证据缺口下生成式 AI 学术信息的社群传播与协作核验机制

——基于多 Agent 模拟、主题建模与扎根分析的混合证据研究

Paper Director 官方体验案例

（研究数据来自智研网真实案例快照；Paper 流程记录为官方预置演示）

摘 要：生成式人工智能进入科研检索、引用核验与知识生产后，信息效率提升与证据可靠性之间形成新的张力。本文围绕一条“生成式 AI 辅助文献检索和引用核验准确率约 94%”但缺少可验证原始来源的争议信息，构建由多 Agent 社会模拟、BERTopic 主题建模、情感分类质量审计与扎根理论分析组成的混合证据链，考察证据缺口如何转化为社群中的提问、回应、信息分享与协作核验行为。研究基于 12 个 Agent、4 类研究群体、2 个实验臂和 10 轮互动，共形成 240 次决策机会，其中 222 条产生可观测文本表达。正式纠偏信息在第 5 轮进入干预臂，12 名成员均具备接触资格，实际有 6 名成员注意到该信息。BERTopic 在 222 条文本中识别出 29 个实质主题和 1 个离群类别，最大主题仅占 7.2%，显示讨论并未收敛为单一议题。扎根分析在 81 个意义单元上形成 67 个最终编码、316 条编码引用和 11 条主轴关系，核心范畴为“证据缺口下的社群协作核验与审慎采纳”。结果表明，证据缺口并不简单导致拒绝或接受，而是通过来源追溯、方法学证据要求、核验责任明确与社群协作形成阶段性的审慎采纳机制。需要强调的是，两实验臂在正式干预前已存在随机轨迹差异，因而本文仅报告描述性差异与干预关联，不作严格因果推断；情感修复运行亦触发类别分布集中预警，仅作为质量控制证据。本文从“可信度判断”推进到“核验行动组织”，为生成式 AI 学术信息治理与科研支持平台的透明分析提供了可复核的机制视角。

关键词：生成式人工智能；学术信息核验；多 Agent 模拟；BERTopic；扎根理论；协作核验

0 引言

生成式人工智能正在从文本生成工具转变为科研活动中的基础性信息接口。研究者使用大模型完成关键词扩展、文献定位、摘要比较、引用核验和方法解释，科研信息的获得速度由此显著提升，但信息是否具有可追溯来源、是否经过同行评议、评价指标是否可复现等问题并不会因为交互效率提升而自动消失。当具体数字、准确率或“经验性结论”在社群中被快速转述时，研究者面对的往往不是简单的真伪二分，而是一个证据尚未闭合的判断场景。

已有扩散研究强调网络关系、重复暴露与群体结构对信息传播的影响[1-3]，可信度研究则关注来源、内容特征与个体判断之间的关系。然而，在学术信息场景中，来源缺失会进一步触发检索、质疑、同行协作、责任划分和制度性纠偏等行动。换言之，研究问题不能只停留于“成员是否相信一条信息”，还需要回答“当证据不足时，社群如何组织核验行动，以及正式纠偏如何进入既有互动结构”。

本文以一条声称“生成式 AI 辅助文献检索和引用核验准确率约 94%且人工复核提升有限”的疑似错误学术信息为实验议题，构建多 Agent 在线研究社群，并将可观测互动文本进一步送入主题建模与扎根分析。研究的目标不是证明某个干预具有严格因果效应，而是利用可追溯的混合证据描绘证据缺口下的互动结构、议题分化和核验机制。

1 理论框架与研究问题

1.1 证据缺口与学术信息可信度

学术信息可信度具有较强的程序性。对于包含具体准确率、效果量或性能比较的主张，仅有内容表面的合理性并不足以支撑研究决策；原始出处、样本范围、评价指标、数据生成过程和同行评议状态共同构成最低证据链。本文将“能够被传播但尚不能被可靠验证”的状态概括为证据缺口。证据缺口并不等同于信息错误，它描述的是当前可见材料不足以完成验证。

