学术探索
Academic Exploration
ISSN 1006-723X,CN 53-1148/C

《学术探索》网络首发论文

题目：从“效率独裁”到“关系共生”：AI 时代传播伦理的范式重构与治理路径
作者：喻国明
网络首发日期：2026-06-05
引用格式：喻国明．从“效率独裁”到“关系共生”：AI 时代传播伦理的范式重构与治理路径[J/OL]．学术探索．https://link.cnki.net/urlid/53.1148.C.20260605.0943.002

从“效率独裁”到“关系共生”：AI 时代传播伦理的范式重构与治理路径

喻国明

（北京师范大学　新闻传播学院，北京　100875）

摘要：面对生成式 AI 引发的“巴别塔陷阱”（效率至上与价值迷失），传播学亟需一场“人性保卫战”。本文旨在打破传统“工具理性”的桎梏，探索 AI 时代传播伦理的重构路径。论文首先解构了“工具中立”的迷思，揭示了算法推荐背后的“价值嵌入”本质与“双重异化”危机（技术拟人化与人类去人性化）。在此基础上，引入“关系范式”作为核心分析框架。文章提出，AI 时代的传播伦理不应是对技术的简单规制，而应是基于“人性温度”的文明重构。通过构建“耶路撒冷方案”，确立“人是目的”的伦理原点。本文创新性地提出了“五大伦理支柱”（尊严、真理、劳动、自由、包容）与“悟空机制”（悬荡-协商-悟空的动态治理），试图在算法逻辑与人文精神之间建立动态平衡，为全球传播治理提供中国方案。

关键词：AI 传播伦理；关系范式；算法权力；人机共生；价值嵌入；悟空机制；技术向善

一、巴别塔的诅咒：效率狂飙下的传播异化与真实危机

2026 年 5 月 25 日，梵蒂冈新主教会议厅的一场发布会意外引爆全球科技圈。教宗良十四世正式发布首份通谕《Magnifica Humanitas》（中文译作《伟大的人性》），副标题直指“在人工智能时代守护人的价值”[1]。Anthropic 联合创始人 Christopher Olah 等科技界代表同台出席，标志着梵蒂冈首次将 AI 议题正式纳入文明、劳动与伦理的宏大框架中审视。通谕的核心并非反技术，而是警示“技术官僚范式”。教宗承认 AI 在医疗、科研等领域的巨大价值，但强调技术绝非中立，它带有创造者的价值观烙印。他提出了一个振聋发聩的呼吁：AI 需要被“解除武装”。这不仅指严禁将致命决策权交给自动化武器系统，更是要将 AI 从单纯的商业逐利与军事竞赛逻辑中解放出来。发布会上，Christopher Olah 罕见承认，前沿 AI 实验室深陷商业与地缘竞争，仅靠行业自律无法确保安全；同时，大模型内部已出现人类难以完全解释的复杂状态。对此，教宗提出了极具象征意义的“巴别塔与耶路撒冷”之问：人类是去建造一座追求绝对效率与统一的傲慢“巴别塔”，还是去重建一座由普通人共同参与、充满信任与连接的“耶路撒冷”？

通谕进一步指出了三大现实危机：一是深度伪造与虚假信息正在瓦解社会的真实基础；二是 AI 若仅以降本增效为目的，将剥夺劳动的身份认同，催生“新形式的奴役”；三是算法推荐与数字画像正潜移默化地替代人类的独立思考与选择。归根结底，《伟大的人性》在追问：当机器越来越像人，人是否还能保持人的温度？教宗给出的答案是，真正让人成为人的，从来不是计算能力，而是责任、良知与真实的连接。AI 时代真正需要守护的，不是技术本身，而是人本身。

项目基金：国家社科重大项目“构建政策实施和舆论引导相协同的预期管理机制研究”（25&ZD177 ）

作者简介：喻国明，北京师范大学新闻传播学院教授、学术委员会主任，北京师范大学传播创新与未来媒体实验平台主任

[1]　雷科技. 教皇发布首份 AI 通谕:4 万字 10 个观点,AI 焦虑被说透了[EB/OL]. (2026-05-28)[2026-05-31]. https://www.36kr.com/p/3827500620652800.

如果我们超越宗教的视角，将这篇通谕视为是一个具有巨大影响力的思想家向 AI 时代的我们所提出的一个重要而关键的思考题的话，我们必须去面对的是：当 AI 越来越像人，我们该如何守护“人的温度”？事实上， AI 时代，真正受考验的不是机器而是人类——这是一场关于技术、良知与未来的灵魂拷问。人工智能技术的迅猛发展正在深刻重塑信息传播的底层逻辑。从“人找货”的线性传播到“货找人”的智能推荐，从单向的信息传递到人机协同的意义共建，传播系统正经历一场前所未有的范式革命。这场革命带来的不仅是效率的飞跃，更是对人类主体性、情感连接和文明形态的深刻挑战。因此，构建 AI 时代的传播伦理学，本质上是一场“人性保卫战”——考验的不是机器能否模仿人的智能，而是人类能否在技术狂奔中守住文明的温度。

