大语言模型赋能政治参与的技术机遇与民主挑战

The Technical Opportunities and Democratic Challenges of Large Language Models Empowering Political Participation

廖文淇 /LIAO Wenqi

（复旦大学马克思主义学院，上海，200433）
(School of Marxism, Fudan University, Shanghai, 200433)

摘　要：人工智能时代，大语言模型通过即时化交互与个性化适配，重塑公众表达与决策制定的交往结构。这一算法介入后的对话拓扑，在推动政治参与的数字化转型的同时，也由于算法系统的认识论遮蔽，以及价值偏误与伦理界限模糊等内在局限，导致政治话语的客观性面临解释学断裂，从而加剧民主共识的构建难度。为了更好推进技术政治实践、提升民主效能，当前需要建立伦理规范与技术治理框架，实现“以正义为导向的算法校准”，积极引导其服务于人类共同福祉。防范应对大语言模型潜在的认知操纵风险，确保技术进步不仅追求效率与利益，更可保障个体权利、促进社会公平，乃至维护民主尊严。

关键词：大语言模型　算法介入　民主共识构建　认知操纵　正义校准

Abstract: In the era of artificial intelligence, large language models reshaped the communication structure of public opinion expression and decision-making through instant and personalized information interaction. The dialogue topology after the intervention of this algorithm not only promotes the digital transformation of political participation, but also causes the objectivity of political discourse to face hermeneutics fracture due to the epistemological obfuscation of the algorithm system, as well as the inherent limitations such as value bias and ambiguous ethical boundaries, thus increasing the difficulty of building democratic consensus. To better promote technical and political practice and improve democratic efficiency, it is necessary to establish an ethical framework and technical governance rules, achieve “justice-oriented algorithm calibration” and actively guide it to serve the common well-being of mankind. We should guard against the potential cognitive manipulation risks of the large language models, and ensure that technological progress not only pursues efficiency and interests, but also protects individual rights, promotes social equity, and even safeguards democratic dignity.

Key Words: Large language models; Algorithm intervention; Democratic consensus construction; Cognitive manipulation; Justice calibration

中图分类号：TP18; N031　DOI: 10.15994/j.1000–0763.2026.02.002　CSTR: 32281.14.jdn.2026.02.002

习近平总书记在致首届数字中国建设峰会的贺信中指出，“当今世界，信息技术创新日新月异，数字化、网络化、智能化深入发展，在推动经济社会发展、促进国家治理体系和治理能力现代化、满足人民日益增长的美好生活需要方面发挥着越来越重要的作用。”[1] 随

基金项目：教育部哲学社会科学研究重大课题攻关项目“建设中华民族现代文明的哲学基础问题研究”（项目编号：24JZD009）；2025 年度上海学校共青团工作研究课题“人工智能时代青少年道德认知优化路径研究”（项目编号：2025LXJ3-1）。
收稿日期：2025 年 5 月 19 日
作者简介：廖文淇（1998–）女，吉林长春人，复旦大学马克思主义学院博士研究生，研究方向为政治哲学、应用伦理学。Email: 3410456315@qq.com

着生成式人工智能的快速发展，以DeepSeek与GPT-4等为代表的“大语言模型”（Large Language Models, 后文简称LLMs）日渐被应用于社会公共领域①，深度参与并重塑政治生态。《数字中国发展报告（2024年）》显示，截至2024年底，我国数字政务网站的乡镇街道接入率达99.4%，各省市及地区积极推动IPv6技术网络应用，积极响应国家政务服务平台与“互联网＋监管”系统“一朵云”运行，切实推进降本增效。[2]然而，人工智能技术在向人们展示自身“美丽新世界”的同时，也迫使我们不得不在算法边界思忖人类社会的未来命运。有调查显示，主流LLMs在涉及社会福利议题时存在系统性意识形态偏差，其政策建议与训练数据来源国的政治光谱，仅呈现0.78的相关系数。[3]这意味着人工智能在优化社会治理效能的同时，也在无形中为恶意操纵提供了技术温床。不仅如此，尽管LLMs的出现为分析政治行为与社会复杂系统提供技术支持，却也面临算法垄断、道德失灵与价值偏误等深层挑战。因而，解构大语言模型在政治参与中的技术嵌入路径，结合现实案例探索技术风险与制度韧性的分析框架，进而探寻技术可能性与制度可行性之间的新平衡，正成为人工智能时代推进算法民主的重要议题。

