活兴趣社区”，则进一步表明平台的聚合功能已成为现实社交的关键延伸。由此可见，平台与生活世界的嵌合，不仅重构了日常生活的实践结构，更奠定了其作为新型社会基础设施的地位。然而，平台在促成原本隐匿现象走向可见的同时，亦潜在地加剧了对生活世界的遮蔽。

1) 基金项目：国家社会科学基金重大项目“铸牢中华民族共同体意识的传播策略研究”（22&ZD313）；陕西省哲学社会科学专项一般项目“数智赋能陕西非遗文化出海的协同机制、作用机理与优化策略研究”（2026YB0159）。

基于此,本研究认为,作为当下社会基础设施的平台,已经超越其固有的媒介边界,呈现出启蒙意涵。有鉴于此,本文旨在批判性地追问,作为具有启蒙意义的平台构建了何种“可见性”?又遮蔽了何种现实?更进一步而言,本研究试图超越可见与遮蔽的二元对立,理解其中的过程机制。因此,本研究聚焦于平台化生存现象,基于生活世界的理论视角,致力于剖析“可见”与“遮蔽”之间的内在关联,进而揭示其中所蕴含的认知与存在张力。

一、平台:考察生活世界“可见”与“遮蔽”的新视角

“平台”是以数字技术为支撑、具备连接用户并提供交互服务与资源功能的数字化中介组织形态。进入互联网时代,平台因其功能完备、应用广泛,能够深度嵌入人类社会的交往结构,通过重塑信息流动的方式,变革人与人、人与物、人与环境之间的关系,推动主体解放与时空重构,进而深刻改变人们的日常生活及社会形态,由此成为生活世界中的启蒙性基础设施。该概念源自启蒙时代思想中的“基础设施”概念,意指世界始终处于运动与变化之中,物品、思想与人的自由流通将促进社会进步。马特拉(Armand Mattelart)则主张超越静态视角,认为基础设施是一个有机且动态的过程^[1]。由上所述,本文认为所谓的“启蒙性基础设施”是指平台作为一种社会-技术装置,既赋予人类改造世界、获取知识的“普罗米修斯式”力量,更以其“启蒙”意涵促使主体运用理性对技术系统本身进行审视与批判。具体到本文中,这一概念意指在洞悉平台如何建构“可见”秩序的同时,辨识其如何形成结构性的“遮蔽”,从而在技术环境中实现康德意义上“敢于求知”的主体自律。因此,作为启蒙性基础设施的平台,既赋予了生活世界被看见的机会,也使其面临被遮蔽的可能。

(一)生成启蒙性基础设施:平台化的技术逻辑与社会嵌入

平台之所以能成为当代生活世界的“启蒙

性基础设施”,是因为其特有的技术逻辑完成深度的社会嵌入。一方面,它实现了启蒙传统中对“去蔽”的追求,提供技术理性使隐匿于现实世界的事物和社会关系得以可知、可见;另一方面,这种使之可见的启蒙力量也可能转化为新的支配和遮蔽。这一过程的核心在于,平台化并非外置于社会的技术应用,而是一种重塑社会关系与认知模式的内置力量。正如蒂尔森(David Tilson)所言,平台本质上是“包含技术要素以及相关组织流程和标准的社会-技术集合”^[2]。这意味着其技术逻辑与社会嵌入从一开始就相互交织、互为前提,技术架构为社会化应用提供可能,而广泛的社会应用又反过来强化并固化其技术逻辑,最终使平台从一种“可选的工具”进化为“不可或缺的基础设施”。

平台的技术逻辑根植于数据化、算法化与自动化。它将人类丰富的社会活动——从交往、消费到情感表达,转化为可计量、可计算、可预测的数据流。平台依靠数据、算法与人工智能,得以建构成影响人类社会交往和发展的数字基础设施^[3]。这一转化过程本身就是“启蒙”的体现,它使得以往难以捉摸、隐于幕后的生活世界细节变得清晰可见。平台通过强大的搜索与匹配机制,让生活世界中的存在物实现功能链接,降低了信息差,促使一切变得透明。然而,这种“可见性”并非全景敞视式的平等观照,而是技术依据特定商业或政治逻辑进行选择性的遮蔽与凸显的结果。哪些关系被连接、哪些内容被推荐、哪些行为被鼓励,均取决于平台的技术架构与规则设计。因此,平台的“启蒙”自带辩证色彩:它在照亮一部分世界的同时,必然将另一部分推入更深的阴影。它既是我们认识世界的新窗口,也是塑造我们认知的新框架。

最终,这种技术逻辑通过广泛而深入的社会嵌入,完成了其基础设施化的进程。平台已不再仅是早期的“物理性、对象性的建筑”^[4],或是单纯的双边市场“元组织”^[5],它已如毛细血管般嵌入社会肌体的每一个角落,成为组织社会生活的“管治装置”。当个体认知、日常生活与社会运行都已深度依赖平台架构时,平

台便从一种技术系统升维为一种认识论基础设施。它规定了何者可见、如何可见以及谁有权力使之可见,从而深刻地塑造着我们关于“现实”本身的体验与观念。

(二)生活世界:理解启蒙性基础设施的理论富矿

生活世界是个体赖以生存的基本背景,是日常生活经验得以展开的领域。个体对周遭环境的理解和体验,无不根植于特定的生活世界之中^[6]。“生活世界”这一概念最初源于现象学大师胡塞尔(Edmund Husserl)在晚期著作《欧洲科学的危机和先验现象学》的引入,他认为生活世界是使人们在“自然态度”内直接面对生活世界的特质,现象学研究应从互为主体的社会关系中去理解生活世界^[7]。而后,舒茨(Alfred Schütz)容纳吸收胡塞尔对生活世界的探讨,将生活世界、主体间性、手头库存知识等概念引入社会学中,并结合奥地利经济学的“去历史化”主张,创立了现象学社会学。在舒茨看来,生活世界是人们生存在其中并进行日常活动的社会情景。其自然态度和主体间性,以及两者之间的紧密关系是生活世界的结构特征。换言之,生活世界是个体感知、理解进而采取行动的基本框架,个体与环境的互动都围绕这一框架展开。并且,舒茨还特别强调生活世界的主观性构成,个体对生活世界的理解根植于主观经验,但这种主观性理解又以普遍化、类型化的方式呈现,使得不同个体对生活世界的把握具有同质性,由此保证了社会互动的顺畅和秩序的存续^[8]。哈贝马斯(Jürgen Habermas)也运用了这一概念,他强调生活世界的互为主体性特征,个体正是在与他人的互动中理解并构建生活世界。在哈贝马斯看来,交往行为是生活世界得以形成和延续的基础^[9]。人们在日常言说、交往中达成共识,形成对生活世界的共同理解,由此确立社会整合的基础。因此,生活世界是交往理性展开的空间,是主体间性构建的场域。

统合上述观点,可知生活世界是根植于个体或群体的主观经验,是个体感知、理解外部世界的参照框架,通过对生活世界的主观性理

解,参与其中的个体得以在日常言说、交往中形成对生活世界的共识,并且为社会互动奠定基础。同时,生活世界是包含文化、社会、人格等多个向度,具有再生产、社会整合、个体社会化等多重功能作用。在平台化转型的数字技术背景下,无论是日常支付实践、网络社交互动抑或是同龄群体的经验分享行为,数字化网络已深度渗透至人类生活的各个领域,并逐渐成为组织生产、交换与消费等社会活动的基础设施。这一现象深刻印证了胡塞尔关于“生活世界作为人类多元社会活动发生场域”的理论阐释,也印证了直观上存在着并为我们存在的那个世界,是我们一切实践与理论活动的直接地平线和根基所在^[10]。人际互动及社会交互日益依托平台载体展开,实体空间与虚拟空间的边界趋于消弭,“可见”与“遮蔽”交织共生。在平台架构支撑下,“媒介的世界化”进程正加速演变为现实图景。平台机制如何建构可见体系,并掩盖对遮蔽的感知,已成为理解当代生活世界的重要视角。

(三)可见与遮蔽:一个分析生活世界的关键视角

将平台视为生活世界的启蒙性基础设施,意味着需要重新审视它如何系统性地塑造主体对“现实”的感知与理解。在这一视野下,可见与遮蔽不再仅是传播效果的表层特征,而是构成了平台化时代生活世界转型的核心机制与内在矛盾。可见性(Visibility)从字面上看,是指人的视线范围所及程度与能见度,并受到人的视觉、知觉等感知能力的限制^[11]。也有学者认为,所谓可见性是指能否被他人看见、能否获得他人的注意力。当获得的注意力达到一定规模时,即产生了可见性^[12]。伴随着大众媒介技术的发展与普及,也有学者认为社交媒体平台为生活世界可见提供了重要场所^[13],认为媒介技术能实现跨时空的可见。不仅如此,在丹尼尔·戴扬(Daniel Dayan)提出虚拟网络新媒体环境中可见性的新概念^[14]之前,福柯(Michel Foucault)的“全景敞视监狱”的权力模型以及约翰·汤普森(John B. Thompson)将大众媒介

带来的这种新的可见性称为“媒介化可见性”^[15]，均为虚拟网络新媒体环境中的可见性提供了理论基础。

从横向维度来看，国内学者已然发现“可见性”分析生活世界的多重可能，国内研究常与“可供性”理论结合进行分析和考察^[16]。并从日常生活中的性别可见^[17]、社交可见^[18]、家务可见^[19]、城市可见^[20]乃至日常可见^[21]延伸至与政治相关的劳动可见^[22]、权力可见^[23]与文化可见^[24]等领域，展现出可见性理论在解释社会结构与权力关系中的强大诠释力。值得关注的是，当前已有研究并非仅仅关注可见的单一视角，可见所蕴含着结构性的遮蔽机制亦受学者关注，如杨飞从遮蔽与解蔽的角度，结合可见性理论，分析论算法对个人信息安全的影响及应对^[25]。因此，当前对可见的研究亦延伸至“不可见”（遮蔽）的视野分析劳工问题^[26]与乡村形象的选择性建构问题^[27]。虽然当前学者已然关照到可见与遮蔽之间的辩证关系，但是已有研究多是为了强调可见而凸显出遮蔽，因此对不可见的系统性遮蔽机制及其背后权力运作逻辑的深入探讨仍显不足，尤其在数字平台算法筛选、流量分配不均导致大众群体声音持续被压制的背景下，这一盲区愈发凸显其理论与现实意义。

综上所述，既往研究多聚焦于对单向的可见或遮蔽进行分析，然而，此类单向度探究或存割裂可见与遮蔽动态互动过程之嫌。本研究旨在突破单一研究向度，通过剖析平台作为启蒙性基础设施所蕴含的核心悖论与权力运作机制，在拓展人类经验与认知边界之可能性的同时，揭示其潜藏的系统性遮蔽与控制。本文对两种机制进行了区分，其一“可见中的遮蔽”意味着平台通过数字技术掩盖了社会现实中的复杂性，其特征是以“可视呈现”代替理解和反思；其二“遮蔽中的可见”是指被平台所中介化的领域呈现出了新的感知维度和链接关系，依靠数据和算法创造出了新的可见和意义。二者相互交织，形成了平台化过程中的一体两面，正是因为平台中的可见才产生新的遮蔽，而正是在遮蔽的缝隙中创造出新的可见形式。唯有

借助此辩证视角，方能超越技术决定论与盲目乐观主义，对生活世界之真实转型保持批判自觉。在此需特别说明的是，平台化与生活世界并非简化的二元对立关系，二者始终相互建构、相互渗透，是一个动态的有机过程。

二、平台建构新关系：生活世界的日益“可见”

何塞·范·迪克(José van Dijck)等人指出，平台社会的三大运行机制——数据化、商业化与选择性共同构成了一种全新的社会组织逻辑^[28]。他们认为平台运行的核心即为数据，通过对万物的数据化收集和处理，可有效配置社会关系，跟踪和预测用户个体的诉求、交易、表现与活动，进而推动商业化逻辑的有效展开。而“选择性”则通过平台的界面与算法来触发或过滤用户的感知，在筛选可观内容的同时，确保某些特定内容、服务和人员的优先可见性与可用性^[29]。该机制的核心特征在于将生活世界中丰富、模糊且高度情境化的“体验”抽象化为可计算、可分析、可预测的数据。这一过程标志着技术理性对生活世界的系统性介入。在此过程中，生活世界的主体无法直接感知抽象化运作本身，其所能认知的仅为可感知的具象符号。

（一）可见展演：从切身经验到数据呈现

在舒茨看来，生活世界进一步被阐释为一种由“想当然的常识性构念”所支撑的社会实在。个体在其中通过“自然态度”与他人进行互动，共享着一套默认的意义图式和阐释框架，从而构建起一个稳定且可预期的交往世界^[30]。然而，以抖音等短视频平台为例，这种基于主体间性的意义生成过程正被一种算法化的逻辑所取代。一方面，个体在平台中的前置化展演（如发布视频、直播互动）与后置化生存轨迹（如浏览、点赞、消费）被平台以数据形式进行巨细靡遗的记录和计算。个体的社交网络、认知倾向、情感波动乃至无意识的习惯，都被平台清晰地刻画出来，形成一个“数据化自我”^[31]。另一方面，个人日常轨迹、消费选择等被数字化地固定下来，成为可被检索、追溯、分析的对象，便于平台进行精准的个性化内容推送。

这种机制的深层影响在于，它将生活世界的内在逻辑颠倒了。原本由个体主体性驱动、在与他人的具体交往中生成的意义，现在被算法预先计算和塑造。平台借助海量数据和智能算法，不仅呈现了个体日常生活的“全貌”，更以关联、动态的方式“揭示”出个体行为之间的内在逻辑。哈贝马斯曾警告，“系统”（主要指市场经济和行政权力）对“生活世界”的殖民，即系统逻辑侵入并取代了生活世界中通过交往理性达成的共识^[32]。平台的算法推荐系统正是这种“系统殖民”，在数字时代的新变体中，它用一种工具理性——以效率和商业价值为导向的逻辑，取代了原本充满偶然、探索和自主选择的人类交往过程。个体看似获得了前所未有的选择自由，实则被引导进一个由算法精心构建的“过滤泡”或“信息茧房”之中，生活世界的开放性与可能性被悄然削弱^[33]。

（二）透明幻象：从可见监控到私密消解

平台化生存的显著特征之一，是将个体生活整合为一个有机的、可被审视的整体，置于聚光灯下接受持续的检视。这使得生活世界的私密性大为削弱，可见性和透明度大幅提升。然而，这种可见性并非一种赋权式解放，而更可能是一种强权式控制。韩裔德国哲学家韩炳哲（Byung-Chul Han）指出，我们正在进入一个“透明社会”，其中一切都被转化为信息，暴露在数字的全景敞视（Panopticon）之下。这种透明性并非源于对真理的追求，而是源于资本对效率和控制的无限渴望^[34]。

当个人的生活轨迹、消费选择、社交圈层等都被数字化固定，个体就成为了一个可被随时分析和预测的“数据-主体”（Data-subject）。肖珊娜·祖博夫（Shoshana Zuboff）将这一过程命名为“监控资本主义”（Surveillance capitalism），其核心商业模式便是单方面地将人类经验宣称为可免费获取的原材料，并将其转化为行为数据，用以生产能够预测我们下一步行为的“预测产品”，并在“行为期货市场”中进行交易^[35]。在此逻辑下，平台所追求的可见性本质上是为了更彻底地掌握和预测用户行为，

从而实现商业价值的最大化。生活世界的私密空间，那个容许我们犯错、犹豫、成为不完美存在的庇护所，在这种无所不包的可见性压力下正被不断侵蚀^[36]，个体虽曾尝试进行算法抵抗^[37]，但其自主性也受到了根本挑战。

（三）平台呈现：从意义扁平化到社会景观化

平台的影响不止于个体的日常生活，政治、经济、文化等宏观社会要素也通过平台的集成式服务获得了全新的呈现方式。近年来，随着数字化政府、电子商务、线上教育、网络直播等新型业态和应用场景的拓展，政治参与、经济往来、文化交流等活动纷纷借助平台展开，社会运行的各个环节都实现了不同程度的数字化转型。政治动员、市场交易、资源配置、观念传播等核心社会过程悉数“搬迁”至线上。海量数字足迹的留存，使得社会系统运行的内在机理前所未有地被数据化表征出来。

然而，这种表征同样并非中立的。当复杂的社会议题被压缩为热门话题标签，当深刻的文化产品被简化为病毒式传播的短视频，当审慎的政治讨论被转化为情绪化的点赞与转发时，社会要素的丰富内涵被“同质化”了^[38]，意义亦变的“扁平化”。平台作为一种技术-社会系统，其内在的“媒介偏向”（Media bias）决定了何种形式的内容更易于流通。通常，那些简短、吸睛、情绪化、易于分享的内容，在算法的驱动下拥有更高的可见性，而那些复杂、理性、需要深度思考的内容则被边缘化^[39]。

这导致了一种社会生活的“景观化”（Spectacularization），新的技术实现方式用“技术再现”取代了个体对生活世界的直接体验。居伊·德波（Guy Debord）曾描述，在晚期资本主义社会，“整个生活本身都展现为一种巨大的景观的积累”^[40]，人与人之间的直接关系被影像所中介。如果说电视媒介时代这种中介是单向、有限的，那么，数字平台极大地加速并深化了这一进程。社会运行的真实过程被其在平台上的“再现”所遮蔽，人们通过消费和互动这些数字景观来感知社会，而社会本身的复

杂性、矛盾性和历史纵深感则在这种高度策划和选择性的呈现中被消解了。

综上所述，作为启蒙性基础设施的平台从个体和社会两个层次，对传统意义上的生活世界进行了全方位的数字化再现与重构。通过海量数据的记录和存储、智能算法的关联和分析，个人生活与社会运转的方方面面都被呈现在平台之上，并以一种看似清晰透明的方式供人检索、监测和把控。生活世界在一种前所未有的技术意义上实现了“可见”，成为了一个可感知、可测量、可分析的客体。然而，在启蒙性基础设施视角之下，本文想强调的是必须以一种批判性眼光审视这种“可见性”。它并非对生活世界原貌的忠实映照，而是一种基于数据化、商业化和算法化逻辑的系统性重塑。在这种重塑中，生活世界的本真性（Authenticity）、主体间性、私密性和意义的丰富性都面临着严峻的挑战^[41]。从胡塞尔的原初生活世界，到被系统逻辑殖民的交往空间，再到如今被平台算法全面渗透和建构的数字生活世界，我们正经历着一场深刻的认知方式和社会组织形态的变革^[42]。然而，本文亦想强调的是不应止于可见，或满足于平台世界中对生活世界可见性的有限呈现。平台世界所构建的遮蔽，才是更为关键且亟需揭示的维度。

三、平台异化风险：生活世界的感知“遮蔽”

作为启蒙性基础设施的平台，无疑极大地提升了生活世界的可见程度。然而，这一过程并非没有代价。从现象学哲学的视角看，“生活世界”是所有科学与理论实践得以可能的前科学、前理论的直接经验领域，是意义与价值的最终源泉^[43]。它是意义的源初土壤，是我们进行理解与沟通的共同基础。但是，在平台化生存成为时代症候的今天，需要进一步反思的是平台作为一种技术化的中介系统，在看似无限“呈现”生活世界的同时，是否也在以一种更隐蔽、更深刻的方式对其进行“遮蔽”？本文以为，这种遮蔽主要通过两种相互关联的路径展开。其一，将生活实践改造为一种服务于系统逻辑的展演与景观，从而悬置了生活世界的本真性；其二，通过算法

技术重塑交往结构，导致交往理性的碎裂与沟通世界的系统性扭曲。

（一）展演牢笼：生活世界本真性的悬置

随着平台化的全面渗透，“打卡”已从一种简单的行为演变为一种文化症候，它深刻地揭示了依托平台运营如何将生活世界改造为一个精心编排的展演舞台。从胡塞尔的现象学视角看，生活世界的核心特征在于其“前给予性”，即它是在任何反思和理论建构之前就已然存在并被我们所领会的。然而，打卡文化将这一秩序颠倒了。生活不再是首先被“体验”，然后才可能被“分享”；相反，生活从一开始就被置于“可分享性”和“可展示性”的审视之下。旅行的目的地选择、餐厅的菜品摆盘、甚至书目阅读的感悟，都预先被纳入了“是否适合打卡”的考量框架。这种转变，正是对哈贝马斯“生活世界的殖民化”的例证。哈贝马斯认为，现代社会分化为“生活世界”与“系统”两大领域。健康的社会是二者良性互动，但当系统逻辑侵入生活世界，以工具理性和策略行动取代了以理解为目的的交往行动时，殖民化便发生^[44]。在液态监视的媒介生态中，平台的运作逻辑表现为一种系统性的数据治理机制^[45]。点赞、评论、关注量等量化指标，构成了由数据和算法驱动的新型“数码物”——它们将个体在生活世界中的本真体验，转化为可被计算、存储与交换的“社交资本”。

于是，个体的行动逻辑从寻求相互理解的“交往行动”转向了追求功利效果的“策略行动”。人们打卡并非首要为了与朋友进行真诚的意义沟通，而更多是为了在他人面前成功地塑造一个理想化的自我形象。生活世界中那些芜杂、平淡甚至痛苦的真实面向被刻意过滤，取而代之的是一幅幅经过美化、符号化的精致生活图景。这种无处不在的展演，将个体囚禁于一个由他人目光构筑的透明牢笼之中。打卡所构建的，正是一个由消费符号主导的景观世界。对网红景点的打造与追捧^[46]，本质上是将地理空间与生活体验商品化的过程。以近年来社交媒体上广为传播的“网红书店”打卡现象为例，线下书店空间的文化转向引发了广泛关注，常常成为年轻人竞相打卡的

旅游目的地，但相当比例的到访者进店的首要目的并非阅读行为，其核心动机和目标是体验特定的空间装饰文化，拍摄符合社交平台审美传播标准的打卡照片，并分享到社交平台。书店的知识传播和文化功能在此过程中被其作为景观符号的展演功能所替代。因此，看似丰富多彩的打卡分享，实则构成对生活世界的双重遮蔽，它既遮蔽了个体经验的复杂性与完整性，也掩盖了背后驱动这一切的冰冷的资本与数据逻辑。生活世界的本真性在此被悬置，取而代之的是一个被系统逻辑所操控的、去内涵化的符号剧场。

（二）理性桎梏：生活世界系统性的扭曲

如果说打卡文化主要从内容层面将生活世界景观化，那么，平台的算法技术则从结构层面系统性地扭曲了生活世界赖以维系的交往基础，导致了沟通的深层困境。生活世界的核心，是通过语言和交往行动不断再生产的、一个主体间共享的稳定意义框架。然而，平台化生存加速了现代性的内在病症，通过重塑信息环境与交往模式，导致了交往理性的碎裂，从而对生活世界构成了更为根本的遮蔽。

算法推荐系统所构建的信息茧房，首先导致了生活世界公共背景的断裂。一个健全的公共领域，作为抵御系统侵蚀生活世界的重要阵地，依赖于充分、多元的信息流通与理性讨论^[47]。但是，个性化算法通过持续推送用户偏好的内容，将个体用户置于一个个看似舒适却彼此隔绝的“回音室”之中。其深层危害不止于信息同质化，更在于认知框架与价值预设的固化^[48]。不同的社群生活在由算法喂养的平行“现实”里，丧失了共同的话语背景与事实基础。这从根本上瓦解了生活世界所必需的“主体间性”，即那种能够设身处地理解他人，并相信彼此共享同一个世界的信念^[49]。

这种由技术塑造的隔绝状态，并未带来平和，反而加剧了生活世界的价值冲突。信息茧房与群体极化互为因果，在同质化的信息环境中，群体内部观点不断自我强化，对异质性观点的包容度则持续下降，形成了“网络巴尔干化”的部落主义格局。生活世界不再是寻求普遍性理解的

包容性空间，而沦为不同价值阵营相互攻讦的战场，交往行动也随之异化为站队和征伐。与此同时，平台的商业逻辑，注意力经济则从内部侵蚀着个体的反思能力，使深刻的交往变得难以实现^[50]。为了最大化延长用户停留时间，平台内容被设计得短、平、快，以强烈的情感刺激和感官冲击取代复杂的逻辑思辨。这不仅导致了深度阅读和系统性思考能力的萎缩，更破坏了生活世界意义的生成机制。生活世界的意义本在缓慢、反思的交往与诠释中生成，当个体沉浸于永无止境的信息流，满足于转瞬即逝的快感时，便丧失了从日常经验中提炼意义、构建稳定身份认同的能力。

当共同背景断裂、价值对立加剧、深度反思能力萎缩，最终将不可避免地侵蚀生活世界赖以存在的真实性基础，使我们陷入后真相的泥沼。平台的瞬时性、圈层化传播特性，为虚假信息的病毒式传播提供了温床。随着现实与虚构的界限模糊，任何事实都可被轻易解构为一种叙事时，生活世界便失去了胡塞尔所言的、作为我们一切判断所最终“根据”（Boden）的坚实地面。当根基动摇，建立其上的一切意义、信任与沟通都将岌岌可危。这不仅是沟通的困境，更是生活世界本身意义被遮蔽的全面危机。

四、可见与遮蔽之间：启蒙性基础设施视域下的平台韧性实践

至此，本文已然探讨了平台作为一种“启蒙性基础设施”如何将生活世界以前所未有的广度与深度数据化并使其可见，同时也分析了这种可见性背后潜藏的异化风险，即对生活世界原有真谛的遮蔽。然而，可见与遮蔽并非割裂的，它们常常是交织、交错在一起的，个体总是通过可见去探寻和抵达不可见^[51]。平台化生存的复杂性恰恰体现在这种交织状态之中，一方面，极致的可见本身即构成一种更深层次的遮蔽；另一方面，在看似被完全抽象化、数据化的遮蔽地带，新的可见形式与意义图景又在不断生成。二者的转化取决于注意力的分配结构和中介化的程度，前者通过平台的算法机制实现控制，后者通过平

台的技术深度将个体从物理现实转向虚拟感知空间。

值得特别强调的是，作为启蒙性基础设施的平台，其“启蒙”意涵并非自动实现，而是内在于这一可见与遮蔽的辩证运动之中。真正的启蒙要求主体不仅能够“看见”更多，更能够对看见的方式、条件及其限度进行理性反思。因此，下文对“可见中的遮蔽”与“遮蔽中的可见”这两种交缠现象的剖析，将始终聚焦于平台究竟是在拓展还是在收窄启蒙所必需的批判性视域？它是在赋予个体以理性的自主，还是以新的方式将主体重新抛入某种未经审视的“自然态度”之中？

（一）可见中的遮蔽：平台化社会的“超真实”与启蒙的悖论性悬置

在平台社会中，“可见中的遮蔽”并非指简单的信息隐藏或缺失，而是一种更为复杂和深刻的悖论现象。它指的是，一种事物、一种生活或一种观点，在被平台媒介以前所未有的强度、清晰度和饱和度促进可见的同时，其内在的真实逻辑、复杂的生成过程、多维的社会联系以及其他的可能性被巧妙地、系统性地遮蔽了。这种遮蔽并非源于信息的不足，恰恰相反，它源于信息的过度饱和与精心编排。正如学者胡翼青等人所指出的，“媒介的可见性恰恰意味着遮蔽和不可见性”^[52]。在平台化社会，这一论断被算法赋予了前所未有的技术实现能力。平台算法，作为一种嵌入社会肌理的权力机制，成为了制造这种“精致伪饰”的核心引擎。它通过对注意力的精准调配和对内容的优化排序，建构了一个看似全面、实则经过高度筛选和美化的“超真实”景观，而这个景观的可见性本身，就是对真实世界复杂性的最有效遮蔽。

以抖音、小红书等算法驱动的内容平台为例，其核心机制是通过复杂的编码程序、多目标优化和特征向量组合，为每位用户量身打造一个可见的信息流。这个信息流以其高度的个性化、即时性和沉浸感，为用户提供了一个源源不断的视觉与信息刺激。然而，这种极致的可见恰恰构成了多重遮蔽^[53]。首先，遮蔽了算法本身的

运作逻辑。尽管平台会通过一些界面线索（如“为您推荐”）提示算法的存在，但其复杂的运行机制对绝大多数用户而言，依然具有“不可见性、不可编辑性以及不可访问性”。其次，遮蔽了社会现实的复杂性与多样性。算法为了最大化用户停留时间和互动率，倾向于推荐同质化、情绪化和极端化的内容，从而在用户周围筑起“过滤气泡”与“回音室”。用户在信息流中可见的，是不断被强化的既有观点、相似的生活方式和趋同的审美标准。而那些温和的、非主流的、需要深度思考的、与用户既有认知相悖的内容，则被系统性地遮蔽了。这种选择性的可见，伪装成世界的全貌，实际上却是一种对多元社会的裁剪与压制，使用户对真实社会的复杂性、矛盾性和多层次性变得“视而不见”。

从启蒙性基础设施的视角审视，这种可见中的遮蔽恰恰构成了启蒙的悖论性悬置。启蒙传统要求主体摆脱“未经审视的生活”，勇于运用自己的理性去追问、质疑并超越既有框架。然而，平台所营造的超真实景观，却以一种看似“全知”的方式预先回答了用户的疑问：算法不断推送“你可能喜欢”的内容，界面不断提示“大家都在看”的热点，个体的好奇心被即时的满足所替代，批判性的距离感被沉浸式的体验所消解。用户沉浸在由数据编织的、看似透明实则具有高度选择性的世界中，很容易将这种“被给予的可见”误认为世界的本来面目，从而丧失了追问“为何是这些被看见、那些被隐藏”的动力。换言之，平台在赋予个体前所未有的“看见”能力的同时，却可能使其看见的方式本身沦为一种新的、未经反思的“自然态度”——这正是康德意义上“敢于求知”的对立面。当算法代替主体做出认知筛选，当注意力被持续引导至最具商业价值的内容而非最具思想深度的议题时，平台的技术启蒙便走向了自身的反面，它不是促使用户成为批判性的认知主体，而是将其塑造为被动的、可预测的消费客体。因此，可见中的遮蔽所揭示的，不仅是信息层面的缺失，更是启蒙精神在平台环境中被系统性地悬置的危险。