1.2 社群传播与协作核验

在线研究社群中的核验具有分布式特征。不同成员拥有不同的信息检索能力、规范意识、工具经验和关系位置。当单个成员无法完成溯源时，提问、分享检索路径、请求方法学细节和转发正式核验信息会把个体判断转化为集体行动。由此，核验机制至少包含三个层次：个体层面的风险感知与行动选择、关系层面的提问—回应—传播，以及制度层面的正式纠偏与责任归属。

1.3 研究问题

据此提出三个研究问题：RQ1，证据缺口出现后，不同研究群体表现出怎样的可观测互动行为结构？RQ2，讨论议题是否围绕来源核验、方法学证据与责任归属形成稳定的主题簇？RQ3，多轮互动中哪些编码与主轴关系能够解释“从证据缺口到协作核验再到审慎采纳”的机制链？

2 研究设计与数据

2.1 多 Agent 社会模拟

实验设置 12 个 Agent，分为 AI 高频科研使用者、普通青年研究者、方法与学术规范关注者、信息检索与学术信息专业人员四类群体，每类 3 人。实验包含对照臂与正式纠偏干预臂，各运行 10 轮，共形成 240 次决策机会。Agent 可选择公开表态、提问、回应、信息分享或保持沉默。最终 222 次决策形成非空文本，18 次保持静默。两臂各包含同一组角色原型，但由于生成式 Agent 以非零温度独立运行，正式干预前已经出现随机轨迹差异，因此后续比较只解释为描述性差异与干预关联。

2.2 正式纠偏与关系暴露

第 5 轮向干预臂发布学术信息服务机构的正式核验与纠偏信息。12 名 Agent 均满足接触资格，其个体暴露概率约为 0.64-0.82，实际 6 名 Agent 记录为“noticed”。全实验保留 90 条社会关系边和 12960 条暴露记录，使成员间接触、可见性和信任关系可以被回溯。这里的 50%注意比例只描述该次模拟运行，不外推为现实群体中的一般效应。

2.3 跨工具混合证据链

社会模拟的 222 条非空表达被作为统一语料进入 BERTopic 主题建模、情感分析和扎根分析。BERTopic 保留不同主体、轮次与实验臂中的社会性重复表达，以避免把群体传播本身误删为技术重复。扎根分析仅使用公开与定向表达，不纳入内部推理文本，共切分为 81 个意义单元。情感分析执行两次算法运行，并保留质量告警用于审计。

2.4 证据边界

本文严格区分“可观测表达”和内部状态。由于当前模拟版本中的部分数值状态没有形成可信的动态更新，本文不使用私人立场、风险感知等状态变量绘制演化趋势，也不以其支持因果结论。主要证据来自可观测行为、文本、关系暴露、主题归属和扎根编码。

表 1 研究设计与数据结构

维度	设置	规模/结果	证据用途
主体	4 类研究群体	12 Agent	异质互动主体
实验结构	对照臂/干预臂	2 臂×10 轮	描述性比较

互动	公开表态/提问/回应/分享/沉默	240 次决策；222 条文本	行为与文本证据
关系	信任/影响/接触/可见性	90 条关系；12960 条暴露	传播结构
正式纠偏	第 5 轮机构核验信息	12 人具备资格；6 人注意	干预关联
主题建模	BERTopic	29 实质主题+5 离群文本	议题结构
扎根分析	开放编码—主轴关系	81 单元；67 编码；316 coding	机制建构