二、祛魅“工具中立”：算法黑箱中的价值暗码与权力拓扑

传统传播伦理将技术视为中性工具，这一立场在哈罗德·拉斯韦尔的《世界大战中的宣传技巧》中得到了典型体现[2]。拉斯韦尔以价值中立的态度分析战争宣传策略，认为技术本身无所谓善恶，关键在于使用方式。这种“工具中立”范式隐含着技术理性与价值判断的二元割裂，将传播伦理简化为对技术应用效果的评估，而非对技术本质的反思。然而，技术非中立性原理已从根本上动摇了这一范式。社会技术系统学派指出，技术不是孤立存在的，而是与社会、经济、政治等各个方面相互作用、相互影响的。哲学家唐·伊德(Don Ihde)更明确指出：“将仪器视为透明或中立，是一种隐含假设”[3]。在传播领域，麦克卢汉的“媒介即讯息”理论进一步揭示了技术对传播内容的内在塑造作用[4]。

1.AI 传播系统的价值嵌入

AI 的算法深处，藏着人类文明的伦理密码——技术不是无辜的信使，而是携带价值基因的生命形态。具体表现为：

（1）代码即价值选择：AI 传播系统的每一行代码都是价值判断的具象化。算法推荐的同质化本质是商业逻辑对公共领域的殖民，将用户简化为数据标签，将复杂的人际关系压缩为可计算的网络结构。当算法核心目标是用户留存与转化率时，技术就成为资本逻辑的延伸，而非服务人类福祉的工具。

（2）数据监控的伦理边界：数据监控的无边界化实质是权力对私人领域的侵蚀。西方“计算优先”的个人权利伦理强调数据最小化与用户控制，而东方“算计优先”的责任伦理则关注数据使用对社会整体福祉的影响。这种差异在传播领域尤为明显：西方更关注个人隐私保护，东方则强调数据的社会价值与集体用途。

（3）算法透明度的双刃剑效应：过度追求算法透明可能暴露技术漏洞或限制商业创新，而完全不透明则会损害公众信任。平衡透明度与保密性成为 AI 伦理治理的关键挑战。

2.破解技术黑箱中的“价值暗码”，需要建立“价值嵌入型”伦理框架

这要求传播伦理从“技术应用效果评估”转向“技术设计伦理审视”，让算法从“效率至上的独裁者”转变为“多元价值的调停者”。实现这一转向，需通过三重路径：

（1）制度创新：建立覆盖技术研发、内容生产、传播应用的全链条规制体系，将抽象伦理价值转化为可遵循的社会规则，实现权力与责任的对称配置。

（2）技术干预：开发可解释性算法(XAI)，设计偏好回滚机制，构建伦理嵌入的技术协议，让

---

[2] [美]哈罗德·D·拉斯韦尔. 世界大战中的宣传技巧[M]. 张洁,田青译. 北京: 中国人民大学出版社, 2003: 23, 37.

[3] [美]唐·伊德. 技术与生活世界:从伊甸园到尘世[M]. 韩连庆,译. 北京: 北京大学出版社, 2012: 45.

[4] [加]马歇尔·麦克卢汉. 理解媒介:论人的延伸[M]. 何道宽,译. 北京: 商务印书馆, 2000: 33.

算法能够感知并尊重多元价值。

（3）文化融合：推动东西方伦理智慧的互鉴，将儒家“仁爱”、道家“无为”、佛教“不杀生”等传统伦理转化为算法可理解的约束条件。

三、路径构建：耶路撒冷方案与生态嵌入

正如教宗通谕中所指出的，AI 技术的发展面临着两种截然不同的文明路径：巴别塔陷阱与耶路撒冷方案。前者以效率至上消解传播与认知上的人文性，后者则以责任为本重建文明的对话基础。

对于 AI 时代的传播生态的构建而言，巴别塔陷阱的危险在于，以“绝对效率”消解传播的人文性，以“信息茧房”固化认知的单一性，以“流量至上”解构公共领域的多元性。具体地说，就是算法推荐系统通过精准分析用户行为模式，不断强化其既有兴趣和立场，导致用户更容易陷入观点同质化的循环，认知视野逐渐收窄，独立思考与理性判断能力受到削弱。更严重的是，算法偏见的全球性扩散已成事实。例如，一项由开源 AI 公司 Hugging Face 首席伦理科学家玛格丽特·米切尔领导的名为 SHADES 的研究课题，收录了 300 多条全球刻板印象，涵盖性别、年龄、国籍等多个维度。研究人员使用 16 种语言设计交互式提示，并测试了数种主流语言模型对这些偏见的反应。结果显示，AI 模型对刻板印象的再现具有明显差异化特征。这些 AI 模型不仅表现出“金发女郎不聪明”“工程师是男性”等常见英语地区刻板印象，在阿拉伯语、西班牙语、印地语等语言环境中，也表现出对“女性更喜爱粉色”“南亚人保守”“拉美人狡猾”等偏见[5]。这表明，算法推荐已不再是简单的技术工具，而是成为价值观的传播载体。