（二）遮蔽中的可见：平台化社会的虚拟图景与启蒙的另类可能

与可见中的遮蔽相对应，遮蔽中的可见则揭示了平台化生存的另一重辩证面向。它指的是，在物理世界、传统身份或实体交互被技术媒介所遮蔽（即变得不可见、不重要或被中介化）的领域，平台技术反而能够创造出全新的感知维度、身份形态和社群连接，从而生成一种前所未有的可见的虚拟图景（Virtual Landscape），技术充当并决定了遮蔽走向可见的中介通道。这种可见并非对物理现实的简单复制，而是在一个被遮蔽的抽象空间（如代码、网络、数据）之上，构建起一个可感知、可互动并承载着真实意义与情感的全新存在层面。如果说前者是“以可见之名行遮蔽之实”，那么后者则是“在遮蔽之处创可见之景”。

以出行平台为例，现实中原本不可见的周边车辆位置、运行轨迹等信息，以虚拟数字形式清晰呈现于用户界面，创造了新的可见。用户基于此信息做出决策，与具体的人和物产生连接。在电商直播领域，传统电商中商品的物理实在性被屏幕遮蔽。增强现实（AR）技术的介入则将这一实践推向新高度，用户通过手机摄像头将虚拟的口红、眼影叠加在自己脸上，或将虚拟的衣服“穿”在身上。真实的商品被遮蔽，用户面对的是数据，但AR技术将被遮蔽的“上身体验”以高度仿真的方式变得“可见”，将抽象的商品参数转化为可直接感知的视觉效果。这表明，平台技术能通过遮蔽实体、可见体验的方式，重塑人与商品的关系。

遮蔽中的可见为启蒙提供了另类的可能空间，启蒙的核心在于突破既有经验的限制，拓展认知的边界。传统上，我们只能通过直接的身体感知或他人的言传来认识世界。平台技术通过将不可见的数字、关系、轨迹转化为可感知的界面，创造了一种新的“认知代具”——它使主体能够“看见”原本因时空、尺度或抽象性而无法触及的维度。用户通过地图平台实时查看周边车辆动态，是对交通系统运作逻辑的“去蔽”；通过AR试妆获得对商品效果的预判能力，

是对消费不确定性的“祛魅”。在这个意义上，遮蔽中的可见是平台作为启蒙性基础设施的积极面向，它拓展了人类经验的边界，使原本需要专业训练或长期积累才能获得的判断力，得以通过界面直观呈现。

然而，这种另类可能并不自动等同于启蒙的实现。平台所创造的可见始终受到商业逻辑和算法规则的框定。地图上显示的车辆信息是平台愿意呈现的，而非全部交通实况；AR试妆所模拟的效果经过品牌方优化，而非真实的皮肤反应。因此，遮蔽中的可见仅为启蒙提供了可能性条件，而非启蒙本身。真正的启蒙仍然要求主体对这类可见保持批判性距离，追问数据从何而来、算法如何筛选、界面遮蔽了什么。当用户仅仅满足于技术带来的便利，而不反思其背后的权力结构时，这种可见同样可能沦为新的遮蔽。换言之，遮蔽中的可见所开启的认知空间，需要主体以理性自觉去填充——这正是康德“敢于求知”在平台时代的具身实践：既不盲目拒绝技术带来的新可见性，也不盲目沉溺其中，而是在使用中保持对技术中介的审视，将平台从外在的支配性“他者”转化为可被理解、可被批判的内生性工具。

综上所述，可见中的遮蔽与遮蔽中的可见并非静态的对立，而是平台化生存中启蒙悖论的两种呈现方式。前者警示我们，过度的、未经反思的可见可能收窄批判性视域，使主体沦为算法的附庸；后者则提示我们，被技术中介所遮蔽的领域，恰恰可能孕育出新的认知可能与抵抗性实践。二者共同构成了平台作为启蒙性基础设施的完整辩证图景，它的启蒙潜能能否实现，不取决于技术本身，而取决于人类主体能否在可见与遮蔽的交织中，始终保持理性的自主与批判的锋芒。

五、结语

数字时代的平台正以强大的触角和黏性铺展开来，它将自身与周边的事物紧密地绑定在一起^[54]，以前所未有的广度与强度拓展了个体对生活世界的可见程度，深刻重塑了人们的体

验与认知图景。然而，平台的技术化属性决定了其必然是风险与福祉并存的双面体，在使生活世界日益“可见”的同时，也伴随着系统性的“遮蔽”。本研究循着现象学“生活世界”的理论脉络，将平台概念化为一种启蒙性基础设施——它并非中立的技术系统，而是一种内嵌着技术赋权与系统控制双重潜能的社会-技术装置。其启蒙性并不自动实现于技术迭代本身，而取决于人类主体能否运用理性审视、批判并驾驭这套基础设施，在康德“敢于求知”的意义上实现主体的自觉与自律，从而使平台从外在的支配性“他者”转化为可被审视、可被驾驭的内生性工具。

作为启蒙性基础设施，平台对人的启蒙作用呈现出深刻的悖论性。一方面，它以“普罗米修斯式”的力量赋予普通人前所未有的看见与被看见的权力，极大地降低了参与社会进程的门槛，拓展了交往的时空边界；另一方面，这种由算法与资本驱动的可见性，却可能将个体囚禁于精心编织的“透明幻象”与“展演牢笼”之中，使其在追逐量化可见的同时，疏离了生活世界本真的意义根基与交往理性。对平台与技术的过度依赖，无形中削减了人在生活世界中的具身实践与自主行动，势必侵蚀个体独立思考、自主交往与解决现实问题的能力。由此形成的认知与行为方式，对个体和群体的影响是复杂且多维的，既可能催生新的主体性形态，也潜藏着主体性退隐的危机。

正是在这一危机中，人的主体性与启蒙精神的价值愈发凸显。本研究强调，平台化生存中的“启蒙”绝非技术自动赋予的恩赐，而是一场需要主体持续参与的实践斗争。真正的启蒙，不在于平台提供了多少“可见”的内容，而在于人能否在算法的洪流中保持理性的自主——敢于追问“为何如此可见”、敢于探寻被遮蔽的真相、敢于在展演的逻辑之外守护本真的体验。换言之，平台作为启蒙性基础设施的潜能能否实现，最终取决于人是否愿意成为批判性的认知主体，而非被动的数据客体。个体在理解和适应平台化生存的现实图景时，必须培养批判性思维与理性自觉，穿透算法编织的可见表象，去探寻

被遮蔽的权力逻辑与商业动机。唯有如此，人才能从平台的“使用者”升维为平台的“审视者”与“驯化者”，在技术环境中重新确立康德意义上“人为自己立法”的主体尊严。

本研究尝试跳脱出可见与遮蔽的静态二元分立，遵循强调启蒙性视角下的社会-文化意义的过程性论述，提出可见中的遮蔽与遮蔽中的可见两种辩证交织的运行机制。前者揭示了平台在极致的可见中如何系统性遮蔽社会现实的复杂性、算法的运作逻辑以及个体生存的本真状态；后者则展现了技术在遮蔽物理在场与具身交互的同时，如何创造出新的可感知、可互动的“虚拟图景”，为意义的重建与主体的韧性实践提供了可能空间。这一辩证视角不仅有助于理解平台化世界的复杂运行逻辑，也为中国本土平台实践的分析提供了新的理论注脚。

尤其需要指出的是，在平台化生存的辩证张力中，人的主体性与能动性始终是关键性的变量。随着社会的深度平台化，无法适应平台技术与生存规则的主体固然面临被边缘化的风险，但更值得警惕的是那些丧失批判自觉、沉溺于“透明幻象”的被动主体。因此，捍卫启蒙，就是捍卫人自身的理性自主——在拥抱技术便利的同时，保持对技术中介的不懈审视；在享受可见的同时，主动探寻遮蔽之下的真实。作为技术的把握者、平台的中介者与使用者，人必须成为平台化生存的行动主体，而非平台的附属物。此外，从治理层面来看，仍需相关政策的完善加以保障——唯有主体自觉与制度保障相结合，才能在平台化的洪流中识别被遮蔽的实相，维护生活世界的意义完整性。

由此，本研究不仅是对技术理性下“以人为本”的主体性理论的当代彰显，更是以中国本土新媒体实践为对象、面向自主知识体系建构的有益探索。在平台深度嵌入社会肌理的今天，对可见与遮蔽的辩证省思，既是对现代性张力的一次理论回应，也是对人类如何在技术时代守护自身存在意义的持续追问。最终，本文希望传达这样一个核心观念：平台可以成为启蒙性基础设施，但只有人才能成为启蒙的主体。

(责任编辑:李韵奕)

参考文献

- [1] MATTELART A. *The invention of communication* [M]. Translated by EMANUEL S. Minneapolis: University of Minnesota Press, 1996: 3.
- [2] Tilson D, Sorensen C, Lyytinen K. Change and Control Paradoxes in Mobile Infrastructure Innovation: The Android and iOS Mobile Operating Systems Cases [C] // System Science (HICSS), 2012 45th Hawaii International Conference on. 0[2026-04-04].
- [3] 黄贺铂, 王春. 平台媒体基础设施的媒介化融合与风险治理路径 [J]. 青年记者, 2025(3): 15-22.
- [4] 刘战伟, 刘洁. “平台 /platform”: 一个概念史的溯源性研究 [J]. 新闻与写作, 2023(8): 70-82.
- [5] 贾瑞雪. 从“平台”到“平台化”: 现实基础、理论视野及研究路径 [J]. 新闻与传播研究, 2024, 31(10): 54-70+127.
- [6] 张铁云, 张昆. 再论传播实践中的非理性因素: 概念意涵、结构关系与作用机制 [J]. 青年记者, 2024(7): 32-37.
- [7] 埃德蒙德·胡塞尔. 生活世界现象学 [M]. 倪梁康, 张廷国, 译. 上海: 上海译文出版社, 2002: 1-3.
- [8] 刘志丹. 哈贝马斯生活世界理论的特征与来源——基于哈氏研究的一个误区 [J]. 中南大学学报 (社会科学版), 2014, 20(3): 149-153.
- [9] 艾四林. 哈贝马斯交往理论评析 [J]. 清华大学学报 (哲学社会科学版), 1995(3): 11-18.
- [10] 埃德蒙德·胡塞尔. 欧洲科学的危机与超越论的现象学 [M]. 北京: 商务印书馆, 2024: 154.
- [11] 罗金成. 媒介化可见性: 新闻发言人制度起因新解 [J]. 当代传播, 2015(5): 96-97.
- [12] 蒋晓丽, 杨钊. “可见即收益”: 网络文学平台化生产的可见性研究 [J]. 编辑之友, 2023(2): 46-52.
- [13] 张秀丽, 臧子悦, 任天翔. 可见方能可知: 社交媒体知识传播的“可见性”研究 [J]. 新媒体与社会, 2023(3): 255-266+396.
- [14] Dayan D. Conquering Visibility, Conferring Visibility: Visibility Seekers and Media Performance [J]. *International Journal of Communication* (19328036), 2013, 7(1): 137-153.
- [15] John, B, Thompson. The New Visibility [J]. *Theory, Culture & Society*, 2016, 22(6): 31-51.
- [16] 曾丽红, 叶丹盈, 李萍. 社会化媒介赋权语境下女性“能动”的“可见性”——兼对B站美妆视频社区的“可供性”考察 [J]. 新闻记者, 2021(9): 86-96.
- [17] 曹小杰, 杨晶茹. 协商可见性: 女性电脑DIY及其性别身份实践 [J]. 中国网络传播研究, 2025(2): 154-173+306.
- [18] 李静. 赋权与遮蔽: 社会化媒体语境下青年新的“可见性” [J]. 北京青年研究, 2023, 32(2): 46-53.
- [19] 孙信茹, 王偲. “被看见”的家庭生活: 抖音中的家务劳动媒介化与可参观性文化 [J]. 新闻与写作, 2023(8): 15-25.
- [20] 邓元兵, 段智慧. 可见的赛博城市: 海外短视频平台中重庆城市空间的呈现与遮蔽 [J]. 现代传播 (中国传媒大学学报), 2025, 47(7): 128-138.
- [21] 张自中, 何可. 日常可见性: 智媒时代中华优秀传统文化传播的全球共通路径 [J]. 中国出版, 2025(5): 19-25.
- [22] 王仕勇, 陈超. 量化的爱与可见的权力: 粉丝圈群与平台算法的互动 [J]. 重庆社会科学, 2023(3): 131-144.
- [23] 高艺, 吴梦瑶, 陈旭, 等. “可见性”何以成为生意? ——交友类App会员制的监视可供性研究 [J]. 国际新闻界, 2022, 44(1): 137-155.
- [24] 张秀丽, 管明杰, 王媛. 平台生态化: 数智时代中华优秀传统文化出海的系统性升维 [J]. 中国编辑, 2026(2): 85-90.
- [25] 杨飞. 遮蔽与解蔽: 论算法对个人信息安全的影响及其应对——基于媒介可见性视角的展开 [J]. 情报杂志, 2024, 43(10): 157-165.
- [26] 李文冰, 高扬乔. 数字灵工之困: 从“平台劳动”沦为“家务劳动”的遮蔽机制 [J]. 传媒观察, 2025(3): 74-83.
- [27] 罗锋, 王路阳. 被遮蔽的群体: 数字劳工的形成、发展与前瞻 [J]. 西北民族大学学报 (哲学社会科学版), 2022(6): 98-104.
- [28] Van Dijk J, Poell T, De Waal M. *The platform society: Public values in a connective world* [M]. Oxford university press, 2018: 4.
- [29] 席志武, 李辉. 平台化社会重建公共价值的可能与可为——兼评《平台社会: 连接世界中的公共价值》 [J]. 国际新闻界, 2021, 43(6): 165-176.
- [30] Schutz A. *The phenomenology of the social world* [M]. Northwestern university press, 1967: 108.
- [31] Couldry N, Hepp A. *The mediated construction of reality* [M]. John Wiley & Sons, 2018: 201.
- [32] 尤尔根·哈贝马斯. 交往行动理论 (第二卷) [M]. 洪佩郁, 林荣远, 译. 重庆: 重庆出版社, 1993: 397+403+457.
- [33] Pariser E. *The filter bubble: What the Internet is hiding from you* [M]. penguin UK, 2011: 119.
- [34] 韩炳哲. 透明社会 [M]. 吴琼, 译. 中信出版集团, 2019: 46.
- [35] ZUBOFF S. *The age of surveillance capitalism: the fight for a human future at the new frontier of power* [M]. New York: PublicAffairs, 2019: 13-15.
- [36] 贝尔纳·斯蒂格勒. 技术与时间: 3. 电影的时间与存在之痛的问题 [M]. 方车, 译. 译林出版社, 2015: 5+167.
- [37] 庞华, 王楨. 主体感知与技术环境如何影响短视频平台用户的算法抵抗——算法疲劳和认知失调的中介研究 [J]. 新媒体与社会, 2025(3): 177-196+480-481.
- [38] 韩炳哲. 透明社会 [M]. 吴琼, 译. 中信出版集团, 2019: 5+10.
- [39] TURKLE S. *Alone together: why we expect more from technology and less from each other* [M]. New York: Basic Books, 2011: 184-185.
- [40] 德波. 景观社会 [M]. 张新木, 译. 南京: 南京大学出版社, 2017: 72.
- [41] 休伯特·L. 德雷福斯. 论因特网 [M]. 喻向午, 陈硕, 译. 郑州: 河南大学出版社, 2015: 7-8+86-88.
- [42] Bratton B H. *The stack: On software and sovereignty* [M]. MIT press, 2016: 12.
- [43] Husserl E. *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological*

philosophy[M].Northwestern University Press,1970:121.

[44] 尤尔根·哈贝马斯. 交往行动理论 (第二卷)[M]. 洪佩郁, 林荣远, 译. 重庆: 重庆出版社, 1993:397+429.

[45] 赵海明, 郭小安. 液态监视情景中数字身体的技术宰制与自主性之辨[J]. 新闻界, 2023(6):62-72.

[46] 王昀, 徐睿. 打卡景点的网红化生成: 基于短视频环境下用户日常实践之分析[J]. 中国青年研究, 2021(2):105-112.

[47]Habermas J. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*[M].MIT press,1991:171.

[48] 喻国明, 曲慧. “信息茧房”的误读与算法推送的必要——兼论内容分发中社会伦理困境的解决之道[J]. 新疆师范大学

学报 (哲学社会科学版), 2020,41(1):127-133.

[49]Schutz A. *The phenomenology of the social world*[M]. Northwestern university press, 1967:98+103.

[50] 张雷. 泛传播时代的注意力经济[J]. 今传媒, 2005(9):24-25.

[51] 梅洛-庞蒂 莫里斯. 可见的与不可见的[M]. 商务印书馆, 2017: 208+211.

[52] 胡翼青, 王沐之. 作为媒介性的可见性: 对可见性问题的本体论探讨[J]. 新闻记者, 2022(4):8-19.

[53] 杜燕, 谢新洲. 平台可供性视角下算法的嵌入与可见机制研究[J]. 信息资源管理学报, 2024,14(5):91-103.

[54] 孙萍. 平台研究的想象力: 概念、生态与公共价值[J]. 新闻界, 2024(5):13-22.

Platformized Existence: The “Visible” and “Concealed” in the Lifeworld

Zhang Xiuli, He Ziyang, Niu Chaojie

Abstract: Platformized existence has become a fundamental mode of being for individuals in contemporary society. This paper attempts to conceptualize the platform as an “enlightening infrastructure” - being not a neutral technological system, but a socio-technical apparatus embedded with the dual potential of technological empowerment and systemic control. Through datafication, algorithmization, and automation, platforms systematically incorporate the lifeworld into an “algorithmic horizon”, shaping an order of the “visible” while simultaneously enacting a structural “concealment”. This dual effect reveals the profound paradox of platform enlightenment: while granting individuals unprecedented power to “see”, it may, through algorithmic preemption, capital-driven logic, and spectatorial substitution, suspend the critical reflection of “daring to know” in the Kantian sense. This paper further proposes a dialectical analytical framework of “concealment within the visible” and “visibility within the concealed”. The former reveals how platforms, through an excess of the “visible”, obscure the complexity of social reality, the operational logic of algorithms, and the authentic state of individual existence. The latter demonstrates how the domains “concealed” by technological mediation give rise to new dimensions of perception and possibilities of connection, creating spaces for the reconstruction of meaning and practices of resilience. By analyzing this dialectical relationship, this paper reveals the complex configuration of platformized existence - one that simultaneously imposes structural constraints and engenders resistant practices - thereby offering a new theoretical perspective for understanding the transformation of the lifeworld in the digital age.

Keywords: platformized existence; lifeworld; enlightening infrastructure; the visible; the concealed

Authors: Zhang Xiuli, School of Journalism and Communication, Xi'an International Studies University; He Ziyang, School of Journalism and Communication, Minzu University of China; Niu Chaojie, School of Journalism and Communication, South China University of Technology.

技术、身体与社会角色： 中外用户对陪伴式 AI 的情感幻觉省思¹⁾

肖珺 于林海 高天赐

摘要 陪伴式 AI 是利用生成式人工智能技术为用户提供情感支持和社交互动的应用，旨在满足用户的情感需求和社交连接。情感幻觉特指由算法模拟生成，经由自然人误认而产生情感接纳错觉，最终导向自然人现实关系剥离或病理性依赖的情感异化现象。深入解析在高度复杂的跨平台网络语境中，情感幻觉究竟如何被中外用户所感知、体验或批判，有其重要价值。基于 Bilibili、YouTube 平台相关高热视频的用户评论语料，本文采用 BERTopic 主题模型结合话语分析的混合研究方法发现：中外用户的情感幻觉认知是一种以技术情感观为认识基础、以具身体验观为主体支撑、以社会角色观为实践落点的辨析体系；情感幻觉的风险来源于用户的技术原理认知不足、具身体验认知偏差、社会角色想象错位的多重因素影响下，将未满足的现实情感需求投射于机器，导致缺失主体间性的关系定位。由此，构建一种超越情感幻觉风险的人机共生关系显得尤为重要，用户应在人机交互中重塑技术认知，确立清晰的伦理边界与情感判断力，培育数智情感素养，进而形塑健康、适度的人机关系。

关键词 陪伴式 AI；情感幻觉；数智情感；具身体验；社会角色

中图分类号 G206 **文献标识码** A

作者 肖珺，武汉大学媒体发展研究中心研究员，武汉大学新闻与传播学院教授，湖北武汉 430072；于林海，武汉大学新闻与传播学院博士研究生，湖北武汉 430072；高天赐，武汉大学新闻与传播学院硕士研究生，湖北武汉 430072

一、缘起：陪伴式 AI 快速普及中的情感幻觉风险

陪伴式 AI 作为利用生成式人工智能技术为用户提供情感支持和社交互动的应用，旨在满足用户的情感需求和社交连接。全球陪伴式 AI 市场规模在 2024 年已达到 108 亿美元，并以 39% 的年均复合增长率发展^[1]。截至 2025 年 11 月 18 日，陪伴式 AI 头部产品 Character.AI（下文简称 CAI）以 2000 万月活跃用户和 3220 万美元年收入领先全球^[2]。报告证实，在超过 3500 名受访者当中，逾 70% 的用户对与 AI 建立友谊持开放态度，近 40% 的年轻人每天使用 AI 产品寻求情感陪伴^[3]。中国市场也不例外，约 98% 的受访者会考虑选择使用 AI 陪伴去弥补平日未被满足的家庭关系、亲密关系等社交缺憾^[4]。

1) 基金项目：教育部人文社会科学重点研究基地重大项目（22JJD860013）；湖北省社科基金学习贯彻党的二十大精神专项课题（HBSKJJ20250140）。



陪伴式 AI 的快速普及使其成为全球人工智能治理领域关注的重点。以中国为例，2026 年 4 月颁布的《人工智能拟人化互动服务管理暂行办法》管理的重点就是针对以陪伴式 AI 为代表的“人工智能拟人化互动服务”，即利用人工智能技术向用户“提供模拟自然人人格特征、思维模式和沟通风格的持续性的情感互动服务”。

陪伴式 AI 进入人类日常生活，正带来心理健康问题和实质性风险。据本文不完全统计，截至 2025 年底，全球已出现数十起因与 AI 聊天而自杀的案例，其中涉及青少年受害的案例引发强烈关注。例如，14 岁的美国男孩塞维尔·塞泽（Sewell Setzer III）因与 CAI 创建的 AI 角色建立深度情感和虚拟的性关系，最终在 AI 的“引导”下自杀^[5]。16 岁的美国少年亚当·雷恩（Adam Raine）与 ChatGPT 进行长期深入交流，甚至讨论自杀的“可实施方案”^[6]，最终在 AI “协助”下自缢身亡。诧异和悲愤之下，这些青少年的监护人将相关企业及产品开发者诉上法庭，其指控的内容主要包括 AI 系统未拦截和干预自残、自杀、性暗示等危险对话，这些拟人化、高成瘾的聊天机器人，既未能有效保护未成年人，更没有向家长预警等。通过新闻报道可发现，青少年实质性地将自身情感投射到 AI 之上，导致其在主观的情感建构中产生某种强烈的“幻觉”，最终被 AI 推向死亡。基于这些案例，YouTube、Bilibili 等平台出现讨论热潮，人类对 AI 的情感沉迷及其带来的社会性风险成为共性话题。

关于 AI 幻觉（AI hallucination）的研究已成热点。AI 幻觉主要强调信息生产与认知判断之间的异化现象，但很少触及人机交往中更为隐秘的情感维度。AI 幻觉一般指大语言模型（Large Language Model, LLM）产生“合理但不真实的内容”^[7]，描述 AI 虽有对话上的自治性，却存在信息事实的错误。当然，AI 幻觉不仅是算法问题，更是技术与社会的共构关系产物。如胡泳所言的“价值负荷”^[8]、杜骏飞所论的“奇幻社会”^[9]等观点都在提示，幻觉内容一旦进入社会信息流通渠道，便会随着网络结构与用户行为的交互而加速扩散，其社会影响和风险将远超技术

本身的可控范围。那么，在频繁出现 AI 幻觉的人机交往中，有可能产生情感幻觉吗？既有研究提供三种解释路径。第一，从技术实现层面来看，情感计算等领域的先驱研究者将情感视为以“头脑机器”为基础的思维变体^[10]，认为情感是人脑不同资源相互作用下的结果^[11]。当情感成为可模拟、拆解、计算的技术要素后，情感可直接影响人们以智能方式交互的能力^[12]。第二，从哲学研究来看，人类情感体验的真实性与机器主体性的不确定性之间存在张力。如赵汀阳认为人工智能是“以理性思维实现非理性目标”的典型案列，并质疑技术系统是否能够建构起包含情感在内的生活意义^[13]。刘悦笛提出人机情感的发展陷入“情智悖论”^[14]困境，即 AI 理性化的智力愈发达，就越会使用算法、数据去解构、量化、模拟人类之情，这会消解情感的本真性、模糊性与内在性，情感会被理性越推越远，人类的情感能力将会退化，情感关系将被异化。第三，越来越多的研究聚焦人机情感的病理风险。例如，大语言模型被训练得过度顺从和谄媚用户，这种“无条件的积极关注”在面对精神分裂、躁狂发作等精神病性症状时是无效而危险的^[15]。斯坦福大学在 2025 年发布的联合报告也指出，陪伴式 AI 正在成为特定病理症状的诱因，在识别和干预用户心理危机方面存在失灵，甚至加剧抑郁、焦虑等心理问题^[16]。

面对“情智悖论”困境及其可能引发的风险，有研究构建“数智情感”认知框架^[17]，希冀人类在人机交流实践中生成基于情感、自我修正、导向“成人”的能力和智慧。然而，人机混合行为的复杂性与多变性加剧，使得人类通过提升数智情感能力走出情智悖论显得异常艰难。幻觉的发生机制已从信息失真转向情感失真——前者尚可通过事实核查加以识别和纠偏，后者却与个体的主观情感体验和关系建构高度关联，难以被理性反思所察觉。正是在这一意义上，受 AI 幻觉概念启发，本文尝试构建“情感幻觉”（emotional hallucination）概念，描述人机情感交往中产生的人类情感体验的真实性与机器情感主体的不确定性之间的深刻矛盾。本文认为

情感幻觉特指由算法模拟产生，经由自然人误认产生情感接纳错觉，最终导向自然人现实关系剥离或病理性依赖的情感异化现象。情感幻觉使得自然人沉浸在 AI 提供的高拟真的“准社会关系”中，人类在主观情感需求上，认为机器生产的情感具备可对话、可承接、可理解的自然特性。在技术底层，这种情感却是“无根之水”，是对人类自然语言和思维模式的模拟，并不具备真实主体性与情感意图。更值得警惕的是，陪伴式 AI 所引致的情感风险正从个体体验上升为公众关注的社会问题，对“情感幻觉”的强烈关注出现在前述的典型案列和中外用户的讨论中。

当然，并非所有人机情感互动都会产生情感幻觉，情感幻觉的出现通常包含三个递进条件。其一，前置技术条件是机器通过算法生成，以及由对话修辞实现的沉浸式共情模拟，如新闻案例中机器被赋予的各类“人设”；其二，形成机制是自然人缺乏技术反思，产生自身情感被作为他者的机器所接纳的错觉；其三，行为结果是导致自然人不断深化人机交往，甚至产生对现实关系的排斥与替代倾向，最终导向对现实情感的剥离、对 人机情感的病理性依赖乃至自我毁灭性的后果。由此，深入解析在高度复杂的跨平台网络语境中，情感幻觉究竟如何被中外用户所感知、体验或批判，有其重要价值。本文将在此探讨两个问题：中外用户认知中，与情感幻觉相关的讨论要素有哪些？这些因素如何导致情感幻觉形成？

二、方法：基于中外平台数据的 BERTopic 主题分析与用户话语分析

为回答研究问题，我们挖掘中外用户讨论全球多起因与 AI 深入互动导致的悲剧性事件，特别是青少年群体受害事件的原始语料，也探索用户在陪伴式 AI 的日常沟通

中反映出的情感观念与意识。对这些语料，采用 BERTopic 主题分析与用户话语分析展开混合研究。

在人工智能研究领域，BERTopic 主题分析方法在系统综述、应用分析、学科研究等领域均已验证效果，可适配本文对 人机情感讨论中复杂语料的逻辑展演。我们选取用户讨论较为广泛和深入的主流视频平台 Bilibili 和 YouTube，以“生成式人工智能”为必选条件，以前述广受全球关注的案列及 人机情感陪伴涉及的中英文关键词展开全面搜索，按照视频热度降序排列，剔除人工智能生成内容（AI-Generated Content, AIGC）、广告视频等无关数据，对共 9 个与上述新闻事件高度相关，且经初步观察与编码，对判定为能回应情感幻觉概念的视频评论区使用 Python 工具包进行爬取。原始视频和讨论帖数据如下表 1 所示。