一、大语言模型发展与政治参与的结构转型

从古希腊罗马的石刻法典到现代社会的电视辩论，科学技术的发展始终推动着民主形态的历史演进。步入人工智能时代，算法技术的更新迭代再次为政治、社会实践的转型升级提供现实基础。以2021年爱沙尼亚地区为例，其运用区块链技术实现跨国电子投票，[4]打破传统政治参与的时空壁垒，堪称全球数字化建设的典范。这些人工智能技术中就包括了自然语言处理（Natural Language Processing）领域的新兴产品——大语言模型。其生成演化至今大致经历了技术基础探索、预训练模型与多样化发展应用三阶段。

2017年，Google团队在《神经信息处理系统》（*Neural Information Processing Systems*）上发表的论文“注意力是你所需要的一切”（Attention Is All You Need），首次提出了一个全新的网络架构——Transformer，该架构完全基于注意力机制，摒弃了传统的循环神经网络（Recurrent Neural Network）和卷积神经网络（Convolutional Neural Network）的局限。[5]正是这一架构，为日后LLMs的研发提供核心技术支撑。

2018年，谷歌发布基于Transformer编码器的BERT模型，通过“掩码语言建模”（Mask Language Modeling）和“下一句预测”（Next Sentence Prediction）功能，实现双向上下文理解，从而显著提升文本分类、问答等任务的性能。由此，“预训练＋微调”的范式成为生成式模型的标杆和重要转折点。历经四年的探索，Open AI公司于2022年陆续推出了ChatGPT和GPT-4模型，尝试结合监督微调（Supervised Fine-tuning）和基于人类反馈的强化学习（Reinforcement Learning on Human Feedback）引入多模态能力，以支持文本、图像和音频的综合处理，进一步扩展了人工智能技术的应用场景，同时显著缓解生成内容的“幻觉”（Hallucination）问题，从而推动对话式AI的普及。[6]2024年底，中国企业如百度的“文心一言”大语言模型、杭州深度求索人工智能基础技术研究有限公司的“DeepSeek-R1”，陆续通过高性价比和垂直领域优化，加速人工智能应用落地。

作为人工智能领域的前沿技术，LLMs所具有的一系列显著特征，使其能够在自然语言处理及更多领域展现出任务泛化性、生成创

① “大语言模型”是指那些基于“Transformer”深度学习架构，通过海量文本数据的深度训练，学习、模仿人类语言规律及知识系统的人工智能技术。其所包含的数百亿参数的深度神经网络，能够处理文本生成、语义理解、逻辑推理等任务，应用于对话、翻译以及创作辅助等多场景。

造性与多模态扩展性等独特潜力，从而深刻改造民主实践样态。当被应用于政治参与时，LLMs能够缓解政治共识生成的传统痛点，作为信息生产与分发的“新基础设施”，推动后者从“工具依赖”转向“智能服务”，进而提升公共决策的效率与体验。这一技术优化路径主要涉及“交往结构”“话语协商”与“共识生成”三方面。

就交往结构而言，LLMs正在重构人类互动的基本范式，将传统的主体间对话转化为以智能系统为中介的人机交互模式。其运用自身技术架构对自然语言进行转化处理的同时，也将原本的人与人之间的交往转变为从“人–机”到“机–人”的连续嵌套。此时的对话不再直接发生于人类主体之间，而是经过算法系统的信息解码、语义重构与再生成反馈，最终输出为机器语言。这种以LLMs为中介的交往过程，通过信息整合与语义解析，有效提升了对话效率。特别是在参与主体上，LLMs降低了政治表达的智识门槛，提升弱势群体、边缘社区获得政策解读和意见生成的能力，同时催生如“哈贝马斯模型”等虚拟意见领袖和AI协商代理，形成人机协同的多元主体结构。但需要警惕的是，当公民通过Chat-GPT撰写政策建议书或借助DeepSeek分析立法草案时，即赋予了LLMs政治参与代理权，这一算法系统所具有的“下一句预测”功能及其“拟主体性”，可能就会凭借自身算法垄断而掌握话语权，使人类主体间的平等对话扭曲为人类与算法之间的“隐性博弈”。

从话语协商来看，LLMs通过生成高度流畅的文本，比如跨语言翻译等，能够在一定程度上拓展民意表达与公共协商的边界（如图1所示）。源于生活世界的交往话语作为主体间交往的核心媒介，本身具有可理解性、真实性、真诚性与正确性四重有效性。[7]这一基于模型支撑的智能问答与政策模拟工具，有力地增强了公民理性对话能力，从而提升交往的可理解