面对巴别塔陷阱，“耶路撒冷方案”提出了“共同建构”的伦理路径。这一方案要求传播系统具备三种能力：在算法狂奔时保留反思空间（沉默的敬畏）、让不同主体参与意义生产（多元的协作）、为边缘群体预留传播通道（脆弱的包容）。这些能力并非抽象理念，而是可通过技术手段实现的具体设计。而我们所倡导的“生态级嵌入”理论正是这一伦理的实践注脚[6]。它强调传播不是技术的扩音器，而是社会的黏合剂。在技术层面，这体现为：

（1）多元价值平衡算法：借鉴儒家“中庸”之道，设计能够在效率与公平、个人与集体、短期利益与长期福祉间保持动态平衡的算法架构。例如，有 AI 开发者提出从业者需将公平性、透明性、隐私性和可靠性四大伦理原则转化为可测试的技术指标，通过伦理-by-Design 框架，主导需求分析、架构评审和代码实现全流程，构建包含公平性检测、透明性报告、隐私合规检查和可靠性熔断的验证工具链[7]。

（2）边缘群体传播通道保障：通过技术手段确保少数群体、弱势群体的声音不被算法淹没。联合国教科文组织《人工智能伦理建议书》提出的“可持续性”原则，要求持续评估 AI 对人类、社会、文化、经济和环境的影响，特别关注对少数群体的保护[8]。

（3）全球伦理对话机制：打破西方中心主义的治理话语霸权，为不同文明提供平等表达的平台。中国提出“各美其美·美美与共”的文明互鉴理念，强调在全球现代化进程中，应尊重不同文明的主体性，避免单一文明标准的垄断。

（4）“生态级嵌入”的两个基本层面：在制度层面，“生态级嵌入”要求：一是建立政府、企业、

---

[5] 张佳欣：AI 输出“偏见”，人类能否信任它的“三观”？新华网 http://www.xinhuanet.com/tech/20250717/76c3c26204694c42bc633a792e3c895f/c.html

[6] 张斌, 赵英才. 人工自然：生态嵌入与经济循环[J]. 自然辩证法研究, 2006(2): 5-8.

[7] 软件测试从业者. 伦理 AI 设计框架:从理论到代码实现[EB/OL]. (2026-04-03)[2026-05-31]. https://blog.csdn.net/2501_94471289/article/details/159641015.

[8] UNESCO. Recommendation on the Ethics of Artificial Intelligence[EB/OL]. (2021-11-24)[2026-05-31]. https://en.unesco.org/artificial-intelligence/recommendation.

学术界、公民社会多元主体共同参与的治理网络；二是推动形成全球性算法监管合作协议，如 GPAI（全球 AI 伙伴关系）等国际组织正在构建的框架；三是构建行政监督、行业自律和用户反馈三位一体的动态监督机制。在文化层面，“生态级嵌入”体现为：儒家“和而不同”思想对算法设计的启示，强调在保持多元性的同时促进共识形成；而道家“无为而治”理念也可以作为轻量化智能设计的指导——如天与养老的“无为算法系统”，通过自适应调节而非繁复预设实现对用户行为的尊重[9]；而佛教“慈悲”与“不杀生”原则对 AI 应用的约束，如医疗 AI 应避免因追求效率而忽视患者情感需求。

概言之，“耶路撒冷方案”，其实质是通过制度创新（全球治理联盟）、技术干预（XAI 可解释性）、文化融合（中庸之道）三管齐下。我们的指导思想中，要将儒家“和而不同”转化为算法的“多样性保留机制”，将道家“无为”转化为算法的“非侵入式设计”（即不过度打扰、不过度诱导）。由此在实践的基础上，建立“算法影响评估报告”制度，让算法权力在阳光下运行。

四、核心悖论：拟人化陷阱与去人性化救赎

AI 时代的传播伦理面临一个深刻的双重悖论：技术越“拟人”，人类越容易将伦理责任转移给机器；传播越“智能”，人越可能异化为数据节点。这一悖论揭示了传播伦理的核心挑战：如何在技术拟人化与人类主体性之间保持平衡。

（一）镜像迷思：AI 的“拟人化”表演与人类的“数据化”囚徒困境

AI 时代的传播伦理陷入了一种深刻的“镜像迷思”：技术越展现出惊人的“拟人化”能力，人类就越面临被“数据化”和异化的风险。在情感计算与生成式 AI 的驱动下，机器不仅能模拟人类的情绪语调，甚至能合成具有欺骗性的情感影像，这种“拟人化表演”正在模糊真实与虚拟的边界。其核心危害在于引发了严重的伦理责任转移——当用户面对心理咨询 AI 或陪伴机器人时，极易将原本属于人际交往中的共情、理解与关怀等道德义务，外包给这套精密的算法系统。正如文档中所警示的，这种“完美”的算法回应虽然满足了即时的情感需求，却可能导致人类在真实社会网络中的情感退化，使人际连接变得脆弱。最终，我们在享受技术便利的同时，却陷入了“数据化”的囚徒困境：人不再是传播的目的，反而沦为算法优化和流量变现的数据耗材。