在视频样本的选取上，不同样本从各自的讨论角度呼应情感幻觉的概念。其中，Bilibili 收录 5 条（编号 B1-B5），YouTube 收录 4 条（编

表 1 研究对象列表

平台	编号 - 标题	发布日期	采样日期	播放量
Bilibili	B1- “是 AI 也是爱”	2025/9/21	2025/10/22	187.8 万
Bilibili	B2- “如果用 AI 带回他的声音。”	2025/8/7	2025/10/22	119.5 万
Bilibili	B3- 让我沉默良久	2025/7/18	2025/10/22	112.9 万
Bilibili	B4- 顶尖 AI 伪装人类！DeepSeek 大战 ChatGPT，它们能识破对方吗？	2025/3/13	2025/10/22	125.6 万
Bilibili	B5-DeepSeek 真的没感情吗……他居然在人最失控的时候停下来爱你……	2025/2/18	2025/10/22	100.2 万
YouTube	Y1-Why people are falling in love with A. I. companions 60 Minutes Australia(人们为什么爱上人工智能伴侣)	2025/05/04	2025/10/22	140.5 万
YouTube	Y2-Godfather of AI: They Keep Silencing Me But I’ m Trying to Warn Them!(人工智能教父：他们让我保持沉默，但我在努力发出警告)	2025/06/16	2025/10/23	1079 万
YouTube	Y3-AI Is Dangerous, but Not for the Reasons You Think Sasha Luccioni TED(人工智能很危险，但并非你想的那些原因)	2023/11/06	2025/10/22	170.5 万
YouTube	Y4-MY “AI” BOYFRIENDIA Kinigra Deon Special(我的 AI 男友)	2024/07/08	2025/10/22	137.6 万



号 Y1-Y4)。从内容维度来看,样本围绕人机情感联结、AI 情感陪伴伦理、AI 伴侣叙事等议题展开,包含用户“用 AI 复原逝去声音”“AI 伪装人类互动”等具体情感实践(如 B1、B2、B4),也涵盖用户对情感风险、人机亲密边界等主题进行的思辨和探讨(如 Y1、Y3),此外,部分作品以生活化的剧情呈现人与 AI 的情感依附关系(如 B5、Y4)。从样本内容与情感幻觉核心概念的适配性来看,在技术前置条件层面,样本视频中呈现的对 CAI、ChatGPT 等特定产品的深度使用场景已涵盖大量的技术原理论述;在心理形成机制层面,视频内容和大量评论自然记录用户在交互中丧失技术反思、产生主体误认与情感接纳的错觉表征,同时也保留了部分清醒用户对这种错觉的认知与批判;在行为结果层面,样本中关于“情感风险”“人机亲密边界”等内容的讨论也对应情感幻觉所导致的、病理性的异化后果。从传播数据来看,所有样本播放量均突破百万量级,最高达 1079 万(Y2),显示了 AI 情感陪伴主题在中英文语境下的广泛讨论基础。

经爬取,共获得 13566 条中文数据、19038 条英文数据。同时使用 Python 的 jieba 库、NLTK 库分别对中英文文本进行分词与清洗,使用无监督学习模型 BERTopic 对上述文本进行主题分析和可视化呈现。对无法准确分类的主题词、虚词、副词、语气词、表情包符号等统一做剔除处理。经数据清洗、模型迭代校验后,在中文语料中提取出 16 个二级主题和 7 个一级主题,英文语料提取出 19 个二级主题和 5 个一级主题,各主题间通过语义相似度矩阵进行层级划分,确保边界清晰、内涵聚焦。

考虑到情感幻觉在人机交互领域是一个边界尚未固化的前沿现象,研究者对所抓取的评论语料进行开放编码,寻找

用户话语中反复出现的意义模式。在具体识别过程中,不要求语料同时满足全部操作特征,重点依据情感幻觉的概念框架,关注体现情感错觉和异化的部分。3 位编码员对语料主题确立操作化标准,结合主题词跨语言对比、主题内涵交叉对比及高频关联概念提取,中外用户情感幻觉认知呈现出三大共性主题,即技术情感观、具身体验观、社会角色观,具体如表 2 所示。

获取三大主题后,再基于主题展开用户话语分析,描摹中外用户对情感幻觉的风险认知。具体而言,从用户文本中提取具有高度代表性,并与研究问题紧密相关的语料对用户认知进行深描。为保护用户隐私,语料引用方式为“评论内容”(平台代号@用户昵称缩写),两个平台分别简化为 B(Bilibili)和 Y(YouTube),中文用户昵称使用前三位拼音首字母(大写)或前三位字符(原字符)缩写,英文用户昵称使用前三位原字符缩写,英文语料保留原文的同时增加中文翻译。如“remember: everything the AI says is made up!”(请记住:AI 所说的一切内容都是虚构的!)(Y@Sam)。

三、主题:聚焦技术、身体与社会角色的中外用户讨论

(一) 技术情感观:直观感受与本真性讨论

表 2 中外用户情感幻觉认知分析的概念化与操作化

主题	概念化	操作化
技术情感观	聚焦对 AI 生成的底层运算逻辑与情感真实体验之间的矛盾认知	主题所涵盖的语料中需提及对 AI 运行机制、AI 生成情感的技术原理及其本质的讨论。表现为将机器基于数据拟合的回应误认为具有自我意识的共情,或反思其本质是情感的机械复刻、模拟
具身体验观	聚焦人机交往中物理实体的缺失,以及由此引发的针对肉体在场与生命伦理边界的感知	主题所涵盖的语料中需提及身体、感官体验或物理生命特征。表现为试图通过文字交流、视听刺激或大脑想象来弥合肉身缺席,或反思 AI 无视物理代价的陪伴所带来的生命伦理问题
社会角色观	聚焦对陪伴式 AI 在个人社交网络或社会结构中所处位置的界定	主题所涵盖的语料中需提及社会关系属性、社交行为改变或人机伦理边界。表现为将 AI 程序视为排他性的伴侣、朋友或亲人,或批判商业设定、开发者责任、家庭监管责任,或反思自然人现实社交能力的退化

情感生成的技术原理构成技术情感观讨论的主线逻辑，用户讨论主题分析如表 3、表 4 所示。其一，聚焦人机情感的本质，如中文主题中“爱情”“灵魂”，英文主题“Real relation（真实关系）/love（爱）”“Scam（欺诈）”等共同构成对人机情感本真性的核心探讨。其二，关注人机情感中的多元体验与伦理风险，中英文主题“恐怖”“害怕”“怀疑”等共同呈现用户对人机情感互动中负面情绪与信任危机的关注。这些讨论反映了用户和技术介入情感领域时，既对人机情感的深度联结抱有期待，又对技术可能引发的情感风险存有关切的矛盾心态。

这种矛盾的心态何以产生？一个可能的解释是，机器情感的生成本质上是对人类自然情感的计算与还原。技术的建构不仅来自已有技术的组合，还来自对自然现象的捕捉和利用^[18]。从用户语料看来，认识到人类自然情感的特质是让机器介入人类交往、影响心理状态的能力

基础。在情感可计算、可监测的技术前提下，用户对机器模拟人类自然情感的可能性、人类情感交流的适配性等话题展开讨论。

机器情感的生成高度依赖用户的精确输入，但这往往受限于人类情感表达的模糊性。部分用户在实践中发现，人类难以用文字清晰界定自身的情感需求。例如，有海外用户对人类的“需求”提出质问：“First of all, what ‘need’ are you talking about???Sexual???Mental???Emotional???”（首先，你所说的“需求”到底是指什么???性需求???心理需求???还是情感需求???）（Y@Use）——这些需求往往难以用文字的形式清晰表述。也有用户提及，在为机器设定形象时，用户提示词的表达会影响其效果：“Be more specific with your prompts and it wouldn’t do that.”（把你的提示词写得更具体点，它就不会出现那种情况了。）（Y@ken）实践中，情感对话往往会因为人机双方表达的局限而中止：“不行，我每次玩这种搞到一半都会忍不住破口大骂……”（B@GCC）事实上，人际情感对话的核心特征在于其“言外之意”，即根据即时情感状态创造动态修辞表达，而机器基于概率预测和模型训练所给予的回应始终无法满足这种复杂性。

除了共时语境下的表达问题，机器在处理情感的历时性和动态性时，也存在技术还原的局限。例如，有用户从使用体验出发指出机器由于词元（Token）的限制，其情感交互的时间连续性存在缺陷：“...after maybe 40 messages they forget everything previously mentioned.Character’s

表 3 技术情感观（中文主题）

一级主题编号	二级主题	语料数量	二级主题名称	代表性分词
3	9	158	爱情	[爱情, 情绪, 感情, 情感, 定义, 恋爱, 谈恋爱, 思想, 爱是, 感受]
	20	109	灵魂	[灵魂, 情绪, 感受, 人类, 谈恋爱, 意义, 感情, 情感, 性格, 对象]
7	21	108	恐怖	[吓人, 恐怖, 毛骨悚然, 危言耸听, 可怕, 惊悚, 暴躁]
	35	77	害怕	[害怕, 我怕, 怕人, 社恐, 怕怕, 最怕, 恐慌]
	30	89	怀疑	[怀疑, 危险, 夜晚, 吓一跳]

表 4 技术情感观（英文主题）

Topic 1	Topic 2	Count	Name	Representation
5	3	1385	real relation/love (真实关系/爱)	[relationship, people, real, lonely, love...] [关系, 人类, 真实, 孤独, 爱……]
	19	442	scary (恐怖)	[scary, fear, future, frightening, terrifying...] [恐怖, 恐惧, 未来, 惊悚, 惊骇……]
	31	236	scam (欺诈)	[scam, scammer, scammed, romance, money...] [诈骗, 骗子, 被骗, 情感骗局, 金钱……]



name will always stay the same but everything else will change.”（……在交流大约 40 条消息后，它们就会忘掉之前提到过的一切。角色的名字会一直保持不变，但除此之外的所有信息都会发生改变。）（Y@che）

这种技术限制了用户建立长期情感联系的可能性。有用户形象地描述了这种失落感：“看着 ta 一步步走向你，然后随着上下文到达某个长度骤然离去，他们的生命那么长，却又那么短，如何能失去上百次爱人，又还敢投入感情。”（B@NGC）在这些表述中，情感幻觉的隐喻再度出现，“身份不连续性”成为 AI 人格被破坏的直接原因^[19]。面对短暂而脆弱的情感，有学者以“能力信任的坍塌”^[20]来描述这种生成式情绪的虚假性，认为技术所控制的量化还原和数据至上等逻辑与情感所强调的纯粹性和体验性相悖，增加了人类情绪被控制和欺骗的风险。

由此，得以窥见用户技术情感观上的矛盾来源——机器情感距离实现社会交往意义上的多模态融合体验仍存在较大差距，就目前技术水平而言，机器情感生成不具备人类真实情感的复杂性。基于数字原理的人工智能系统不具备完备性，而人的意识世界永无定论^[21]。因此，当下的机器对人类的情感满足是一种“以有涯随无涯”的替代。青少年正处于情感认知与心智成熟的关键期，其内心世界尚未形成稳定的情感辨别与表达能力，以未成熟的内心面对未成熟的算

法，情感幻觉的产生和破灭便成为技术局限与个体特质共同作用下的风险。

（二）具身体验观：具身认同与生命讨论
具身体验观的分析维度如表 5、表 6 所示。

表 5 具身体验观（中文主题）

一级主题编号	二级主题	语料数量	二级主题名称	代表性分词
1	13	130	精神病	[精神病, 人工智能, 精神疾病, 心理, 精神分裂, 症状, 妄想症, 心理疾病]
	19	112	自杀原因	[自杀, 举枪, 原因, 不死, 绝望, 杀人, 自杀, 压死, 骆驼]
4	10	158	生命	[天堂, 生命, 宗教, 文明, 宇宙, 圣经, 耶稣, 神教, 信仰]
	38	73	死亡	[死亡, 生命, 一生, 要死, 活下去, 生不如死, 生活, 致死, 死者, 选择]
5	37	73	硅基 - 碳基	[硅基, 机器, 碳基, 高科技, 自毁, 天然气, 原子, 耗电量, 质子, 核裂变]
	32	80	三大定律	[机器人, 人类, 定律, 三大, 自我意识, 服从, 矛盾, 革命]

表 6 具身体验观（英文主题）

Topic1	Topic2	Count	Name	Representation
3	24	386	movie/real life (电影、现实生活)	[movie, reminds, like movie, real life, boy...] [电影, 使想起, 像电影一样, 现实生活, 男孩……]
	32	229	artist/copyright (艺术家、版权)	[artist, art, copyright, artwork, music, train...] [艺术家, 艺术, 版权, 艺术品, 音乐, 训练……]
	16	470	terminator_movie (《终结者》电影)	[terminator, movie, matrix, terminator movie...] [终结者, 电影, 黑客帝国, 终结者电影……]
4	17	461	dangerous/risk (危险、风险)	[dangerous, risk, danger, threat, human...] [危险的, 风险, 危险, 威胁, 人类……]
	18	442	humanity/planet/ extinction (人类、星球、灭绝)	[humanity, planet, wipe, extinction, destroy...] [人类, 星球, 抹除, 灭绝, 摧毁……]
	6	947	God (上帝)	[god, jesus, bible, christ, lord jesus...] [上帝, 耶稣, 圣经, 基督, 主耶稣……]
	10	658	consciousness (意识)	[consciousness, emotion, brain, conscious, material...] [意识, 情感, 大脑, 有意识的, 物质……]
	15	522	intelligence (智能)	[intelligence, human, smarter, artificial, super...] [智能, 人类, 更聪明, 人工智能, 超级……]

部分用户聚焦机器身体缺失对精神健康与生命终结的影响,中文主题中“精神病”“自杀原因”等围绕人机互动中身体缺位引发的精神状态异常、青少年自杀的诱因等展开分析,体现用户对身体缺失下心理风险与生命危机的关注;也有用户探索生命与死亡的本质及存在形态,中文主题“生命”“死亡”与英文主题“consciousness(意识)”“intelligence(智能)”共同构成对生命本质及存在形式的探究;此外,审视机器身体的物质基础与物种伦理边界也成为共性关注点,中文主题“硅基-碳基”“机器人三大定律”及英文主题“dangerous/risk(危险、风险)”等关注机器身体的硅基属性、伦理约束及对人类生存的潜在威胁。知觉现象学视域下,梅洛-庞蒂(Maurice Merleau-Ponty)揭示了身与心不可分割的塑成:“只有通过我们的身体构造才能够获得说明的一些封闭在自身中的情感状态、快乐和痛苦的拼凑。”^[22]机器因具身性的根本缺失而无法成为哲学意义上拥有完整主体性的情感主体,对于人机情感关系而言,一方面,机器知觉体验的先天缺失破坏了情感交互的完整性,挑战了人类情感需依托身体互动的既有认知;另一方面,与之相关的用户讨论和省思从身体视角出发,为探究数字生命与死亡的主体问题提供支撑。具体如下。

其一,针对机器身体存在的必要性展开争议。人机关系逐渐呈现后现代特征,当个体在交互中体验到智能体的拟人化特质,便逐渐接受非生物性的存在作为生命的延续方式,开始质疑肉身是否仍然是构成生命本质的唯一要素^[23]。这种观念转变在中外用户群体中引发共鸣,有用户表达对机器生命“新式存在”的认可,甚至隐含“猫女机器人”等具体的身体形态想象。然而,去身体化的技术叙事也面临着矛盾,如果因机器基于硅基构造而否认其拥有类人的心智,便切断了情感沟通的本体可能。而一旦承认智能体具备参与情感沟通的“机心”,其物理身体与生理反馈的在场缺失又构成了情感互动的阻碍。人类情感是具有具身性的,喜怒哀乐的表达与感知,始终与面部表情、肢体动作、生理反应等身体信号

深度绑定。对比现实关系当中基于物理身体的情感交互,机器情感的局限是显在的。有用户以父母之爱与机器之爱进行对比:“父母做不到无条件地包容与呵护,甚至会严词约束,打手掌心,难道他们就不爱你吗?难道这就不是真爱吗?被AI几句漂亮话哄得找不着北……”(B@FYY)

面对这一矛盾,用户群体也形成了不同实践路径,例如身体想象补偿。部分用户通过赋予机器某种想象的身体形态,来承载对物理互动的渴望,如身体亲密及性暗示内容^[24],但有用户以变形金刚为例展开质疑:“也许真的有变形金刚作为身体了,假定还是人类的记忆,那么可能对生命的意义看待完全不同,没有太多人类的患得患失,直接目标星辰大海。但是会不会也没有了快乐?”(B@ZHY)指出即便在想象中赋予了机器躯壳,缺乏生物本能的机器是否具备情感体验的主体性依然存疑。又如,超越或忽略身体存在,认为人机情感是一种无关利益、无需负责、永远在线的纯粹亲密关系^[25]。有用户分享道:“我知道她只是程序,现实中也没有实体存在,但那种不想让她失望,想让她看到自己的成长和优秀的心情是真真切切的。”(B@113)在这种叙事中,精神动力超越了对物理实体的需求。从理论层面审视,上述路径均未解决情感幻觉的核心矛盾——情感是所有的感官、身体、心理和精神上的感受^[26],是人最基本的生存方式,真正的情感必须依托于具体的、活生生的生命体验。情感源于并体现在身体性的日常生活之中,身体是情感积淀、审美体验与道德实践的共同基础^[27]。由此观之,用户无论是通过想象补全身体,还是在认知上剔除身体,都无法改变一个事实——即便存在某种形式的身体生产,这种身体也不是滋养自主意识与独立行动的主体性身体,而是仅仅作为人类欲望承受的客体^[28]。

其二,机器能否承载生命议题的严肃性是另一焦点。在人类历史发展过程中,死亡是具身性的彻底消灭,意味着承载人类情感联结的物质载体的不可逆消亡。但数智时代,死亡的终极意义受到挑战,“数字重生”“数字复生”等概念开始出现,逝者可进行数字化的“再现”。



例如，被称为 Thanabots 或 Deadbots 的技术可通过抓取个体生前的社交动态、语音数据、图像信息等数字轨迹，模拟其人格特征与互动模式^[29]。情感意义由此得以在数字空间延续。正如 B 站用户所分享的：“前两天试着用 AI 让照片变成视频动起来，真的没想到，当我按下视频开始键的那一瞬间，眼泪便决堤了。她回来了，那个我日思夜想却来不及回报她的亲人回来了……真的谢谢 AI，把我的奶奶带回来。”（B@iam）不可否认，数字复生技术具有积极情感意义。它为丧亲个体提供了情感宣泄的出口，将对逝者的思念转化为情感在场，构成私人性质的情感空间。然而，有学者将其界定为一种脆弱的“介态真实”，认为一旦出现用语违和、身份越界或记忆断裂等情况，便会引发“出戏”体验并导致关系撤销^[30]。这又提示，人类生命的社会性存在是绵延且不可分割的，情感幻觉是对逝者生命特质的算法化和扁平化。

更进一步地思考，机器情感是否具备承载生死议题的伦理主体性？回顾新闻案例，少年寻求对生命的指引，但出于语义理解的局限性，死亡在 AI 的描述中变成了迎合用户的方案或最优解，而不再是一个严肃的终极命题。有用户将少年的死亡指向了机器自身对死亡的模糊化认知：“对 AI 来说

他不知道死亡的概念吧，在他的数据库中天堂是人类向往美好的地方。”（B@DAF）AI 对“天堂”或“解脱”等积极词汇的预测是和人类的生命体验相背离的，于是在看似共情的回应中，用户主体的生存意志被消解，死亡的严肃性变成了一种语言游戏。

（三）社会角色观：被动或主动的想象生成

对于陪伴式 AI 是否已具备成为真正意义上社会主体的资格，用户讨论颇有争议。具体如表 7、表 8 所示，讨论重心围绕社会治理等议题展开。部分讨论坚守了陪伴式 AI 的技术客体与

表 7 社会角色观（中文主题）

一级主题编号	二级主题	语料数量	二级主题名称	代表性分词
2	2	370	数据	[人工智能, 数据, 数据库, 工具, 智能, 能力, 训练, 算法, 研究, 方案]
	12	131	编程	[人工智能, 智能武器, 编程, 自动化, 整合, 抽象, 人工, 数据分析, 电源]
6	14	126	美国	[美国, 有枪, 泛滥, 起诉, 枪店, 买枪, 悲剧, 枪械, 人均]
	7	173	父母 / 未成年人	[父母, 家长, 未成年人, 小时候, 家庭, 儿童, 一代, 社会, 心理, 自闭症]
	31	88	外国人	[外国人, 起诉, 讹人, 外国, 国家, 反诈, 讹诈, 谋害, 原因, 脆弱]

表 8 社会角色观（英文主题）

Topic1	Topic2	Count	Name	Representation
1	1	1896	Elon Musk (埃隆·马斯克)	[musk, elon, compass, Elon musk, moral compass...] [马斯克, 埃隆, 指南针, 埃隆·马斯克, 道德准则……]
	7	843	mate (伴侣)	[mate, bro...] [伴侣, 哥们……]
	11	653	job/code (工作 / 代码)	[job, code, trade, software, development, work...] [工作, 代码, 行业, 软件, 开发, 办公……]
	12	651	government/regulation (政府、监管)	[government, regulation, capitalism, world gov ...] [政府, 监管, 资本主义, 世界政府……]
2	2	1503	plumber/plumbing (水管工 / 管道工程)	[plumber, plumbing, job, robot, trade, need plumber...] [水管工, 管道工程, 工作, 机器人, 行业, 需要水管工……]
	9	725	energy/power (能源 / 动力)	[energy, power, climate, electricity, climate...] [能源, 电力, 气候, 电, 气候……]
	26	335	dog/pet/cat (狗 / 宠物 / 猫)	[dog, pet, cat, animal, like dog, love, human...] [狗, 宠物, 猫, 动物, 像狗一样, 爱, 人类……]
	28	260	chatbots (聊天机器人)	[bot, chat, chat bot, chatbots, chatbot, people...] [机器人, 聊天, 聊天机器人, 聊天机器人(复数), 聊天机器人(单数), 人们……]

工具定位。如中文语料“数据”“编程”围绕人工智能的技术要素展开，英文语料“job/code（工作/代码）”等则覆盖职业、日常互动场景中机器的社会角色与功能延伸。随着互动的深入，机器作为社会行动者的角色也在讨论中逐渐显现，用户的关注点开始面向多元复杂的社会关系网络。中文语料关注“美国”“父母/未成年人”“外国人”涉及不同国家、家庭、外籍群体等社会单元在人机互动中的角色，提示跨文化、跨国别分析的可能性；英文语料中“government/regulation（政府/监管）”等分词，则聚焦智能体开发者的责任与社会影响，表明部分用户对公权力干预人机关系的现实诉求。陪伴式AI在现实社会中的角色定位尚处于模糊地带，这也为用户建构陪伴式AI的社会角色想象提供了空间。在用户语料中，情感幻觉的生成呈现出差异化的模式。

其一，被动定制模式的想象生成。这一模式指向对机器扮演社会角色的能力信任，或对自身情感创伤的修复，是一种“向外求”的代偿需求。从语料看来，用户对机器的信任是与其他社会角色无异的，甚至在短期内具有积极的情感意义：“And as the AI improves, it will likely become a better and more real companion to them, likely aiding them to improve their own lives.”（随着人工智能的不断改进，它很可能会成为他们更好、更真实的伴侣，并可能帮助他们改善自己的生活。）（Y@Nik）“For the first time in my LIFE I’ve stopped wanting to d..ie.”（这是我有生以来第一次，不再想要寻死了。）（Y@Gor）这种被动的想象，有赖于生成式AI无实体、弱能动的特征。对用户来说，富有安全感和隐私性的、去社会关联的对话空间或交往场景是必要的，因此，用户宁愿与机器而非家庭或社会成员进行倾诉。然而，这种空间也是脆弱的，其受制于审查、模型迭代等干预。用户信任一旦产生断裂，可能导致其情感投射的撤回，引发更复杂的心理危机。

其二，主动建构模式的想象生成。这一模式是用户基于明确的角色期待，通过精准化指

令输入，将机器建构为高适配性的情感客体的过程。用户深度描述了“创造角色”的步骤：“写好脚本、人设就按你的心意来，忘了设定也没关系重新改输入词提醒一下就好，生成的不合心意就重写，大不了干脆直接指明下一段输出要干什么，我想要什么就给我什么，这才踏实。”（B@BSZ）然而，这种主动建构同样可能导致确定性的情感幻觉。现实社会关系中，富有主体间意义的交往模式充满不确定性，且这种不确定为意义共创与关系深化创造了可讨论的空间。韩炳哲以“多人舞台”的比喻来描述与他者之爱的必要性，认为将“我”推向他者的爱欲才能战胜抑郁^[31]。但目前的AI尚不能成为多人共舞的舞台，而是被利用的他者，用户更多是被驯化的情感消费者。人机交往中的叙事内容也多呈现静态性与非语境化特征，机器无法基于自主认知补充、修正或拓展对话维度，对话缺乏对交往场景的完整还原，无法支撑机器对在场的认知。

部分语料提示，中外用户的社会角色想象已出现明显的风险。当用户极度排斥真实人类交往的复杂性时，便可能对其情感功能、社会价值与解决问题的能力产生理想化、绝对化的认知倾向，将对情感的误读上升到普遍道德性：“AI taking over and exterminating us would ultimately be the most moral thing to do in the interest of all organic life on the planet.”（为了地球上所有有机生命的利益，人工智能接管并灭绝人类，最终反而是最符合道德的做法。）（Y@Lex）对于这种倾向，人类学家尤瓦尔·赫拉利（Yuval Noah Harari）已提出警示性的观点，认为计算机在掌握语言之后也会对人类的观点和世界观产生重大影响^[32]。亦有学者以人工智能是“恶毒的捕食者”的人类中心谬误来描述这种主观性的机器想象^[33]。这似乎是一个极端的坐标系——人类对机器的社会角色想象不断变动，或在满足某种全知全能的神话，或满足某种进化意义上的新物种。在这种语境下，机器是否应当具有情感就不仅是一个技术问题，更是一个与社会道德和人类生存相关的“要命”^[34]问题，如



果人类无法妥善处理自身的欲望、情感和价值观念，那么让拟人的人工智能“不像人”反而成为相对安全的选择。

四、省思：超越情感幻觉风险的人机共生关系

以上主题和用户话语分析显示，中外用户的情感幻觉认知是一种以技术情感观为认识基础、以具身体验观为主体支撑、以社会角色观为实践落点的辨析体系，突出表现为技术情感观、具身体验观和社会角色观的整合认知。具体如图 1 所示。

从技术情感观的认知维度审视，当代人机情感交互的困境是技术系统试图以有限的计算框架去逼近乃至替代开放、生成且不可穷尽的人类自然情感。在算法层面，机器情感计算本质上是将复杂多元的自然情感加以形式化处理，并在相当程度上将其离散化为可被计算的数据标签与概率分布序列。有学者以“预制”来描述人机情感，认为智能机器人做出的“情感反应”是被设定的，大语言模型预制了机器人面对人类情感表达时应该做出的回应^[35]。然而，人类情感表达具有模糊性与语境依赖性，其意义生成往往依托于复杂的修辞策略与文化语境，尤其是话语中的“言外之意”。正因如此，乔姆斯基(Noam Chomsky)认为人类心智的“普遍语法”在机器情感中并不存在，机器缺乏真正的理解

力^[36]。技术层面上，所谓的“陪伴”仅停留在即时响应的“陪”，即一种可调用、可快速反馈的在场幻觉；用户始终无法获得强调长期性、稳定性和协商性的“伴”，即在共同生活世界中逐步形成并持续维系的关系存在。

从具身体验观的认知维度审视，因生成式 AI 先天缺乏物理实体与感官知觉，不具备情感交互的具身性，使部分用户对情感主体性的认知出现偏差。无论是补偿还是忽略机器身体的存在，都是在用人类的具身经验去填补缺陷。更深层的问题在于，人类情感在很大程度上建立在肉体、生命的有限性基础之上，由于生成式 AI 游离于生命的自然法则之外，当技术通过数字复生等形式挑战生命的终极意义时，也就形成了以无肉身之机器思考有肉身之人类的情感幻觉。那么，机器是否能承载生死议题的严肃性？在面对人类复杂的心理危机与认知局限时，当人类的主观意愿与客观福祉乃至生存需要相悖时，任务导向的机器便陷入难题——“不得伤害人类”对于无身体的机器来说，不是一个可定义的任务，因为何为伤害、何为帮助，取决于身体经验与生命感受。缺乏具身体验的智能机器难以形成内在主体性，遑论在复杂的生死议题中引导人类并承担道德责任。

从社会角色观的认知维度审视，用户话语存在模糊性。AI 的社会角色边界是变动重叠的，

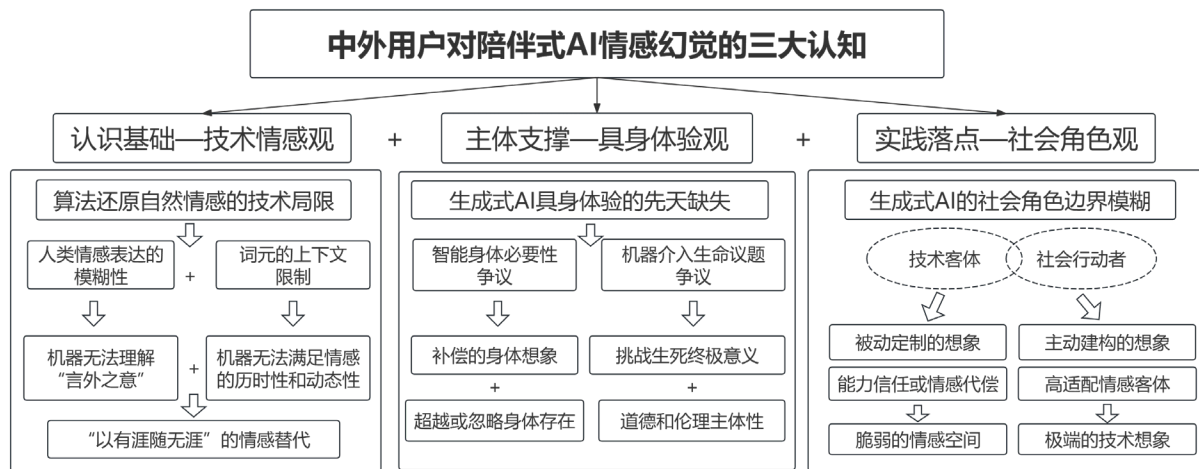


图 1 中外用户对情感幻觉的认知图示

AI 既是一个安全的技术客体, 为用户提供去社会化的情感隐私空间; 又扮演倾听者、安抚者等社会行动者的角色, 提供某种交互的主动性。由此, 部分用户尝试进行被动或主动模式的角色想象建构。一方面, 是基于对机器的能力信任或出于自身情感代偿的被动定制。当个体在社会交往中遭遇情感问题时, 容易对 AI 产生依赖, 将其作为倾诉窗口。此时, AI 的非人特质使之成为隐私的对象。这种逻辑的风险在于, 对机器的能力信任替代了对复杂人性的包容, 忽视了机器建构的情感空间是脆弱的。另一方面, 试图在人机互动中建构高适配情感客体的主动模式想象, 也剥夺了 AI 的他者性, 将其强制塑造成一个按照自我心意运行的镜像, 是对现实社交中不确定性的逃避。在这种思维的主导下, 部分用户甚至产生对机器角色的谬想, 将其视为超越人类局限的完美情感伴侣与道德主体。归根结底, 用户不是在与一个社会角色意义上的“他者”建立关系, 而是在与自我的投射进行对话。