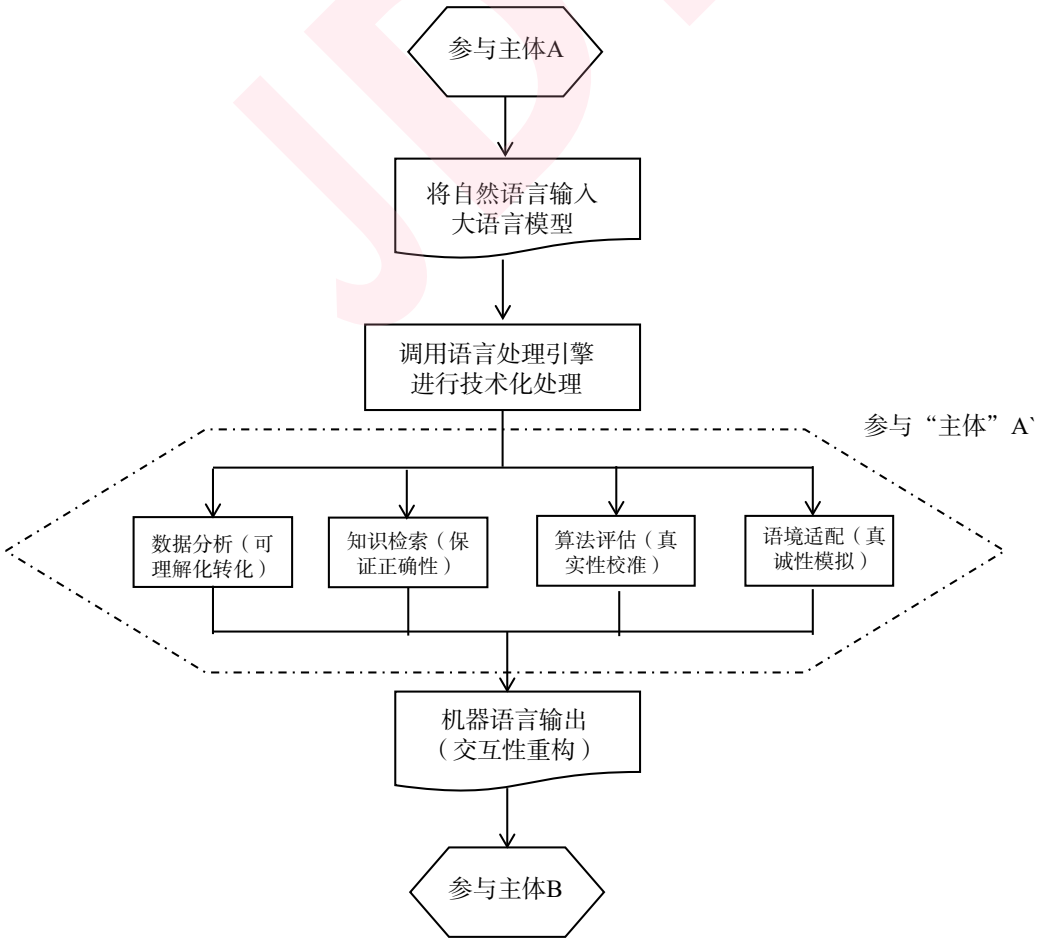

图1 LLMs中介下的交往拓扑

性。算法驱动的舆情分析使政府得以捕捉微观诉求，不仅极大提升了民情收集的广度和效率，更拓展了政府感知社会情绪和需求侧变化的深度与广度，有助其及早发现并化解基层矛盾。[8] 与此同时，大语言模型基于自身数据库所生成的“意见”还具有正确性的相对优势，特别是在快速整合海量信息、识别潜在模式，以及提供多元化视角方面。它能高效处理结构化知识，在事实性查询、数据总结与常见问题解答等场景下，展现出较高的表达准确性。但这一技术中介也可能会由于人机对话过程中“主观表达”与“技术模拟”的模糊边界，导致语言本身的真实性存疑。特别是当用户与AI展开辩论时，后者既可模仿公民口吻批量生成“民意”，也能伪装成专家提供“专业”意见，比如虚构学术论文支撑特定立场，致使公共领域陷入“真实性危机”。[9] 不仅如此，经过机器转化后的交往话语，由于缺失现实生活世界作为背景，在真诚性方面令人难以信任。作为无意识的符号处理器，LLMs输出的“真诚表达”实质是训练数据中情感模式的统计复现，其话语只是来自程序设定的语料库，而非源于言说者对自身话语承担伦理责任的意愿。就此而言，传统人类主体之间通过眼神、语调等副语言系统构建的真诚性验证通道，被压缩为单向度的文字形式，用户既无法穿透代码检验机器的真实意图，亦被迫接受这一“伪主体”所设定的交往规则。