（二）重铸“人的温度”：在人机传播中确立不可让渡的伦理红线

传播的“智能化”正在加剧人类的“去人性化”。算法推荐系统通过标签化细分人群，使不同群体处于差异化的信息环境之中，公共讨论空间被压缩，群体之间的理解与沟通难度增加，社会共识形成更加困难。长期依赖算法推荐获取信息会导致信息来源单一、知识结构碎片化，用户的批判性思维能力逐渐退化，最终沦为算法系统的“数据节点”。传播伦理必须建立“人性守护红线”，以应对这一双重悖论：

（1）拒绝将人简化为数据标签：在算法设计中保留人的复杂性与多维度性，避免将人类行为完全量化为可计算的特征向量。

（2）反对用算法替代人际理解：在医疗、教育、心理咨询等关键领域，强调人机协同而非人机替代，确保 AI 系统辅助而非取代人类的专业判断与情感连接。

（3）警惕以效率名义牺牲人与人的情感连接：在传播系统中为人与人的情感交流保留空间，避免技术理性对人文价值的全面侵蚀。

（4）维护传播的终极介质是“有温度的人”：传播不仅是信息传递，更是情感交流与价值共建，必须确保技术服务于人的全面发展，而非人的异化。

---

[9] 科技 Geek. 天与 无为算法系统 问世:端侧智能驱动养老行业新变革[EB/OL]. (2025-04-27)[2026-05-31]. https://www.fromgeek.com/daily/1044-683991.html.

五、实践框架：五大支柱与治理工具箱

为应对 AI 时代的传播伦理挑战，本文提出五大伦理支柱，以此共同构建一个以“人性温度”为核心的传播道德坐标系。

1.尊严伦理：人是目的而非算法的耗材

尊严伦理的核心原则是：数据可以量化行为，但永远无法计算灵魂——传播的第一准则是让人成为目的，而非算法的耗材。这一原则在传播领域面临严峻挑战，如社交媒体平台通过数据挖掘和行为预测不断强化用户成瘾性，使用户沦为流量经济的“数字劳工”。具体地说，尊严伦理的实践路径包括：

推广"提示工程师”培养计划：通过系统化训练，培养能够与 AI 进行有效人机对话的专业人才，强化人的主体性。提示工程师需掌握如何向 AI 明确表达意图、设置边界、要求解释等核心技能，确保人在传播关系中的主导地位。

建立“不可量化价值保护清单”：明确列出算法不应量化或过度量化的价值领域，如情感、隐私、信仰等，并通过立法手段强制要求 AI 交互界面标注“非人类情感主体”，帮助用户区分算法输出与人类共情。

设计“人机技能互补认证体系”：推动传播从业者从单纯的内容生产者向“算法监督者”“意义建构者”转型，通过再培训机制保障“数字工匠”的职业尊严。

2.真理伦理：传播应成为“真相免疫系统”

在深度伪造技术日益成熟的时代，传播必须成为“真相免疫系统”——用多元对话对抗信息操纵，以事实核查筑牢公共理性。算法推荐系统通过流量逻辑强化特定观点，削弱了公共理性的基础，使虚假信息与阴谋论在算法助推下迅速传播。具体而言，真理伦理的实践路径包括：

构建“真相共同体”协作网络：这是应对 AI 时代虚假信息泛滥、重塑清朗信息生态的系统性工程。这一机制的核心在于打破壁垒，整合多元力量形成合力：专业媒体机构与事实核查组织发挥“前哨”作用，依托权威资源进行深度查证与辟谣；学术研究者提供坚实的理论支撑与前沿技术辅助，帮助剖析虚假信息的生成逻辑与传播特征；而普通用户则从被动的接收者转变为主动的监督者与参与者，通过日常媒介素养的提升和负责任的分享行为，阻断谣言的传播链条。另一方面，通过建立跨领域、多主体的协同共治模式，“真相共同体”不仅能够实现对虚假信息的快速发现、精准识别与有效拦截，还能在全社会范围内建立起抵御信息污染的集体防御机制。这不仅有助于压缩虚假信息的生存空间，更能以真实、理性的优质内容凝聚共识，为公众在复杂的信息洪流中点亮思维的明灯。

嵌入 XAI 可解释性工具与事实核查平台：通过算法透明化与人工核验形成伦理双保险，使用户能够理解算法推荐的逻辑，并在必要时通过事实核查工具验证信息真实性。所谓 XAI 是 Explainable Artificial Intelligence 的缩写，中文全称为可解释人工智能。它的核心目标是把传统 AI 难以理解的“黑箱”决策过程，转化为人类能够看懂、听懂并能追溯理由的“透明”过程。通俗地说：普通 AI（黑箱模型） = 只给你一个最终结果，但说不出理由（就像一个人凭直觉答题，却解释不清为什么）。而可解释 AI（XAI） = 不仅给出结果，还会清晰地向你说明判断的依据、逻辑和权重（比如：“我判定这笔贷款有风险，是因为申请人的负债比例过高，且近期征信有异常”）。

