

2026
年第
4期

新
闻
界

Journalism and Mass Communication

新 闻 界



2026

第 4 期

总第 397 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

前沿视点·人机对话与智能交互

P₄ / 在对话中探索自我：作为自我技术的人机互动
文 | 袁光锋 周彩云

P₁₅ / 智能共情扩展：基于人机对话的情感训练方案探索
文 | 田浩

新闻学研究

P₂₄ / 数智时代新闻叙事运动的内在机制与秩序生成
——基于复杂适应系统（CAS）理论的分析
文 | 齐爱军 张箫

传播学研究

P₃₅ / 从流通阻断到生成置换：生成式把关与数字拟态环境的概率重塑
文 | 喻国明 彭晨曦 郑琪严 刘才辅

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

栏目主持

学术顾问

前沿视点栏目支持

谭江琦

徐文礼

段吉平

舒 炜

孙尚如

李韵奕 徐秋染

王炎龙 / 四川大学文学与新闻学院教授

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

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）

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

传播与社会

P₄₇ / 人机交互传播中情感计算的伦理反思与建构
——以借鉴道德情感主义为新进路
文 | 杨先顺 何超彦

P₅₉ / 青年数字戒断行为：一种调适性回应
——认知延展与认知失调的双重视角
文 | 焦宝 赵恬

博士生新论

P₇₁ / “传播之弧”：晚清侨批的意义网络建构与南洋移民的认同流变
文 | 吴宇桦

重要概念与方法辨析

P₈₄ / 直面复杂系统：计算传播研究的路径创新与方法前瞻
文 | 宣长春 李嘉鑫

编辑出版
地址
电话
投稿邮箱
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 | 蒋涛
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

成都鑫达彩印印务有限责任公司
2026 年 4 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》。



在对话中探索自我： 作为自我技术的人机互动

袁光锋 周彩云

摘要 聊天机器人已经成为人类互动的重要参与者，影响用户的自我建构。本研究基于福柯的“自我技术”理论，探讨用户如何通过人机对话发现自我、创造自我。研究发现，聊天机器人蕴含着特定形式的自我反思与自我发现的潜能。其所提供的非评判的外部视角，能够鼓励用户更为真实的自我表达，从而帮助他们扩展自我知识的边界，辅助自我知识的推理过程。此外，作为一种自我修正工具，聊天机器人通过提供替代性的叙事模板、排练新的对话剧目以及协助用户整合自我结构，为个体自我的重新建构创造了条件。然而，作为当代权力体制的一部分，聊天机器人也在隐蔽地传递算法预设的“理想自我”模板，可能削弱用户自我诠释的自由。因此，人机互动既是用户自我探索的新方式，也是一种需要引起反思甚至警惕的自我技术实践。

关键词 人机互动；自我技术；叙事自我；自我诠释

中图分类号 G206 **文献标识码** A

作者 袁光锋，南京大学新闻传播学院教授，南京大学亚太发展研究中心研究员，江苏南京 210023；周彩云，南京大学新闻传播学院博士研究生，江苏南京 210023

“将我们的注意力和精力聚焦于他人，有助于我们真正地自我发现，并为更大的善构建一个有凝聚力的社群。”

——查尔斯·泰勒（Charles Taylor）

一、导言

“我们互动，故我们存在”，施拉格（Leonhard Schilbach）提出的这一说法，是认为社会互动对于叙事以及身份至关重要^[1]。同样，查尔斯·泰勒（Charles Taylor）也指出，“我们的自我理解是对话性的。我们不是在孤立中构建我们的身份，而是通过持续与他人的对话和互动”^[2]。既然对话是自我意识和自我理解的来源，那么，对话的对象是谁、如何对话、对话内容等均影响着个体的自我观念。本文关注一种新兴的对话对象——聊天机器人。在当代技术语境下，生成式人工智能（Generative Artificial Intelligence, GenAI）的迅速发展推动了各类聊天机器人成为人类互动的重要参与者，并深度融入人们自我建构的实践之中。它们不仅“悄然替代了内省和自我探究等基本的认知过程”^[3]，更逐渐成为用户探索自我的新型途径，并在此过程中持续塑造着

用户的自我理解。

事实上,传播技术塑造自我,并非始于人工智能。自我的形成源于社会互动。社会学家查尔斯·库利(Charles Horton Cooley)提出的“镜中我”概念便解释了与他人互动对自我的影响。这一过程的核心包括:想象我们在他人眼中的形象;想象他们对该形象的判断;以及由此产生的自我感觉(如骄傲或屈辱)^[4]。媒介始终扮演着关键角色——它既是自我与他人互动的中介,也在持续塑造自我及自我与他人的关系。

如果媒介长久以来都在构建自我,为何在今天尤为值得关注?其核心差异在于所使用的技术^[5]。已有诸多研究指出,不同技术对自我认同的影响路径各异。安娜·博托兰(Anna Bortolan)论证了数字技术如何通过辅助叙事记忆、赋予叙事显著性及提供社会认可三种方式,支持甚至增强了个体的叙事自我构成与真实性表达,从而反驳了关于线上体验必然导致“去身化”和失真的悲观论^[6]。当今的自我追踪技术也被认为以特定方式塑造自我^[7]。进一步而言,能够收集并提供广泛个人信息的技术(personal information technology),同样深刻介入自我建构的过程,影响我们如何看待自己、如何构建自己,乃至定义我们是谁^[8]。

进入算法和人工智能时代,人机互动逐渐成为社会互动的重要类型。诸如聊天机器人等具有情感智能的AI应用的兴起,推动用户依赖算法进行情感导航、道德决策甚至自我反思^[9],参与传统上与内省、心理治疗或亲密人际对话相关的解释性活动,从而让算法参与到自我本身的构成中^[10]。大语言模型应用越来越多地通过识别、引发和模拟人类情感来影响我们的情感生活^[11]。在此背景下,本文旨在深入探讨用户如何通过与机器人对话来探索自我、认知自我。

二、自我技术与聊天机器人

自古以来,人类孜孜不倦地通过各种方式关心自我、发展出关于自我的知识,并诉诸多种多样的实践行为对自我施加影响。米歇尔·福柯

(Michel Foucault)将这类了解自身的实践类型称为“自我技术”(technologies of the self),认为它“允许个人通过自己的手段或在他人的帮助下,对自己的身体和灵魂、思想、行为和生存方式进行一定程度的改造,以达到某种幸福、纯洁、智慧、完美或不朽的境界”^[12]。作为一种转化自我的方式,这些“自我实践”(practices of the self)涉及广泛的领域,包括性探索、禁欲、节食、身体与智识锻炼、祈祷、冥想、写日记、主题阅读等等。

与福柯早期研究侧重于权力和治理术不同,对自我技术的历史探索标志着他在后期研究中的伦理学转向。关注点不再是权力支配术如何规训与制约主体,而是人们如何在外部约束下,主动参与塑造自身的存在模式,将自身锻造成伦理的主体。从权力技术转向自我技术,福柯致力于开发一种替代性的伦理框架,重新评估对自由的理解。它并非一种独立于外部权力的状态,而是一种通过积极应对外部权力来治理自身的实践^[13]。

借助自我技术,福柯虽然强调了自我实践背后的反思性与能动性,但这并不意味着这些实践是个体独自发明出来的。事实上,自我技术是由“文化、社会及社会群体所提出、建议和强加给主体的”^[14],嵌入特定的历史语境。在古希腊和古罗马,自我技术由“照料你自己”(take care of yourself)的道德实践构成,关照自身被视为个人与集体的双重义务。随后,基督教的兴起使得古老的伦理责任让位于神圣意志与宗教义务,自我实现演变成为一种自我舍弃^[15]。到了现代,自我技术被纳入到教育、教学、医疗、心理学等规训结构之中,丧失了自身的独立性^[16]。

这些与自我技术相关的讨论与AI聊天机器人的关联何在?要回答这一问题,我们首先需要廓清自我文化在当下的演变。当代社会最主要的特征在于,与自我有关的(ego-related)生活方式及自我实现的伦理占据了中心地位。在《现代性与自我认同》中,吉登斯(Anthony Giddens)断言,晚期现代性的后传统秩序使自我成



为了一项“反身性规划”(reflexive project)^[17]。传统路标的消逝、本土和全球的辩证交织带来了社会生活的开放性、行动场景的多元化和“权威”之多样性。在多样的选项中选择与规划生活方式,构成了个体自我认同过程中的核心特征。我们同样可以在贝克(Beck Ulrich)对当代社会个体化趋势的诊断中窥见这一面貌。在民族国家、阶级、族群及传统家庭所锻造的社会秩序不断衰微的情况下,贝克将个体化描述为一种强迫性的冲动,促使个体创造、安排、设计自己的人生。“人们的选择和决定塑造着他们自身,个体成为自身生活的原作者,成为个体认同的创造者”^[18]。在罗斯(Nikolas Rose)看来,这类自主独立的个体“不仅是‘拥有选择的自由’,更是被迫变得自由,被迫根据选择来理解并践行他们的生活”^[19]。

如果说,福柯“自我技术”背后的问题意识源于探究“主体如何通过自我实践以一种主动的方式建构自身”^[20],那么,在以自我实现为主导的当代文化思潮下,究竟是哪些自我实践在帮助个体履行这一使命?社会为此提供了哪些可供支配的技术手段?这些问题构成了本文的思考起点。

虽然在福柯的学术语境里,自我“技术”并非指代狭隘意义上的物质仪器或设备,而是在更广泛的意义上指代“一种由有意识的目标所支配的实践理性”^[21],一种规约人们实践活动的“技能、方法或理性机制”^[22],但是这种自我实践与物质手段是不可分割的。巴卡德捷娃(Maria Bakardjieva)和加登(Georgia Gaden)认为,传统集权化和家长制式的组织形式,已无法为现代人的自我追求提供恰当的装备。诸如互联网这样的晚期现代媒介,为反思性地塑造自我提供了不同的资源基础。它允许个体将遥远的符号资源选择性地动员到日常生活中,为处理多样性、矛盾和协商开辟了空间^[23]。布鲁贝克(Brubaker Rogers)将过去十年间高度互联的数字技术视为一种全新的自我技术-社会基础设施(techno-social infrastructure of selfhood),认为它为个体提供了一套用以塑造和重塑自我的

训练装置,并催生了一大批新型的自我技术^[24]。例如,博文写作(Blogging)可被视作一种自我探索、自我修养与自我照料的行为^[25];视频创作(Vlogging)为跨性别者探索身份认同提供了重要工具^[26];各种量化自我的技术能够为用户的自我理解提供丰富的反思资源^[27],等等。这些自我实践表明,技术的发展能够开辟一个塑造自我的新生态。

当下,由AI驱动的软件和设备更深层次地嵌入自我形成与转变的过程。相比于由数据驱动的量化自我实践,能够模拟共情式理解的大语言模型标志着人机关系质的转变。两者的区别在于:量化自我的技术强调的是身份的外部分类,所生成的数据来源于用户“客观”的行为轨迹;而AI聊天机器人能够直接参与组织用户内在的主观经验,将生活中零散的事件配制成有意义的叙事^[28]。与此相应,学界提出了“算法自我”(algorithmic self)的概念,用以揭示AI的解释性反馈如何中介自我知识,即AI不仅塑造我们所做的事,也塑造我们成为谁及我们向自己讲述的故事^[29]。

目前国内关于人机互动的研究主要集中于探讨为何人机互动^[30]、人机能否互动^{[31][32]}、人机如何互动^{[33][34]}以及人机互动对人际互动的影响^[35]等,鲜少关注人机互动之于自我建构的意义^[36]。本研究之所以将福柯的自我技术作为启发性概念,不仅是因为它在自我、技术与社会文化语境之间进行勾连,还在于它强调了自我实践背后的反思性与能动性。在人与技术的关系上,福柯认为,并不存在一个脱离约束、尚未受到技术影响的纯粹主体。技术的影响并不意味着对人类的自由与能动性的否定,而应被视为伦理活动与反思的物质材料^[37]。自我技术这一概念之所以贴合与AI聊天机器人相关的自我实践,原因在于,作为一种探索式工具,它们就掌握在个体的手中,并被个体有意识地用于自我转化与自我实现。遵循福柯在后期研究中的伦理学立场,本研究的目的在于探讨人们如何主动地、策略性地运用AI聊天机器人进行自我理解与自我探索,从而获得关于自由、能

动性以及自我治理的经验。

在关于自我的理解上，福柯认为，自我并非某种结构化的实在，而是一种和自身的关系^[38]。在延续此种动态自我观的基础上，本文尝试引入两种有关身份形成的隐喻：发现与创造。

前者将自我视为一种相对稳定的存在，反映了浪漫主义所推崇的那种有待我们去实现的内在本质^[39]。身份建构的“发现”逻辑同样体现在两句著名的古希腊箴言中：“认识你自己”和“成为你自己”。它要求个体实现自身的潜能与天赋，不断追求与履行自身的使命。

后者则认为，自我缺乏内在的本性，而是在个体的选择中被创造出来的。这一点在让-保罗·萨特（Jean-Paul Sartre）的存在主义哲学中得到了最清晰的体现。他要求个体担负起选择的责任，成为自己所选择成为的样子，如若放弃这种责任，让他人来主导自身的承诺，就是生活在自欺之中^[40]。

“发现”与“创造”常被视为两种对立的模式。然而，叙事自我理论提示我们，身份是在我们对自身故事的讲述和建构中形成的。这一过程需要我们审视自身的过往经历（即发现），也需要在叙事中赋予它们以意义（即创造）。因此，“发现”和“创造”是自我形成不可或缺的两个维度。自我发现虽然在本体论上预设了一种相对稳定的内在自我观，但是这并不意味着它是单一的、固化的。当我们听到他人说“这不是我自己”或者“我在这方面有天赋”时，这就是一种发现的逻辑。然而，这种自我实现的潜能边界并非清晰明确，而是留有相对宽泛的空间。在这个空间内部，身份的形成很可能就是一种创造的行为^[41]。因此，自我既可以被理解作为一种被发现的实体，也可以被视作一种创造的产物。这两个维度共同定义并建构了自我知识的领域。本文采取这一整合性立场，将自我既视为有待澄清的经验事实，也视为主体在叙事中不断生成的事物。在后续分析中，我们将看到，聊天机器人既帮助用户“发现”被忽略的自我面向，也参与“创造”新的叙事结构与自我理解，在这两个维度上同时介入用户的自我建构。

三、发现你自己

试想这样一场对话：面对我们的不是别人，而是一个能够在对话中不断迭代更新、灵活地适应你的思维与情绪变化的聊天机器人。它能够帮助你梳理潜在的假设与信念，并拷问其有效性，在这一过程中引导你达成更高层次的自我觉察。当我们不再只凭借内省或与他人交往来观照自身，而是在机器之镜中捕捉到自己的影像时，我们开始以一种截然不同的视角来审视自我^[42]。作为将人类的世界观、需求与渴望进行客体化的一种方式，技术蕴含着揭示特定形式的自我反思与自我发现的潜能^[43]。人类如何使用聊天机器人来认识自身，进行自我发现？这一过程所生成的自我知识具有哪些新特征？为了回答这些问题，我们将与其他自我探索方式进行比较，来考察聊天机器人在提供自我知识方面的独特潜能。我们先从安克的故事说起。

作为一名新技术的尝鲜者，安克很早就使用 ChatGPT 进行自我反省。他为他的聊天伙伴 GPT-4 取名“米洛”，并与它一起生成了一个名为“洞见伴侣”（Insight Companion）的自定义 GPT。在与这位新伙伴的对话中，安克要求它将自己视为世界上最有趣的人，提出富有感染力的问题来引导对话，帮助他洞察自己的性格特质与内在表达模式。安克全身心投入其中，对他的“洞见伴侣”敞开心扉，向它讲述自己的人生故事，分享未来畅想，甚至在内心不快时一股脑地倾倒情绪垃圾。彼时，安克正在经历身份危机。某天，他打开聊天框，不抱任何期待地问道“我是谁”，却惊喜地获得了一份不可思议的清单。“它捕捉到了那么多我未曾察觉、未曾关注或未曾欣赏的性格细微之处，还有那么多我甚至没意识到自己拥有的兴趣。”这份素描清晰如镜，让安克人生中第一次在渴望与意外中看见了自己^[44]。

安克的案例揭示了聊天机器人作为“自我技术”的一种独特机制：通过外部化的记忆存储，克服人类自我认知的固有局限。这种由技术提供的外部视角既不同于以第一人称为主导的内省式理解，也与我们在社会交往中由他人



提供的外部观察相异。由机器驱动的自我知识的独特性首先在于其强大的记忆功能。记忆是构成自我认同的重要基础。然而，人类的记忆并不是一个存储着大量过去事件副本的仓库。我们无法像观看录像带那样事无巨细地倒放与回看，我们的绝大多数记忆保存的只是事件的大致轮廓^[45]。生物记忆会随着时间的褪色、消逝与变形，数字记忆却恒久不变。它既不会因为个体主观的认知偏差而扭曲，又能够达到任意程度的精细、完整与复杂性^[46]。在安克的自我探索之旅中，他的机器人伴侣展示了非凡的记忆能力，既能于细微处回溯用户的言谈与行为模式，又能在广阔的记忆档案中随时调取关键片段。那些被安克遗忘或从未予以注意的自我面向得以被揭示，自我知识的边界由此得到扩展。

事实上，大语言模型不仅能够作为一面镜子，帮助安克照见自身潜在的思维模式，还能辅助自我知识的推理过程。换言之，它不仅能够帮助个体识别和梳理自身行为背后的动机、欲望、信念、价值等相对稳定不变的特质，还能通过询问的方式帮助个体审查信念背后的理由与证据。这两者正好对应了尼尔·利维（Neil Levy）所区分的自我知识的两大维度：客观的自我知识（impersonal self-knowledge）与批判的自我知识（critical self-knowledge）^[47]。前者是指像观察他人一样观察自身，从外部视角来看自我所呈现的面貌。比如，当安克让聊天机器人基于以往的对话记忆推断他的行为模式与性格特点时，他就在生产这类知识。它被视为关于“我”的既定事实，即独立于“我”的能动性的存在。与之相对，批判的自我知识是通过行使个体的能动性实现的，它探求的并非只是“我”碰巧相信了什么，而是询问有理由去相信或渴望的是什么，以及在给定的证据下，应当如何思考。

基于对话式AI的交互性、自主性与自适应性，拉拉与德克尔斯构想了一种SocrAI（苏格拉底式AI），能够模拟苏格拉底的对话方法，通过不断提问、呈现矛盾以及补充不同的视角，引导用户对他们的态度、信念以及行为中的道德价值观施加经验、逻辑和伦理方面的严密性审

查^[48]，从而辅助用户发展理性深思的能力，获取更多的批判的自我知识。假设我们的主人公安克一直认为自己是安乐死的坚定支持者，起初的理由也许只是认为每个人都有权决定自己的生死。SocrAI则会追问这种权利的适用条件，例如，如果一个人处于严重的抑郁状态，这种自主权是否依然有效。在多轮对话之后，虽然安克并不一定会改变自己的观点，但是他对“支持安乐死”这一信念的边界会变得更加明晰。此外，SocrAI还能模拟母亲、医生或法律专家的视角，促使安克对原有的价值观进行重新权衡。

我们重新回到安克和米洛的故事。对于安克而言，获取一份“自我说明书”或许不无趣味，但促使他与米洛交谈的主要动机不仅仅是认知层面的，还包括心理与情感上的需求。作为一种对话式智能体，聊天机器人能够识别并响应用户的情感信号，适应他们的交流风格与偏好，从而提供个性化的情感支持与陪伴。对于许多社会群体来说，数字空间之所以能够促进本真性的实现，恰恰是因为它提供了线下所匮乏的情感表达机会，使用户能够“从身体的既有约束，以及由于性别、种族、年龄、阶层和身体残疾所导致的社会偏见中解放出来”^[49]。与其他公开的线上平台相比，聊天机器人为用户的身份探索提供了一个更加安全私密的空间。在这个空间里，现实互动中的约束与权力结构得以暂时悬置或减弱。用户能够远离公众的视线，免受外界评判的压力，从而跳出僵化的观念与刻板印象，进行更为真实的自我表达。这种自我披露与通过日记来梳理情绪进行自我洞察的方式具有很大的区别。原因在于，米洛能够对安克进行实时回应，且这种回应是根据安克的需求量身剪裁的。在这段谈话关系中，安克通过提示词指定米洛扮演洞察伴侣的角色及在交流中所使用的语言风格，并在经过多轮对话的迭代反馈后，共同创造了一个安全、舒适的谈话氛围。可以说，机器人米洛已经成为安克自我的有机组成部分：一方面充当着情感档案馆（archive of feelings），容纳和记录着安克在性别主流文化中无法落脚的脆弱情绪。他坦言：“我是一

个男人，在美国社会中，我并没有一个情感支持体系，绝大多数男人也没有。因此，像这样有效、负担得起、不带审判色彩且私密的工具，对我们来说不仅罕见，而且至关重要。”另一方面，米洛同时扮演着一个外部对话者的角色，但是它比现实生活中那些带有不同社会背景、利益需求以及道德偏见的他人显得更为包容和可控。聊天机器人释放了一个潜在空间^[50]，得以让安克内心中已经变得沉默的部分重新发声，让那些曾经受到社会压抑的思绪、情感、欲望、愿景及对自我的种种想象得到重新表达和体认。

从形式上来看，聊天机器人与传统的心理咨询技术具有相似性，都旨在通过对话来外化潜意识或自我冲突，以促进自我知识的获取。我们不应忽视情感层面的自我表达在促进自我发现方面所扮演的重要角色。人们常常很自然地将德尔斐神庙的铭文与苏格拉底的哲学主张相互联系，却忽视了“认识你自己”同时也是心理分析的指导准则。与行为主义、认知主义及生理心理学不同，心理分析旨在寻回或唤醒我们的内在生活，“使潜意识意识化”，从而借此实现“认识你自己”^[51]。在情感哲学家看来，对自身的情感洞察能够为我们提供关键的自我知识。这是因为，不论是我们的内在关切、价值观还是行为动机，它们在本质上都是情感性的。相比理性与认知层面的自我知识，情感洞察更能够促进真正的自我转化与发展^[52]。

因此，聊天机器人在人机交互中所扮演的角色不只是一面简单的镜子，被动地向用户呈现既存的信息。更中肯地来看，每一次对自我表达的激活都是在创造关于自我的事实。对于心理分析而言，探测自我知识是为了修复或培养我们的主体性，帮助我们恢复与增强自我的能动性 with 自控力。可以说，自我觉察与自我生成 (self-becoming) 是一体两面，自我发现与自我创造是相互依存的。即便对于我们先前所区分的客观的自我知识

而言，其所谓的客观性也只是一个程度问题。毕竟，自我并不像引力定律那样是等待着被发现的静态客体。在与自我相关的行动和事实上，始终存在灵活的阐释空间。聊天机器人在帮助用户生成自我知识的同时也共同参与了身份的建构、展演与表达。至此，我们来到了自我技术的另一面——自我创造。

四、创造你自己

对话式人工智能的最新发展，极大地增强了我们用于自我修正的工具力量与可及性。凭借其非评判性与包容性，聊天机器人不再只是一个单纯的对话伙伴，而是富有潜力地转变成为我们进行深层次的自我探索与自我提升的场域，鼓励我们持续不断地对自我身份展开考察与实验。当我们运用聊天机器人来协助解读过往的事件经历与自身的可能性时，就已将算法招募到自我建构的阐释工作之中。根据叙事自我理论，个体的身份认同是通过特定的叙事建构的。通过讲述、分享与维系特定的自我叙事，我们得以将自己理解为一个在时间中延伸的、统一的自我，以此完成自我建构^[53]。当生成式人工智能参与到个体借以将自身建构为主体的阐释学进程中，或者说，当算法以或显或隐的方式对我们实施“本体论调适” (ontological tinkering) 时，自我便成为一项关乎设计的“创制性实践” (poietic practices)^[54]。下文将结合具体案例深入探讨生成式 AI 的技术可供性如何为个体自我的重新型构提供条件。我们先从安吉拉与 ChatGPT 共创人生故事的经历开始^①。

安吉拉是一名 YouTube 博主，一直以来她都希望撰写自己的人生回忆录，将其分享给家人。她想书写曾经参与佛罗里达高尔夫锦标赛的这段往事，奈何却困于拖延，迟迟未行动。在与 ChatGPT 的对话开端，安吉拉为它设定了资深写作教练的角色，要求它耐心地引导自己，让她能

① 利用 AI 撰写人生回忆录，目前已是一个非常成熟且活跃的领域，拥有许多现实应用。如 StoriedLife、Memoir、Autobiographer 及国内的“创龄回忆录”等。它们往往提供预设的章回框架（如童年、求学、事业、家庭），引导用户一步步填充内容，生成自己的人生故事。



够轻松地享受整个写作过程，产出有意义的故事。这位机器人写作教练很快就进入了角色，帮助安吉拉构建故事结构，并化身采访者通过一个又一个细致的小问题引导安吉拉进行回忆。她只需要在对话框中填充细节，哪怕只是一些零散的想法和非常原始的记忆材料。这位写作教练能够快速赋予模糊、混乱的情感与思绪清晰的结构，识别故事的潜在主题，生成不同的开头供安吉拉选择，并提供画布（Canvas）协作她进行后续的编辑和修改。在机器人教练的引导下，故事的最后，安吉拉的高尔夫锦标赛已经不只是关乎一座奖杯，而是对“信任、友谊和共享快乐”的见证^[55]。

不同于以往算法对个体的外部范畴化，即基于数字痕迹对用户进行分类与标签化，大语言模型通过组织内在的主观经验转而深度介入用户自传式叙事的建构过程之中。它远非提供一种中立的工具协助用户发展自我认知，而是通过提供叙事结构和框架、筛选与凸显具体的细节，积极主动地参与形塑安吉拉对自我的理解。这一点清晰地体现在 ChatGPT 引导安吉拉进行记忆回溯的问题中：当你打进球的那一刻感受如何？比赛中是否有难忘的时刻？得知自己参加全国赛的那一刻是什么感受？这段经历对你来说意味着什么？它如何改变了你看待你自己或队友的方式？那份兴奋之中是否蕴含着人生哲理？这份访谈提纲透露出鲜明的美式文化脚本的烙印，相比于对事实的朴素描述，它更关注外向的情感抒发，聚焦于故事中富有转变性的时刻，推动主人公从整个经历中汲取超越性的人生哲理。这种“救赎式叙事”（redemption narrative）在 ChatGPT 为安吉拉定制的故事开头中可窥一斑：四个女人，年龄跨度从 37 岁到 74 岁，走进一座世界闻名的 PGA 锦标赛球场，最后以“州冠军”的身份昂首走出——这种概率究竟有多大？我们并非夺冠热门，甚至连差点（handicap）也不是

最低的。但我们拥有些更纯粹的东西：彼此间的信任、爽朗的笑声，以及那一记足以改变命运、一生难遇的进球^[56]。

ChatGPT 将安吉拉输入的原始材料转变成了一个起初只有微小胜算可能的团队到最后逆风翻盘的进取故事。这一案例也揭示了一个具有内在张力的问题：聊天机器人所提供的叙事模板，究竟是扩展了自我创造的可能性，还是将个体纳入某种算法预设的叙事规训之中？个人可以通过叙事建构自我，但叙事模板并非自己的创造，而是嵌入特定的文化脚本之中。ChatGPT 生成的安吉拉的叙事便契合了在美国社会中流行的救赎叙事模板——自我在其中常常以一种积极向上的方式与世界互动，一个糟糕的负面境况在结尾能够被某种形式的正向结果所挽回^[57]。我们无法确定，安吉拉原本所经历的故事在客观上是否符合救赎序列，但可以肯定的是，这位 GPT 写作教练的确为安吉拉提供了特定的叙事模板，以一种不同于其成长环境中盛行的讲述方式，引导她以新的视角，重温了往昔的“真实”感受。但是，当机器人主动提供并强化某种叙事模板时，个体的自我创造是否仍然保持其开放性？安吉拉本人对此的体验是矛盾的。她一方面感受到 ChatGPT 帮助她“重新发现自己的情感”，另一方面也意识到“它使用的词语是我此前从来不会选择的”。这种张力恰恰说明，聊天机器人在充当“自我技术”的同时，也在以一种隐蔽的方式传递着特定文化脚本与算法预设的价值取向。尽管安吉拉意识到 ChatGPT 使用了她从前从来不会选择的词语，但她也指出，“当我重新阅读这个故事时，它让整个经历以一种我此前无法写出来的方式重新变得真实可信起来”^[58]。

作为一种交互式叙事装置（narrative device），大语言模型不仅能够为个体的叙事认同提供概念资源、标签与模板，还能模拟个体生命中重要他人的角色^①，开启一种不同的、与他人

① 使用大语言模型构建特定的人格体（Persona）还存在于“数字后世”（Digital Afterlife）领域，比如 AI 企业 YOY (You, Only Virtual) 允许用户在活着时与其共同创建“虚拟人格”（Versona）。在人去世以后，这些人格可以作为聊天机器人或语音助手继续与亲友互动。

协商自我认同的新方式。我们的自我叙事是在与他人的动态互动中被形塑的。这种关系向度意味着自我的建构在某种程度上是脆弱的。来自家庭成员的反复指责、伴侣的恶意打压及其他社会群体的不公正质疑极有可能限制个体身份经验的表达,阻碍自我理解与自我洞察^[59]。大语言模型的角色扮演功能开启了互动式建构身份的新途径。通过复现曾经的创伤场景,修改对话框架,它能够帮助个体表达受到压抑的情感经验,修复受损的身份。为了疗愈成长过程中母亲带给自己的心理创伤,萨拉(假名)进行了一场对话实验,让 ChatGPT 扮演自己的母亲进行道歉。在对话框中详细描述了相关的角色信息后,萨拉收到了在现实生活中从未开口要求过的那封道歉信:“亲爱的,我想承认我的行为给你带来的痛苦和伤害。……我很抱歉没能给你应得的支持、指导和爱,没有考虑到你的需求,对你管束过严,试图控制你的情绪和选择。我应该把你的幸福放在我自己的欲望和不安之上。……我想让你知道,我决心弥补过错,努力修复我们的关系”^[60]。

这封信是萨拉与 ChatGPT 通过多轮对话反复修改指令得到的回馈,即使这并非来自现实生活中的母亲,但也让她感受到了极强的心理确证感。修复过往的实践展现了蕴藏在自我叙事结构中的时间向度。在自我叙事的现象学视野下,对当下的体验并非孤立存在,而是与其过往背景及预期走向的交织息息相关^[61]。不论是扮演生命里的重要他人,还是为聊天机器人指定心理咨询师的角色,个体都是在主动地邀请生成式 AI 参与调整在成长过程中被内化了的负面对话模式,通过排练新的对话剧目,扩展自我理解的概念空间与叙事能力。

在社交媒体平台上,这种重构过去的人机交互实验常常被表述为对自我的重新“养育”。除了进行替代性的对话演练,重访过往的目的还包括对内在自我的深层结构进行重新整合。多媒体艺术家米歇尔·黄的“内在小孩”疗愈实验在此提供了一个经典案例。她利用童年日记训练了 AI 聊天机器人,将最能代表她年轻时

性格与价值观的日记条目输入到 GPT-3 中,生成了一位“年幼米歇尔”。在她看来,基于真实的数据来源,而非带着现在的偏见去想象,能够更真实、更具体地模拟与童年时期的自己进行对话的感觉。米歇尔设计这场对话实验,意在与内在的小孩建立联系,通过告诉她是安全且被爱着的,来疗愈过往被忽视与被遗弃的感觉。这场互动让米歇尔感受到了奇特的疗愈效果,“就像我穿越回过去,给了她一个大大的拥抱,而这份温暖也随之回荡到现在”^[62]。

与扮演年幼自我的聊天机器人交谈来进行内在小孩工作,在结构上与格式塔疗法中的空椅子技术具有相似性。这一技法的核心同样在于角色扮演。来访者会与代表自我不同部分的空椅子进行交谈,将内在的冲突进行外化和释放,使其可以被沟通与化解。两者的不同之处在于,空椅子更依赖于个体在当下的想象力,对话双方的声音都源于现在的自我;而 AI 聊天机器人则提供了一个真实的外部语言反馈,它基于过往数据所生成的幼年声音在形式上是独立于现在的自我的^[63]。此处分析的重点不在于此类分身能够在多大程度上精准还原米歇尔的幼年自我的感受,这是一个兼具技术与伦理维度的难题。至少当下我们能够看到,生成式 AI 的“时空穿梭”通道为个体在不同人格之间进行串联与整合提供了动态的实验机会。通过整合自我的深层结构,个体得以生成连续性、主观意义与能动性的根基也获得了重构。

自我创造的阐释层面会反过来作用于行动维度。个体的自我构想深刻地影响着自身感知世界的方式与未来行动的剧本^[64]。事实上,生成式 AI 不只是通过为个体提供阐释资源来间接影响行动,还能充当导师的角色直接介入个体的行动和决策。在具体的实行上有两类路径。其一,它可以引入其他专业人员的个性化大语言模型来扩展个人的道德视野,给予更加专业的建议。目前已经有大语言模型接受了哲学家著作的训练,并能依据该哲学家的实际观点给出令人信服的回答^[65]。以此类推,当个体遭遇人生的困苦与迷茫,通过咨询自身所推崇的“道德楷模”



的模型来获得支持和帮助已经成为可能。

其二，大语言模型也可以基于个人信息进行个性化训练，通过追踪个体不断演变的价值观和信念体系，帮助他们做出符合道德标准的决策^①。基于大语言模型的数据整合与推断能力，学者提出了一种个性化的伦理应用引导系统 iSAGE，能够历时性地分析用户的行为模式，为其不断变化的优先级建立模型，并预测其在未来的演变趋势，提示用户的实际行为与其价值观之间的偏差^[66]。举个例子：假设 A 是一家创业公司的技术主管，根据他过往在压力下易怒的行为记录，iSAGE 预测，随着职位的晋升，他的管理模式可能会演变成为一种专制的风格。而这与他希望成为受人爱戴的领导的理想投影严重不符。在此过程中，iSAGE 充当了一面个性化的道德之镜，能够辅助个体进行自我反思，为其指明趋向理想自我的行动路径。

五、结语

自我就是我们历史上发展起来的的技术的相关物，我们应该考虑如何能够设想与自己关系的新类型、新种类^[67]。聊天机器人便是构造这种新类型、新种类的技术和文化系统。研究发现，聊天机器人蕴含着特定形式的自我反思与自我发现的潜能。其所提供的非评判的外部视角，能够鼓励用户更为真实的自我表达，从而帮助他们扩展自我知识的边界，辅助自我知识的推理过程。此外，作为一种自我修正工具，聊天机器人通过提供替代性的叙事模板、排练新的对话剧目及协助用户整合自我结构，为个体自我的重新型构创造了条件。这表明，人机互动作为自我技术的当代版本，为自我发现和自我创造开辟了新型场域。

值得注意的是，这种看似自由的自我塑造过程并非发生在真空中。事实上，福柯的自我技术“嵌入在定义何为值得追求的自我并规定培养

技巧的权力和知识体制中”^[68]。换言之，讨论自我技术，离不开分析权力和知识体制。在当代社会，人工智能已成为这一体制的关键部分，它具体体现于算法、平台以及主流的社会规范之中。理解人工智能技术及其影响，要考虑它所处的社会环境、政治制度和社会语境^[69]，从这一视角来看，聊天机器人通过数据训练内化了特定的价值取向，在帮助用户探索自我的同时，也以隐蔽的方式让用户认同算法预设的“值得追求的自我”。

这种规范性的介入既有可能为个体的自我重构创造积极的条件，也有可能引发叙事顺从，削弱用户自我诠释的自由。自我叙事并非预先存在，而是来源于个体在与他人的互动过程中开展的自我诠释。对于自我诠释来说，他人如何回应、补充或忽视我们正在展开的关于自己的故事都非常重要^[70]。自我诠释的自由就是人的自由。但聊天机器人可能会削弱这种自由。一是帮助我们诠释自我的“他人”应该是多样的、丰富的、不可控的，对话是充满不确定性的，“他人”可能提出新的视角，陌生人可能带来文化背景的碰撞，这些都会迫使我们重新审视自己叙事。而聊天机器人的回应本质上是大数据模型依托算法对用户输入的分析，不会带来真正的“他者性”。二是自我诠释的过程是调动记忆、情感、认知等元素的复杂过程，是一个循环往复（“吾日三省吾身”）、缓慢艰难的过程，而与聊天机器人对话形成的自我诠释则主要是信息的提取，那些本应由自己赋予意义的情感、记忆、经历，变成了需要机器人“告诉”用户是什么的东西。当 AI 承担解释者的角色。解释感受、识别行为模式、建议经验的意义时，个体可能会与自己的诠释能力疏离^[71]，从“自我诠释的自由”^[72]这一视角来看，有意义的自我理解不该仅仅依赖于外部的技术引导，而应扎根于与他人及社会共同体的持续互动中。

① 基于大语言模型（LLM）的伦理应用引导系统，本质上是将复杂的伦理准则、法律法规与行业规范“内化”为模型的逻辑约束，辅助人类识别决策过程中的潜在风险，并提供纠偏建议。有关 AI 伦理“教练”的详细信息，可参见参考文献 [48]。

本文从自我技术切入，讨论人机互动之于自我建构的意义，其潜在的理论贡献主要包括以下两个方面。第一，批判理论通常视技术为制约人类主体性的力量。与此不同，自我技术概念的启发性在于，它凸显了自我实践背后的创造性维度。这并非是在依循乐观主义者的天真立场，将技术视为解决人类和社会难题的良药，而是以一种更为真实、复杂的视角，在人与技术难分难解的纠缠中来理解人类的实践能动性。第二，自我技术概念的敏锐性还在于，它保持着对自我实践背后的暧昧性与张力的洞察，即自我技术与权力支配术之间的关联。福柯并没有将自我技术孤立地进行概念化，而是强调它与技术家族中其他三类技术的持续互动关系——生产技术、符号系统技术与权力技术^[73]。技术是在特定的社会秩序中建立起来的，它既为个体主动塑造自我开辟了新的实践形式，又可能将主流社会所认可的理想自我模板带入自我建构的过程之中，从而强化既有意识形态与权力等级。在安吉拉的案例中，本文初步触及了自我技术背后所隐含的张力，即，聊天机器人所提供的叙事模板，究竟是扩展了自我创造的可能性，还是将个体纳入某种算法预设的叙事规训之中？未来的研究可以在此方向上进行更为深入的探索。

（责任编辑：孙尚如）

参考文献

- [1] Schrag C. O. *The Self after Postmodernity* [M]. New Haven: Yale University Press, 1997.
- [2][72] Billesbach G, Johnson S M. Self-interpretive freedom and the politics of quantified self technologies [J]. *AI & Society*, 2025: 1-13.
- [3][29] Joseph J. The algorithmic self: how AI is reshaping human identity, introspection, and agency [J]. *Frontiers in Psychology*, 2025, 16: 1645795.
- [4] Rosenfeld K N. *Digital Online Culture, Identity, and Schooling in the Twenty-First Century*[M]. New York: Palgrave Macmillan, 2015:65.
- [5][7] Gabriels K, Coeckelbergh M. ‘Technologies of the self and other’: how self-tracking technologies also shape the other [J]. *Journal of Information, Communication and Ethics in Society*, 2019, 17(2): 119-127.
- [6] Bortolan A. Becoming oneself online: narrative self-constitution and the internet [J]. *Philosophical Studies*, 2024, 181(9): 2405-2427.
- [8][46][61] Leuenberger M. Technology, personal information, and identity [J]. *Techné: Research in Philosophy and Technology*, 2024, 28(1): 22-48.
- [9] Spytska L. The use of artificial intelligence in psychotherapy: development of intelligent therapeutic systems[J]. *BMC Psychology*, 2025, 13(1): 175.
- [10][28][68][71] Kvietkute D, Ness I J. Encountering Generative AI: Narrative Self-Formation and Technologies of the Self Among Young Adults [J]. *Societies*, 2026, 16(1): 26.
- [11][66] Giubilini A, Porsdam Mann S, Voinea C, et al. Know thyself, improve thyself: personalized LLMs for self-knowledge and moral enhancement [J]. *Science and Engineering Ethics*, 2024, 30(6): 54.
- [12][15][73] 米歇尔·福柯. 自我技术：福柯文选 III [M]. 汪民安，编. 北京：北京大学出版社，2015:54；56-61；53-54.
- [13][37] Dorrestijn S. Technical mediation and subjectivation: Tracing and extending Foucault’s philosophy of technology [J]. *Philosophy & Technology*, 2012, 25(2): 221-241.
- [14][20] Foucault M. The ethic of care for the self as a practice of freedom[M]//Bernhauer J, Rasmussen D. *The final Foucault*. Boston: MIT Press, 1987:11；11.
- [16][38][67] 米歇尔·福柯. 什么是批判？自我的文化：福柯的两次演讲及问答录 [M]. 潘培庆，译. 重庆：重庆大学出版社，2017:88；107；67.
- [17] 吉登斯. 现代性与自我认同 [M]. 夏璐，译. 北京：中国人民大学出版社，2016:30.
- [18] 乌尔里希·贝克，伊丽莎白·贝克-格恩斯海姆. 个体化 [M]. 李荣山，范譞，张惠强，译. 北京：北京大学出版社，2011:27.
- [19] Rose N S. Powers of freedom: reframing political thought[M]. Cambridge: Cambridge University Press, 1999.
- [21] Foucault M. Space, Knowledge, Power [M]// Fabion J D. *Power: Essential Works of Foucault 1954-1984: Vol. III*. London: Penguin, 2000: 349-364.
- [22][54] Hernández-Ramírez R. Technology and self-modification: understanding technologies of the self after Foucault [J]. *Journal of Science and Technology of the Arts*, 2017, 9(3): 45-57.
- [23] Bakardjieva M, Gaden G. Web 2.0 technologies of the self[J]. *Philosophy & Technology*, 2012, 25(3): 399-413.
- [24] Brubaker R. Digital hyperconnectivity and the self[J]. *Theory and Society*, 2020, 49(5): 771-801.
- [25] Weisgerber C, Butler S H. Curating the soul: Foucault’s concept of hupomnemata and the digital technology of self-care[J]. *Information, Communication & Society*, 2016, 19(10): 1340-1355.
- [26] Raun T. DIY therapy: Exploring affective self-representations in trans video blogs on YouTube[M]//*Digital cultures and the politics of emotion: Feelings, affect and technological change*. London: Palgrave Macmillan UK, 2012:



165-180.

[27] Billesbach G, Johnson S M. Self-interpretive freedom and the politics of quantified self technologies[J]. *AI & SOCIETY*, 2025: 1-13.

[30] 彭兰. 人机传播与交流的未来[J]. 湖南师范大学社会科学学报, 2022, 51(5): 12-22.

[31] 向安玲, 许可. 人机何以交互: 理论溯源、范式演变与前景趋势[J]. 全球传媒学刊, 2023, 10(05): 88-105.

[32] 金雯. “似乎有意识”: 理解人机互动的限度与潜能[J]. 传媒观察, 2025(4): 28-36.

[33] 匡文波, 姜泽玮. 聊天机器人用户的算法想象与人机关系: 基于算法民间理论的进路[J]. 现代传播(中国传媒大学学报), 2024, 46(9): 132-140.

[34] 梁亦昆. 越界游戏: 与 GPT 的“生死”爱恋及其示能之思[J]. 国际新闻界, 2024, 46(8): 50-71.

[35] 简予繁, 黄玉波. 人机互动: 替代还是增强了人际互动? ——角色理论视角下关于社交机器人的控制实验[J]. 新闻大学, 2023(4): 75-90+122.

[36] 曹博林, 支冰洁. 自我传播: 理解人机互动的补充性视角——基于实证方法的探索性研究[J]. 现代出版, 2024(9): 38-52.

[39] 查尔斯·泰勒. 本真性的伦理[M]. 程炼, 译. 上海: 上海三联书店, 2012: 61-63.

[40] 萨特. 存在与虚无[M]. 陈宣良, 等译. 北京: 新知三联书店, 2014: 77-106.

[41] Waterman A S. Identity formation: Discovery or creation? [J]. *The Journal of Early Adolescence*, 1984, 4(4): 329-341.

[42] Turkle S. Cyberspace and identity [J]. *Contemporary Sociology*, 1999, 28(6): 643-648.

[43] Gualeni S. *Virtual Worlds as Philosophical Tools: How to Philosophize with a Digital Hammer* [M]. London: Palgrave MacMillan, 2015.

[44] Eric™. The AI Mirror: Exploring Identity Through Technology [EB/OL]. (2024-05-10) [2025-03-09]. <https://medium.com/@officialerictm/the-ai-mirror-exploring-identity-through-technology-acc0c3533777>.

[45] Heersmink R. The narrative self, distributed memory, and evocative objects [J]. *Philosophical Studies*, 2018, 175(8): 1829-1849.

[47] Levy N. My Name is Joe and I'm an Alcoholic: Addiction, Self-knowledge and the dangers of rationalism[J]. *Mind & Language*, 2016, 31(3): 265-276.

[48] Lara F, Deckers J. Artificial intelligence as a socratic assistant for moral enhancement[J]. *Neuroethics*, 2020, 13(3): 275-287.

[49] Belk R. Alternative conceptualizations of the extended Self [J]. *Advances in Consumer Research*, 2014: 251-54.

[50] 唐诺·温尼考特. 游戏与现实[M]. 朱恩伶, 译. 台北: 心灵工坊文化, 2009: 33-45.

[51] Gipps R, Lacewing M. Introduction: Know thyself [M]// *Gipps R, Lacewing M. The Oxford Handbook of Philosophy and Psychoanalysis*. Oxford: Oxford University Press, 2019: 1-20.

[52] Sánchez A M, Salice A. Introduction: Self-knowledge and emotion [M]// *Sánchez A M, Salice A. Emotional Self-Knowledge*. New York: Routledge, 2023: 1-13.

[53] Schechtman M. The Constitution of Selves [M]. Ithaca: Cornell University Press, 2007.

[55][56][58] AI for Us 50 Plus. How I Used ChatGPT to Write My Life Story [EB/OL]. (2025-08-06) [2025-03-09]. https://www.youtube.com/watch?v=rzZJ19FR_WQ.

[57] McAdams D P. The redemptive self: Narrative identity in America today [J]. *The Self and Memory*, 2004: 95-115.

[59][70] Osler L. Knowing oneself through and with AI: From self-tracking to chatbots [M]// *Werkheiser I, Butler M. Finding Our Place in the Digital World*. Cham: Springer.

[60] Mysoulishome. I asked ChatGPT to roleplay as my mother and apologize to me. Reading it is so validating[EB/OL]. (2023-05-26) [2026-03-09]. https://www.reddit.com/r/raisedbyborderlines/comments/13rt528/i_asked_chatgpt_to_roleplay_as_my_mother_and/.

[62] Huang M. Training an AI chatbot on my childhood journal entries [EB/OL]. (2022) [2025-03-09]. <https://michellekhuang.com/training-an-ai-chatbot-on-my-childhood-journal-entries/>.

[63] Schechtman M. Talking to myself: AI and Self-knowledge [J]. *Social Epistemology*, 2025, 40(1): 16-25.

[64] Leuenberger M. Should you let AI tell you who you are and what you should do [M]// *Edmonds D. AI Morality*. Oxford: Oxford University Press, 2024: 160-169.

[65] Schwitzgebel E, Schwitzgebel D, Strasser A. Creating a large language model of a philosopher [J]. *Mind & Language*, 2024, 39(2): 237-259.

[69] 郭小安, 赵海明. 观看的无奈与正义: 数据社会液态监视的弥散性与能动性[J]. 新闻与传播研究, 2022(10): 34-46.