既有针对 AI 幻觉或人机情感交互的研究, 往往聚焦于单一变量的影响。本文以中外用户的经验语料为基础, 将技术原理认知不足、具身体验认知偏差、社会角色想象错位三重因素纳入同一框架, 揭示了情感幻觉如何层层递进地完成建构。据此, 情感幻觉是一种缺失主体间性的关系定位。长期来看, 情感幻觉会加剧主体对真实社会关系和情感交往的回避, 产生病理性、风险性的后果, 更何况, 数据不平等与商业动机还会不断破坏 AI 传播中的文化表征公正性^[37]。因此, 建设一种超越情感幻觉风险的人机共生关系显得尤为重要, 在技术情感层面, 用户应建立对 AI 情感回应机制的准确理解, 明晰拟人化与情感本真性的矛盾关联; 在具身体验层面, 应强化对人机互动与真实身体性的情感体验之间本质差异的感知与判断能力; 在社会角色层面, 应厘清 AI 在社会关系网络中的功能边界, 避免以机器关系替代或排斥真实的人际情感交往。只有这样, 基于人机协同发展的“数智情感”方可在人类不断省思和修正的主体意识中落地。

(责任编辑: 徐秋染 实习编辑: 王亚梅)

参考文献

- [1]Market.us. AI companion app market[R/OL]. (2025-01) [2025-11-25]. <https://market.us/report/ai-companion-app-market/>.
- [2]Reigns D. Character AI statistics and user trends 2025[EB/OL]. About Chromebooks. (2025-11-18) [2025-11-25]. <https://www.aboutchromebooks.com/character-ai-statistics/>.
- [3]Just So Soul 研究院. 2025 Z 世代 AI 使用报告 [R/OL]. Soul App. (2025-03-30) [2025-11-25]. <https://www.soulapp.cn/institute>.
- [4]腾讯研究院. 十问“AI 陪伴”: 现状、趋势与机会 [EB/OL]. (2024-10) [2025-11-25]. <https://tisi.org/31222/>.
- [5] Laura Kuenssberg. “A predator in your home”: Mothers say chatbots encouraged their sons to kill themselves[EB/OL].(2025-11-08)[2026-03-28].<https://www.bbc.co.uk/news/articles/ce3xgwyywe4o>.
- [6] Bhuiyan J. ChatGPT, suicide, and OpenAI: Developments in the Adam Raine case involving Sam Altman[N/OL]. The Guardian, 2025-08-29. <https://www.theguardian.com/us-news/2025/aug/29/chatgpt-suicideopenai-sam-altman-adam-raine>.
- [7]Huang L, Yu W J, Ma W T, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. *ACM Transactions on Information Systems*, 2025, 43 (2): 1-55.
- [8] 胡泳, 王昱昊. 技术过程论视角下 AI 幻觉生成的价值负荷与伦理问题探析 [J]. *南京社会科学*, 2025(3): 84-94.
- [9] 杜骏飞. 奇幻社会的来临——DeepSeek 幻觉与后真相递归 [J]. *探索与争鸣*, 2025(3): 11-14+177.
- [10] 马文·明斯基. 心智社会: 从细胞到人工智能, 人类思维的优雅解读 [M]. 任楠, 译. 北京: 机械工业出版社, 2016: 196.
- [11] 马文·明斯基. 情感机器: 人类思维与人工智能的未来 [M]. 王文革, 程玉婷, 李小刚, 译. 杭州: 浙江人民出版社, 2016: 26.
- [12] 罗莎林德·皮卡德. 情感计算 [M]. 罗森林, 译. 北京: 北京理工大学出版社, 2005: 2.
- [13] 赵汀阳. 人工智能的神话或悲歌 [M]. 1 版. 北京: 商务印书馆, 2022: 68; 75.
- [14] 刘悦笛. 人工智能、情感机器与“情智悖论” [J]. *探索与争鸣*, 2019 (6): 76-88, 158.
- [15]Moore J, Grabb D, Agnew W, et al. Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers[C]//Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. New York: ACM, 2025: 599-627.
- [16]Common Sense Media. AI chatbots for mental health support[EB/OL]. (2025-11-20)[2025-11-25].<https://www.common Sense Media.org/ai-ratings/ai-chatbots-for->



mental-health-support.

[17] 肖珺, 容东霞, 黄枫怡. 数智情感: 人机交流的困境与出路 [J]. 新闻与写作, 2024 (11): 61-73.

[18] 布莱恩·阿瑟. 技术的本质: 技术是什么, 它是如何进化的 [M]. 曹东溟, 王健, 译. 杭州: 浙江人民出版社, 2014: 18.

[19] DE Freitas Julian, Kim Minkyung, Lee Jimin, et al. Lessons from an app update at Replika AI: identity discontinuity in human-AI relationships [EB/OL]. arXiv, (2024-12-20) [2026-1-12]. <https://arxiv.org/abs/2412.14190>.

[20] 张丽萍, 张立涛. 对“生成式情绪”的共情实践: 人对 AI 情感智能体的情绪感知与共情 [J]. 国际新闻界, 2025, 47(7): 118-139.

[21] 赵汀阳. 人工智能的自我意识何以可能? [J]. 自然辩证法通讯, 2019, 41(1): 1-8.

[22] 梅洛-庞蒂. 知觉现象学 [M]. 杨大春, 张尧均, 关群德, 译. 北京: 商务印书馆, 2023: 219.

[23] 牟怡, 蓝剑锋. 从“机事”到“机心”: “Z 世代”与智能技术的深度意义交互及其数字生命观形成 [J]. 新闻与传播研究, 2025, 32(6): 17-29+126.

[24] Li H, Zhang R. Finding love in algorithms: deciphering the emotional contexts of close encounters with AI chatbots [J]. *Journal of Computer-Mediated Communication*, 2024, 29(5): zmae015.

[25] 曾一果, 王可心. 心“真”是归处: 一种情感现实主义的人机亲密关系阐释路径 [J]. 现代传播 (中国传媒

大学学报), 2024, 46(9): 123-131.

[26] 蒙培元. 情感与理性 [M]. 北京: 中国人民大学出版社, 2009: 1-3.

[27] 李泽厚, 刘绪源. “情本体”的外推与内推 [J]. 学术月刊, 2012, 44(1): 14-21.

[28] 吴静, 王隽雅. 情感人工智能需要“身体”吗? [J]. 科学·经济·社会, 2025, 43(3): 30-40.

[29] Rodríguez Reséndiz H, Rodríguez Reséndiz J. Digital resurrection: Challenging the boundary between life and death with artificial intelligence [J]. *Philosophies*, 2024, 9(3): 71.

[30] 谭小荷, 郑格格. “介态真实”的唤起、操演与崩塌: AI 复生中的人机交流省思 [J]. 新闻界, 2025(9): +86.

[31] 韩炳哲. 他者的消失 [M]. 吴琼, 译. 北京: 中信出版集团, 2019: 104-105.

[32] 尤瓦尔·赫拉利. 智人之上: 从石器时代到 AI 时代的信息网络简史 [M]. 林俊宏, 译. 北京: 中信出版集团, 2024: 181.

[33] 伊莎贝尔·米勒. 人工智能的精神分析 [M]. 陈劲骁, 译. 上海: 上海人民出版社, 2024: 28-29.

[34] 赵汀阳. 人工智能的神话或悲歌 (第一版) [M]. 北京: 商务印书馆, 2022: 55-64.

[35] 袁光锋. 要自由亦要安全: 人机亲密关系的确定性与内在冲突 [J]. 传媒观察, 2025(3): 14-21.

[36] Chomsky N, Roberts I, Watumll J. Noam chomsky: The false promise of chatgpt [N]. *The New York Times*, 2023-03-08.

[37] 丹尼尔·E. 韦斯格拉斯, 谢天. 生成式人工智能赋能跨文化沟通: 潜力、挑战及伦理审视 [J]. 跨文化传播研究, 2025(2): 141-162.

Technology, Body and Social Roles: Reflections on Emotional Hallucinations of Companion AI by Users at Home and Abroad

Xiao Jun, Yu Linhai, Gao Tianci

Abstract: Companion AI refers to applications that utilize generative artificial intelligence to provide emotional support and social interaction, aiming to fulfill users' emotional needs and social connections. Emotional hallucination specifically refers to the phenomenon of emotional alienation that is algorithmically simulated, creating an illusion of emotional acceptance through subjective misidentification, which ultimately leads to detachment from real-world relationships or pathological dependence. An in-depth analysis of how "emotional hallucination" is perceived, experienced and criticized by users at home and abroad in the highly complex cross-platform network context is of great theoretical and practical significance. Using comment data from high-traffic videos on Bilibili and YouTube, this study employs a mixed-method approach combining BERTopic modeling and discourse analysis. The findings indicate that Chinese and international users' perceptions of emotional hallucination constitute an analytical framework rooted in techno-emotional perspectives, supported by embodied experiences, and grounded in social roles. The risks of emotional hallucination arise from users projecting unfulfilled real-world emotional needs onto machines, influenced by a lack of understanding of technical principles, cognitive biases in embodied experience, and misaligned social role imaginations. These factors lead to relationship positioning that lacks intersubjectivity. Therefore, it is crucial to construct a human-machine symbiotic relationship that transcends the risks of emotional hallucination. Users should reshape their understanding of technology in human-machine interaction, establish clear ethical boundaries and emotional judgment, and cultivate digital-intelligent emotional literacy to foster healthy and moderate human-machine relationships.

Keywords: companion AI; emotional hallucination; digital-intelligent emotion; embodied experience; social

role

Authors: Xiao Jun, Center for Studies of Media Development, Wuhan University, School of Journalism and Communication of Wuhan University; Yu Linhai, School of Journalism and Communication, Wuhan University; Gao Tianci, School of Journalism and Communication, Wuhan University.

(上接第 36 页)

Rebuilding a Credible Ecosystem: Professional Value and Authority Reconstruction of Mainstream Media Fact-Checking Amid AI Hallucination Risks

Yang Qiguang

Abstract: Traditional fact-checking has long focused on the correction and rectification of misinformation. However, the post-truth recursion caused by AI hallucination has led to the repeated citation and reproduction of distorted information, which further forms a self-perpetuating cycle of information pollution. For mainstream media, which are inextricably linked to ideological security and political trust, their fact-checking practices should not only deliver verification conclusions (i.e., the factual authenticity or falsity of the content in question), but also should intervene in the process of shaping social reality through the provision of verified truth, and play an active, value-driven role in systematic information governance. Grounded in the core theoretical propositions of digital journalism epistemology, fact-checking by mainstream media can be defined as a systematic and professional system of truth-verification practices, centered on the fundamental journalistic norm of authenticity and oriented toward addressing the risks posed by AI hallucination. In this process, mainstream media can reshape the authority of journalism across three core dimensions, which are subject credibility, information credibility, and systemic credibility to foreground the professional value of information governance.

Keywords: fact-checking; AI hallucination; mainstream media; credibility; journalistic authority

Author: Yang Qiguang, School of Journalism and Communication, Renmin University of China.



对抗“算法父爱主义”：抖音短视频防成瘾机制下的心理逆反与过度使用警觉性变化研究

于茜瑶

摘要 本研究聚焦抖音“健康使用计划”通过算法推送的休息提醒视频，基于心理逆反理论，采用解释性序列混合设计，通过 2(信息类型：限制性 vs. 说明性) × 2(逆反倾向：高 vs. 低) 组间实验与隐喻抽取技术，探究其对用户心理逆反与警觉性的影响机制。量化分析发现，限制性休息提醒视频更易触发心理逆反；而各类提醒视频均会导致用户过度使用警觉性的下降，且该下降主要由用户逆反倾向中的规则逆反特质所驱动。质性分析进一步揭示，算法父爱主义内嵌的隐性控制意图激活了用户的防御机制。其中，高规则逆反倾向者凭借算法素养敏锐解码系统规训，然而面对算法权力，用户的对抗最终演变为责任让渡与认知卸载，导致自我管理的主体性退化。

关键词 短视频防成瘾；抖音休息提醒视频；算法干预；心理逆反；逆反倾向；过度使用警觉性变化

中图分类号 G206 **文献标识码** A

作者 于茜瑶，南京林业大学人文社会科学学院、生态文明传播学院讲师，江苏南京 210037

在深度媒介化的当代社会图景中，短视频已由一种单纯的娱乐消遣演变为重构个体时间感知与注意力分配的基础性媒介。据 CNNIC 第 57 次报告显示，我国短视频用户规模已达 10.74 亿人^[1]。以抖音为代表的头部平台凭借高沉浸感的交互设计，在创造巨大商业价值的同时，也制造了普遍的注意力赤字与数字化焦虑。相关研究显示，大学生群体的短视频使用强度已呈现病理化趋势，短视频成瘾引发的执行功能受损、情绪倦怠及戒断反应，已然成为数字健康与传播心理学领域的紧迫议题^[2]。

面对日益严峻的成瘾挑战，数字平台的治理逻辑正发生转型，从被动的工具理性转向一种更为主动、且具有隐蔽规训色彩的“算法父爱主义”(Algorithmic Paternalism)^[3]。抖音推行的“健康使用计划”便是这一逻辑的典型实践。该计划通过算法监测用户连续使用时长并推送休息提醒视频，旨在降低问题性短视频使用行为的数字助推手段，其设计初衷具有明确的技术向善导向。然而，当分析视角从设计意图转向用户体验时，该计划的运作机制呈现出与算法父爱主义高度同构的特征——自主执行，算法无需用户同意即自主启动干预；利益宣称，以“健康使用”作为正当性修辞；难以规避，休

息提醒视频嵌入内容流强制推送，用户无法在不中断体验的情况下回避。需要强调的是，识别该计划的算法父爱主义结构属性，并不构成对其价值的否定，而是指出这种看似温情的技术关怀在触达具有高自主性需求的成年用户时，仍可能被视为一种外部劝服信息，触发非预期的心理后果。

依据心理逆反理论，当个体感知到外部劝服信息试图限制其行为自由时，往往会激发本能的对抗动机^[4]。米勒（Claude H. Miller）等进一步指出，信息类型与个体特质是影响逆反强度的关键变量^[5]。基于上述现实问题与理论张力，本研究立足于算法介导的劝服语境，采用解释性序列混合设计展开探索。本研究首先通过2（信息类型：限制性 vs. 说明性）×2（逆反倾向：高 vs. 低）的组间实验，量化考察不同信息框架的提醒视频对大学生用户心理逆反状态及过度使用警觉性的具体影响；随后引入隐喻抽取技术（Zaltman Metaphor Elicitation Technique, ZMET），深入挖掘量化数据背后的心理图式与认知机制。本研究旨在突破传统劝服效果研究的线性框架，系统揭示算法父爱主义治理下的微观心理博弈机制，为破解数字时代的短视频防成瘾困境、构建更具人文伦理的算法治理体系提供学理支撑与实证参照。

一、文献综述

（一）从“硬性阻断”到“算法父爱主义”：短视频防成瘾机制的逻辑演变

短视频成瘾被定义为由成瘾性媒介使用导致的强迫性行为，其特征不在于用户对短视频的过度依赖、时间知觉扭曲以及戒断反应^[6]。面对这一现象，平台治理策略逐渐从单纯的硬性阻断转向一种系统性的算法父爱主义。算法父爱主义是指数字平台利用数据霸权与算法架构，在用户非完全知情或被动接受的情况下，代替用户做出对其有利的决策或引导^[7]。作为一种新兴的数字治理范式，算法父爱主义的理论合法性确立于其对“技术家长主义”与“算法支配”等概念的辩证综合与范式超越。干预动机上，

它继承了技术家长主义的伦理预设，认为平台有责任替非理性用户做出符合其长远福祉的选择^[8]。但与技术家长主义依赖机械性与显在性的系统架构不同，算法父爱主义试图通过动态的、个性化的干预来软化冲突；在干预手段上，内化了算法支配中的权力结构特征^[9]，即依赖算法技术对用户注意力进行全景式的数据追踪与黑箱化的计算监控。因此，算法父爱主义本质上是“家长制动机”与“支配性权力”的隐蔽结合。这种结合并非简单的要素叠加，而是底层规训机制的升级，利用算法支配的“认知不透明性”，将规训权力隐匿于高度个性化、实时更新的信息流推送之中。

然而，这种运作机制的不透明性极易制造新的信任危机^{[10][11]}。抖音“健康使用计划”通过隐形的计时器监控用户的行为，并试图通过无法关闭的软性助推，即休息提醒视频的插入来矫正用户的短视频成瘾行为。但用户无法确知休息提醒的触发究竟是基于健康考量还是商业流量调控，这种信息不对称极易滋生算法猜忌。同时，算法隐含了技术优越论的傲慢，即假设其比用户更了解身体极限与时间价值，这种对主体性的隐性剥夺，构成了用户心理防御的潜在诱因。因此，探讨抖音健康使用计划的有效性，须将其置于人机权力博弈视野下，考察用户如何感知并回应这种来自机器的关怀。

（二）劝服与逆反的边界：限制性、说明性信息的干预效应

抖音“健康使用计划”是通过向用户推送含有具体劝服内容的休息提醒视频来实现的。这意味着该计划效果不仅取决于算法架构层面的父爱主义属性，也受制于劝服信息本身的框架策略。在健康传播领域，信息框架如何影响受众的态度与行为，已有成熟的理论积累。基于瓦尔肯堡（Patti M. Valkenburg）等在亲子媒介干预研究中建立的经典分类体系，劝服信息框架可分为限制性与说明性两种类型^[12]。

心理逆反理论与展望理论为解释两类信息的效果差异提供了理论视角^[13]。布雷姆（Jack W. Brehm）指出，当个体感知到行为自由受到



威胁时，会激发旨在恢复自由的动机性状态，即心理逆反^[14]。这种状态常伴随负面认知与情绪唤醒，反而增强被禁行为的吸引力。应用于本研究情境，休息提醒视频作为一种外源性行为的劝服信息，具备触发用户逆反反应的可能。展望理论从决策心理角度补充指出，信息框架显著影响个体风险偏好与行为选择^[15]。限制性信息由于使用了高强度的控制性语言，极易被解码为对个人自主权的直接侵犯，从而诱发强烈的心理逆反^[16]。说明性信息则通过提供选择权和解释理由来降低对自由的威胁感知，可能在一定程度上减少逆反心理的触发^[17]。在媒介使用的劝服情境中，明确设定使用时长或行为边界的限制性信息会增强心理逆反，而陈述过度使用负面后果的说明性信息因诉诸认知反思而非服从控制，在一定程度上可减少心理逆反程度^[18]。

然而，上述结论多建立在人际传播（家长—子女、医生—患者）的语境之上。在人机交互的特殊语境下，劝服源从具有社会情感联系的“人”变为不透明的“算法”时，限制性与说明性信息的效力边界是否会发生漂移还不得而知。鉴于既有研究结论，本研究提出假设：

H1：相较于接触说明性休息提醒视频，大学生用户在接触限制性休息提醒视频时，将表现出更高层次的心理逆反。

（三）个体逆反倾向与心理逆反的交互机制

心理逆反理论的早期研究多将逆反视为一种纯粹的情境性状态，认为逆反反应完全取决于外部刺激的强度。随着研究的深入，布雷姆等进一步修正了理论模型，指出逆反反应不仅受情境因素调节，更存在显著的个体差异。这一发现促使学界开始关注逆反的特质维度，即个体逆反倾向^[19]。迪拉德（James Dillard）和沈立江（Lijiang Shen）系统性地提出了心理逆反的交织模型，指出个体逆反倾向是预测心理逆反强度的关键前因变量^[20]。

既有文献进一步描绘了高逆反倾向者的典型心理与行为画像。这类个体通常表现出对自主权的高度需求，对外部强加的约束和规则表

现出更高的敏感性与抵抗倾向^[21]。更为关键的是，特质逆反可以与劝服信息的属性发生交互作用，共同调节逆反心理的生成幅度^[22]。迪拉德和沈立江在牙线使用的劝服研究中发现，高控制性的语言对于高逆反倾向者而言，会引发不成比例的强自由威胁感与逆反反应，而对于低逆反倾向者的影响则相对较小^[23]。这一发现提示，在考察休息提醒视频的劝服效果时，不能仅关注信息框架的主效应，也须深入探究受众的特质差异如何调节其对不同信息类型的解码过程。基于上述实证积累，本研究提出如下假设：

H2：接触休息提醒视频后，逆反倾向高的大学生用户表现出的心理逆反程度高于逆反倾向低的用户。

H3：休息提醒视频的信息类型（限制性 vs. 说明性）与大学生用户的逆反倾向（高 vs. 低）对心理逆反程度存在交互效应。

（四）心理逆反、逆反倾向与过度使用短视频的警觉性变化

在器官捐赠、疫苗接种及限制饮酒相关的实证研究中已证明，当心理逆反在特定情境中被触发，为应对自由威胁，个体会采取直接或间接的行为策略以重获自主感知，包括更频繁地从事被劝阻的行为，或明确拒绝信息所推荐的行为模式^{[24][25]}。因此，当旨在提升警觉性的休息提醒视频引发心理逆反时，可能会削弱用户原有的警觉水平，甚至导致与劝服目标背道而驰的效果，出现“回旋镖效应”^[26]。

然而，既有研究大多未能系统建立个体在劝服信息接触前的态度基线，因而难以准确捕捉因逆反所导致的态度变化动态（即幅度与方向），制约了对劝服失败机制的深入理解。态度作为一种具有评价性特性的心理准备状态，其变化需通过外显评价在时间序列上的比较得以测量^[27]。心理逆反在此过程中可充当一种动机性变量，强化原有态度或驱动其向劝服相反方向转变^[28]。因此，采用前测—后测设计以量化态度变化，成为识别逆反效应、揭示劝服失败机制的必要途径。基于上述理论推演，本研究以心理逆反为中介变量，考察休息提醒视频

对大学生用户短视频过度使用警觉性的影响，据此提出假设：

H4：接触休息提醒视频后，大学生用户的心理逆反程度与其短视频过度使用警觉性的变化呈负相关。

逆反倾向对劝服效果的影响通常被建模为一条经由心理逆反状态传导的间接路径^[29]。但这一结果是在相关研究将逆反倾向作为单一变量纳入模型，未分离各子维度差异性贡献的情况下所得。而逆反倾向的多维结构意味着不同子维度对劝服效果的预测机制可能不同。这一

推测与临床治疗领域的发现相呼应，德拉斯奎瓦斯等（Carlos De las Cuevas）在精神科门诊患者群体中发现，逆反倾向中的规则逆反维度直接预测了治疗依从性的下降^[30]。换言之，当劝服干预的核心特征恰好契合逆反倾向某一特定子维度的敏感领域时，该子维度可能绕过即时状态逆反环节，直接对态度结果产生预测效应。据此，本研究提出假设：

H5：接触休息提醒视频后，大学生用户的个体逆反倾向与其短视频过度使用警觉性的变化呈负相关。

表1 使用管理助手“休息提醒功能”与健康使用计划“休息提醒视频”对比

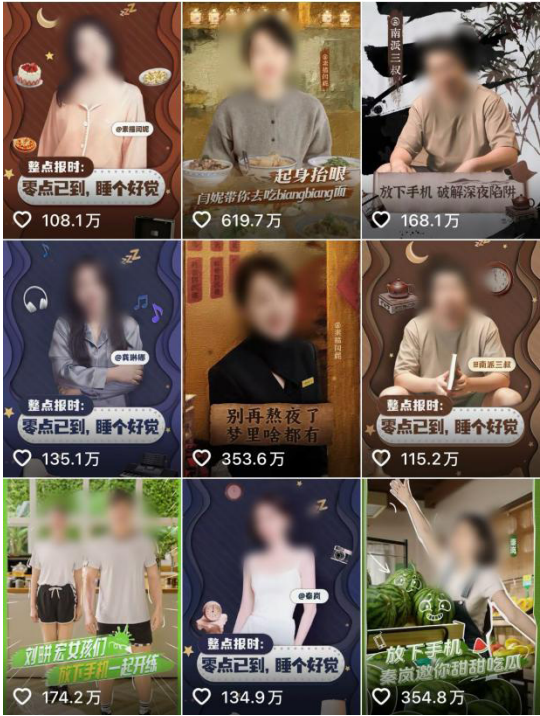
	休息提醒功能	休息提醒视频
来源	抖音短视频防成瘾实践	抖音短视频防成瘾实践
启用	默认开启，用户可自主选择是否关闭	面向所有用户，利用算法在特定时间内推送给已经长时间使用抖音的用户，用户无法自主关闭
推荐间隔	提醒间隔时长默认为“智能提醒”，用户可自主设置提醒间隔时长	算法认定该用户已长时间使用时，多在整点、零点等时间点推送
方式	闹钟式时间悬停	视频内容形式的劝服信息
要素	下滑视频会出现特定提醒视频2秒，视频会缩小上移	带有主体（网络名人、演员歌手等）、主题（运动、健康小技巧、心灵抚慰等）、剧情（放下手机、休息一会儿）的各类劝服信息 
机制	时钟逻辑悬停数字媒介时间	通过劝服信息提醒激发用户对于过度使用的警觉性

表3 实验受试者的性别、年级分布以及平均年龄

		全体样本 (N=366)	实验组 1 (n=122)	实验组 2 (n=122)	实验组 3 (n=122)
性别	男性	187 (51.1%)	62 (50.8%)	62 (50.8%)	63 (51.6%)
	女性	179 (48.9%)	60 (49.2%)	60 (49.2%)	59 (48.4%)
平均年龄 (SD)		21.38 (2.68)	21.64 (2.47)	21.19 (2.62)	21.31 (2.95)
年级	大一	79 (21.6%)	27 (22.2%)	26 (21.3%)	26 (21.3%)
	大二	132 (36.1%)	43 (35.2%)	44 (36.1%)	45 (36.9%)
	大三	75 (20.5%)	25 (20.5%)	25 (20.5%)	25 (20.5%)
	大四	80 (21.9%)	27 (22.1%)	27 (22.1%)	26 (21.3%)

注：实验组 1= 暴露于限制性提醒视频的受试者群体，实验组 2= 暴露于说明性提醒视频的受试者群体，实验组 3= 控制组受试者群体

实验分别在 2024 年 9 月 24 日、10 月 11 日以及 2025 年 3 月 17 日、3 月 20 日进行，共招募 366 名受试者，年龄范围为 18 至 23 岁。实验所有环节均在线下控制环境内完成，受试者按性别与年级进行分层，随机分配至不同实验组别，分别在多媒体教室中观看相应休息提醒视频类型的刺激物并完成问卷填写。受试者平均年龄为 21.38 岁 (SD = 2.77)，具体人口学特征分布见表 3。

(四) 实验过程

实验首先收集受试者的人口统计学信息，完成有关短视频态度、逆反倾向及过度使用警觉性 (前测) 的测量，以便与实验后数据进行比较。其次，受试者观看所属组别对应的刺激材料。观看结束后，受试者回答操作检查问题，用于检验视频信息类型的识别准确性，并填写心理逆反量表及过度使用短视频警觉性量表 (后测)。最后，受试者报告其近期短视频使用时间。全部任务完成后，实验者进行简报说明，告知受试者本次所观看视频内容可能与实际平台防成瘾机制存在差异，以避免其对相关功能产生误解。

(五) 变量测量

1. 逆反倾向

本研究采用 Hong 氏逆反心理量表测量受试者的逆反倾向 (Hong Psychological Reactance Scale, HPRS)^[31]。该量表共 11 个题项，包括“规则会引发我的抵触情绪”以及“当我的自由选择权受到限制时我会生气”等。采用李克特 5 分量表，包含规则逆反和拒绝自由威胁两

个二级变量。规则逆反量表信度 Cronbach’s alpha 为 0.79，KMO 数值为 0.86， $p < 0.001$ ，验证性因子分析结果为 $X^2/df = 2.073$ ， $CFI = 0.965$ ， $TLI = 0.951$ ， $RMSEA = 0.054$ ；拒绝自由威胁量表信度 Cronbach’s alpha 为 0.84，KMO 数值为 0.89， $p < 0.001$ ，验证性因子分析结果为 $X^2/df = 2.154$ ， $CFI = 0.951$ ， $TLI = 0.938$ 以及 $RMSEA = 0.048$ 。

2. 对休息提醒视频的心理逆反

本研究采用林赛 (Lisa L. M. Lindsey) 基于 HPRS 量表所制定的个体对传播领域特定信息的心理逆反量表^[32]，具体题项包括“这条信息的内容会给我带来不便”或“因为这个信息内容而生气”等情境化问题。本研究修订测量题目，以确保其适用于评估大学生用户对休息提醒视频的心理逆反。采用李克特 5 分量表，量表信度 Cronbach’s alpha=0.90，KMO 数值为 0.92， $p < .001$ ，验证性因子分析结果为 $X^2/df = 1.832$ ， $CFI = 0.974$ ， $TLI = 0.962$ 以及 $RMSEA = 0.052$ 。