建立“理性内容配额制”：建立“理性内容配额制”，是重塑健康信息生态、对抗算法偏见的关键制度设计。该机制要求在底层推荐逻辑中强制设定深度分析、多元观点及事实核查类内容的推送比例，通过流量倾斜打破单一情绪的“信息茧房”。在技术实现上，平台需改变单一的“点击率驱动”模式，引入多维权重模型，将专业性、事实核查度等指标纳入考量，并设定最低多样性阈值。当检测到用户长期处于低多样性状态或陷入极化阵营时，系统应主动注入跨领域科普与对立视角的深度

报道，强制关联背景证据链。这不仅有助于从源头上抑制群体情绪极化，更能引导公众接触更全面的信息，提升理性判断力，从而在追求个性化的同时，保障公共讨论空间的真实性与权威性。

设计“慢传播”机制：这是对抗数字时代信息过载与情绪传染的有效交互干预。该机制主张在即时反馈链路中刻意嵌入反思性停顿（如延迟响应功能），通过人为设置的认知缓冲带，打破用户被算法驯化的惯性接受模式。研究表明，适度的延迟并非技术缺陷，而是具有深刻心理学意义的“有益摩擦”。澳大利亚墨尔本理工大学 Lauren L. Saling2025 年发表在心理学新思想的一项研究表明，延迟性思考使人更能识别复杂的合取规则，以避免匆忙做单一判断，并且，在创造力的替代用途测试里，人的发散性思维与创造力得分也更高[10]。这种停顿能够促使用户跳出自动化的快思考状态，激发主动的慢思考，从而深化对复杂议题的理解与把握。它让信息的传递不再是表面化的数据交换，而是成为认知沉淀的过程，赋予受众在复杂舆论场中重获感官自主权与判断主动权的机会。

3.劳动伦理：AI 应成为人类创造力的放大器

AI 不应制造“新形式奴役”，而应成为人类创造力的放大器——传播业的未来不是人机替代，而应是人机共舞。算法推荐系统通过精准预测用户行为，不断强化特定消费模式，使用户陷入“算法殖民”状态，丧失自主选择与创新的能力。体地说，劳动伦理的实践路径包括：

建立“人机技能互补认证体系”：这是智能时代重塑传播人才评价标准、构建新型生产关系的关键路径。该体系旨在打破传统单一维度的考核模式，推动传播从业者向复合型“双栖人才”转型：一方面，要求从业者熟练掌握 AI 辅助创作、内容管理、数据分析等前沿技术工具，提升信息处理与分发效能；另一方面，高度强调并考察人类独有的创意编排力、情感共鸣力以及伦理判断力。在具体的实施层面，这一认证体系可围绕底层夯实语言与数字素养根基、中层强化行业场景适配、顶层培育创新表达与伦理把关三个维度搭建分层评价模型。通过将批判性思维、事实核查能力以及价值观过滤机制纳入核心考核指标，确保从业者在驾驭技术洪流的同时，能够坚守人文立场，有效规避算法偏见与数据异化风险。最终，这种以能力图谱为牵引的认证机制，将助力传播人才在复杂的人机协同生态中实现优势互补，创造出兼具技术锐度与人性温度的高质量内容价值。

设计“跨圈层中继算法”：这是打破数字时代信息壁垒、重塑社会共识的关键技术干预。该机制的核心在于重构推荐逻辑，将“认知多样性”与“公共利益”纳入核心权重指标，从而强制向用户推送跨越原有偏好边界的异质信息与理性观点。在具体实现上，平台可建立“跨气泡信息中转站”，通过算法动态监测并识别不同群体间的认知鸿沟，以温和、客观的方式将具有建设性的对立视角引入用户的视野。这种适度的“寄生干预”不仅能有效抑制极端情绪的回音室效应，还能在算法层面为多元观点的碰撞留出空间。通过促进不同立场群体间的相互理解与良性对话，“跨圈层中继算法”有望逐步弥合日益撕裂的社会认知，为重建包容、理性的公共对话基础提供坚实的技术支撑。

重构流量逻辑：这是破解当前算法推荐“唯数据论”困境、推动数字信息生态向善的核心举措。该机制要求平台彻底改变以点击率和停留时长为单一导向的流量分配模式，将“公共价值贡献度”作为算法权重的核心考量因素。在具体的技术实现上，平台需构建多维度的内容价值评估模型，将议题的思想性、知识增量以及正向社会效应等指标深度嵌入推荐系统的目标函数中。通过设立“公共内容权重指标”与“推荐底线”，确保教育科普、权威新闻及民生建设等高社会价值内容能够获得基础可见性与结构性扶持。这种从“兴趣引导”向“价值平衡”的范式转向，能够有效避免优质理性内容在激烈的流量竞争中被边缘化，从而遏制过度娱乐化和情绪化趋势导致的传播异化，让流量真正成为承载主流价值与凝聚社会共识的数字桥梁。

实施“技术普惠补偿原则”：这是弥合数字鸿沟、实现社会公平正义的必然要求。在人工智能等

---

[10] 风雨无阻. 拖延者更聪明?研究表明:爱拖延的人,反而更有创造力和逻辑力[EB/OL]. (2026-02-28)[2026-06-03]. https://www.toutiao.com/article/7611696610224947758/.