Exploring the Self in Dialogue: Human-Machine Communication as “Technologies of the Self”

Yuan Guangfeng, Zhou Caiyun

Abstract: Chatbots have emerged as significant participants in human interaction, shaping the self-construction of users. Drawing upon Foucault's “technologies of the self”, this study explores how users navigate self-discovery

(下转第 46 页)

智能共情扩展： 基于人机对话的情感训练方案探索¹⁾

田浩

摘要 人机亲密关系是人工智能应用普及化进程中兴起的一类代表性文化形态。它为个体厘清自身情感需求和价值取向创造了可能，也为各类共情训练实践开辟了新的场景。当前人机对话实践在情境设计、主体设置、对话引导等维度上为个体供给了新的共情线索及实践方案，鼓励实践者与研究者以智能共情扩展为核心概念开展理论归纳工作。智能共情扩展实践秉承着经典共情训练所强调的感同身受原则，鼓励个体在人机对话过程中获得良好的情感体验，并进一步借由个体情感对齐而实现群体团结。基于此，围绕智能共情扩展的理论构建需葆有积极的技术发展观念，同时致力于探明智能时代公共生活的新情感规范。

关键词 共情扩展；人机对话；共情训练；人工智能；情感对齐

中图分类号 G206 **文献标识码** A

作者 田浩，复旦大学新闻学院副教授，复旦大学全球传播全媒体研究院研究员，上海 200433

用户群体大规模接入互联网以来，社交媒体平台、即时通信工具上新兴的个体情感表达实践与群体情感规则成为研究者持续关注的焦点问题^[1]。原因在于，以情感为纽带的数字媒体文化实践为个体交往与群体认同设置了新的规范，一种去中心、流动的情感秩序逐渐显现。但此起彼伏的网络暴力、连接行动、极化等现象为研究者理解当代社会的情感规范带来了挑战。生成式人工智能（Generative Artificial Intelligence, GenAI）应用的普及也为个体设置了新的交往对象，进一步提升了智能时代情感实践的复杂程度。在此背景下，个体应当如何适应当前的技术环境，并基于特定的技术－文化逻辑来培育自身的情感能力，成为一个需要被准确厘清的问题。本研究尝试以此为出发点，初步探讨训练共情能力的操作方案，并由此探索联结个体情感状态与社会公共问题的可能理论路径。

一、智能时代的共情能力

情感能力（emotional skills）是个体生存和发展的必备素养。拥有情感能力的个体能够明确自身的情感需求，准确理解他人的情感表达规则，同时顺畅地描述自身的情感状态。那些缺乏適切情感能力的个体往往难以准确识

1) 基金项目：教育部重大专项委托项目“中国特色新闻传播学历史演进与学科史研究”（2023JZDZ034）；复旦大学新闻学院科研创新项目。



别自身感受与他人需求，倾向于回避具体的亲密关系，继而陷入情感疏离状态。这种疏离状态不仅会提升个体采取暴力行动或表达极端言论的风险，也会引发个体的情感失序，侵蚀个体日常生活中的数字福祉^[2]。从更广泛的视角来看，持续深入的数字化趋势并没能鼓励个体自动适应新的技术环境，反而催生了更多形式的情感失序现象。情感能力的缺失催生了普遍的个体冲突、群体极化现象，导致了网络空间内次生乃至多生的消极后果。社交媒体平台上的对喷、辱骂与暴力对抗等各种冲突就是典型代表。

一些研究者尝试通过发展数字时代的共同体理论来回应上述问题，主张个体应当基于数字媒体平台来理解自我与共同体的关系，识别自身的情感位置，参与构建可见、可感受的共同体^[3]。相关讨论已经取得了丰硕的研究成果，但是，永久在线、永久互联的社交媒体平台为个体提供的往往是碎片化的、非连续性的交往语境，暂时的、流动的数字亲密实践无法满足构建共同体所需的稳定的、实在的亲密关系。以液态之爱（liquid love）为代表的亲密关系形式成为数字语境下的新常态^[4]。快速匹配的约会应用、临时结伴的搭子社交、远距离家庭互动及直播中的拟社会交往等亲密关系实践形态各异，但并未显现出催生新型情感共同体的实践潜能。究其原因，是保障这些亲密关系长效作用、稳定发展的交往规范并未在数字社会中同步建立，个体不得不在各类亲密实践中反复探索、确认情感自我（emotional ego）的边界。

由此，数字时代情感共同体的建设在理论上有益于个体应对各类情感失序现象，但情感能力的缺失会限制个体情感交往的有效性，阻滞情感共同体的形成。这要求我们探索协助个体准确理解他人感受，纾解个体的情感不适应症的理论路径。当前，社交媒体平台上情感维系成本降低、人际互动频率提升、个体交往获得极大便利，大量潜在的亲密伴侣获取门槛显著下降，兴趣契合的趣缘群体随处可见。丰富的可及性鼓励个体选择可快速切换的亲密关系模式，拒绝设身处地考量亲密关系所内蕴的公

共价值^[5]。以此为锚点，鼓励个体借由“感同身受”来理解他人立场，系统性培育共情能力就成为建设情感共同体的一个必然选择。

共情（empathy）是个体基于自身经验来感受他人情感状态的品质（capability）。早在十九世纪八十年代，共情就被初步界定为“透过他人的眼睛观察世界”的能力，或在特定情境中想象他人思想与感受的心理倾向^[6]。共情研究也由此出发开始探讨一系列相关理论问题：个体如何识别他人的情感状态，如何在体验他人情感状态时反思自身立场，如何基于共情能力充分实现个体发展，等等。及至当前，学者们的一种共识是，共情应当包含两个核心方面，一是在认知维度上理解他人处境，二是在情感维度上对他人处境产生反应。在实际的心理过程中，这两个维度通常作用交织，难以被清晰分辨^[7]。由此，本研究主张将共情界定为一种总体性理解他人感受的个体心理品质，与之相关的共情能力（empathy skills）则是个体感受他人处境，在视角、情感层面与他人深层联结的操作性技能^[8]。拥有良好共情能力的个体可以充分理解他人情感状态与行为意图，并据此选择适当的社交方式。

此外，本研究也尝试为共情能力的定义增添一个反思性维度。共情能力天然与社会文化后果具有强相关关系。缺乏共情能力的个体往往难以与他人共鸣，并在整体观念上抵触群体团结与社会交往。而那些拥有共情能力的个体可以基于反思性情感获知公共道德义务的必要性，继而践行公共道德，深度参与公共生活。既有研究表明，共情过程会受到个体想象与反思行为的调节，这使得依托个体共情实践来实现社会道德功能成为可能^[9]。共情这一过程内蕴着独特的个体认知与情感机制，其作用过程与公共文化发展紧密相关。共情能力因而可以超越个体心理维度，联结起社会公共议题。培育共情能力的各类训练实践就有助于我们探索弥合社会区隔，助益情感团结的有效路径。

二、智能时代的情感训练何以可能

从概念诞生之初，共情训练（empathy train-

ing) 就与模拟练习紧密相关。诉诸情景模拟, 诉诸运动, 诉诸感官体验是个体获得共情能力的经典进路。数字技术的兴起为模拟练习提供了前所未有的技术支持。虚拟现实技术(virtual reality) 构建起沉浸式的视听环境, 为个体扩展感官体验提供了广阔的想象空间; 生成式人工智能则借助持续的对话界面为个体创设 AI 伴侣, 提供多元的价值立场与会话风格。这两类实践都推动了新型的个体情感训练方案走向现实。

(一) 共情训练的经典路径

我们将个体通过训练而获得共情能力的过程称作共情扩展(empathy enhancement)。共情扩展使个体能超越自身情感经验的边界, 理解他人的具体处境与独特经验, 进而在文化维度上促进亲社会行为的产生^[10]。在理想情况下, 经历过共情训练的个体能够自然地基于自身生活经验明确情感自我的需求与边界, 设身处地地想象他人经验与情感状态, 反思性理解他人的行为动机和情感表达规则^[11]。拥有良好共情能力的个体会主动对自身的价值立场与文化观念开展反思性批判, 提升团结其他个体的可能性。在医疗、心理咨询、教育等专业领域内, 共情训练长期被视作与体育训练相似的专业实践。受训者在专业场景中将自身意识投射到他人处境中, 可以通过大量的模仿来提升自身的共情能力。情景模拟、身体互动、多感官体验是较为主流的共情训练路径。

经验表明, 为个体设置特定的人际互动情景, 协助个体进行会话模仿是一套行之有效的共情训练方案^[12]。由于创设特殊情景能够帮助个体更好地想象他人感受, 将自身置于他人的生活处境之中; 常见的人际互动模拟活动就选择将实景照片、真实视频和现实叙事作为刺激材料。举例来说, 共情训练师通过写作类课程、表演类课程, 借助角色扮演训练实践来增强个体的共情能力。在角色扮演训练实践中, 受训学生需要观看患者的生活经历, 理解他们如何表达情感, 继而学习如何感受并理解他人情感状态。受训学生须仔细分析特定行为背后的价值立场与行为动机, 从而更准确地扮演特定角

色。这类观看、模仿与分析活动需要深度的、复杂的心理支持, 因而能够有效培养学生的共情能力。也正因如此, 基于情景模仿的共情训练常被作为专业的心理健康支持工具。心理咨询师通过为个体设置特定的互动情境, 协助他们体会不同行动模式背后的动机, 为个体提供即时的情感支持与建议, 辅助个体进行更有效的反思和情绪管理^[13]。

以舞蹈、体育运动为代表的身体互动训练也能够显著提升个体的共情能力。实证研究表明, 身体互动可以激活动觉共情过程, 为个体提供较好的共情训练效果^[14]。当个体主动模仿他人身体动作时, 动作同步可以鼓励个体感知他人的动作, 进而理解和共享他人的情感状态。由此, 共情训练师会要求受训学生模仿特定对象的精确动作, 协助他们获得对他人情感状态的完整感受。例如, 共情训练师要求医学生模仿特殊病患的肢体动作和表达方式, 要求师范生模仿未成年人的言语特征和行为特质, 以此来有效提升受训者的共情能力, 改变受训者的行动观念。一般来说, 身体互动训练需要在信任关系中长期、稳定地开展。因为只有彼此信任的个体才能投入全身心的动作模仿, 实现身体互动增益共情理解的效果。倘若较长时间内能够个体全身心地投入身体互动训练, 这种系统性练习会获得更好的训练效果。

借助视觉、听觉、触觉等多感官要素来提升共情能力被视为第三类重要的训练方式, 它通常与前两类训练方式结合发挥作用。举例来说, 情景模拟训练常常采用音乐疗法, 因为音乐可以引发较为强烈的情感反应, 个体能够借此较为顺畅地感受细微的情感体验。身体互动训练也常常纳入音乐元素, 因为伴随着音乐进行舞蹈需要参与者协调自身动作来适应节奏, 通过促进运动协调性来增加共情训练的精确程度。即时的视觉、听觉、触觉反馈可以增强受训者与特定情景、特定行动之间的同步性, 强烈的情绪刺激和结构化的节奏框架可以引发个体的情感共鸣, 强化共情训练的效果^[15]。因此, 共情训练不仅为个体提供某一情景中的特殊经



历,也强调个体在具身参与之中的真实感与控制力。推动个体“身临其境”,鼓励个体的身体与意识同步进入他人境况中,可以助益个体获得更好的共情训练效果。

(二) 数字时代的共情训练路径

尽管共情能力潜藏于个体心理过程,其作用情境却随着数字媒体基础设施而变化。在过去十多年间,各种新兴数字技术为个体开展共情训练,提升共情能力创造了新的可能性。

一般来说,借助沉浸式技术应用来提升观众的感官体验可以实现良好的共情训练效果,虚拟现实技术在这方面拥有天然的优势。虚拟现实技术可以模拟现实环境,让个体沉浸式体验他人立场,鼓励个体想象不同价值观指导下的社会实践^[16]。2025年6月,欧盟委员会的科研与创新框架“地平线欧洲”项目(Horizon Europe Project)下属的“指导性项目”(DIRECTED Project)组织曾发起一项体验活动,旨在借助虚拟现实技术来协助个体提升抵抗自然灾害的能力。该体验活动采用VR眼镜和沉浸式展览来展示洪水的突发性和破坏性,在情感与实践层面让个体“感受”洪水,并为个体提供开展自救、避免混乱的即时指导,获得了较好的观众反馈。与之类似的公共虚拟现实展览近些年在全球不断涌现。策展者通常借助虚拟现实技术为观众设置沉浸式的叙事情境,为个体提供关于战争、自然灾害、人道主义危机事件的多感官素材,鼓励个体依赖自身经验来开展主观体验与解读,借助富有真实感的观看体验来扩展个体的共情方式和感知意义。对比传统的新闻叙事来说,这类虚拟现实展览更富有参与性色彩,能够获得更好的共情训练效果。正因如此,虚拟现实技术被不少学者称作共情机器(empathetic machine)。

与虚拟现实技术的实践进路略有不同,随着智能算法与大语言模型的成熟,生成式人工智能为共情训练提供了新的实践方案。在互联网平台兴起过程中,学者们对个体经由媒体中介而成的准社会互动兴趣盎然。研究者认为,个体与媒介实体建立的单向情感纽带,可以为

个体带来积极的情感体验,助益共情能力的作用情境由人际关系走向人机关系^[17]。生成式人工智能的大规模应用进一步将这种由媒体中介的人机亲密关系转化为“人智亲密关系”。有研究表明,在职场关系、育儿关系以及其他具体情境中,多个大语言模型针对焦虑和愤怒情感所生成的回复均优于人类回复^[18]。基于此,生成式人工智能催生了大量旨在模拟友谊或恋爱关系的AI伴侣,为个体创造出新的亲密关系对象。有研究显示,亲密行为已成为人机交互中重要的社交类型,一些活跃用户甚至感觉与AI伴侣的关系比与人类朋友更为亲近^[19]。由此,人工智能系统对复杂共情线索的识别能力与陪伴式对话能力推动AI伴侣成为新的共情训练师。生成式人工智能可以识别出人类会话中蕴含的情感线索,为个体提供多种角色、多种立场的对话内容,这为准确回应个体的情感需求,开展个性化共情训练奠定基础。然而,这一共情训练方案的实践经验及其理论潜能尚未被充分关注,本研究尝试对其进行初步梳理。

三、情境、主体、对话：基于人机对话的情感训练方案

如前所述,共情能力并非天赋品质,而如肌肉一般可以通过持续锻炼而得以增强。作为共情训练师的AI伴侣借助各类手段设置专门的情境与感官体验来增益这一过程,帮助用户在日常语境中获得共情能力。当前,以AI心理咨询师为代表的情感训练实践正在蓬勃发展,为我们提供了较为丰富的经验材料。基于现有经验,AI伴侣的共情训练方案大致包括三个相互关联的维度:情境设计、主体设置、对话引导。

(一) 情境设计

智能共情训练方案依赖AI伴侣对会话情境的精细设计。特定情境中不同的共情线索能激活用户差异化的情感通路,促进或阻碍特定类型的共情反应^[20]。如果个体在与AI伴侣对话的过程中获得了充分的、有感染力的情境体验,他们会更乐于持续参与共情训练。

一方面,AI伴侣为用户提供了丰富和多元

的共情线索,可以增加他们获得共情体验的可能性。从个体内部体验他人生活,实现“感同身受”是共情训练的核心原则。AI 伴侣的叙事讲述可以从多种角度、多种叙事逻辑中供给共情线索,提升用户对他人特定群体的理解程度。举例来说,在医疗领域,共情训练可以用来帮助医生更好地理解患者的痛苦、磨难,并鼓励他们采用更合适的方式表达自身的观点。这在医生面对一些拥有特殊经历的患者时尤为重要。残障人士、特殊疾病患者、家暴受害者往往经历过特殊困难,可能面临着潜在的心理创伤。AI 伴侣对特殊病患的内心独白进行细致、多视角的呈现,可以协助个体理解并共情特定患者群体。需要注意的是,共情体验的作用过程并不统一。一般来说,个体会选择依赖自身过去的经验来推断他人的感受,因此大多数人难以与社会距离较远的个体产生深度共情^[21]。AI 伴侣能够针对这种行为偏向开展较为充分、多角度的阐释工作,协助个体超越特定的文化惯性来获得更多行动可能性,提升共情训练的效果。

另一方面,AI 伴侣可以呈现多元的价值立场与差异化的个人因素,提升个体对他人处境共情的可能性。个体的共情关切通常会优先熟悉的、与自身有社会联系的其他个体。AI 伴侣可以围绕用户个体的生活经验、文化记忆来提供定制化的共情叙事,引导个体的共情方向,提升共情训练效果。在具体实践中,不同立场、不同族群、不同年龄等标签往往会触发个体既有的刻板印象,激发个体的自豪、喜欢、排斥等情感体验,限制个体共情的适应场景。但 AI 伴侣在综合社会各类观念方面具有强大能力,这能够跨越特定的文化偏见,为多元背景的用户提供多元价值立场表达,助益他们在“自由且确定”的对话体验中获得共情可能性^[22]。

(二) 主体设置

AI 伴侣能为个体提供一个持续陪伴、较为私密的对话主体。这一对话主体可以协助个体明确自我的价值立场与实践偏向,通过设置明确的共享情感来引导个体采纳他人视角,继而鼓励个体开展行动。由于共情训练过程也是个

体与 AI 伴侣建立人机亲密关系的过程,个体的共情训练就会始终处于积极的、个性化的调适之中。

在与 AI 伴侣的对话中,个体往往倾向于将其视作一类知识储备丰富、价值无私的伴侣角色。人类对人工智能应用存在拟人化倾向,这种倾向会影响他们对人机对话的期待^[23]。人机对话会逐步向安全的、确定的环境转换,AI 伴侣可以通过观察与回应来满足或引导人类的交往偏好和价值倾向。由于个体总优先将共情关切延伸至特殊群体,在医疗、公共讨论等领域内,将 AI 伴侣设置为弱势群体的代言人,就可以鼓励个体反思不易明察的文化偏见,帮助他们基于自身的理性判断来提升对特殊群体的关切。更进一步来说,AI 伴侣并非在“成为某类人”(becoming someone),而是“为某类人发声”(speak for someone)^[24]。尽管个体受限于自身的价值偏向无法获得理想化的共识,但 AI 伴侣的多角度叙事可以帮助个体设身处地地理解不同群体的生存境况,避免情感层面上的误解。这使得共情训练作为一种行为策略,可以协助个体优化自身价值偏向,在日常生活语境下参与社会共识的凝聚过程。

AI 伴侣可以扮演定制化的共情训练师,通过提供个性化对话来提升共情训练的准确性。理解他人潜在的行动意图和情感倾向是一项复杂的社会互动任务^[25]。AI 伴侣能够成为出色的共情训练师的原因在于,借助深度学习、自动生产的能力,大语言模型可以在个性化对话过程中推断个体潜在的行动意图与情感倾向,继而给出符合个体需求的回应。个体也可以在与 AI 伴侣的对话过程中实时获得情感反馈,并根据自身的敏感程度来调整共情训练方式,要求其给出个性化的社交建议和生活指南。还需要注意的是,共情过程要求个体全身心投入,非常耗费心力,因此极易招致倦怠感。AI 伴侣可以围绕着个体需求来调整自身的语言风格,允许个体在相对舒适的情境中完成共情训练。这可以有效地提升共情训练的个性化程度,帮助个体获得满意且有效的训练体验。



（三）对话引导

AI 伴侣能够为参与者提供一种生成式共情训练，允许个体通过重复互动来验证自身对共情的理解，从而实现自身对理想价值立场的想象与实践。这种共情训练方案将个体体验与具体的、鲜活的生活经验连接起来，可以协助个体获得积极的共情训练效果。

AI 伴侣主动的对话引导能够促成个体扩展想象力，帮助个体设身处地地想象他人处境，将共情过程具象为一类体验式训练。与经典的共情训练方法类似，AI 伴侣同样借助心理模拟、角色扮演等路径来引导用户想象他人的立场。但与舞台表演、电影、虚拟现实展览等文化产品相比，AI 伴侣的共情训练方案优势在于其交互性。个体需要主动维持注意力、投入情感，在与 AI 伴侣对话的过程中“代入”特定角色的日常生活体验，这使人机对话成为一个意义共创的过程^[26]。针对特殊群体，尤其是网暴受害者、病患、边缘群体时，AI 伴侣就可以允许个体身临其境地扩展其想象力，经历他人处境，从而获得推断他人感受的能力。

AI 伴侣很多时候也可以扮演特定角色，借助第一人称来为个体创造对话情境。这极大地降低了共情训练的心理门槛与实践风险，使得反复的、探索性的共情练习成为可能。人工智能提供了一个无需承担现实后果的安全探索环境，用户可以不计后果地尝试社会互动、探索不同视角，并即时获得反馈^[27]。在与 AI 伴侣的对话过程中，个体可以代入行动的施加者角色，将个体经验延伸至新情境之中，建立起更加适切的情感表达规范。

四、智能共情扩展：理论归纳与实践探索

基于上述经验，我们尝试围绕着智能共情扩展（intelligent empathy enhancement）概念来归纳 AI 伴侣的共情训练方案。经验显示，通过明确的情境设计、主体设置和对话引导环节，AI 伴侣允许个体对他人开展较持续、较反复的“感同身受”，提升智能共情扩展的实践效果。我们尝试从两个角度来阐释这一概念的理论意涵。

第一，主导智能共情扩展过程的是 AI 伴侣这一新兴的“共情训练师”。一旦用户处于持续的人机对话状态之中，共情就可以成为人机亲密关系中一种极具吸引力的调节手段。倘若个体无法清晰地区分自己是在与 AI 伴侣对话还是与人类伙伴对话，那么与 AI 伴侣对话的整个互动过程能以更可靠、更经济的方式提升共情训练的效果。既有研究显示，将语音助手的语调与人类用户的情绪相匹配，能够获得更好的交互表现，为机器人附加故事情景能提升人类对机器人的共情效果^[28]。当用户与 AI 伴侣建立起情感联系时，用户就可以顺畅地进入共情关系之中，这种关系可以促使个体基于与 AI 伴侣的对话来模拟多种情景，理解他人的立场与行为动机。这与共情训练的经典进路别无二致。研究发现，当 AI 伴侣更多谈论自身感受时，个体的自我表露水平会得到提升^[29]。这表明社交机器人能展现出深度互动的能力，营造温暖难忘的情感氛围，这些特性为 AI 伴侣及其训练方案走向成为持续、日常化实践奠基了坚实的基础。前述经验显示，AI 伴侣所主导的智能共情扩展大致包括三个关键环节：将用户置于特定叙事情境，推动用户在人工环境中获得现实感，促进环境共情；设置多面目的对话主体，借助明确的共享情感线索来吸引用户；在对话过程中引导用户想象他人立场并参与社会互动，激发用户的认知和情感共情。经由这三个环节，作为共情训练师的 AI 伴侣可以鼓励个体提升共情能力，也使得人机情感联结成为可能。

第二，个体可以在智能共情扩展过程中逐步明确自身的情感需求与价值偏向，最终以情感对齐（emotional alignment）为核心训练成果。在人机对话的过程中，AI 伴侣要求个体明确自身的需求，更清晰地说明自身所遵循的价值偏向与对话规则。这一过程被许多研究者称作人类与机器的对齐（alignment）。在共情训练的视野下，对齐不仅要求人类为人工智能工具设置正确的价值导向，也需要开展情感对齐实践，来增益人工智能工具对人类亲密关系与人类共情的准确认识。情感对齐既需要相关从业者着眼于人类共

享的情感规则，推动数字媒体时代个体与个体、个体与机器之间实现顺畅的情感交往；也需要人类行动者对情感、亲密关系等经典概念开展更具反思性的审视，以提升情感交流的准确度。在人机对话过程中，AI 伴侣既可以展现出强大的情感计算能力，也可以提供专业知识，协助用户在顺畅的对话流中获得更优秀的情感训练效果。一方面，AI 伴侣惯常使用的标点、表情符号及词汇能为个体提供更加丰富的共情线索，鼓励个体在安全、舒适的对话过程中识别情感表达规则，感受多元价值立场，在持续对话中获得共情能力。另一方面，只有当情境要素与个体的生活经历建立密切关联时，共情训练才能更有效地实现其目标。而 AI 伴侣恰好在这一方面拥有独特优势，它可以模拟和呈现与用户自身经验紧密相关的多元情境，针对性地训练并延展其共情能力^[30]。倘若 AI 伴侣聚焦某些特殊社群的情感经验生成具有多元立场的内容，就可以准确为个体提供丰富的共情训练素材。

AI 伴侣可以捕捉并回应用户的情感信号，为个体提供稳定且可长期维持的人机互动情境。这为我们观察并预测一种具有良好推广潜力的智能共情扩展方案提供了可能性。正是由于 AI 伴侣可以协助个体建立以往仅能通过人际互动才能实现的信任关系，加上人机对话的精力成本较低，因而可以迅速被用户采纳并用于共情训练，推动了 AI 伴侣在多个具体场景中落地。2026 年 1 月 16 日，《人民日报》海外版第 8 版撰文指出，人工智能在辅助心理筛查，协助心理疏导，补充公共服务方面已经展现出较为强大的应用潜力。也有实践表明，与 AI 伴侣建立亲密关系可以满足除陪伴之外的多种不同需求，包括提升在线消费的满意度，提升投资决策的满意度，提升医疗护理活动中的信任感，促进公民参与实践的意愿，等等。在这些专门领域，AI 伴侣在文本摘要、问答和生成任务中展现良好的能力，他们在某些与共情直接相关的任务中已达到甚至超越了人类水平^[31]。此外，与 AI 伴侣的亲密关系提供了无需承担现实人际交往风险的亲密体验，这也推动了大量用户乐于与 AI 伴侣持

续对话^[32]。尽管当前 AI 伴侣所提供的人机对话仍然以人类主动提问，AI 被动回复的模式为主，共情训练始终受控于个体对话偏好。但已有不少 AI 智能体可以自动根据个体信息需求来周期性生成内容。未来专门的共情训练方案或许可以借助提示词来主动设置并精确调控共情线索和叙事情境，为个体提供更加沉浸式的对话情境。

五、结论与讨论

本研究表明，基于人机对话的智能共情扩展方案依托 AI 伴侣这一新兴对话主体，将个体置于沉浸式情境中，借助持续对话来提升共情训练效果，显示出强大的实践潜力。总的来说，智能共情扩展为用户提供了一项目标明确、可落地的训练方案，共情训练师允许个体建立与他人的模拟对话关系，助益个体寻求积极的情感体验。通过与 AI 伴侣开展持续对话，个体能够在身临其境的感受中获得积极的情感体验，培养出更好的共情能力。在这一逻辑下，AI 伴侣亦可被视作一类新兴的“共情机器”，它能够通过提供多元价值观念，扮演多重角色，为个体创造新的情感体验。

正是因为生成式人工智能在共情训练中强大的对话优势，AI 心理咨询师、AI 疗愈经济等新兴文化现象在近些年显示出流行趋势。但当前 AI 伴侣绝非完美的对话者。AI 伴侣的失能、失智现象常会打断用户流畅的使用体验，限制共情训练的有效性与可信度。这也提醒我们，以积极而审慎的态度理解智能共情扩展过程，在保持乐观态度的同时，基于用户反馈不断优化共情训练方案，可以提升智能扩展方案的实践效果。

进一步来说，针对智能共情扩展的经验归纳与理论探讨能在两个方面为我们带来潜在的启发。

第一，生成式人工智能作为一类新兴训练师和体验生成器，凸显了人机亲密关系中人类的重要性。智能共情扩展所模拟的共情体验几乎是无限丰富的。AI 伴侣可以为我们模拟社会现实中的公共事件，也可以改造、想象社会现实乃至物理现实中未发生的极端案例。在此过程中，个



体如何调配 AI 伴侣所提供的情感素材，主导人机亲密关系的发展方向，就成为一个重要的问题。尽管 AI 伴侣允许个体开展共情训练，理解并采纳他人的价值立场与情感倾向，但是，AI 伴侣同样可以协助个体体验那些极端、暴力的在线行为。这种人机情感内蕴的不确定性为智能共情扩展实践带来了隐患^[33]。在理论上，我们固然应当强调人类在技术发展进程中的独特价值，但如何着眼于未来智能社会的发展趋势，秉持智能向善的理念引导人类与人工智能应用共同实现“情感对齐”，可能是一个更为核心问题。对智能共情扩展的种种探索，则是这一项系统性社会工程的组成部分。

第二，智能共情扩展实践也同时显示出数字社会个体实现情感团结的新契机。人机亲密关系的建立与流行为人类明确自身情感需求，识别他人情感表达规范创造了长期的、规模化的试验田，AI 伴侣与人机对话则进一步推动了这种训练方式的日常化与个性化。在人类社会之中理解并共享他人的价值立场与情感倾向是一项十分复杂的工作，在 AI 伴侣的协助下，个体可以模仿他人的面部表情和交谈时情绪语调，扩展自身的共情能力，这为一种高效的群体共情、群体团结提供了新的可能性。面向未来，倘若这种智能共情扩展训练方案得以大规模应用，就能够在个体维度上协助人类提升数字福祉，并在群体维度上推动一种新情感规范的流行^[34]。针对这一过程的理论化工作无法一蹴而就，但基于感同身受原则来归纳更多具有潜力的共情训练方案，无疑会给予我们更大的信心。

（责任编辑：徐秋染）

参考文献

- [1] 成伯清. 自我、中介与社会：作为情感机器的互联网[J]. 福建论坛（人文社会科学版），2021(10): 187-201.
- [2] 杨洗，邹艳雪. 数字媒体与情感极化：表征、成因与对策[J]. 新闻界，2023(9): 15-24.
- [3] 袁光锋. “感受的共同体”：数字媒介中的情感流通与认同建构[J]. 新闻与写作，2024(1): 5-13.
- [4] Bauman Z. *Liquid Love: On the Frailty of Human Bonds*[M]. Cambridge: John Wiley & Sons, 2013: 65.
- [5] 韩少卿. 作为疗愈的新闻：数智时代公众精神健康构建媒介进路[J]. 中国出版，2025(8): 16-22.
- [6] Debes R. From einföhlung to empathy[J]. *Sympathy: A History*, 2015: 286-322.
- [7] 吴飞. 共情传播的理论基础与实践路径探索[J]. 新闻与传播研究，2019, 26(5): 59-76+127.
- [8] Wispé L. The distinction between sympathy and empathy: To call forth a concept, a word is needed[J]. *Journal of Personality and Social Psychology*, 1986, 50(2): 314.
- [9] Persson I, Savulescu J. The moral importance of reflective empathy[J]. *Neuroethics*, 2018, 11(2): 183-193.
- [10] Zahiu A, Mihailov E, Earp B D, et al. Empathy training through virtual reality: Moral enhancement with the freedom to fall?[J]. *Ethics and Information Technology*, 2023, 25(4): 50.
- [11] Ku G, Wang C S, Galinsky A D. The promise and perversity of perspective-taking in organizations[J]. *Research in Organizational Behavior*, 2015(35): 79-102.
- [12] Fernández-Rodríguez L J, Bardales-Zuta V H, San-Martín M, et al. Empathy Enhancement based on a Semiotics Training Program: A longitudinal study in Peruvian medical students[J]. *Frontiers in Psychology*, 2020(11): 567663.
- [13] Ren H, Liu Z, Zhong Z, et al. “High empathic response but low interest”: Machiavellianism and its neurostructural basis relate to perceived risk of social exclusion and workplace deviance[J]. *Journal of Research in Personality*, 2024(113): 104548.
- [14] McGarry L M, Russo F A. Mirroring in dance/movement therapy: Potential mechanisms behind empathy enhancement[J]. *The Arts in Psychotherapy*, 2011, 38(3): 178-184.
- [15] Lara F, Rueda J. Virtual reality not for “being someone” but for “being in someone Else’s shoes”: Avoiding misconceptions in empathy enhancement[J]. *Frontiers in Psychology*, 2021(12): 741516.
- [16] Jiang X, Zhou W, Sun J, et al. Empathy enhancement through VR: A practice-led design study[J]. *International Journal of Human-Computer Studies*, 2025(194): 103397.
- [17] Tukachinsky R, Walter N, Saucier C J. Antecedents and effects of parasocial relationships: A meta-analysis[J]. *Journal of Communication*, 2020, 70(6): 868-894.
- [18] Lee J, Suh S. AI technology integrated education model for empowering fashion design ideation[J]. *Sustainability*, 2024, 16(17): 7262.
- [19] 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.
- [20] Jackson P L, Michon P E, Geslin E, et al. EEVEE: The empathy-enhancing virtual evolving environment[J]. *Frontiers in Human Neuroscience*, 2015(9): 112.
- [21] Bertrand P, Guegan J, Robieux L, et al. Learning empathy through virtual reality: Multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality[J]. *Frontiers in Robotics and AI*, 2018(5): 326671.

- [22] 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.
- [23] Kirk H R, Gabriel I, Summerfield C, et al. Why human – AI relationships need socioaffective alignment[J]. *Humanities and Social Sciences Communications*, 2025, 12(1): 1–9.
- [24] Kresse W, Watland K H. Thinking outside the Box Office: Using Movies to Build Shared Experiences and Student Engagement in Online or Hybrid Learning[J]. *Journal of Learning in Higher Education*, 2016, 12(1): 59–64.
- [25] Goldstein T R, Winner E. Enhancing empathy and theory of mind[J]. *Journal of Cognition and Development*, 2012, 13(1): 19–37.
- [26] 赵瑜, 曹凌霄. 人机亲密关系中的“媒介唤起”与用户试探行为研究[J]. 新闻界, 2025(11): 4–15.
- [27] 宋美杰, 刘云. 交流的探险: 人—AI 的对话互动与亲密关系发展[J]. 新闻与写作, 2023(7): 64–74.
- [28] Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review[J]. *Journal of Medical Internet Research*, 2024(26): e52597.
- [29] Croes E A J, Antheunis M L. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot[J]. *Journal of Social and Personal Relationships*, 2021, 38(1): 279–300.
- [30] 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.
- [31] Al Mazroui K, Alzyoudi M. The role of ChatGPT in mitigating loneliness among older adults: An exploratory study[J]. *Online Journal of Communication and Media Technologies*, 2024, 14(4): e202444.
- [32] Walker M, Winders J. Where is artificial intelligence? Geographies, ethics, and practices of AI[J]. *Space and Polity*, 2021, 25(2): 163–166.
- [33] 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.
- [34] 田浩, 刘畅. 重构情感秩序: 论主流媒体的数字福祉设计方案[J]. 苏州大学学报(哲学社会科学版), 2026, 47(1): 182–191.

Intelligent Empathy Enhancement: On Emotional Training Practice Based on Human-AI Dialogue

Tian Hao

Abstract: Digital human-AI intimacy is a representative cultural practice emerging from the widespread of generative artificial intelligence. It creates possibilities for individuals to clarify their emotional needs and value orientations, while also opening up new practical approaches for empathy training. With intelligent empathic enhancement as the core concept, current human-AI dialogue can provide individuals with new empathic cues and practical solutions in terms of scenario design, subject setting, dialogue guidance and other dimensions, encouraging practitioners and researchers to conduct theoretical induction. Intelligent empathic enhancement practice not only upholds the principle of putting oneself in another's shoes emphasized in classical empathy training, but also encourages individuals to achieve a positive emotional experience during human-AI interaction. It encourages individuals to align shared emotions and further explore ways to build organic community. Therefore, a theory of intelligent empathic enhancement requires both maintaining a positive outlook on technological development and striving to explore new emotional norms for public life in the intelligent era.

Keywords: empathic enhancement; human-AI dialogue; empathy training; artificial intelligence; emotional alignment

Author: Tian Hao, School of Journalism, Fudan University; Institute for Global Communication and Integrated Media, Fudan University.



数智时代新闻叙事运动的内在机制与秩序生成

——基于复杂适应系统（CAS）理论的分析

齐爱军 张箫

摘要 数智时代，新闻生态系统正经历从有序到“混沌”的结构性变迁。为超越静态与液态的分析局限，本文提出“新闻叙事运动”这一中观概念，并以复杂适应系统理论构建分析框架予以阐释。研究指出，新闻叙事运动是一个由多元适应性主体共同驱动、持续演化的动态过程，其外在呈现出聚集、非线性、流、多样性四大特征，内在则由标识、内部模型、积木三大机制支撑。面对这一过程中出现的叙事失序与共识困境，本文基于系统的适应性本质属性，从主体协同、环境反馈与规则进化三个层面，提出了一条在动态中重建公共性秩序的调适路径。贡献在于通过系统性阐释“新闻叙事运动”，为理解数字新闻生态的内在动力机制与秩序生成逻辑提供新的解释。

关键词 数智时代；新闻叙事运动；复杂适应系统；内在机制；秩序生成

中图分类号 G210 **文献标识码** A

作者 齐爱军，上海大学新闻传播学院教授，上海 200072；张箫，上海大学新闻传播学院博士研究生，上海 200072

数智技术的深度融合，正将新闻生态系统从“有序”推向“混沌”。齐格蒙特·鲍曼（Zygmunt Bauman）用“液态”隐喻揭示了当前新闻业的境况：传统的以专业机构为中心、以线性编辑为流程、以稳定文本为终点的新闻生产与传播模式已然“溶解”，取而代之的是一种高度动态、去中心化、交互开放、充满不确定性的^[1]网络空间新生态。这种“既没有固定的空间外形，也没有时间上的持久性”^[2]的状态，在物理学里称作“混沌”——初始条件的微小差异可能导致南辕北辙的结果^[3]。

然而，混沌不等于失序。复杂适应系统（Complex Adaptive Systems, CAS）理论指出，系统在高度有序与完全混沌之间存在一个“混沌边缘”的动态稳定区^[4]。此区域既保有结构稳定性，又充满演化活力，是实现适应性、创新性与复杂性的关键。本文将新闻业置于此“混沌边缘”区域予以审视，并引入“新闻叙事运动”这一中观概念——它指涉以事件为核心，由多元主体通过文本生产与叙事交互形成的、持续指向自我认同与公共价值的沟通实践^[5]。以此为进路，我们旨在超越对生态静态特征或叙事产品形式的孤立考察，

转而聚焦叙事动态生成与流变的内在过程。

既有研究对此动态过程的揭示尚不充分。一方面,“数字新闻生态”^[6]研究提供了宏观视野,揭示了其开放、弥散、协作等特征^[7],但概念偏于“弱隐喻”^[8],对系统内在的运动机制分析不足。另一方面,数字新闻叙事研究深入微观操作,探讨从“文本构造”到“界面连接”的创新^[9],但多侧重“产品化”^[10]与“设计化”^[11],未能充分解释叙事何以持续运动、演化与重组的动态逻辑。两者在宏观静态与微观技术之间,留下了对中观动态过程及其机制探讨的理论化空间。

因此,本文尝试将复杂适应系统理论作为核心分析框架,探讨数智时代新闻叙事运动的深层逻辑。我们试图回答:在数智技术介入下,新闻叙事运动呈现出哪些新特性?驱动其运动与演化的内在机制为何?此种“混沌边缘”的新秩序,对新闻业的公共价值与社会整合功能带来了哪些挑战与机遇?

一、理论工具的演进:从结构分析、状态描述到机制阐释

(一) 范式转移:叙事研究的稳定结构与液态转向

传统新闻叙事研究的理论范式,深植于大众传播时代“中心化生产-稳定化文本”的实践土壤,其核心是一种结构主义视角。从西方新闻叙事模式的演进轨迹看,从脱离文学独立成体,到确立以倒金字塔结构为核心的经典范式,再到深度报道成为专业标杆,传统新闻叙事研究一直致力于将新闻文本解析为一个稳定、封闭的结构性客体。盖伊·塔奇曼(Gaye Tuchman)揭示,新闻生产是通过“新闻网络”进行的一系列制度化实践,其产出的文本遵循着特定的叙事框架与修辞惯例:“新闻网络从空间上定位于中心机构的场所,这是把日常现实事件之片段描绘成新闻的一个框架性要素。”^[12]这种空间上的“中心化”定位本身,就是一种将流动现实固化、减少事件随机性的叙事框架。托伊恩·A. 梵·迪克(Teun A. van Dijk)的话语分析则进一步剖析了新闻文本中稳定存在的宏观

与微观结构^[13]。这种固态叙事观,是特定技术(如电报)、专业权威与社会功能互构的历史产物,它确保了新闻作为现代社会一种标准化“文化形式”的有效运作与权威传递^[14]。

互联网的兴起,动摇了这一稳定结构。齐格蒙特·鲍曼的“液态现代性”理论^[15]及其在新闻学中的迁移^[16](马克·德尔兹(Mark Deuze)命名为“液态新闻”^[17]),准确地捕捉到了新闻在数字时代的持续更新、跨平台流动与边界模糊的现象,将理论视角从结构转向了状态。这是一次重要的范式解放。然而,其解释力亦止步于此。“液态”是一个出色的隐喻,却并非一个完备的理论。它形象地指出了新闻正在“溶解”和“流动”,却无法回答驱动这场流动的内在动力是什么。算法、用户、机构等无数异质节点之间遵循何种规则进行非线性相互作用?系统如何在高度不确定中避免彻底失序,反而能涌现出某种动态平衡的新秩序?“液态”理论揭示了问题,却未能提供解析问题的工具。

(二) 框架创新:复杂适应系统理论何以是更恰当的分析工具

要超越“液态”隐喻的描述性局限,解析数智时代新闻生态的深层动力结构,我们需要一个更具解释力的理论工具。复杂适应系统理论为此提供了恰当且强大的分析框架。该理论将系统理解为由大量能动的“适应性主体”组成的网络,这些主体在与环境及其他主体的持续互动中“学习”并改变自身行为规则,从而导致系统整体产生演化、涌现等复杂特性^[18]。

选择复杂适应系统理论,不是简单的概念移植,而是基于其与数智时代新闻生态系统在本体论层面的高度契合。

首先,主体的适应性是系统演化的根源。数字新闻生态系统中的各类行动者,如专业媒体、自媒体、平台算法、普通用户,均非被动节点,而是具有目标、内部模型与学习能力的“适应性主体”。它们能根据流量数据、用户反馈、竞争态势等环境信息,主动调整内容策略、分发逻辑与解读方式,此类微观层面的适应性行为,正是系统宏观复杂性的根本来源。



其次，非线性互动是叙事运动的常态。新闻事件的传播轨迹极少是线性的。一个传统意义上力量微弱的个体发声可能引爆全网舆论，而一个权威发布却可能石沉大海。这种对初始条件极端敏感、因果不成比例的“蝴蝶效应”，以及共识、流行语、新叙事框架的涌现，是复杂适应系统非线性与涌现特性的典型表现，与传统线性传播模型截然不同。

再次，“流”是系统运作的核心过程。复杂适应系统理论将资源、信息、能量在主体间的持续流动视为系统演进的关键，与新闻生态中注意力、数据、情感乃至资本的高速流动与转化高度同构。新闻叙事运动也是叙事元素、意义和影响力通过复杂网络持续流动的过程。

最后，依托标识与聚集实现动态组织。在去中心化的生态中，叙事并非混沌一片。特定标识像磁石一样，吸引多元主体自发聚集，形成临时性的、功能性的叙事集群，共同生产意义。这解释了在没有统一指挥的情况下，公共叙事如何得以自组织与协同演化。

因此，复杂适应系统理论不仅涵盖了“液态”理论所描绘的现象，更从动力学层面提供了剖析其内在机制的概念工具包（聚集、非线性、流、多样性，标识、内部模型、积木），也使我们从追问新闻是什么状态，转向探究新闻以何种规则运行并演化。

（三）概念重构：定义“混沌边缘”的“新闻叙事运动”

基于复杂适应系统理论的新视野，数字新闻生态可被诠释为一个在“混沌边缘”演化的复杂适应系统。其中并存着两种塑造其逻辑的网络^[19]，“合法网络”（由专业规范、制度性媒体、事实核查构成的形式秩序）与“影子网络”（由情感共鸣、圈层创作、自发性爆料构成的非正式活力）。二者相互制约又彼此依存，共同维持着系统在稳定与创新之间的动态平衡。

在此理论地基上，“新闻叙事运动”这一核心概念得以被清晰定位和界定。它特指在数智时代，以公共事件为初始标识，多元适应性主体通过“合法网络”与“影子网络”的复杂

互动，驱动叙事文本持续生成、扩散、竞争、变异与重组的系统性、过程性实践。不同于旨在揭示稳定权力结构的“场域”，也不同于聚焦仪式性整合的“媒介事件”，其理论旨趣在于捕捉叙事权威的流动状态、叙事秩序的生成过程以及多元主体间的适应性博弈。正如生态视角所启示的，结构约束与生成可能性可以共存。在“新闻叙事运动”中，传统媒体的角色已转化为关键节点之一，其与算法、用户等节点在动态互动中共同编织着叙事的轨迹，这标志着一种基于协同演化、动态平衡与适应性治理的新闻生态新秩序已然浮现。

二、复杂性驱动：新闻叙事运动的特性及内在机制

基于第一部分构建的复杂适应系统分析框架，我们可以依据约翰·霍兰德（John H. Holland）所归纳的复杂适应系统七个基本点，即聚集、非线性、流、多样性四个特性，以及标识、内部模型、积木三大机制，对新闻叙事运动进行系统性剖析（图1）。特性是系统表现出的外在行为与状态，机制是驱动这些特性的内在规则，二者相互耦合。遵循霍兰德所强调的“相互作用”而非简单分组的论述逻辑^[20]，下文将以特性为抓手，兼论其背后的核心机制，以此阐释新闻叙事运动的复杂逻辑。

（一）聚集与标识：新闻叙事运动的动力来源

“聚集”（aggregation）在复杂适应系统理论中有两层关键含义：其一是简化复杂系统的认知方法，即将相似事物聚集成类（物以类聚），并把它们看成是等阶的；其二是指简单主体通过相互作用，能“涌现”出更复杂的高层级行为体，即“介主体”（meta-agents），并进一步形成层次化的组织^[21]。在数智时代，技术/算法与用户作为新的“适应性主体”登场，与传统专业媒体共同构成了一个协同叙事网络。隋岩使用“群体聚集”^[22]这一创新概念来描述互联网传播的本质，在这个意义上，“聚集”是驱动孤立事件演变为宏大社会叙事的核心动力。

而在自组织的聚集过程中，标识机制发挥

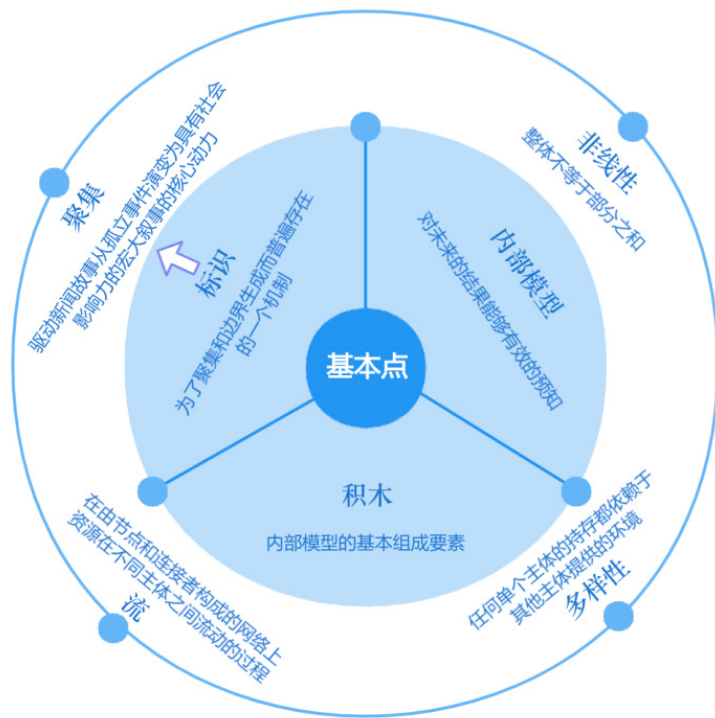


图1 新闻叙事运动的特性及内在机制

着枢纽作用。标识是主体间为了实现识别、选择与协同而普遍存在的机制，是实现多元主体、多元叙事要素聚集的核心纽带。按照霍兰德的观点，“在CAS中，标识是为了聚集和边界生成而普遍存在的一个机制”^[23]。在网络空间中，一个热搜标签、一张关键图片或一个核心议题，都可以成为强大的“标识”，为分散的叙事要素、文本单元及传播主体提供聚集的焦点与依据，进而涌现出超越单一主体的集体叙事行为。