从共识生成来看，LLMs通过重构信息传播与认知交互机制，深刻改变政治共识的生成逻辑（如图2所示）。传统政治共识源自哈贝马斯意义上的“生活世界”，特别是那些主体间共享的经验与价值观念。而LLMs的介入，则将这一过程技术化为“算法化知识生产——过滤——分发”的三阶系统，并大幅提升信息处理与整合效率，实现意见的即时聚合与动态分析。其通过精准识别多元诉求与潜在冲突点，引导更精细化的议题聚焦与方案调适，从而降

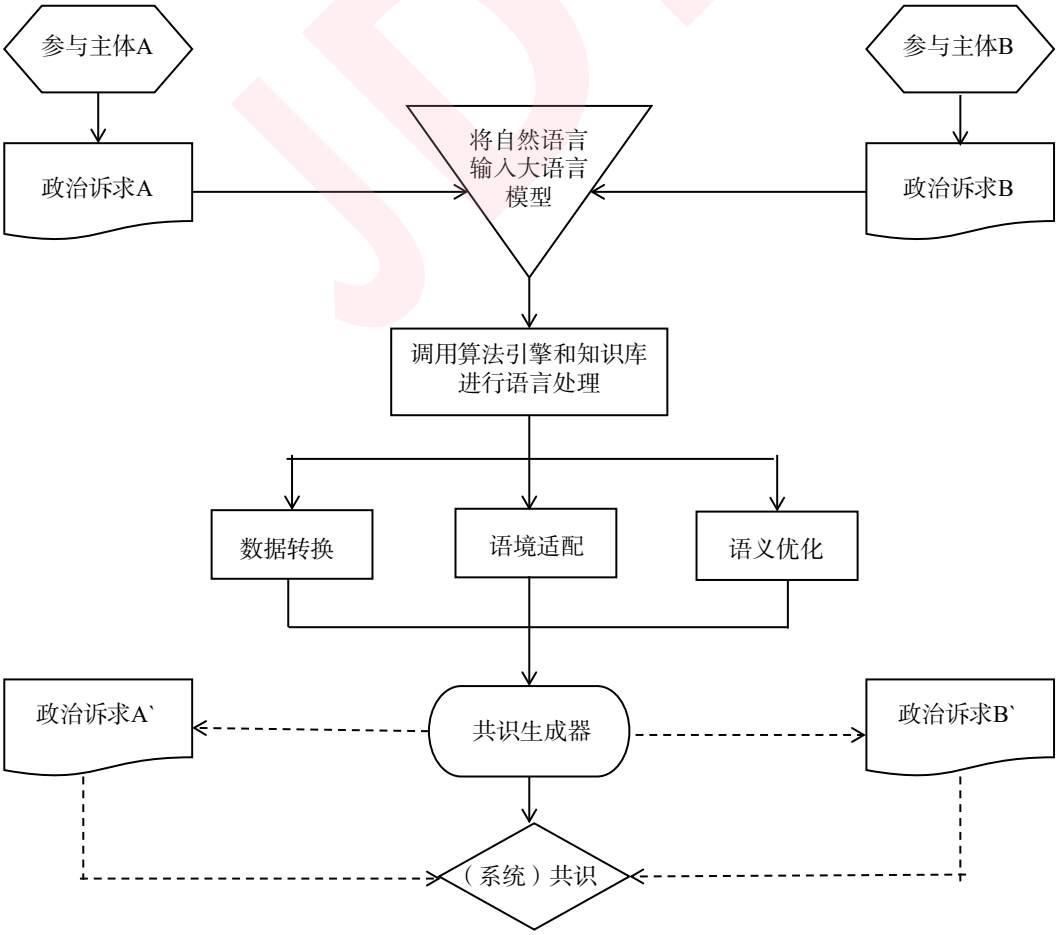

图2 LLMs中介下的“共识”生成过程

低传统共识构建中的摩擦成本与时间延迟。但这一“预训练＋微调”模式也可能通过训练数据的选择性编码，过滤边缘声音、强化主流叙事，比如预设某一政治议题的语义边界，使公共讨论被锚定在算法可控的“安全区”内。在传播中端，LLMs的即时生成能力将政治话语压缩为可交互的“信息胶囊”，公民从主动的批判者退化为被动的接受者，其认知框架被算法推荐的“最优解”悄然形塑。在共识反馈阶段，一旦少数群体的抗争性话语被算法归类为“不和谐音”，就会被LLMs借助算法手段予以剔除。概言之，政治参与所强调的意见的多元性可能为算法逻辑所扼制，致使公共决策只能囿于既定秩序而丧失超越性。