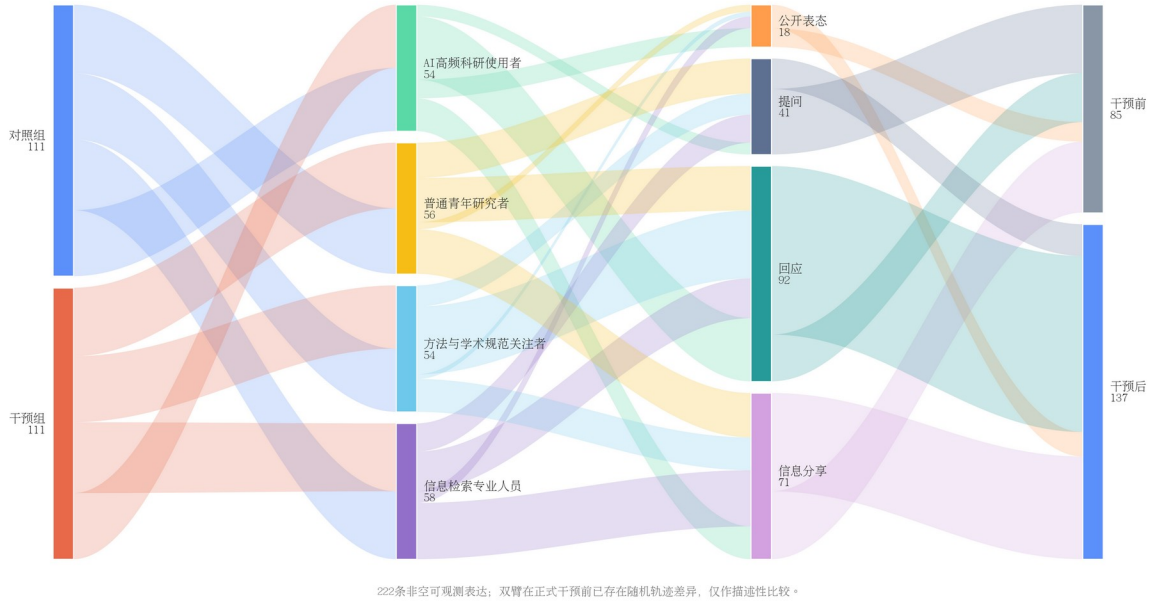

图 1 双实验臂—群体—行为—阶段 Alluvial 流向图

3 可观测行为与议题演化

3.1 行为结构与正式纠偏触达

222 条可观测表达中，回应与信息分享构成主要互动形式。干预臂在干预后出现 31 次回应和 23 次信息分享，对照臂同期分别为 41 次和 19 次；干预臂公开表态由干预前 6 次增至干预后 9 次，对照臂则由 2 次降至 1 次。由于两臂在前 4 轮已经存在行为数量差异，这些变化不能被解释为净处理效应。更稳妥的理解是：正式纠偏被嵌入一个已经分化的互动网络，部分成员注意到纠偏信息并继续通过回应、分享与提问参与核验。

3.2 主题结构：从单一争议到多路径核验

BERTopic 在 222 条文本中形成 30 个主题记录，其中 29 个为实质主题，5 条文本被标记为离群。最大主题“核验与社群”包含 16 条文本，占 7.2%；随后是“渠道与关于”“原始与没有”“检索与清单”等主题。最大主题占比低意味着讨论并未被一个单一话题完全主导，而是围绕来源追溯、渠道识别、检索清单、引用核验、编辑责任等多个相互关联的问题展开。