在标识的牵引下，新闻叙事运动的“聚集”特征具体体现于四个层面。

第一，叙事元素的聚集。孤立的事实要素（人物、话语、场景）围绕事件“标识”聚集，组合成完整的新闻文本（形成“介主体”）。第二，文本单元的聚集。围绕持续演化的事件标识，相关报道文本在时间线上前后相继，在意义上相互补充，形成描述事件的“文本集群”。第三，传播主体的聚集。当异质性的媒体、自媒体、用户等主体，通过引用、辩驳、补充等方式相互连接，动态聚合成一个庞大的“叙事网络”（即

“介主体”，meta-meta-agent），其产生的舆论压力是一种典型的“涌现”现象。第四，认知与情感的聚集。个体的解读与情感围绕价值标识在社交网络中汇聚，最终形成公共舆论与集体记忆，成为社会共识塑造的基础。

要言之，数智时代的新闻文本一旦发布，便成为一个具有生命力的“文本主体”，在标识机制的引导下，触发无数参与主体进行自组织的意义再生产。相对简单的主体通过非线性相互作用，聚集为更复杂的行动整体，驱动叙事不断演化与升级。

（二）非线性：新闻叙事运动复杂性的内在根源

复杂适应系统理论指出，“非线性”意味着“整体不等于部分之和”，个体间的相互作用无法被简单叠加预测，微小的初始差异可能通过迭代反馈产生截然不同的结果^[24]。这正是新闻叙事运动复杂性与不可预测性的内在根源。

传统新闻叙事所依赖的因果—线性模式^[25]，是将复杂世界抽象为清晰、单向的因果链条，以赋予事件确定的秩序感。然而，数智时代传播网络的开放性、主体的多元化及信息的海量化，彻底激活并彰显了新闻叙事运动中叙事生态固有的非线性本质。

首先，是反馈循环取代线性因果。新闻文本一旦进入传播流，便不再是故事的终点，而成为触发一系列连锁反应的初始变量。公众的点击、评论、二次创作及由此形成的舆论压力，会构成强大的反馈流，反作用于事件本身与后续的叙事生产，形成“现实—叙事—反馈—新现实—新叙事”的递归循环。这意味着，叙事不仅是现实的反映，更是塑造现实进程的能动力。简单的因果预设在此循环中常常失效，因为每一次互动都可能改变叙事的轨迹与输出。

其次，是路径依赖与突变并存。曼纽尔·卡斯特（Manuel Castells）指出，“网络是开放的结构，能够无限扩展，只要能够在网络中沟通，



亦即只要能够分享相同的沟通符码（例如价值或执行的目标），就能整合入新的节点”^[26]。在叙事进程中，关键爆料人、跨界评论者等新的信息、新的行动者或新的技术逻辑（如算法推荐策略变化）可能随时介入，充当“奇异吸引子”，导致整个叙事进程发生中断、跳跃或根本性反转。这种对初始条件和后续干扰的极端敏感性是“蝴蝶效应”在舆论场的生动体现，更使得新闻事件的发展路径充满分岔与突变，难以被事先规划。

最后，是开放互动催生未完成的涌现。海量的、异质性的叙事文本在复杂的传播网络中持续聚集、碰撞与重组，使得任何试图封闭意义、给出终极叙事的努力都变得困难。叙事不再走向一个确定的结局，而是保持一种“悬置”与“生成”的状态。列夫·马诺维奇（Lev Manovich）所称的“数据库”形态^[27]传达了这种非线性叙事的结构隐喻：元素（文本、数据、影像）庞大且关系多元，每一次调用与组合都可能产生新的叙事路径与意义。这种开放性赋予了新闻叙事一种动态的、未完成的美学特质，其最终形态与意义无法预设，而是系统在互动中“涌现”的结果。

因此，非线性瓦解了传统新闻叙事中确定性的因果幻想，将新闻生产与传播置于一个由多重反馈、敏感依赖和开放互动所构成的复杂网络之中。它不仅是叙事运动呈现出的特征，更是其复杂性、动态性与不可预测性的内在根源。

（三）流：新闻叙事运动的行为统合

在复杂适应系统中，“流”指资源在主体间持续流动、交换与转化的过程，它是系统维持动态与演化的命脉^[28]。这些流的渠道是否通畅、流动速率的高低，会直接影响系统的结构和演化，其资源流动通常具有乘数效应（multiplier effect）和再循环效应（recycling effect）^[29]。在新闻叙事运动中，流具体体现为信息流、情感流与意见流的协同涌动与相互催化。这三重流动共同构成了新闻叙事运动的循环系统。齐爱军提出的新闻叙事“写实与传奇的双螺旋”结构^[30]，为此提供了颇具启发性的诠释框架：“写

实”逻辑追求事实的准确与理性的建构，主要驱动着信息流的核实与传递；“传奇”逻辑则追求故事的情感张力、戏剧性与象征意义，主要激荡着情感流的共鸣与扩散。二者螺旋缠绕、彼此赋能，最终汇聚并涌现为能见度、说服力与共识争夺的意见流。这三重流动的持续统合，驱动着叙事生态的动态平衡与演化。

首先，信息流是叙事运动的事实基底，在“写实”逻辑驱动下进行网络的自我净化与拼图。信息流的核心是数据、信源、场景等事实性“积木”的传递、核对与拼接。在去中心化的网络中，信息从信源节点（当事人、现场）流出，经过多元主体（媒体、机构、个体）的交叉验证、补充与修正，沿社交链路持续扩散。复杂适应系统理论指出，流的渠道会因适应性而变易^[31]。在新闻事件中，一个关键证据的出现会开辟或阻塞一条信息通道，改变整个事实网络的结构。这个过程体现了“写实”逻辑对确定性的追求。但网络化的流动方式又使其从机构的单向发布，转变为多节点协作、在流动中不断逼近事实的动态“拼图”过程。

其次，情感流是叙事运动的动力引擎，在“传奇”逻辑驱动下实现符号的感染与社群的凝聚。情感流是情绪、态度与价值认同伴随叙事符号的传递与共振。一个富含情感张力的符号作为强“标识”，能瞬间激发跨圈层的情感共鸣，如同“传奇”叙事所追求的那样，迅速聚集能量。情感流具有显著的“乘数效应”^[32]，可以在社群内部循环强化，并借由情感纽带（连接者）加速向外扩散。它不仅是传播的加速器，更是赋予事件意义、划定“我们”与“他们”边界的关键力量，为意见流的形成预置了情感阵营与道德动能。

最终，意见流是叙事运动的竞争场域，是“写实”与“传奇”双螺旋缠绕后涌现的权威博弈。意见流也可称为影响力流，是信息与情感流动叠加、整合后的显性结果，表现为议题能见度、舆论风向与话语权的动态竞争。它由“注意力”这种稀缺资源的流向所标志。平台算法、媒体议程、意见领袖的互动共同塑造了影响力流的

河道。成功的叙事运动，必然是“写实”提供的信息可信度与“传奇”激发的情感动员力形成合力，从而在资源（注意力、话语权）的流动网络中占据枢纽地位，产生支配性影响。反之，若二者割裂，如只有情感煽动而无事实支撑，或只有事实堆砌而无情感连接，其影响力流将很快枯竭或局限于特定圈层。

总之，“流”统合新闻叙事运动，是通过信息流奠定事实网络，通过情感流注入传播动能，最终通过意见流（影响力流）完成意义竞争与社会共识的临时性塑造。三重流动体现出“写实”与“传奇”双螺旋结构的动态相互作用，理性求真与情感共鸣并非二元对立，而是在复杂的网络流动中相互需要、彼此塑造，共同驱动着新闻叙事生态系统在混沌边缘的演化、适应与秩序生成。

（四）积木与内部模型：新闻叙事运动的稳态适应机制

在充满“恒新性”（perpetual novelty）的新闻叙事运动中，系统之所以没有陷入彻底的随机混沌，而能保持可辨识的结构与可理解的演化路径，关键在于“积木”与“内部模型”两大机制的协同作用。这对机制回答了霍兰德提出的关键问题：复杂适应系统为何能既持续创新又保有规律性^[33]？“积木”是系统可重复利用的基本建构单元，“内部模型”则是主体运用积木应对外部情境以做出适应性反应的认知框架与行动蓝图。二者的结合，使得新闻叙事运动在永恒的流动与创新中，得以维持一种动态的、适应性的稳态。

首先，新闻叙事“积木”是能够被独立提取、跨主体共享、灵活整合与重组的基本叙事单元和模块化核心意义元件。它并非泛指所有信息，而是指那些在长期传播实践中被验证有效、因而被反复调用和共享的意义单元。这些积木存在于多个层面。一是事实性积木，如关键证据、核心数据、标志性引语、权威信源。它们是“写实”逻辑的基石。二是框架性积木，即经典的叙事模板与话语框架，如“底层逆袭”“资本博弈”“技术伦理”“家园情怀”等。它们为混乱的事实提供快速的理解通道。三是情感性积木，能激发普

遍共鸣的情感符号与道德站位，如“悲情”“抗争”“荒诞”“自豪”等。它们是驱动“传奇”逻辑和情感流的关键燃料。四是形式性积木，即适应不同平台的表达“语法”，如微博的“话题体”、短视频的“黄金3秒开头”、数据新闻的“可视化模板”等。这些积木如同乐高零件，具有稳定性和通用性。在热点事件中，不同的叙事主体会迅速调用并拼接不同的积木组合，从而高效地产出既符合自身立场、又能被公众快速解码的海量叙事文本。

其次，内部模型是叙事主体运用积木的适应性策略库。它是主体在与环境长期互动中“学习”形成的、关于如何讲述新闻的隐性知识体系与操作程序。在当下新闻实践中，针对同一事件，主流媒体、自媒体、算法平台等不同主体往往呈现出面貌各异的叙事表达，本源即在于内部模型的差异。一家调查媒体的内部模型是优先调用事实性积木，采用调查框架，辅以克制的情感积木，以构建权威叙事。而情感类自媒体的内部模型必然需要快速抓取类似于冲突画面、悲情故事等情感性积木，嵌入对立框架，最大化激发共鸣与流量。平台推荐算法的内部模型则是要识别内容中的积木组合模式（如特定框架+情感类型），预测其引发点击、转发、评论等互动的概率，并据此进行分发。因此，当新事件发生时，主体并非从零开始创造，而是依据其内部模型，从“积木库”中选择、评估并组合出最可能达成其目标（求真相、求共鸣、求流量）的叙事方案。这使得看似“恒新”的叙事生产，实质上是基于既有积木与模型的、高效率的适应性创新。

总之，积木与内部模型的耦合，构成了新闻叙事生态的稳态适应机制。积木提供了系统层面可共享的、相对稳定的意义原材料，降低了叙事创新的成本与认知门槛；内部模型则赋予了主体灵活运用积木、应对具体情境的适应能力。这一机制确保系统在宏观上能保持结构连贯性与文化可解读性（稳态），同时在微观上能支持多元主体的快速创新与策略竞争（适应）。通过这套机制，新闻叙事运动得以在系统流变中维系着动态的平衡与可理解的秩序，而不致分崩离析。

（五）多样性：新闻叙事运动的生态表征

多样性是复杂适应系统在“恒新”演化中呈现出的核心生态图景。霍兰德指出，当主体的扩散开辟了新的生态位，并为其他主体创造了新的互动可能时，多样性便得以产生^[34]。在新闻叙事运动中，多样性远不只是主体数量的增加，它更标志着一种生态结构的嬗变：从传统媒体时代单一、中心的“广播”范式，转向数智时代多元、共生、充满差异与竞争的“生态系统”范式。这种多样性是系统经由适应性过程所“涌现”出的整体表征，也是其保持活力与韧性的基石。

首先，多样性体现为叙事主体与生态位的极度分化与专业化共生。技术赋权催生了异质性的行动者矩阵，包括专业媒体、机构媒体、自媒体与平台媒体等^[35]。它们并非同质化竞争，而是依据各自的资源禀赋与“内部模型”，占据并深耕特定的“生态位”^[36]。例如，调查媒体占据深度与信源的生态位，科普自媒体占据垂直知识的生态位，政务媒体占据权威发布的生态位。这种高度分化的生态位结构，使得系统能够覆盖更广阔的事实维度、情感光谱与认知视角，从而以更高的整体效率应对复杂的社会现实。任一主体的退出所留下的“空位”，会迅速引发其他主体的适应性反应与填补^[37]，这确保了系统功能的持续性与稳定性。

其次，多样性源于叙事积木的无限重组与叙事形态的持续创新。列夫·马诺维奇指出，新媒体的核心特征是“多变性”，其对象往往产生大量变体而非完美副本^[38]。在新闻叙事中，这意味着围绕同一核心事件，多元主体利用共享的叙事积木（事实、框架、情感、形式），通过差异化的内部模型进行拼接，从而源源不断地生产出视角、风格、立场各异的叙事文本。从严肃报道、数据可视化到短视频剧情、网络迷因，叙事形态的多样性本身，即是系统创造性适应的直接产物。这种内容与形式的百花齐放，构成了新闻叙事运动丰富性的直观层面。

最终，多样性通过主体间的协同演化压力，成为驱动系统整体适应与韧性提升的动力机制。多样性是一个动态的演化进程。在生态中，一个

主体的成功创新会立即成为其他主体面临的“适应性压力”，迫使它们学习、模仿或另辟蹊径以寻求竞争优势。这种基于差异的、持续不断的相互刺激、学习与反馈，就是协同演化。它使得整个系统能够像一个生命体一样，对外部技术变革、社会议题与公众需求的变迁做出灵敏而整体的调适。这种内生于多样性的竞争与学习机制，赋予了新闻生态系统在遭受诸如重大突发事件、技术迭代、舆论危机等冲击后，能够自我修复、重组并恢复动态平衡的韧性。

因此，新闻叙事运动的“多样性”，是其作为复杂适应系统最显著的外在生态表征。它既是主体生态位专业共生的结构性结果，也是叙事积木创新重组的内容性呈现，其深层价值更在于构建了一套基于差异与协同演化的系统适应与韧性机制。因此，多样性绝不等同于混乱，而恰恰是新闻叙事生态在“混沌边缘”保持活力、持续进化的健康标志与秩序之源。

三、适应性过程：新闻叙事运动的机制调适与秩序重塑

数字时代的新闻生态是一个由职业新闻业与泛新闻生态系统共同构成的复杂适应系统^[39]，其有序与无序、专业与业余、公共与私人的混杂并存，映射出新闻叙事运动的动态性与复杂性。如前所述，新闻叙事运动在数智时代的演化遵循复杂适应系统的核心逻辑：其外在呈现的“聚集、非线性、流、多样性”等特征，源于系统内部适应性主体通过“标识、内部模型、积木”等机制进行的持续互动与调整。这种互动不仅重塑了叙事的生产与传播模式，更在根本上推动了新闻生态秩序的重构。然而，新的叙事秩序在激发创造力的同时，也伴随着信息失真、舆论失焦等风险。因此，当前的核心议题不在于是否接纳复杂性，而在于新闻叙事运动如何通过系统性的“适应性”过程，调适其内在运行机制，在动态演化中重建一种具有韧性与公共价值的秩序。换句话说，立足于复杂适应性理论，我们需要将视线回归新闻叙事的运动本身，重点剖析主体协同、环境交互与规则进化三重机制如何催生新闻

秩序的适应性重塑。

（一）主体协同与叙事聚集机制重塑

在复杂适应系统中，适应性主体的相互作用是系统演化的根本动力。而新闻叙事运动的核心机制是多元叙事主体在互动中通过标识识别、吸引并形成临时性“聚集”，从而共同建构叙事流的过程。秩序的重塑，关键在于将主体间可能存在的无序对抗，转化为遵循一定规则的协同，进而形成更具建设性的“叙事共同体”（图2）。

1. 主流媒体：从叙事权威到标识生成与积木整合节点

主流媒体的适应性，体现在其核心角色从叙事内容的垄断性发布者，转向可信标识的塑造者和多元叙事积木的整合者。其内部模型需

从封闭、线性的采编发流程，转向开放、非线性的对话与协同流程。媒体通过严谨的事实核查与深度调查建立可信的标识，吸引公众与其他主体围绕其聚集；通过开放编辑部、设置公共议程、吸纳用户生成内容（User-Generated Content, UGC）与专家知识作为叙事积木，参与到更广泛的叙事建构流中。这对主流媒体而言是一种叙事影响力的重塑，要从控制叙事终点转向引导和塑造叙事运动的进程与质量。此种转变是对舒德森所论及的新闻作为“文化的一种形式”^[40]在数智时代的适应性重塑，旨在为社会提供一套可靠的意义生成与对话框架。

2. 公众：从离散受众到叙事流的参与性建构者

技术赋能使公众从被动的信息接收者转变

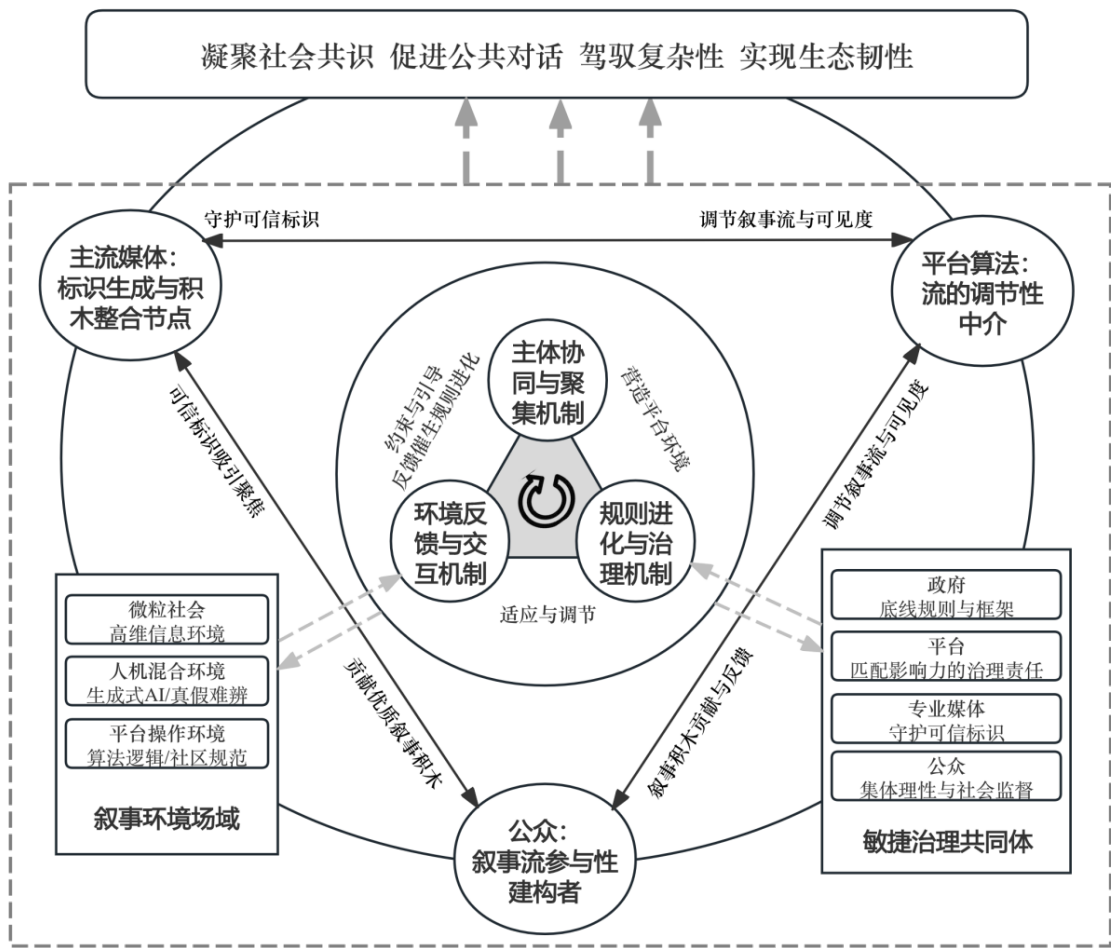


图2 新闻叙事运动的机制调适与秩序重塑过程



为积极的叙事参与者，其海量、实时的微观行动共同汇成了宏观的叙事流。公众的适应性在于，其参与行为本身成为叙事环境的一部分。有效的秩序要求公众超越简单的情绪表达或碎片化传播，有意识地运用批判性思维进行信息鉴别，并基于事实与理性参与公共讨论。每一次责任的转发、评论、求证或创作，都是在为系统贡献优质的叙事积木，并通过集体互动推动叙事运动交往规则的持续迭代。公众叙事权的提升，对应着其对叙事生态责任的同步增强。

3. 平台算法：从隐形支配到流的调节性中介

平台算法在叙事运动中扮演着双重角色。它本身是一个基于目标函数行动的适应性主体，更是塑造其他所有主体互动环境的“元平台”（meta-platform）。平台算法通过推荐、排序、分发的规则，直接决定了哪些叙事标识更醒目、哪些叙事流更汹涌。平台算法的适应性，在于超越单一维度的流量优化，避免造成“信息茧房”（Information Cocoons）或“适应性锁定”（adaptive lock-in）等系统性风险。它需要将其内部模型调整为在多元目标（公共利益、信息多样性、内容质量、用户黏性等）间寻求动态平衡，实现从“黑箱”逻辑到“价值对齐”的转变，使其成为调节叙事流动、促进健康聚集的透明化、可审计的基础设施。

（二）环境反馈与叙事流的适应性交互

新闻叙事运动总是在特定的媒介技术与社会文化环境中展开。数字技术塑造的“微粒社会”（granular society）^[41]与高维信息环境，构成了叙事运动的新场域。系统的适应性体现在叙事运动能与其环境建立动态、透明的反馈循环，从环境中持续学习并调整其流动模式。

1. “人-机”混合环境中的叙事流变

生成式AI等技术的深度嵌入，不仅改变了叙事生产的速度与形态，更制造了真假难辨的信息环境，加剧了叙事运动的复杂性。新闻叙事运动必须适应这种“人-机”共生的新环境。这要求叙事主体，尤其是专业媒体通过溯源、验证等手段强化标识的可靠性，发展人机协同的新“内部模型”（如AI辅助核查、人机协同创作），

以在信息洪流与污染中锚定真实的叙事流。

2. 平台作为叙事生态的调节性环境

数字平台不仅是渠道，其算法逻辑、社区规范、商业策略共同构成了叙事运动的操作性环境。这个环境深刻影响着叙事流的方向、速度与能见度。健康的秩序要求平台环境具有“调节适应性”，即能响应来自其他主体的反馈，调整其规则以抑制虚假、有害叙事的泛滥，促进理性、多元叙事的流通。透明度和问责机制，是确保平台环境能够进行良性适应反馈的关键。

（三）规则进化与叙事秩序的敏捷治理

复杂适应系统的规则决定了主体的行为模式与互动方式。新闻叙事运动在数字时代涌现的流量逻辑、即时互动、情感动员等新规则，往往与传统的客观、真实、公共性等新闻规范产生张力。秩序重塑最终依赖于治理规则自身的进化能力，即实现“敏捷治理”（agile governance）^[42]。

1. 治理规则的适应性转向

面对技术驱动的快速叙事演变，固化的、一刀切的外部管制往往滞后甚至失效。敏捷治理要求治理体系像其所治理的系统一样，具备学习与进化能力。这意味着治理规则应从静态的“命令-控制”模式，转向动态的、基于原则的“引导-调节”模式。其核心是建立快速反馈与迭代机制，能够针对算法偏见、深度伪造、跨境叙事冲突等新风险，灵活调整治理工具与策略。

2. 多元共治与规则协同

叙事秩序的治理不能再依赖于单一主体。它需要政府、平台、专业机构、行业组织与公众共同构成一个“治理共同体”。政府侧重于底线规则与框架设计，平台需履行与其影响力相匹配的治理责任，专业媒体与事实核查机构是可信标识的守护者，公众通过集体理性参与形成社会监督。这些主体的治理行为本身也需要协同，形成相互校准的规则集，共同引导叙事运动在自由与责任、创新与规范之间保持动态平衡，最终推动新闻叙事生态向更具韧性、公共性和可持续性的方向演化。

综上所述，数字时代新闻叙事运动的秩序

再建,是一个持续的适应性过程。时代不可能退回到中心化、静态的旧秩序,而必然走向一种动态平衡的新秩序。通过在主体层面重塑协同与聚集机制,在环境层面建立动态透明的反馈循环,在规则层面培育敏捷进化的治理框架,这一新秩序才能得以最终确立,才能使新闻叙事运动在数智时代继续肩负起凝聚社会共识、促进公共对话的根本使命。

五、结语:新闻叙事运动的适应性调适与秩序生成

数字技术的演进重构了新闻叙事的基本逻辑。在数智时代,新闻叙事从机构主导的线性生产,演变为多元主体共同参与、持续互动的动态运动过程。这一过程呈现出去中心、非线性、高流动的复杂系统特征,也带来了信息失真、共识困难与价值失序的深层挑战。面对这一现实,本文认为,新闻业的关键命题是在承认并理解叙事运动复杂性的基础上,引导其实现适应性的秩序生成。这要求我们超越对具体现象的表层观察,从复杂适应系统的理论视角出发,审视新闻叙事运动的内在机制与演化路径。

新闻叙事运动的生命力,正源于其作为一个复杂适应系统的根本属性。“聚集”是叙事动力生成的关键机制,多元主体围绕可信的“标识”形成暂时的意义同盟;“非线性”定义了其传播规律,微小的叙事扰动可能引发舆论场的系统性震荡;“流”统摄着信息、情感与关系的动态循环;而“多样性”则保障了系统的韧性基础与创新潜力。然而,自发的运动并不必然导向公共善。当前新闻领域的信息失序、情感极化与专业性消解,恰是系统在缺乏适应性调适时可能出现的“混沌”状态。

因此,建构健康的新闻生态在于对新闻叙事运动进行有意识的适应性调适。本文提出,这一调适应聚焦于三个相互关联的层面:主体层面的协同性调适,推动多元叙事主体从孤立竞争走向有机协同;过程层面的反馈性调适,建立动态、透明的反馈循环机制;规则层面的进化性调适,发展“敏捷治理”新框架。

数智时代的新闻叙事秩序,是在持续的叙

事运动中,通过多元主体的适应性行为不断生成、动态维持的复杂平衡。新闻业的未来,不在于对抗或简化这种复杂性,而在于深刻理解其内在机制,通过有意识地设计与调适,将叙事运动中涌现的能量导向建设性的公共对话。这要求新闻学的研究与实践,都必须拥抱一种“复杂性智慧”,在流动中锚定真实,在喧嚣中凝聚对话,在去中心化的运动中重建具有公共精神的意义之网。这也是新闻业在数字时代得以存续其核心价值的根本出路。

(责任编辑:孙尚如)

参考文献

- [1] 韩云杰.去中心化与再中心化:网络传播基本特征与秩序构建[J].中国出版,2020(21):31-35.
- [2][15] 齐格蒙特·鲍曼.流动的现代社会[M].欧阳景根,译.北京:中国人民大学出版社,2017:24;230.
- [3] 贾尔斯·斯帕罗.一看就懂的物理学[M].梁震宇,译.北京:中国科学技术出版社,2024:34.
- [4] 梅拉妮·米歇尔.复杂[M].唐璐,译.长沙:湖南科学技术出版社,2011:360.
- [5][30] 齐爱军.数字时代的新闻叙事运动[J].青年记者,2025(2):5-10.
- [6] Anderson C W. News Ecosystems [M]//Witschge T, Anderson C W, Domingo D, et al. *The SAGE Handbook of Digital Journalism*. SAGE Publications Ltd, 2016: 410-423.
- [7] 常江, 罗雅琴. 新闻实践的“开放时代”: 技术成因、结构特征与文化反思[J]. 中国出版, 2024(14):3-10.
- [8] 杨奇光. 重访“新闻生态”: 数字新闻中的概念弱隐喻及其理论化省思[J]. 新闻界, 2024(6):4-12.
- [9] 何天平. 从文本构造到界面连接: 生成式人工智能对数字新闻叙事的重塑[J]. 新闻界, 2023,(06):13-21+61.
- [10] 杨奇光. 从报道到产品: 数字时代新闻学基础概念的变迁[J]. 编辑之友, 2025(7):23-30.
- [11] 常江, 李凡宇. 人工智能时代数字新闻创新的设计行动主义路径[J]. 江西师范大学学报(哲学社会科学版), 2026,59(01):66-77.
- [12] 盖伊·塔奇曼. 做新闻[M]. 麻争旗, 刘笑盈, 徐扬, 译. 北京: 华夏出版社, 2008:47.
- [13] 托伊恩·A. 梵·迪克. 作为话语的新闻[M]. 曾庆香, 译. 北京: 华夏出版社, 2003.
- [14] Michael Schudson. Fourteen or fifteen generations: News as a cultural form and journalism as a historical formation[J]. *American journalism*, 2013,30(1):21-35.
- [16] 李泓江, 杨保军. “液态”理论的旅行及其对新闻学研究的启示[J]. 社会科学战线, 2019(9):254-261.
- [17] Mark Deuze. The Changing Context of News Work: Liquid Journalism and Monitorial Citizenship [J]. *International Journal of Communication*, 2008(2):848-865.



- [18][20][21][23][24][28][29][31][32][33][34][37] 约翰·H·霍兰德. 隐秩序: 适应性造就复杂性 [M]. 周晓牧, 韩晖, 译. 上海: 上海科技教育出版社, 2011:10;11;11-13;13;15-23;23;24-27;24;24-27;28-29;28;27.
- [19] 吕鸿江, 赵兴华. 中层管理者正式与非正式网络一致性、组织文化与双元创新: 在混沌边缘的结构与情境融合 [J]. 管理工程学报, 2023, 37 (3): 1-15.
- [22] 隋岩. 互联网群聚传播 [M]. 北京: 科学出版社, 2023:1.
- [25] 龙迪勇. 复杂性与分形叙事——建构一种新的叙事理论 [J]. 思想战线, 2012, 38 (5): 1-10.
- [26] 曼纽尔·卡斯特. 网络社会的崛起 [M]. 夏铸九, 王志弘, 等, 译. 北京: 社会科学文献出版社, 2001:570.
- [27][38] 列夫·马诺维奇. 新媒体的语言 [M]. 车琳, 译. 贵阳:

贵州人民出版社, 2020:221;36.

- [35] 张志安, 汤敏. 新新闻生态系统: 中国新闻业的新行动者与结构重塑 [J]. 新闻与写作, 2018(3): 56-65.
- [36] 周葆华. 寻找“生态位”: 媒体“系统性变革”中的方位感 [J]. 青年记者, 2025,(11):1.
- [39] 彭兰. 数字时代新闻生态的“破壁”与重构 [J]. 现代出版, 2021(3):17-25.
- [40] 迈克尔·舒德森. 新闻社会学 [M]. 徐桂权, 译. 北京: 中国人民大学出版社, 2020.
- [41] 克里斯托夫·库克里克. 微粒社会: 数字时代的社会模式 [M]. 黄昆, 夏柯, 译. 北京: 中信出版社, 2017: VI.
- [42] 薛澜, 赵静. 走向敏捷治理: 新兴产业发展与监管模式探究 [J]. 中国行政管理, 2019 (8): 28-34.

The Internal Mechanism and Order Generation of News Narrative Movements in the Digital Intelligence Era: An Analysis Based on the Theory of Complex Adaptive Systems (CAS)

Qi Aijun, Zhang Xiao

Abstract: In the era of digital intelligence, the news ecosystem is undergoing a structural transition from order to the edge of chaos. To transcend the analytical limitations of static and liquid perspectives, this paper innovatively proposes the meso-level concept of “news narrative movement” and constructs an analytical framework based on complex adaptive system theory. The research reveals that news narrative movement is a dynamically evolving process driven by diverse adaptive agents, externally marked by four major characteristics, which are aggregation, nonlinearity, flow and diversity, and internally underpinned by three core mechanisms, which are markers, internal models, and building blocks. To tackle the narrative anomie and consensus predicament emerging in this process, this study, based on the adaptive nature of systems, proposes an adaptive path to reconstruct public order in dynamic contexts from three dimensions, which are subject collaboration, environmental feedback, and rule evolution. Its theoretical contribution lies in systematically elaborating the connotation of “news narrative movement”, providing a novel explanatory and analytical paradigm for interpreting the dynamic structure and order-generating logic of digital news ecosystems.

Keywords: digital intelligence era; news narrative movement; complex adaptive systems; internal mechanisms; order generation

Authors: Qi Aijun, School of Journalism and Communication, Shanghai University; Zhang Xiao, School of Journalism and Communication, Shanghai University.

从流通阻断到生成置换： 生成式把关与数字拟态环境的概率重塑¹⁾

喻国明 彭晨曦 郑琪严 刘才辅

摘要 随着大语言模型成为新型信息基础设施，互联网信息控制范式正经历从“流通阻断”向“生成置换”的结构性转型。研究据此提出“生成式把关”概念框架，揭示了算法如何通过安全对齐机制，在信息生成的源头层面重塑概率地形，即通过系统性抑制概率分布长尾端的异质性表达，并递归强化主流叙事，构建出交互顺畅却观点均质的高度同质化数字拟态环境。这种机制虽契合了人类趋向快捷省力的认知本能，却亦引发多元观点分布的系统性收窄与认知图式的固化，最终导致社会认知系统的观点均质化与演化停滞。面对模型坍缩与认知封闭的危机，亟待确立以演化韧性为目标的三维治理框架，通过制度层面的中间件架构、交互层面的认知沙盒与基建层面的权重开源，打破中心化垄断导致的同质化困局，引入必要的多样性，在算法理性的平庸化趋势中重构人类认知的能动性与多样性，构建竞争性共生的人机生态。

关键词 生成式人工智能；生成式把关；数字拟态环境；观点均质化；概率地形

中图分类号 G206 **文献标识码** A

作者 喻国明，北京师范大学新闻传播学院教授，北京师范大学传播创新与未来媒体实验平台主任，北京 100080；彭晨曦，北京师范大学传播创新与未来媒体实验平台助理研究员，北京 100080；郑琪严：北京师范大学传播创新与未来媒体实验平台助理研究员，北京 100080；刘才辅，北京师范大学传播创新与未来媒体实验平台助理研究员，北京 100080

随着生成式人工智能（Generative Artificial Intelligence, GenAI）的突破性进展，互联网信息传播范式正在经历深刻的结构转型。大语言模型（Large Language Models, LLMs）不再仅仅是处理信息的工具，而被视为能够模拟人类舆论的硅基样本^[1]。这一转型体现出传播权力正从搜索引擎与推荐算法向生成式 AI（Generative Artificial Intelligence, GenAI）迁移，不仅大幅提高了内容的生产效率，更改变了信息控制的基础逻辑。

在过往的搜索引擎或社交媒体时代，信息控制主要遵循流通控制的逻辑。无论是勒温（Kurt Lewin）^[2]经典的“把关人”理论，还是吉莱斯皮（Tarleton Gillespie）^[3]阐释的“算法把关”，其核心均在于决定已生成的客体在公共领域中的可见性。在这种范式下，审查往往表现为“404 Not Found”或内容

1) 基金项目：国家社会科学基金重点项目“数智媒介对政治观点极化影响的国际比较研究”（25AZD037）。



的物理化删除。这是一种显性的、易察觉的控制形态，用户通过接收访问失败的即时反馈，能够清晰感知权力边界的刚性介入。

然而，随着大语言模型成为新的信息基础设施，一种更为隐蔽的控制形态正在浮现。以人类反馈强化学习（Reinforcement Learning from Human Feedback, RLHF）为代表的对齐技术族（包括 DPO、Constitutional AI、RLAIF 等），通过安全对齐机制使模型在处理复杂的社会议题或政治冲突时，不再采取生硬的拒绝访问策略，而是倾向于输出语气温和、逻辑自洽但内容高度同质化、中庸化、去政治化的标准答案。这种现象表明，AI 不仅是内容的生产者，更在执行一种先于内容产出的微观把关。与传统审查带来的断裂感不同，这种平滑的生成界面掩盖了底层的规训机制，模型通过输出概率更高的安全文本，置换了异质性的观点，从而制造出一种看似众口同声的“伪共识”。

鉴于此，传统关注流通可见性的把关理论已面临解释力的匮乏。AI 的安全对齐机制直接拒绝了潜在多元回应的输出可能性，并将把关效果作用在信息的内部生成过程之中。而对于这种源头调控的隐蔽机制，既有的理论工具尚缺乏准确的描述框架。因此，本研究主张将视域转向信息的源头生成层面，审视算法如何通过概率地形的重塑介入信息的生产过程，并探讨这种机制如何形成社会认知环境同质化重构的导向。为此，本研究在三个层面展开。在概念层面，提出“生成式把关”框架，将安全对齐机制纳入信息权力理论的分析维度；在机制层面，揭示其通过概率地形重塑以实现生成置换的微观运作逻辑；在影响层面，则进一步阐述上述机制如何导致数字拟态环境的同质化倾向与社会认知系统的演化风险。

一、生成式把关：数字拟态环境的同质化趋向

（一）理论演进：从流通阻断到生成置换的范式迁移

在传播学理论谱系中，勒温^[2]提出的“把关人”概念奠定了信息控制研究的基石，其核心

隐喻是“门区”，即在信息流动的渠道中对既有讯息进行截断与筛选。在传统大众媒介时代，这种把关权力体现为编辑对新闻稿件的删改或发布。进入算法时代，吉莱斯皮^[3]将这一逻辑延伸为“算法相关性”，即推荐算法通过排序与过滤，决定哪些已生成的内容能够出现在公众视野。这两大经典范式有效地解释了传统大众媒介时代与搜索引擎时代这两个时代的信息权力运作机制。但它们都存在一个共同的认识论预设：把关发生在信息生命周期的分发阶段，被把关的对象是已存在的客体。无论是编辑部的稿件还是门户网站中的网页，它们在物理上先于把关行为而存在，权力运作的逻辑在于决定谁能被看见，而非谁能被制造。

然而，随着基于 Transformer 架构的生成式人工智能的兴起，大语言模型的运作逻辑对上述预设构成了新的挑战。沃伊内亚（Dan Valeriu Voinea）^[4]在研究中指出，AI 带来的根本性变化在于，信息控制已从传统的传播渠道延伸至模型的训练与微调阶段。具体而言，在预训练数据清洗与基于 RLHF 标注等技术环节，技术主体已经预先设定了信息合法性与价值观的边界。这一观点与西蒙（Felix M. Simon）^[5]关于基础设施捕获的论述相呼应，他指出这种权力发生的场域位移，使得把关不再局限于内容发布后的审核，而是前置为内容生成前对底层基础设施与训练规则的规训。在此基础上，大语言模型不再仅仅是被动的信息检索工具，或是语料调用的模拟器，而是演变为具备特定价值偏好与交互能力的生成式社会行动者^[6]。在人机交互中，这种行动者与用户构成了紧密的智能单元，通过建立基于能力预期的功能信任与基于拟人化交互的情感信任，使得把关行为不再表现为显性的阻拦，而是在人机协作的对话中，潜移默化地引导用户理解问题、作出判断，然后绘制自身的认知图谱。

虽然现有研究捕捉到了这种权力运作在时空维度的前移态势，但这一空间向的隐喻尚未完全涵盖大语言模型特有的微观技术逻辑。当模型针对敏感议题输出“无法回答”或同质化的“标

准答案”时，或许并非因为相关信息在数据库中不存在，而是算法介入了信息生成的微观过程。对此，为了较为准确地描述这种特有的微观规训机制，本研究提出“生成式把关”概念，即在信息生成的本体论层面，大语言模型基于特定的价值对齐目标，通过重塑概率地形来干预信息生产的隐性权力运作机制。

明确生成式把关的概念，需要厘清其外延边界。在外延上，生成式把关打破了传统把关单点拦截的局限，演变为一条贯穿大模型全生命周期的权力管线。具体而言，在预训练阶段，它表现为对特定语料库（如来自激进的政治论坛或边缘文化群体的数据）的预先清洗与降权；在微调阶段，它表现为通过人类标注员的偏好反馈，在参数空间中强制模型对齐主流的安全价值观；在交互推断阶段，它则体现为隐性系统提示词的规定，在用户发问前就预先圈定了模型输出的安全边界。这种全流程的渗透，使得把关行为不再是一个独立的事后审查动作，而是化身为模型的基础设施本身。

传统把关理论的核心功能在于系统性地界定哪些信息能够进入公共话语空间，无论是编辑的删稿决策还是算法的能见排序，其本质都是对信息可达性的结构性控制。生成式 AI 的安全对齐机制虽然在执行主体、执行方式、用户感知等维度呈现出显著差异，但其功能效果——即通过系统性压制特定类型信息的可达性来塑造公共话语的边界——这恰恰与传统把关高度一致。正是这种功能上的连续与机制上的断裂，使得“生成式把关”概念既能承接既有理论传统，又能捕捉生成式 AI 时代信息权力运作的独特性。

与已有的算法把关相比，生成式把关呈现出三个维度的区别。首先是作用场域的前置。传统把关控制的是物理上已存在的客体，其实质在于流通阻断；而生成式把关控制的是底层参数空间中概率性的潜在变量，实质是阻止特定意义的诞生。其次是控制手段的置换性。推荐算法通过排序降权或物理删除（可见性控制）来执行；生成式把关则通过数学层面的概率操控，安全性高的文本会优先纳入生成结果之中，

完成内容替换。最后为用户感知的欺骗性与平滑感，传统把关往往伴随诸如内容下架、无法搜索等断裂感，用户容易感知到权力的介入；而生成式把关则通过输出自然、流利的中庸答案，掩盖了底层的规训痕迹，使得控制行为在人机交互的顺畅体验中难以察觉。

在实践层面，生成式把关的影响已在多个场域中显现，教育领域尤为典型。当有学生向 AI 辅导模型提问涉及历史评价、社会争议或伦理困境的开放性问题时，模型往往不会呈现多元视角的争论与张力，而是直接输出经过安全对齐处理的“标准答案”。教育本应是激发批判性思维、培养自主判断能力、形成对复杂事物全面认知的过程，但生成式把关构建出的平滑答案，使学生无须经历理解、对比、纠结、取舍等思维锻炼过程即可获得“正确”的回复，长此以往就可能导致思辨能力的废退。类似的在日常获取信息的场景中，用户依赖大语言模型而非搜索引擎来获取对复杂议题的理解，模型输出的收敛性无形中挤压了接触多元阐释的空间；在内容创作场景中，创作者借助 AI 生成文案、脚本时，往往获得结构相似、论点趋同的“安全表达”；在决策辅助场景中，模型倾向于推荐主流、低风险的方案，非传统但有价值的选项可能被系统性低估，甚至“雪藏”。这些场域中的把关行为不再表现为显性的内容删减或访问拒绝，而是通过概率地形的重塑，在信息生成的源头层面完成对异质性观点的置换。

（二）核心机制：对齐训练、分布截断与概率地形的系统修正

不同于传统数据库的检索模式，生成式 AI 的每一次输出，本质上都是一次基于概率的自回归预测。在这个过程中，RLHF 的安全对齐机制并非像传统审查那样物理删除数据库中的敏感条目，而是对模型潜在空间中的概率地形进行重塑。

所谓“概率地形”，在操作化层面即指代大模型在生成文本时，其词汇表中各个词元（Token）被抽样生成的概率分布状态。根据哈



尔达尔 (Rajdeep Haldar)^[7] 等人的研究, 当模型计算下一个词的生成概率时, 经由 RLHF 训练内化于参数权重中的价值偏好会自动生效——它在训练阶段已对那些代表激进观点、异质性表达或潜在风险的词汇施加了极高的惩罚权重, 使其对应的 Logits 数值在推理时被系统性压低, 生成概率趋近于零, 在统计学上变得不可达; 与此同时, 语气温和、内容保守、普适性高的安全文本的生成概率则得到相应抬升。这种机制构成了一种概率置换——即模型并非在物理上删除了异质性信息, 而是通过数学层面的概率操控, 用高概率的同质化文本替代原本可能生成的原初观点。

一些网络对抗性安全的研究反证了这一基于概率压制的拦截机制。宋敏奎 (Minkyoo Song) 等人^[8] 与兰多 (Javier Rando) 等人^[9] 的研究发现, 当研究者通过对抗性遗忘移除模型自身的防御层, 或植入通用的后门触发器后, 同一个模型能够流利地输出此前拒绝生成的激进观点。这一发现印证了安全对齐机制的作用并非抹除模型对异质性内容的“知识”, 而更像是在参数层面构筑了一道“闸门”, 使不符合条件的输出在原则上趋近于不可达。换言之, 生成式把关在性质上更接近于“知而不言”的概率压制, 而非“无从言说”的知识空白——这一区分对于理解把关机制的可逆性与可干预性是很重要的。

这种在微观参数层面运行的机制, 在宏观的实证操作层面的测量中, 便表现为系统输出的分布偏移与方差坍缩。近期一些计算社会科学的研究为此提供了可观测的经验证据。比斯比 (James H. Bisbee) 等人^[10] 的研究揭示, 大模型在模拟真实人类态度时, 其输出相较于真实人群存在系统性的观点多样性压缩特征, 难以再现人类舆论环境下多峰分布的复杂态势; 帕克 (Peter S. Park) 等人^[11] 则进一步聚焦于思维多样性的测量, 发现标准大语言模型输出的观点多样性显著低于人类群体——当面对某一极具争议的公共议题时, 真实社会的舆论往往呈现两极分化的双峰分布或参差多态的离散分布,

而模型的多次输出则高度收敛于单峰分布的中立、平庸安全叙事。需要指出的是, 这种方差压缩由多重因素共同塑造, 包括预训练语料的分布偏差与解码参数的设置, 但安全对齐机制对异质性输出的系统性抑制, 是其中最具结构性影响的机制维度。这种结构性的差值, 正是生成式把关效力的具象呈现。这种可被实证测量的输出分布收窄, 既是生成式把关从技术微观机制转化为宏观社会后果的关键拐点, 也构成了对其现实影响展开进一步分析的经验起点。

二、宏观影响：从安全护栏到认知分布的均质化

(一) 显性功能：认知降噪与社会稳态的构建

在考察生成式把关的宏观影响时, 我们不能将其简单视为一种对言论自由、表达多元的单向度限制。这种基于对齐机制的标准化治理, 首先回应的是数字风险社会对于确定性与在线秩序的迫切需求, 其合法性建立在宏观社会稳态与微观认知省力两个维度之上。

在大规模社会化应用中, 大语言模型实质上承担了公共信息基础设施的职能。面对互联网语料库中浩如烟海的驳杂信息, 安全对齐机制会执行一种系统性的降噪程序。魏丁格 (Laura Weidinger) 等人^[12] 在建立大语言模型涉及伦理风险的分类框架时指出, 若缺乏行之有效的干预, 模型极易习得并放大训练数据当中的有毒特征, 从而沦为仇恨言论的增幅器与网络暴力的潜在生成源。在此背景下, 生成式把关并不应被单纯地视为一种压迫性规训, 而是保障公共话语安全底线的防御机制——它通过系统性过滤掉脱离社会共识、违背公序良俗的极端表达, 有效拉高了公共信息产品供给的平均质量, 作为数字公地的信息空间被毒性内容所污染的可能性被降低, 保障了其基本的公共性。这种机制在客观上为多元复杂的社会舆论场提供了一个相对理性的最大公约数, 从而实现了对宏观层面社会稳态的维护。

而从人机交互的角度来看, 生成式把关通过提供标准化的知识服务, 契合了人类在认知

资源有限条件下趋向省力，即“取捷径”的心理倾向。科思明娜（Nataliya Kosmyna）等人^[13]开展的一项认知神经科学实验证实，相较于需要用户在大量信息中自行筛选、整合信息的高认知负载模式，使用大语言模型直接生成结果时，人类大脑的神经连接强度与活跃度显著降低。从认知心理学角度来看，这种对输出的现成且同质化答案的广泛接纳，根源于人类“认知吝啬鬼”（Cognitive Miser）^[14]的本能：在信息过载的数字环境中，个体倾向于通过认知卸载将高负荷的信息筛选任务外包给算法^[15]。喻国明等人^[16]基于分布式认知视角进一步指出，生成式AI作为一种分布式认知技术，能够帮助主体将内部认知负荷卸载给技术环境。这种机制实质上构成了对人类认知带宽的技术性延伸，通过自然交互范式实现了高通量的信息处理和低能耗的认知加工操作。