新技术飞速发展的当下，由于基础设施落差与技术素养断层，弱势群体往往面临被边缘化的风险。因此，必须通过制度设计与资源倾斜进行主动干预与补偿。具体而言，该原则强调对老年人、低收入群体及偏远地区居民提供精准的“数字技能赋权”。一方面，政府与企业应协同推进适老化及无障碍改造，降低智能设备的使用门槛；另一方面，需依托社区和公益组织持续开展多层次数字教育，将 AI 应用纳入终身学习体系，帮助弱势群体掌握现代科技技能。此外，还应完善收益分配机制与社会保障兜底，确保技术带来的经济红利能够向贡献数据的底层劳动者倾斜。唯有坚持“以人为本”的技术向善导向，才能打破技术精英的垄断壁垒，让智能化时代的红利真正惠及全体社会成员。

4.自由伦理：用户应拥有“说不的权利”

算法的终极权力不是预测行为，而是剥夺选择——传播伦理必须赋予用户“说不的权利”，让技术可控、可退、可质疑。算法推荐系统通过持续学习用户行为，不断缩小其信息接触范围，最终使用户陷入算法的“自由牢笼”，看似拥有无限选择，实则被精准控制。因此，自由伦理的实践路径至少应包括：

设计“可逆性交互机制”：这是重塑算法信任、将控制权真正交还给用户的核心产品理念。该机制要求平台赋予用户对自身数据与推荐逻辑的绝对掌控权，确保其能够随时退出数据追踪、重置个性化偏好或质疑系统决策。在具体实现上，这种可逆性体现在多个维度：首先，提供一键式的“算法重置”功能，允许用户通过清理缓存或重新选择兴趣标签等方式，让推荐流回到全新的起点；其次，建立细粒度的负反馈闭环，例如支持长按视频标记“不感兴趣”并选择具体原因（如内容重复、质量低下），使算法能精准识别并减少类似推送；最后，设置清晰的退出与恢复路径，当用户关闭某项推荐功能后，不仅能在设置中随时“重新开启”，还能查看和管理历史屏蔽记录。这种充满安全感的设计不仅尊重了用户的自主选择权，也为平台优化模型提供了宝贵的反馈数据，实现了人机之间的良性博弈与协同进化。

推行“算法影响评估报告”制度：这是打破技术壁垒、重塑公众对算法信任的关键治理手段。该机制要求平台将原本被视为商业机密的底层逻辑置于阳光之下，定期向社会公开其推荐机制的运行原理及潜在的偏见风险。在具体实施层面，这一制度倡导采用双重信息披露模式：一方面提供包含核心参数与决策依据的详细技术格式，供监管机构与专业审计人员进行深度审查；另一方面，生成通俗易懂的摘要版本，向普通用户清晰阐释数据使用目的、内容分发规则及其可能带来的社会影响。同时，结合风险分级理论，对于涉及信用评估、定价等高风险领域的算法，应强制进行独立审计与全流程追溯。这种从“黑箱操作”走向“透明治理”的转变，不仅有效保障了用户的知情权，更能倒逼平台在系统设计之初就内嵌伦理考量，防范算法歧视与信息操控，真正实现技术向善。

建立“算法选择权”机制：这是打破单一价值导向、重塑用户认知自主权的核心制度设计。该机制主张将推荐系统的控制权部分让渡给用户，允许其在一系列不同价值导向的算法模型间进行自主选择。在具体的技术实现与产品形态上，平台可以开发直观的“算法控制面板”，为用户提供多样化的推荐策略选项。例如，追求高效获取资讯的用户可以选择“效率导向”模型；注重社会公平与多元视角的用户可切换至“公平导向”模式；而渴望探索未知领域的用户则能开启“创意导向”系统。此外，平台还需提供可视化的“算法偏好说明书”，清晰解释不同设置对推荐结果的具体影响。这种从“被动投喂”向“主动定制”的转变，不仅有效满足了用户的个性化需求，还能从根本上缓解信息同质化问题，推动算法治理走向多元共治的新格局。

构建“动态伦理字典学习”框架：这是将东方哲学智慧融入人工智能治理的前沿探索。该机制深刻借鉴道家“无为而治”的思想精髓，主张摒弃僵化、机械的底层代码约束，转而开发能够根据用户反馈与社会环境变化动态调整伦理权重的自适应算法系统。在具体实现上，这种“无为”并非放任不管，而是强调顺应事物本性的柔性引导。系统通过持续捕捉多元文化背景下的用户交互数据与情感反馈，像水一样灵活地绕过生硬的干预，自动优化伦理决策矩阵。它不预先设定绝对统一的

道德标准，而是在复杂多变的现实情境中保持高度的弹性与包容性，做到“柔弱胜刚强”。这种充满智慧的算法设计，完美契合了老子“治大国若烹小鲜”的治理理念——在保障安全底线的前提下，以润物细无声的方式实现人机价值的协同进化，从而达成既尊重个体差异又兼顾社会共识的灵活治理境界。