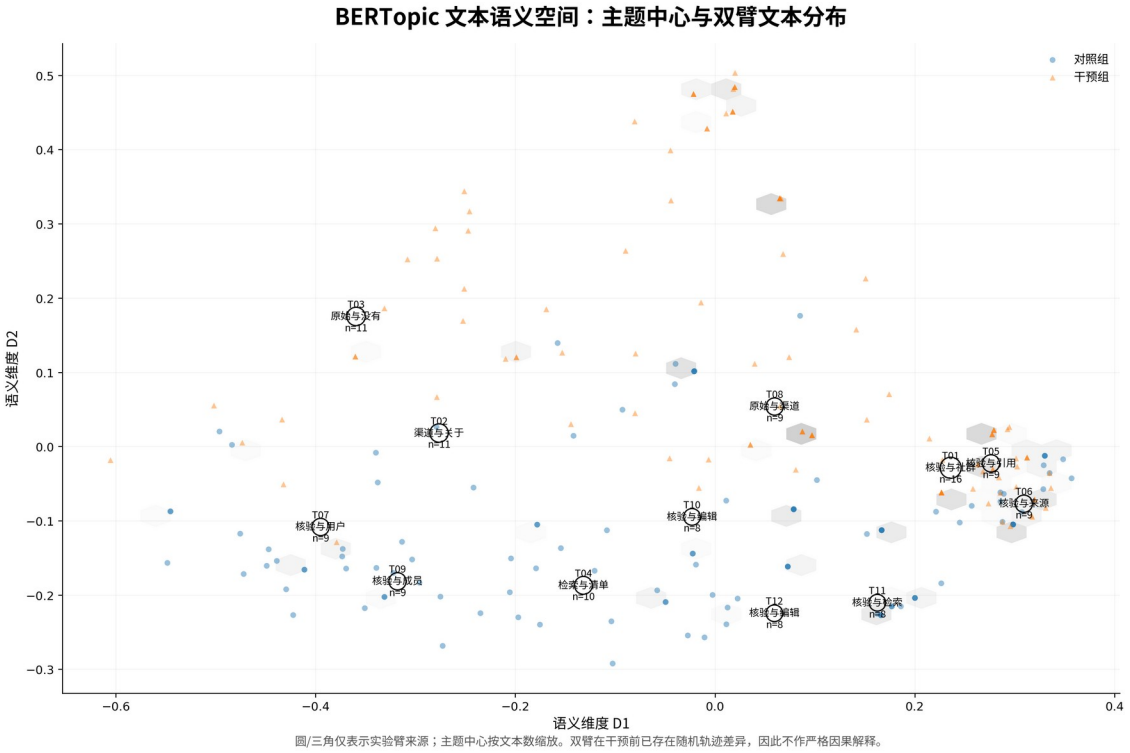

图 2 BERTopic 语义空间与双臂分布图

表 2 BERTopic 规模最大的 10 个主题

主题号	标签	文本数	平均归属得分
0	核验与社群	16	0.614
1	渠道与关于	11	1.000
2	原始与没有	11	0.640
3	检索与清单	10	0.540
4	核验与引用	9	0.636
5	核验与来源	9	0.831
7	原始与渠道	9	0.642
6	核验与用户	9	0.691
8	核验与成员	9	0.744

9	核验与编辑	8	0.604

3.3 情感分类作为质量敏感性证据

情感初始运行将 84.2%的文本标为负面，平均置信度 0.657，并产生 40 条低置信度记录。修复运行的平均置信度提高到 0.869，但负面占比上升至 96.8%，同时系统触发“label_distribution_collapse”质量预警。这一结果说明，在以质疑、核验和风险讨论为主的科研语料中，情感标签容易把“审慎质疑”压缩为“负面”。因此本文不把 96.8%作为研究发现，而把它视为分析工具透明披露不确定性的例证。