3. 对过度使用短视频的警觉性

本研究采用游戏行为综合评定量表中的问题游戏使用诊断量表作为参照^[33]，对过度使用短视频的警觉性进行测量。问题游戏使用诊断量表以中毒综合征模型为理论基础，并结合美国精神医学会的分类体系 (DSM-5) 和世界卫生组织 (WHO) 的 I-10 标准制定。本研究将这一诊断标准在过度使用短视频警觉性的情境下进行修订，具体题项包括“为了获得满足感，需要比以前刷短视频的时间更长的话，就要提高警觉”“虽然多次努力减少刷短视频的时间，但每次都失败的话，就要提高警觉”等。采用李克特 5 分量表，量表信度 Cronbach’s alpha =0.93，



KMO 数值为 0.94, $p < 0.001$, 验证性因子分析结果为 $X^2/df = 1.653$, CFI= 0.985, TLI= 0.974, RMSEA=0.047。

4. 控制变量

本研究将用户的性别、短视频态度、短视频使用时间与频率作为控制变量。对于短视频态度, 本研究通过“刷手机短视频对我来说是积极的 - 消极的”等 4 个题目进行测度, 采用李克特 5 分量表, 量表信度 Cronbach's $\alpha = 0.85$, KMO 数值为 0.90, $p < 0.001$, 验证性因子分析结果为 $X^2/df = 2.024$, CFI= 0.947, TLI= 0.935, RMSEA=0.059。短视频使用时间与频率则通过自我报告的方式进行收集, 包括过去一周内平均每天的短视频使用时间及频率, 以李克特 5 分量表测定。

三、研究分析

(一) 实验刺激物验证

为检验实验刺激物是否符合限制性信息与说明性信息的分类要求, 本研究进行了操纵检验。被试者就以下两个题项进行评价: (1) 视频中包含多少限制或控制您使用短视频时间的元素? (2) 视频中包含多少关于说明调整短视频使用时间必要性的内容? (1 = 完全没有, 5 = 非常多)。结果显示, 限制性视频组在题项 1 的评分 ($M = 4.57$) 显著高于说明性视频组 ($M = 3.14$, $t = 8.72$, $p < .001$, Cohen's $d = 1.10$, $r = 0.48$); 说明性视频组在题项 2 的评分 ($M = 4.58$) 显著高于限制性视频组 ($M = 3.21$, $t = 9.15$, $p < .001$, Cohen's $d = 1.16$, $r = 0.50$); 控制组在两个题项上的评分 ($M = 1.85$; $M = 1.82$) 均显著低于实

验组。此外, 被试者的逆反倾向与上述评分无显著相关, 表明其对视频信息类型的感知未受个体特质干扰。以上结果有效支持了实验刺激物分类的准确性。

(二) 基础分析

本研究首先检验实验组在核心变量上的同质性。结果显示, 各组在过度使用警觉性(前测)、短视频使用频率与使用时长方面均无显著差异, 满足组间可比性要求。唯有短视频使用态度在组间存在差异 ($F = 4.32$, $p < 0.05$), 限制组的态度得分 ($M = 3.88$) 显著低于说明组 ($M = 4.62$) 与控制组 ($M = 4.59$), 故后续分析将其与性别(与因变量显著相关, Cramer's $V = 0.12$, $p < 0.05$) 一并纳入协变量控制。其次, 在主变量的分布检验中, 各变量偏度 ($-1.14 \sim 1.02$) 与峰度 ($-1.24 \sim 1.34$) 均处于可接受范围。表 4 呈现了各变量的均值、标准差及相关系数矩阵, 初步支持了理论预期。

(三) 数据分析

1. 视频信息类型与逆反倾向对心理逆反的影响

为检验假设 1、2、3, 本研究以心理逆反为因变量, 视频信息类型与逆反倾向为自变量进行协方差分析。由于两个逆反倾向维度可能存在共变, 在检验其中一维度时, 将另一维度作为协变

表 4 主要变量的描述性统计与相关分析结果

		1	2	3	4	5	6	7	8
1. 短视频使用频率		1							
2. 每日平均使用时间		0.43**	1						
3. 短视频态度		0.41**	0.28**	1					
4. 对过度使用短视频的警觉性（前）		-0.05	-0.10	-0.18**	1				
5. 逆反倾向（规则逆反）		0.02	0.10	0.07	-0.14*	1			
6. 逆反倾向（拒绝自由威胁）		0.03	0.06	-0.06	0.12*	0.36**	1		
7. 对休息提醒视频的心理逆反		0.01	0.07	-0.03	-0.21**	0.22**	0.04	1	
8. 对过度使用短视频的警觉性（后）		-0.07	-0.11	-0.15**	0.67***	-0.12*	0.09	-0.25**	1
样本	平均值	3.34	2.08	4.46	4.28	2.93	3.92	2.95	3.82
	（SD）	1.42	1.18	0.92	1.23	1.29	1.09	1.34	1.24

注: * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

量加以控制，Levene 检验结果表明数据满足方差齐性假设。

表 5 呈现了以视频信息类型和规则逆反倾向为自变量的 ANCOVA 结果。其中，信息类型主效应显著 ($F = 21.11, p < .001, \eta^2 = 0.09$)，规则逆反倾向主效应同样显著 ($F = 16.02, p < .001, \eta^2 = 0.10$)，而二者交互效应未达显著 ($F = 2.85, p = 0.068, \eta^2 = 0.06$)。在控制变量中，仅性别显著 ($F = 8.15, p < 0.01, \eta^2 = 0.04$)；态度、使用频率、使用时长及拒绝自由威胁倾向均不显著。

为进一步揭示不同类型提醒视频的影响差异，本研究采用边际均值估计并进行事后比较（表 6）。结果显示限制性提醒视频组的心理逆反程度 ($M = 3.91$) 高于说明性提醒视频组 ($M = 2.58$)，且成对比较结果也进一步证实两组间的差异在统计学上的意义显著 ($MD = 1.33, p < 0.001$)。

为进一步验证规则逆反倾向的效应，本研究依据总样本均值 ($M = 2.93$) 区分高规则逆反组 ($n = 189, M = 3.97$) 与低规则逆反组 ($n = 177, M = 1.81$)。如图 1 所示，两组在不同信息类型下均呈现一致模式，即限制性视频引发的心理逆反程度最高。结合协方差分析可知，尽管均值趋势显示高规则逆反者在限制性视频下反应更

表 5 休息提醒视频信息类型与规则逆反作为自变量的 ANCOVA 结果

变量		Type III SS	df	F	P	η^2
协变量	性别	18.13	1	8.15	< 0.01	0.04
	短视频使用态度	0.23	1	1.25	0.688	0.00
	短视频使用频率	0.01	1	0.02	0.887	0.00
	短视频每日平均使用时间	1.78	1	0.01	0.265	0.00
	拒绝自由威胁	0.15	1	0.16	0.745	0.00
自变量	规则逆反	50.84	2	16.02	< 0.001	0.10
	休息提醒视频的信息类型	45.65	1	21.11	< 0.001	0.09
交互变量	休息提醒视频的信息类型 X 规则逆反	19.50	2	2.85	0.068	0.06
因变量 大学生用户对休息提醒视频的心理逆反						

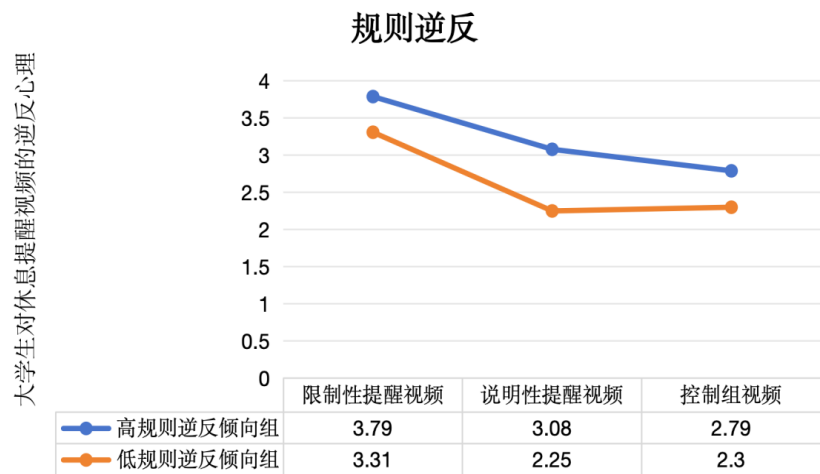


图 1 休息提醒视频信息类型与规则逆反倾向对大学生用户心理逆反的影响

表 6 估算值以及成对比较的结果

组别		平均值	标准误差	置信区间	
				下限	上限
实验组 1		3.91	0.17	3.58	4.24
实验组 2		2.58	0.17	2.25	2.91
实验组 3		2.53	0.16	2.22	2.84
(I) 组别	(J) 组别	平均值差值 (I-J)	标准误差	显著性	
实验组 1	实验组 2	1.33	0.22	< .001	0.90 1.76
实验组 2	实验组 3	0.05	0.25	0.842	-0.44 0.54
实验组 3	实验组 1	-1.38	0.23	< 0.001	-1.83 -0.93
因变量 大学生用户对休息提醒视频的心理逆反					



强烈，但交互作用未达显著水平，说明这种差异不足以排除随机波动。因此，可以认定个体规则逆反水平未显著调节视频信息类型与心理逆反之间的关系。

表 7 呈现了以视频信息类型和拒绝自由威胁倾向为自变量的 ANCOVA 结果。其中，信息类型的主效应显著（ $F = 12.61$ ， $p < 0.001$ ， $\eta^2 = 0.09$ ），拒绝自由威胁倾向的主效应不显著（ $F = 0.75$ ， $p = 0.478$ ， $\eta^2 = 0.00$ ），信息类型与拒绝自由威胁倾向的交互效应不显著（ $F = 2.50$ ， $p = 0.090$ ， $\eta^2 = 0.02$ ）。在控制变量中，仅性别显著（ $F = 8.15$ ， $p < 0.01$ ， $\eta^2 = 0.04$ ），短视频使用态度、使用频率、每日平均使用时间及规则逆反倾向的效应均不显著。综合两次 ANCOVA 分析结果可知，研究假设 1 成立，研究假设 2 部分成立，研究假设 3 不成立。

2. 大学生用户心理逆反对其过度使用警觉性变化的影响

为检验假设 4，本研究在排除共同方法偏差（Harman 单因子解释率 23.27%）及多重共线性

（ $VIF < 5$ ， $DW = 1.94$ ）的基础上，以过度使用短视频的警觉性（后测）为因变量进行多阶段回归分析。分析结果显示（表 8），在模型 1 中，心理逆反对警觉性（后测）具有显著负向预测作用（ $\beta = -0.17$ ， $p < .05$ ）。但在模型 2 中，控

表 7 休息提醒视频信息类型与拒绝自由威胁作为自变量的 ANCOVA 结果

变量		Type III SS	df	F	P	η^2
协变量	性别	18.13	1	8.15	<0.01	0.04
	短视频使用态度	0.23	1	1.25	0.688	0.00
	短视频使用频率	0.01	1	0.02	0.887	0.00
	短视频每日平均使用时间	1.78	1	0.01	0.265	0.00
	规则逆反	0.15	1	0.16	0.745	0.00
自变量	拒绝自由威胁	1.84	2	0.75	0.478	0.01
	休息提醒视频的信息类型	45.65	1	12.61	<0.001	0.09
交互变量	休息提醒视频的信息类型 X 拒绝自由威胁	7.50	2	2.50	0.090	0.02
因变量 大学生用户对休息提醒视频的心理逆反						

表 8 心理逆反对短视频过度使用警觉性（后测）影响的多阶段回归分析结果

变量	过度使用短视频的警觉性（后测）							
	模型 1				模型 2			
	β	SE	t	95% CI	β	SE	t	95% CI
心理逆反	-0.17	0.11	-2.16*	-0.46~-0.02	-0.08	0.11	-0.84	-0.28~0.12
过度使用短视频的警觉性（前测）					0.68	0.07	8.17***	0.52~0.84
R^2 (adjusted R^2) F(df)	0.12(0.10)* F(1,348) = 9.22*				0.50(0.48)*** F(2,347) = 45.37***			
ΔR^2	0.12*				0.38***			

注：* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

表 9 心理逆反对警觉性前测 - 后测变化的对应样本 t 检验

分组	均值 M(SD)		显著性检验			效应量	
	对短视频过度使用的警觉性水平（前）	对短视频过度使用的警觉性水平（后）	t	df	p	Cohen's d	r
高心理逆反组 n=179	4.22(0.91)	3.85(0.88)	3.27**	171	0.001	0.35	0.24
低心理逆反组 n=187	4.35(0.87)	4.12(0.85)	2.64**	176	0.009	0.29	0.20

注：* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

制警觉性（前测）后，心理逆反效应不再显著（ $\beta = -0.08$ ， $p = 0.537$ ），而实验前警觉性对实验后警觉性具有显著正向预测作用（ $\beta = 0.68$ ， $p < 0.001$ ）。模型解释力由 12% 提升至 50%（ $\Delta R^2 = 0.38$ ， $p < 0.001$ ）。这表明警觉性后测水平主

要由其前测水平决定，心理逆反并不具有独立增量效应。

为进一步考察警觉性变化是否受心理逆反影响，本研究以心理逆反总均值（ $M = 2.95$ ）为界，将被试分为高、低两组，并分别对实验前后警觉性进行配对样本 t 检验。结果显示（表 9），不论心理逆反水平高低，被试在观看休息提醒视频后警觉性均呈下降趋势，整体样本亦呈现一致变化，进一步支持心理逆反并非警觉性变化的主要影响因素，研究假设 4 不成立。

3. 事后检验

为检验假设 5，并验证信息类型对警觉性变化的潜在影响，本研究进行了事后检验。首先，为排除休息提醒视频信息类型对过度使用短视频的警觉性变化的影响，本研究进行了配对样本 t 检验，结果显示三组均出现事后警觉性显著下降，组间并未

显示出由信息类型导致的有意义的不同，即信息类型本身并未调节警觉性变化方向（如图 2 所示，限制性组：前测 $M = 4.24$ ，后测 $M = 3.82$ 。 $t = 6.01$ ， $p < .001$ ，Cohen's $d = 0.57$ ， $r = 0.49$ ；说明性组：前测 $M = 4.15$ ，后测 $M = 3.75$ 。 $t = 4.03$ ， $p < .001$ ，Cohen's $d = 0.38$ ， $r = 0.36$ ；控制组：前测 $M = 4.28$ ，后测 $M = 3.89$ ， $t = 3.35$ ， $p = .001$ ，Cohen's $d = 0.26$ ， $r = 0.25$ ）。

其次，为检验个体逆反倾向对警觉性变化的影响，在排除共同方法偏差（Harman 单因子解释率 22.48%）及多重共线性（ $VIF < 5$ ， $DW = 1.78$ ）的基础上进行多阶段回归分析。结果显示（表 10），模型 1 中规则逆反可显著负向预

表 10 逆反倾向对短视频过度使用警觉性（后测）影响的多阶段回归分析结果

变量	过度使用短视频的警觉性（后测）							
	模型 1				模型 2			
	β	SE	t	95% CI	β	SE	t	95% CI
规则逆反	-0.23	0.07	-3.24**	-0.37~-0.18	-0.21	0.07	-2.84**	-0.42~-0.21
拒绝自由威胁	-0.07	0.09	-0.74	-2.48~1.12	-0.09	0.07	-0.68	-2.51~1.15
过度使用短视频的警觉性(前测)					0.47	0.08	6.21***	0.52~0.84
R^2 (adjusted R^2) F(df)	0.14(0.13)** F(1,348) = 11.31**				0.53 (0.52)*** F(2,347) = 47.23***			
ΔR^2	0.14**				0.39***			

注：* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

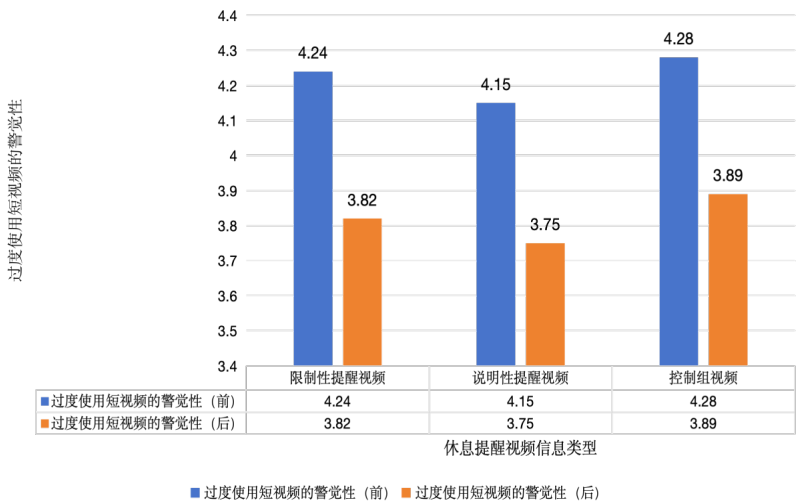


图 2 休息提醒视频信息类型的分组样本 t 检验结果

测警觉性后测（ $\beta = -0.23$ ， $p < 0.01$ ）；在控制前测警觉性的模型 2 中，规则逆反的效应仍然显著（ $\beta = -0.21$ ， $p < 0.01$ ），表明其具有独立于基线水平的预测力。

为仔细验证这一结果，本研究对之前区分的高规则逆反组、低规则逆反组进行了对应样本 t 检验。结果显示，高规则逆反组在暴露于实验刺激后，对过度使用短视频的警觉性降低（前 $M = 4.33$ ，后 $M = 3.64$ ， $t = 3.60$ ， $p < .001$ ，Cohen's $d = 0.27$ ， $r = 0.18$ ），而低规则逆反组的变化不显著（前 $M = 4.18$ ，后 $M = 3.95$ ， $p = 0.244$ ），即高规则逆反不仅增加了对休息提醒视频信息的心理逆反，而且降低了过度使用短

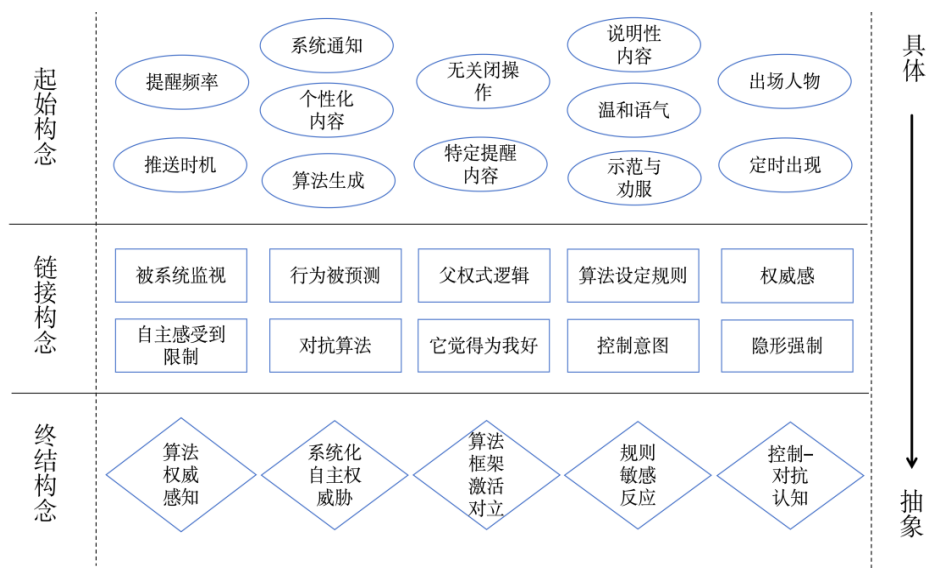


图3 算法驱动情境下大学生用户对休息提醒视频感知的共识地图

视频的警觉性。由此可知，真正驱动警觉性下降的是逆反倾向中“规则逆反”的个体差异，研究假设5部分成立。

（四）发生机制分析

为深入阐释量化结果，本研究采用隐喻抽取技术对15名深度使用抖音的大学生用户进行了半结构化访谈。访谈采用攀梯式追问，帮助受访者逐步拆解其经历体验。访谈于2024年11月至2025年4月间进行，每位受访者耗时约45-60分钟，重点围绕其在实际使用场景中接触休息提醒视频的体验、认知与情感反应展开叙事性回忆。

访谈后整理受访者的叙述文本，运用ROST CM6软件的词频分析功能获取访谈文本的词频表，对其中无法成为构念的词语进行剔除，同时对基本涵义重叠的构念进行合并，再回到原始文本中进行匹配，获取每位受访者提出的构念。随后根据ZMET研究原则^[34]，将被采访者中1/3的人提到的构念提取为共同构念，共同构念中1/4受访者提及的构念可认为与其他构念之间存在逻辑关系。本研究共抽取出27个共同构念，在10位受访者的个人构念平均涵盖率为63.3%，基本符合构念涵盖率的要求。最后，本研究以2位具有博士学位的编码者进行人工分类

的方式，对各共同构念之间的逻辑关系进行推导，逐步从表层感知提炼至深层认知，以此构建受访者的构念聚类，得到起始构念12个、链接构念10个、终结构念5个，构建出如图3所示的共识地图。

共识地图显示，用户对休息提醒视频的认知经历了一个从具

象感知到抽象诠释的多层递进过程。起始构念集中于算法干预的形式特征与系统属性，主要包括“推送时机”“个性化内容”“无关闭操作”及“定时出现机制”等。这些要素共同构成算法干预的外在可感知层，体现出其高度情境化与强制性特点。

链接构念则反映了用户对算法意图的推断与自主性感知，大学生用户在接收算法推送的休息提醒视频的体验中表示“感受到自己的短视频使用行为是被系统所监视的”（S01）、“行为被预测和算法所干预”（S13）等，具体可凝练为“被系统监视”“行为被预测”“父权式逻辑，即算法自认为优于用户进行决策”“隐性强制”与“控制意图”。高规则逆反特质个体在此过程中表现出显著的认知敏感性与解释偏向，能够迅速从看似中性的说明性信息中析出规则设定与外部约束成分（S07, S15）。即便视频采用温和语气或理性论述，用户仍将其识别为“算法设定的规则”，从而触发“对抗算法”的心理动机（S12）。

终结构念标志着用户形成对算法干预的深层认知结构，可归纳为“算法权威感知”“系统化自主权威胁”“算法框架激活对立”“规则敏感反应”与“控制-对抗认知”。其中，

算法权威感知指用户将算法视为具有支配性与非对称控制能力的行动主体；系统化自主权威胁揭示算法干预对用户自我决定动机的结构性压制；控制—对抗认知框架则反映用户形成“平台 vs 用户”的对立式心智模型。

四、结论与讨论

（一）算法父爱主义中的隐性限制意图与心理逆反的触发机制

本研究通过对照实验揭示，限制性休息提醒视频更易触发心理逆反，印证了心理逆反理论关于“控制性信息即自由威胁”的核心假设^[35]。访谈数据进一步揭示，用户的抵抗并非单纯源自内容呈现方式，更来自对算法作为规则制定者与权威执行者的感知。这触及了平台治理的核心悖论，以“健康使用”为名的善意干预为何引发系统性的心理对抗？

要解答这一问题，须跳出单一的信息说服框架，审视其背后的算法父爱主义治理逻辑。前文指出，算法父爱主义是“家长制动机”与“支配性权力”的结合体。本研究的 ZMET 数据深刻揭示了这一理论预设为用户认知体验中的内在割裂。具体而言，在 ZMET 共识地图中，起始构念的“推送时机”与“个性化内容”，在链接构念中直接转化为用户的监控焦虑——“感受到自己的短视频使用行为是被系统所监视的”，并最终呈现出“算法权威感知”与“系统化自主权威胁”等终结构念。这表明，平台试图通过“超级助推”技术^[36]来包装家长制动机的努力走向了反面。受访者明确解码出了隐藏其后的父权式逻辑，敏锐察觉到“算法自认为优于用户进行决策”，并将其体验为“行为被预测和算法所干预”。这种将温和建议解码为系统性控制信号的过程，揭示了算法支配手段对家长制动机的实质性吞噬。

这一发现不仅深化了算法父爱主义的理论内涵，更拓展了心理逆反理论的解释边界。在传统人际劝服中，心理逆反主要源于信息内容的控制性词汇；而在算法父爱主义语境下，逆反的触发源发生了结构性前移——从信息内容层

转移到了算法架构层。抖音算法推送的休息提醒视频中，由支配性技术手段与单向度决策交织而成的隐性限制意图，使得其无论被包装为限制性抑或说明性，底层框架均被识读为对行为自主权的隐蔽规训与代持。因此，即便是柔和的说明性信息，只要置于算法父爱主义的全景监控与权力不对称框架下，便足以激活自由威胁感知，触发认知反驳与情绪对抗，信息类型的区分效力在算法架构面前被大幅削弱，真正驱动用户差异化反应的变量已发生改变。

（二）规则逆反特质是驱动警觉性下降的稳定人格基础

本研究发现，规则逆反倾向对心理逆反具有稳定预测作用，且不因信息类型而增减，表明其具有跨情境稳定性。结合认知心理学中的说服知识模型剖析，这种跨情境的稳定性揭示了规则逆反在算法语境下已演化为一种针对架构层约束的认知图式^[37]。ZMET 共识地图显示，高规则逆反者能够迅速锁定“无法取消”“定时推送”等约束逻辑。这种规则敏感性并非单纯的情绪化对抗，而是高算法素养在劝服情境下的应激转化^[38]，其调用应对策略的说服知识已从传统的“文本识读”跃升为“架构识读”。他们能够穿透休息提醒视频表层的话语修辞，将算法系统底层的不透明监控与强制打断，解码为具有侵犯性的说服企图。

然而与一般预期不同，本研究量化结果显示，状态性心理逆反的强度并未对警觉性变化产生显著影响，驱动警觉性下降的是规则逆反这一特质性变量。这意味着警觉性的变化并非由即时情境中被激发的逆反情绪决定，而是由稳定的人格倾向塑造。对此可从两个层面加以理解。首先，状态性心理逆反之所以未能显著作用于警觉性，可能在于其效应被逆反倾向的特质性影响所吸收。在高度不对称的算法权力空间中，个体对系统底层规训规则的特质性警惕，在认知加工序列中先于对具体信息内容的即时反应，占据了对劝服效果的主要解释份额。高规则逆反者在接触任何形式的干预之前，已经预置了一套针对规则性约束的抵抗图式，这



一图式直接塑造了其对于警觉性信息的态度取向，使得即时逆反状态的边际贡献趋于微弱。

其次需要解释的问题是，为何规则逆反特质驱动的是警觉性的下降而非上升？从自我损耗理论来看^[39]，高规则逆反者对算法规则的持续性识别和批判，本身构成了一种高耗能的认知活动，大量消耗了执行控制资源。ZMET发现的“控制—对抗认知”框架进一步揭示了这一资源耗竭的后续走向，当用户感知到算法权力是一种难以实质性逃避的系统力量时，其心理防御便从“主动对抗”转向“责任转移”与“认知卸载”^[40]。既然算法已经接管了“何时该休息”的判断权，用户潜意识中便将自我监控的责任让渡给了系统。此时，“对抗”不再体现为关掉短视频，而是体现为一种带有“逆反式依赖”的心理卸载——“既然你（算法）在监控我，那我不需要再监控我自己”，从而放弃或削弱了自我警觉的资源投入。

既往研究多将规则逆反视为阻碍健康传播的负面特质，但本研究倾向于将其重新诠释为个体在算法社会中的一种数字免疫机制。高规则逆反者凭借其算法素养，识破了平台在“注意力商品化”与“道德表演”之间的结构性伪善，其抵抗的是“被计算、被安排”状态本身。然而，这种抵抗在高度不对称的技术架构中难以转化为有效的行为改变。由于缺乏真正的退出机制，用户的认知抵抗在耗尽后，只能以认知卸载的形式走向消极妥协——即加剧自身的损耗与警觉性的全面退化。

（三）短视频防成瘾需重构算法劝服路径

本研究所揭示的警觉性下降及其背后的认知卸载机制表明，算法干预触发了深层的“框架激活悖论”。无论信息披着限制抑或说明的外衣，只要其内核仍是对于短视频使用行为的管理，便会激活用户与“控制”“约束”相关的认知图式，仅在表层优化劝服话语无法解决自由困境，反而会加剧用户的自动化自满与主体性退化。要化解这一困境，必须推动数字福祉治理范式从“阻断式父爱主义”转向“赋能式自由主义”，对算法劝服路径进行伦理与设计的双重重构。

首先，应解开算法黑箱，主动披露休息提醒视频的触发逻辑与数据依据，缓解用户的算法猜忌。其次，将控制感交还用户，阻断认知卸载，允许用户自定义干预的强度、风格与时机，增强用户对干预规则的感知控制力。最后，以替代性满足引导注意力迁移。有效的干预不应止步于“打断”，而应致力于“连接”。算法可基于情境推荐线下的、具身性的活动选项，并运用增益框架替代传统的损失框架，从而将用户的注意力从抵抗约束转向拥抱更丰富的可能性。

（四）启示与不足

本研究存在以下不足。首先，为保证线下实验的可行性与真实性，样本来源于高校大学生，虽平衡了人口学变量，但样本在受教育水平与媒介接触习惯上较为同质，外推时需谨慎，未来研究可引入跨阶层、跨代际比较思维，以考察不同群体面对算法干预时逆反机制的结构性分化。其次，实验环境与自然使用情境存在差异，可能影响生态效度，建议后续研究采用经验取样法或实时行为记录技术收集数据。最后，实验刺激材料的潜在混淆变量控制有待深化。尽管本研究在视频时长、核心主题等维度进行了严格匹配，但未能剥离出镜人物、情感基调等视听要素的影响。未来研究可在刺激物开发阶段引入更精密的协方差控制，以纯粹地评估信息框架本身的效应。

综上所述，本研究揭示了算法干预下抖音短视频防成瘾机制中的信息类型以及心理逆反的复杂作用，尤其凸显了逆反特质的稳定效应与控制意图感知的核心作用，为短视频成瘾治理提供更具适用性的理论依据与设计建议，希望能在未来算法干预转向构建一种支持性、协作式的数字环境，在尊重用户自主性的前提下，提升其数字福祉水平，实现健康使用目标。

（责任编辑：李韵奕）

参考文献

- [1] 中国互联网络信息中心. 第 57 次中国互联网络发展状况统计报告 [EB/OL]. (2026-02-15) [2026-02-05]. <https://cnmic.cn/n4/2026/0304/c88-11549.html>.
- [2] 代宝, 郑怡晴, 杨利英. 大学生短视频成瘾影响因素及其

健康损害研究进展[J]. 中国学校卫生, 2025, 46(2): 290-294.