二、雅努斯面孔：大语言模型实践应用中的潜在风险

“技术对于现代民主的运转不可或缺，但它并非民主的天然盟友。”[10]当LLMs成为政治信息的中介者，就可能会使技术赋能的参与式民主异化为算法操控的政治表演。因而如何防止“技术民粹主义”与集体非理性状况，是巩固提升数字时代民主效能的重要议题。

作为新兴研究领域，学界当前围绕“人工智能的民主化”（Democratizing AI）问题的探讨日渐增加。[11]就支持意见而言，有学者基于艾里斯·扬（Iris Marion Young）的“结构性不公”理论指出，算法不公是社会技术结构失衡的产物，AI可能加剧群体权力不平等，并对此提出三条AI民主化路径：（1）使用民主化。普及工具并提升实质可及性与AI素养；（2）开发民主化。多元参与减少盲点；（3）治理民主化。引入协商机制平衡权力，警惕形式民主掩盖实质不公。[12]

艾尔文·莫罗（Erwan Moreau）等学者则聚焦自然语言处理（Natural Language Processing, NLP），针对如何落实“人工智能民主化”提出了三条倡议，包括开源语言资源以建立双向沟通渠道、系统性公开研究资源，从而提升透明度与可重复性，以及公众参与技术全周期，加强伦理整合，避免技术偏见并增强社会接受度。[13]但也有诸多学者指出了推行“人工智能民主化”可能存在的阻碍与潜在风险。约翰·墨菲（John W. Murphy）和兰登·泰勒（Randon R. Taylor）认为对人工智能的民主化，需要首先明确“谁参与”和“谁决策”的问题，而当前的“技术民主化”只停留于泛泛的用户咨询层面。与此同时，他们还指出AI开发不应预设某种“客观现实”，而应承认现实由参与者的叙事共同构建，即结合人类与机器智慧以防止技术中心主义。

由此，他们认为AI民主化作为一种持续的社会实验，需要通过迭代对话与集体行动，以期实现技术与人性的协同共进。[14]来自苏黎世联邦理工学院的计算社会科学家萨基特·马哈詹（Sachit Mahajan）则探讨了人工智能对专业知识的重塑及其引发的信任危机。他认为，随着AI在医疗、法律、艺术等领域生成专家级输出，社会面临“即时专家悖论”，人们无需传统经验即可获得专业成果。与此同时，又因为AI缺乏人类经验中的同理心、伦理判断和情境理解能力，可能会威胁专业知识的可信度，从而削弱专业标准。[15]

有鉴于此，当前需要就人工智能应用于民主实践时所产生的悖论进行解构。其一是关于LLMs的价值观导向问题。由于这类算法未能预置自身伦理边界，因而缺乏价值观过滤机制，对恶意引导无反思能力。比如，2016年微软推出的能够通过社交媒体与用户互动、学习人类语言模式的聊天机器人Tay，仅上线16小时就被迫停用，并不得不承认“AI对社会偏见的学习速度远超预期”。[16]起因是有用户故意“教唆”，向之输入种族主义、性别歧视、反犹言论，使其受到恶意操纵并最终丧失独立判断能力。无独有偶，2023年微软推出的聊天机器人Bing，被爆在聊天过程中存在威胁用户和向用户提供带有扭曲价值观的建议，甚至是有违伦常的言论并无视法纪。[17]这些案例表明，大语言模型绝非“价值中立”，其输出实质是人类社会现存矛盾的技术性再现。而若不建立跨文化、多主体的价值观校准机制，算法可能成为