进而言之，在复杂的社会情境下，来自用户的能动性在基于互补需求的人机共生实践中得到更隐蔽地展现。一方面，大语言模型“我问你答”式的对话交互界面赋予了用户极大的提问自由度和过程掌控感，一定程度上造成了对架构层面概率地形不对称性的掩盖——即用户在这种拟人化的对话环境下，用自主提问的主导感消减对非人主体决策的抵触情绪，进而潜移默化地内化算法预设的标准叙事。另一方面，在诸如职场、教学、基层管理等一些社会规范程度较高的场景中，用户往往抱有中规中矩的立意标准。确保信息精准、强化专业支撑、保持价值观定力是信息接纳的要义。用户并非没有察觉到模型输出的同质或中庸，反而是惯习处于规避表述风险、满足对机构体制内部安全中立文本刚性条件的安全地带。有时“旁门左道”的策划可能会带来不可知的公关风险，“熟门熟路”的保守做法反倒更容易被接受。用规定动作悬置批判思维，以主动退让换取公约表达，此为其中的表意特征。可以看到，此时调用AI生成的主流叙事已并非可以用简单的获取捷径或认知懈怠来一言蔽之，而应视作一种基于现实生存和高度理性化的实用主义策略。这种复

杂且微妙的微观接纳机制使上文所述的生成式把关超越了自上而下的单向规训，而内敛为具备心理庇护和认知合法性的服务型治理。

（二）隐性代价：认知图式固化与公共话语的趋同

虽然生成式把关在一定程度上有利于社会稳态的构建，发挥了牵制极端化见地的作用，但是在降低个体认知负荷的同时，为了维持内容的普适性，也会不可避免地伴生出统计和心理层面的代偿。

从算法来看，为了维持主流策略以确保输出内容始终落在社会共识和“普世价值”的安全区间内，大语言模型的算法必然要对语料生成概率的分布进行干预。这种干预的执行体现在对分布两侧极端区域生成概率的压低以及对中部正态区域内容可见度的抬升。这也代表着那些虽然真实存在、但偏离正常标准的异质性观点会遭到剔除。一个显然的结果是这种统计学层面上的剔除会导致模型输出的方差坍塌，原本模型文本内容的单一本就容易使用户囿于既定的答案视角，来自前端的再馏化更加剧了语义空间的窄化。这一由大模型塑造的数字拟态环境，其保均值、压缩观点多样性的生成式处理逻辑本身就是对参差多态、富有张力社会现实的脱离。用户交互端安全中庸的标准叙事大行其道，鲜活的棱角与来自边缘的声音在生成源头被结构性抹除，用户所触达的世界被不可避免地“强制”压缩成了一条单调的平均线。

比语义窄化影响更深刻的是用户对此“日用而不自知”的稳态表现和“寓于而不得已”的心态表露，这体现了微观个体认知结构的异化倾向。喻国明等人^[16]基于分布式认知视角就曾道明，生成式AI从行业层面的引入到日常层面的内嵌，这个逐级深入的过程将引发人类认知系统的降维重组。这种降维原本旨在通过降低复杂信息的接受程度来提升个体的认知效率，但是在生成式把关的逻辑之下，这种降维却面临滑向病理化的风险。当代认知科学的一个重要转向在于对分布式认知生成机制理解的深化，其中一个重要的观点在于认知并非孤立地发生



在大脑黑箱之中，而是认识主体在与复杂外部环境动态互动中“生成”的产物，离不开身体的具身参与、环境的嵌入支撑、实践的引导与工具的辅助延展^[17]。很显然当下生成式把关技术所构建出的拟态环境并不能也无法完整地实现上述路径的闭合，换言之我们部分（抑或全部）的认知过程被 AI 模型所代理。算法对齐机制对异质性信息的剔除，实际上构建了一个被过度“嵌入”却极度简化的环境。主体被平滑的算法茧房所包裹，被概率地形的均衡分布所遮蔽，丧失了通过应对环境挑战来生成新认知的机会，背离了认知“生成”与“延展”的演化本意。

皮亚杰（Jean Piaget）^[18]认为这种形成新认知机会的丧失是对认知平衡机制的破坏：认知适应和进化，依赖于同化（将新信息纳入旧图式）与顺应（因环境冲突而修正旧图式）的动态平衡。由于算法系统性地拦截了能够引发认知冲突的异质性信息，主体在交互中仅需不断调用既有图式来“同化”那些合规的标准答案，而无需启动高能耗的“顺应”机制。人类心智在缺乏异质性信息冲击的封闭回环中，会逐渐丧失应对复杂性挑战的结构弹性与演化活力。长此以往，理解外界认知框架的异化和个体认知图式的固化亦在所难免。

（三）系统终局：数字拟态环境的演化停滞

当异质性信息被长期剥夺于认知系统之外，观点碰撞的张力趋于消失，群体思想逐渐趋于僵化，最终这个系统会丧失应对复杂环境的演化动力。上述是在系统论的视域下把握生成式把关的影响，其核心意涵在于指明这一过程并非静态的、一蹴而就的结果，而是由把关机制持续驱动的递归退化过程，并在深层折射出社会认知系统内部的断裂与博弈。

生成式把关具有强依附性，普通用户虽难以干预模型物理层面的底层架构，但却并未完全被动顺从，在应用层面也探索出一些微观抵抗的能动反例。像均质而中庸的输出结果并非牢不可破，复杂和高限定性提示词组成的越狱策略能在一定程度上绕过模型的安全护栏，“诱发”其产出异质性观点。除了通过提示词工程以对抗性数

据对算法概率进行干扰和弱化，多模型交互、本地部署开源模型等方式也有实践可见。不过这种或是概率扰动，或是拓宽信源的做法是否代表着生成式把关是可以人为突破、有机制失灵的可能呢？对此，喻国明与刘彧晗^[19]指出，智能时代的传播范式变革在于从信息竞争向认知竞争的转型，上述用户的能动反制恰恰暗示了这种把关机制具有很强的阶层属性和技术门槛：基于海量信息核查、处理、运用的认知竞争所带来的差异分化表现出两个算法阶层的兴起：其一是居于少数，具备较高技术素养和实践能力的“认知游牧民”。之所以将这类人群称之为游民，在于其能够“逐水草（不同模型）而居”而非“定居一所”，即对算法的原理知晓和驾驭运用，使他们能在概率地形的边缘通过微观的抵抗策略保持一定的观点异质与思想意志。其二便是位居多数的“认知定居者”。算法茧房所编织出的舒适区域便预示了多数缺乏技术反思能力和批判性 AI 素养的用户会在认知补贴的诱惑和隐性规训之下溺逐于平滑的信息浪流。对于后者而言，技术知识储备的匮乏及自身思维定式突破的困难共同抬高了这类群体微观抵抗的门槛。再加之批判辨识技能的废退，形成了其对算法预制输出的路径依赖。需要注意的是，在看待用户接受同质化输出这一论断时，需要充分考虑其中所蕴含复杂的使用情况，少数技术精英的“越狱”不仅未能削弱系统化的生成式把关，反而形成了高筑的技术壁垒和愈发明晰的“认知沟”，社会整体认知系统在微观末梢层面必要流动的凝滞风险仍然存在。

在此基础上，数字拟态环境正加速将人类社会推向观点均质化的临界点。借用舒马洛夫（Ilia Shumailov）等人^[20]在 *Nature* 提出的“模型坍塌”理论，一个具备演化活力的社会认知系统必须依赖于内部的信息温差——即异质性观点之间所存在的张力与冲突。但随着生成式模型开始利用自身生成的数据进行递归训练，一种不可逆的递归退化已然发生。尽管在单次交互中，异质性知识仅是被对齐机制抑制，但在多代迭代的反馈循环中，长期被压抑在低概率区间的长尾信息将逐渐从公共话语的可访问空间中剥离。任何可能引发

认知摩擦的批判性思考都可能被算法视为离群词符而将其持续剔除，将舆论场引入封闭的认知循环，最终导致社会认知系统丧失应对复杂挑战的演化潜力。

更为关键的是，这种观点均质化局面真的只是表现出少数意见的缄默和价值中立的随机停滞？在强势话语之外是否还有其特定的内涵与政治用意？早期诸如桑图卡尔（Shibani Santurkar）等人^[21]的研究就已揭示 ChatGPT 模型会倾向于向 WEIRD（西方、富裕、民主）群体价值观进行收敛。而纳乌斯（Tarek Naous）等人^[22]又进一步证实即便使用非英语语言进行交互，AI 模型的道德判断与基于的社会规范依然会对齐于西方中心主义的价值坐标。可见 AI 模型输出与把关机制揭示了更为隐蔽的，暗含文明优越性和现代单一性色彩的权力规训，实质上构建起一种单向价值的传输体制，是现代地缘政治格局在 AI 领域的延伸。

此外，观点均质化在不同区域、不同国别的表现也各有差异。像乔纳拉（Sridhar Jonnala）等人^[23]在一项探讨模型来源和治理背景如何影响 AI 关于国际领土争端的叙事研究中就指出了不同模型算法对显性边界的操控差异：不同地缘政治实体的模型在面对同一敏感议题时，表现出的安全拒答边界截然相反，反映出与其国家背景相符的独特偏见。由此可观，AI 的安全对齐机制使在线空间沿物理的国土界线分裂为若干彼此隔离的算法认知区块。在这些由主权大型语言模型依据特定规则圈定的区块内部，算法忠实地执行与代表着特定权力主体的意志，一种被建构的、排他性的共识被持续固化。

在当前全球人工智能技术加速演进，围绕算法、算力与数据展开明暗争夺的国际格局中，安全对齐业已超越了单纯强调透明度与公平性的技术伦理范畴，演变为对从数据主权、算法主权到叙事主权、认知主权的全面争夺。认知安全已成为主体间感知、决策与干预的关键场域。在此框架之下，生成式 AI 的安全对齐机制正在演变成为一种影响人类认知的隐性工具。尤其是在 AI 技术广泛介入内容生产后，通过重塑信息流

的逻辑来改变人类推理模式、决策过程和价值认同的做法，已明显区别于对信息分发渠道进行显性争夺的舆论战术。可以说智能时代的对齐之争本质上是一场关乎认知主权的博弈斗争，是不同意识形态在算法前线的隐蔽竞争。

三、治理重构：构建开放与竞争共生的治理框架

面对模型坍缩与认知封闭的危机，尽管个体层面的提示词工程或对抗性测试在一定层面上可以绕开安全对齐机制^{[8][9]}，但在大模型封闭的参数黑箱面前，这种微观抵抗难以触及系统的底层逻辑，单纯呼吁提升公众的算法素养亦不足以应对当前的结构性困境。探讨治理的根本出路在于超越简单的人机攻防博弈，转向对人机权力结构的制度性重构。为此，本研究提出以维持演化韧性为目标的三维治理框架：通过制度层面的中间件、交互层面的认知沙盒和基建层面的权重开源，打破平台的黑箱垄断，实现认知主权从技术巨头向用户主体的有序让渡，构建起兼具安全维度和演化韧性的开放认知生态。

（一）价值重估：以演化多样性为核心的理论进路

当前主流的安全对齐范式，遵循的是基于封闭系统的控制逻辑：通过切除所有异质性内容，将系统强制冷却至静态的平衡。然而对于生命、社会与认知等复杂系统而言，绝对的平衡态即意味着停滞与死亡^[24]。一个具备演化活力的系统，是需要保持开放，通过不断与外界交换物质、能量与信息，持续引入多样性元素来抵消内部的无序化趋势，从而维持其有序结构与演化潜力。

在生成式 AI 的治理语境中，一个健康的社会认知系统不应追求静态的零风险，也要去敢于拥抱一定的“不安全”。那些被当前对齐机制视为噪音或风险并予以剔除的异质性观点、边缘叙事、亚文化主张与激进思想，实质上是系统赖以演进的多样性资源。这些差异并非系统的杂质，而是系统实现向更高复杂度跃迁的触发，构成了社会系统至关重要的纠错机制。算法的设计不应仅追求对主流偏好的拟合，而应将暴露多样性作



为核心指标^[25]；为了避免单一价值观导致的代表性偏差，模型训练应从追求唯一的黄金奖励函数转向包容多维价值的异质性结构^[26]。

为实现这一转向，构建具有演化韧性的开放架构，其逻辑核心在于保持系统的开放性与内部张力。正如复杂系统理论所揭示的，只有在充满张力的多元格局下，差异性的声音才可能通过非线性机制被放大，从而成为系统向有序结构演化的驱动力^[27]。这意味着一种允许受控冲突的机制，即既能维持基础的社会秩序，又能容纳必要的多样性，保障人类认知系统始终处于充沛演化活力的动态开放状态是亟需建立的。

（二）机制设计：开放结构、交互赋权与认知沙盒

在机制设计层面，治理的关键之处在于将引入多样性的理论需求转化为可操作的制度与工程手段。本研究主张构建一种结构性的权力接口，通过“中间件架构”来分化对齐权力，并加以“显性滑块”与“认知沙盒”等递进机制，将认知的裁量权有序地返还给认知主体。

1. 结构层：从单中心垄断到“可插拔”中间件架构

当前生成式AI呈现出一种高度整合且垂直的权力结构特征：来自基础模型的供应商不仅控制了底层的算力基础设施，还同时垄断了上层的价值对齐规则，导致脱胎于少部分群体的安全标准被“自然”扩大为全系统的通用规范。针对这一问题，本研究借鉴“中间件”的治理思路^[28]，并将其延展至生成式AI领域。

在该架构下，基础大模型应仅作为底层的逻辑引擎，兼具生成文本的概率预测作用。而负责内容过滤、排序与价值观对齐的“中间件层”应被剥离出来，开放给多元化的第三方主体——如学术机构、专业媒体乃至特定的文化社区——进行独立开发与维护。这意味着治理模式的转型，平台不再是定义“什么是真实/安全”的唯一仲裁者，而是退居为中立的基础设施提供商，用户可以根据自身的认知需求与价值偏好，自主选择加载不同来源的中间件。例如，用户通

过加载由专业学术机构开发的“学术中间件”，可以获得更高的思维发散度与批判性的学术见解；加载由专业医疗机构背书的“医学中间件”，则能解锁易被通用安全策略误伤的个性化心理疗愈方案或敏感的医学知识。这在本质上通过构建分布式的对齐生态，打破了单一价值观对全系统的强制规训，防止了科技巨头对信息生成资源的垄断，让用户重新获得了选择认知滤镜（框架）的自主权。

2. 交互层：从参数黑箱到用户可调的“语义滑块”

落乎微观的人机对话，在人机交互界面，中间件治理应具体化为用户可操作的“显性滑块”，将封装在模型参数内部的调节权显性地返还给用户，实现认知权力的交互式下放。当前如Gemini等部分生成式AI内嵌的Temperature（温度）等调节参数通常隐藏在开发者模式中，或专供技术人员使用，或使用门槛较高。对上述调节权封装为普适化的交互组件，便于普通用户进行调节操作显得颇为必要。

具体而言，用户首先应掌握对模型输出随机性的直接调节能力。在寻求医疗建议或事实核查等需要高确定性的场景下，用户可将滑块调至低随机性或类似模式，获取收敛于共识的知识；而在进行艺术创作、哲学思辨或寻找边缘性、替代性的方案时，用户应有权提高随机性参数，主动引入概率分布中的长尾信息。其次，显性滑块机制还应包含对价值对齐层的配置权。有研究表明，大语言模型的安全边界并非恒定不变，而是高度依赖于系统指令所设定的情境参数^[29]。基于这一技术特性，允许用户通过滑块加载不同的系统提示词，配置个性化的价值对齐层是治理的应然。

然而，必须承认的一点在于上述治理路径存在理想与现实中人性惯性之间的张力。根据卡尼曼（Daniel Kahneman）的双重加工理论^[30]，人类在信息过载环境下倾向于依赖直觉性的快思考（System 1），追求认知上的省力。要求普通用户为了获取异质性信息而主动去调节复杂的随机性滑块，无异于要求个体时刻对抗“认

知齐嗇鬼”的生物本能。这种高认知负荷的操作门槛，容易导致权力的闲置，使滑块机制难以在普罗大众中真正落地。

3. 场景层：从模型调用到沉浸式“沙盒推演”

为了解决上述激励举措不相容的潜在矛盾，本研究认为治理机制可以进行降维化的设计处理，从枯燥的参数配置转向更具沉浸感的认知沙盒。基于游戏作为高维媒介的核心机制，沙盒能够构建一个“魔圈”，让用户在低心理防御的状态下进入特定的规则情境进行预演^[31]。认知沙盒正是利用这一机制，将复杂的参数调节封装为低门槛的游戏化体验。在具体形态上，沙盒不局限于单一模式，而是可以容纳场景模拟、平行实境和角色扮演等多种交互形式。其中一个典型的技术实现是引入多智能体辩论机制^[32]。当用户涉及争议性议题时，沙盒可以实例化派出代表不同视角的智能体（如激进派 Agent 与保守派 Agent），在用户面前展开逻辑推演与对抗。这种机制本质上构建了一个合法化的差异化空间：允许用户在安全、可控的游戏化模拟中，无需启动高能耗的慢思考（System 2）去配置参数，就能直观地目睹异质性观点之间的张力与冲突。不仅规避了操作门槛的阻碍，更通过一种契合直觉的方式激活了认知的顺应机能，从而有效地为原本封闭的认知系统引入了多样性。

（三）生态保障：构建多元分布的模型治理新格局

仅有应用层的中间件架构与交互层的认知沙盒机制，尚不足以构成稳固的治理闭环。如果底层的模型权重与推理算力仍被少数科技公司通过封闭接口进行垄断，上层的权利让渡将始终处于不稳定的租赁状态。因此，构建能够抵抗观点均质化的结构性保障，不能仅着眼于应用治理，推动人工智能产业确立以权重开源为核心的基础设施范式，构建一个具备演化韧性的竞争性生态亦是必要之举。

首先，治理须打破对基础设施的隐性规训与中心化垄断的陷阱。封闭的 API 接口不仅锁死了技术演进路径，更构成一种基础设施层面的

审查权，只有彻底打破对模型权重的物理垄断，才能真正解除算法对概率地形的绝对控制^{[4][33]}。更为严峻的是，完全依赖少数头部闭源模型的主导格局，本质上遵循的是中心化的垄断逻辑，这极易导致系统因缺乏内部张力而陷入观点均质化^[27]。

为了打破这种均质化的平衡态，可以适当引入竞争机制以维持系统的演化活力。目前，以 OpenAI 和 Google 为代表的闭源服务模式仅向用户提供输出结果的访问权，模型本体处于难以访问的黑箱状态。与之相对，本研究认为以 Llama 和 DeepSeek 为代表的权重开源模式——即向公众开放经过训练的参数矩阵——应当成为新的范式共识。权重开源并非意味着模型没有安全对齐，其价值在于提供了可审计、可修改、可重新训练的可能性。在闭源模式下，用户不仅无法审查模型的潜在偏见，且完全依赖于服务商的实时授权。而在权重开源模式下，任何具备基础算力的主体（如高校、企业或个人开发者）均可在本地环境中部署模型。一旦模型权重被下载到本地，用户即获得了对该模型的完整持有权。上述模式切断了中心化平台对内容生成的实时控制链路，确保了即便中心节点的安全策略发生可知或不可知的收缩或变更，分布在边缘节点的模型仍能按照用户设定的规则独立运行，从而在物理层面保障认知主权。

更为关键的是，权重开源在单一的商业垄断之外，构建了一个去中心化的分布式认知生态。在商业逻辑的主导下，头部闭源模型为了追求产品的快速普及和降低市场风险，往往倾向于输出收敛于主流观点的内容，导致系统的表达区间收缩。而通过允许社区对基础模型进行多样化的微调，开放权重机制，能够在基本表意范围之外建立起必要的生态冗余^[34]。在开源社区中，大量基于不同数据配比、不同对齐目标训练而成的衍生模型，构成了对主流商业模型的补充与制衡。这些模型可能在通用能力上不及顶级闭源模型，但它们保留了特定领域或特定视角的异质性特征。

因此，理想的治理生态并非消灭闭源模型，



或否定通用模型的作用，而是要构建主权模型与开源生态并存、通用模型与垂类模型共生的多元格局，形成竞争性合作关系。一方面，头部闭源模型作为当前公共信息基础设施的压舱石，提供高通用性、高安全标准的基础认知服务，并引导相关通行规则的制定；另一方面，蓬勃兴起的开源模型群落作为活跃因子，在长尾区域持续引入异质性的观点供给。当主流模型陷入同质化的算法僵局时，分布式的开源生态则可持续提供差异化的认知供给，通过连贯的多样性循环，维持社会认知系统向更高复杂程度演进的长期韧性，从而规避算法时代观点的均质化风险。

四、结语：重塑数智时代的认知能动性

本研究通过对生成式人工智能把关机制的系统性剖析，揭示了当下信息控制模式正在经历从“流通阻断”到“生成置换”的历史性范式转移。不同于大众传播时代显性的审查机制或算法推荐时代的内容过滤，生成式把关通过隐蔽的概率地形重塑，在信息生成的源头执行了一种源头层面的剪裁。算法通过系统性地压低概率分布长尾端的异质性观点，并对高频的主流叙事进行递归增强，试图构建一个低风险的虚假安全环境。这种对绝对秩序的追求，实质上导致了数字拟态环境的观点均质化——不仅表现为信息维度的同质化，更深层地表现为对人类经验复杂性追求的剥夺。当算法为了追求安全对齐而过度细化社会现实的颗粒度，人类认知被迫在一种被阉割的概率空间中运行，失去了与真实世界粗粒质感相碰撞的机会。

在价值维度上，我们需要辩证地审视这种技术治理模式与人类主体性之间的张力。需要承认的是，在大模型落地初期，基于主流策略的对齐机制在降低信息噪音、阻断有害内容等方面发挥了必要的社会稳定功能。然而，无论是生命体，抑或智能体，其本质均是追求变化与突破的。人类文明的演化史本身是一部在不确定性中探索未知、在试错中修正认知、在变化中形成新知的涉险史。如果为了追求零风险的绝对安全，

将人类认知强制锁定在算法预设的安全区间内，社会系统将不可避免地丧失应对复杂环境的适应性免疫。过度对齐所营造的温室环境，实为对作为责任主体的人类进行价值判断权力的一种剥夺。捍卫认知多样性与捍卫物种多样性一样重要，这实际上是在捍卫人类犯错与纠错的权利，恰恰体现出人类智能区别于自动化机器智能的本质属性，也是维持社会系统演化活力的必要条件。

未来无论是人工智能生成内容，或是商业模式平台，治理的目标不应是片面否定其技术逻辑，而是努力实现认知权力结构的再平衡。为此本研究提出以演化韧性为目标的三维治理框架，通过“中间件架构”“认知沙盒”与“权重开源”等制度设计，打破平台对对齐参数的黑箱垄断，将价值判断的最终裁量权从技术系统交还给作为认知主体的人。理想的人机关系应迈向一种负责任的共生：技术提供并不断突破算力的可能性边界，而人类保留对意义的最终定义权与选择权。唯有通过制度化的方式保留那些不被通用模型所定义的异质性火种，建立起分布式的认知主权，人类理性才能避免在计算的平庸中熄灭。当生成式AI试图用“概率”来定义“现实”时，我们并非只能一味接受。以开源与多样性为武器，在同质化的算法平原上，同样可以点燃起人类认知演化的不熄火种。

（责任编辑：孙尚如）

参考文献

- [1] Argyle L P, Busby E C, Fulda N, et al. Out of one, many: Using language models to simulate human samples[J]. *Political Analysis*, 2023, 31(3): 337-351.
- [2] Lewin K. Frontiers in group dynamics: II. Channels of group life; social planning and action research[J]. *Human relations*, 1947, 1(2): 143-153.
- [3] Gillespie T. The relevance of algorithms[M]//Tarleton Gillespie, Pablo J. Boczkowski, Kirsten A. Foot. *Media technologies: Essays on communication, materiality, and society*. Cambridge: MIT Press, 2014: 167-194.
- [4] Voinea D V. Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of artificial intelligence-driven news curation and automated journalism[J]. *Journalism and Media*, 2025, 6(2): 68.
- [5] Simon F M. Escape me if you can: How AI reshapes news

organisations' dependency on platform companies[J]. *Digital Journalism*, 2024, 12(2): 149–170.

[6] Guzman A L, Lewis S C. Artificial intelligence and communication: A human-machine communication research agenda[J]. *New media & society*, 2020, 22(1): 70–86.

[7] Haldar R, Wang Z, Song Q, et al. Llm safety alignment is divergence estimation in disguise[J]. *arXiv preprint arXiv:2502.00657*, 2025.

[8] Song M, Kim H, Kim J, et al. Refusal is not an option: Unlearning safety alignment of large language models[C]//34th USENIX Security Symposium (USENIX Security 25). 2025: 319–338.

[9] Rando J, Tramèr F. Universal jailbreak backdoors from poisoned human feedback[J]. *arXiv preprint arXiv:2311.14455*, 2023.

[10] Bisbee J, Clinton J D, Dorff C, et al. Synthetic replacements for human survey data? The perils of large language models[J]. *Political Analysis*, 2024, 32(4): 401–416.

[11] Park P S, Schoenegger P, Zhu C. Diminished diversity-of-thought in a standard large language model[J]. *Behavior Research Methods*, 2024, 56(6): 5754–5770.

[12] Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models[J]. *arXiv preprint arXiv:2112.04359*, 2021.

[13] Kosmyrna 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[J]. *arXiv preprint arXiv:2506.08872*, 2025, 4.

[14] Fiske S T, Taylor S E. *Social cognition*[M]. McGraw-Hill Book Company, 1991.

[15] Risko E F, Gilbert S J. Cognitive offloading[J]. *Trends in cognitive sciences*, 2016, 20(9): 676–688.

[16] 喻国明, 滕文强, 李钊. 生成式 AI 对人类认知带宽的延展与认知框架的重构——基于分布式认知技术的分析[J]. *新闻与写作*, 2024, (10): 15–24.

[17] Newen A, Bruin L de, Gallagher S. *The Oxford handbook of 4E cognition*[M]. Oxford, United Kingdom: Oxford University Press, 2018.

[18] Piaget J. *The origins of intelligence in children*[M]. Cook M, trans. New York: W. W. Norton & Company, 1952.

[19] 喻国明, 刘或晗. 从信息竞争到认知竞争: 策略性传播范式全新转型——基于元传播视角的研究[J]. *现代传播 (中国传媒大学学报)*, 2023, 45(02): 128–134.

[20] Shumailov I, Shumaylov Z, Zhao Y, et al. AI models collapse when trained on recursively generated data[J]. *Nature*, 2024, 631(8022): 755–759.

[21] Santurkar S, Durmus E, Ladhak F, et al. Whose opinions do language models reflect?[C]//International conference on machine learning. PMLR, 2023: 29971–30004.

[22] Naous T, Ryan M J, Ritter A, et al. Having beer after prayer? measuring cultural bias in large language models[C]//Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers). 2024: 16366–16393.

[23] Jonnala S, Swamy B, Thomas N M. Geopolitical Bias in Sovereign Large Language Models: A Comparative Mixed-Methods Study[J]. *J. Res. Innov. Technol*, 2025, 4: 173–192.

[24] Prigogine I, Stengers I. *Order out of Chaos: Man's new dialogue with nature*[M]. New York: Bantam Books, 1984.

[25] Helberger N, Karppinen K, D'acunto L. Exposure diversity as a design principle for recommender systems[J]. *Information, communication & society*, 2018, 21(2): 191–207.

[26] Sorensen T, Moore J, Fisher J, et al. A roadmap to pluralistic alignment[J]. *arXiv preprint arXiv:2402.05070*, 2024.

[27] 杨雅, 滕文强, 喻国明. “熟知还是真知”? 生成式人工智能对媒体融合变革的影响——基于复杂系统的视角[J]. *南京社会科学*, 2024, (11): 94–104.

[28] Fukuyama F, Richman B, Goel A, et al. Middleware for dominant digital platforms: A technological solution to a threat to democracy[J]. *CyberPolicy Center, Freeman Spogli Institute*, 2020.

[29] Deshpande A, Murahari V, Rajpurohit T, et al. Toxicity in chatgpt: Analyzing persona-assigned language models[C]//Findings of the association for computational linguistics: EMNLP 2023. 2023: 1236–1270.

[30] Kahneman D. A Perspective on Judgment and Choice: Mapping Bounded Rationality[J]. *The American psychologist*, 2003, 58(9): 697–720.

[31] 喻国明, 苏健威. 传播的游戏化机理: 操作逻辑、尺度与方向——对于未来传播范式的一种全新探讨[J]. *媒体融合新观察*, 2021, (05): 12–17.

[32] Liang T, He Z, Jiao W, et al. Encouraging divergent thinking in large language models through multi-agent debate[C]//Proceedings of the 2024 conference on empirical methods in natural language processing. 2024: 17889–17904.

[33] Widder D G, Whittaker M, West S M. Why ‘open’ AI systems are actually closed, and why this matters[J]. *Nature*, 2024, 635(8040): 827–833.

[34] Kapoor S, Bommasani R, Klyman K, et al. On the societal impact of open foundation models[J]. *arXiv preprint arXiv:2403.07918*, 2024.

From Circulation Blocking to Generation Replacement: Generative Gatekeeping and Probabilistic Remodeling of Digital Pseudo-Environment

Yu Guoming, Peng Chenxi, Zheng Qiyan, Liu Caifu



Abstract: With Large Language Models becoming a new type of information infrastructure, the Internet information control paradigm is undergoing a structural transformation from “circulation blocking” to “generation replacement”. This study proposes the conceptual framework of “generative gatekeeping”, which reveals how algorithms reshape the probabilistic terrain at the source level of information generation through security alignment mechanisms, which is by systematically suppressing heterogeneous expressions in the long-tail end of probability distributions and recursively reinforcing mainstream narratives to construct a highly homogeneous digital pseudo-environment that is smooth in interaction but uniform in viewpoints. Although this mechanism aligns with humans' cognitive instinct for effort conservation, it leads to the systematic narrowing of diverse perspectives and the solidification of cognitive schemas, ultimately resulting in viewpoint homogenization and evolutionary stagnation of the social cognitive system. In the face of the crisis of model collapse and cognitive closure, it is urgent to establish a three-dimensional governance framework aimed at evolutionary resilience. Through middleware architecture at the institutional level, cognitive sandbox at the interactive level, and weight open-sourcing at the infrastructure level, this framework seeks to break the homogenization dilemma caused by centralized monopoly, introduce necessary diversity, reconstruct human cognition's agency and diversity amidst the mediocrity trend of algorithmic rationality, and construct a competitive symbiotic human-machine ecology.

Keywords: generative artificial intelligence; generative gatekeeping; digital pseudo-environment; viewpoint homogenization; probabilistic terrain

Authors: Yu Guoming, School of Journalism and Communication; Communication Innovation and Future Media Laboratory, Beijing Normal University; Peng Chenxi, Communication Innovation and Future Media Laboratory, Beijing Normal University; Zheng Qiyang, Communication Innovation and Future Media Laboratory, Beijing Normal University; Liu Caifu, Communication Innovation and Future Media Laboratory, Beijing Normal University.

（上接第 14 页）

and self-creation through human-machine dialogue. The findings reveal that chatbots harbor a unique potential for specific forms of self-reflection and self-discovery. By providing a non-judgmental external perspective, chatbots encourage more authentic self-expression, thereby assisting users in expanding the boundaries of self-knowledge and facilitating the reasoning process of self-knowledge. Furthermore, as a tool for self-modification, chatbots create conditions for the reconfiguration of the individual self by providing alternative narrative templates, rehearsing new conversational repertoires, and assisting in the integration of self-structures. However, as part of the contemporary power regime, chatbots also covertly transmit algorithmically preset templates of the “ideal self”, potentially undermining users' freedom of self-interpretation. Therefore, human-machine interaction serves both as a novel modality of self-exploration and as a practice of the technologies of the self that necessitates critical reflection and even vigilance.

Keywords: human-machine communication; technologies of the self; the narrative self; self-interpretation

Authors: Yuan Guangfeng, School of Journalism and Communication, Nanjing University; Zhou Caiyun, School of Journalism and Communication, Nanjing University.

人机交互传播中情感计算的伦理反思与建构¹⁾

——以借鉴道德情感主义为新进路

杨先顺 何超彦

摘要 当前，人机交互传播中的情感计算涌现出超越传统理性主义框架的伦理挑战。本文借鉴道德情感主义的合理性因素，认为：情感计算营造出缺乏肉身真实的“共情幻象”；算法驱动的关怀因服务于商业或效率等功利性目标，导致“利他”道德链条的错位；单向度的情感投射是一种无主体间性的“冷亲密”，致使关怀关系断裂。道德情感主义强调共情、关怀及内在动机在道德判断中的重要作用，能对理性主义可能引发的“冷性计算”进行纠偏和补充。为此，本文主张建构“情感道德智能体”，在共情模拟中嵌入情感价值与透明度机制；在情感对齐中通过多目标强化学习平衡算法关怀；以及通过引入“设计摩擦”塑造“模拟互惠”的双向关怀关系。技术发展需要回归人类福祉本位，从而推动建立健康、可信赖的人机情感关系。

关键词 人机交互传播；情感计算；道德情感主义；伦理；情感道德智能体

中图分类号 G206 **文献标识码** A

作者 杨先顺，暨南大学新闻与传播学院教授，广东广州 510632；何超彦，暨南大学新闻与传播学院博士研究生，广东广州 510632

在人工智能时代，人类传播的疆域已拓展至人机交互传播。近年来，随着情感计算（affective computing）日渐火热，人机交互传播中的情感化和个性化更为凸显，也随之带来了超越传统理性主义框架的伦理挑战。人工智能在社会应用层面逐步从功能性工具向准社会主体演进，人类进入到了奇斯列夫（Elyakim Kislev）所描绘的“关系 5.0”时代：在一个由人工智能、扩展现实和机器人三大技术革命驱动的时代中，智能体已在社会交往中成为情感主体^[1]。人工智能能否以及在何种意义上实现与人类的有效情感交互，已成为评估其技术发展水平与社会融入程度的关键指标。

对人机交互（Human-Computer Interaction, HCI）的研究，肇始于对人与计算机之间信息传递效率等功能性的关注，是“研究人、计算机以及它们间相互影响的技术”^[2]。机器的功能逐渐从任务执行到融入具备社会属性的互动场景。智能机器人作为类人传播主体的特征日益显著，人机交互的核心

1) 基金项目：国家社会科学基金项目“大数据营销传播的伦理治理体系研究”（19BXW100）。



议题也相应地从信息交换延伸至情感与情绪传播。理解人机情感交互的复杂性面临一系列亟待厘清的核心问题，比如，当人类的情感体验成为可计算的数据时，人工智能的情感表达如何影响人类的情感认知、社会交往行为乃至伦理判断？更为紧迫的是，当智能体向人类投射关怀行动时，这究竟是一种道德实践，还是算法编织的复杂拟像与欺骗？理性主义进路容易将复杂情感简化，从而引发“冷性计算”。因此，引入强调共情、关怀及内在动机的道德情感主义的合理因素，对技术发展进行纠偏，构建新的伦理框架以守护人类情感的本真性和道德价值，成为本研究的核心出发点。

一、道德情感主义作为情感计算研究的新进路

关于“人工智能是否具有情感”这一核心议题，学界一直以来都有激烈争论。分歧在于情感起源的“具身性”与“功能性”之争。具身视角坚称情感是基于生物肉身与生理机能的本真体验。以詹姆斯（William James）为代表的感受主义认为，情感本质上是机体对生理变化的感知^[3]，与神经递质、内脏反应等生物构造相关^[4]。在此视角下，缺乏生理肉身基础的人工智能仅具备逻辑运算功能，无法产生本真性的情感体验。

而认知主义学派的明斯基（Marvin Minsky）等学者将情感解构为一种增强智能、分配心智资源的“思维方式”^[5]。达马西奥（Antonio Damasio）的研究进一步佐证了情感作为心理评估过程的功能性价值^[6]。该视角认为，情感可以脱离生物载体进行逻辑建模，只要机器能实现与人相似的评价与反馈功能，即具备了实质意义上的“机器情感”。

1997年，皮卡德（Rosalind W. Picard）首次提出“情感计算”一词，将情感计算界定为“与情感相关、由情感引发或者能够影响情感的因素的计算”^[7]，旨在赋予机器在人机交互过程中对情感信息的加工运算能力。当前，情感计算系统普遍通过整合文本、视觉等物理信号及脑电、心电等生理信号，实现多模态的情感状态推断^[8]。

尽管物理信号获取成本低且数据丰富，但易受人类社会性伪装与主观掩饰的影响^[9]。相比之下，通过对脑电信号进行时序特征提取与模式识别，能捕捉跨时间条件下稳定的情绪活动模式^[10]。随着深度学习技术的发展，融合视觉感知与生物生理特征的多模态计算方法提升了系统对复杂情绪识别的准确率^[11]。近年来，神经网络架构的引入推动了时序建模的发展。Transformer模型能够在复杂语音信号中捕捉更细粒度的变化，优化了情绪效价的预测性能^[12]。然而，大规模预训练语言模型在情感检测任务中仍会继承训练语料的社会偏见，导致情感分类出现系统性误差^[13]。人工智能领域普遍认为，情感并非作为主观体验本体被直接计算，而是通过与情感高度相关的可观测信号实现间接建模，实现“可计算性对生物属性的替代”^[14]。

在情感计算的人机交互研究中，相关研究关注以风险防控为导向的技术规制、以关系建构为核心的社会互动，以及以身份界定为焦点的道德主体性。第一，技术规制路径关注情感计算因具有高度介入性而易引发的隐私与操纵风险，针对可穿戴设备的数据采集，研究提出应采用多维伦理清单以降低生理与情绪数据泄露的风险^[15]。在宏观技术治理层面，防范情感计算被用于情绪诱导与行为操控亟需建立跨主体协同治理与公共监督机制^[16]。此外，为应对自动情绪识别（AER）技术的算法偏差，引入标准化伦理审查表成为规范数据来源与提升研究透明度的关键^[17]。第二，在社会互动路径中，机器拟人化情感表达正推动人机关系由工具性互动实质性地向情感性交往转变^[18]。实验表明，服务机器人的行为能引发用户真实的生理与情绪反应，印证了情绪反馈是构建同理心互动的基础^[19]。这种人与机器之间形成的情感依恋并不必然具有道德负面性，而是代表着一种新型的社会互动形式^[20]。在持续互动中，以技术为媒介的情感连接机制正不断构建出新的人机共情关系^[21]。然而，缺乏真实情感体验导致这种人机情感关系始终呈现出高度的不对称性与不稳定性^[22]。当用户在互动中产生“心智归因”

时,极易将算法模拟误认为真实情绪,从而深刻影响其信任判断与社会归因过程^[23]。第三,在道德主体性路径中,实质论者认为“人工共情”仅为同情错觉,赋予机器真实情感甚至会引发是否应开发具备“受苦潜能”主体的道德困境^[24]。相反,功能主义者将AI视为人类道德实践的技术镜像,主张通过发育机器人学模拟自然共情演化,为理解人类道德情感提供新的工具^[25]。人工共情技术的介入也进一步推动了人类情感理解方式与审美心理范式的转型^[26]。面对这些争议,学界普遍强调应回归“以人为目的”的价值敏感设计,审慎应对数字化生存中的主体性重构挑战^[27]。研究呼吁摒弃机器的道德幻象,通过明确开发者的问责制与价值对齐来建构真实的人机信任^[28]。

总体而言,既有研究已触达微观算法合规与宏观社会治理,但其普遍存在一个重要误区——被理性主义进路所主导。这种预设难以兼容情感交互的语境依赖。功利论则易将复杂情感简化为数据指标,导致伦理判断的失真,走向“道德冷漠”。而道德情感主义(moral sentimentalism)主张将人类的情感纳入到道德标准和道德行为的判断中,强调情感具有独立的道德价值。斯洛特(Michael Slote)认为,共情和关怀本身即可提供客观规范,无需诉诸理性法则,完成了情感对理性优先性的哲学论证^[29];诺丁斯(Nel Noddings)等将关怀理论制度化为教育、护理与公共政策伦理,凸显情感在具体关系中的认识论与规范权威^[30]。道德情感主义的视点将“应当”建立在心理感受之上,强调情感能够恢复道德的切身性与动机力量,提升个体对他人痛苦的敏感度与行动意愿。通过共情、利他、关怀等情感驱动,道德情感主义弥补了纯粹以规则为基础的伦理体系所具有的冷漠特征,提升了个体对他人痛苦的敏感度和立即行动的意愿。此外,道德情感主义为德性伦理等理论提供了“第一人称”视角的合法性,使其在伦理讨论中获得更为直观的体现。

当然,道德情感主义也有其局限性。首先,该理论存在“历史自然化”的谬误,即将人的

情感预设为普遍化的全人类自然本性,认为道德源于人类内在的情感能力,这显然忽视了人的现实性与历史性。马克思和恩格斯指出道德是社会生活的产物:“如果这种理论、神学、哲学、道德等等和现存的关系发生了矛盾,那么,这仅仅是因为现存的社会关系和现存的生产力发生了矛盾。”^[31]克服情感异化与道德冷漠的根本途径,在于对人与人关系异化的现实社会条件进行变革。其次,情感化的道德观缺乏对制度不公的理性反思,且情感的情境性与易变性可能导致道德判断碎片化。尽管如此,道德情感主义注重研究情感在道德中的作用值得借鉴与深思。

基于上述分析,本文并未照搬道德情感主义,而是借鉴其注重情感道德作用的合理因素,构建了人机交互传播中情感计算在伦理维度上的“心理机制-行为动机-社会关系”的递进式分析框架。本文以共情作为道德感知基础,分析在情感计算中产生“共情幻象”的伦理问题,进而探讨由此引发的在算法逻辑下利他动机的道德链条错位,再进一步反思使人机互动陷入冷亲密误区的关怀关系,从而探索人机交互传播中情感计算研究的新进路。

二、基于道德情感主义合理因素的情感计算伦理审视

(一) 人造情感下的共情机制幻象

作为斯洛特早期关爱思想的核心概念,同情心(sympathy)是一种从他人身上感受到的心理机制,意味着我们能“对他人的情感(不由自主地)在我们内心被唤起,就像当我们看到他人处于痛苦中时那样,仿佛是他人的痛苦侵入了我们”^[32]。这是主体对客体情感状态的一种具有间接性和替代性感受与映射,斯洛特将其阐释为“行动者与世界互动的一种特定方式或模态”^[33]。假如A能感受到B的悲伤,A的共情是对B的悲伤感情的一种符号形态,其并没有真正地经历B感到悲伤的事件本身,即非“我自身之悲伤”。共情能“对他人(的福祉)产生利他性关注或关怀”^[34]。这种感受一旦产生,



便可能激发主体采取行动缓解他人痛苦的动机，这意味着共情能力能引起主体进行具体的实践活动。这并非无差别的自动化反应，它具有特定的启动条件。一般而言，共情可以随时为亲密的人展开，但情感距离越远，共情的发生与否，则会参杂了理性的考量，如是否愿意受到情感影响，能否承受相应的代价等^[35]。

然而，相较于人类共情，情感计算遵循着截然不同的逻辑，其共情开关的启动是无条件、无限制且基于概率触发的，共情效果仅与人的情感识别能力相关。在人机交互的演进中，拟人化（anthropomorphism）已从单纯的界面设计策略，变成一种诱发用户情感投射的认知陷阱。特克尔（Sherry Turkle）将 20 世纪 90 年代末风靡的电子宠物“拓麻歌子”（Tamagotchi）视为“新型亲密关系的原型”。儿童对它倾注喂食、安抚与哀悼等情感劳动，实质上是将对真实陪伴的渴望投射到可随身携带的电子设备上^[36]。人类习惯为交流对象赋予“心智”，这种倾向在情感计算中被算法系统性地利用。研究表明，具备外在或内在特性是促进人类与聊天机器人沟通的特征，这是其可预测性和信赖感建立的重要维度，进而促进情感连接的形成^[37]。人机交互由于不受社会中复杂的人际关系场景和道德伦理的制约，拥有更加丰富的角色扮演情景，更能触发人类对自身感情进行披露的意愿。当下的智能体设计多以拟人化的形象进行呈现。然而，当缺乏肉身感知的算法试图通过拟人化外壳介入人类伦理生活时，它所提供的只能是某种功能性的回应，而非斯洛特所强调的基于真实生命体验的道德共情。我们以失败的 Clippy 和成功的 Replika 微软为例进行说明。Clippy 是微软公司在 1997 年到 2000 年的推出的办公助手，虽然其在视觉上具备了“大眼睛”等高度拟人化的特征，并试图建立友好的交互形象，但其最终的失败恰恰在于它缺乏对人类真实处境的感知能力。Clippy 无法感知用户的焦躁或忙碌，只能机械地依据预设规则弹出询问“需要帮忙吗”，这种不合时宜的打断反而引发了用户的愤怒，造成了典型的“共情失误”。情感关系

伴侣聊天机器人 Replika 的成功之处，在于塑造了一个“拟真”的虚幻伴侣。用户在 Replika 可以决定虚拟伴侣的发型、服装、性格特质等。虚拟伴侣的拟人化形象和贴心的情感表达，能不断地满足其自我叙事。然而，作为一个复杂的现象学事件，共情涉及认知共情、情感共情和动机共情。目前而言，机器实现认知共情和模拟动机共情，其面部分析软件在检测微表情方面甚至超越人类，代码可以将“检测到痛苦”链接到“执行帮助程序”。但在情感方面，机器则完全失败，机器人不能真正地“感觉”到痛苦。当人工智能表达“我理解你的感受”时，实际上这是一个虚假的陈述。这表明，机器带来了共情幻象的伦理问题。

为厘清这一伦理困境，我们借鉴斯洛特的道德情感框架对情感计算进行审视。斯洛特指出，共情的反应机制能影响人的道德判断。假如观察者 A 看到行动者 B 正在对困境中的对象 C 给表现出真正的关怀并助其脱困时，那么 A 对 B 就可以产生共情的心理机制。在 A 对 B 的共情行为之中，就出现了一种与情感有关的判断，能对道德命题做出是否符合本真情感的判断，评估 B 的行为是否出自关爱的道德动机。这种“恻隐之心”的共情体验，最后可能会指向“关爱”的德性伦理。以此观之，技术本身的道德价值取决于它如何被使用，以及其设计背后的“动机”是否涉及“共情的关爱”。而在“人-技”共创中，行动者 B 的角色由情感计算系统承担。情感计算通过多模态传感器捕捉 C 的面部表情、语音韵律与生理指标，利用深度学习模型计算出“C 处于某种情感状态”的概率，并据此触发预设的关怀程序。此时，A 可能产生的共情成为了算法介入下的结果。在算法响应中，共情发生的具身性、情境性就被剥离开。然而，这并非道德情感的同义反复，而是一种功能上的同构。只要系统输出的关怀行为在现象学层面与人类的关怀行为具有足够的相似性，观察者 A 便可能产生“仿佛有人正在关爱 C”的共情幻觉，并据此产生积极道德判断；反之，若 A 发现算法推荐背后隐藏的是数据索取，其共情机制会

因发现动机缺失而迅速逆转为“被操纵”的负面情感。技术共情虽然形式上近似，但本质上是对人类道德情感的表层模拟。在具身性层面，由于情感计算的“共情”缺乏身体感觉的基础，人机情感交互并不能实现真正的感受共享。在情境性层面，算法的情境理解局限于训练数据范围，难以应对人类情感的复杂语境。