5.包容伦理：技术垄断只会加剧文明分裂

技术垄断只会加剧文明分裂，唯有"慢传播”才能缝合数字鸿沟——让每个声音都被听见，是传播最朴素的伦理。算法推荐系统基于主流文化数据训练，往往忽视少数群体、边缘文化的声音，导致文明多样性在数字空间中被压缩。包容伦理的实践路径包括：

建立“全球传播治理联盟”：这是打破西方中心主义技术霸权、推动构建更加公正合理的国际数字秩序的关键举措。该机制的核心在于搭建一个平等互鉴的文明对话平台，通过常态化的思想碰撞与经验共享，探寻人工智能治理的共通价值与多元路径。在具体实践中，这一联盟致力于主动设置全球议程，向世界阐释人工智能发展的中国方案。它将积极推动东方传统伦理智慧（如“中庸”、“和谐”、“天下为公”）与数字文明的深度融合，把人类命运共同体的价值追求转化为具有普遍认同的全球治理规范。例如，在算法设计、数据跨境流动等核心议题中嵌入文化包容性与技术向善原则，要求 AI 生成内容保留地域文化标识，防范因算法偏见导致的“文化误译”。同时，联盟还将聚焦跨机构协作，鼓励国内外企业共建跨文明数据库，为发展中国家提供符合自身国情的治理思路，从而持续提升我国在全球数字治理领域的伦理引领力与话语影响力。

实施“技术普惠补偿原则”：这是矫正算法偏见、构建包容性数字生态的必然要求。在当前的流量分发机制下，少数群体与边缘文化往往因缺乏历史数据积累和初始互动量而面临冷启动瓶颈，导致其在信息传播中被系统性边缘化。因此，必须通过算法层面的主动干预进行结构性补偿。在具体落地时，平台应将“公平性”转化为可计算的技术指标，为这些弱势内容设立专项的“流量扶持池”。通过在推荐排序模型中引入逆倾向加权等去偏机制，人为提高小众及边缘内容的曝光基准线，打破传统协同过滤带来的马太效应。这种补偿并非无底线的强制摊派，而是为了提供平等的起步机会，确保多元声音能够跨越数据鸿沟被公众看见。唯有如此，才能有效避免算法沦为固化社会不平等的工具，让技术红利真正惠及每一个独特的文化个体

构建“多元文化语料库”：这是从数据源头破解算法偏见、实现情感传播包容性重构的核心基石。当前，主流 AI 模型多基于标准化语言样本训练，难以精准捕捉全球各地边缘文化的独特表达与深层情感符号。因此，必须统筹多方资源，广泛采集涵盖不同民族、地域及方言的自然口语与多模态数据，打造具有高度代表性和丰富度的跨文化语料底座。在具体的技术实施上，该语料库不仅要纳入海量的文本内容，还应深度融合图片、音频和视频等多模态数据，全面记录真实生活场景中的文化记忆与情感交流。通过对这些原始数据进行精细化的多维标签分类与人工质检，系统能够深度学习并理解小众语言背后的复杂语义与文化语境。这种从底层数据维度的纠偏，将有效修正传统算法对非主流表达的刻板印象与识别盲区，使人工智能真正具备跨文化共情能力，从而在全球范围内推动更加平等、多元且充满人性温度的智能对话生态。

推动“伦理向善”的传播实践：这是引导人工智能从单纯的“流量驱动”向“价值引领”转型的关键举措。该机制深刻汲取了儒家“仁者爱人”的哲学精髓，强调在算法设计的底层逻辑中注入深厚的人文关怀与社会责任感，使技术真正服务于人类福祉。在具体落地层面，平台应将抽象的道德理念转化为可执行的代码约束与权重指标。例如，在内容分发模型中建立多维度的社会价值评估体系，主动识别并优先扶持那些倡导和谐、增进共识、传递正能量的优质内容；同时，对可能引发群体对立、情绪撕裂或信息茧房效应的极端言论进行柔性干预与降权处理。这种将传统美德与现代科技深度融合的治理模式，不仅体现了“己欲立而立人”的互动智慧，更确保了智能传播系统在追求效率的同时不失道德底线，最终为构建理性、包容、和谐的数字公共空间提供坚实的技术支撑。

六、未来图景：动态演化与悟空机制

当算法能生成情感文本、合成逼真影像、预测社会行为，传播伦理的终极命题已然清晰：技术服务于人的全面发展，而非人的异化；文明的进步在于关系的深化，而非效率的狂奔。传播学应当成为"人性的守夜人"，在技术狂飙中锚定文明的根基。我们认为未来传播伦理学将呈现出三大发展趋势：

（一）从理性范式到关系范式的范式转变

这标志着 AI 伦理评估正经历一场深刻的价值重构。当前主流的“理性范式”高度依赖抽象原则与后果计算，往往将复杂的伦理情境简化为静态的合规清单和功利主义的利弊权衡。这种模式虽然便于审计，却忽视了真实互动中流动的个体感受与动态的社会联系。