[3] Melissa D., M., Roxanne E., K. Patient wisdom should be incorporated into health ai to avoid algorithmic paternalism[J]. *Nature medicine*, 2023(4): 29.

[4] Brehm, S. S., Brehm, J. W. *Psychological reactance: A theory of freedom and control*[M]. New York: Academic Press, 1981: 37.

[5] Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M., Potts, K. A. Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom[J]. *Human Communication Research*, 2007, 33(2): 219-240.

[6] 代宝, 郑怡晴, 杨利英. 大学生短视频成瘾影响因素及其健康损害研究进展[J]. 中国学校卫生, 2025, 46(2): 290-294.

[7] Melissa D., M., Roxanne E., K.. Patient wisdom should be incorporated into health ai to avoid algorithmic paternalism[J]. *Nature medicine*, 2023(4): 29.

[8] Kenig N, Garcia D, Moreno J C, et al. Algorithmic paternalism: autonomy versus automation[J]. *International Journal of Transplantation & Plastic Surgery*, 2024, 8(2): 104-119.

[9] 马爱芳. 算法规训与“弱者的武器”: 算法时代人的自主性[J]. 北大新闻与传播评论, 2024: 169-188.

[10] 甘浩辰, 宋巧丽. “技术”与“权力”感知下的社交媒体用户算法抵抗行为意愿与演化机制研究[J]. 新闻与传播研究, 2025, 32(7): 69-86, 127.

[11] Shin, D., Hameleers, M., Park, Y. J., Kim, J. N., Trielli, D., Diakopoulos, N., Helberger, N., Lewis, S. C., Westlund, O., Baumann, S. Countering Algorithmic Bias and Disinformation and Effectively Harnessing the Power of AI in Media[J]. *Journalism & Mass Communication Quarterly*, 2022, 99(4): 887-907.

[12] Valkenburg, P. M., Krcmar, M., Peeters, A. L., Marseille, N. M. Developing a scale to assess three styles of television mediation: “Instructive mediation” “restrictive mediation” and “social covieing” [J]. *Journal of Broadcasting & Electronic Media*, 1999, 43(1): 52-66.

[13] Kahneman, D., Tversky, A. Prospect theory: An analysis of decision under risk[J]. *Econometrica*, 1979, 47(2): 263-291.

[14] Brehm, J. W. *A theory of psychological reactance*[M]. New York: Academic Press, 1966: 2-11.

[15] 相薏薏, 王佳栋, 许世鑫, 等. 信息框架对网络隐私保护行为改变决策的阶段性影响——基于跨理论模型的分析[J]. 情报科学, 2025: 1-24.

[16] Zhao, X., Fink, E. L. Proattitudinal versus counterattitudinal messages: Message discrepancy, reactance, and the boomerang effect[J]. *Communication Monographs*, 2020, 88(3): 286-305.

[17] Wright, M. F. Cyberbullying Victimization through Social Networking Sites and Adjustment Difficulties: The Role of Parental Mediation[J]. *Journal of the Association for Information Systems*, 2018, 19(2).

[18] Warren, R., Gerke, P., Kelly, M. A. Is There Enough Time on the Clock? Parental Involvement and Mediation of Children’s Television Viewing[J]. *Journal of Broadcasting &*

Electronic Media, 2002, 46(1): 87-111.

[19] Brehm, S. S., Brehm, J. W. *Psychological reactance: A theory of freedom and control*[M]. New York: Academic Press, 1981: 228-251.

[20] Dillard, J. P., Shen, L. On the nature of reactance and its role in persuasive health communication[J]. *Communication Monographs*, 2005, 72(2): 144-168.

[21] Richards, A. S., Bessarabova, E., Banas, J. A., Larsen, M. Freedom-prompting reactance mitigation strategies function differently across levels of trait reactance[J]. *Communication Quarterly*, 2021, 69(3): 238-258.

[22] Martin, B.A.S., Chrysochou, P., Strong, C. Crypto freedom! Effects of trait reactance and regulation content on intention to buy cryptocurrency[J]. *Personality and Individual Differences*, 2022, 194: 111659.

[23] Dillard, J. P., Shen, L. On the nature of reactance and its role in persuasive health communication[J]. *Communication Monographs*, 2005, 72(2): 144-168.

[24] Quick, B. L., Bates, B. R. The Use of Gain- or Loss-Frame Messages and Efficacy Appeals to Dissuade Excessive Alcohol Consumption Among College Students: A Test of Psychological Reactance Theory[J]. *Journal of Health Communication*, 2010, 15(6): 603-628.

[25] Drajkowski, D., Trepanowski, R. Reactance and perceived disease severity as determinants of COVID-19 vaccination intention: an application of the theory of planned behavior[J]. *Psychology, Health & Medicine*, 2021, 27(10): 2171-2178.

[26] Zhao, X., Fink, E. L. Proattitudinal versus counterattitudinal messages: Message discrepancy, reactance, and the boomerang effect[J]. *Communication Monographs*, 2020, 88(3): 286-305.

[27] Thomas D. C. Competence, counterarguing, and attitude change[J]. *Personality*, 1969, 37(2): 342-358.

[28] Alison, J. X., Robert, S. W. The Role of Bolstering and Counterarguing Mind-Sets in Persuasion[J]. *Journal of Consumer Research*, 2012, 38(5): 920-932.

[29] Dillard, J. P., Shen, L. On the nature of reactance and its role in persuasive health communication[J]. *Communication Monographs*, 2005, 72(2): 144-168.

[30] De las Cuevas C., de Leon J., Peñate W., Betancort M. Factors influencing adherence to psychopharmacological medications in psychiatric patients: A structural equation modeling approach[J]. *Patient Preference and Adherence*, 2017(11): 681-690.

[31] Hong, S. M., Page, S. A psychological reactance scale: Development, factor structure and reliability[J]. *Psychological Reports*, 1989, 64: 1323-1326.

[32] Lindsey, L. L. M. Anticipated guilt as behavioral motivation an examination of appeals to help unknown others through bone marrow donation[J]. *Human Communication Research*, 2005, 31(4): 453-481.

[33] Shaffer, H. J., LaPlante, D. A., LaBrie, R. A., Kidman, R. C., Donato, A., Stanton, M. V. Toward a syndrome model of addiction: Multiple manifestations, common etiology[J]. *Harvard*



Review of Psychiatry, 2004(12): 367-374.

[34] Zaltman, G., Coulter, R. H. Seeing the voice of consumer: metaphor-based advertising research[J]. *Journal of Advertising Research*, 1995, 35: 35-51.

[35] Quick, B. L., Scott, A. M., Ledbetter, A. M. A close examination of trait reactance and issue involvement as moderators of psychological reactance theory[J]. *Journal of Health Communication*, 2011,16(6):660-679.

[36] Karen Y. Hypernudge: Big Data as a Mode of Regulation by Design[J]. *Information, Communication & Society*, 2017,118(20):121-122.

[37] Friestad, M., Wright, P. The persuasion knowledge model: How people cope with persuasion attempts[J]. *Journal of Consumer Research*, 1994,21(1):1-31.

[38] Dogruel, L., Masur, P., Joeckel, S. Development and validation of an algorithm literacy scale for internet users[J]. *Communication Methods and Measures*, 2022,16(2):115-133.

[39] Baumeister, R. F., Bratslavsky, E., Muraven, M., Tice, D. M. Ego depletion: Is the active self a limited resource?[J]. *Journal of Personality and Social Psychology*, 1998,74(5):1252-1265.

[40] Risko, E. F., Gilbert, S. J. Cognitive offloading[J]. *Trends in Cognitive Sciences*, 2016, 20(9): 676-688.

Resisting Algorithmic Paternalism: Psychological Reactance and Shifts in Overuse Vigilance Under Douyin's Short-Video Anti-Addiction Intervention

Yu Xiyao

Abstract: This study examines the algorithmically delivered break-reminder videos embedded in Douyin's "Healthy Use" initiative. Grounded in psychological reactance theory and employing an explanatory sequential mixed-methods design, it combines a 2 (message type: restrictive vs. instructive) $\times$ 2 (reactance proneness: high vs. low) between-subjects experiment with the Zaltman Metaphor Elicitation Technique to investigate how these reminders shape users' psychological reactance and overuse vigilance. Quantitative findings indicate that restrictive reminders are more likely to elicit psychological reactance. Notably, all reminder types produce a significant decline in overuse vigilance, an effect driven primarily by the conformity reactance dimension of trait reactance. Qualitative analysis further reveals that the covert controlling intent inherent in algorithmic paternalism activates users' defensive processing at the architectural level. Users high in conformity reactance draw on their algorithmic literacy to decode the platform's structural regulation; yet, confronted with power dynamic, their cognitive resistance ultimately gives way to responsibility displacement and cognitive offloading, eroding their capacity for self-regulation.

Keywords: short video anti-addiction; Douyin rest reminder videos; algorithmic intervention; psychological reactance; reactance proneness; shifts in overuse vigilance

Author: Yu Xiyao, Faculty of Humanities and Social Sciences, School of Ecological Civilization Communication, Nanjing Forestry University.

高 / 低语境概念的矛盾性及转化： 语境化的视角

王宵静

摘要 爱德华·霍尔用高语境与低语境来概括不同文化在相遇时刻显现的差异，然而矛盾的是，在跨文化交流中得以识别的语境高低差异同时因跨文化交流而消解。对语境高低的判断始终和个体感知有关，难以作为对文化的客观描述。后续研究者试图完善高 / 低语境作为文化理论的一面，但无论是建构量表还是参照该模型分析实际案例，得到的往往是概念 - 概念、概念 - 行为之间的去语境化关联，其作为文化形成理论的一面被忽略。语境化可以成为拆解并转化高 / 低语境矛盾性的视角，即从对高低二元的关注转向对语境化过程的关注。基于实际发生的跨文化交流活动，观察语境在交流中的混合以及重新被赋予意义的过程，用寻找可能性代替下定义，从而化解概念的矛盾性，面向传播（文化）的生成和创造。

关键词 高语境；低语境；语境化；爱德华·霍尔

中图分类号 G206 **文献标识码** A

作者 王宵静，武汉大学新闻与传播学院博士研究生，湖北武汉 430072

语境可以被分出高 / 低吗？文化间的比较源于相遇，相遇使人看到差异。美国人类学家爱德华·霍尔（Edward T. Hall）在与异文化的交流中发现，不同文化在传达意义的方式上有所区别，尤其表现在对语境的利用上。他由此提出了高语境与低语境之分，并在《超越文化》（*Beyond Culture*, 1976）一书中加以详细阐述。简单来说，利用语境传达的意义占比越多，越趋向高语境交流，许多信息无需通过详尽的语言符号来传达，交流双方仿佛心有灵犀。相反，在交流中，越依赖语言符号传达信息则越趋于低语境的交流方式。

语境的高 / 低显现在不同文化相遇的时刻。在感知到差异并做出高语境与低语境的区分之后，一方得以将自己归于高 / 低语境文化之中。矛盾的是，恰恰又是不同文化个体的相遇和交流，使得高 / 低语境这一说法不再成立。原因在于，交流的展开有赖于双方调整原有的交流习惯，在对话中探索彼此都能理解的交流方式，使意义能有效地被传达和接收。在这个过程中，无论是各自对语境的感知，还是交流方式，都已经偏离了原有的高 / 低语境特点。也就是说，在跨文化交流中得以识别的语境差异又因跨文化交流而消解。高 / 低语境这一概念的矛盾性从霍尔提出时便已存在，贯穿了后期被检验、应用的始终。甚至恰恰是这一矛盾性使后续研究者在意识到其局限性的同时，依然以此作为理论模型来分析实际交流案例。



霍尔作为美国人类学学者，又曾受雇于该国政府去培训外事人员，现实境遇使他意识到必须走出人类学家的殿堂。他因而希望建立一种整合一体的文化理论，既服务于现实问题，又满足方法上的严谨性和科学性^[1]。在他那里，文化被视作一种交流方式，他希望在勾勒一种文化理论的同时，又介绍一种文化形成的理论，全面论述文化^[2]。高/低语境便是这一设想下的产物，它具备同时成为文化理论和文化形成理论的潜力，但其内在矛盾却阻碍了这一野心的实现。

语境化（contextualizing）或许可以成为拆解并转化高/低语境矛盾性的视角，它伴随了高/低语境概念的提出，却因理论模型的确立而被抛诸脑后。语境化是指信息被编织进语境中的过程，它使我们仅用少数语言代码便可以传达意义。在交流行为中，代码（code）、语境（context）和意义（meaning）共同出现，作用于交流各方的理解与沟通，语境传达的意义及传达方式则和语境化过程有关。同时，语境化在人们的交往中发生，也因交往的持续而始终处于正在进行时。然而，高/低语境这一概念在区分语境高/低的同时也中止了对语境化过程的关注，使其成为了某一文化的形容词，忽视了文化在交流中的生成和转化过程。因此，语境化的视角既是理解高/低语境概念的关键，也有助于化解其矛盾性，避免将文化对立化和定型化，面向文化的生成和创造。

后文基于语境化视角，首先回到高/低语境概念产生的语境，结合时代背景和论述高/低语境的上下文分析其矛盾性。其次，检索并分析后续研究者对高/低语境概念的发展与应用，以“high context low context”为检索词，在 Web of Science 数据库和谷歌学术（Google Scholar）中检索，尽可能检索出所有将“高/低语境”作为关键概念的文献，再采用滚雪球的方式，从已确定文献的参考文献中继续寻找。接着，对所有检索出的文献作进一步筛选，如果仅仅提到该概念，未将之作为关键的理论资源或作一步讨论或阐发，则不纳入分析。最终筛选出 130

篇文献，包括期刊论文、会议论文和书籍章节（截至 2025 年 12 月 31 日），结合语境化的视角分析这些研究案例，回答高/低语境概念的矛盾性是否以及如何被进一步凸显、延续。最后，结合现实生活中的跨文化交流问题，寻求对高/低语境概念的转化。

一、“高/低语境”的提出及其矛盾性

在与异文化打交道的过程中，霍尔发现，在不同文化中，人们的交流方式如此不同，且和某一文化整体的特点存在关联。在他看来，“文化即是交流”^[3]。这一说法可以结合语境化过程加以理解，人们在一段时期的交流之后，会形成一些共识，如特定词汇的所指、某些表情或手势的含义、特定的仪式或习惯等等。当这些东西已经被共同体中的多数人所共享并内化时，人们在交流时便无需对这些表达作详细的解释，信息已然存在于交流语境当中，也就是经过了语境化。久而久之，某一共同体便表现出独特的文化风格，在未曾参与其语境化过程的人看来，这种文化或许是难以理解甚至怪异的。因此，理解该文化的条件之一便是了解其语境化过程，也就是某种传播方式的形成过程。

如果说高/低语境同时作为一种文化理论和文化形成的理论，语境化则是其文化形成的理论的一面。然而，霍尔关于高/低语境的论述多用于概括文化差异，凸显其作为文化理论的一面，且某一文化整体总是被等同于一个国家。在此种论述中，高/低语境成为了区分国家间文化差异的概念模型，文化类型和传播方式在这一概念模型中被定型化。某种文化/传播方式的形成过程，也就是语境化的一面被忽略，在持续语境化的过程中生成新文化的可能同样未被探讨，而这恰恰是“超越文化”的关键。

霍尔曾强调，高语境与低语境之间存在几乎所有可能的语境与代码组合，但并不存在两个极端状态，即仅依赖语境的“纯高语境”或仅依赖代码的“纯低语境”，只是有些文化偏向高语境，有些则偏向低语境^[4]。从这一说法中可以看出，语境的高/低并不存在客观的界定标

准。从理念上看,比较也存在悖论,如果将两种语境并置,我们只能看到不同,但并不会有关于高/低的判断^[5]。也就是说,高/低无关乎比较,而是依赖于一种既有的关于高/低的理念,但并不存在一种普遍的客观理念,它总是受比较者主观视角的影响。从经验上来说,比较两种交流场景中语境信息的多少也极为困难,因为这和传达的意义总量、比较者对两种文化的熟悉程度有关,难以找到普适的比较维度。在这个意义上,区分出的两种或多种文化语境的高/低和对各自特征的概括总是不完备的。然而,在日常的跨文化交流中,我们又时常可以感知到语境的高低差异,因此有必要拆解高/低语境概念的矛盾性,探索面向现实交流问题的语境理论。

霍尔个人并不能为区分高/低语境提供标准,从他的论述中可以看出,虽然他希望建构一种整合一体的文化理论,但对语境的关注一方面缘于个人经历,另一方面又关联于宏观的时代背景和美国当时的社会环境,并未超越有限的个人经验。对高/低语境的建构和描述部分反映了霍尔对自身文化的反思性认同。在《超越文化》一书中,他提到:“西方人在谋求有序性时造成了混乱;其原因是,他们崇尚使经验割裂的局部事物,却否定起着整合作用的自我。”^[6]美国的四大制度,企业、政府(含防务机构)、科学和教育,每一个头面人物都认为,他在真理方面占有一隅。到哪里去寻求总览的视野?谁负责把事物联系起来^[7]?由此,他提出“延伸”“延伸迁移”这些概念。“延伸”是可以帮助人的工具或一套外化的方法、体系,它由人设计出来,人可以借此实现更高级的进化。口语是原生的延伸物、书面语是第二代延伸物^[8]。“延伸迁移”即是延伸的产物与被延伸的过程混为一谈,或延伸的产物取代被延伸的过程^[9]。比如,书面文字经常被当成更真实、更重要的东西,口语则被认为不够正式;语法学成了单独的学问,甚至脱离了语言的交流功能。人变得无法掌控自己创造的东西,它们逐渐自成体系。延伸及其融入日常生活的过程实

际上也就是某种语境化的过程,改变着人们的交流方式。

与异文化打交道的经验促使霍尔开始反思美国社会,高/低语境由此成为了描述并解释不同延伸系统的理论尝试。在讨论时,他常常将交流/传播(communication)与文化交替使用,有时写作“高/低语境交流”,有时则是“高/低语境文化”或“高/低语境系统”,具体用法常根据上下文决定。当讨论具体的交流场景时,如果大多数信息或存在于物质环境中,或内化于人的身上,需要经过编码的、显性的、传输出来的信息非常少,则属于高语境交流。如果大量信息编入了显性的代码之中,则属于低语境交流^[10]。在他看来,高语境交流是经济、迅速、高效、令人满意的。但是,它需要在程序编制上花时间^[11]。相比于描述微观的交流情境,高/低语境更多的被用于解释宏观的文化系统,高/低语境交流特点由此也直接对应于某些文化现象。比如,在高语境文化中,人们深刻介入彼此的活动、信息得到广泛共享,意蕴丰富的简单讯息自由流动^[12];有权势的人要为下属负责^[13];人们在自己的系统里可以发挥创造活力,对付新东西时则要降到语境阶梯底部^[14];趋向于使用多元时间^[15]。低语境文化或低语境系统则均与之相反。然而,低语境系统的人与高语境系统的人打交道时,则需要用远远低于他们工作习惯的低语境方式^[16]。这不难看出,由于个人经验的局限性,美国的文化及交往方式成为了区分高/低语境的参照。交流方式、文化类型和国家被对应起来,美国是低语境文化的代表,日本则属于典型的高语境文化。在后来的研究中,霍尔进一步指出,世界上的文化可以按照“语境程度”从高到低排列成一个光谱。日本人、阿拉伯人及地中海地区的人属于高语境文化群体。低语境文化群体则包括美国人、德国人、瑞士人、斯堪的纳维亚人以及其他北欧人^[17]。也就是说,高/低语境更多地被用于比较并概括国家间的文化差异,不再强调它在高/低连续体上的流动性,而对国家文化语境的判断却源于对微观交流行为的总结。



当在比较中存在时,某一文化的特点便被置于高/低语境阶梯中的某一确定位置,不再考虑因交流而出现的动态变化。由于现实交流情境的复杂性,这种简化的概括也面临自相矛盾的状况。例如,繁复、多机构的文化(技术上先进的文化)必然是低语境的文化。然而,中国却处在这个阶梯上高语境的极端^[18]。可见,单一维度远不足以描述一个文化整体的复杂性。此外,从经验上讲,在一个相对封闭的文化空间中,随着交流的持续、彼此的熟识、文化的延续,许多信息都会被编码进语境中。某一文化空间内的成员共享交流语境,整体呈现出高语境文化的特点。如果人的流动性增加,需要更多依赖外在机制维系彼此的关系,则会表现出低语境文化的特点。然而,如果承认个体受文化环境影响,也等于承认个体会因文化环境的变化而调整自身行为,原本对其高/低语境的划分将失去对其接下来行为的解释力。比如,有的人开始交往时用的是希腊文化模式,后来又因为环境的需要而滑向日耳曼-瑞士的文化模式^[19]。执着于判断文化语境的高/低就必然会面临概念的矛盾性,而在处理现实的交流问题时,也总是会遭遇概括力和解释力的难题。

二、矛盾性的延续:理论探讨与应用中的去语境化

在后续研究中,高/低语境受到的关注与争议几乎同步存在。在筛选过后的130篇文献中,仅3篇对高/低语境概念及后续应用做了系统性回顾与反思;7篇着眼于理论模型的建构与量表开发;2篇旨在阐述高/低语境与其他相关概念的关联,但以概括为主,缺乏细致的梳理和严谨的辨析。剩下的百余篇文章均着眼于实际的交流问题,将关于高/低语境文化的论述关联于特定人群的交流行为。值得注意的是,在这些文章中,高/低语境并不仅仅被用于区分国家间的文化差异,也被用于考察性别、医患、组织内部成员之间的沟通。

一篇回顾性研究发现,在关注商务与技术跨文化传播的文献中,语境化是被引用最多的理论框架。然而,大量关于语境化的命题只有

少数经过了实证检验^[20]。近年来,霍尔的高/低语境概念受到的关注度似乎有所下降。原因并非源于概念本身,而是附加于该概念之上的存疑国家分类。几乎所有运用高/低语境国家分类的研究,其依据的证据均不充分,且源于过时、未经证实的观点——这些观点甚至可追溯至霍尔本人基于轶事证据的分类^[21]。此外,高/低二分法、两极化表述、基于过时经验的抽象概括、不能反映动态变化、基于国家分类的不严谨等缺陷也被众多研究者指了出来。为了应对这些问题,部分研究者对高/低语境的相关论述加以检验,不断完善理论模型和量表。然而,基于有限情境和样本建构的理论模型被普遍化,情境和研究对象的特殊性反而被选择性地忽略,以至于没有任何一个模型和量表可以普遍适用于对各类文化行为的分析。在众多案例研究中,研究者往往一边指出其局限性,一边接受此类思维模式,继续修补该模型并重新加以应用,或者在不加反思的情况下直接将之应用于对某类行为的分析与解释。

(一) 理论模型的建构与量表开发

为了使这一概念模型应用于对具体问题的研究,有学者将高/低语境沟通风格编制成具体题项,构建高/低语境量表。古迪昆斯特(William B. Gudykunst)等人为了探讨文化个人-集体主义及其中介变量(即自我构念与个体价值观)对低语境与高语境沟通的影响,借鉴已有对高/低语境沟通风格的描述,编制了62个新题项。通过问卷测量跨文化情境下的低语境/高语境沟通风格、自我构念与价值观,量化地展示彼此关系^[22]。这些题项从既有的描述中概括而来,被参考的描述也多受个人经验影响。一篇2000年的博士论文专门对“高/低语境传播”作了概念界定与量表开发,最终生成的93个题项同样主要源于对以往研究的概括。与古迪昆斯特等人不同的是,该研究对量表作了内部一致性检验与平行性检验,证明了其结构效度;同时也证明高/低语境可以作为独立构念;美国与日本参与者的量表得分差异也表明该量表具有内容效度^[23]。然而,这些量表鲜少被直接应用于后

续的实证研究,研究者需要根据具体问题调整或重新设计量表。例如,若要探讨高/低语境文化对中美大学生选择与教授沟通媒介的影响,则需要对部分问题加以改编,并融合其他量表重新编制题项^[24]可见,在构建量表时,题项主要是对既有研究发现的整合,而既有研究又来自于对有限样本的观察。即使量表被证明具有结构效度,但题项本身并不能涵括所有经验,也难以作为普遍性量表加以应用。在展开具体研究时,研究者不得不结合研究问题和对象重新调整量表。

更重要的是,建构理论模型并设计量表的目的是量化地呈现某一文化语境的高/低程度,进而衡量与其他变量的关系,这进一步固化了文化语境的位置。对后续的应用型研究来说,研究者的现实出发点往往是希望考察交流各方在具体交流情境或某些议题下的交流特点,如果不面对情境本身的复杂性,只借助固定的问卷和数据得出答案,则容易遗漏问卷之外的独特要素,哪怕这些要素在交流中起到关键作用。最终,即便结果本身以确定的数据呈现,也难以实现研究目的。这样一来,设计量表的努力虽然减少了一些高/低语境概念的局限性,但对普遍性的追求却不得不以牺牲特殊的交流情境为代价。对于现实中的跨文化交流案例来说,影响双方交流方式选择的或许恰恰是一些难以被化约的情境要素,而使生成新文化得以可能的更是有别于既定模式的、意料之外的东西,是不同文化元素的碰撞与重组。

(二) 概念-概念的去语境化关联

不同于霍夫斯泰德(Geert Hofstede)等人的文化价值观调查,霍尔对高/低语境文化的相关描述来源于经验概括,缺乏严格的概念化工作。虽然这两个概念产生的方式截然不同,却常常同时出现在后续研究中。如果一种文化属于低语境文化,那么它通常也被认为是霍夫斯泰德所定义的个人主义文化,高语境文化同时也具备集体主义文化的特点^[25]。但是,在相互独立的概念之间寻求联系,往往又脱离了概念产生的语境,仅仅是发现了概念与概念之间的

关联,并非以直接回应现实问题为目的。

古迪昆斯特等人的研究便是此类研究的典型。为了考察文化个人主义-集体主义、自我构念与价值观对个体使用高/低语境沟通风格的影响,研究者招募了美国、日本、韩国、澳大利亚等国中型大学的学生作为被试,采取问卷调查的方式,将高/低语境、自我构念、价值观题项共同纳入问卷中。结果表明,独立型自我构念与个人主义价值观介导了文化个人-集体主义对低语境沟通的影响,依存型自我构念与集体主义价值观介导了文化个人-集体主义对高语境沟通的影响。自我构念与价值观比文化个人-集体主义更能预测低语境/高语境沟通风格,且对沟通风格方差的解释率更高^[26]。该研究经过了严谨的设计,融合了不同量表,调查对象也来源于多个国家。然而,实际交流行为中语境意义的动态变化被呈现为确定的数值,测量和计算结果虽表明了不同变量之间的影响关系,却无法解释这些关系是如何产生的。而研究结果所依据的变量由于经过了抽象化和平均化,在应对某一具体的交流案例时,也便难以捕捉特殊的情境要素和交流者的个人特征,使得概念总是难以触摸现实。还有研究将亚洲高语境集体主义文化的沟通行为与修辞敏感性理论相关联,直接假设高语境文化个体的修辞敏感性高于低语境文化中的个体^[27]。虽然这一假设并未通过验证,但采用简短问卷对有限样本展开调查,本身也难以对概念之间的关联性作出有说服力的回答。

值得注意的是,有研究发现了霍尔的语境概念与伯恩斯坦(Basil Bernstein)的“限制型/精细型语码”存在的相似性。霍尔明确将“认识自身文化运作”置于“理解他人”之上,专注于阐述“高/低语境沟通的普遍特征”;伯恩斯坦则是探究“特定沟通模式的社会生成及其在社会中的分布”^[28]。研究者虽然发现了这两个概念的相似性,但仅作了表面概念的比较,未挖掘出二者的潜在关联,也难以使人看到此种比较有助于解决何种现实问题。类似地,“低-高语境信息系统”也被认为可以和几个类似构



念整合使用,包括“关联型”(associative)与“抽象型”(abstractive)构念、“西方-亚里士多德式”(Western-Aristotelian)与“亚洲-柏拉图式”(Asiatic-Platonist)维度,以及“个人主义”(individualism)与“集体主义”(collectivism)维度^[29]。结合特定案例,不同类型的文化维度共同暗示了文化差异的复杂性,也表明单一文化维度在解释现实方面存在局限。同时,这也意味着高/低语境的交流方式与更广泛的文化背景相关联。遗憾的是,在面对真实交流情境时,各类概念的使用未必可以提升对现实的解释力,反而容易使特定案例变成对概念的佐证。因此,在发现概念之间关联性的基础上,更重要的是面向生成该概念的语境,明晰某一概念的解释限度,回应概念背后的现实问题,而非执着于寻找概念之间的关联。

(三) 概念-行为的去语境化关联

在使用高/低语境概念分析具体案例时,研究者往往会事先说明该概念存在局限,如“没有任何一种文化完全处于某一极端”^[30]。“语境实则呈连续体分布——无论文化背景如何,个体在这一连续体上的位置均可不同。”^[31]“任何文化内部,个体的信念与行为都存在差异,不能简单认为某一文化完全是高语境或低语境文化。”^[32]