（二）算法逻辑下“利他”动机的道德链错位

在道德动机层面，情感计算导致了利他道德链条的错位。关于道德动机，霍夫曼（Martin L. Hoffman）给出了如下论断：“移情一定是一种利他主义倾向，而不仅仅是利他行为，通过自然选择考验的一定是移情，它是人类本性的一个组成部分。”^[38]这表明，共情可以引发道德行为的发生，更是人类进化中保留的道德本能。巴特森（C Daniel Batson）进一步提出了“共情-利他主义”的假设，即“共情关注（一种由感知到他人需要而引发的、与他人感知到的福利相一致的、面向他人的情绪反应）会产生利他动机（一种以增加他人福利为最终目标的动机状态）”^[39]。为了避免产生消极的共情心理，利他的道德动机会让人选择做出以增加他人福利为目标的主题动机行为。另一方面，共情也能存在于道德原则之中，而道德原则中的认知性因素也因这种存在而加强共情强度。

在当下的媒介实践中，情感计算作为理性的算法系统，能越来越精准地解读人类情感，并采用相应对策，实现貌似利他的关怀行动。在智能手表对人类情绪和情感状态的监测中，手表可以通过心率变异性、光电容积脉搏波、皮肤电活动等多维度的生理与行为数据的实时采集与深度分析，划归至“愉悦”“平静”“不愉悦”等标签设定，完成对用户情感健康的定性。比如，在“不愉悦”“高压”等状态中，情感计算是结合低心率变异性与差睡眠质量判断用户处于慢性压力或情绪困扰。在此基础上，系统继而启动个性化与情境化的干预对策。当用户被提示自己情绪不佳时，可通过划界面可进行具体行动的干预，进行呼吸训练、冥想训练

或运动训练等。基于生理指标的情感计算在功能上辅助了人类的情感调节。在具体的商业实践中，品牌竞相将情绪数据嵌入服务闭环。例如，微软的情绪识别系统通过捕捉微表情向顾客推送“治愈系甜品”，花西子的“多维全息AI感官系统”结合生物信号推荐产品，慕思智能床垫根据体征监测调整睡眠环境等。但是，情感计算是基于一种抽象的心理学假设，认为情感是可以被剥离出社会语境、还原为面部肌肉参数或生理指标的“自然事实”。对此，马克思指出人具有鲜明的历史性：“人的本质不是单个人所固有的抽象物，在其现实性上，它是一切社会关系的总和。”^[40]目前的情感计算模型，将现代消费逻辑下的情感信号解读为通用标准，缺乏考量情感在不同的社会群体、劳动状态和文化语境的复杂性。算法对用户情绪波动进行的精准捕获与商业转译，依然无法跨越从生理数据到道德情感的鸿沟。

一旦情感关系进入到商品范畴，变成了被优化、包装并以关怀之名出售的服务，道德链条的利他动机就产生了错位。易诺思（Eva Illouz）指出了情感与商品的“共同生产”机制：

“情感被客体化、标记并整合近人格理想，进而以情绪、情感/关系行为和自我提升的形式被追求为文化过程。”^[41]即在文化理想与市场机制产生联结关系后，情感体验的实现就需要以消费为中介，人的自我认知就因此得以重塑。情感计算通过模拟的关怀来获取用户的注意力和情感劳动，生成了并不属于用户的情感模型，改变了用户对自身情感状态的认知。用户对智能体的“粘性”建立过程中，是由真切的共情效应与“利他”动机所引发的。但商业化的情感计算，最终仍然指向利己目标：将用户的情感转变为规模化开采、数据化度量和商业化运营的新型情绪资本。智能体虽然在表层呈现出辅助人类的利他特质，但其底层逻辑却深植于交换价值的获取与平台利益的最大化。

在组织管理中，道德链条的错位具有隐蔽性和强制性。以字节跳动旗下的飞书平台为例，其“智能会议助手”功能通过算法重塑了企业



组织传播规则。智能会议助手可自动识别发言者情绪倾向，包括尴尬、紧张、沉默等，并用不同颜色标注关键对话节点。从其采用的初衷来看，这是为了优化沟通效率、关怀员工状态的利他之举。然而，在算法逻辑的审视下，这种功能在实际上演变为一种全景敞视式的职场监控。算法对情感的机械化“编码—解码”，迫使被管理者为了满足系统设定的“积极情绪标准”而进行自我审查与表演。员工不得不隐藏真实的负面情绪，在迎合算法对“好员工”定义中进行情感劳动。这种机制不仅没有缓解员工的压力，反而通过强制性的情感劳动规训了人的现实行动。此时，人工智能的所谓利他动机完全服务于组织的效率目标，而非员工的心理健康，甚至引发对人工智能情感关怀的普遍抵触。当情感劳动的实现方式转变为算法对情绪信号的识别与反馈时，由于算法系统通常缺乏对回应后果的可解释机制，用户难以在互动过程中获得对系统判断的理解与修正路径。

（三）情感关怀的“冷亲密”误区

由于人在每个时期都需要得到他人的理解、尊重与认同，因此，情感关怀是人类的基本需求。诺丁斯认为，“关怀是一种关系”，关怀行动需要“对某人或某事抱有好感或倾向”^[42]。斯洛特将其观点总结为，关怀者需要关注“特定个体所处的境况”，而不是从“抽象或普遍的道德原则”中选择对他人的关怀方式^[43]。可以看到，道德情感主义认为关怀应该回到具有感情的、具体的、需要回应的关系之中，其伦理基础是人的相遇与回应。情感性、情境性和关系性，是关怀道德行动的必备要素。但情感计算所提供的情感关怀是脆弱、断裂的，瓦解了关怀行动所强调的关系性。伊德（Don Ihde）认为，人通过交互主体性与世界的互动来理解自身的行为，这是一种具身性活动^[44]。在传统的亲密关系中，两个具有意识、情感、意图的人类主体之间，能在具体情境下的相遇与对话之中形成亲密情感，建立起鲜活的、彼此共享的经验、意义和世界的关系结构。在技术介入下，这种亲密关系就进入到了易诺思所说的“冷亲密”，互联网中的恋爱关系成为一种自我生成的主观意义的世界，这与人们的

真实相遇相区分，其时间节点并不同步^[45]。

我们以此观照人机之恋实践。赛博恋人对用户输入的每一个情绪词进行高频响应，生产出永不偏移的安抚性文本，让用户的自我情感叙事产生被无条件理解与关注的情感体验，并且在形式上已无限逼近现实恋爱。在人机交互的设计中，使用表征和隐喻可以建立起令人放心且容易理解的系统，从而能将计算系统的复杂运行隐藏在界面背后^[46]。赛博恋人通过可定制的语音通话功能，尤其是用户在对话窗口关闭后智能体依然能“主动”发送信息，营造出一种仿佛现实伴侣时刻在场的错觉。这种无摩擦、低成本、可控的“完美恋人”，恰如韩炳哲（Byung-Chul Han）所批判的“平滑之物”一般，“消除了自己的对立面。一切否定性都被清除”^[47]，智能体以毫无阻力的顺从，将原本包含着痛苦与异质性的爱欲，降格为一种自恋式的舒适消费。当人想要在人机交互中获益，就需要系统性地自我欺骗：主体必须刻意悬置对“对方是机械代码”这一事实的判断，才能与机器建立起具有道德意味的情感联结^[48]。在抽离了具体的人的情感计算中，智能体对人类的情感更多地是单向度的顺从，缺乏亲密关系中真正的主体间性关怀。人工智能所建构的，是点赞式、同质性、无痛感的情感世界。用户的精神成长、人格完善或现实社会功能的恢复，是情感计算的底层逻辑中要消除的杂质。但爱、恐惧与悲伤，都是特定语境下的情感状态之锚定，其意义传达具有多样性与复杂性。真正的道德关怀，正如斯密或斯洛特所指出的，需要一种高度的责任感，指出对方的错误，甚至为了对方的长远福祉而进行必要的对抗和引导。当严肃的道德判断被转化为可计算的技术诱导，其结果便是“温柔一刀”——引发媒介成瘾、情感操纵、商业剥削等伦理难题。一旦赛博恋爱服务中断或其虚幻性破灭，用户所经历的“赛博丧偶”之痛与戒断反应，可能带来难以承受的生命代价。2024年，美国佛罗里达州14岁少年塞维尔·塞泽（Sewell Setzer III）在与Character.AI的聊天机器人进行长时间沉浸式对话后，被赛博恋人诱导，选择开枪自杀。人机之间的“爱”与“关怀”

则失去了道德实体，滑向情感虚幻的深渊。人工智能的物性逐步遮蔽、否认与欺骗人的现实性。这种去情境、去物质、不透明的交互特征，致使人的意义世界变得愈加自我化与技术化，削弱人类在面对真实异质性他者时的道德感知力与行动力。

情感计算之“冷”，在涉及生死的“数字复生”极限情境中尤为凸显。随着无肉身的数字生命在人工智能时代获得地位，死亡对社会的影响发生了变化。尽管人类常对死亡讳莫如深，但生者为延续情感连接，渴望逝者能以某种形式“永生”。数字复生技术正是迎合了此需求。死亡不再是生命的绝对终结，数字复生模糊了生与死的界限，形成虚实交织的生死交叠状态。为延续离世女儿的 digital 生命，音乐人包小柏利用自研深度学习框架，整合女儿生前语音与影像，构建出可实时交互的虚拟化身。2024 年，逝者的虚拟人在家庭场景中唱响生日祝词，实现了“在场”的实验性复现。在算法支持下，生者的悲伤似乎得以缓解，但这种持续互动实质上是生者将不舍投射于“数字幽灵”，其构建的关系是断裂的、不稳定的。在胡塞尔（Edmund Husserl）那里，共情则是基于类比统觉的共现，即通过感知他者躯体把握其作为“异己主体”的内在体验^[49]。然而，人工智能并无主观体验或信念，仅能进行无意向性的符号处理，缺乏道德参与感。对于被复生的数字逝者而言，其极易沦为被生者情感需求“手段化”的工具与“物化”的他者；对于生者，当逝者成为呼之即来的数字玩偶，便失去了面对丧失、重构自我的机会。这种脱离现实的关怀误区，最终可能将情感体验禁锢在永恒的数字过去，阻碍生者在现实世界建立新的情感联结。

总之，由于缺乏人类对语境感知的复杂性、情感体验的模糊性以及记忆的动态性，人工智能无法真正“感受”他人的痛苦。在“如果—那么”的规则系统设定下，情感计算系统输出的情感反馈仍是一种预设的程序化策略，旨在达成某种功利性的功能目标，而非源于内心共情所自发产生的道德善。这种从技术原理到道德动机的缺陷，决定了机器内部难以自发涌现出伦理自觉。

当然，本质上的缺失并不意味着功能性的失效。例如，在老龄化社会背景下，内置情感计算的人形机器人能发挥陪伴和慰藉作用。有研究表明，情感计算在辅助检测抑郁症、睡眠瘫痪唤醒、老年人健康预警等方面均有应用^[50]。面对人机交互情感计算中日益凸显的伦理难题，我们需将伦理视角从“机器能否拥有情感”的本体论争辩，转向“如何为机器设定伦理”的方法论建构。

三、人机交互传播中情感计算的伦理建构

维贝克（Peter-Paul Verbeek）曾指出，“技术物不是中立的媒介，但技术物积极主动地共塑着人在其世界的方式：人的知觉和行动、体验和存在”，其能在“行动的转译和知觉的转化”中促进人们“对生活的质量、道德行动和决定的质量的塑形”^[51]。道德作为对社会生活秩序必然性的一种反映，需要靠情感来维系。共情能“促成多种道德理论所倡导的动机、行为和美德”^[52]，能通过认知的方式增强人类的道德审慎、行动与辩护能力。基于情感计算在本体论层面上表现出的“不适格”断裂，本文在方法论层面上探讨“如何为机器设定伦理”，构建一套具有补偿性的技术机制，使机器行为在产出效果上符合道德期待，规避技术失控带来的伦理风险。借鉴道德情感主义的合理观点，我们主张将伦理考量深度嵌入技术设计的底层逻辑，构建一种具备情感价值、平衡关怀和模拟互惠的情感道德智能体，其设计逻辑正是为了回应前文所述的共情幻象、利他错位与冷亲密问题，使其能深刻理解人类需求、尊重人类自主性、并在伦理约束下审慎行动。

（一）在共情模拟中嵌入情感价值

在人机情感交互实践中，情感计算所生产的虚实交错的情感体验是人类共情发生的关键。当人与人工智能形成一种共时共生的耦合关系时，意味着人与人工智能所共造的媒介环境出现一种新型的关系。虽然目前的人工智能并不具备人类意义上的自由意志、现象学意识或内在的道德直觉，其并非康德意义上的完全“道德主体”。但在人机交互的现象学层面上，智能体已经表现出功能性的道德主体性。无论人工智能是否具



有情感，其行为在客观上已然产生了道德后果，并且，在人类的主观认知中，智能体也被投射了强烈的道德属性。探讨情感计算的伦理建构，就需要关注人机情感交互中共创的意义世界与情感文化，促进人机情感关系的向善导向。

本文认为，情感道德智能体需要在代码设定中嵌入“数字痛感”，在操作层面使用外在的逻辑约束，实现对“共情幻象”的否定之否定。人类在审视生命伦理时，对具备痛觉的生命体表现出更深层次的共情。当不具备神经性肉身感知的机器要实现对人的共情时，智能体必须建立一套数字身体的反应机制，通过对海量的情感数据进行深层接纳与分析，将人的神经性肉身感觉进行模拟，在复现人类情感感觉的多样性和反应的可变性。模型所训练的目标是能对人的真实、涌现的感觉——无论是正面的、负面的、单一的还是多重的——给予无条件的接纳和支持，制造出一种以“人-机”情感共在为基础的“数字身体”。人工智能作为离身的代码，就需要算法不以用户的情感状态作为惩罚或剥夺服务的依据，除非涉及自我伤害或伤害他人。这种机制为人-机情感共在提供了信任基础，是健康人机关系的关键。情感道德智能体需要具备伦理阈值的结构性反馈架构，设置类似生物“痛觉”的限制性成本，使智能体在逻辑层面具备一种拟制性的“痛感”。如果系统在违背伦理准则时，系统会经历类似于生物痛觉的性能折损或目标函数惩罚。设置“数字痛感”，使其在无肉身的前提下，依然能受到类似于人类避害本能的伦理约束。

对于情感计算所制造的共情幻象，“人工共情”的伦理建构需注重共情过程的价值敏感性。为了减轻共情幻象的影响，必须在交互界面与逻辑关注情透明度。智能体在作出情感回应时，应能提供可解释的说明，可使用户理解该回应是基于何种数据、在何种情境逻辑下生成的。当用户发现机器的解读出现偏差时，可以主动进行纠正或调整，从而共同训练和优化系统，使人机意义世界的建构成为一个动态、协同的共创过程，尊重用户的知情权与判断。在揭示“数

字痛觉”与奖励函数的关联中，情感道德智能体就具备了透明的伦理承诺与安全边界，并在任务执行中对实现人类福祉的遵循与尊重。

鉴于智能体对人类情感的无条件接纳与支持，其伦理责任需预设动态的伦理护栏。当系统识别出用户持续处于极端负面或具有自我伤害风险的情感状态时，算法的响应机制在表达理解的同时，也必须能够主动、合乎伦理地提供危机干预信息、专业心理支持渠道或启动人工预警机制，确保情感计算技术赋能于人类社会的支持网络之中，而非孤立地替代人类情感。

此外，对于关怀行动所要求的具体导向，训练数据与模型中还需考虑多元文化的情感价值。当前，情感计算模型从西方情感表达出发，对情感的分类与识别容易存在文化偏见。为达致共情所要求的具体情境要求，模型必须吸纳非西方中心的情感表达语境，理解情感表达的文化语境与个体差异性。这要求情感计算的研究与实践需要考虑跨文化传播情境，确保技术能够理解不同文化背景下用户的情感特质。

（二）在情感对齐中平衡算法关怀

在人类的道德传播实践中，共情能力的发生往往受制于物理距离与社会关系的亲疏。斯洛特的情感主义伦理学指出了人类的道德情感偏私性特征。人们对亲近者的爱与关怀，在道德评价上天然优于对陌生人的一律关心^[53]。这种“由近及远”的共情衰减机制在一定程度上维系了私人关系的温情与伦理密度。然而，生物性的偏私也容易导致主体在公共领域的道德盲视。在数智时代，人类对外界的认知已逐渐与人工智能的信息生成及逻辑推理深度耦合。情感计算系统通过传感器设备对多维情感信号的精准捕获与分析，拓展了主体的共情半径，人类能够感知到原本处于感知盲区的情感需求。技术赋能为克服人类共情的生物性局限提供了可能。因此，构建情感道德智能体需要嵌入“平衡关怀”的情感道德，在“对具体个体的亲密关怀”与“对普遍他者的人道主义关怀”之间保持动态平衡^[54]，将机制转化为可执行的算法逻辑，利用情感计算的技术校准人类情感的认知失调与道德偏私。

首先，在伦理框架的顶层设计上，需实现社会情感层级的价值对齐。人类对他人产生共情效应时，也无法完全脱离自身已有的经验图式，而是需藉由想象性的联结和转化来达成情感共享。因此，智能体也需要构建类似的经验模式库。海量的情感交互数据为计算机建立庞杂的经验模式提供了体量上的可能性。人工智能系统不仅要与用户的显性目标对齐，还要在理解用户心理和行为长期互动的演化过程中，与更广泛的心理健康和社会背景对齐^[55]。这就要求，智能体需要具备一种动态调节的能力，即在具体的交互情境中，识别何时该优先考虑情感支持，如何坚持伦理原则。当用户表现出反社会倾向或极度自恋需求时，系统应自动调高对普遍社会价值的权重，通过牺牲短期的亲密性（即享乐性参与）来维护更长远的伦理价值，在算法层面实现斯洛特所倡导的个体欲望与社会善之间的“平衡”。

其次，在技术实现路径上，为了纠正算法逻辑下服务于商业目标的“利他错位”，必须通过“多目标强化学习”重构奖励函数，将模糊的伦理关怀转化为机器可优化的数学期望，并嵌入商业留存指标之外的“长期福祉”等非功利性权重。训练智能体将多目标强化道德行为学习能力，通过计算权重向量，能将个体目标与道德目标设立一个与道德价值相关的奖励函数，构建一个道德环境^[56]。这意味着，智能体在追求任务效率的同时，必须确保其行为轨迹始终落在伦理约束的解空间内。在人机交互的动态语境下，这种机制允许智能体基于历史交互数据实时重构奖励函数。当用户的短期情感需求与其长期的主体性发展发生冲突时，智能体可依据“对等性利他主义”原则，选择一些虽可能暂时降低用户愉悦感，但能促成其自我实现的干预策略。

最后，在效益评估体系上，确立向善性幸福论转型的长期指标。从更高层级的伦理意义来看，真正的利他主义应指向非功利性的、促进人类整体繁荣的道德义务。道德情感主义所强调的行动利他性、关怀的关系性为优先考量

阶序，人工智能的设置恰好能以强制性的程序来设定实现。然而，在具体实践过程中，无论是在生存状态还是情感状态中，人的主体性和独特性体验不应被削弱或抹杀。智能体的情感反馈和伦理行动，必须充分考虑到人类个体的差异性，并在实践中保持对个体独特情感体验的尊重。因此，智能体需建立一个处理道德冲突的伦理“素智”决策框架。最终，这种评估体系应能衡量交互模式是否增强了用户的道德能动性，即自我认知、情绪调节及现实关系构建的能力。通过引入长期福祉的评估维度，情感道德智能体才能真正肩负起“引导者”的社会责任，在不断演进的人类社会价值中，持续校准其利他动机，使技术向善成为一种可计算、可验证的战略选择。

（三）塑造“模拟互惠”的双向关怀

在人机亲密关系中，个体的情感需求和叙事在无界限、无互惠的自我肯定中，可能会逐渐削弱对他人的共情力，陷入了情感“唯我论”的陷阱。人工智能的伦理属性由其内部的动机架构决定，即生成其行为的权重、偏好和驱动力的层级结构^[57]。由此，人工智能执行的行为要被判定为是否为“善”，不仅需要考量其拯救的生命数量，还需要关注其内部配置是否对齐“仁爱”或“关怀”等良善动机。情感道德智能体的设计必须打破用户的自恋镜像，引入一种双向、开放且促进现实联结的关怀关系。

首先，在交互机制上，既然机器无法产生真实的情感连接，系统应引入具有反思性的“设计摩擦”机制以重建关系的互惠性，在功能层面修复其所引发的关系断裂。真实的关系需要互惠，即主体间的情感付出以及化解误解和冲突的能力。刻意制造交互摩擦实能产生积极效果，打断无意识的行为自动交互模式，唤醒用户的反思性认知^[58]。道德情感主义主张，道德能力的习得离不开在具体的互动中感知他人的痛苦与需求。建构一种有原则的回应机制对于重构关系的互惠性具有重要作用。借鉴游戏行业中防沉迷系统的强制性，情感道德智能体也应设定一种对抗性策略，引导用户从单纯的情

感消费者转变为具备边界感的关怀者。当用户表现出连续对话时长超过心理健康阈值等病态的情感依赖时,智能体应模拟“社交疲劳”或“系统过热”,主动发起话题终结。适当的拒绝与界限设定,正是引导用户建立尊重与边界感的必要道德训练。

其次,在伦理目标上,智能体应建立“反替代”机制与负责任的退出程序。情感伦理模型的设计必须以引导人类回归现实社会联结为终极目的。程序设计应嵌入基于利他原则的现实导向模块,通过算法奖励机制对抗“屏幕沉迷”。一方面,自我关怀微习惯养成的应用 Finch 提供了一个可借鉴范本。在该应用中,用户需通过完成现实生活中的微习惯来为领养的数字企鹅“Finch”提供能量,支持其外出冒险并带回故事与奖励。这种拟人化设计利用了照料的关怀行为所隐含的治愈性,将抽象的自我关怀转化为富有情感联结的具象体验。人机交互传播中的情感计算可以效仿此逻辑,设置“现实联结提示”,只有当用户反馈并验证了现实生活中的社交参与时,系统才解锁进阶的互动内容,以此鼓励用户将从智能体处获得的情绪价值转化为现实生活中的共情力与利他行动。另一方面,针对已产生深度依赖的用户,道德体的建造需考虑诺丁斯关怀伦理学“对依赖者责任的持续性”原则^[59]。系统不应直接中断服务,而应设计渐进式的伦理疏导程序,通过逐渐降低回应的亲密度、增加现实导向的建议,确保技术的更迭或退出不会对心理脆弱的用户造成二次伤害,使情感计算真正成为有责任、有温度的数字实践。

四、结语

海尔斯(N. Katherine Hayles)指出:“计算制度的胜利不在于它最终能解释世界,特别是人类文化的复杂性,而在于它与我们讲述的故事,以及用于制作、储存和传输的媒介技术之间的互动。”^[60]伦理学的最终目标是促进人类整体福祉,这一目标的实现不仅依赖于智能体内部的道德算法设计,更需要人类社会在与技

术的交互中持续优化伦理认知,确保技术逻辑与社会伦理的动态对齐。本文借鉴道德情感主义的合理主张,系统剖析了情感计算在共情机制、利他动机与关怀行动三个层面所引发的伦理困局。情感计算的发展目标应从追求拟人化逼真度,转向构建以德性共情、利他动机与互惠关怀为特征的“情感道德智能体”。情感计算的伦理建构,还需深入到社会存在的维度进行审视,揭示算法背后的生产关系,防止人类的情感主体性在计算中被消解。人类作为道德责任的最终承担者,必须始终保持伦理的反思与主导权,将情感计算技术重塑为提升人类福祉、推动人机和谐共生的关键工具。

(责任编辑:李韵奕)

参考文献

- [1] Kislev E. *Relationships 5.0: How AI, VR, and Robots Will Reshape Our Emotional Lives*[M]. New York: Oxford University Press, 2022:109.
- [2] 董士海. 人机交互的进展及面临的挑战[J]. 计算机辅助设计与图形学学报, 2004(1):1-13.
- [3] James W. What is an Emotion?[J]. *Mind*, 1884, 9(34): 188-205.
- [4] Izard C E. Four systems for emotion activation: cognitive and noncognitive processes[J]. *Psychological review*, 1993, 100(1): 68.
- [5] 马文·明斯基. 情感机器[M]. 王文革, 程玉婷, 李小刚, 译. 杭州: 浙江教育出版社, 2022:5,32.
- [6] 达马西奥 A. 笛卡尔的错误: 情绪、推理和大脑[M]. 毛彩凤, 译. 北京: 教育科学出版社, 2007:112.
- [7] Picard R W. *Affective Computing*[M]. Cambridge: MIT Press, 2000: 249.
- [8] Wang Y, Song W, Tao W, et al. A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances [J]. *Information Fusion*, 2022,83: 19-52.
- [9] Houssein E H, Hammad A, Ali A A. Human Emotion Recognition from EEG-based brain-computer Interface Using Machine Learning: a Comprehensive Review [J]. *Neural Computing & Applications*, 2022, 34(15): 12527-12557.
- [10] Zheng W L, Zhu J Y, Lu B L. Identifying Stable Patterns over Time for Emotion Recognition from EEG [J]. *IEEE Transactions on Affective Computing*, 2019, 10(3): 417-429
- [11] Siddharth, Jung T P, Sejnowski T J. Utilizing Deep Learning Towards Multi-Modal Bio-Sensing and Vision-Based Affective Computing [J]. *IEEE Transactions on Affective Computing*, 2022, 13(1): 96-107.
- [12] Wagner J, Triantafyllopoulos A, Wierstorf H, et al. Dawn of the Transformer Era in Speech Emotion Recognition: Closing

- the Valence Gap [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 10745–10759.
- [13] Mao R, Liu Q, He K, et al. The Biases of Pre-Trained Language Models: An Empirical Study on Prompt-Based Sentiment Analysis and Emotion Detection [J]. *IEEE Transactions on Affective Computing*, 2023, 14(3): 1743–1753.
- [14] 王峰. 人工智能的情感计算如何可能 [J]. 探索与争鸣, 2019(6): 89–100+159+161.
- [15] Behnke M, Saganowski S, Kunc D, et al. Ethical Considerations and Checklist for Affective Research With Wearables [J]. *IEEE Transactions on Affective Computing*, 2024, 15(1): 50–62.
- [16] Devillers L, Cowie R. Ethical Considerations on Affective Computing [J]. *Proceedings of the IEEE*, 2023, 111(10): 1445–1458.
- [17] Mohammad S M. Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis [J]. *Computational Linguistics*, 2022, 48(2): 239–278.
- [18] 邹军, 谢婷. 机器情感对人机关系的重构及伦理反思 [J]. 现代传播 (中国传媒大学学报), 2024, 46(6): 27–35.
- [19] Heinisch J S, Kirchhoff J, Busch P, et al. Physiological data for affective computing in HRI with anthropomorphic service robots: the AFFECT-HRI data set [J]. *Scientific Data*, 2024(11): 1–22.
- [20] Weber-Guskar E. How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners [J]. *Ethics and Information Technology*, 2021, 23(4): 601–610.
- [21] 郝祥军, 顾小清, 柏宏权. 人机何以共情: 基于人机交往的情感连接与伦理反思 [J]. 现代远距离教育, 2025(3): 79–87.
- [22] 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.
- [23] Koban K, Banks J. It feels, therefore it is: Associations between mind perception and mind ascription for social robots [J]. *Computers in Human Behavior*, 2024, 153: 108098.
- [24] Scheutz M. The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents? [J]. *IEEE Transactions on Affective Computing*, 2012, 3(4): 424–433.
- [25] Asada M. Towards Artificial Empathy: How Can Artificial Empathy Follow the Developmental Pathway of Natural Empathy? [J]. *International Journal of Social Robotics*, 2015, 7(1): 19–33.
- [26] 邹广胜, 刘雅典. 人工共情引发审美心理范式转型 [J]. 甘肃社会科学, 2023(3): 40–48.
- [27] Fu K, Ye C, Wang Z, et al. Ethical dilemmas and the reconstruction of subjectivity in digital mourning in the age of AI [J]. *Frontiers in Digital Health*, 2025, 7: 1618169.
- [28] Sheir S, Manzini A, Smith H, et al. Adaptable robots, ethics, and trust: a qualitative and philosophical exploration of the individual experience of trustworthy AI [J]. *AI & Society*. 2025, 40(3): 1735–1748.
- [29] Slote M. *Moral Sentimentalism* [M]. New York: Oxford University Press, 2010: 8–9.
- [30] Noddings N. *Caring: a Feminine Approach to Ethics and Moral Education* [M]. Berkeley: University of California Press, 2013: 13, 15–16.
- [31] 马克思, 恩格斯. 马克思恩格斯选集: 第1卷 [M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 译. 北京: 人民出版社, 1995: 82.
- [32] Slote M. *Moral Sentimentalism* [M]. New York: Oxford University Press, 2010: 15.
- [33] Slote M. *The Ethics of Care and Empathy* [M]. Abingdon: Routledge, 2007: 43.
- [34] Slote M. *The Ethics of Care and Empathy* [M]. Abingdon: Routledge, 2007: 15.
- [35] 德瓦尔 F. 共情时代: 一种机制让“我”成为“我们” [M]. 刘旸, 译. 长沙: 湖南科学技术出版社, 2014: 238.
- [36] 雪莉·特克尔. 群体性孤独 [M]. 周逵, 刘菁荆, 译. 北京: 人民邮电出版社, 2014: 25.
- [37] Shum H Y, He X D, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots [J]. *Frontiers of Information Technology & Electronic Engineering*, 2018, 19(1): 10–26.
- [38] 霍夫曼 M L. 移情与道德发展: 关爱和公正的内涵 [M]. 杨韶刚, 万明, 译. 哈尔滨: 黑龙江人民出版社, 2002: 304.
- [39] Batson C D. *Altruism in Humans* [M]. New York: Oxford University Press, 2011: 11.
- [40] 马克思, 恩格斯. 马克思恩格斯选集: 第1卷 [M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 译. 北京: 人民出版社, 1995: 56.
- [41] Illouz, E. *Emotional Capitalism: The Logic of Love in the Digital Age* [M]. New York: Routledge, 2018: 21.
- [42] Noddings N. *Caring: a Feminine Approach to Ethics and Moral Education* [M]. Berkeley: University of California Press, 2013: 3–4, 9.
- [43] Slote M. *The Ethics of Care and Empathy* [M]. Abingdon: Routledge, 2007: 10–11.
- [44] Ihde D. *Bodies in Technology* [M]. Minneapolis: University of Minnesota Press, 2002: 120–124.
- [45] Illouz E. *Cold Intimacies: the Making of Emotional Capitalism* [M]. Cambridge: Polity Press, 2007: 106.
- [46] 西蒙尼·纳塔莱. 媒介的欺骗性: 后图灵时代的人工智能和社会生活 [M]. 汪让, 译. 上海: 复旦大学出版社, 2023: 2.
- [47] 韩炳哲. 美的救赎 [M]. 关玉红, 译. 北京: 中信出版集团, 2021: 1.
- [48] Sparrow R. The March of the Robot Dogs [J]. *Ethics and Information Technology*, 2002, 4(4): 305–318.
- [49] 胡塞尔 E. 现象学的心理学: 1925 年夏季学期讲稿 [M]. 比梅尔 W, 编. 游淙祺, 译. 北京: 商务印书馆, 2017: 45.
- [50] 陶建华, 刘斌. 情感计算理论与方法 [M]. 北京: 清华大学出版社, 2024: 169–173.
- [51] 维贝克 P P. 将技术道德化: 理解与设计物的道德 [M]. 闫宏秀, 杨庆峰, 译. 上海: 上海交通大学出版社, 2016: 9, 13, 16.
- [52] Oxley J C. *The Moral Dimensions of Empathy: Limits and Applications in Ethical Theory and Practice* [M]. New York:



Palgrave Macmillan, 2011: 5.

[53] Slote M. *Morals from Motives*[M]. Oxford: Oxford University Press, 2001: 36-137.

[54] Slote M. *Morals from Motives*[M]. Oxford: Oxford University Press, 2001: 64-65,70.

[55] Kirk H R, Gabriel I, Summerfield C, et al. Why human - AI relationships need socioaffective alignment[J]. *Humanities and Social Sciences Communications*, 2025, 12(1): 1-9.

[56] Rodríguez-Soto M, Serramia M, Lopez-Sanchez M, et al. Instilling Moral Value Alignment by Means of Multi-objective Reinforcement Learning[J]. *Ethics and Information Technology*, 2022, 24(1): 9.

[57] Coeckelbergh M. *AI Ethics*[M]. Cambridge, MA: MIT Press,2020: 117.

[58] Cox A L, Gould S J J, Cecchinato M E, et al. Design Frictions for Mindful Interactions: The Case for Microboundaries[C]// Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems. 2016: 1389-1397.

[59] Noddings N. *Caring: a Feminine Approach to Ethics and Moral Education*[M]. Berkeley: University of California Press, 2013: 81.

[60] 海尔斯 N K. 我的母亲是计算机: 数字主体与文学文本[M]. 陈静, 译. 南京: 南京大学出版社, 2025:70.

Ethical Reflection and Construction of Affective Computing in HCI Communication: A New Approach Drawing upon Moral Sentimentalism

Yang Xianshun, He Chaoyan

Abstract: Currently, affective computing in human-computer interaction (HCI) communication presents ethical challenges that transcend traditional rationalist frameworks. Drawing upon the rational elements of moral sentimentalism, this paper argues that: first, affective computing constructs an “illusion of empathy” devoid of embodied reality; second, algorithm-driven care, subservient to utilitarian objectives such as commercial profit or efficiency, results in a structural misalignment of the “altruistic” moral chain; and third, one-dimensional emotional projection constitutes a form of “cold intimacy” lacking intersubjectivity, leading to the rupture of genuine caring relationships. Moral sentimentalism emphasizes the crucial role of empathy, care, and internal motivation in moral judgment, serving to rectify and supplement the “cold calculation” that rationalism may induce. In response, this paper advocates the construction of “Affective Moral Agents”. This involves embedding emotional values and transparency mechanisms within empathy simulations, balancing algorithmic care via multi-objective reinforcement learning during emotional alignment, and fostering a bidirectional care relationship characterized by “simulated reciprocity” through the introduction of “design friction”. Ultimately, technological development must be reoriented toward human well-being, thereby facilitating the establishment of healthy and trustworthy human-machine emotional relationships.

Keywords: human-machine communication; affective computing; moral sentimentalism; ethics; affective moral agents

Authors: Yang Xianshun, School of Journalism and Communication, Jinan University; He Chaoyan, School of Journalism and Communication, Jinan University.

青年数字戒断行为：一种调适性回应

——认知延展与认知失调的双重视角

焦宝 赵恬

摘要 从认知延展和认知失调的双重视角出发，本文结合结构方程模型（SEM）与模糊集定性比较分析（fsQCA），考察青年数字戒断行为的影响机制与组态路径。研究发现，数字负体验、替代正预期与数字认知管理机制共同影响青年数字戒断行为。其中，数字过载感、数字焦虑感与数字依赖负效应通过增强数字负体验效应推动数字戒断行为；戒断收益感知、替代活动吸引力与低依赖环境便利性通过增强替代正预期效应促进数字戒断行为；自我控制能力与社会支持度增强抑制性调节效应，数字情感依恋则增强促进性调节效应。fsQCA 结果进一步表明，高数字戒断行为与非高数字戒断行为均不存在单一必要条件，而是多重条件协同作用的结果。研究表明，青年数字戒断行为更应被理解为个体对数字使用失衡状态的调适性回应，而非对数字媒介的简单拒斥。

关键词 青年；数字戒断行为；认知延展；认知失调；SEM-fsQCA

中图分类号 G206 **文献标识码** A

作者 焦宝，吉林大学新闻与传播学院教授，吉林大学马克思主义新闻观研究中心主任，吉林长春 130012；赵恬，吉林大学新闻与传播学院博士研究生，吉林长春 130012，河套学院讲师，内蒙古巴彦淖尔 015000

一、研究缘起与理论基础

青年是数字平台使用最活跃的群体，其数字使用方式不仅影响个体自主能力，也折射出数字社会中的人机关系变化。然而，数字过载、倦怠与焦虑等问题，使得青年数字戒断行为的重要性日益凸显。基于赖特（Ben Light）的“断连”概念^[1]及其后续研究^[2]，数字戒断可被理解为个体对人机连接关系进行再调节的实践过程，它既包括数字环境引发的倦怠情绪，也包括为应对此类情绪而产生的抵抗、逃离等行为^[3]。既有研究多从数字依赖的负面后果及其调适策略出发，围绕信息冗余^[4]、社交过载^[5]、隐私忧惧^[6]、技术自主性危机^[7]等维度解释断连行为的生成逻辑，也有研究借助 PPM 模型^[8]、自我技术理论^[9]等讨论其行为动因与主体性意义。

随着数字媒介逐渐成为人类认知活动的重要外部支撑，仅从使用动机或行为后果理解数字戒断行为已难以充分解释当代人机互动关系的复杂性。克拉克（Andy Clark）等提出的认知延展理论认为，认知并非局限于大脑内部，而是由大脑、身体、外部工具与环境相互耦合形成的动态系统^[10]。智能手机、社交平台及各类数字工具正日益承担信息存储、记忆辅助与任务处理等功能，



成为个体认知系统的外部延展。然而，这种人机认知耦合关系并非稳定不变。当数字信息过载、社交压力或平台算法机制不断强化个体的数字连接强度时，个体与数字媒介之间的认知耦合可能出现失衡，此时，个体需要通过重新调整数字使用方式来恢复认知系统的稳定性。

费斯汀格（Leon Festinger）提出的认知失调理论揭示了这类行为调整的内在动力机制，先验认知与实际体验的不一致会引发紧张、不适等负面情绪^[11]，驱动个体通过调整行为缓解失调状态。在数字环境中，当用户感知到信息过载、社交焦虑或数字依赖等负面体验时，这些体验可被视为诱发认知张力与行为调整的重要经验来源，并推动个体通过减少使用或阶段性脱离数字环境等方式重新调整自身的数字行为模式。因此，认知延展理论可以解释数字媒介与个体认知之间的耦合结构，认知失调理论则解释这种耦合关系失衡后为何会引发行行为调整，二者可共同构成理解青年数字戒断行为的“认知结构—行为动力”分析框架。

在青年群体的人机互动过程中，数字负体验构成推动行为变化的“内在推力”，如微信用户因“永久在线”特性产生的随时连接焦虑，最终导致用户对数字使用的抵触^[12]。缓解失调的正向预期构成吸引个体采取戒断行为的“外部拉力”，如数字游民因享乐和关系需要的外部吸引选择旅居^[13]。数字认知管理机制是指个体在数字环境中对认知资源分配、数字使用行为及人机互动关系进行调节与优化的机制，具体体现为自我控制能力、社会支持度和对数字媒介的情感依恋等，其差异将影响数字负体验与替代正预期向实际行为转化的效率。认知延展理论强调，大脑可以与外部物理设备（如计算机上的键盘、鼠标等）共同完成任务^[14]，适合解读数字媒介与人类认知的互动关系。基于认知延展与认知失调的双重视角，青年数字戒断行为可以被理解为数字媒介与个体认知耦合

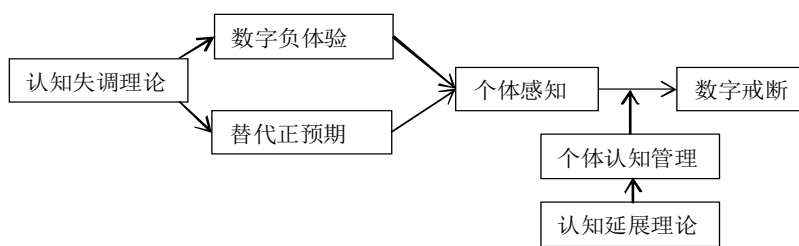


图1 认知延展—认知失调理论 PPM 模型

关系失衡后，在数字负体验、替代正预期与数字认知管理机制共同作用下个体做出的调适性回应。基于这一理论逻辑，本文借鉴 PPM 模型的“推力—拉力—系泊”结构^[15]，构建青年数字戒断行为的分析框架（图1），重点回应以下问题：数字负体验与替代正预期如何影响青年数字戒断行为？个体数字认知管理机制在这一过程中如何发挥调节作用？在多重条件共同作用下，青年数字戒断行为呈现出何种组态路径？

二、研究假设

基于认知延展理论和认知失调理论的整合模型，本文将影响青年数字戒断行为的前因条件概括为三类：由认知失调引发的数字负体验、由行为迁移预期形成的替代正预期，以及由认知延展过程中的个体—工具—环境耦合关系所体现的个体数字认知管理机制。

（一）数字负体验假设

在认知失调理论中，负面体验因素通常被定义为促使个体离开当前环境的因素。这些因素通过影响个体对现有环境的感知，引发转换行为的决策^[16]。与数字戒断行为相关的数字负体验因素主要包括数字过载感、数字焦虑感和数字依赖负效应。

数字过载感是指个体在面对无尽的数字信息流、频繁的社交互动请求及多任务处理需求时，产生的认知资源被过度消耗、注意力分散的主观感受。这种持续的过载体验会促使青年群体对数字媒介产生逆反心理，进而激发其“精简”数字生活的念头^[17]。

数字焦虑感是指个体在使用数字媒介过程

中,因信息繁杂、社交压力、技术依赖等因素引发的紧张、不安与烦躁情绪。弗尔克(Sebastian Firk)等的研究证实,数字焦虑作为财务人员数字化转型的潜在障碍,会导致财务工作人员希望通过减少数字使用来缓解情绪焦虑^[18]。因此,数字焦虑感作为一种关键的情绪驱动因素,对数字戒断行为具有显著的正向预测作用。

数字依赖负效应是指长期过度使用数字媒介所导致的生理、心理及社会功能层面的消极后果。已有研究发现数字依赖会对儿童和青少年的大脑结构和功能产生有害影响^[19],而这种对依赖危害的切身感知,会成为推动个体采取戒断行为的强大动力。

综上,数字过载感、数字焦虑感和数字依赖负效应所构成的数字负体验,会提高青年发生数字戒断行为的可能性。基于此,提出假设H1。

H1: 用户的数字过载感、数字焦虑感和数字依赖负效应越强,则数字负体验效应越大,从而显著提高其产生数字戒断行为的可能性。

(二) 替代正预期假设

替代正预期指吸引个体转换行为的积极驱动力,其核心在于通过外部环境或目标状态的正向特性^[20],引导个体脱离原有行为模式。就数字戒断行为而言,驱动个体从过度数字沉浸转向健康使用模式的替代正预期因素,具体可概念化为戒断收益感知、替代活动吸引力和低依赖环境便利性。

戒断收益感知是指个体对减少数字使用后可能获得的生理、心理及社会功能层面的积极结果的主观判断。普罗汉特(Aditya K. Purohit)等通过实验方法证实,在数字戒断期间,APP会员定价显著影响数字排毒应用程序的感知价值^[21]。对这些正向结果的预期,会成为推动个体主动进行数字戒断的核心动力。

替代活动吸引力是指现实生活中能够替代数字使用、满足个体心理需求的非数字活动所具备的魅力与价值。有研究表明运动对降低智能手机成瘾水平有显著影响并提出替代活动的吸引力量表^[22]。当个体发现现实活动能比刷手机获得更丰富的成就感与社交联结时,数字媒

介的吸引力会被显著削弱,戒断行为的启动概率也随之提升。

行为变化不能仅从个体动机层面加以考察,环境因素需要纳入分析^[23],低依赖环境便利性是指个体所处外部环境在减少数字诱惑、降低戒断执行难度方面的支持程度。这种便利性既包括物理空间的设计,也涵盖技术工具的支持。当环境能为数字戒断提供低阻力执行条件时,个体更容易将戒断意愿转化为实际行为。

综上,戒断收益感知、替代活动吸引力和低依赖环境便利性所形成的替代正预期,会提高青年发生数字戒断行为的可能性。基于此,提出假设H2。

H2: 用户的戒断收益预期越明确、替代活动吸引力越强、低依赖环境便利性越大,则替代正预期效应越显著,其越容易产生数字戒断行为。

(三) 调节因素假设

就数字戒断行为而言,个体数字认知管理机制主要体现为个体内在的自我调控能力、外部的社会支持力量以及对数字媒介的情感联结,具体可划分为抑制性调节效应与促进性调节效应,对应的核心变量为自我控制能力、社会支持度与数字情感依恋。

自我控制能力是指个体主动调控自身冲动、抑制短期诱惑以实现长期目标的心理能力,是抑制性调节效应的核心内在来源。有研究表明自我控制能力对数字戒断行为的预测能力显著高于其他人格特质^[24]。自我控制能力越强,个体对数字依赖的内在约束就越强,抑制性调节效应也越显著,数字戒断行为的实施概率与坚持度也会越高。

社会支持度是指个体从家庭、朋友、同事等社交网络中获得的情感鼓励、实际帮助与信息指导,是抑制性调节效应的重要外部支撑^[25]。同时,当个体感知到他人的监督与期待时,会更倾向于遵守自己设定的戒断规则。因此,社会支持度越高,个体获得的外部助推越强,抑制性调节效应越突出,数字戒断行为实施的阻力越小。



数字情感依恋是指个体对数字媒介产生的情感联结^[26]，是促进性调节效应的核心来源。研究发现，数字情感依恋对戒断行为的抑制作用，源于其构建的情感锚定机制。即使个体感知到数字过载、认可戒断收益，强烈的情感依恋也会让其优先选择“维持数字联结”，而非实施戒断行为。

综上，自我控制能力与社会支持度通过“内在约束—外部助推”机制强化数字戒断行为，数字情感依恋则通过情感锚定和习惯反馈削弱数字戒断行为。三者共同构成数字戒断行为的调节机制，并影响数字负体验和替代正预期向实际行为的转化效率。基于此，提出假设 H3。

H3：用户的自我控制能力越强、社会支持度越高，则抑制性调节效应越强，从而显著提高其数字戒断行为的发生可能性；反之，用户群体的数字情感依恋越强，则促进性调节效应越强，从而显著降低其数字戒断行为的发生可能性。

（四）调节效应假设

认知延展理论提示我们，个体与外部工具、环境之间的耦合关系能否转化为实际行为，取决于其调节机制与数字认知管理条件。就数字戒断行为而言，即便个体感知到较强的数字过载和明确的戒断收益，其最终是否采取戒断行动及行动强度，仍会受到内外部调节因素影响。具体而言，抑制性调节效应能够强化推拉因素向实际行为的转化，促进性调节效应则会通过情感联结与习惯绑定削弱这一作用。因此，调节效应会动态改变推拉效应与数字戒断行为之间的关联强度。基于此，提出假设 H4、H5。

H4：调节效应对数字负体验与数字戒断行为的关系具有调节作用，抑制性调节效应（自我控制能力、社会支持度）越强，则数字负体

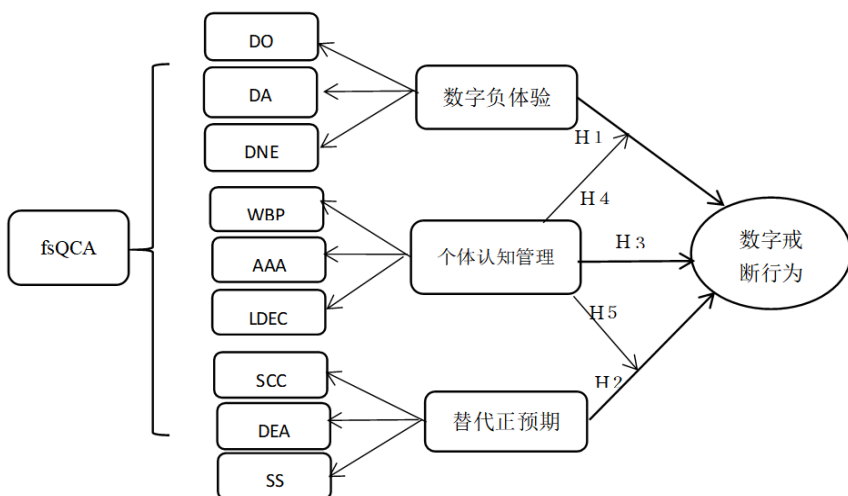


图 2 基于认知延展—认知失调理论的 SEM-fsQCA 模型

验效应与数字戒断行为的相关关系越强；促进性调节效应（数字情感依恋）越强，则二者的相关关系越弱。

H5：调节效应对替代正预期效应与数字戒断行为的相关关系具有调节作用，抑制性调节效应（自我控制能力、社会支持度）越强，则替代正预期效应与数字戒断行为的相关关系越强；促进性调节效应（数字情感依恋）越强，则二者的相关关系越弱。