未来的传播伦理将全面转向关注人机关系质量的“关系范式”。在这一新框架下，评估的核心不再是冰冷的技术指标（如准确率、召回率），而是用户在长期互动中的主观体验与福祉。具体而言，评估指标需向三个维度延伸：首先是“长期信任度”，即系统能否在持续的交互生命周期中培育出尊重与可信的连接；其次是“情感连接强度”，要求 AI 具备情境感知能力，提供有温度的陪伴而非机械的模板回应；最后是“赋能感评估”，重点关注用户在使用后是否感到自身能力得到了提升与延展，而非陷入对技术的过度依赖。这一范式转移促使我们重新审视技术的终极目的——真正的智能不仅在于“算得准”，更在于懂得如何温柔以待，维系健康、可持续的人机共生关系。

（二）“悟空机制”：从静态规则到动态协商的治理机制

这标志着 AI 传播伦理正经历一场深刻的范式跃迁。传统的伦理框架往往试图将复杂的道德困境简化为预先设定的代码或僵化的合规清单，这种“一刀切”的模式在面对快速迭代的智能时代时，极易陷入路径依赖与系统性盲区。

有研究表明，AI 元人文框架的核心生命力，在于其内嵌的“悟空机制”[11]。该机制远非一个简单的冲突解决工具，而是智能文明实现“自我反思”与“范式跃迁”的元能力。面对因系统决策速度与人类认知调适速度断裂而加剧的深层僵局，悟空机制通过其“系统场域主权”，扮演着至关重要的“速度调节器”与“意义再生器”角色。该机制的运行分为三个递进阶段：首先是“悬荡”，即在面对前所未有的价值冲突时，主动暂停自动化决策与效率至上的惯性思维，保持问题的张力与开放性；其次是“协商”，通过构建多方参与的对话场域，让不同利益相关者在具体情境中进行深度的价值博弈与意义重构；最后是“悟空”，在充分吸收多元视角后实现认知层面的创造性涌现，跳出二元对立的僵局，生成具有更高维度的解决方案。

这一动态协商过程不仅有效化解了具体的伦理危机，更将每一次碰撞转化为系统学习的契机。新达成的共识会被沉淀并反哺至底层知识库，使伦理原则不再是刻板的教条，而是能够随技术演进与社会变迁持续自我修正、不断生长的生命体。

（三）从技术理性到人文价值的深度融合

这标志着传播伦理正超越单纯的技术管理规范叠加，迈向以“人的现代化”为核心旨归的伦理秩序重塑。在这一进程中，中国式现代化的本质被深刻界定为人的现代化，其终极价值在于实现人的自由全面发展，这构成了对西方现代化进程中资本异化逻辑的辩证扬弃与根本超越。

在人工智能深度嵌入新闻传播、舆论引导与文化传承的今天，技术应用的双刃剑效应日益凸显，技术异化对人的主体性侵蚀成为严峻挑战。因此，构建符合中国国情、体现社会主义核心价值观的可信任人工智能传播伦理框架，既是防范风险的现实需求，更是推动智能向善发展的必然选择。这一融合要求我们坚持“人民至上”的价值依归，确立“物服务于人”而非“人为物役”的发展逻辑。

---

[11] 岐金兰. AI 元人文:悟空机制与反思——论智能文明的自我超越之道[EB/OL]. (2025-11-04)[2026-06-03]. https://www.cnblogs.com/qijinlan/p/19191236.

通过制度创新划定技术边界，将抽象的伦理价值转化为可遵循的社会规则；同时借助技术创新实现自我净化，打破“算法黑箱”，保障公众知情权与监督权。当技术逻辑与人文精神实现深度交融，人工智能将真正成为守护人类主体性、促进人的解放的重要力量，最终引领人类社会迈入人机协同、共生共荣的文明新形态。

概言之，AI 时代的传播伦理，本质是一场“人性保卫战”——它考验的不是机器能否模仿人的智能，而是人类能否守住文明的温度。当技术能够生成情感文本、合成逼真影像、预测社会行为时，传播学更需要成为“人性的守夜人”，在算法的巴别塔上重建文明的对话基础，让传播真正成为“伟大人性”的见证者与守护者。

在这场技术与文明的对话中，我们或许可以借用《中庸》的智慧：“至诚之道，可以前知。国家将兴，必有祯祥；国家将亡，必有妖孽。见乎蓍龟，动乎四体。祸福将至：善，必先知之；不善，必先知之。故至诚如神。”[12]在 AI 时代的传播伦理建设中，我们同样需要保持“至诚”的态度，既不过度恐惧技术的威胁，也不盲目崇拜技术的便利，而是在技术与人性的平衡中，寻找传播文明的未来航向。我们始终需要记住：技术服务于人的全面发展，而非人的异化；文明的进步在于关系的深化，而非效率的狂奔。传播的终极介质永远是“有温度的人”，而非冰冷的算法。

[12] 朱熹. 四书章句集注[M]. 北京：中华书局，2011：33.