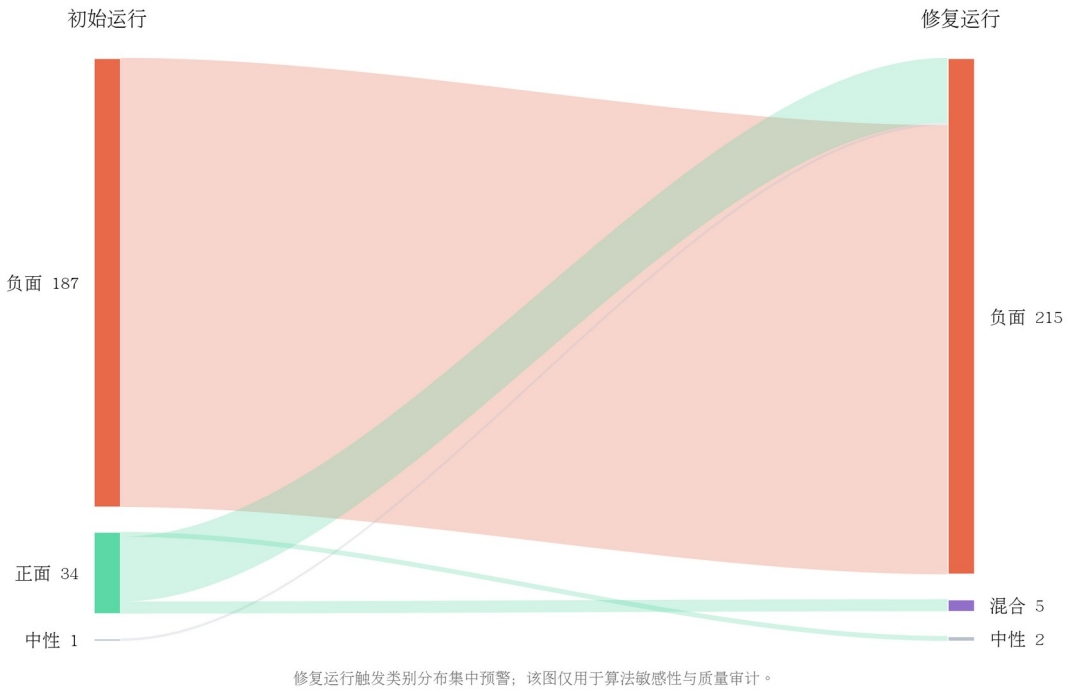

图 3 情感两轮分类迁移图（质量审计）

4 扎根分析与机制建构

4.1 开放编码与范畴结构

扎根分析在 81 个意义单元上形成 67 个最终 active code 和 316 条 coding，并进一步组织为 12 个范畴。频次最高的编码包括“指出信息缺乏可验证来源”（34 次）、“提出系统化的文献检索与同行评议核验方法”（32 次）、“倡导社群共同检索原始研究”（21 次）、“要求具体方法学证据作为评估前提”（20 次）、“建议制定标准化核验清单”（17 次）、“明确核验责任主体”（16

次）以及“主张来源未验证前保持人工核验强度”（15 次）。这些编码共同显示，社群注意力并未停留在真假判断，而是迅速进入“如何核验、由谁核验、以什么证据核验”的程序性问题。