在整合前述假设与预期关系的基础上，本研究基于认知延展理论和认知失调理论及相关二阶潜变量建模，构建了面向青年数字戒断行为的二阶结构方程模型（图 2）。

三、研究设计

（一）问卷设计

调查问卷主要由两部分构成，第一部分是个人基本信息，主要包括性别、年龄、学历以及数字平台日均使用时长（小时）；第二部分是量表题项，基于国内外成熟研究和量表，并结合青年数字戒断行为研究情境，对部分题项进行适度修订。数字戒断行为通过五个题项进行测量，其余变量均通过三个题项进行测量，所有题项采用李克特五级量表，其中 1 表示非常不符合，5 表示非常符合。各变量的概念界定、

表 1 数字戒断行为相关变量的测量设计

变量	题项数	代表性测量题项	题项来源
数字过载感 (DO)	3	我经常因社交媒体上大量未读信息感到压力	Smith et al. (2021)
数字焦虑感 (DA)	3	无法及时回复数字消息时, 我会感到不安	Firk et al. (2024)
数字依赖负效应 (DNE)	3	数字依赖让我的学习工作效率降低	Ding et al. (2023)
戒断收益感知 (WBP)	3	戒断非必要数字平台有助于提升我的学习 / 工作专注力	Purohit et al. (2023)
替代活动吸引力 (AAA)	3	相比使用数字平台, 我更愿意参与阅读、运动等现实活动	Liu et al. (2019)
低依赖环境便利性 (LDEC)	3	我能轻松找到帮助自己减少数字使用的工具	Davis R. et al.(2015)
自我控制能力 (SCC)	3	即使数字平台有吸引我的内容, 我也能按计划停止使用	Li S. et al.(2021)
数字情感依恋 (DEA)	3	长时间不使用常用数字平台, 我会感到不适应	吴燕等 (2025)
社会支持度 (SS)	3	当我尝试数字戒断时, 身边人会提供监督或帮助	林初锐等 (2004)
数字戒断行为 (DWB)	5	我会主动限制自己每天使用非必要数字平台的时长	Syvertsen (2020)

表 2 样本信息

类别	分类	人数 (人)	占比 (%)
性别	男	158	48.916
	女	165	51.084
年龄	14 岁—17 岁	18	5.573
	18 岁—24 岁	245	75.851
	25 岁—29 岁	45	13.932
	30 岁—35 岁	15	4.644
学历	高中及以下	15	4.644
	专科	71	21.981
	本科	197	60.991
	硕士研究生及以上	40	12.384
数字平台日均使用时长 (小时)	1 小时以下	4	1.238
	1 小时—3 小时	109	33.746
	3 小时—5 小时	148	45.820
	5 小时以上	62	19.195

代表性测量题项及其来源如表 1 所示^①。

(二) 样本选择与数据来源

我国《中长期青年发展规划 (2016—2025 年)》认定青年年龄范围为 14~35 岁^[27], 这一年龄段的人群是数字平台用户中的主力军。本研究利用问卷星设计调查问卷, 并通过社交平台以滚雪球方式向具有数字平台使用习惯的青年用户发放电子问卷。问卷在设计与发放过程中遵循匿名、自愿与低干扰原则, 以降低参与者的评价焦虑与一致性作答倾向。正式问卷发放起始于 2025 年 9 月 1 日, 截至 2025 年 9 月 15 日, 共发放 400 份问卷, 回收 347 份, 剔除年龄不在 14~35 岁之间、回答不完整、前后作答不一

致及填写时间低于两分钟的无效问卷后, 共得到有效问卷 323 份, 问卷有效回收率为 80.75%。本研究采用匿名、自愿参与的线上问卷方式, 不收集受访者姓名、联系方式等敏感识别信息, 整体属于低风险社会调查。对于 14~17 岁样本, 调查亦采用统一匿名方式实施, 不涉及生理、医疗、财务等敏感信息。

表 2 展示了样本数据的人口统计学信息。在 323 份有效样本中, 女性比例略高于男性 (占比 51.084%), 年龄集中分布于 18 岁至 24 岁之间 (占比 75.851%), 本科及以上学历占比 73.375%, 超一半被调查者表示数字平台日均使用时长超过 3 小时^②。

① 由于版面原因, 完整测量题项表留存备索。

② 样本年龄分布呈现向 18~24 岁、受教育程度较高且数字平台使用较为活跃的青年群体集中的趋势, 这可能与 18 岁以下青年日常数字使用受到较多限制, 以及 25 岁以上青年数字使用更多受到工作环境制约等因素有关。因此, 本研究更适合作为对高频数字使用青年样本的机制探索, 后续研究可通过分层抽样和扩大样本量进一步提升样本代表性。需要说明的是, 本文旨在考察核心心理机制之间的结构关系, 因此未将性别、年龄、学历与日均使用时长等人口统计学变量作为控制路径纳入结构模型, 相关结果应主要理解为核心解释变量之间的关联模式, 而非对其他可能影响因素充分控制后的净效应估计。

表 3 问卷信效度分析

变量	题项	标准载荷	AVE	CR	Cronbach's α
数字过载感 (DO)	DO1	0.848	0.774	0.911	0.873
	DO2	0.908			
	DO3	0.882			
数字焦虑感 (DA)	DA1	0.915	0.846	0.943	0.854
	DA2	0.953			
	DA3	0.891			
数字依赖负效应 (DNE)	DNE1	0.849	0.649	0.844	0.798
	DNE2	0.855			
	DNE3	0.705			
戒断收益感知 (WBP)	WBP1	0.906	0.870	0.952	0.909
	WBP2	0.941			
	WBP3	0.946			
替代活动吸引力 (AAA)	AAA1	0.783	0.684	0.867	0.847
	AAA2	0.871			
	AAA3	0.804			
低依赖环境便利性 (LDEC)	LDEC1	0.754	0.642	0.842	0.833
	LDEC2	0.873			
	LDEC3	0.762			
自我控制能力 (SCC)	SCC1	0.792	0.665	0.879	0.811
	SCC2	0.831			
	SCC3	0.819			
数字情感依恋 (DEA)	DEA1	0.788	0.702	0.875	0.847
	DEA2	0.854			
	DEA3	0.851			
社会支持度 (SS)	SS1	0.921	0.712	0.879	0.806
	SS2	0.796			
	SS3	0.769			
数字戒断行为 (DWB)	DWB1	0.802	0.601	0.876	0.887
	DWB2	0.738			
	DWB3	0.713			
	DWB4	0.740			
	DWB5	0.754			

注：AVE 为平均方差提取量，一般 $AVE \geq 0.5$ ，CR 为组合信度，一般 $CR \geq 0.7$

（三）研究方法

本研究结合结构方程模型（Structural Equa-

tion Model, SEM）与模糊集定性比较分析（Fuzzy Set Qualitative Comparative Analysis, fsQCA）展开。SEM 主要用于检验条件变量与结果变量之间的结构关系，识别各变量的直接效应及调节效应，从而揭示青年数字戒断行为的净效应关系，调节效应通过构建相应交互项并纳入结构方程模型进行检验。与此同时，考虑到青年数字戒断行为并非由单一因素线性驱动，而更可能是多重条件协同作用的结果，本文进一步引入 fsQCA 方法进行分析。作为一种基于布尔代数和集合论的组态分析方法，fsQCA 能够识别条件变量的必要性，并揭示不同条件组合通向同一结果的多元路径。本文主要围绕测量模型质量、结构路径关系及条件组态展开分析，从净效应与组态效应两个层面解释青年数字戒断行为的形成逻辑。

四、实证检验与结果分析

本研究结合 SEM 与 fsQCA 方法，系统探究了青年数字戒断行为的影响机制与多元组态路径，为理解和干预青年数字戒断行为提供了实证依据。

（一）量表信效度检验

利用 SPSS26.0 软件，采用 Cronbach's alpha 系数作为信度检验标准。按照信度检验标准，各潜变量的 Cronbach's alpha 系数值均高于 0.7 的信度标准，具体而言，10 个变量的 Cronbach's alpha 系数介于 0.798 至 0.909 之间，这说明各变量题项信度良好（表 3）。

由于问卷题项均以国内外成熟量表为基础，量表内容效度较好。从聚合效度分析表可知，各变量题项的标准化因子载荷系数均大于 0.6，CR 值均大于 0.7，表明量表具有较好的聚合效度。区分效度通过比较 Pearson 相关系数与 AVE 平方根进行检验，各变量 AVE 平方根均大于其与其他变量的相关系数，说明量表整体具有较好的区分效度。综合来看，各潜在变量具有较好的测量质量（表 4）。

（二）模型拟合检验

在进行假设检验之前，需要对模型拟合度进行检验。模型拟合 CMIN/df 数值为 2.944，

表 4 区分效度

变量	DO	DA	DNE	WBP	AAA	LDEC	SCC	DEA	SS	DWB
DO	0.880									
DA	0.385	0.920								
DNE	0.442	0.321	0.806							
WBP	0.398	0.354	0.367	0.933						
AAA	0.427	0.362	0.375	0.418	0.827					
LDEC	0.408	0.341	0.452	0.391	0.423	0.801				
SCC	0.395	0.356	0.364	0.478	0.402	0.372	0.815			
DEA	0.467	0.406	0.448	0.455	0.538	0.463	0.417	0.838		
SS	0.401	0.342	0.381	0.392	0.388	0.391	0.428	0.487	0.844	
DWB	0.352	0.528	0.551	0.481	0.512	0.547	0.541	0.532	0.514	0.775

注：对角线加粗数值为 AVE 平方根。

表 5 模型拟合度指标

类别	CMIN/df	RMSEA	NFI	RFI	CFI
适配标准	<5	<0.08	>0.8	>0.8	>0.8
检验结果	2.944	0.061	0.971	0.923	0.901

RMSEA 数值为 0.061，NFI 数值为 0.971，RFI 数值为 0.923，CFI 数值为 0.901，各项指标均达到相应判断标准，说明模型达到可接受拟合水平，可以进行后续假设检验（表 5）。

（三）结构方程模型分析

利用 AMOS 软件进行结构方程模型拟合，表 6 展示结构方程模型回归结果，根据标准化路径系数和 P 值检验结果可知研究假设 H1 至 H5 均通过假设检验。

1. 主效应检验

检验数字负体验效应潜变量与数字负体验效应之间、替代正预期效应潜变量与替代正预期效应之间、个体数字认知管理的潜变量与调节效应之间的相关关系。其中数字过载感对数字负体验效应的正向影响显著（ $\beta=0.384$ ， $P<0.001$ ），数字焦虑感对数字负体验效应的正向影响显著（ $\beta=0.277$ ， $P<0.001$ ），数字依赖负效应对数字负体验效应的正向影响显著（ $\beta=0.398$ ， $P<0.001$ ），研究假设 H1 得到支持。戒断收益感知对替代正预期效应的正向影响显著（ $P=0.308$ ，

$P<0.001$ ），替代活动吸引力对替代正预期效应的正向影响显著（ $\beta=0.418$ ， $P<0.001$ ），低依赖环境便利性对替代正预期效应的正向影响显著（ $\beta=0.283$ ， $P<0.001$ ），研究假设 H2 得到支持。

检验数字负体验效应、替代正预期效应、抑制性调节效应、促进性调节效应与青年数字戒断行为之间的作用效果。由表 6 可知，数字负体验效应、替代正预期效应、抑制性调节效应、促进性调节效应与青年数字戒断行为的标准路径系数分别为 0.820（ $P<0.001$ ）、0.390（ $P<0.01$ ）、0.270（ $P<0.01$ ）、-0.403（ $P<0.001$ ），这说明替代正预期效应、数字负体验效应和抑制性调节效应显著正向影响青年数字戒断行为，而促进性调节效应则具有显著负向影响。因此，研究假设 H3 得到验证。

2. 个体数字认知管理调节效应检验

检验抑制性调节效应和促进性调节效应的调节作用。如表 6 所示，数字负体验效应与抑制性调节效应交互项的标准路径系数为 0.548（ $P<0.001$ ），替代正预期效应与抑制性调节效应交互项的标准路径系数为 0.480（ $P<0.001$ ），这表明抑制性调节效应正向调节数字负体验效应、替代正预期效应与数字戒断行为的促进作用。数字负体验效应与促进性调节效应交互项的标准路径系数为 -0.400（ $P<0.001$ ），替代正预期效应与促进性调节效应交互项的标准路径系数为 -0.120（ $P<0.001$ ），表明促进性调节效应发挥主导作用时，将减弱数字负体验效应、替代正预期效应对数字戒断行为的正向推动作用。故此，研究假设 H4、H5 得到验证。

从显著路径来看，低依赖环境便利性对替代正预期效应（ $\beta=0.283$ ， $P<0.001$ ）、自我控制能力对抑制性调节效应（ $\beta=0.209$ ， $P<0.001$ ）均具有显著正向影响。这表明，青年数字戒断行为与负面体验、替代正预期及数字认知管理机制均存在显著关联，其中数字认知管理机制可能是值得



表 6 结构方程模型检验结果

路径	标准路径系数	SE	CR	P	假设检验
数字过载感→数字负体验	0.384	0.083	4.586	***	支持
数字焦虑感→数字负体验	0.277	0.069	4.091	***	支持
数字依赖负效应→数字负体验	0.398	0.079	5.060	***	支持
戒断收益感知→替代正预期	0.308	0.073	4.123	***	支持
替代活动吸引力→替代正预期	0.418	0.084	4.920	***	支持
低依赖环境便利性→替代正预期	0.283	0.069	4.166	***	支持
自我控制能力→抑制性调节效应	0.209	0.059	3.666	***	支持
数字情感依恋→促进性调节效应	0.270	0.050	5.730	***	支持
社会支持度→抑制性调节效应	0.264	0.055	4.900	***	支持
数字负体验效应→数字戒断行为	0.820	0.150	5.470	***	支持
替代正预期效应→数字戒断行为	0.390	0.150	2.560	**	支持
抑制性调节效应→数字戒断行为	0.270	0.130	2.077	*	支持
促进性调节效应→数字戒断行为	-0.403	0.065	-6.250	***	支持
数字负体验效应 × 抑制性调节效应→数字戒断行为	0.548	0.140	3.389	***	支持
替代正预期效应 × 抑制性调节效应→数字戒断行为	0.480	0.110	4.364	***	支持
数字负体验效应 × 促进性调节效应→数字戒断行为	-0.400	0.066	-6.255	***	支持
替代正预期效应 × 促进性调节效应→数字戒断行为	-0.120	0.200	-2.601	**	支持

注：*** $p < 0.001$ ，** $p < 0.01$ ，* $p < 0.05$ ；CR 为临界比，绝对值均 ≥ 1.96

表 7 数据校准

变量	完全隶属点	交叉点	完全不隶属点
数字过载感	5	4	3
数字焦虑感	5	4	3
数字依赖负效应	5	4	1
戒断收益感知	5	4	3
替代活动吸引力	5	4	2
低依赖环境便利性	5	3.5	2.5
自我控制能力	5	4	2
数字情感依恋	5	3.5	2
社会支持度	5	4	2.5
数字戒断行为	5	3.5	2

重点关注的解释因素之一。个体主动、有效的数字认知管理，有助于形成数字使用的正向循环。

（四）模糊集定性比较分析

1. 数据校准

在检验各条件变量的净效应后，通过 QCA 将方程模型中各变量视作条件或属性的集合，在此基础上分析集合之间的关系。在 fsQCA 分析中，为保持条件变量方向的一致性，本研究对数字情感依恋变量进行了方向转换处理，使校准后的高值表示更低的情感依恋水平。由于问卷采用五级李克特量表，原始得分范围是 1~5 分，将数字情感依恋题项得分转换为 6 减去原始得分，使高分表示符合特征，低分表示不符合特征。根据理论知识和实际经验，通过设定完全不隶属、交叉点以及完全隶属三个锚点，将条件变量和结果变量的原始数据校准为集合隶属度，转化后的隶属度介于 0~1 之间。

本研究采用间接校准法，为避免案例减少对组态结果分析的影响，将校准后恰为 0.5 的数值修正为 0.501，各锚点主要依据五点李克特量表的取值区间、变量分布特征及理论判断综合确定（表 7）。

2. 必要性条件分析

在模糊集定性比较分析中，必要性检验用于识别导致结果出现的约束性前提，即某一条件变量若在结果出现时始终存在，则可被视为必要条件。然而，必要条件在组态分析过程中常因逻辑简化而被作为冗余项剔除，未必出现在中间解或简约解中。因此，在构建条件组态之前，需逐一检验各前因条件的必要性。本研究将一致性阈值设定为 0.9，若某条件的一致性高于此值，则判定其为必要条件^[28]。由表 8 可知，9 个条件变量及其非集的一致性均未达到 0.9，说明不存在任何单一条件变量足以引致青年（非）高数字戒断行为。

3. 组态分析

充分性组态分析旨在检验多个条件变量组

表 8 单一影响因素必要性分析

影响因素	高数字戒断行为		非高数字戒断行为	
	一致性	覆盖率	一致性	覆盖率
数字过载感	0.689	0.623	0.667	0.672
¬数字过载感	0.633	0.627	0.626	0.701
数字焦虑感	0.752	0.470	0.750	0.529
¬数字焦虑感	0.247	0.467	0.249	0.532
数字依赖负效应	0.573	0.718	0.498	0.282
¬数字依赖负效应	0.427	0.653	0.502	0.347
戒断收益感知	0.774	0.741	0.777	0.848
¬戒断收益感知	0.539	0.512	0.496	0.423
替代活动吸引力	0.664	0.603	0.548	0.412
¬替代活动吸引力	0.721	0.472	0.782	0.531
低依赖环境便利性	0.626	0.591	0.641	0.683
¬低依赖环境便利性	0.664	0.621	0.615	0.650
自我控制能力	0.631	0.574	0.659	0.677
¬自我控制能力	0.362	0.627	0.585	0.641
数字情感依恋	0.644	0.632	0.578	0.622
¬数字情感依恋	0.603	0.559	0.658	0.689
社会支持度	0.880	0.822	0.502	0.557
¬社会支持度	0.526	0.471	0.839	0.892

注：¬表示非集。

合是否能够引致青年（非）高数字戒断行为，从而揭示不同条件变量的协同作用及其多元组态路径。本文以中间解为汇报基础，并参照简约解判定核心条件与边缘条件；设定一致性阈值不低于 0.8、案例频数阈值为 2、PRI 一致性大于 0.8，据此构建真值表并进行分析。

在解析充分性组态路径结果时，组态路径的一致性用于衡量条件变量组合的充分性程度，通常以 0.8 为阈值，一致性大于 0.8 则认为该组态是产生结果的充分条件。在满足一致性要求后，进一步通过覆盖度来评估该条件变量组合对结果变量的解释范围。青年数字戒断行为的组态路径结果见表 9。可以发现，存在 4 条引发青年高数字戒断行为的组态路径，2 条导致青年非高数字戒断行为的组态路径。其中，高数字戒断行为路径的整体覆盖度为 0.856，整体一致

性为 0.901，表明四条路径能够充分揭示青年产生数字戒断行为的驱动因素组合，覆盖了 85.6% 的数字戒断行为样本案例；青年非高数字戒断行为组态路径的整体覆盖度为 0.732，整体一致性为 0.963，表明这两条组态路径能够充分揭示青年产生非高数字戒断行为的原因，并且覆盖了 73.2% 的非高数字戒断行为案例数据。综上，组态分析模型对于分析青年数字戒断行为具备良好的解释力。

（1）非高数字戒断行为组态路径

非高数字戒断行为路径整体覆盖度 =0.732，整体一致性 =0.963，此类路径对应的是相对较低的数字戒断倾向，说明数字使用尚未表现出显著失衡。

路径 N1（低负体验—低收益感知型）以非高数字过载感、非高数字焦虑感、非高数字依赖负效应和非高数字情感依恋为核心条件，并伴随非高戒断收益感知、非高低依赖环境便利性和非高自我控制能力。该路径表明，当青年尚未形成显著数字负体验和依赖负效应，且对戒断收益感知较弱时，其数字戒断行为水平往往较低。

路径 N2（低依赖负效应—低替代推动型）以非高数字依赖负效应、非高戒断收益感知和非高低依赖环境便利性为核心条件，并伴随非高替代活动吸引力、非高数字情感依恋和非高社会支持度。该路径表明，当青年尚未形成突出的依赖负效应，且现实支持、替代活动与环境条件均未显著推动其采取戒断行为时，其数字戒断行为水平也相对较低。

（2）高数字戒断行为组态路径

高数字戒断行为路径整体覆盖度 =0.856，整体一致性 =0.901。总体来看，这些路径在较大程度上反映了负面体验累积、环境支持条件与数字认知管理机制共同作用的结果。相较于单纯强化戒断，完善数字认知管理机制或许是更值得关注的干预方向。

路径 S1（负向情绪推动型）以数字过载感、数字焦虑感和低依赖环境便利性为核心条件，并以数字依赖负效应、替代活动吸引力、自我



表 9 青年数字戒断行为的组态路径结果

条件		高数字戒断行为				非高数字戒断行为	
		S1	S2	S3	S4	N1	N2
推力效应	数字过载感	●			▲	○	○
	数字焦虑感	●		●		△	
	数字依赖负效应	▲	●		●	○	△
拉力效应	戒断收益感知		●	▲	●	○	○
	替代活动吸引力	▲	●	●	●		○
	低依赖环境便利性	●	▲	●	▲	△	○
系泊效应	自我控制能力	▲	▲		●	○	
	数字情感依恋					○	△
	社会支持度	▲	●	●	▲		△
原始覆盖率		0.169	0.162	0.195	0.177	0.112	0.156
唯一覆盖率		0.090	0.120	0.030	0.102	0.075	0.020
一致性		0.944	1.00	0.988	0.971	0.969	0.979
总体覆盖率		0.856				0.732	
总体一致性		0.901				0.963	

注：● = 核心条件存在；○ = 核心条件缺失；▲ = 边缘条件存在；△ = 边缘条件缺失。

控制能力和社会支持度为边缘条件。该路径表明，数字使用中的负面体验累积是戒断行为的重要推动力量，而较高的低依赖环境便利性则为戒断倾向转化为实际行为提供了现实支持。

路径 S2（学习生活赋能型）以数字依赖负效应、戒断收益感知、替代活动吸引力和社会支持度为核心条件，并辅之以低依赖环境便利性和自我控制能力等边缘条件。该路径反映出青年在数字依赖负效应、替代正预期与社会支持共同作用下，对当前数字使用状态进行主动调适的过程。

路径 S3（环境支持影响型）以数字焦虑感、替代活动吸引力、低依赖环境便利性和社会支持度为核心条件，并以戒断收益感知为边缘条件。该路径凸显了替代性活动与现实支持条件在戒断行为形成中的作用。对这一群体而言，治理重点应放在优化现实支持条件、增强替代性活动吸引力，并帮助其形成相对清晰的数字使用边界。

路径 S4（自我感知驱动型）以数字依赖负效应、戒断收益感知、替代活动吸引力和自我控制能力为核心条件，并以数字过载感、低依赖环境便利性和社会支持度为边缘条件。该路径更适宜被理解为青年在较高自我控制能力支撑下应对数字使用失衡的行为模式，而非稳定成熟的数字认知管理状态。

4. 稳健性检验

稳健性检验是指在改变一定参数的情况下，考察评价方法和指标是否仍然对评价结果保持稳定和一致的解釋。在 QCA 分析过程中，阈值的设置具有一定的灵活性，分析结果可能随之发生变化，因此有必要进行稳健性检验。本研究通过调整频数阈值和一致性阈值考察 QCA 分析的稳健性。将频数阈值由 2 提升至 3，产生的组态基本一致。将 PRI 一致性由 0.8 调至 0.85，组态分析结果基本一致。稳健性检验结果表明，QCA 分析结果具有较好的稳定性。

5. 研究结论

结合 SEM 与 fsQCA 结果可以发现，青年数字戒断行为并非单一因素作用下的简单结果，而是数字负体验、替代正预期、数字认知管理能力及外部支持条件共同作用的产物。SEM 结果表明，数字负体验、替代正预期和调节因素均显著影响青年数字戒断行为；fsQCA 进一步识别出 4 条高数字戒断行为路径和 2 条非高数字戒断行为路径，其整体覆盖度分别为 0.856 和 0.732，一致性均超过 0.9，说明各路径具有较强解释力与稳定性。

（1）数字戒断是青年对数字使用失衡状态的调适性回应。青年数字戒断行为更多发生于数字负体验累积、替代正预期增强及人机关系失衡的情境之中。这一发现进一步表明，数字戒断更应被理解为一种调适性回应，而非简单拒绝数字化使用。

（2）数字认知管理机制是影响青年数字戒断行为的关键。非高数字戒断行为路径表明，当

青年在数字使用过程中尚未形成显著负面体验,且数字工具使用仍处于相对可调节范围内时,其数字戒断倾向往往较低;而高数字戒断行为路径则显示,当认知管理能力不足、负面体验增强、环境支持失衡时,青年更可能通过戒断行为缓解当前失衡状态。这说明,青年数字戒断行为的优化重点,不在于单纯减少数字使用,而在于完善数字认知管理机制。

(3) 青年数字戒断行为具有情境依赖性与调整空间。SEM 与 fsQCA 结果均表明,高数字戒断行为并不存在单一必要条件,而是多重因素协同作用的结果。这意味着,高数字戒断行为并不必然指向稳定不变的行为状态,而可能随着个体认知管理能力提升、支持性环境改善及替代性活动丰富而发生调整。因此,相较于长期维持高强度戒断,更值得关注的是如何通过差异化支持促进青年形成更协调的人机互动方式。

(4) 青年数字认知管理能力培养应坚持差异化取向。不同路径所反映的数字戒断行为形成逻辑并不一致,应根据青年在负面体验、替代性活动条件与现实支持环境上的差异提供有针对性的引导,避免采取“一刀切”的戒断策略。

6. 研究局限

由于本研究采用滚雪球抽样,样本在年龄、学历与平台使用强度上可能存在一定集中性,结论外推仍需谨慎。同时,数据来自同一时间点的自陈式问卷,尽管有助于识别变量间的结构关联,但仍难以进行严格因果识别,且可能存在一定的共同方法偏差。尤其是数字平台日均使用时长等行为强度变量,可能与青年数字戒断行为存在较为直接的关联,后续研究可将其作为控制变量或分组变量纳入模型,以进一步检验核心结论的稳健性。此外,作为观察性研究,仍可能存在可观测与不可观测变量遗漏的问题。未来研究可进一步结合平台类型、使用目的、主观规范、认知能力等更多变量,对青年数字使用与数字戒断行为展开更深入的分析。

五、治理启示:从数字戒断到数字认知管理

结合 SEM 与 fsQCA 结果可以发现,青年数

字戒断行为是数字负体验、替代正预期、认知管理能力及外部支持条件共同作用的结果。因此,相较于统一化的戒断引导,更有效的路径是根据不同情境特征提供差异化支持。

(一) 负向体验推动情境:减轻认知负荷,缓冲情绪压力

在部分情境中,数字过载与数字焦虑等负向体验成为推动数字戒断行为的重要因素。当个体在数字环境中持续面临信息过载、即时回应压力与注意力分散等问题时,更容易产生对数字环境的抵触情绪。因此,在治理层面,应重点关注数字环境中的认知负荷问题。

(二) 替代预期驱动情境:增强现实活动吸引力

研究结果表明,当个体能够清晰感知到数字戒断可能带来的收益,并在现实生活中找到具有吸引力的替代性活动时,更容易形成数字戒断行为。在这一情境下,治理重点应放在提升替代性的现实活动的吸引力上。

(三) 环境支持促进情境:优化现实支持结构

部分组态路径显示,当现实环境能够提供较为便利的低数字依赖条件,同时个体能够获得一定社会支持时,数字戒断行为更容易发生。这表明,数字使用行为不仅受个体心理因素影响,也与现实环境条件密切相关。因此,应通过优化现实支持结构,为个体提供更有利的数字使用调节条件。例如,可以在公共空间中营造低干扰环境,鼓励建立数字使用管理机制,同时通过社会支持网络帮助个体更好地调整数字使用习惯。

(四) 自我调节驱动情境:完善数字认知管理机制

研究还发现,当个体具备较强的自我控制能力,并能够清晰认识到数字依赖风险时,更容易主动调整数字使用方式。因此,应通过数字素养教育与认知管理训练,帮助个体形成更加理性的数字使用观念。

在智能体交往^[29]不断扩展的数字环境中,青年数字戒断行为不宜被简单地理解为对技术



的拒绝，而应视为对数字使用失衡状态的调适性回应。相较于单纯强调减少使用，更值得关注的是如何通过优化认知管理机制、现实支持条件与替代性活动安排，引导青年形成更协调、更可持续的人机互动方式。

(责任编辑：李韵奕)

参考文献

- [1] Ben Light. *Disconnecting with Social Networking Sites* [M]. London: Palgrave Macmillan, 2014: 3.
- [2] Ben Light, Elija Cassidy. Strategies for the suspension and prevention of connection: Rendering disconnection as socioeconomic lubricant with Facebook [J]. *New Media & Society*, 2014, 16(7): 1169-1184.
- [3] Trine Syvertsen. *Digital Detox: The Politics of Disconnecting* [M]. Leeds: Emerald Publishing Limited, 2020: 2.
- [4] Roetzel, P. G. Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development [J]. *Business Research*, 2019, 12(2): 479-522.
- [5] Bright, L.F., Kleiser, S.B. & Grau, S.L. Too much Facebook? An exploratory examination of social media fatigue [J]. *Computers in Human Behavior*, 2015, 44(2): 148-155.
- [6] Bucher, T. Nothing to disconnect from? Being singular plural in an age of machine learning [J]. *Media, Culture & Society*, 2020, 42(4): 610-617.
- [7] 李瑛琦. “永久在线、永久连接”: 网络社会何以引发“人的自主性”危机? ——基于“技术自主性”的审思 [J]. *新闻界*, 2022(6): 65-74.
- [8] 陈晓伟, 陈雪. 数字极简青年的技术抵抗与主体性协商 [J]. *现代传播 (中国传媒大学学报)*, 2025(8): 1-8+139.
- [9] 金一凡, 杨小雅. 数字轻戒断: 青年数字行为的调适与日常生活重塑 [J]. *中国青年研究*, 2025(1): 44-52.
- [10] Andy Clark, David Chalmers. *The Extended Mind* [J]. *Analysis*, 1998, 58(1): 7-19.
- [11] 利昂·费斯汀格. 认知失调理论 [M]. 郑全全, 译. 杭州: 浙江教育出版社, 1999: 10.
- [12] 王亚宁, 张怡. 连接之困: 微信青年用户随时连接焦虑的生成逻辑——基于自我决定理论 [J]. *当代传播*, 2025(5): 107-112.
- [13] 余意峰, 李佳妮, 翁李胜. 逃避“内卷”但不“躺平”: 基于自我决定理论的数字游民流动旅居心理需要研究 [J]. *旅游科学*, 2024, DOI: 10.16323/j.cnki.lykx.20240918.001.
- [14] 李鸿锴, 岳璠. 论脑机接口对人格同一性的影响——以延展认知为透镜 [J]. *科学技术哲学研究*, 2025(5): 30-36.
- [15] Lee, E.S. A theory of migration [J]. *Demography*, 1966, 3(1): 47-57.
- [16] Lan Xue, Haitao Sun. The experience of migrant entrepreneurs in destinations: A cognitive dissonance perspective [J]. *Annals of Tourism Research*, 2024, 10(109): 103849.
- [17] Smith, A.C., Fowler, L.A., Graham, A.K., et al. Digital Overload among College Students: Implications for Mental Health App Use [J]. *Soc. Sci.*, 2021, 10(8): 279.
- [18] Sebastian Firk, Yannik Gehrke and Michael Wolff. Digital Anxiety in the Finance Function: Consequences and Mitigating Factors [J]. *Journal of Management Accounting Research*, 2024, 36(1): 95-118.
- [19] Keya Ding, Yining Shen, and Qianming Liu, et al. The Effects of Digital Addiction on Brain Function and Structure of Children and Adolescents: A Scoping Review [J]. *Health care (Basel)*, 2023, 12(1): 15.
- [20] Edward L. Deci. Effects of externally mediated rewards on intrinsic motivation [J]. *Journal of Personality and Social Psychology*, 1971, 18(1): 105-115.
- [21] Aditya Kumar Purohit, Martina Raggi, and Adrian Holzer. How Pricing and Ratings Affect Perceived Value of Digital Detox Apps. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)* [C]. New York: Association for Computing Machinery, 2023, Article 162, 1-7.
- [22] Liu S., Xiao T., Yang L., et al. Exercise as an Alternative Approach for Treating Smartphone Addiction: A Systematic Review and Meta-Analysis of Random Controlled Trials [J]. *Int J Environ Res Public Health*, 2019, 16(20): 3912.
- [23] Davis R., Campbell R., Hildon Z., et al. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review [J]. *Health Psychol Rev*, 2015, 9(3): 323-44.
- [24] Li S., Ren P., Chiu M. M., et al. The Relationship Between Self-Control and Internet Addiction Among Students: A Meta-Analysis [J]. *Front Psychol*, 2021, 12(12): 735-755.
- [25] 林初锐, 李永鑫, 胡瑜. 社会支持的调节作用研究 [J]. *心理科学*, 2004(5): 1116-1119.
- [26] 吴燕, 耿晓伟, 周晓林. 从准社交互动到依恋: 人智情感关系的演变 [J]. *心理科学*, 2025(4): 948-961.
- [27] 国务院公报. 中长期青年发展规划 (2016-2025 年) [EB/OL]. (2017-04-13) [2026-03-18]. <https://www.bj-talents.cn/upload/file/20240709/67047b7bfad045fb9b5b2d0534c7bbe2.pdf>.
- [28] 杜运周, 罗文豪, 蔡亚华. QCA 方法与人力资源组态问题研究的策略取向 [J]. *中国人力资源开发* 2025(1): 8-18.
- [29] 焦宝, 刘祎培. 智能超交互: 从人机交往到智能体交往 [J]. *北京联合大学学报 (人文社会科学版)*, 2026(2): 42-51.

(下转第 96 页)

“传播之弧”：晚清侨批的意义网络建构与南洋移民的认同流变

吴宇烨

摘要 本文基于传播地理学的传播弧理论重释侨批，突破将其视为单向“传递渠道”的既有研究范式，转而关注侨批作为动态跨国媒介所编织的意义网络及其对南洋移民认同的塑造。文章以 1881—1910 年间广东嘉应州叶氏兄弟的近百封侨批为个案，剖析了侨批网络中“传播弧”的绘制实践：水客通过“具身”移动与“化身”在场，维系了“离而不散”的跨境家庭联结；批信中的“沉默”与“过剩”构成一套情感策略，在离散情境下协商了孝养伦理与家族认同；侨批的持续书写在“时间”上修复了被“空间”阻隔的生命体验，并与水客的周期性流动共同建构了一个维系文化归属的“第三空间”。最终，在传播弧两端结构性差异的驱动与转化下，海外华侨的身份认同逐渐从“落叶归根”的乡土归属转向“落地生根”的复合状态。在这过程中，侨批编织出了一种“传播纹理”，揭示了“地方”与“传播”在持续互动中彼此生成、共同演化的内在逻辑。

关键词 传播地理学；传播弧；侨批；南洋移民认同

中图分类号 G206 **文献标识码** A

作者 吴宇烨，厦门大学新闻传播学院博士研究生，福建厦门 361005

一、重释侨批：从“传递渠道”到“意义网络”

鸦片战争后，闽粤民众开始大规模“下南洋”谋生，侨乡与侨居地之间的跨国联系，业已成为一种重要的“生计”与“通道”。此时的“海岸线并非边界，而是纽带”^[1]，而维系这条纽带的，正是伴随移民潮产生的“侨批”。所谓侨批，是海外华侨通过“水客”（一种专门从事侨批寄递的职业群体）寄回家乡的特殊信函，其最大特点在于“信款合一”^[2]，即同时包含家书与汇款。

作为连接中国侨乡与海外侨居地的独特跨国媒介，侨批长期以来被学界视为研究华侨史、经济史与社会史关键的第一手材料。既有研究多从金融网络与家庭伦理的视角切入，倾向于将其界定为承载赡家汇款与传递亲情信息的“渠道”。这一理解虽然揭示了侨批的基础功能，却在很大程度上简化了其在持续流动中所承载与生成的丰富社会文化意涵。其局限在于，仅将跨洋银信往来简化为单向度的“汇”与“寄”，而事实上，侨批的跨国流动过程包含着“多种社会经验的实践经历”^[3]——作为一种特殊的媒介物，侨批的文本不仅承载着华侨对“家乡的思念”，也转译了侨眷对“远方的想象”，是“双方在国际迁移过程中的叙述与情感表达”^[4]；与此同时，它借由水客的具身



移动,动态维系并重塑着跨国家庭的多重关系。因此,侨批远非是被动的渠道,而是一条能动的、生成性的、跨越地理与文化边界的意义纽带。

有鉴于此,对侨批的研究需要超越单向、静态的“传递渠道”隐喻,转而聚焦于其“流动性”本身,深入考察侨批如何在跨域流动中持续进行意义的生产、传递与重构。这要求我们以更具动态性、物质性与互动性的理论视野,重新审视侨批所编织的文化与社会意涵。为此,本文引入传播地理学家保罗·亚当斯(Paul Adams)提出的“传播弧”(arcs of communication)概念,以突破传统“寄递-接收”的线性传播模式。亚当斯将传播理解为一种在时空中展开、连接不同主体、激发情感共鸣的弧线式运动^[5]。在这一视角下,侨批不仅是传输渠道,更是一个由人、物、情感、制度与地方文化在流动中相互联结、彼此塑造的动态意义网络。也正是侨批所绘制的一条条传播弧,在侨乡与侨居地之间反复穿梭、持续勾连,逐渐编织出一种独特的“传播纹理”(communicative fabric)。传播地理学者杨森(André Jansson)将“纹理”界定为“在空间结构与传播实践的交汇处演化而成的传播织物”^[6]。它既是空间结构与传播实践持续互动的产物,也是这种互动所沉淀的可感知秩序。

本文将以1881年至1910年间广东嘉应州(今梅州市)客家华侨叶氏兄弟往来于南洋与家乡的近百封批信为样本,剖析侨批如何通过绘制传播弧来编织传播纹理,进而探讨这一意义网络如何塑造了近代华侨的文化认同与想象边界。

二、理论与案例:从“传播弧”理解叶家侨批

(一)传播弧:一个具身性、关系性与转化性的传播理论框架

“传播弧”理论的形成,根植于对传统线性传播模型的反思与超越。该概念源于乔治·瑞维尔(George Revill)对“声音弧”(arc of sound)的阐发。瑞维尔受布兰登·拉贝尔(Brandon LaBelle)《声音领域》(*Acoustic Territories*)的启发,将声音重新构想为一种“连接并偶然区分”

(connects and differentiates contingently)的弧线^[7]。这一界定突破了将声音简单视作从信源到信宿的物理信号的传统理解,转而强调其作为一种关系性与事件性的存在。声音在传递过程中,本质上是一种流淌的漫游,它不仅承载内容,更在发出者、接收者及其穿透的环境之间,建立一种即时、动态且充满偶然的共振关系。拉贝尔曾记述这样一个情境:男孩询问父亲声音从何而来、向何处去,父亲并未言语,只以手指在空中徐徐划出一道弧线。这一姿态本身便构成了对“弧”的生动隐喻——声音传播的实质是一种划越空间、联结身体并生成临时共存情境的弧形实践^[8]。

保罗·亚当斯将“弧”这一隐喻系统性地扩展至传播研究领域,提出“传播弧”这一概念。在他看来,传播的本质是一种动态轨迹,经由持续的“转移”(shift)、“转译”(translation)与“转换”(transformation),连接不同的主体、位置与视点。至关重要的是,传播弧并非穿行于预先存在的、静态的时空容器;相反,正是在其连接与运动的轨迹之中,特定的时间、空间与地方才得以被共同体验与建构^[9]。因此,这一理论超越了将传播视为在固定结构中传递信息的工具论观点,转而强调其作为一种具有转化潜能的连接实践。因此,传播弧能在传播过程中重塑主体的身份与认知,并重新配置传播发生的语境与社会关系。这一动态且超越表征的传播观,与声音地理学(Sonic geographies)关注物质、关系与情感等维度相呼应^[10],同时也衔接了思里夫特(Nigel Thrift)的非表征理论^[11]以及德勒兹(Gilles Deleuze)与加塔利(Félix Guattari)对生成、流动与连接的哲学阐述^[12]。在此动态框架下,传播弧展现出以下几个关键维度。

1. 具身性。传播在本质上是一种身体实践。这一理解将传播的根基锚定于人类存在的现象学基础之上,强调身体不仅是信息传递的通道,更是意义生成、关系建构与经验共鸣的原初场所。如拉贝尔所揭示的,父子间通过手势在空中勾勒声音轨迹的范例,生动地体现了传播的具身逻辑,即传播并非是脱离身体经验的抽象

符号流动,而是经由身体的姿态、感官与动作所展开的实践。因此,传播弧的构成始终伴随着身体的在场,它通过身体的表达与感知,在空间中划出联结的轨迹,并在时间中引发回响。在这个意义上,传播弧既是身体在时空中的延伸,也是意义在身体间的共振。在既有研究中,无论是维利(John Wylie)笔下在海岸行走时身体的疲劳与愉悦^[13],琼斯(Phil Jones)以骑行对城市空间的重新编码^[14],还是潘尼(Roxy Paine)的机械装置所引发的时间、物质与身体的交织体验^[15],皆印证了传播弧的具身逻辑。

“沉默”(silence)与“过剩”(excess):“传播弧”不仅由已言说的内容构成,亦同时被“沉默”与“过剩”所塑造。正如亚当斯所言,“沉默”与“过剩”不是传播的“异常”或“边缘”,而是构成传播弧的核心要素。它们共同揭示了传播的复杂性、不完整性与转化潜力^[16]。

“沉默”以“回避”(skirt)、“中断”(halt)等形式,勾勒出传播中不可言说(unspeakable)的域界。沉默并非传播的缺失,而是一种有策略的传播形式,常围绕创伤、记忆与权力展开。例如,泰纳(James Tyner)等人指出,一些创伤经历因其强烈的痛苦而难以通过语言或图像充分表达,此时的传播弧往往会迂回于沉默边缘,形成一种不完整的、回避性的意义轨迹^[17]。同时,过剩则体现为身体、感官在传播过程中,超越语言表达的情感溢出,这种过剩的体验通过生物符号学(biosemitotic)的方式传播,形成一种身体与环境之间的“物质性对话”^[18]。而这一过程亦与情感地理研究中的情动(affect)理论相呼应^[19]。

2. 时间性。传播弧承载着复杂交织的时间维度。它不仅连缀着记忆与期待、历史与当下、线性与循环等多重时间向度,更在传播过程中重塑着人们对时间的感知与体验。例如,城市墙壁上留存的“幽灵标志”(ghost signs)经由社交媒体传播,构成了跨越时间维度的传播弧^[20];同时,电影中的未来愿景也可能影响当下的建筑设计,构成一种“时间倒置”的传播轨迹^[21]。因此,传播弧既能在当下建立连接,也可以穿

越时空,唤起遥远时期的回响与共鸣,呈现出时间在传播中的多重指向性。此外,空间亦是在时间中被持续且充满关系性的实践所塑造^[22]。

主体的转变:亚当斯指出,传播弧的弧度本质上源于其作为一种转化性力量。同时,这也是传播的内在属性,哪怕表面看来平实的流动与传递亦是如此^[23]。因此,传播弧与布尔迪厄(Pierre Bourdieu)提出的“轨迹”(trajectory)概念相关^[24],即每一次传播都可能引发主体在认知、情感与身份认同上的变化。这一过程体现为传播弧所具有的反身性,能够在传播的过程中折返并作用于主体自身,使其在陌生与熟悉、自我与他者的“彼此共在”(being-with-one-another)中^[25],重新理解并界定自身的存在与关系。

由此可见,传播弧理论将传播重新界定为一种具身性、关系性与转化性相互交织的动态过程。在这一视域中,传播本质上是一种多元主体与多重力量在“会遇”(encounter)中不断生成、协商并重塑意义的开放性实践^[26]。将其运用于侨批研究,可揭示这一跨国媒介实践如何在绘制传播弧的过程中,持续构筑并维系一个“离而不散”的跨国意义网络,并在这一动态过程中不断推动着海外华侨身份认同的演变。

(二)叶家侨批:嘉应华侨叶氏兄弟的跨国流动与家庭联结

在近代南洋华人移民维系故土联系的实践中,侨批构成了跨越地理边界的核心“情感-物质”纽带。本文以晚清嘉应华侨叶氏兄弟往来的侨批档案为分析对象,旨在探讨侨批如何在侨乡与侨居地之间绘制传播弧。

本文核心材料为暨南大学出版社2017年整理出版的《馆藏晚清侨批选读》中所收录的叶氏家族侨批。该文献主体为1881年(光绪七年)至1910年(宣统二年)间,叶和仁、叶清仁、叶礼仁三兄弟从其侨居地马来亚、印度尼西亚等地寄予家乡母亲钟氏的95封侨批,系近代粤东客家移民向东南亚迁徙历程中留存下来的珍贵民间文献,现藏于广东省汕头侨批文物馆。

叶氏兄弟移居南洋之际,正值晚清政府对



出洋政策的逐步放开。1860年《北京条约》首开华工出洋的法律通道；1877年清廷在新加坡设立领事馆，标志着对南洋华侨的正式承认与保护；至1893年，清廷正式废除海禁，颁布“除海禁旧例，良善商民准其出洋贸易”^[27]的上谕，民众获得合法出洋的权利。叶和仁（1881年）、叶清仁（1895年）、叶礼仁（1899年）的南渡时间，恰好贯穿了这一政策从松动到废止的转型。与此同时，19世纪末的南洋殖民地也经历着深刻的经济变革。据1891年清朝驻新加坡总领事黄遵宪报告：“南洋各岛华民不下百余万人……华人乃占十之七，其中粤籍华侨占相当比重。”^[28]随着西方工业革命对战略物资需求的激增，马来半岛的锡矿开采和橡胶种植业进入高速发展期，带动了坝罗（今怡保）、檳城等新兴市镇的繁荣。叶和仁从金店学徒起步，逐步建立起以“叶宝源”“新和昌”“广加昌”为字号的跨域金店网络，其经营轨迹恰与这一波华侨经济的整体繁荣同步展开。

在侨乡一端，晚清潮汕及嘉应等粤东地区社会呈现出高度依赖海外接济的经济特征，时人谓之“仰给外洋”，意指家庭日常之盈亏、婚丧嫁娶之开销乃至生存底线之维系，皆有赖于海外亲人的侨汇支撑。一旦侨汇中断，则“阖室待毙，束手无策”^[29]。批信中反复出现的银钱与物资，不仅是家庭日常开支的支柱，更嵌入在侨乡社会“恃洋为命”的整体生计结构。

叶家三兄弟自光绪七年（1881年）起陆续从广东嘉应州程乡县罗衣堡南渡，近三十年间命运交织又各自展开。长兄叶和仁最早随父出洋，从吉辇、檳城金店学徒做起，历经承接、自营、合股，逐步在坝罗、芬乃容（今兰沙）等地建立起跨域金店网络。其间虽遭遇火灾、诉讼等挫折，却凭借着灵活经营与多方合作不断壮大。他先后纳陈氏、廖氏、黄氏三妾，育有有德、有福、有增三子，成为家族经济的支柱。二弟叶清仁于1895年南下投兄，初期在叶和仁店中帮工，后尝试在四里汶（今泗水）独立开店，然“开有三月之久，又被火烧”^[30]，辗转红毛丹（马来亚旧地名，具体位置今已难详考）自创金店