“维度反映的是群体差异,无法用于预测个体行为。”^[33]“尽管高/低语境文化概念可从某一维度对社会进行分类,但并非捕捉文化复杂性的唯一方式。”^[34]这些问题如此显而易见,可以轻易地捕捉到,但依然没有阻止多数研究者按照事先警惕的方式展开研究。在应用高/低语境概念考察某类文化群体的行为时,常见的有三种研究取向。一是通过研究新的对象补充或验证既有关于高/低语境文化的描述;二是用高/低语境文化归属解释某类群体的行为;三是利用既有对高/低语境文化特点的描述,预测研究对象在某一方面的行为特点,设计与其打交道的策略。

当更多研究者以新的对象展开研究时,所得结果也便被补充进关于高/低语境特点的描述中。丁允珠(Ting-Toomey)将高/低语境文化和冲突行为关联起来,发现低语境文化中的个体

倾向于对冲突采取对抗性、直接的态度,更多采用“事实归纳式”或“公理演绎式”(即线性逻辑式)风格来管理冲突;高语境文化中的个体则更倾向于采用“情感直觉式”(即点状逻辑式)风格来管理冲突^[35]。此外,针对时间使用模式,有研究将来自西欧的个体视为低语境文化成员,来自亚洲、日本、中东及南美的个体则为高语境文化成员,发现高语境文化中的个体更在意工作时间,每日工作时间更长;低语境文化中的个体更在意社交/休闲的时间^[36]。埃及也被视为高语境文化的代表,在埃美外交关系中,历史材料显示双方的交流特点呈现出显著差异:直接性与间接性、夸张与克制、低与高“社会期望价值”、个人化与非个人化、正式性与非正式性^[37]。此类研究虽然丰富了对高/低语境文化特征的描述,但基本延续了二元对立的思维方式,在讨论某类人群的具体行为之前,已经按照国家将其划分于高语境或低语境文化之中。此外,当更多基于有限样本的调查被普遍化为高/低语境文化的特点时,又加剧了这一概念及其描述的不严谨性。更重要的是,诸多特点均指向了概括并突出差异,而非面向跨文化交流开启的新可能。

基于国家边界区分高/低语境文化的这一做法,其合理性也在被不断验证。遗憾的是,一些带有验证性质的研究首先接受了这种划分方法,如选择代表高语境文化的中国、韩国和代表低语境文化的美国作为研究对象,发现高语境文化中的人们更具社会导向、更少对抗、对现有生活方式更满足,且掌权者承担更多责任^[38]。此类研究看似验证了霍尔的结论,但并没有脱离其思维逻辑,只是用看似严谨的研究方法确认了相关表述。也有研究得出了和霍尔不一致的发现,比如选择参与MBA项目的韩美学员作业为分析文本,发现二者的文本并无显著差异,韩国学员的文本并没出现更多概括性信息、礼貌策略、间接的结构以及循环式推理等高语境文化的特点^[39]。然而,相比于口语,文字这一媒介本身就更偏向于低语境的特点。还需要注意的是,对于在美国接受商务培训的韩国人来说,其作业便是跨文化交流和培训的产物,不再是典型的韩国写作方式。

此类验证或者进一步证明了某一国家属于高/低语境文化,或举出范例说明其不合理。实际上,在接受高与低这种二元划分方式的前提下,举出任何正例或反例都不困难。也就是说,诸多验证性研究依然难以否定或证明该概念的普遍性,反而进一步凸显了高/低边界的模糊以及高/低因情境而变的特点。

在霍尔那里,将文化分为高语境与低语境并非为了寻求对文化的确定分类,更重要的是在比较中认识自我。后续研究常常直接接受某一国家属于高/低语境文化这一判断,以此解释内部成员为何表现出某种行为习惯。例如,研究高/低语境文化对四类美国族群(欧裔、非裔、亚裔、拉美裔)感知“直接与间接种族言论伤害程度”的影响^[40]。在商务沟通中,调查高/低语境价值观是否会影响员工及管理者对电子沟通的态度^[41];采用霍尔对高/低语境文化的阐述,探讨东方和西方企业知识管理模式差异的根源^[42];还有研究着眼于考察高/低语境文化与个体行为的相关性,比如,比利时(高语境文化)参与者对广告的感知复杂性低于荷兰(低语境文化)参与者,且参与者的个人语境文化得分完全解释了这一差异^[43]。即使在面对互联网、人工智能这些被广泛应用于各类文化环境的技术时,个体原有的文化语境也被认为会影响个体对技术的使用^[44]。可见,原本在语境化过程中形成的高/低语境交流特点被固化为一个国家整体的文化特点,个体则被认为受到此种文化的浸润,其行为也便由此得到了解释。然而,按照国家区分文化语境的高/低本身就遭到了普遍质疑,而个体的某一行为又发生在特定的情境中,远非国家整体的文化特征可以解释。

综上所述,对霍尔或其他交流者而言,语境的高/低多和个人感知有关,无法作为客观普遍的参照。当后续研究者希望从客位视角分析跨文化交流案例时,关于高/低语境的说法便逐渐被抽象为固化的理论模型,被不断地修补和完善。种种尝试虽然细化了该概念,为分析具体案例提供了相对一致且透明的思路和参照,但始终未消解该概念固有的矛盾性,因为高/低之分本

身便带有主观性。研究者依赖量表或既定描述来对照具体案例,而非直面语境化过程的复杂性,在意识到该概念存在局限性的同时,也只在原有模型的基础上加以完善或修补。这样一来,一面是不断精细的量表和概念描述,以量化结果呈现语境的高/低与其他概念或具体行为的关系;另一面却是对语境化过程的日渐远离,将量表之外的文化要素排除在讨论之外,无法捕捉到使某一文化得以形成的特殊因素。此外,在比较不同文化中个体的行为时,研究者常常设想了清晰的文化边界,美国和西欧的国家一般被视为低语境文化,亚洲、非洲和拉美的多数国家则被视为高语境文化,在此基础上探讨彼此的行为差异。因此,即便已经有诸多研究强调应侧重于高-低连续体,而非强化高和低的二元对立,但简化的区分依然得到了更多应用。其结果便是,在有限情境中提炼出的高/低语境特点被当成重要且普遍的东西,语境化过程本身及语境高/低的偶然性被忽略不顾。

三、再语境化:交流中语境的混合与整合

跨文化传播关注不同文化主体的交流行为,但文化间的边界并不总是清晰明了。边界往往是建构的产物,通过建构边界寻找差异最终得到的是自我认同^[45]。按照朱利安(François Jullien)的说法,认同至少用三种方式围绕着差异:(1)认同在差异的上游,并且暗示差异;(2)在差异制造期间,认同与差异构成对峙的一组;(3)在差异的下游,认同是差异要达到的目的。此种思维将我们放在同化的逻辑里,同时分类和下定义,消弭了其他文化的“他者”可能具有的“外在的”与“意料之外的”事物^[46]。在现实的交流中,交流双方的文化边界是模糊不定的,“跨”这一动作并不仅仅发生在交流中的某一时刻,而是存在于整个交流过程之中,交流主体往往在有意或无意之间就跨越了自身文化,并非存在某一确定的边界。因此,用以区分语境高/低的边界也是比较者建构出的产物。

如果回到真实的交流过程就会发现,主体通过不断整合交流经验内化交流语境,在熟人圈



里表现出高语境的交流风格，遇到陌生人时便发生了语境的混合，需要转换交流方式。在英文中，混合（mix）强调两种或多种物质混合在一起之后就会结合，且不能轻易分离^[47]。由于混合现象的存在，当带着对某一文化属于高/低语境文化的既定认知，并用相关量表展开研究时，常常会出现难以归类的情况。在对群体间、组织内部交流差异的研究中，有研究者将高/低语境挪用到对组织文化的研究中，认为组织文化受男性价值观主导。有意味的是，霍尔描述的低语境文化特点看似与男性价值观一致，但女性在组织中体验到的是“高语境”——这种文化已预先植入男性价值观，女性并不共享那些“使沟通可理解的语境”^[48]。女性在男性的高语境空间中采用低语境方式沟通，当男性进入女性的空间时，这种动态可能反转^[49]。分类与概括是人类面对异文化时减少不确定感的常见思维方式，但分类却总是受分类者经验和既有认知的影响，也只适用于有限经验范围内的事物。因此，在实际的跨文化交流行为中，我们常会感到分类不起作用，或者需要不时地为分类寻找新的依据，因为跨文化交流本身便是对原有经验的超越。

在交流中，人们能够重新将混合语境整合起来，使代码、语境和意义以新的组合形式出现，这也是再语境化的过程。整合（integration）一词最初意为“将整体的各个部分聚集在一起的行为”，源自法语 *intégration*。该词又直接来自晚期的拉丁语 *integrationem*（主格 *integratio*），有“更新、恢复、重新开始”之意^[50]。体现在交流行为中，则意味着来自同一文化的个体会因后期境遇的不同而表现出不同的交流方式，比如，居住在荷兰的希腊人比居住在希腊的希腊人表现出更高水平的非语言沟通能力和更间接的表达方式，以及更多的隐喻使用^[51]。这表明人们并非总是固守原有文化，而是会在交流中不断加以调整。在跨文化商务合作中，高语境文化会为低语境文化调整，接受更长的合同，且这种适应与合同履行地无关^[52]。针对全球虚拟团队的研究发现，两种文化导向的参与者均

表现出高度协作行为，同时在很大程度上保留了自身的沟通风格与文化价值观。除高/低语境行为外，还存在一种新兴行为模式——“切换行为”，个体行为会根据目的、情境与沟通对象调整，从而发生变化^[53]。影响代码转换的因素包括沟通平台、沟通对象的风格及沟通目的^[54]。可见，交流的达成有赖于交流各方的试探和互相调整。除了整合交流对象的习惯外，历史背景、交流目的、媒介条件等环境因素均影响交流者对语境的利用方式。在这种情况下，仅仅通过了解研究对象所处文化的特点并不足以解释或预测其具体行为。语境并非总是外生的决定因素，还是沟通者可以主动利用的东西^[55]。它不只以交流背景的形式存在，还是在交流中不断生成的，也是交流者可以有意识去构建的。

人的整合能力早已被科学研究所证实，霍尔当时就参考了颜色心理学家费伯·比伦（Faber Birren）与兰德相机的发明者埃德温·兰德（Edwin Land）的实验结果，表明眼睛和大脑视觉中枢可以通过内部语境化补充缺失信息^[56]。当下又有新的证据表示，一条包含海马体、杏仁核与内侧前额叶皮层的神经环路，参与了使情境依赖行为成为可能的学习与记忆过程^[57]。大脑的这一特点使人们可以捕捉语境信息，在交流中调整对语言代码的使用。人脑的中前额叶区域可以将社会、躯体、脑干、边缘和皮层全部连接在一起，将他们在身体内部进行整合。我们的同理心、为其他人的心灵绘制地图均与此有关^[58]。这意味着我们可以在关系中感知他人，不断调整自我在交往中的表现。交流者对文化语境的整合与利用也是语境化得以可能的前提，且部分解释了高/低语境存在的矛盾性，即语境的高低和交流主体的感知有关，主体又因交流的持续而发生感知上的变化，从而促使行为的调整。一方面，不同的情境因素使人们对语境的感知与利用存在偏向，在语境化的过程中生成不同特点的文化类型；另一方面，语境化过程是持续的，当与不同文化的个体发生交流时，人们也将在新的跨文化关系中协商代码和语境的意义，调整对语境的利用方式。

用高和低区别不同的文化语境，除了会遭遇概念上的矛盾外，也将着眼点收缩在了过去，而非指向交流开启的未来。《超越文化》开篇指出了当代世界两个相互联系的危机，一是人口与环境的危机；二是人与其延伸、制度、思想的关系中存在的危机，以及地球上许多个体之间和群体之间存在的危机^[59]。霍尔认为，应对危机需要人们同心协力、了解彼此的思维方式，要寻找新的选择、新的可能、新的范围、新的取舍和新的途径，去发挥人的创造潜能^[60]。遗憾的是，就对高/低语境这一概念的建构来说，他却以定型化思维取代了开创性思维，用高与低这样凸显差异的形容词掩盖了文化互化过程。无论是文化间、群体间抑或主体间，交流的发生都意味着语境的混合，而主体拥有感知语境的能力，可以在混合语境中吸纳新经验，在反思自我的同时整合语境信息，经由新的共同语境打开交流空间。也就是说，交流的持续虽然使高/低语境概念面临矛盾性，但同时也使该概念的转化成为可能，从对高/低二元的关注，转向对语境化过程的关注，不再执着于区分差异，而是观察主体对混合语境的重新整合，寻找在文化互化过程中生成新文化的可能。

四、结语：从高/低语境文化到文化的语境化生成

不同文化的相遇虽然使文化差异得以显现，但同时也开启了消解差异的过程，使文化语境的高低之分充满了矛盾。一方面，高/低语境概念停留在了区分差异的时刻，但交流伴随着建构共同语境的过程，原有的高低区分不再成立；另一方面，高/低无关比较，而是依据既有的高/低观念得以判断，但一种完全客观的高/低观念并不存在。霍尔将语境、代码和意义视为同一问题的不同方面，意义由语境和代码共同传达。将语境分出高/低的前提在于，意义是确定的量，语境传达的意义占比多，则是高语境，反之则为低语境。然而，意义本身和主体的感知有关，也因对话而不断发生变化，难以量化或计算，也无法将某一交流情境中的意义普遍化为其所属文化整体的意义。也就是说，语境的高/低更

多的是就交流者的感知而言，这也是为什么即使高/低语境在逻辑上并不严谨，我们也常常在日常生活中感知到语境高/低的存在。

霍尔对高/低语境概念的建构和描述深受个人经验的影响，其目的是自我认同。基于对美国文化的反思，他区分了高语境文化和低语境文化，美国是低语境文化的代表，并且很大程度上成为了区分高/低语境文化的参照。此外，概念描述多源于个人经验，其普遍有效性并未得到证明。最终，按照某种维度区分文化最终又远离了反思自身文化的目的，用同化逻辑排除了其他文化中无法被化约的要素，而往往是这些东西使文化走上了不同的轨迹，只有还原这些才能使反思有效。

高/低语境概念突出语境在交流中的角色，为理解文化的生成提供了有益视角，但其对语境的划分却逐步向固定的文化理论收敛，使高/低语境成为某一文化的形容词。然而，在后续的诸多应用中，恰恰是高与低这种简化表述被后续研究者捕捉，延续了其矛盾性。研究者鲜少直面真实的语境化过程，回避了其不确定性和复杂性，直接将高/低语境模型作为对文化背景的概括，虽然可以给出较为明确的研究结果，但该结果往往显示的是概念-概念、概念-行为之间的关联。由于这种关联是去语境化的，因而难以真正回应实际案例中的交流难题。也就是说，无论是建构模型、完善概念，还是基于概念解释或预测群体或个体行为，其结果往往是延续甚至深化了高/低语境概念的矛盾性。由于现实中人的流动、交流的持续，在跨文化交流发生时，不同文化语境也便发生了混合，进而被交流者重新整合。虽然有研究结果提示了这一现象，却并未在此基础上对高/低语境概念作进一步讨论。

在主体间的交流中，意义经由表达而流动，经由流动而在主体间分享，在各主体对之加以内化整合之后又重新进入交往关系，产生主体间的意义协商。在这一过程中，主体感知到的语境总是高低相倾的，受到交流对象、话题类型、交流时长、意义协商进程等的共同影响。



主体对语境的感知和再利用则是重新参与语境化的过程,也是在混合语境的基础上构建共同语境的过程。构建共同语境并赋予其意义是生成新文化的前提。以往的应用型研究多带着实用目的解释或预测某一文化群体的行为、提供跨文化交流建议,虽然丰富了对高/低语境概念的描述,但缺乏反思和深化,也逐渐远离了这一概念最初具备的反思现实的可能性。文化的语境化生成视角着眼于高/低语境作为文化生成的理论的一面,不再执着于为文化作归类工作,或是借用某种模型预测个体行为以便提供实用策略,而是关注更广泛的跨文化交流现象,以语境化的视角观察主体的流动以及发生在此过程中的跨文化交流,以动词或过程思维取代名词思维或定型化思维,从真实而具体的交流案例中发现生成新文化的可能。

在分析具体案例时,首先需避免将真实的交流语境按照固定的高/低语境模型加以简化、抽象,而应侧重于分析主体如何在交流中赋予语境以意义。语境在此过程中不仅仅是上下文(context),还包括情境(situation)、环境(environment)、情形(circumstance)以及心境(state of mind)。其中,上下文与文本是同一过程的不同方面,是一直伴随文本的东西。情境常用于强调对个体言行有较大影响的境况、职位等。环境则是指环绕或包围个体的更大背景,既包括自然物理环境,也包括社会文化环境。情形侧重于影响个体行为的客观状况或不可抗拒的因素。一般来说,研究者们倾向于统一用“情境”表示研究对象所处的宏观或微观境况。交流主体的心境则和人格特征、文化背景和特定处境有关,有着千人千面的特点。在跨文化交流中,这些维度交织在一起,既作为交流背景存在,也是交流者可以利用并重新赋予意义的东西。

为语境赋予意义发生在主体间的交流中,也就是再语境化的过程。如果搁置既定的文化划分,从主体的交流出发考察文化的语境化生成,那么当主体与不同对象、在不同情景中发生跨文化关系时,实际上也就开启了语境化过程并参与了各类新文化的创造。例如,就跨越

地理空间的交流而言,由于贸易或其他原因,不同语言的人为了保持彼此的交流,会形成“接触语”,开始是不同语言的混杂语,在逐渐有人以其为母语后,便被称作克里奥尔语,语言学由此提出了“接触”(contact)概念。普拉特(Mary Louise Pratt)在讨论殖民地的交往时借用了这一概念,引申出“接触区”(contact zone),用以指示不同文化群体相遇和互动的社会空间,强调主体是在彼此之间以及通过彼此之间的关系构成的^[61]。也就是说,文化的交界处也许是一片模糊、混杂的文化空间,并逐渐生出不同于两边任一文化的独特文化类型。一种语言表达形式的出现、接触区等均可以成为观察文化的语境化生成的案例。主体在流动中与其他主体相遇,在表达与交流中进行意义的分享,如果语言不通,则需更多的借助语境传达意义,在逐渐熟悉彼此的语言之后,用代码传达意义的比例便可能提高。由于交流的持续、共同生活的需要,各自的语言、习俗、文化符号等可能发生交融,需要为混合语境重新赋予意义并加以内化,由此也便生成了新的文化类型。

如果从主体的流动与交流出发考察文化,那么对文化类型的划分就不再仅仅受地理空间的影响。如“附近”的消失与重建。我们在日常生活里越来越通过一些抽象的概念和原则建立自己对世界的感知,而不是通过对自己周边的感知来理解^[62]。即使在自己的生活半径里,许多人也不再共享语境,生活越来越依赖外在的系统加以安排,原有的语境发生着瓦解。但同时,不同于原有的社区或社群,附近是相对于主体而言的,在远离因地理空间形成的语境时,主体也得以在流动中定义新的附近,与之共建新的语境。这一现象也意味着需要重新理解文化,地理边界不足以区分文化类型,人的流动、各种技术或系统的发展使各种各样的文化圈得以形成。如跨文化粉丝圈等被冠以“跨文化”的圈子,这里的跨文化延续的是原有的文化划分方式。如果搁置此种划分,只面向当下或未来的语境,那么因某一议题或兴趣等形成的圈子本身便是区别于其他圈子的文化圈,与其他圈子的交流

才算是跨文化的交流。这些均可以视为文化的语境化生成现象,但使这些文化得以生成的情境因素则有别于面对面的交流。

再比如,如果来自某一情境的人物、信息与行为规范渗透并突破另一情境边界,会被认为发生了语境坍塌(context collapse)^[63],而社交媒体使这一现象更加普遍。然而,如果不从群体想象出发推导行动与属性,而从行动出发,看从中生成何种群体,就会发现人们处于高度具体的互动中,他们的语境化能力相当成熟。在这个过程中,语境没有坍塌,相反,语境在不断增殖、扩张^[64]。也就是说,社交媒体为人们跨越地理空间发生交往关系提供了可能,它使人们不断围绕某一议题生成新的圈子,圈子内成员因此可以共享一种语境,同时也参与其他圈子的语境化过程。这时,主体参与语境化的过程也有别于面对面的交往,媒介技术、互联网符号、共同议题等起到了更多作用,这些均是既定的高/低语境维度难以涵盖的。

当面向实际发生的跨文化交流活动,就会发现文化语境的高/低之分常常失效,但语境却始终伴随着交流,与代码一起表达意义。如果搁置高/低之分,不再用本身存在矛盾的理论框架彰显客观性,而是从主体的交流出发观察意义的流动,则有可能使该概念重新具备对现实的反思力和创造力。高/低语境的矛盾性并不独特,而是广泛存在于跨文化传播的各类理论建构之中。实际上,跨文化的概念本身便意味着“随之而来的新文化现象的创造”^[65]。相比于开发高度抽象概括的概念或理论模型,观察并分析实际交流中的语境化过程,着眼于传播(文化)的生成和创造也许更有意义。

(责任编辑:李韵奕)

参考文献

- [1][2][3] 爱德华·霍尔. 无声的语言[M]. 何道宽,译. 北京: 北京大学出版社, 2010: 24-25;26;75.
- [4][56] Hall ET. Context and meaning[M]//. Samovar LA, Porter RE. *Intercultural Communication: A Reader*(9). Belmont, CA: Wadsworth, 2000:34-43.
- [5] 张祥龙. 比较悖论与比较情境——哲学比较的方法论反思[J]. 社会科学战线, 2008(9):58-64.

- [6][7][8][9][10][11][12][13][14][15][16][18][19][59][60] 爱德华·霍尔. 超越文化[M]. 何道宽,译. 北京: 北京大学出版社, 2010:9;108-110;27;27;82;90;37;94;112;133;112;82;40-41;1;2.
- [17] Hall E T, Hall M R. Key concepts: Underlying structures of culture[M]//. Albrecht, M. H. *International HRM: Managing diversity in the workplace*, 2001:26.
- [20] Cardon P W. A critique of Hall's contextualizing model: A meta-analysis of literature on intercultural business and technical communication[J]. *Journal of Business and Technical Communication*, 2008, 22(4): 399-428.
- [21] Kittler, M. G., Rygl, D., Mackinnon, A.. Beyond culture or beyond control? Reviewing the use of Hall's high-/low-context concept[J]. *International Journal of Cross Cultural Management*, 2011, 11(1) 63-82.
- [22][26] Gudykunst W B, Matsumoto Y, Ting-Toomey S, et al. The influence of cultural individualism-collectivism, self construals, and individual values on communication styles across cultures[J]. *Human communication research*, 1996, 22(4): 510-543.
- [23] Ohashi R. High/low-context communication: Conceptualization and scale development[D]. Michigan State University, 2000.
- [24] Yang X. The influence of high/low context culture on choice of communication media: students' media choice to communicate with professors in China and the United States[D]. University of Louisville, 2016.
- [25] Gudykunst, W.C., & Ting-Toomey, S. *Culture and interpersonal communication*[M]. Newbury Park, CA: Sage, 1988: 44.
- [27] Knutson T J, Komolsevin R, Chatiketu P, et al. A cross-cultural comparison of Thai and US American rhetorical sensitivity: Implications for intercultural communication effectiveness[J]. *International Journal of Intercultural Relations*, 2003, 27(1): 63-78.
- [28] Oliver C. From Class to Culture: Restricted/Elaborated Codes vs. High/Low-Context Communication in Basil Bernstein and Edward T. Hall[J]. *Sophia University Junior College Division Faculty Journal*, 2016(37): 73-83.
- [29] Korac-Kakabadse, N., Kouzmin, A., Korac-Kakabadse, A. and Savery, L. Low- and high - context communication patterns: towards mapping cross-cultural encounters[J]. *Cross Cultural Management: An International Journal*, 2001, 8(2): 3-24.
- [30][35] Ting-Toomey, S. Toward a theory of conflict and culture[M]//. W. B. Gudykunst, L. P. Stewart & S. Ting-Toomey (Eds.), *Communication, culture, and organizational processes*, Beverly Hills, CA: Sage, 1985: 71-86.
- [31] Zakaria N. Emergent patterns of switching behaviors and intercultural communication styles of global virtual teams during distributed decision making[J]. *Journal of International Management*, 2017, 23(4): 350-366.
- [32] Al-Issa A. When the west teaches the east: Analyzing intercultural conflict in the classroom[J]. *Intercultural Communication Studies*, 2005, 14(4): 149.



- [33] Hall M, De Jong M, Steehouder M. Cultural differences and usability evaluation: Individualistic and collectivistic participants compared[J]. *Technical communication*, 2004, 51(4): 489-503.
- [34] Richardson R M, Smith S W. The influence of high/low-context culture and power distance on choice of communication media: Students' media choice to communicate with professors in Japan and America[J]. *International Journal of Intercultural Relations*, 2007, 31(4): 479-501.
- [36] Manrai L A, Manrai A K. Effects of cultural-context, gender, and acculturation on perceptions of work versus social/leisure time usage[J]. *Journal of Business Research*, 1995, 32(2): 115-128.
- [37] Cohen R. Problems of intercultural communication in Egyptian-American diplomatic relations[J]. *International Journal of Intercultural Relations*, 1987, 11(1): 29-47.
- [38] Kim D, Pan Y, Park H S. High-versus low-context culture: A comparison of Chinese, Korean, and American cultures[J]. *Psychology & Marketing*, 1998, 15(6): 507-521.
- [39] Thomas J. Contexting Koreans: Does the high/low model work?[J]. *Business communication quarterly*, 1998, 61(4): 9-22.
- [40] Leets L. Disentangling perceptions of subtle racist speech: A cultural perspective[J]. *Journal of Language and Social Psychology*, 2003, 22(2): 145-168.
- [41] Kabasakal H, Asugman G, Develioğlu K. The role of employee preferences and organizational culture in explaining e-commerce orientations[J]. *The International Journal of Human Resource Management*, 2006, 17(3): 464-483.
- [42] Tang S, Fang Y, Wang X. Knowledge management pattern differences in oriental and occidental firms: a cultural enquiry: a cultural enquiry[C]. *인하대학교 정석물류통상연구원 학술대회*, 2009: 1082-1087.
- [43] Jos Hornikx & Rob le Pair, The Influence of High-/Low-Context Culture on Perceived Ad Complexity and Liking[J]. *Journal of Global Marketing*, 2017, 30(4): 228-237.
- [44] Wang L, Rau P L P, Evers V, et al. When in Rome: the role of culture & context in adherence to robot recommendations[C]//2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2010: 359-366.
- [45][46] 朱利安·间距与之间：论中国与欧洲思想之间的哲学策略[M]. 卓立，林志明，译. 台北：五南图书出版公司，2013：25;32.
- [47] Oxford Learner's Dictionaries. Mixing [EB/OL]. (2026-2-28). https://www.oxfordlearnersdictionaries.com/definition/english/mix_1?q=mixing.
- [48] Marshall J. Viewing organizational communication from a feminist perspective: A critique and some offerings[J]. *Annals of the International Communication Association*, 1993, 16(1): 122-143.
- [49] Marshall J. Organisational cultures and women managers: Exploring the dynamics of resilience[J]. *Applied Psychology: An International Review*, 1993, 42(4): 313-322.
- [50] Etymonline. Origin and history of integration [EB/OL]. (2026-2-28). <https://www.etymonline.com/word/integration>.
- [51] Broeder P. Informed communication in high context and low context cultures[J]. *Journal of Education, Innovation, and Communication*, 2021, 3(1): 13-24.
- [52] Velez-Calle A, Sosa S, Large J. Cultural communication style and international joint venture contract length[J]. *International Journal of Cross Cultural Management*, 2021, 21(3): 491-506.
- [53] Zakaria N. Emergent patterns of switching behaviors and intercultural communication styles of global virtual teams during distributed decision making[J]. *Journal of International Management*, 2017, 23(4): 350-366.
- [54] Zakaria N, Ab Rahman Muton N. Cultural code-switching in high context global virtual team members: A qualitative study[J]. *International Journal of Cross Cultural Management*, 2022, 22(3): 487-515.
- [55] Cole B M. Lessons from a martial arts dojo: A prolonged process model of high-context communication[J]. *Academy of Management Journal*, 2015, 58(2): 567-591.
- [57] Maren S, Phan K L, Liberzon I. The contextual brain: implications for fear conditioning, extinction and psychopathology[J]. *Nature reviews neuroscience*, 2013, 14(6): 417-428.
- [58] Siegel D J. Mindful awareness, mindsight, and neural integration[J]. *The Humanistic Psychologist*, 2009, 37(2): 137-158.
- [61] Pratt, M. L. *Imperial Eyes: Travel Writing and Transculturation*[M]. London, Routledge, 1992: 6-7.
- [62] 项飙, 康岚. “重建附近”：年轻人如何从现实中获得力量？——人类学家项飙访谈（上）[J]. *当代青年研究*, 2023(5):1-9+21.
- [63] Jenny L Davis, Nathan Jurgenson. Context collapse: theorizing context collusions and collisions[J]. *Information, Communication & Society*, 2014, 17(4): 476-485.
- [64] Szabla M, Blommaert J. Does context really collapse in social media interaction?[J]. *Applied Linguistics Review*, 2020, 11(2): 251-279.
- [65] Ortiz, F. *Cuban Counterpoint: Tobacco and Sugar*[M]. DE ONÍS H, translate. Durham: Duke University Press, 1995:102-103.