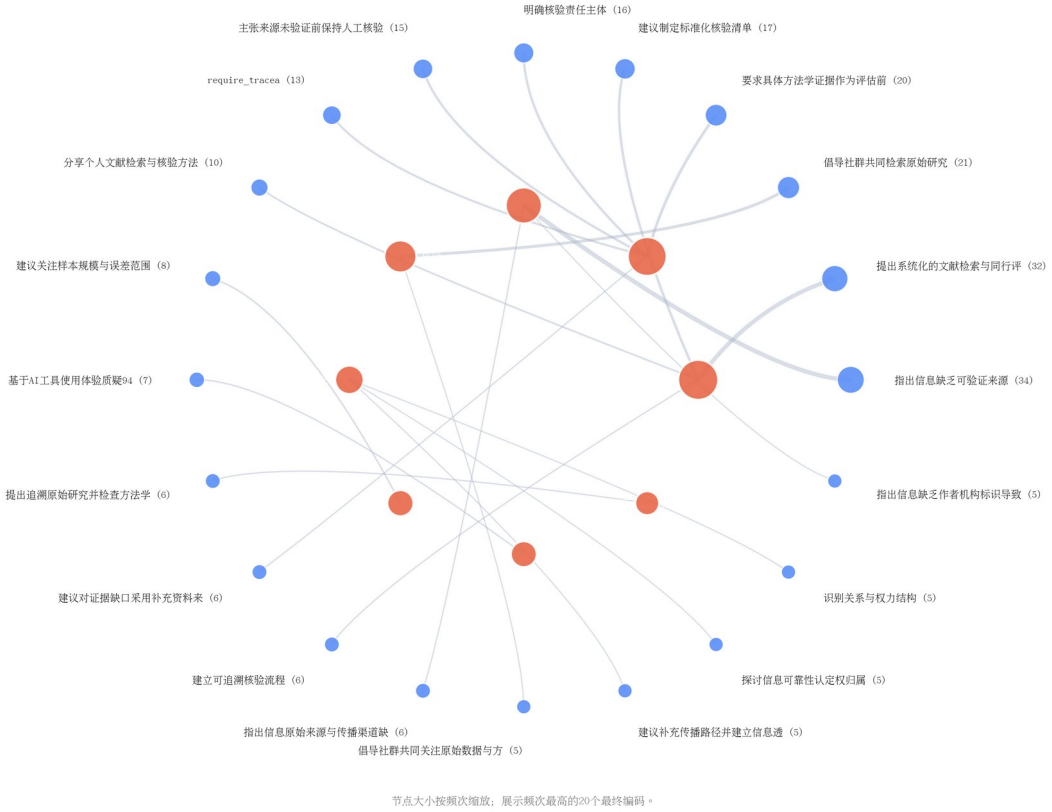

图 4 扎根范畴—高频编码环形关系图

表 3 频次最高的 10 个最终编码

编码	频次	所属范畴	定义摘要
指出信息缺乏可验证来源	34	指出信息缺乏可验证性	指出该 AI 科研信息声称 94%准确率，但原始出处和研究方法细节
提出系统化的文献检索与同行评议核验方法	32	提出系统化核验方法	提出通过多个学术数据库和引文索引使用关键词组合检索，限定时间
倡导社群共同检索原始研究	21	倡导社群协作核验	建议社群成员共同寻找原始研究或评测报告，分享核验经验，并询问
要求具体方法学证据作为评估前提	20	要求证据与责任明确	提出需要具体的测试集、样本量、评估指标和误差分析等证据，才能
建议制定标准化核验清单	17	提出系统化核验方法	建议共同拟定标准化核验清单，包括原始来源检索、方法学细节核查
明确核验责任主体	16	要求证据与责任明确	提出由编辑或指定人员牵头核验，社群成员协助提供线索，以建立明
主张来源未验证前保持人工核验强度	15	要求证据与责任明确	主张在信息来源验证之前，不轻易降低人工引用核验强度。

require_traceable_source_details	13	要求证据与责任明确	要求用户提供原始出处或可访问的方法学细节，否则只能视为未经验
分享个人文献检索与核验方法	10	提出系统化核验方法	分享自己常用的核验做法：通过 Google Scholar 或 P
建议关注样本规模与误差范围	8	确认资料缺失情况	建议社群成员在引用前核查原始出处，并关注样本规模、测试集构成

4.2 主轴关系与核心范畴

11 条主轴关系将分散编码连接为较清晰的机制路径。第一类关系由“缺少原始出处和方法学细节”指向“数据库检索、同行评议核验与方法学检查”；第二类关系由“缺乏标准化流程和责任主体”指向“制定核验清单并明确编辑或平台牵头”；第三类关系由“个人 AI 使用经验与幻觉引用”指向“在来源未验证前维持人工核验强度”；第四类关系由“正式机构纠偏信息出现”指向“增强未验证判断，但仍要求继续追溯原始研究”。综合这些关系，核心范畴被概括为“证据缺口下的社群协作核验与审慎采纳”。