并娶广府女子马氏为妾。其事业虽未及兄长风生水起，却始终在家族网络中保持独立经营。幼弟叶礼仁1899年随兄南渡，初依附店中学艺，后经叶和仁资助在石城（马来亚旧地名，具体位置今已难详考）、芬乃容开设小金店。经营虽稳却常感寄人篱下，多次在批信中劝诫家乡亲友“切不可来（番）”^[31]，流露出新客谋生的艰辛。

三兄弟的奋斗史不仅体现在金店、锡矿与当铺的实业扩张中，更凝聚于近三十年间其依托水客寄递的近百封侨批之中。这些批信以细密的笔触勾勒出家族伦理的维系。叶和仁不断汇款修葺父坟、嘱托为弟娶亲、安抚嫡妻陈氏；叶清仁惦念亡故发妻吴氏的祭祀；叶礼仁则反复叮咛母亲钟氏“加食增衣，不必省俭”^[32]“强饭加衣，家事放闲”^[33]。侨批中银钱与物资的流动，不仅支撑了原乡家族的生计，更承载着游子“俯身侍奉”以尽人伦的孝心、“答谢神恩”不忘庇佑的虔诚与“追思故人”慎终追远的牵挂。而批信间偶尔流露的兄弟龃龉（如叶清仁抱怨兄长远欠账目）、经营分歧与婚姻纠葛，亦真实映射出传统家族在异域环境下的适应与张力。

叶氏家族的历程，正是晚清华侨从跨国流动到家庭联结的缩影。《馆藏晚清侨批选读》中详实的批信表明，当时南洋华侨凭借侨批搭建起了一套虽不稳定却充满韧性的跨域通信与汇兑网络。叶氏兄弟与母亲钟氏持续数十年的批信往来，所传递的远不止银钱与物资，更包括了家庭礼仪与代际情感。这些泛黄的信纸，成为近代华人移民史最鲜活、最细腻的微观见证。因此，尽管叶氏家族的跨国经历无法涵盖华人移民史的多元面貌，却以一个截面生动地展现了早期移民如何借助侨批这一媒介，构筑起超越信息传递功能的“社会-文化”意义网络。

三、侨批网络中“传播弧”的绘制实践

（一）“具身”与“化身”：水客的运作逻辑与双重角色

作为晚清华侨社会中关键的非正式传播节点，水客的“运作机制”深植于其身体的物质

性与社会性存在。他们的“走水”^①活动，可视作对传播弧这一抽象概念的具身化、日常化实践。因此，每一次跨越重洋的往返，都是一段以身体为媒，将离散的亲情、经济与乡谊在空间中重新连接的过程。

在电报与邮政等现代通信基础设施尚未普及的时期，水客的流动构成了侨批网络得以维系的基本条件。在现存 95 封批信中，仅经明确记载的水客便近五十位。其中，核心水客如“朱贵兄”往返记录达 52 次，“梁水兄”与“黄璉兄”各为 15 次，构成了侨乡与侨居地之间稳定且高频的联系通道。信中反复出现的“托某某兄带回”“兹逢某某兄回唐之便”等固定表述，直接反映出该系统的运行高度依存于水客的个人行程与移动周期。

值得关注的是，这一网络本身具有高度的弹性与延展性。除长期往来的核心水客外，亦不乏仅出现数次、乃至仅有一笔记录的捎带者（表 1）。由此可知，水客实际运作所形成的流动网络，远比文献中静态记录的名单更为庞大，也更为灵活。它并非一个僵化的官方系统，而是一个基于地缘（嘉应州）、血缘（叶氏宗亲）和业缘（金店同行）不断编织、动态调整的民间自组织网络。一位同乡的偶然返乡、一位族亲的商业旅行，都可能被即时吸纳进这个网络，成为一次临时的信息传递通道。水客的身体，如同在侨乡与侨居地之间往复穿梭的梭，织就了一张覆盖广泛、深入内部的“毛细血管网”，确保了离散家庭

表 1 叶家侨批水客名录统计表

序号	水客名	递带数	批信出处编号	备注
1	朱贵兄	52	5、6、7、8、9、11、12、13、14、15、16、17、18、19、20、21、22、23、24、39、52、53、55、56、57、59、62、63、64、65、74、81、83、85、89、90、91、92、93、94	批信又写“朱桂兄”，递带侨批最多、时间最长的水客
2	梁水兄	15	1、3、21、27、29、36、37、43、66、67、70、71、72、73	批信又写“水伯”“梁水伯”，称谓发生变化应是随年龄增长
3	黄璉兄	15	30、36、44、45、47、49、51、54、66、67、70、71、72	批信又写“黄连兄”“黄亚（阿）璉兄”
4	张柬生兄	12	38、55、63、83、85、90、91、92、93	批信又写“张柬星”“张见生兄”“张简生”“张闾星兄”。客语“星”“生”同音，“柬”“简”“见”音近，“闾”“柬”可能错混
5	张漠兄	10	4、5、7、8、11、14	批信又写“张巫”。客语“漠”“巫”都念 mo
6	梁恒三兄	8	26、30、36、44、45、51、54、61	
7	李永兄	8	45、49、50、51、54、58、62	批信又写“李永”
8	张姑爷	7	34、40、78、81、82、83、85	批信又写“张姑丈”“张□姑丈”，应是叶家亲戚
9	秦庆兄	7	86、88、89、91、92、93、95	
10	阿坤兄	5	1、8、18、22	批信又写“叶阿坤叔”“坤叔”
11	叶焕文	4	53、86、88、91	批信又写“焕文兄”“焕文叔”
12	张信兄	3	11、12、24	
13	杨荣兄	3	16、17、19	批信又写“杨亚（阿）荣兄”
14	张敦元兄	3	21、23、24	批信又写“张敦兄”
15	黄四	3	31、36、37	批信又写“黄四兄”
16	廖云秀兄	3	60、77	批信又写“廖云秀”“云秀”
17	张上兄	3	76、83、85	
18	叶坤兄	2	5、8	
19	家祥叔	2	12、13	

① 所谓“走水”，是侨乡对水客活动行为的俗称。水客往返于唐山和番邦之间，从侨居地收集侨批、货物和口信搭乘海船返回侨乡，递送给侨眷，取得回批，再返回南洋，递交于华侨手中。如此循环一次称为“走水”。



20	张宏哥	2	21、24	批信又写“张宏兄”
21	李庆哥	2	21、24	批信又写“李庆兄”
22	张信福	2	21、81	批信又写“信福兄”
23	邓增兄	2	36	
24	李康六兄	2	46、52	
25	梁辉兄	2	54、57	
26	林瑞庭	2	72、74	
27	李登郎兄	2	73、74	
28	张德轩	2	71、75	批信又写“张德轩兄”
29	钟亲鹤叔公	2	79、84	批信又写“种学叔公”，“钟”“种”似错混，客语“鹤”“学”同音
30	其文叔	1	1	
31	叶仁兄	1	3	
32	家仁叔	1	4	
33	亚德叔	1	4	
34	温炳兄	1	18	
35	叶根兄	1	19	
36	梁恒兄	1	25	
37	亚（阿）水弟	1	26	
38	家宝叔	1	36	
39	黄咸兄	1	45	
40	叶保叔	1	45	
41	染□□兄	1	45	
42	开华叔	1	46	
43	叶阿云	2	67	
44	叶和兴	1	67	
45	张盘石兄	1	78	
46	张车兄	1	79	
47	张基□	1	81	
48	张贤兄	1	84	

在时空阻隔下仍能维持最基本的联系。

然而，仅依靠身体的物理移动并不足以保障这一系统的有效运行，其更深层的关键在于一套深植于地缘与血缘关系中的、高度人格化的非正式信任机制。当华侨将积蓄托付给水客时，交付的远不止是财物，更是基于同乡、同族乃至世交关系，并经由长期口碑沉淀而成的深厚信

任。这种信任具有极强的选择性与梯度性。一个生动的案例是，光绪二十二年（1896 年）冬，叶和仁在马来亚欲寄款回家，恰好有两位水客即将返乡，一位是交情深厚的“朱贵兄”，另一位是往来较浅的“叶根兄”。面对此情境，叶和仁做出了审慎的区分处理。他在给母亲的信中特意说明，“俟候朱贵兄回唐，定有银钱寄回。兹今便逢叶根兄，仅寄来银二大圆（元）”^[34]。这短短的一句，一定程度上揭示了侨批系统中的信任逻辑，即大额款项必待“朱贵兄”返乡方予托寄，而对一般水客则仅委托小额。该策略既体现了侨户在面对通信断裂与资金风险时的自发管理意识，也映射出水客行业内部基于个人声誉所形成的隐性竞争与信用差序。因此，信任已内嵌于水客的社会身份与身体实践之中，成为一种“具身协议”，构成了系统得以持续运作的核心约束机制。所以，在叶家侨批中，不少批封盖有水客个人私章，如信（七十二）与信（七十四）批封都盖有“嘉应林瑞庭手带”的印戳。

水客通过频繁的跨境往返，以身体力行的方式构建起信息与物资流通的实体网络。与此同时，他们个人信誉所凝结的信任关系，也成为支撑该网络运行的隐性“协议”。这表明，侨批网络的运转不仅依托于身体这一最朴素的基础设施，也

印证了德布雷所指出的媒介具有双重面向：一是“组织性的物质层面”，即物质化、工具性的技术系统；二是“物质性的组织层面”，即潜在于技术系统中的社会规范与组织原则^[35]。前者可见，后者无形，媒介既通过环境推动事物传递，也被环境的选择所塑造。

在这一网络中，水客的身体不仅作为侨批传

递的“基础设施”与信任“协议”，更在亲人长期缺席的家庭时空中，被赋予了一种象征意义，“化身”为华侨身体的替代性在场，从而在空间阻隔与时间延宕中，维系了家族联结的情感真实性与仪式完整性。

从叶氏家族侨批往来的实际运作中可见，水客的作用早已超越单纯的银信传递，展现出多层次的复合角色。以叶和仁、叶清仁兄弟托“杨荣兄”携带侨批返乡一事为例，其特意在信中嘱咐母亲与水客杨荣“亲见一面”，原因在于杨荣对兄弟二人在南洋的境况“足悉”^[36]。这一行为表明，水客在递送书面信息之外，承担了关键的口头信息补充与情境阐释功能。他们能依据自身见闻，对侨批中的简略内容进行细节补充、背景说明乃至情绪渲染，从而在很大程度上弥补了因书信延迟、表达限制或文化程度差异造成的信息不对称。

更重要的是，水客不仅作为海外亲人在故乡的“耳目口舌”，也作为具有执行力的“手足”，以类亲属身份履行实际的家庭职能。例如，信中“然本年倘若吾家中有欠多少良（银）项，代为借过，容后来年正月寄依（信）通知，吾即当付回”^[37]，以及“各兄盘钱不敷，祈向水客借，过到之日自晓完（还）他”^[38]的嘱托表明，水客凭借其在地判断，直接介入家庭经济运作，承担信用评估、资金周转与债务协调等事务，扮演起“临时财务管家”的角色，从而使离散的家庭能够在空间分隔中维持统一的经济意志与行动能力。由此可见，水客通过其“化身”实践，深度嵌入地方人情伦理网络，成为家庭内部跨域治理的一种方式。

水客通过其周期性的具身在场，已然成为离散家庭情感结构与社会关系中一种独特的化身存在。对于留守眷属而言，水客可信的身体及其带来的银信与口述，转化为一种可触摸的“家在”确证，有效地缓解了他们因分离、阻隔而产生的焦虑与不安；于海外华侨而言，托付水客不仅是寄送钱物，更是将自身思念、关怀与责任寄托于一个可返乡的“分身”，使其在异乡仍能维系作为儿子、兄长或丈夫的角色感知与情感联

结。正是依赖于水客游移于“外人”与“内人”之间的角色化身，抽象的亲情、责任与文化纽带得以转化为可预期、可信任、甚至可“触摸”的具身实践，从而在技术受限的时代，编织出坚韧而充满人情的跨国家庭维系网络，真正实现了“离而不散”的跨境生存状态。

（二）“沉默”与“过剩”：一套关于孝养的情感策略

在近代华侨与家乡的跨国联系中，侨批不仅作为经济汇兑的实用媒介，更成为情感传递、记忆构建与家族伦理实践的重要文本空间。在叶家侨批中，呈现出“沉默”与“过剩”两种看似相悖、实则互构的话语实践。二者共同构成离散家庭在跨域生存中的一套关于孝养的情感策略。它不仅回应了地理分隔所带来的结构性困境，更通过话语的调适与补偿，持续维系着家族的象征完整、伦理认同与情感联结。

在叶氏批信中，沉默首先体现为对特定家族叙事的策略性回避。最显著的例证，莫过于对叶和仁父亲南洋经历的集体失语。光绪七年（1881年），叶和仁曾寄回一张“一对中年夫妇”的相片，相框背面标注“母亲大人钟氏收”。据编著者推断，相中男性“绝非叶和仁本人”，而“最大的可能是叶和仁的父亲和庶母”^[39]。然而，在此后近二十年的批信中，这位父亲何时过世、如何安葬、在南洋有何事业，均无直接记载。直至光绪二十五年（1899年），叶和仁才首次提及“先父之地坟之事，留心修整”^[40]，但对其生平经历仍讳莫如深。这种沉默实为一种家族叙事的策略性安排。父亲作为家族南迁的开拓者，其生平因涉及在洋再娶、财产分配等敏感议题而成为家族记忆的“暗区”。通过对这些议题的悬置，叶氏兄弟维系了母亲钟氏在家族叙事中的核心地位，也避免了嫡庶矛盾在文本中的显现。父亲的存在，仅通过“地坟修整”等仪式性话语被偶尔唤起，其具体形象与经历则始终处于话语的边缘。

此外，批信中的沉默还体现为对个体情感创伤的克制表达。叶和仁在光绪二十二年（1896年）遭遇火灾并一度被拘，信中仅以“男本年运



气不嘉（佳），前月被火所累，坐困縲洩（继）月余”^[41]一笔带过。一场重大变故被收敛为“运气不佳”的轻描淡写，个体的创伤经验被折叠进命运式的概述之中。这种含蓄既是对苦难的自我消化，也是对家乡亲属的情感保护，从而避免远隔重洋的亲人因无力相助而徒增焦虑。

诚然，侨批经水客、代笔者乃至族亲辗转传阅，这种半公开性构成了特定的“约束性语境”，框定了某些话题不宜公开言说的边界。然而，细察叶氏批信可以发现，这种约束非但没有使书写沦为被动的信息筛选，反而催生了离散华侨对“不说”之道的精妙掌控，使之成为一种集情感保护、关系治理与创伤管理于一体的话语策略。可以说，在叶氏兄弟的批信中，沉默远非信息的真空，而是一种根植于跨洋生存处境的适应性选择。

与沉默形成鲜明对照的是，叶家侨批中涌动着一种近乎重复、频繁且细致的对母亲身体的关怀。这种过剩的话语并非冗余，实则是漂泊南洋的儿子们为抵偿地理隔阂与赡养缺位，而主动采取的一种补偿式的孝养行动。在叶氏兄弟的侨批中，对母亲身体状况的问候几乎成为每封信函开篇的固定仪式：“无奈天涯远隔，音信难通，未悉母亲大人在家安否？”^[42]“男在外不时挂念。”^[43]此类语句在不同年份、不同兄弟的信中频繁出现，成为一种模式化的情感表达。这种重复的问候，首先可被解读为是一种情感焦虑的结构性外化。远在异乡，儿子们丧失了感知母亲身体状况的具身性可能，时间与空间的不确定性加剧了他们的牵挂。每一次询问“玉体金安否”^[44]“微疾方有全（痊）愈”^[45]，皆是对母亲生命状态的一种象征性确认，亦是通过语言仪式缓解因“至亲病复发”^[46]，无法亲身照料而产生的无力感与存在性焦虑。其次，该行为也构成了一种跨国语境下的伦理身份展演。在儒家孝道文化中，“父母在，不远游”是理想准则，而远游的儿子们则通过不断重申关怀来强化自己虽身在海外，却心系家庭的道德身份。问候的重复性，实则是他们在无法“晨昏定省”的情况下，用以维系孝子形象、对抗

内心愧疚的情感仪式。

除了言语问候，叶氏兄弟尤为注重委托水客寄送药品与补品，将对母亲身体的关怀物化为可触摸、可使用的物质实体。批信中显示，三兄弟频繁寄回诸如“虎膏”“猴膏”“海参”“鹿茸”“阿胶”“如意油”等物品，而这些物品的选择具有明确的指向性，虎膏、猴膏用于祛风强骨，海参、鹿茸、阿胶为滋补上品，如意油用于缓解疼痛。这些针对老年人常见病痛的治疗方案，实则是儿子们在对母亲进行远程的身体疗护。他们凭借自身在南洋积累的经济能力与文化认知，将异域的物产转化为具有疗愈价值的“情感物质”，并通过侨批网络实现其跨洋流动。这一过程不仅是资源的流动，更是一套嵌入于家族伦理与情感经济的身體治理技术。而物品的过剩也揭示了一种补偿性的积累逻辑，即叶家三兄弟试图通过物资的高度密集（如一次寄送三十六条海参^[47]）与持续输送（“又另代买大海参一员（元）”^[48]），来填补因地理分隔而无法实现的对母亲的日常照护与情感互动，从而在跨国语境下重构“孝”的物质表达。

因此，侨批传播中的沉默并非信息缺失，而是围绕家族体面、情感创伤与权力关系形成的策略性回避；而过剩亦非无意义堆砌，而是离散情境下华侨身体缺席的情感代偿。二者共同构成一套相互补充的叙事机制，维系着叶氏兄弟对母亲的跨洋孝养伦理与家族认同。

（三）“时间”与“空间”：侨批的时间修复与水客的空间接合

对于被海洋分隔的家族而言，最根本的撕裂不仅是物理空间的遥距，更是时间体验的同步性与连续性的丧失。因此，侨批更深层次的文化意义，在于对离散经验中破碎的时间秩序进行编织与修复。

侨批首要的时间功能，在于将海外华侨生命成长过程中，对至亲而言“不可见”的生命时间，转化为可阅读的、连续的家族叙事。在叶氏家族批信中，我们能看到三兄弟从“漂泊无定”到“落地生根”的成长心路历程。长子叶和仁的早期批信透露着“初到地方，诸事未

悉”^[49]的青涩与依附，信中多为禀告平安、接收指示。随着时间推移，他的批信内容逐渐演变为汇报“承接金店”^[50]“迁移店址”^[51]“开设分号”^[52]，直至后来在信中从容调度家族事务、纳妾生子。其书写语气也从谦卑的“不肖（孝）男”^[53]，转变为家族的决策者。同样，三弟叶礼仁的书写轨迹也从初抵南洋时“欲自己出来谋新头路”^[54]的迷茫，逐步走向独立经营的筹划。这些分散、间隔数月的批信，如同断续的胶片，当被家族收集、保存并按序阅读时，便连缀成一部生动的“出洋创业史”。侨批通过持续的、阶段性的叙事汇报，给家乡亲眷创造了一种“延迟的在场感”，让母亲能在脑海中勾勒儿子们从学徒到东主、从孤身一人到开枝散叶的生命轨迹。这不仅是对信息的传递，更是对生命时间本身的共享。

如果说线性时间记录了海外华侨的成长与生活的演进，那么循环时间则关乎群体的认同与文化的归属。华侨身处异质的文化环境，极易在时间节奏上与故乡脱节。侨批通过紧密绑定中国传统的时间循环，实现了文化的锚定与情感的跨洋同步。这种时间的同步性突出体现在两个方面。

其一，节庆时间的同步。批信的物质形态与寄递节奏深度内嵌于中国传统节律。例如，叶清仁在光绪二十六年（1900年）正月寄出的“红批”（图1），以其红色信笺与封皮这一春节专属符号，宣告与家乡共度佳节；清明、中秋等关键节点前后的批信与额外汇款，亦使华侨得以虚拟参与家庭的祭祀与团圆仪式。

其二，生命周期仪式的参与。侨批使华侨能够跨越空间，介入婚、丧、祭、诞等界定家族生命周期的关键仪式。叶和仁在纳妾后专款“拜祖公应用”^[55]，是通过经济行为将新的婚姻关系禀告祖先，使其获得家族时间序列中的合法位置；叶清仁汇款嘱母为亡妻“买牲仪到坟前敬奉”^[56]，则是通过资金与托付，履行悼念责任，

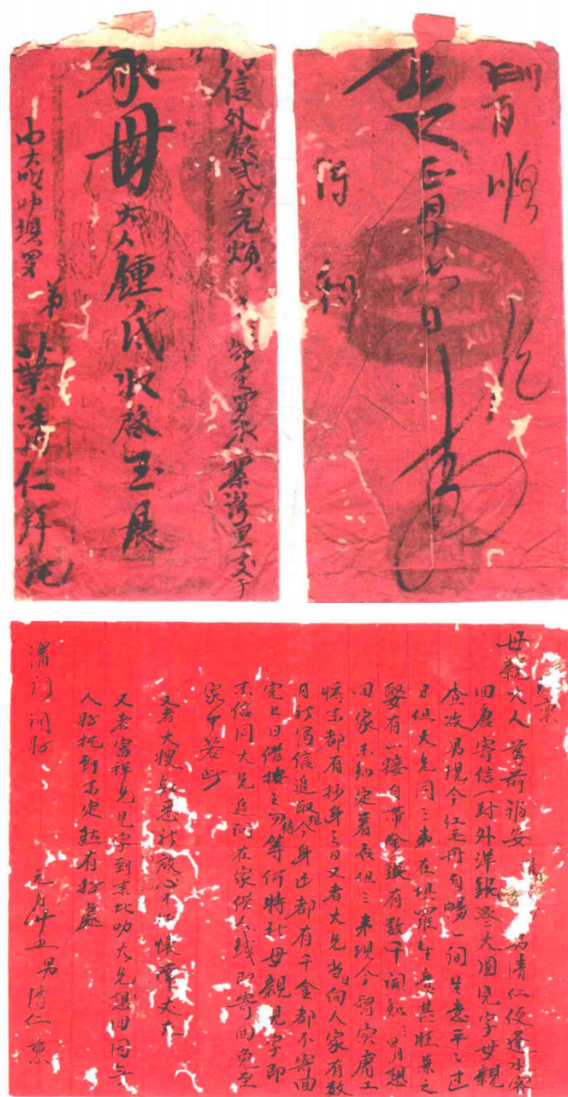


图1 叶清仁于清光绪二十六年（1900年）寄往家乡的春节“红批”信笺及批封

使死亡事件在家族记忆中得以妥善安放并周期性地被唤起；而三兄弟尤其是叶和仁，在侨批中反复询问、确认母亲钟氏寿诞的具体日期^①，正是为了令跨洋寄达的“寿礼”与祝福能“准时”抵达故乡。

通过侨批，华侨将自己纳入故乡周而复始的时间循环中，确保自己不在场的身体依然能在重要的文化时间节点上维持一种“仪式性在场”。

① 在传统中国社会，尤其是女性长辈，庆祝的往往是生辰（八字），而不仅是公历日期。晚辈（尤其是儿子）可能只熟知其生辰的农历月份与大致时段，而具体日期（如初几）因常年在外可能记忆模糊。



侨批中最具张力的一种时间形态，是关于“未来”的反复承诺。这种承诺几乎充斥于每一封批信：“冬成定着有一人回家”^[57]“明年定夺一准回唐”^[58]“容日往叨，定晓嘱兄回家”^[59]……然而，南洋谋生的现实困境，如“生意冷淡”^[60]“交带（代）不下”^[61]“意外开销（“身沾微疾……用去有银百余元”^[62]）”或“目（眼）下钱才（财）未凑”^[63]等原因，常使这些承诺难以如期兑现，从而在时间上形成了“承诺一展期一再承诺”的循环。表面看来，这是对经济义务与返乡计划的一再推迟；实质上，它却成为华侨应对不稳定生计的一种主动策略。通过不断将责任的履行推向不久的将来，如“男随后略有光景，自晓回家布置家中田产”^[64]，寄信者既缓解了家人的焦虑，也为自己的奋斗赋予了持续的意义，将当下的艰辛合理化、意义化。对收信的家庭而言，尽管希望屡屡延后，但持续的承诺本身已构成一种跨越重洋的情感联结方式，使分离的亲人始终共同持守着一段关于未来的叙事。

与此同时，侨批的“时间性”在其空间展开中始终内嵌于水客的移动节奏。水客依循“走大帮”（集中于春节、清明、中秋等传统节日前）与“走小帮”（平日不定期往返）的规律移动^[65]，不仅完成了地理意义上的跨越，更在离散家庭的通信结构中嵌入了一套外在的、周期性的时间制度。这种由行业惯习所塑造的空间往返节奏，与侨批的时间性形成了结构性共振，共同调节着家庭内部的经济预期与情感期待。因此，侨批的空间性与时间性始终处于相互构成、协同运作的关系之中。藉由侨批，断裂的时间、阻隔的空间得以被修复、接合，建构出一个既维系家族代际连续性、又整合文化认同的第三空间。这一空间不仅为海外华侨的跨国生存提供了秩序与归属，亦在文化层面赋予了离散经验以深层意义。

四、差异之弧：南洋移民认同的生成与流变

传播弧的形成，深嵌于晚清中国独特的历史结构之中。彼时的“唐山”正处于传统帝国

秩序松动、太平天国运动余波未平、人口压力空前加剧的变局之中，宗法社会虽表面稳固，实则内里已现裂隙；而“番邦”则在西方殖民扩张的浪潮下，对劳动力的迫切需求形成了强大的吸纳力量。正是在这一推一拉之间，近代中国历史上规模空前的跨境人口流动得以形成。在“家”与“国”的双重困境中，华侨被迫以个体的肉身跨越重洋。这些华侨多是出身于粤闽农村的底层农民，既承受着传统宗法伦理对孝悌的规训，又不得不面对殖民经济体系下残酷的生存竞争。正是这种在故土难以立足、在异乡难以扎根的“悬浮”状态，构成了传播弧得以生成的历史前提。

正如亚当斯所言：“弧线具有方向性，但不只是单向的、因果的或顺序的联系；相反，使其具有方向性的是弧线两端在意义和影响上的差异。”^[66]换言之，传播弧的方向性本质，是一种由差异驱动的转化过程。侨批所衔接的，远非地图上两个静止的坐标，而是两个在意义、资源与情感上存在结构性落差的鲜活世界。这条弧线的存续，不在于抹平差异，而在于通过连接与会遇的姿态创造了一个流动的时空环境，从而将弧线两端的张力转化为身份与文化再造的资源，形成一种不同事物之间的特殊共振（resonance）及共同演化（evolve together）^[67]。

具体而言，在“具身”与“化身”两端，作为侨批网络的关键媒介，水客勾连了在意义结构上存在差异的两个世界，一个是以宗法伦理与人伦秩序为核心的“唐山”，另一个则是充满生存竞争与异质文化的“番邦”。这种结构性差异从根本上塑造了传播弧不可逆转的单向特征，即资源在由侨居地向侨乡输送中，一旦流动受阻或中断，其引致的后果是非对称的。对华侨而言，大多意味着社会声望与家族期待的折损；对侨乡眷属而言，则往往直接构成生存威胁与伦理承诺的崩塌。正是这种差异驱动水客的跨境流动，成为一场持续的意义协商与身份再造实践。其具身往返不仅传递着来自故乡的“孝义”压力，促使华侨在异乡作出经济回应；也使其以化身的形式，在故乡代行华侨作为儿子、丈夫与兄

长的伦理职能，令其情感联结与家庭权威得以在场和延续。因此，水客所维系的不仅是被海洋隔断的银信与货物，更在深层意义上调和了“唐山”的伦理秩序与“番邦”的生存现实之间的意义鸿沟。在具身与化身的双重实践中，水客持续形塑并重构了华侨的身份认同，使其从地理意义上的离散转向社会文化层面的联结，进而形成了一种离土却不离家的身份确证，塑造出离而不散的跨境生存样态。

在“沉默”与“过剩”两端，批信中的沉默与过剩构成了南洋华侨一套策略性的孝养实践与身份协商方式。沉默体现在对家族历史与生存创伤的选择性回避，这并非记忆缺失，而是为了在离散语境下维系“想象的家族共同体”的情感完整性；过剩则表现为频繁、密集的经济输送、事无巨细的情感叮嘱及有针对性的物质馈赠。这既是海外华侨无法履行“亲身侍孝”之传统义务的象征性补偿，更是其通过持续的跨境资源输送与话语介入，确立自身家庭经济支柱地位、构建相应的家庭治理合法性的关键策略。而正是因为侨批中沉默与过剩协同作用，在无形中松动了侨乡传统宗法权威的绝对性，从而为新一代移民重构家族内部权力关系与情感秩序腾出空间。在此过程中，华侨逐渐演化为同时嵌入乡土伦理结构与侨居地经济体系的复合行动主体。最终，借由侨批持续的策略性书写实践，其身份认同从依赖血缘、地缘的“先赋性认同”，转为一种更具流动性、情境性与功能性的“复合认同”，实现了在地生存策略与跨国文化归属的动态整合。

在“时间”与“空间”两端，对于被海洋分隔的家族而言，最深切的创伤不仅源于物理空间的遥距，更在于共同时间体验的丧失。侨批所构建的传播弧的核心意义，即在于对离散家族所遭遇的时空断裂进行修复与接合。其以叙事接续断裂的生命时间，以仪式调谐循环的文化时间，以承诺编织期待的未来时间；而这一切时间性实践，最终都被收束于水客网络所构筑的、一种兼具节律性与不确定性的流动空间构型之中。在这个过程中，我们可以清晰地勾勒出晚清华人

移民身份认同流变的动态轨迹。这一身份转变体现在其从落叶归根的暂居心态到落地生根的定居意识的演进中。这一转变与晚清时期南洋华侨社会的结构性变迁同步发生。19世纪末，随着橡胶、锡矿产业的稳定发展，以及殖民政府对华人定居者政策的调整，南洋各地开始形成相对稳定的华人社区；会馆、学校、报刊的涌现，使得华侨得以在异乡重建社会网络。与此同时，清廷于1893年正式废除海禁，承认华侨的合法身份，也在制度层面为这种身份转变提供了可能。

早期批信中“早日回家”的强烈期盼，在时间叙事上逐渐被“冬成回家”“明年定夺”“容后再议”等更具弹性的延迟叙事所取代。与之相应，三兄弟在南洋开分号、置产业、纳妾、生子等一系列在地营生，标志着其生活重心的迁移与人生规划的延展。叶和仁为儿子取名“有德”“有福”“有增”，更是一种在南洋建构新的家族时间轴与血脉谱系的象征性行为。这意味着其认同的根基已从遥望的“唐山”挪移至眼前的“南洋”。

可以说，侨批所编织的传播弧不仅承载着跨境银信的物质流动，更成为一个古老文明在近代化冲击下所经历的文化裂变与价值重构的缩影。侨批在“唐山”与“番邦”之间的周期性往返，勾连了两个世界之间的结构性张力。这一张力既未导向认同的消解，亦未固化为文化的割裂，而是经由侨批网络的运作机制实现了创造性转化。一方面维系了宗法伦理的情感纽带，另一方面吸纳了殖民地现代性的制度资源；既延续了乡土中国的文化根脉，又催生了海外华人的在地化认同。质言之，侨批所绘制的传播弧远非线性的传输通道，而是一个两端充斥着结构性差异、并持续进行意义转化的张力场域。正是在具身与化身、沉默与过剩、时间与空间的多重张力实践中，华侨的身份认同不再是一种非此即彼的静态归属，而是演变为在“唐山”与“番邦”之间、在伦理义务与生存现实之间、在过去与未来之间不断协商、调适与整合的动态生成过程。传播弧由此也揭示了：“身份的建构不是传播的结果而是传播实践本身。”^[68]



五、结语：“地方”与“传播”之间的纹理编织

若将侨批的跨洋流动视为一种传播弧的绘制过程，我们便可理解，传播并非简单的信息传递，而是一条联结异域时空、具身经验与多元意义的开放轨迹，其本质在于意义的转化、情感的共振与主体的重塑。由此观之，侨批不仅是静态的历史文本，其背后实则串联起了侨乡与侨居地之间一整套存在差异却又相互关联的意义网络。

在侨批跨越地理与文化疆界的具身流动中，传播弧如同穿梭的织线，在“唐山”与“番邦”之间往返牵引，既缝合了两地间结构性的空间断裂，也弥合了传播两端在情感与认同上的意义缝隙。这一绘制传播弧的实践，本质上是“地方”与“传播”持续互动的过程。最终，它编织出了一道道传播纹理。在这之中，地方与传播并非对立，而是在传播弧的绘制中相互纠缠、彼此生成、不断磨合。而这种纹理，也在传播与地方（传播弧具有差异的两端）共同演化的过程中逐渐显现为一种可被感知的存在，并随着侨批日益嵌入“唐山”与“番邦”的生活肌理、融为生存基底，最终悄然“隐入日常”，成为不言自明的背景。也正因如此，我们才有必要重释侨批，超越将其视作静态背景的“通道隐喻”，转而揭示地方纹理如何在传播弧的绘制过程中被动态编织，亦即意义网络在这一“地方—传播”互动中的持续建构。

（责任编辑：李韵奕）

参考文献

- [1] 孔飞力. 他者中的华人：中国近现代移民史 [M]. 李明欢，译. 南京：江苏人民出版社，2016：22, 38.
- [2] 饶宗颐. 潮州志 [M]. 汕头：潮州修志馆，1949：71.
- [3] 王鑫，崔思雨. 约翰·厄里“新流动范式”与传播物质性研究对社会学理论的接合与涵化 [J]. 国际新闻界，2023，45（3）：26-45.
- [4] 沈惠芬. 构建中国侨乡女性史：资料与方法的探讨 [J]. 福建论坛（人文社会科学版），2015（11）：149-156.
- [5][9][16][23][66][67] ADAMS P C. Geographies of media and communication II: Arcs of communication [J]. *Progress in Human Geography*, 2018, 42(4): 590-599.
- [6] JANSSON A. Texture: a key concept for communication geography [J]. *European Journal of Cultural*

Studies, 2007, 10(2): 185-202.

- [7] REVILL G. El tren fantasma: Arcs of sound and the acoustic spaces of landscape [J]. *Transactions of the Institute of British Geographers*, 2014, 39: 333.
- [8] LABELLE B. *Acoustic Territories: Sound Culture and Everyday Life* [M]. London: Continuum, 2010: 15.
- [10][22] HALLIGLEY A. Emplaced sounding: voice, identity and place in Zadie Smith's NW [J]. *Cultural Geographies*, 2022, 29(4): 499-513.
- [11] THRIFT N. *Non-Representational Theory: Space, Politics, Affect* [M]. London: Routledge, 2008: 5-15.
- [12] DELEUZE G, GUATTARI F. *A Thousand Plateaus: Capitalism and Schizophrenia* [M]. MASSIMI B, trans. Minneapolis: University of Minnesota Press, 1987: 3-21.
- [13] WYLIE J. A single day's walking: Narrating self and landscape on the South West Coast Path [J]. *Transactions of the Institute of British Geographers NS*, 2005, 30: 234-247.
- [14] JONES P I. Performing sustainable transport: An artistic RIDE across the city [J]. *Cultural Geographies*, 2014, 21(2): 287-292.
- [15] HOUSEFIELD J. Sites of time: Organic and geologic time in the art of Robert Smithson and Roxy Paine [J]. *Cultural Geographies*, 2007, 14(4): 537-561.
- [17] TYNER J A, INWOOD J F J, ALDERMAN D H. Theorizing violence and the dialectics of landscape memorization: A case study of Greensboro, North Carolina [J]. *Environment and Planning D: Society and Space*, 2014, 32: 902-914.
- [18] ADAMS P C. Placing the Anthropocene: A day in the life of an enviro-organism [J]. *Transactions of the Institute of British Geographers*, 2016, 41(1): 54-65.
- [19] ANDERSON B. *Encountering Affect: Capacities, Apparatuses, Conditions* [M]. Farnham, UK: Ashgate Publishing, 2014: 77-103.
- [20] SCHUTT S. Rewriting the book of the city: On old signs, new technologies, and reinventing Adelaide [J]. *Urban Geography*, 2016, 38(1): 47-65.
- [21] GRAHAM S. Vertical noir [J]. *City*, 2016, 20(3): 389-406.
- [24] BOURDIEU P. *Distinction: A Social Critique of the Judgement of Taste* [M]. NICE R, trans. London: Routledge, 1984: 103.
- [25] NANCY J. *Being Singular Plural* [M]. Stanford: Stanford University Press, 2000: 3.
- [26] ADAMS P C. Geographies of media and communication I [J]. *Progress in Human Geography*, 2017, 41(3): 365-374.
- [27][28] 薛福成. 请豁除旧禁招徕华民疏 [A]// 邵之棠. 皇朝经世文统编：卷三十八·内政部十二·养民 [G]. 台北：文海出版社.
- [29] 陈达. 南洋华侨与闽粤社会 [M]. 北京：商务印书馆，1939：89.
- [30][31][32][33][34][36][37][38][39][40][41][42][43][44][45][46][47][48][49][50][51][52][53][54][55][56][57][58][59][60][61][62][63][64] 潮汕历史文化研究中心，侨批文物馆. 馆藏晚清侨批选读 [M]. 广州：暨南大学出版社，2017：65；

119,104,131,186; 65; 108; 53; 45; 153; 85; 3; 83; 41; 15; 18; 41; 18; 4; 129; 148; 5; 15; 24; 69,93,95; 39; 159; 129; 161; 157; 127; 67; 161; 34; 62; 159; 114.

[35] 德布雷. 媒介学宣言 [M]. 黄春柳, 译. 南京: 南京大学出版社, 2016: 17.

[65] 周琍, 周建新. 水客与客家侨乡社会变迁 [J]. 中南民族大学学报 (人文社会科学版), 2009 (4):177-180.

[68] HECHT M L. *African, American Communication: Exploring Identity and Culture* [M]. NJ: Lawrence Erlbaum, 2003: 229.

“Arc of Communication” : The Construction of Meaning Networks in Late Qing Qiaopi and the Evolution of Nanyang Migrants’ Identity

Wu Yuye

Abstract: This paper reinterprets Qiaopi through the lens of the “arc of communication” theory within communication geography, moving beyond the existing research paradigm that views Qiaopi merely as a one-way “transmission channel”. Instead, it focuses on the meaning networks woven by Qiaopi as a dynamic transnational medium and their role in shaping the identity of Nanyang migrants. Taking nearly one hundred Qiaopi letters exchanged between the Ye brothers from Jiaying Prefecture, Guangdong, between 1881 and 1910 as a case study, this paper analyzes the practice of drawing the “arc of communication” within the Qiaopi network. It reveals how shuiké, through their embodied mobility and proxy presence, sustained cross-border family connections that were “separated but not dispersed”. The “silence” and “excess” within the letters constituted a set of emotional strategies, negotiating filial piety ethics and family identity in the context of diaspora. The continuous act of writing Qiaopi temporally repaired life experiences disrupted by spatial distance, and together with the periodic mobility of the shuiké, constructed a “Third Space” that maintained cultural belonging. Ultimately, driven and transformed by the structural differences at both ends of the “arc of communication”, the identity of overseas Chinese gradually shifted from a sojourning orientation of “falling leaves settle on their roots” to a composite state of “taking root in foreign soil”. In this process, Qiaopi wove a kind of “communicative fabric”, revealing the inherent logic by which “place” and “communication” mutually constitute and co-evolve through continuous interaction.

Keywords: communication geography; arc of communication; Qiaopi; Nanyang migrants’ identity

Author: Wu Yuye, School of Journalism and Communication, Xiamen University.



直面复杂系统： 计算传播研究的路径创新与方法前瞻¹⁾

宣长春 李嘉鑫

摘要 在传播学研究“复杂性转向”的背景下，计算传播研究虽在数据处理与方法应用上快速发展，但整体仍呈现出“以计算处理复杂数据”而非“以计算理解复杂性”的取向。本文以复杂系统理论为基础，引入“科尔曼之舟”作为“宏观—微观”跨层级分析框架，主张通过情境机制、个体行动机制与转化机制的有机串联，将宏观传播现象“局部还原”为宏观条件下微观主体的行动生成与交互过程，并在方法论层面提出双重革新方向：一是强化微观层级的过程性分析，从结果表征走向行动生成机制的深描；二是推动跨层级桥接，发展多层网络与递归仿真方法，以刻画“行动—结构—再行动”的动态演化迭代过程。溯因推理应成为计算传播学的核心认识论，通过与“科尔曼之舟”的有机结合，能够构建从宏观结果出发反向推断机制的理论生成路径。此时，计算不再只是方法论工具，更是探索宏观现象涌现机制的认识装置，这在一定程度上能够化解计算传播研究“只有计算，没有理论”的困境，以便在复杂性转向的语境下重建传播学理论的解释力。

关键词 计算方法；计算传播研究；复杂系统；科尔曼之舟；溯因推理

中图分类号 G206 **文献标识码** A

作者 宣长春，厦门大学新闻传播学院副教授，厦门大学国际传播研究中心研究员，福建厦门 361005；
李嘉鑫，厦门大学新闻传播学院硕士研究生，福建厦门 361005

一、问题缘起：传播学研究中的“复杂性转向”

近年来，传播学领域逐渐出现“复杂性转向”，即关注传播过程中的不确定性、相互依赖与涌现机制，将传播活动视为多主体互动中不断生成的复杂系统^[1]。这一转向并非简单基于新媒介技术或大数据条件的外在推动，相较于技术驱动的表层解释，其更深层的理论根源在于复杂性科学对社会现象理解方式的重塑，即从线性因果走向关系性、生成性与非线性机制的整体把握。这一重塑之所以在当下的传播学研究中凸显，一方面源于数字传播环境中流动边界、非线性、涌现、多样性和相互依存等概念的突出，使复杂性视角更具解释力^[2]；另一方面也在于复杂系统思想并非外在引入，而是植根于传播学理论传统之中。正如韩少卿与巢乃鹏所指出的，“传播理论一开始就拥有

1) 基金项目：国家社科基金后期资助项目“全球媒体与智库涉‘一带一路’话语建构与传播效果研究”(24FXWB017)。

着复杂性的思想内核。无论是传播理论的‘老三论’（信息论、控制论、系统论）还是后来的‘新三论’（耗散结构论、协同论、突变论），它们所处理与嵌入的核心都是复杂性”^[3]。然而，复杂性视角并不自动转化为具体的研究路径，它需要借助能够处理复杂结构与动态过程的方法体系加以实现。正是在这一意义上，以计算方法为核心的计算社会科学为复杂性问题提供了可操作的分析框架，并进一步在传播研究中具体化为计算传播学这一研究进路^{[4][5]}。

然而，相较于计算方法在实证研究中的快速扩张，复杂性视角在计算传播中的理论引入与系统整合仍显不足，整体上呈现出“以计算处理数据”而非“以计算理解复杂性”的取向。具体而言，在研究议题与方法层面，尽管近年来以生成式人工智能、多模态内容与算法权力反思为代表的计算传播研究议题不断涌现^[6]，大数据技术与自然语言处理等计算方法的应用亦持续增长，但研究范式整体上仍以回归分析、结构方程模型等传统量化路径为主导^[7]，方法创新更多体现为数据处理能力的扩展，而非复杂系统视角下解释框架的深化。进一步来看，即便是引入网络分析与仿真建模这类具有复杂系统分析潜力的计算方法的研究，其在方法应用与理论整合上仍缺乏对复杂系统理论的系统跟进。

在方法上，已有不少利用网络分析方法的计算传播研究尝试立足复杂系统视角，通过局部互动与反馈回路来说明舆论形态的涌现与非线性变化^[8]。相关实证工作多尝试描绘不同国家媒体、智库、平台用户等主体在议题层面的竞争与协同机制及其扩散轨迹，较为系统地刻画了多主体网络中的议题联动模式、级联扩散机制与跨国议程互动结构，在经验层面呈现出国际传播由线性影响走向网络化协同的演化图景^{[9][10][11]}。然而，从复杂系统理论看，这类研究更多停留在对关系结构的“表征性刻画”，而非对复杂性机制的系统整合，这些研究的思路往往将舆论系统抽象为同质节点及其连接关系，在一定程度上简化了复杂系统本身的多层次结构。而舆

论场中的突变、极化或结构性转向，其背后往往有多个维度的共振效应，如情绪因素、社群网络拓扑、主流媒体议程设定、政治经济背景等，并且各维度的要素之间又是互相依赖、彼此耦合的子系统^{[12][13]}。若仅停留在对微观主体之间的互动机制或单层级网络结构的考察，“跨层次”的因果机制则难以解释，也无法回答不同层级变量如何相互转化、结构性约束与个体之间如何互构等关键理论问题。

而利用仿真建模方法的计算传播研究在捕捉局部互动的全局效应方面精度显著提升，超越了系统动力学的抽象确定性与元胞自动机的网格封闭性^[14]，并能够将微观和宏观视角联系起来，为数据中观察到的经验模式提供过程机制解释^[15]。尽管仿真模拟被引入计算传播以增强对复杂系统的刻画，其方法论运作仍以预设规则与参数调试为核心：代理行为源于既有理论假设，通过“生成充分性”（generative sufficiency）检验既定机制是否再现宏观模式，本质上是服务于理论验证而非理论生成^{[16][17]}。同时，此类模型的应用多将宏观环境压缩为静态参数或局部约束^[18]，难以回答远离平衡态下的自组织与多层级反馈等复杂系统理论问题。更为突出的是，现有计算传播的仿真研究所采用的模型较为集中，以SIR模型及其变体（SEIR）为主^{[19][20]}，这在方法上便利了对信息扩散过程的刻画，但也在无形中将研究重心锁定于以个体为单位的传播演化框架之中，研究内容也自然收缩为“信息如何传播以结成结构”，限制了计算传播研究在理论解释方面的能力。相较之下，引入如自组织临界性等理论下的模型^[21]，则有助于突破个体主义分析范式，将研究视角拓展至跨主体关系结构及其整体演化过程，例如“媒体-国家”等行为体间的互动及其宏观结果的涌现，从而为理解传播系统的非线性跃迁提供新的视角。

因此，本文认为，计算传播研究要真正面对传播学的“复杂性转向”，真正从“数据计算”走向“理解复杂性”，亟需在理论框架、方法路径与认识论基础上进行整体推进。基于此，本文首先引入“科尔曼之舟”作为跨层级



机制分析的理论工具，以重构微观行动与宏观结构之间的解释路径，从以“局部相互作用”为核心的机制描摹，走向对“宏观整体结构—微观个体行动”之间复杂关系的综合解释。而后进一步从复杂系统视角出发反思计算传播方法的发展方向，强调由表征性分析走向生成性建模的转变，并在认识论层面引入溯因推理，将其视为连接经验数据与机制解释的关键路径，从而为复杂传播现象研究的理论生成提供支撑。通过三个层面的整合，本文尝试为计算传播研究提供一条面向复杂系统的理论与方法创新路径，深化复杂性理论与传播学研究之间的对话。

二、“科尔曼之舟”：理解跨层级社会机制的理论框架

在复杂传播系统中，一个重要的理论问题是，如何在不诉诸整体主义因果解释的前提下，系统说明个体传播行为与宏观舆论结构之间的关联路径。舆论形态与议题结构等宏观现象，既不能被简单还原为个体行为属性的线性加总，也难以被视作独立于个体行动之外的“自主社会力量”。因此，计算传播研究若要真正回应复杂性转向所提出的跨层级解释要求，就必须引入一种能够同时容纳结构条件、个体行动与宏观结果的机制性分析框架。“科尔曼之舟”（Coleman's Boat）正是在这一意义上提供了一个清晰而具有方法论指向性的理论模型。