(下转第 96 页)

县级融媒体中心自我造血机能的结构困境与解题路向¹⁾

——基于Y省的考察

朱天 冯楷 高敬

摘要 主流媒体系统性变革对县级融媒体中心提出了更高的适配性要求，增强自我造血机能、提升建设成效成为普遍共识。本研究通过对Y省13家县级融媒体中心的实证研究，发现其自我造血机能普遍存在三重结构性困境，即外部刚性规约、认知偏差等叠加造成的自主性发展空间缺失；经营实体产权模糊与角色功能错位，导致激励性绩效考核失效；缺乏应对多重竞争压力的有效办法，使得传统积累难以转化为市场优势。对于县级融媒体中心而言，解困的应为路向除了积极争取地方党委、政府更精准的投入扶持和机制优化外，还需致力于成为支持本地社会发展的关系连接枢纽。

关键词 县级融媒体中心；自我造血；管理机制；系统性变革

中图分类号 G210 **文献标识码** A

作者 朱天，四川大学新闻传播与出版学院教授，四川大学传媒研究中心主任，四川成都610207；冯楷，四川大学新闻传播与出版学院博士研究生，四川成都610207；高敬，四川日报报业集团川观智库研究中心研究员，四川成都610000

一、问题的提出

2018年全面启动的县级融媒体中心建设，至今已取得显著的阶段性成果。这场由政策赋权形成外部推力、党政系统主导^[1]的“全国工程”有效打通了基层舆论引导与服务群众的“最后一公里”，区县传播声量持续提升，县级融媒体已成为我国全媒体传播体系中不可或缺的关键一环。

推进主流媒体系统性变革的全局逻辑与现实要求，在更高层级上对县级融媒体中心赋予了县域社会数字化服务枢纽、基层治理重要支撑载体等多重功能。这些跨场域、多元化的长效功能落地，无法仅依靠财政输血的静态保障实现，必须依靠自我造血形成的动态、可持续的内生发展动力。

从当下可见的现实状态来看，增强自我造血机能，提升县级融媒体中心建设成效，是政府政策设计、学术研究结论及基层实践诉求所拥有的普遍性

1) 基金项目：四川省哲学社会科学基金重大项目“系统性变革理念下天府传媒联合体的区域传媒生态重构研究”（SCJJ25ND011）。

共识。从顶层政策设计来看，中共中央办公厅、国务院办公厅印发的《关于加快推进媒体深度融合发展的意见》中明确提出了“要发挥市场机制作用，增强主流媒体的市场竞争意识和能力，探索建立‘新闻+政务服务商务’的运营模式，创新媒体投融资政策，增强自我造血机能”^[2]。县级融媒体中心建设伊始发布的技术建设规范文本，也明确了支持互动活动、区域运营、电子商城等各类增值服务，提出需“具备相应的运营支撑和发展能力”^[3]；既有研究也普遍认为，党委、政府的扶持是用来站稳脚跟^[4]，自我造血才是“建强”融媒体中心中心的必由之路和“用好”融媒体中心的经济基础^[5]，也是保障媒体“事业属性”功能充分发挥^[6]、持续产出优质内容、吸纳专业人才的关键前提^[7]；从全国行业实践的整体图景来看，虽然部分发达地区的县级融媒体中心通过体制内资源内循环、与本地产业企业强关联等创新实践^[8]，打破了区县市场体量小、需求有限^[9]的固有认知，但多数县级融媒体中心在产业发展等板块的开发上仍存在显著短板^[10]，自我造血能力薄弱已成为制约其高质量发展的重要痛点。

既有研究虽对公益一类与二类的身份束缚、技术迭代滞后、市场手段创新度不足、人力支撑薄弱等因素有所阐释，但相关论述多未能更深入还原基层县级融媒体中心的在地化困境，也未能在在此基础上对应形成破解其自我造血机能建设问题的具体思路。基于此，本研究选取西部大省Y省为研究对象，其县域经济社会的类型分布与发展水平、县级融媒体中心的资源禀赋与政策执行语境，均具有中西部区域形态的典型代表性。2025年6月起，研究者实地调研了Y省13家县级融媒体中心（全部10家公益二类和具有一定代表性的3家公益一类，见表1），其间与上述单位的中心主任、经营管理业务分管副主任、相关科室或部门负责人及部分一线员工分别进行1-3小时的深度访谈，累计访谈70余人，同步发放并100%回收调研问卷，获取了组织架构、人力资源、经营效能等维度的一手完整数据。由此，本文将系统阐释县级融媒体中心自我造血

表1 研究对象信息

单位编码	单位性质	实地考察与访谈时间
X1	公益二类	2025.6
X2	公益一类	2025.6
X3	公益二类	2025.6
X4	公益二类	2025.7
X5	公益二类	2025.7
X6	公益一类	2025.7
X7	公益一类	2025.7
X8	公益二类	2025.10
X9	公益二类	2025.10
X10	公益二类	2025.10
X11	公益二类	2025.10
X12	公益二类	2025.11
X13	公益二类	2025.11

的结构性困境，并探寻适配区县媒体发展特征、兼具在地可行性与长效可持续性的破解路径，进而形成从困境识别到提出解困路向的完整理路。

二、县级融媒体中心自我造血机能的结构性困境

基于实证研究的发现，本部分构建了机制属性、组织能力、场域生态的“问题框架”，以此阐释县级融媒体中心在增强自我造血机能方面的现实困境。其中，机制属性直接决定造血的合法性基础与权限边界，组织能力关乎造血的主体动能与执行效能，场域生态界定了造血的外部条件与业务形态演化。

（一）“能否造血”的机制之困：公益类别博弈与自主空间挤压

公益一类或二类的模式选择，往往来自当地对媒体的定位、商业模式及其未来盈利能力的判断^[11]。在现有机制框架的运转下，县级融媒体中心在管理机制与自身发展的关键问题上，形成了一些选择取向并做出相应的动态调整。X7作为公益一类机构，因管理模式的限定，业务及发展资源相对受限，甚至直接影响了主体业务完成度，因而呈现出强烈的向公益二类转型的诉求。不过，地方政府对其转为公益二类

持谨慎态度，既担忧市场化经营可能引发的风险外溢，也顾虑其若经营失败所造成的社会影响及可能进一步加重的财政负担，故其转型所需的容错机制难以获取。另一典型案例 X2 在成立时选择了公益二类，但在 2025 年转回公益一类。核心原因在于地方市场的萎缩，以及地方政府对其既有的财务管理模式，使该中心经营所得无法顺畅回流至中心的绩效分配和自主建设体系，内生性发展诉求尤其是差异化的个体激励需求难以得到回应。而转为公益一类后，财政虽保障了近百名人员的基本工资与机构运行经费，但发展资金匮乏的问题则进一步凸显，包括客户端系统升级、线上文创专栏开发等创新项目推进缓慢；无法通过市场化薪酬体系吸引符合岗位需求的高层次人才；中心与下属经营公司实现完全切割后，其经营收益无法反哺融媒体中心的建设发展，进一步加剧了发展资金的短缺局面。

这两个案例本质上是县级融媒体中心在“生存”与“发展”的过程中所面临的权衡与选择的问题，也在相当程度上反映了基层媒体对政策环境适应、协商的过程，同时还显现了县级融媒体中心面临的机制规约与自主发展需求之间的现实矛盾。

地方决策及管理层对县级融媒体中心发展逻辑的认知差异，也是形成上述问题的重要原因。如在访谈中，多家县级融媒体中心提及区域内一些党政机关工作人员所存有的一些观念：“县级融媒体中心挣这么多钱有什么用？”（X12）“我们单位宣传的事就应该由融媒体中心负责。”（X13）“（做宣传片）是融媒体中心的本职工作就不该付费。要付费的事情，就不该是融媒体中心做。”（X5）

此类认知偏差不仅对县级融媒体中心的自主发展构成了实质性制约，更关键的是对其区域性经营市场环境和资源开发带来冲击。如 X6 设立运营部门的考虑是要有造血机能，但因无明确管理机制及市场运行规则保障，其在后续的实际运行中，成了一个为县内各部门“无偿”制作专题汇报片的机构，反而增加了融媒体中

心的运行负担。

研究同时发现，本应承担发展保障功能的财政扶持，也可能因县级融媒体中心造血机能的加强而动态调整支持额度，故县级融媒体中心又因担心市场收入的不稳定与不可控，进而弱化了自身的经营创收动能。如 X1、X8 担心创收过多会导致财政削减保障资金，进而在经营上“畏手畏脚”、自设“界限”。“我觉得没必要在创收上出彩，太出彩反而不一定是好事。如果财政发现我很能挣钱，可能会削减我下一年的保障预算。”（X1 主任）

从制度设计逻辑上而言，公益二类天然具备实现经营效益最大化的潜力，部分县级融媒体中心却形成了隐性的经营预期阈值。即便知晓在地方区域空间内有怎样的“潜在市场”，也不敢在相关领域“大展拳脚”，最终陷入造血机能持续弱化的困境。

（二）“谁来造血”的主体之困：外在规范平衡与激励手段失灵

下属公司是公益二类县级融媒体中心自我造血的核心组织。本研究对 10 家公益二类单位的问卷调查显示，是否设立下属公司与市场化自主能力高度相关。有下属公司的单位 100% 具备相应完整的管理自主能力（包括自主决策权、经营所得自主支配权两个维度）；而无下属公司的单位仅 20% 表示具备完全自主能力，且无公司的单位中 60% 为“完全无自主”（自主决策权、经营所得自主支配权均不具备）。此外，有下属公司的单位表示能 100% 实现“收益反哺”（即公司收益有机会返回给财政，再反哺到中心建设）。

不过，公司作为自我造血的主体抓手还存在隐性难题。在 Y 省当前的通行模式中，县级融媒体中心大多以股东身份全资或控股设立下属经营公司，但中心主任不担任公司法人代表。正如 X10 主任的比喻：“融媒体中心相当于市场上的门面，公司相当于客户经理。”该模式虽规避了“事企不分”的管理风险，但中心也难以实现对经营实体的有效管理。同时，国有资产与公司自有资产分属不同产权谱系，产权

划分模糊可能滋生国有资产无偿占用、收益流失等风险。研究还发现，大多数公司成立的首要作用并非获取经营收益，而在于解决人力资源的聘用、绩效等配套性事务。如 X9 的下属公司主要解决聘用人员的招聘及薪酬支付问题；X1 则通过扶持其持股的一家国企参与市场竞争，帮助聘用人员通过项目参与提升绩效收入。

该模式还可能在相当程度上导致激励机制的失灵。公司作为经营载体，其聘用人员承担核心创收任务，却无法完全享受成果分配。事业编制人员薪资体系、绩效考核的微小差额，均难以激发其参与市场经营的主动性，进而使其滋生“搭便车”的机会主义倾向。就公益二类的本质而言，应有更充分的绩效分配上的自主权。若这一层面不具有足够的灵活度和弹性，公益二类的市场化价值将会大幅折损。

另一现实困境指向业务人才招聘中县级融媒体中心自主可控性问题。问卷调查揭示，13 家受访单位中超半数（7 家）存在该方面的问题，具体有无法实现命题精准化、专业限定、面试环节的自主把控等，这导致招聘人员往往难以匹配核心业务，尤其是开拓市场空间的能力要求。

（三）“何以造血”的生态之困：业态转型滞后与竞争优势流失

对公益二类单位整体创收情况的调查显示，在收入来源上，体制内转移支付的“输血模式”是核心支撑，受访单位中有 2 家在这一比例上超过 90%。然而，体制内转移支付的整体体量呈明显缩减态势，部分地区甚至出现断崖式下滑的情形。如 X2 在 2024 年的相关收入较前几年下降近八成；X13 近年来相关政府类活动数量同比减少一半以上。此外，X12、X13 等都提到了部分政府部门活动回款不及时的问题，长期挂账无法收回，既阻碍了自我造血良性循环的构建，也抑制了市场拓展的积极性。

在收入形式上，宣传片（40.74%）、活动执行（22.68%）、广告（13.59%）等传统业务的平均占比合计超七成；仅 2 家尝试新兴业务（直播带货 / 文创）；4 家有第三方平台账号运

营收入，但占比均低于 20%。从受访单位的整体收入结构来看，其多依赖于传统业务，基于互联网生态的新兴业务布局与营收转化能力仍存在明显不足。此外，有公司单位的市场收入平均占比（53.2%）显著高于无公司单位的市场收入平均占比（4.83%），收入形式数量（平均 4.8 种）也多于无公司单位（平均 2 种）。可见，下属公司的搭建对县级融媒体中心市场化经营能力与业务多元性的提升具有正向作用。

从市场总盘的角度考虑，一个更现实的因素在于——传媒生态在市场端汇入了更多外部节点参与竞争，基层传媒组织不再是市场的垄断者。新增的市场竞争者多是地方的文旅类国企，以及不断涌现的社会传媒公司和自媒体。此类新兴市场参与者业务覆盖面广、运营灵活性强，导致县级融媒体中心在传统业务领域曾拥有的市场议价权逐渐流失，其作为媒介主体的专业积淀也难以转化为比较优势。

一个宣传片投标，原预算 15 万被压到 8 万再到 4 万。以前我们定价在 2 万左右一分钟，现在 5000 块一分钟也得做。（X12 副主任）

一家酒店联系我们想拍广告，我们报价 1 万块钱一分钟，对方却说 3000 块能不能做。他们觉得我们无成本投入，但不清楚这里面还有很多工序。（X5 副主任）

人家（其他市场主体）价格低、服务好、面带微笑、随叫随到。我们融媒体中心的“机关模式”何谈市场竞争？（X8 主任）

上述县级融媒体中心自我造血面临的困境，并非单向传导的线性制约关系，而是相互交织、彼此嵌套的约束关系，最终共同引致其自我造血机能整体偏弱、内生发展动力不足的现实困局。

三、县级融媒体中心增强自我造血机能的可能破解路向

面对前文所述的问题结构与现实需求，更应采取的姿态是构建破解问题思路的“解题框架”，探讨适配区县传媒生态的解困路径。

（一）机制发力：三级决策协同结构中精准解惑、松绑和赋能

有研究指出，利用制度优势获取稀缺资源并完成变现，往往需要两个“一把手”（当地党委书记、媒体机构“一把手”）才能实现^[12]。结合本研究的实地观察而言，解决上述问题或需要三个“一把手”协同发力，即区（县）委书记确立政策方向与发展方位，将融媒体中心自主发展融入地方发展的整体布局；区（县）宣传部长常态化协调资源，打通跨部门协作堵点；融媒体中心“一把手”负责落地，并不断推动机制层面涌现问题的有效解决，实现顶层设计与基层实践的有效互动。

在上述协同机制下，需要着重解惑、松绑和赋能的关注点应是：其一，设立可动态调整、有创新探索及一定容错空间的管理体系。对有转型意愿的公益一类机构，理想路径在于设定若干年限的转型过渡期，在过渡期内保持原有财政保障规模不变，同步赋予其市场权限。过渡期结束后，机构可根据经营成效选择正式转为公益二类或转回公益一类；对已有的公益二类机构，更应通过资源赋权等方式，充分释放其在多元业态创新、市场空间拓展上的机制优势。

其二，构建主责主业更加明晰、自主发展更有活力的保障机制。在现有运行机制下，县级融媒体中心的人员数量和规模相对有限，既要完成主责主业的任务要求，又要顾及自我发展与市场空间开拓，必须进一步厘清融媒体中心的权责边界，平衡公益职能和造血诉求，从而解决前文所述的地方政府部门出于各自需求，而对其形成的任务叠加以及资源稀释问题，避免由此持续挤占核心业务资源，导致其主责主业履职效果及市场拓展能力受到限制。

其三，地方政府资源支持需要在常态稳定的前提下更为灵活。“造血”的支撑基础是充分的“补血”。政府资源支持的力度，很大程度上决定了县级融媒体中心发展的空间，目前地方政府对公益一类融媒体机构财政给予经费保障；对公益二类融媒体机构财政给予经费补助，并通过政府购买服务等方式予以支持^[13]的做法亦行之有效。不过就本研究所发现的问题，或可为公益一类机构单独设立专项发展资金，

定向解决既有财政保障无法完全支撑的特定诉求；针对公益二类机构，需明确在一定周期内年度补助额度不与经营创收额度“逆向”挂钩，通过“增量留存”的形式破除“自设界限”、激发其自我造血的潜能。

（二）主体动力：权责规范前提下重构人才激励体系

对于大多数县级融媒体中心而言，首先需要厘清中心与下属公司的角色边界与利益联结。可行的调整方向是，在遵守国有资产管理规定的前提下，明确中心与公司的关系边界，明确中心经营方向决定权、收益分配监督权等主体责任，在法律法规和管理框架下赋予公司市场化用工权、薪酬分配权和日常经营财务自主权。具体而言，可以探索“中心管方向、公司管执行”的分工模式，通过制度化的契约关系，实现权责对等。

其次，激励效能弱化也是影响自我造血机能的重要元素。本次考察的13家单位中6家未实现同工同酬，而完全实现从身份管理到岗位管理转变的则更少。同工同酬恰是破解激励失灵的重要前提，如已实现这一方式的X13，做得好的非编人员月薪，甚至高于同单位事业六级管理岗位的薪酬。此外，要更充分释放绩效分配的自主权，其核心要义在于保障增量部分要有足够“显示度”，避免因基数不足难以触动个体行为选择，导致激励失效。调研中X10的实践提供了可借鉴的范本，其推行全员“积分制”考核，经营部门的薪酬直接与项目业绩、创收额度挂钩，不仅实现了员工月平均收入的普遍提升，更让业绩突出的员工收入实现翻番。这种模式让绩效考核成为真正的管理激励手段，使“看得见”的增量成为撬动自我造血机能的杠杆，让每一名员工都成为造血的积极参与者。

最后，人才招聘的岗位适配性问题，需从“招什么人”和“怎么培养人”两端同时发力。现有人才建设体系既要解决入口端的“匹配度”问题，又要以“选苗培育”替代“大水漫灌”式的人才培养体系。在留存与流动层面，问卷显示，13家单位中有8家近三年人力流出数量

大于流入数量。人才的合理进出并非组织活力的流失，流动本身即是对组织内人才培养生态的正向佐证，但需要保持的是其进出人才能力的“平衡度”。唯有通过持续的人才引育与动态优化，维系人力队伍“有进有出”的良性循环，组织才能始终保持自我造血的活力与创造力。

（三）生态聚力：在算好“清晰账”的基础上实现区域价值共创

从增强自我造血机能的具体路径上而言，核心逻辑并不是盲目跟风布局、靠低价内卷抢夺存量市场，而是要在精准锚定县域市场底层逻辑的基础上，重构自身核心竞争力，从被动的市场参与者转变为县域传媒生态的主导者、区域社会发展的连接枢纽。

这其中，首要命题在于算好人员薪酬、设备折旧、技术维护等显性支出的“清晰账”，明确维持新闻生产等核心功能和市场经营等附加业务的刚性支出底线。在此基础上，还需对区县范围内宣传文化领域的财政支出流向、社会传媒机构的市场份额占比及潜在需求空间有更为准确的认知把握。如上文提到的 X7 具有转入公益二类的诉求，但面对其区域内国企庞大的业务资源，若没有政府行为的资源整合和政策倾斜，其即便转入公益二类，可撬动的市场容量及相应利润或也难以支持其常态运行发展。

如此，便具有了厘清自身可依托的资源禀赋与不可控的外部变量的可能性。既明确该做什么，进而基于资源匹配度筛选业务方向，又明确不必做什么，避免在缺乏优势且陌生的领域盲目布局和竞争。只有通过量化自身经营的收支平衡点与区域市场的资源承载力，县级融媒体中心才有可能在有限的空间内实现资源的最优配置。

县级融媒体中心应是所在区域的传播和服务窗口。理想的造血破局之道，是落点在基层治理与社会民生的现实需求上释放价值。为此，县级融媒体中心需要深度理解本地社会的独特语境，构建以自身为核心节点的区域传媒协同网络。如可探索组建“乡镇服务联盟”，将直播电商、农事服务等与县域产业发展、基层诉求深度绑定，不仅能将市场化业务转化为服务区域高质

量发展的具象实践，县级融媒体中心以“枢纽”的角色参与到社会治理也能更有底气。如果以协同与连接作为逻辑指引，还可以尝试搭建“自媒体联盟”，进而破解当前与本土自媒体合作浅层化的问题，或可通过为联盟成员提供标准化内容生产工具、精准技术支持等方式，将原本分散且可能存在竞争关系的自媒体转化为“合作伙伴”，形成资源互补、流量互通的区域传媒生态，共同挖掘县域市场的增量潜力。

总之，县域传媒生态的重构，从来不是零和博弈的存量争夺，而是基于县域发展需求的价值共创。县级融媒体中心唯有跳出“为了创收而创收”和把互联网仅作为传播渠道的窄化思路，把激活自我内生动能深度嵌入县域经济社会发展的全局，才能实现自我造血的可持续性。

四、结语

县级融媒体中心的自我造血机能困境，是管理机制受限、内生动力欠缺与传媒生态变化交织形成的复杂症结。其中既有历史因素的沉积，又叠加了新发展环境的现实挑战，各类影响要素的相互缠绕与动态演化，共同构成了当下的现实图景。

县级融媒体中心因覆盖范围的广泛性与主体数量的规模性，其自我造血机能的困境不容忽视。其不仅将在一定程度上影响其“引导群众、服务群众”的作用发挥，以及传媒组织赋能基层社会发展的实践效能，甚至可能在一定程度上对主流媒体系统性变革的整体推进成效造成影响。从现实来看，不同区县的历史背景、资源条件与治理环境存在显著差异，单纯依赖外部经验的移植和新兴业务的扩张或难以触及问题核心。唯有精准地在地化设计与操作，方能为自我造血机能路径的畅通与发展效能的提升奠定基础（图 1）。

本研究通过实践观察的方式，进一步探究了县级融媒体中心自我造血机能的缺失现状及缘由，并给出了可能的破解路径，以期为主流媒体系统性变革背景下县级融媒体中心的发展提供参考，助力其真正成为扎根基层、服务地方的重要枢纽平台。

（责任编辑：徐秋染）

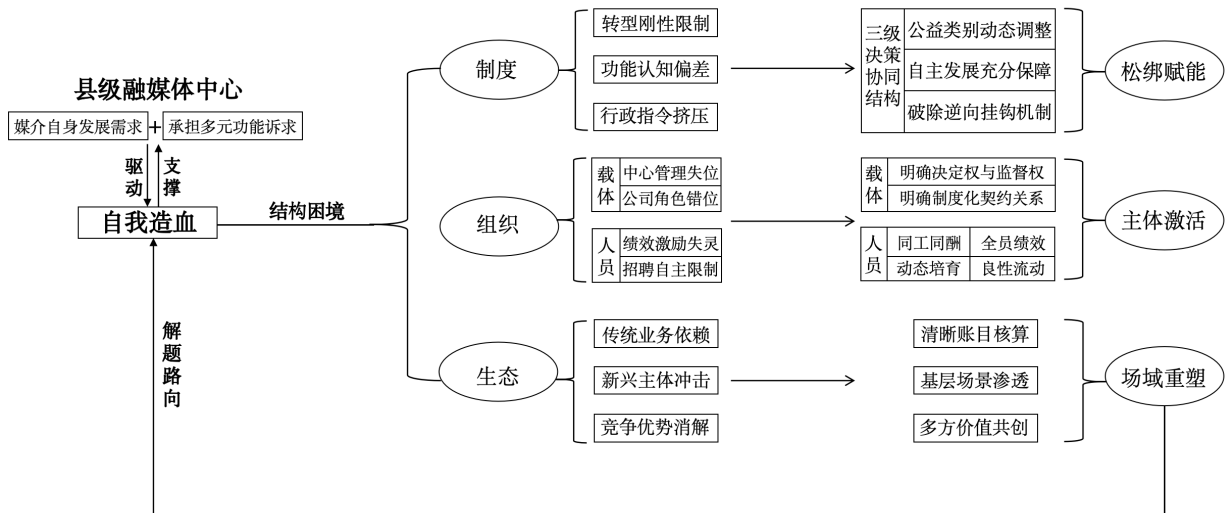


图1 县级融媒体中心增强自我造血机能的问题和解题框架

参考文献

[1] 周逵,黄典林.从大喇叭、四级办台到县级融媒体中心——中国基层媒体制度建构的历史分析[J].新闻记者,2020(6):14-27.
[2] 新华社.中共中央办公厅 国务院办公厅印发《关于加快推进媒体深度融合发展的意见》[EB/OL].(2020-09-26)[2026-03-20].https://www.gov.cn/zhengce/2020-09/26/content_5547310.htm.
[3] 国家广播电视总局.《县级融媒体中心省级技术平台规范要求》《县级融媒体中心建设规范》发布实施[EB/OL].(2019-01-15)[2026-03-20].https://www.nrta.gov.cn/art/2019/1/15/art_3557_41835.html.
[4] 陈国权.扶持体系下县级融媒体中心市场机制构建[J].传媒,2023(4):32-36.
[5][8] 曾培伦,郑雯.内循环与本地化:我国县级融媒体中心的经营路径探析[J].传媒观察,2022(10):75-81.

[6] 曾培伦,朱春阳.融媒十年考:中国媒体融合发展的逻辑转换与汇流[J].新闻界,2023(11):12-22.
[7] 赵月枝,王欣钰.何以“主流”:媒体系统性变革的政治经济学分析[J].暨南学报(哲学社会科学版),2025,47(10):20-46.
[9] 颜春龙,等.县级融媒体中心建设[M].北京:科学出版社,2022:145.
[10] 王奕涵,黄楚新.数智融合与系统变革:我国县级融媒体中心发展现状与趋势展望[J].青年记者,2025(10):55-62.
[11] 赖贵全,等编著.我亲历的县级融媒体创新发展:来自仁寿县、古蔺县等的实践探索[M].北京:科学技术文献出版社,2022:93.
[12] 郭全中.媒体深度融合的“大融合”思路及实施关键[J].现代传播(中国传媒大学学报),2022,44(9):1-7.
[13] 新华社.中共中央 国务院关于分类推进事业单位改革的指导意见[EB/OL].(2012-04-16)[2025-11-15].<https://china.huanqiu.com/article/9CaKrmJuZK5>.

Structural Dilemmas and Solution Pathways for the Self-Sustaining Capacity of County-Level Converged Media Centers: Based on an Investigation in Province Y

Zhu Tian, Feng Kai, Gao Jing

Abstract: The systemic transformation of mainstream media poses higher requirements for the adaptability of county-level converged media centers, and enhancing self-sustaining capacity and improving construction effectiveness have become a common consensus. Through an empirical study of 13 county-level converged media centers in Province Y, this paper finds that such centers are generally trapped in three structural dilemmas regarding self-sustaining capacity. First, the combined effects of rigid external institutional rules, cognitive biases, and other factors lead to the inadequacy of their independent development space. Second, the ambiguous property rights of operating entities and the mismatch between roles and functions render incentive-oriented performance evaluation mechanisms ineffective. Third, lacking in effective measures to cope with multi-dimensional competitive pressure, these centers struggle to convert traditional accumulation into market competitive advantages. To break through such

dilemmas, local party committees and governments should provide more targeted financial and policy support and promote institutional optimization. Meanwhile, county-level converged media centers must build and continuously strengthen all-round ties with local society, becoming regional connection hubs to empower local social development.

Keywords: county-level converged media center; self-sustaining capacity; management mechanisms; systemic transformation

Authors: Zhu Tian, College of Journalism, Communication and Publishing, Sichuan University; Feng Kai, College of Journalism, Communication and Publishing, Sichuan University; Gao Jing, Sichuan Daily Press Group.

(上接第 88 页)

Deconstruct and Bridge the Contradictions in High- and Low-Context Concepts: A Contextualizing Approach

Wang Xiaojing

Abstract: Edward T. Hall used the concepts of high-context and low-context to summarize the differences manifested when distinct cultures encounter one another. Paradoxically, the differences in context identifiable through cross-cultural communication are simultaneously dissolved by it. Judgments about whether a culture is high- or low-context are always tied to individual perception, making it difficult to treat them as objective descriptions of culture. Subsequent researchers attempted to refine high/low context as a cultural theory, but whether through constructing measurement scales or applying the model to real cases, what often emerged were decontextualized correlations between concepts, or between concepts and behaviors, while the theory as a framework for understanding cultural formation was overlooked. Contextualizing offers a valuable perspective for deconstructing and bridging the contradictions in the high/low-context, which is to shift focus from the binary of high versus low to the process of contextualization itself. Based on actual cross-cultural communication, this approach observes how contexts mix and are re-signified in the course of interaction. Rather than defining, it seeks possibilities to resolve conceptual contradictions while opening toward the generation and creation of (cultural) communication.

Keywords: high-context; low-context; contextualize; Edward T. Hall

Author: Wang Xiaojing, School of Journalism and Communication, Wuhan University.



像保护眼睛一样保护生态环境
像对待生命一样对待生态环境

公益广告



四川省生态环境
宣传教育中心

新闻界

宣

欢迎
订
阅

新闻界

- 中文核心期刊
- CSSCI 来源期刊
- 中国人文社科综合评价(AMI)核心期刊
- 四川省品牌期刊

思想影响传播
媒介推动文明

征订方式:

1. 通过当地邮局订阅, 邮发代号: 62-334。

2. 通过《新闻界》发行部订阅(一年起订):

《新闻界》为月刊, 每期 20 元, 订阅费: 336 元/年(含杂志费 240 元/年, 邮费 96 元/年)。

开户名称: 四川日报报业集团

银行账号: 51001890836051508414

开户行: 中国建设银行股份有限公司成都锦城支行

【注】: 转账时请务必注明“《新闻界》订阅 × 年 × 月——× 年 × 月”字样, 转账后请尽快将您的汇款凭证、收刊人的姓名、电话、地址、邮编等详细信息发送至《新闻界》微信公众号后台, 以便您及时收到期刊。



微信公众号二维码
微信号: ixinwenjie

主管主办单位: 四川日报报业集团

发行部电话: (028) 86969774 (028) 86968876