4.3 机制链：证据缺口—协作核验—责任明确—审慎采纳

机制链可概括为四个连续但可反复的环节。其一，具体数字主张缺少可追溯依据，触发证据缺口识别；其二，成员通过提问、分享检索路径和方法学标准组织协作核验；其三，当个体核验无法闭合时，讨论转向责任主体、平台记录与正式纠偏机制；其四，在证据仍不充分的情况下，社群采取“暂不降低人工核验强度”的审慎采纳策略。正式纠偏可以降低部分不确定性，但不能替代原始研究和方法学证据。

5 讨论

5.1 从可信度判断到核验行动组织

本文的主要启示是，生成式 AI 学术信息治理不能只依赖一个“可信/不可信”的终局标签。面对证据缺口，研究社群会产生一系列中间行动：要求出处、提出检索策略、比较个人经验、寻找正式机构意见、讨论责任归属。平台若只提供最终分类而不保留这些过程信息，反而会丢失最具有治理价值的机制证据。

5.2 正式纠偏不是单次信息注入

第 5 轮正式纠偏只有一半具备资格的成员实际记录为注意到该信息，说明制度性信息进入社群仍受可见性、接触概率和既有互动结构约束。更重要的是，纠偏并没有终止讨论，而是被进一步转发、解释和要求核对原始出处。这意味着纠偏机制的作用可能不是“直接改变立场”，而是为既有核验流程提供更高权重的证据节点。

5.3 方法论意义：跨工具结果必须保留可追溯性

本研究将同一批 222 条文本贯穿主题建模、情感分析和扎根分析，主题归属能够逐条回到原始 unit_id，情感元数据可以从 raw 字段恢复，扎根 coding 则对能够唯一匹配的证据片段进行回溯。跨工具分析由此不再是若干互不相干的截图，而形成从原始表达、计算结果到理论编码的证据链。与此同时，无法确定归属的扎根 coding 不被强行映射，理论饱和阈值未达到也不被包装为“已经饱和”，体现了结果透明性。

6 结论、局限与展望

本文围绕证据不足的生成式 AI 学术信息，利用多 Agent 社会模拟和跨工具混合分析描绘了社群核验机制。研究发现可以归纳为三点：第一，证据缺口会激活提问、回应、分享和公开表态等多种互动，而不是简单导致接受或拒绝；第二，主题结构持续围绕来源追溯、方法学证据、检索路径与责任归属展开，显示核验具有明显的程序性；第三，扎根理论将这些行为整合为“证据缺口下的社群协作核验与审慎采纳”，其核心机制是由证据识别推动协作检索，再由责任与制度性信息支撑审慎决策。

研究存在三方面局限。其一，两实验臂在正式干预前已经出现随机轨迹差异，当前单次运行不适合支持严格因果处理效应。其二，部分模拟数值状态没有形成可信的动态更新，本文因此只使用可观测行为和文本证据。其三，扎根分析未达到程序预设的理论饱和阈值，且没有识别出负面案例，理论结构仍需更多异质材料检验。

后续研究可在固定随机种子与配对初始状态基础上增加重复实验，分离自然随机差异与正式纠偏关联；同时将真实研究社群调查、访谈与平台日志引入证据图，检验协作核验机制在不同学科和不同权威结构中的适用边界。对于科研支持平台而言，更具实践价值的方向是把“来源状态、验证过程、责任主体和质量告警”作为一等数据对象，而不是仅输出单一概率或情感标签。

正式纠偏注意率：$R_{noticed} = N_{noticed} / N_{eligible} = 6 / 12 = 0.50$

参考文献

[1] Granovetter M S. The strength of weak ties[J]. American Journal of Sociology, 1973, 78(6): 1360-1380.

[2] Centola D. The spread of behavior in an online social network experiment[J]. Science, 2010, 329(5996): 1194-1197.

[3] Vosoughi S, Roy D, Aral S. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.

[4] Rogers E M. Diffusion of Innovations[M]. 5th ed. New York: Free Press, 2003.

[5] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure[EB/OL]. arXiv:2203.05794, 2022.

[6] Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory[M]. 2nd ed. Thousand Oaks: Sage, 1998.