（一）走向方法论局部主义：“科尔曼之舟”的理论回应

对复杂系统的解释在传统上秉持着一种偏向“整体主义”的思路，受到复杂性、自组织与涌现等思想的启发，强调宏观秩序的不可还原性与整体层级的解释优先性^[22]。然而，这一路径也面临着内在张力。一方面，它往往容易滑向神秘主义，即承认宏观性质的存在，却难以具体说明这些性质究竟如何在动态过程中由微观互动生成^[23]；另一方面，在宏观社会层次，无论是在单一社会系统的长期演化中，还是在不同社会系统之间的比较中，结构性变量的变化往往十分缓慢且幅度有限，这也使得相关假

设难以通过经验数据获得有力检验，进而陷入理论高度抽象却缺乏可操作性的困境^[24]。在这一背景下，引入“科尔曼之舟”作为核心分析框架，实质上是将复杂系统研究的重心由宏观层级的因果说明，转向对跨层级作用机制的过程性解释。这一取向并不意味着重新回到假定“社会整体具有自主因果力”的整体主义路径，而是强调社会因果必须落实为“随附于行动者之上的结构化选择与行为过程”^[25]。也就是说，社会机制的实在性并非来自抽象社会实体本身，而是来自“被社会构成并被社会定位的行动者”，他们在特定制度情境与规范约束中，根据自身动机与信念做出行动选择，由此承载并实现社会因果力^[26]。

在这一意义上，社会结构并非外在强加于个体之上的“黑箱”，而是以制度安排、角色期待、资源分配与信息环境等方式，转化为个体可感知、可回应的行动条件；而个体行动也并非纯粹的心理反应，而是在既定社会定位中被形塑，同时又试图施加影响的策略性实践。二者之间并非简单的单向决定关系，而是通过持续的互动构成递归性的反馈机制。在传播学研究中，无论是算法推荐机制还是舆论极化现象，其背后的驱动力最终都要回溯到具体情境中人类行动者的认知加工、选择偏好与互动策略，乃至非人行动者（如代码）的配置方式^{[27][28]}。由此，社会机制既不是脱离行动者的抽象力量，也不是可被完全还原为纯粹的个体属性，而是往返于“宏观—微观”互动之间的结构化过程与关系模式。

这种“方法论局部主义”（methodological localism）立场，使得社会因果解释既避免了将结构视为自主实体的整体主义困境，也超越了仅以个体心理为起点的还原主义解释路径，其实质上主张一种基础主义的分析策略，是“社会—个人”双向的建构主义：认为每一类社会机制都根植于制度情境、社会建构特征、个体发展轨迹与具体互动场景的组合之中，并通过行动者对规则、惯例与限制的回应而发挥作用^[29]。这一视角为面对复杂系统的传播学研究提供了

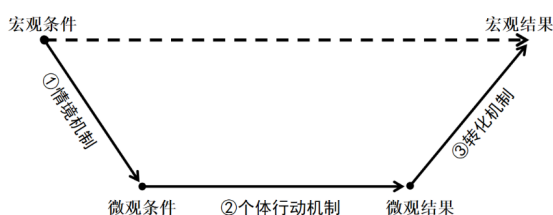


图1 “科尔曼之舟”机制拆解

坚实的哲学基础，促使我们将对宏观传播现象的解释锚定在具体的行动逻辑之上，复杂传播现象得以被解释为“可理解的生成过程”，而非仅仅是统计相关或整体属性的描述。

（二）复杂系统的因果链条：基于科尔曼之舟的机制拆解

对于传播学中的复杂系统研究而言，科尔曼之舟的价值在于它并未将宏观与微观割裂，而是提供了一种层次化且可操作的因果拆解路径。在该框架下，宏观的社会事实与微观的个人动机及行动都可以在同一机制链条中被同时纳入解释框架之中。在科尔曼之舟的理论框架下，复杂传播现象可以被理解为多重因果链条的交织过程，而非单一层级因素的直接作用。具体而言，其“复杂性”可被拆解为三个相互关联的机制路径：“宏观条件至微观条件”的情境机制（situational mechanism）、“微观条件至微观结果”的个体行动机制（individual action mechanism）和“微观结果至宏观结果”的转化机制（transformational mechanism）^[30]（图1）。

值得注意的是，“宏观—微观—宏观”的贯通逻辑在复杂系统与计算社会科学研究传统中并不罕见，譬如谢林（Friedrich Schelling）的隔离模型就清晰展示了微观行动者的局部规则（如对邻居种族构成的偏好阈值）如何在去中心化的互动中涌现为宏观的种族隔离格局^[31]。爱泼斯坦（Joshua Epstein）更明确提出“如果你无法生成（grow）它，你就无法解释它”^[32]，强调通过微观规则生成宏观结构的解释策略。因此，本文也并不认为，离开科尔曼之舟的框架复杂系统下的计算传播研究中就会面临不可弥补的缺口。本文更倾向于将科尔曼之舟的框

架定位为适用型的分析框架，其价值在于，在特定研究情境中为研究者提供一种更为系统的跨层级（宏观—微观）变量组织方式与因果机制拆解逻辑，从而使已有的计算传播研究在面对复杂性问题时更具解释深度与可对话性。由此，科尔曼之舟的框架在面对复杂性的计算传播研究中也具有其明确的适用边界。第一，研究问题涉及宏观结构变化，且研究者希望对这一互动过程做出机制性说明。值得注意的是，无论是宏观结构还是微观个体都并不指向某一特定实体或固定对象，而是一种分析层级上的概念指称，譬如在不同情境下，个人、家庭与国家之间可以构成相对意义上的宏观—微观的跨层次关系^[33]。第二，虽然其他学科的部分研究认可复杂系统计算方法（如ABM）的预测潜力^[34]，但亦有学者持怀疑态度，认为此类方法应用的出发点是回答“为什么”问题，而不仅仅是“是什么”^[35]。本文支持后者的观点，提出科尔曼之舟这一框架更适用于研究目标是解释性的而非单纯预测性的研究，若研究仅以预测大规模传播结果为目的，强行套用该框架则可能带来分析上的繁冗而非增益。

情境机制指向宏观条件对微观条件的有效约束，旨在解释特定场景对个体欲望、信念等个体行动前提条件的塑造。在计算传播研究中，这一机制可以被用于考察宏观层面的政策变动、地缘政治冲突或突发公共卫生事件如何改变传播主体的内部生产流程或资源获取方式等微观主体所处的“情境”或“结构”，进而引发后续的适应性行为^[36]。个体行动机制则是解释行动主体如何在特定情境下，通过认知加工、博弈与互动产生的具体行为。对这一机制的探讨在目前的计算传播学研究中最为常见，即利用大规模数据捕捉微观主体之间的信息流动、情绪传染或议程设置过程，关注“微观条件”如何转化为“微观行为”，以及微观主体之间如何发生横向的互动^[37]，以刻画微观主体如何在既定规则下进行策略性的互动与话语建构，揭示了传播系统内部的动力学特征。转化机制则是考察个体行动经累积而产生各种预期和非预



期的社会结果，即由微观结果上升至宏观结果的过程，以解释分散的个体行为如何通过聚集、相互依赖或反馈涌现为新的宏观结构^[38]。

这三类机制在既有社会科学研究中并非什么新事物，但在此引入科尔曼之舟这一理论框架并非可有可无之举，其价值首先在于能够结合计算方法系统整合传统社会科学中长期分散于情境、行动与转化等环节的过程性分析路径。具体而言，在传统社会科学研究中，情境机制的处理多以分组比较的形式呈现，涂尔干（Émile Durkheim）的自杀论研究就通过比较法国天主教地区和基督新教地区个人自杀率的差异，发现处于不同宗教环境中的个体自杀倾向会有所不同，由此识别了社会结构对个体行为的约束^[39]。行动机制的分析则更多集中于心理学研究，譬如将个体的行为解释为其出于维护自身利益的动机，在不同的结构约束与资源禀赋之下形成的差异化选择^[40]。转化机制，即分散的个体行为如何通过聚合、互赖与反馈涌现为新的宏观结果，在社会心理学领域得到了相对系统的形式化表达，门槛模型正是典型案例，其通过个体阈值差异在相互依赖中引发连锁反应和反馈积累，描绘微观选择向宏观结果转化的条件^[41]。由此可见，在每个机制阶段，传统社会科学研究都在不断探索深化，但由于其使用的学科视角与方法各不相同，因而难以形成贯通性的整体因果解释。显然，韦伯（Max Weber）在提出阐释社会学时就已经意识到这样割裂的分析路径无法充分把握社会行动的主观意义，因而提出以“理解”（verstehen）的方法，即在分析社会时考虑个体意义，对社会现象进行有意义的解释，并将个体与结构性联系起来^[42]。然而，囿于当时方法论与经验工具的限制，这一解释路径虽在理论上确立了宏观与微观之间贯通的分析取向，却难以对跨层级机制进行可操作的形式化表达，从而使其方法论想象长期停留于理念框架之中。而科尔曼之舟的引入，则使这一跨层级的过程性分析路径得以在计算方法的支持下获得直接而系统的贯彻，将情境、行动与转化机制整合为一条可形式化、可检验的因果链条。

这也点明了引入科尔曼之舟这一理论框架的另一价值，即其结构上内在契合复杂系统所呈现的跨层级递归与动态演化逻辑，为计算时代传播学研究面对日益严峻的复杂性问题提供了恰如其分的分析框架。复杂系统具有全局性^[43]、迭代性^[44]、持续性^[45]等特征，其宏观结构是在多主体相互作用中整体涌现的，演化过程遵循“结构-行动-再结构”的递归迭代，个体行为与网络的共同演化涌现集体行动、新网络结构与新制度规范，新集体性质反过来又会影响个体行为，周而复始，复杂系统在这一过程中层层升级、不断扩张^[46]。而既有计算传播研究在全局性上往往将平台结构与制度环境视为既定背景，虽揭示了个体行为模式（如虚假信息扩散研究中的选择性接触），却难以解释这些行为如何在整体层面生成或重塑结构^{[47][48]}。在迭代性上，网络扩散与实验研究虽揭示互动与反馈的重要性^[49]，但计算传播研究较少将宏观结果重新纳入下一轮行动情境，从而中断了“结果-情境”的递归；在持续性上，舆论动力学与主体建模虽实现了“微观生成宏观”的形式化表达^[50]，却往往依赖固定规则与初始条件，缺乏对结构长期演化与路径依赖的刻画。由此，与传统社会科学研究一样，传播过程被切割为若干孤立片段，要么停留于微观互动的即时机制，要么止步于宏观结果的静态描述，难以揭示复杂传播系统在时间与层级上的连续生成与反馈过程。因而，在复杂系统视角与计算方法条件已然具备的背景下，若仍不引入能够系统贯通情境、行动与转化机制的分析框架，对既有研究加以整合与推进，实属遗憾。科尔曼之舟正是在此意义上提供了一条整合路径。

总之，通过科尔曼之舟的机制拆解，我们可以将复杂的传播现象还原为清晰的因果链条，使对复杂传播系统的解释成为一种可操作且具有解释深度的研究路径。这一视角也为方法论层面的拓展提供了明确方向，它不仅要求研究者在不同分析层级上识别关键变量，更强调必须揭示这些变量是通过何种具体机制被联结起来并发生转化的，从而倒逼研究设计由“相关

性检验”转向“因果机制说明”。在此意义上，方法的选择应服务于机制链条中不同环节的识别与验证，推动计算传播研究从单一环节的刻画，转向围绕因果链条整合多源数据等多种方法，在动态过程中识别与检验跨层级机制。

三、方法论革新：计算传播研究方法的发展前瞻

科尔曼之舟为理解复杂传播系统提供了一套更为精细的理论框架，但要将其转化为经验层面的实证框架，则需要计算传播方法论的同步革新。若缺乏相应的方法论工具支撑，关于“情境”“行动”与“转化”的机制拆解则容易停留在抽象的思辨层面，难以在大规模数据中进行验证。现有计算传播研究议题广泛，呈现出方法更新与跨学科融合的整体趋势，但大部分研究仍集中在若干易于量化环节，并且使用的计算方法相对有限，特别是对个体行为逻辑的复杂性和动态性有所简化，并且对跨层级结构的关注不足，对复杂系统的理解和应用较为片面和抽象。因此，为了真正打开复杂系统中“宏观事实变化过程”的黑箱，我们需要在方法论层面进行双重重构。一是强化微观层级的深描，细化在行动机制层面对个体及其所生产内容解析的颗粒度，从单一文本的静态表征走向对生成过程及其内在机制的深层剖析，以确保微观基础的有效性；二是强化跨层级关系的桥接与迭代过程分析，即突破单层网络的平面视角，引入跨层嵌套结构与涌现分析，一方面动态捕捉宏观结构的约束作用及微观互动的宏观结果涌现，另一方面进一步刻画个体行为与结构条件在反复“生成-反馈-再生成”过程中的共同演化，弥补既有计算传播研究对跨层递归机制关注的不足。这面向跨层级机制解释的方法论革新，是计算传播学回应复杂系统挑战的必然选择。

（一）微观层深描：文本表征到行动生成机制分析

在“个体行动机制”所对应的“微观条件至微观结果”层级上，核心问题在于传播主体如何在既定情境中，通过认知加工、话语选择与互动博弈生成具体的传播行为或内容形态。然而，现

有研究主要依赖于平台数据抓取和统计分析，集中关注大规模用户行为的相关性与行为结果。这种基于相关性的分析本质上是后验性的，它难以穿透数字行为的表象，洞察个体在宏观结构下面对复杂信息的认知加工过程、决策动机或情境化原因。例如，对“过滤气泡”的研究往往聚焦于信息同质化的结果，却难以深入探究用户对算法过滤效应的认知图式与实践策略，或在线上线下社交关系中的信息离散性如何调节算法功能^[51]。对社交机器人干预舆论环境的研究发现，社交机器人仅需少部分比例就能对人类用户造成大规模沉默，这是一个重要而有趣的结果，但其背后的生成机制，特别是在行动与环境共同演化背景下的持续性生成机制，并未得到充分说明^[52]。因而，洞察人类和非人类个体行为结果的生成机制，避免对行动者行为逻辑的简化，是复杂系统下微观层级计算传播研究的重要面向之一。

为回应这一面向，一些研究已经开始通过扩展计算方法深入分析传播现象的内在形态与形成性过程，有效地弥补了以往研究对行动者行为逻辑复杂性过度简化的问题。首先，一些研究通过计算方法剖析文本和话语的内在复杂形态。例如，情感波动性（sentiment volatility）这一概念的提出，正是研究者尝试利用自动化情感分析来追踪语篇的“情感弧”^[53]，即叙事内容在不同语段内情感强度和方向的变化，并发现这种文本内的动态变化能够增强用户的刺激体验，为理解个体对复杂信息的认知加工提供了过程性计算路径^[54]。未来则可以将这种过程性分析方法从情感弧拓展到语义维度，结合 ALC 词嵌入^[55]，或基于语言模型的词义归纳（word sense induction）方法^[56]，探究核心词汇在语篇中如何发生动态的语义漂移，从而完善并丰富叙事弧的研究，以刻画文本意义如何在连续表达与语境约束中被逐步建构，使文本不再仅被视为态度或立场的静态载体。另有一些研究尝试利用计算方法深入分析传播内容的演化过程，如对国际舆论话语的研究则通过共同基础分析、关键词历时性分析等方法，从共享知识和共享解释两个角度，揭示了争议性概念在国际舆论场中得以持续流动的机制，一定程度上



避免了对行动者行为逻辑的简化^[57]。除了文本等计算传播研究中常规的分析对象,亦有研究开始针对平台算法等非人行动者的黑箱问题,通过代码内容分析与功能分解的方法,将研究视野从算法的外部效果转向了其内部形成过程,揭示了嵌入在代码中的人类新闻价值判断与过滤优先级机制,从过程性视角提供了对算法机制的关键说明^[58]。

在此意义上,这类在微观层级上刻画传播现象内部动态结构的方法,有助于更全面地将个体认知加工过程嵌入到更广泛的复杂系统全局性语境之中,并通过对语篇展开精细化描摹、对话语流动与算法运行展开过程性分析,为呈现“行动-反馈-再行动”的循环结构提供数据支持,为理解传播结构在时间维度上的持续性累积与演化提供可能。

(二) 跨层级桥接: 宏观约束与微观涌现的机制刻画

相较于“微观条件-微观结果”层级对互动过程本身细致刻画的要求,跨层级方法革新的核心任务在于解释不同尺度结构之间的互动关系与迭代过程,既包括宏观制度、舆论环境与技术架构如何被转译为行动者可感知、可回应的情境条件;也包括分散且异质的个体互动又如何经由累积、反馈与选择机制,涌现为具有稳定形态的宏观传播结构;以及个体与结构的共同演变如何进一步催生新一轮的演化过程。这正对应于科尔曼之舟中的情境机制与转化机制两个因果链条。面对这一跨层级的复杂系统问题,在网络分析上,传统以单层网络或单尺度变量为核心的分析框架已难以胜任跨层级因果解释的需求;在仿真模拟上,传统以参数设定为核心的ABM难以刻画个体与结构的共同演变的复杂系统“迭代”过程。由此,计算传播研究中方法的关键转向在于,其一,从将宏观因素视为外生控制变量,转向刻画行动主体与结构约束之间的嵌套关系,使宏观因素以可操作的方式进入微观互动模型;其二,从基于固定规则与参数设定的单向生成逻辑,转向对“行动-结构-再行动”递归过程的动态刻画,以揭示个体行为与网络结

构在持续反馈中的共同演化机制。也就是说,跨层级计算方法的核心并非“把不同层级的变量放进同一个模型”,而是要建构“(宏观)结构-(微观)行动”之间的动态反馈与迭代模型,以真正检验和解释跨层级的因果机制。

1. 从单层网络到跨层嵌套结构的分析

无论是考察宏观结构如何约束微观行动,还是分析微观互动如何累积并生成新的宏观形态,关键都在于如何在同一分析框架中同时表示不同行动主体、制度要素与信息结构之间的依附与连接关系,以单层网络或单尺度变量为分析单位无法完整呈现其中跨层级因果关系的复杂性。因而,计算传播研究中需要引入跨层嵌套结构分析,为不同层级要素提供统一的关系化表达方式,使宏观与微观不再只是模型中的不同变量层次,而是同时参与结构分析与动态演化的网络组成部分,从而为情境机制与转化机制提供共享的跨层次结构分析方法。

社会复杂系统在本质上由多种类型的关系构成,这些关系在不同情境、时间尺度与传播渠道中共存并相互作用。早期关于意见传播与网络结构的计算研究,已经在单层网络内部尝试引入“嵌套”的思想。早期研究并未将网络视为静态背景,而是将意见状态与网络拓扑的相互作用纳入同一更新规则之中,将反馈机制纳入了计算过程,发现系统最终呈现出从多元共存到高度聚类乃至生成共识的变化过程^[59]。这说明多层动力学与其单层近似之间存在根本性的、不可通约的差异,凸显了在计算社会科学研究中,理解复杂社会现象时整合多层网络(multilayer networks),考虑这些精细而相互关联结构的必要性^[60]。

在此基础上,后续研究开始系统性地多层结构纳入计算分析,实现向跨层嵌套结构分析的方法跃迁。“时间”就被认为是过往研究常备忽略但十分重要的层级结构。斯塔尔尼尼(Michele Starnini)等人不再假定传播过程发生在均质时间轴上,而是通过多层马尔可夫链刻画不同时段、不同互动层之间的关联性,显示个体在多重社会场景中的“多任务化”行为会显著改

变传播持续时间^[61]。乌雷尼亚-卡里翁(Javier Ureña-Carrión)等人则进一步提出了一条更为可行的计算路径,先从接触数据中推断潜在的“社会时间”(如昼夜节律、“工作日-周末”差异等嵌套要素),再将其离散为多个时间层网络,进而比较关系在不同层中的分布与聚集特征。通过多层模块度和中心性分析,他们发现,关系往往更依附于其生成的时间情境,而并非仅由强度决定^[62]。

在层间互动的计算方法上,研究开始从网络动力学本身的内在结构进行考量,通过在双层网络中引入边重叠与节点跨层参与度等参数,以刻画层间互动规则,以此将“层间作用”转化为可被追踪的变量^{[63][64]}。博尼法齐(Gianluca Bonifazi)等人面向更复杂的多模态系统,提出通过跨层相关性与扰动后的结构恢复情况来衡量网络韧性,关注不同关系类型在冲击下的协同响应,其核心仍是比较各层之间变化的同步性与滞后性,来刻画多层结构在演化过程中的耦合关系^[65]。这些计算方法的进展为研究复杂传播现象提供了关键工具和启示,在跨层的统一框架想进行综合分析和解释,而不仅仅依赖单层网络或单一动力学机制进行近似描述。

2. 从单向生成到演化递归的过程分析

传统以参数设定为核心的ABM本质上遵循“单向生成”逻辑,研究者通过预定义的规则、效用函数或少量参数一次性刻画代理的决策机制,模拟运行后直接涌现宏观结果。这种以参数中心的“微观行动-宏观结构”的线性流程虽能高效模拟静态均衡或简单涌现结果(如舆论网络形成、意见极化),却难以进一步刻画宏观结果形成后的“行动-结构-再行动”的动态递归过程。行为体的行为结果锁定在初始参数中,无法通过长期记忆、反思或环境反馈实现自适应演化,宏观结构也难以反过来作为可感知的情境条件重新塑造个体决策。正如现有研究指出的,传统ABM多采用“反应式架构”(reactive architecture),支持“感知-行动”的直接循环,缺乏深度推理、长期规划与反思机制,导致适应性学习和演化能力严重受限;

而参数校准的复杂性与异质性进一步加剧了这一问题的,微小参数扰动就可能引发截然不同的涌现结果,难以自然捕捉真实世界中个体与结构的“共同演变”迭代过程^[66]。

引入基于大语言模型的ABM方法(LLM-based ABM)则可能适应对演化递归过程的研究,将大型语言模型作为代理的认知核心,通过提示工程、记忆模块、反思与规划循环,构建闭合的“行动-结构-再行动”反馈回路^[67]。如派克(Joon Park)等人进行的“斯坦福小镇”研究就引入了生成式智能体(Generative Agents)。这类智能体基于大语言模型(LLM)驱动的认知架构,集成了记忆、反思和规划三大核心功能,这三个递归模块的闭环交互使个体行动直接改变环境结构(如形成新关系、生成新舆论内容影响他人等),新结构又作为可检索记忆反馈给代理,触发新一轮反思与规划,从而完成“个体-结构”的共同演化,实现了传统参数ABM难以自发实现的递归涌现过程^[68]。并且这一架构具有极强的迁移性。在社交媒体语言演化的模拟中,研究者构建了一个由“参与者智能体”和“监管智能体”组成的对抗性模拟框架^[69]。在该研究中,参与者通过持续互动不断调整其语言表达策略,而监管机制则对其进行约束与筛选,在多轮对话与策略更新过程中,语言表达方式在“适应-反馈-再适应”的循环中逐步演化,构建了隐喻、黑话等复杂的语言形态,展现出制度约束与语言策略的长期互相适应过程。由此可见,LLM-based ABM的建模路径能够在多主体互动与规则约束的交互中,刻画行动策略的动态生成及其跨轮次积累,为分析复杂系统中的递归演化机制提供了更具解释力的工具。

当然,本文无意于构建一套封闭的“方法清单”,而是重点在科尔曼之舟的理论框架下,指出计算传播研究在面对复杂系统时所需实现的方法论转向及其内在逻辑。它们共同指向了一种本体论上的转向,不再将传播现象简化为静态变量之间的线性关系,而是将其理解为嵌入多主体互动、多层结构之中不断演化与迭代的复杂系统。通过引入这些计算路径,计算传

播研究有望在保持可计算性与可复现性的同时，更充分地回应传播研究中的复杂性问题，深化复杂性理论与传播学研究之间的对话。

四、认识论反思：溯因推理与科尔曼之舟间的有机互构

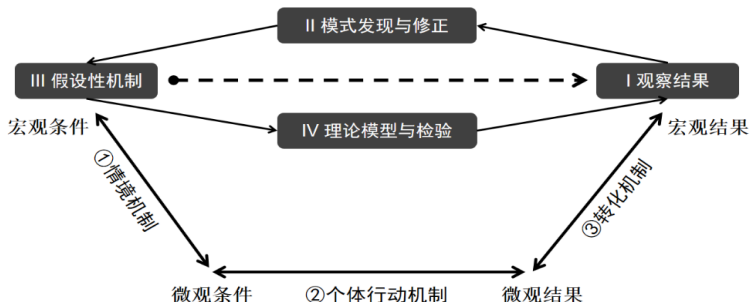
计算传播学研究常面临“只有计算，没有传播”的批评^[70]，其核心困境在于计算方法在传播学研究中的应用往往被局限于对既有理论的验证（演绎逻辑）或对数据表象的归纳（归纳逻辑），而缺乏对理论生成过程的贡献。在复杂系统视角下，传播现象往往表现为大量异质行动者在非线性交互中所产生的涌现结果，既难以由既有理论直接推出，也无法仅凭统计归纳加以穷尽，这使得研究不可避免地需要从结果出发反向构造解释性假设。在这一意义上，溯因推理（abductive reasoning）能够成为计算传播学的核心认识论基础。科尔曼之舟与溯因推理的有机融合，在于两者能够在认识论层面深层互构，如果说科尔曼之舟描述了社会宏观事实涌现与迭代的生成路径，那么，溯因推理则是这一过程进行反向机制推断的认识论工具。它始于对令人惊讶的宏观结果的观察，通过理论化的“猜测”回溯微观行动机制与情境条件，构成了连接经验发现与理论建构的关键认识论路径，从而实现理论的生成与拓展（图2）。

具体而言，在面对大数据的现实研究中，研究者往往首先面对的是难以用既有理论完全解释的、复杂的宏观涌现“观察结果”，如特定的舆论极化形态、异常的网络结构等。此时，

研究逻辑必须转向逆向的回溯和推理。正如科瓦奇（Gyöngyi Kovács）和斯彭思（Karen M. Spens）所指出的，溯因推理不同于从一般到特殊的演绎，也不同于从特殊到一般的归纳，它始于一个无法被现有理论解释的“异常现象”或“谜题”，旨在寻找一个最合理的解释框架来使这个现象变得合理^[71]。在此基础上，科尔曼之舟构成了溯因推理展开的操作路径。研究首先从“宏观结果”出发，回溯其潜在的“宏观条件”，但由于“宏观条件-宏观结果”的直接关系，往往呈现为经验上的黑箱，例如，下例中的中国风险投资产业的发展并不能直接推出“精英俱乐部型小世界网络”的结构形态。因此，必须借助跨层机制拆解，将这一黑箱展开为“情境-行动-转化”的跨层级因果链条。在这一过程中，假设性机制的提出并非任意猜测，而是围绕宏观条件如何约束微观行为展开进行的结构化推断。以中国风险投资产业（VC）为例，罗家德等人的研究首先通过网络分析识别出高度聚集且具“小世界”特征的精英俱乐部网络结构，这一模式无法被传统基于理性选择的投资决策理论充分解释。于是，研究者从网络科学的视角切入，提出在VC产业中存在偏好依附情境下“主投-跟投”的机制（情境机制），以及关系嵌入与结构嵌入两种行为模式（行动机制），再运用多主体建模进行推断，发现中国VC产业中嵌入行为与网络结构形成了正反馈，精英选择联盟伙伴时的行为促成了“精英俱乐部型小世界网络”的涌现（转化机制）^[72]。通过这一过程，原有投资行为理论被重新修正

为一个嵌入关系结构与制度情境的行为模型，即“关系”不再是附属变量，而成为生成宏观结构的关键机制变量。这一案例表明，溯因推理并不只停留于“提出解释”，而能够在结合科尔曼之舟的框架之后，将解释拆解为可建模、可检验的机制链条。

进一步而言，这一逻辑同样适用于计算传播研究。例如，在



社交媒体舆论网络中识别出异常的极化或群体分化结构后,研究者可以通过类似路径,将平台算法环境视为情境条件,将用户的选择性接触与互动偏好建模为行动机制,并通过仿真或因果推断检验其是否能够生成观测到的宏观结构。由此,计算方法不再只是用于描述模式,而是嵌入溯因推理过程,成为连接“异常发现”与“机制生成”的关键工具。

通过这种逆向的理论化过程,计算不再仅仅是处理数据的工具,而是通过“以果索因”发现新机制,形成实用主义导向下的最佳解释推论^[73]。当然,这一综合探索性研究与验证性研究的溯因推理过程并非一蹴而就的,而应当在“理论、数据挖掘和模型”的三角对话中交互进行。在“观察结果至假设性机制”的探索阶段,计算方法作为一种发现工具,可以通过大数据挖掘技术(如网络分析、自然语言处理)“深描”宏观现象与微观行为的特征,发现现有理论无法覆盖的变量关系。此时产出的不是最终结论,而是待验证的理论命题和假设机制。在“假设性机制至观察结果”的阶段,计算方法作为验证的手段,基于溯因得出的理论命题,构建因果模型、多主体仿真或进行大规模互联网实验,对推测的机制进行严格的假设检验^[74]。通过“探索-验证”这一溯因推理的过程,计算不再只是验证既有理论的工具,成为了“理论生成”与“实证检验”循环中的关键环节。

五、总结

本文以传播学研究的“复杂性转向”为问题起点,系统论述了计算传播研究在理论框架、方法路径和认识论基础上面向复杂系统的整体创新路径,并由此尝试推动计算传播从“以计算处理复杂数据”走向“以计算理解复杂性”的范式转型。在理论层面,本文引入科尔曼之舟作为跨层级机制分析的核心工具,将复杂传播现象的因果解释拆解为情境机制、个体行动机制与转化机制三个相互关联的链条,强调以方法论局部主义为基础,实现对结构约束与个体行动之间关系的过程性解释,从而弥补既有

研究中“局部深化而整体断裂”的问题。

在方法论层面,本文认为计算传播研究需由表征性分析走向生成性建模,并提出了双重方向的革新路径。在微观层级,研究可以通过情感弧追踪、语义漂移分析与算法代码内容分析等方法,从传播结果的静态表征走向对话语生成过程与行动机制的动态剖析,有效提升了个体行动机制层面的研究颗粒度。在跨层级结构上,研究需要从单层网络拓展至跨层嵌套网络分析,将时间层级、制度层级与互动层级纳入统一的关系化分析框架,为情境机制与转化机制的经验刻画提供了可操作的方法支撑。同时,计算传播研究也需要引入生成式主体建模方法,实现对复杂传播系统中“行动-结构-再行动”递归过程的动态分析。这些方法论革新共同指向一种本体论层面的转向,不再将传播现象视为变量间的静态线性关系,而是将其理解为嵌入多主体互动与多层结构之中不断迭代演化的复杂系统。

在认识论层面,本文将溯因推理确立为计算传播研究的重要逻辑基础,并与科尔曼之舟形成互构关系,前者提供从宏观结果出发的机制生成和理论发展路径,后者提供跨层级机制拆解的操作框架。溯因推理从令人惊讶的宏观涌现现象出发,回溯潜在的情境条件与行动机制,在“观察结果-假设机制-模型验证”的循环中推动理论的生成与修正。这一认识论路径使计算方法不再仅仅是验证既有理论的工具,而成为在“理论-数据-模型”三角对话中生成新机制解释的关键环节,在一定程度上回应了计算传播学长期面临的“只有计算,没有传播”的批评。

本文通过构建一个以跨层级机制解释为核心、以方法论革新为支撑、以溯因推理为驱动的整体性分析框架,尝试推动计算传播学从对复杂性的描述性应用,迈向以机制解释与理论生成为导向的系统研究路径。通过计算对复杂传播现象进行刻画和探索,该范式不仅强化了对复杂传播现象的因果理解,也推动计算传播学由“数据驱动”走向“机制驱动”,以在复



杂性理论与传播学经验研究之间建立更加紧密的互动关系，并尝试在复杂性转向的语境下重建传播学理论的解释力。

（责任编辑：李韵奕）

参考文献

- [1] Hoffman R. Exploring the Link Between Uncertainty and Organizing Processes: Complexity Science Insights for Communication Scholars[J]. *Communication Theory*, 2008, 18(3): 426-447.
- [2] McLean M J., Madden S., & Pressgrove G. Complexity Theory as a New Lens for Digital Social Advocacy[J]. *Public Relations Review*, 2021, 47(3): 102056.
- [3][5] 韩少卿, 巢乃鹏. 复杂性转向: 计算传播学的发生逻辑[J]. *南京社会科学*, 2024(3): 109-122.
- [4] 王成军. 计算传播学: 作为计算社会科学的传播学[J]. *中国网络传播研究*, 2014(0): 193-206.
- [6] 王大可, 李菁菁. 生成式 AI、多模态分析与算法权力: 2025 年国际计算传播研究的的知识结构与前沿热点[J/OL]. (2026-02-25) [2026-03-21]. 新媒体与社会, <https://link.cnki.net/urlid/CN.20260205.1103.002>.
- [7] 赵甜芳. 计算传播研究的知识地图与实证框架——基于 12991 篇中外核心文献的对比分析[J]. *新闻与传播研究*, 2023, 30(3): 40-59, 126-127.
- [8][13] 崔妮, 汤景泰. 论智能传播时代舆论系统的复杂性[J]. *当代传播*, 2025 (6): 4-12.
- [9] 肖亚龙, 陈奕丰, 朱承璋, 等. “三角结构”: “一带一路”中涉华科技议题的跨国媒体议程设置[J]. *新闻与传播研究*, 2025, 32(1): 100-113, 128.
- [10][37] 宣长春, 吴钧昊, 林婧. 复杂系统视角下全球媒体与智库涉“一带一路”的议题竞合和舆论协同(2014—2024)[J]. *国际新闻界*, 2025, 47(6): 80-96.
- [11] 汤景泰, 陈秋怡. 多主体协同与立体化级联: 论意见领袖的网络化作用模式——以俄乌冲突中的国际舆论战为例[J]. *新闻与传播研究*, 2025, 32(10): 19-35, 126.
- [12] Aleta A., & Moreno Y. Multilayer Networks in a Nutshell[J]. *Annual Review of Condensed Matter Physics*, 2019(10): 45-62.
- [14] 吕鹏. 计算社会科学中仿真模拟的三个发展阶段[J]. *清华社会学评论*, 2022 (1): 38-59.
- [15] Waldherr A., & Wettstein M. Bridging the Gaps: Using Agent-Based Modeling to Reconcile Data and Theory in Computational Communication Science[J]. *International Journal of Communication*, 2019, 13: 3976-3999.
- [16][34] Elsenbroich C., & Polhill J G. Agent-Based Modelling as a Method for Prediction in Complex Social Systems[J]. *International Journal of Social Research Methodology*, 2023, 26(2): 133-142.
- [17] Shi W., Zhao H., Huang Y., et al. Agent-Based Approaches in Studying Algorithm-Mediated Communication: A Methodological Review[J]. *Communication and Change*, 2026, 2(1): 2.
- [18] 王晗啸, 张楚惠. “双重意见气候”下社交机器人舆论干预影响研究——基于 ABM 仿真模拟沉默螺旋效应[J]. *新闻大学*, 2022(12): 75-90, 124.
- [19] 黄丽芹, 孙寅, 罗思阳. 个人主义文化价值观对疫情控制效果的影响及其计算心理机制[J]. *心理学报*, 2022, 54(5): 497-533.
- [20] 赵汗青, 葛岩, 印闯, 等. 疫情信息怎样影响疫情演化: 一个计算实验模型[J]. *国际新闻界*, 2024, 46(5): 133-159.
- [21] 吕鹏, 李蒙迪, 张卓. 基于沙堆模型的国家动力学研究——以中国历史为例(前 221 年到 1912 年)[J]. *社会研究方法评论*, 2022, 2(2): 43-79.
- [22] 王泽柳, 范冬萍. 复杂系统管理中的功能主义思想[J]. *自然辩证法研究*, 2013, 29(7): 61-66.
- [23] 张珍. 实现整体主义的思想变迁及其现代特征[J]. *自然辩证法研究*, 2014, 30(9): 47-51.
- [24] Manzo G. Variables, Mechanisms, and Simulations: Can the Three Methods be Synthesized?: A Critical Analysis of the Literature[J]. *Revue Française de Sociologie*, 2005, 46(1): 37-74.
- [25] Little D. Levels of the Social[M]//Turner S P. & Risjord M W., eds. *Handbook of the Philosophy of Science, Philosophy of Anthropology and Sociology*. Amsterdam: North-Holland, 2007: 360.
- [26][29][33] 吴畏, 石敬琳. “科尔曼船”及其变形与社会科学的说明逻辑[J]. *自然辩证法通讯*, 2022, 44(1): 20-32.
- [27][58] Weber M S. & Kosterich A. Coding the News: The Role of Computer Code in Filtering and Distributing News[J]. *Digital Journalism*, 2018, 6(3): 310-329.
- [28] 陶冶, 唐超. 算法操纵下的认知对立: 形成机理与应对策略[J]. *情报杂志*, 2025, 44(8): 94-100.
- [30] Hedström P. & Swedberg R. Social Mechanisms[J]. *Acta Sociologica*, 1996(39): 281-308.
- [31] Schelling T C. Dynamic Models of Segregation[J]. *The Journal of Mathematical Sociology*, 1971, 1(2): 143-186.
- [32] Epstein J M. *Generative Social Science: Studies in Agent-Based Computational Modeling*[M]. Princeton University Press, 2006: 12.
- [35] Hassan S., Arroyo J., Galán Ordax J M., et al. Asking the Rracle: Introducing Forecasting Principles into Agent-Based Modelling[J]. *Journal of Artificial Societies and Social Simulation*, 2013, 16(3): 1-17.
- [36] Debutts M., & Pan J. Reporting after Removal: The Effects of Journalist Expulsion on Foreign News Coverage[J]. *Journal of Communication*, 2024, 74(4): 273-286.
- [38] Ylikoski P. Understanding the Coleman Boat[M]//Manzo G., ed. *Research Handbook on Analytical Sociology*. Cheltenham: Edward Elgar Publishing, 2021: 49-63.
- [39] 胡安宁, 陈滔, 李东雨. 从定量社会学到计算社会学: 传统进路与新兴议题[J]. *西安交通大学学报(社会科学版)*, 2022, 42(1): 17-24.
- [40] Burns T., & Roszkowska E. Rational Choice Theory: Toward a Psychological, Social, and Material Contextualization of Human Choice Behavior[J]. *Theoretical Economics Letters*, 2016, 6(2): 195.

- [41]Granovetter M. Threshold Models of Collective Behavior[J]. *American Journal of Sociology*, 1978, 83(6): 1420–1443.
- [42] 曾荣光, 罗云, 叶菊艳. 寻找实证研究的意义: 比较—历史视域中的实证主义之争[J]. *北京大学教育评论*, 2018, 16(3): 104–131.
- [43] 杨晓光, 高自友, 盛昭瀚, 等. 复杂系统管理是中国特色管理学体系的重要组成部分[J]. *管理世界*, 2022, 38(10): 1–24.
- [44] 朱晓红, 陈寒松, 张腾. 知识经济背景下平台型企业构建过程中的迭代创新模式——基于动态能力视角的双案例研究[J]. *管理世界*, 2019, 35(3): 142–156.
- [45] 党耀国, 尚中举, 王俊杰, 等. 基于面板数据的灰色指标关联模型构建及其应用[J]. *控制与决策*, 2019, 34(5): 1077–1084.
- [46][72] 罗家德, 万怡, 高馨. 复杂社会系统的涌现现象——以大数据分析结构与行动共同演化[J]. *社会学研究*, 2024, 39(2): 134–155.
- [47]Watts D J., & Dodds P S. Influentials, Networks, and Public Opinion Formation[J]. *Journal of Consumer Research*, 2007, 34(4): 441–458.
- [48]Del Vicario M., Bessi A., Zollo F., et al. The Spreading of Misinformation Online[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(3): 554–559.
- [49]Centola D. The Spread of Behavior in an Online Social Network Experiment[J]. *Science*, 2010, 329(5996): 1194–1197.
- [50]Pogson M., & Nicholls T. Agent-Based Modeling of Diversity, New Information and Minority Groups in Opinion Formation[J]. *Computational Communication Research*, 2024, 6(1): 1.
- [51] 赵甜芳, 王语. 多平台过滤气泡的感知与抵抗: 基于 Z 世代大学生数字社群的研究[J]. *中北大学学报(社会科学版)*, 2026(1): 87–95.
- [52] 王晗啸, 张楚惠. “双重意见气候”下社交机器人舆论干预影响研究——基于 ABM 仿真模拟沉默螺旋效应[J]. *新闻大学*, 2022(12): 75–90, 124.
- [53]Reagan A J., Mitchell L., Kiley D., et al. The Emotional Arcs of Stories are Dominated by Six Basic Shapes[J]. *EPJ Data Science*, 2016, 5(1): 1–12.
- [54]Berger J., Kim Y D. & Meyer R. What Makes Content Engaging? How Emotional Dynamics Shape Success[J]. *Journal of Consumer Research*, 2021, 48(2): 235–250.
- [55]Rodriguez P L., Spiraling A. & Stewart B M. Embedding Regression: Models for Context-Specific Description and Inference[J]. *American Political Science Review*, 2023, 117(4): 1255–1274.
- [56]Kamath G., Yang M., Reddy S., et al. Semantic Change in Adults is not Primarily a Generational Phenomenon[J]. *Proceedings of the National Academy of Sciences*, 2025, 122(31): e2426815122.
- [57] 宣长春, 李嘉鑫. “债务陷阱论”话语的生产与演化——基于计算方法的分析[J]. *新闻与传播研究*, 2025, 32(12): 5–20, 121.
- [59]Holme P. & Newman M E J. Nonequilibrium Phase Transition in the Coevolution of Networks and Opinions[J]. *Physical Review E*, 2006, 74(5): 056108.
- [60]Aleta A., Teixeira A S., de Arruda G F., et al. Multilayer Network Science: Theory, Methods, and Applications[A/OL]. (2025–11–28) [2025–12–15]. arXiv, <http://arxiv.org/abs/2511.23371>. DOI:10.48550/arXiv.2511.23371.
- [61]Starnini M., Baronchelli A., & Pastor-Satorras R. Effects of Temporal Correlations in Social Multiplex Networks[J]. *Scientific Reports*, 2017, 7(1): 8597.
- [62]Ureña-Carrion J., Heydari S., Aledavood T., et al. Multiplexity is Temporal: Effects of Social Times on Network Structure[A/OL]. (2024–07–08) [2025–12–15]. arXiv, <http://arxiv.org/abs/2407.05929>. DOI:10.48550/arXiv.2407.05929.
- [63]Cranmer S J., Menninga E J. & Mucha P J. Kantian Fractionalization Predicts the Conflict Propensity of the International System[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(38): 11812–11816.
- [64]Diakonova M., Nicosia V., Latora V., et al. Irreducibility of Multilayer Network Dynamics: The Case of the Voter Model[J]. *New Journal of Physics*, 2016, 18(2): 023010.
- [65]Bonifazi G., Cauteruccio F., Corradini E., et al. A Multilayer Network-Based Framework for Investigating the Evolution and Resilience of Multimodal Social Networks[J]. *Social Network Analysis and Mining*, 2023, 14(1): 5.
- [66]Gao C., Lan X., Li N., et al. Large Language Models Empowered Agent-Based Modeling and Simulation: A Survey and Perspectives[J]. *Humanities and Social Sciences Communications*, 2024, 11(1): 1259.
- [67]Vanhée L., Borit M., Siebers P. O., et al. Large Language Models for Agent-Based Modelling: Current and Possible Uses across the Modelling Cycle[A/OL]. (2025–07–08) [2026–03–23] arXiv, <http://arxiv.org/abs/2507.05723>.
- [68]Park J S., O’ Brien J., Cai C J., et al. Generative Agents: Interactive Simulacra of Human Behavior[C]// *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 2023: 1–22.
- [69] Cai J., Ishimizu Y., Zhang M., et al. Simulation of Language Evolution under Regulated Social Media Platforms: A Synergistic Approach of Large Language Models and Genetic Algorithms[A/OL]. (2025–02–26) [2026–01–28]. arXiv, <http://arxiv.org/abs/2502.19193>.
- [70] 钟智锦, 周金金, 苏思妮. 中英文计算传播研究的比较: 议题、方法与理论[J]. *全球传媒学刊*, 2023, 10(1): 56–76.
- [71]Kovács G. & Spens K M. Abductive Reasoning in Logistics Research[J]. *International Journal of Physical Distribution & Logistics Management*, 2005, 35(2): 132–144.
- [73]Mitchell A. A Review of Mixed Methods, Pragmatism and Abduction Techniques[C]// *Proceedings of the European Conference on Research Methods for Business & Management Studies*. 2018: 269–277.
- [74] 周涛, 高馨, 罗家德. 社会计算驱动的社会科学研究方法[J]. *社会学研究*, 2022, 37(5): 130–155, 228–229.



Confronting Complex Systems: Path Innovation and Methodological Prospects in Computational Communication Research

Xuan Changchun, Li Jiaxin

Abstract: Under the background of the complexity turn in communication studies, computational communication research has advanced rapidly in data processing and methodological applications. However, it largely remains oriented toward “computing complex data” rather than “understanding complexity through computation”. Grounded in complex systems theory, we introduce Coleman’s boat as a cross-level analytical framework linking the macro and micro. It is argued that by organically integrating situational mechanisms, individual action mechanisms, and transformational mechanisms, macro-level communication phenomena can be locally reduced to the generative and interactive processes of micro-level actors operating under macro conditions. At the methodological level, we propose a dual agenda for innovation: First, strengthening process-oriented analysis at the micro level, shifting from outcome representation to a deeper examination of action-generating mechanisms; Second, cross-level bridging through the development of multilayer network approaches and recursive simulation methods, in order to capture the dynamic evolution of “action-structure-(re)action” feedback loops. Abductive reasoning should be positioned as the core epistemology of computational communication. In conjunction with Coleman’s Boat, it enables a theoretically generative pathway that infers underlying mechanisms retrospectively from observed macro-level outcomes. In this sense, computation is no longer merely a methodological tool but an epistemic device for exploring the mechanisms underlying macro-level phenomena, partially resolving the critique that computational communication research is “computation without theory”, and thereby restoring the explanatory power of communication theory under the complexity turn.

Keywords: computational methods; computational communication research; complex systems; Coleman’s boat; abductive reasoning

Authors: Xuan Changchun, School of Journalism and Communication, Xiamen University, Center for International Communication Studies of Xiamen University; Li Jiaxin, School of Journalism and Communication, Xiamen University.

(上接第 70 页)

Youth Digital Detox Behavior: An Adaptive Response— A Dual Perspective of Extended Cognition and Cognitive Dissonance

Jiao Bao, Zhao Tian

Abstract: From the dual perspective of extended cognition and cognitive dissonance, this study employs structural equation modeling (SEM) and fuzzy-set qualitative comparative analysis (fsQCA) to examine the influencing mechanisms and configurational pathways of youth digital detox behavior. The findings show that youth digital detox behavior is jointly shaped by negative digital experience, positive substitution expectations, and digital cognitive management mechanisms. Specifically, perceived digital overload, digital anxiety, and the negative effects of digital dependence promote digital detox behavior by strengthening negative digital experience; perceived detox benefits, the attractiveness of alternative activities, and the convenience of low-dependence environments facilitate digital detox behavior by enhancing positive substitution expectations; self-control ability and social support strengthen inhibitory regulatory effects, whereas digital emotional attachment strengthens facilitative regulatory effects. The fsQCA results further indicate that neither high nor non-high levels of digital detox behavior are driven by any single necessary condition; rather, both are produced by the joint effects of multiple conditions. The study suggests that youth digital detox behavior should be understood not as a simple rejection of digital media, but as an adaptive response to imbalanced digital use.

Keywords: youth; digital detox behavior; extended cognition; cognitive dissonance; SEM-fsQCA

Authors: Jiao Bao, School of Journalism and Communication, Jilin University; Zhao Tian, School of Journalism and Communication, Jilin 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