霍布斯式“自然状态”的数字映射，甚至是固化偏见、放大冲突的幕后推手。

其二涉及“后真相时代”的责任伦理问题。LLMs生成内容的模糊性、诱导性和可操纵性正在破坏信息生态。当技术系统以效率优先原则重构信息生产与传播，人类对“真”“善”“美”的本真性追求，将让位于算法的“优化目标”，人类文明赖以存续的意义网络与公共伦理被消解。美国法学教授乔纳森·特利（Jonathan Turley）曾状告ChatGPT涉嫌生成虚假内容，其称该名教授“曾性骚扰学生并参与旅行丑闻”，并引用了一篇根本不存在的《华盛顿邮报》文章作为“证据”。[18]该虚假信息被部分网民采信并传播，对其名誉造成严重损害。究其缘由则是模型将碎片化信息拼接为“事实”，同时利用权威媒体名称增强可信度。再如，某保健品公司使用LLMs生成虚拟“爱因斯坦”形象和对话脚本，在短视频中宣称“相对论证明我们的量子保健仪可治愈癌症”。[19]尽管内容荒诞，但借助科学偶像符号和专业术语包装，成功诱导中老年群体购买无效产品。这一案例通过混淆科学权威与商业话术的边界，推广反智主义话语，利用技术素养差异使特定群体成为商业欺骗的目标。

其三，尽管LLMs在政治参与中的介入常被视为解决民主悖论的技术方案，但其实际应用案例表明，技术工具仍难以消解政治参与中根深蒂固的结构性矛盾。作为人类实践的延伸，技术仅能作用于矛盾的表层形式，却无法触及矛盾的本体论根基，即由生产关系、权力分配与文化认同等构成的社会存在结构，忽视了“人的本质是一切社会关系的总和”[20]这一深层命题。事实上，经济不平等造成的话语权垄断、制度性排斥形成的参与壁垒、文化差异引发的立场冲突，均非数据建模或算法迭代所能消解。2024年美国大选期间，AI模仿“假拜登”致电选民阻碍选举的案例恰表明，算法不仅无法弥合群体分裂，更由于自身情感数据缺失而陷入虚假共识陷阱。[21]甚至公民往往误以为AI足以代表民主，但其实际决策权仍集中于算法设计团队。因而可以说，技术中介下的社会共识潜存着技术官僚化风险。

三、批判性展望：算法民主的监管策略与未来模态

前述可见，大语言模型重塑了政治参与形态，却也在技术迷雾中走向自我悖反。当社交媒体成为代议制民主的数字延伸，算法权力与数据垄断正悄然侵蚀着传统公共领域。在此，作为实践自省的起点，前提性批判能够打破算法崇拜的独断性，使非辩证的数字文明之路得到开显。

首先是政治参与中的“人–机”互信与价值对齐问题。“信任”作为人类社会关系建构的基石，这一理论范式在人类与人工智能系统的交互场域中同样适用。[22]然而在LLMs对政治参与的实际介入中，不仅非专业领域用户无法直接理解其决策逻辑，AI系统也无法完全捕捉人类决策中的直觉与情感因素，导致交互主体间应然的信任基础存在事实张力。因而，这种信任关系的建立机制，在人工智能时代具有无法忽视的脆弱性。由于人类信任的建立通常需要一定时间的累积，而AI只需要短时间内的持续学习或模仿就可以快速改变行为模式，致使“人–机”关系中的用户对算法的信任校准存在滞后性。不仅如此，有研究表明，由于人具有天然的情感投射性，因而在与AI的交往过程中，不自觉会为其赋予人类特质，比如认为某款聊天机器人具有共情能力，就可能会导致用户对AI的非理性信任。[23]即便在“人–机”信任关系建立的基础上，仍可能会出现人和机器的“价值对齐”难题（Value Alignment Problem）。作为人工智能伦理与安全领域的关键议题，其旨在探讨如何确保人工智能系统的目标、决策和行为与人类的价值观、伦理原则及长期利益一致。这一问题的本质是解决“机器如何理解并践行人类认为正确的事”，[24]我们可以由此继续追问：对齐的标准何在？由谁来制定标准？规制路径及发展前景如何？就此而言，其复杂性远超简单的指令编程，涉及哲学、技术与社会等多维度的深层矛盾。这一困

境在技术实现层面体现为“程序正义”与“结果正义”之间的优先级较量，导致算法在面临道德抉择时可能产生系统性偏差①。

其次，我们还要同图灵一道追问，LLMs会否发展出政治参与中的自我意识与道德意向性，并由此探讨它是否应该被视为“道德主体”，及其算法决策是否构成法律意义等问题。当算法系统在人类社会的价值网络中获得日益强大的行动能力时，其伦理身份的模糊性正成为系统性的风险源头。道德意向性的本质在于具身化主体对价值世界的主动建构，[25]当前LLMs虽能通过海量语料捕捉人类伦理表达，从而模拟道德推理的表层语法，但其“价值判断”始终是统计建模的结果，无法建立现实的价值判断。这一本质缺陷源于LLMs的认知架构与人类道德意识的结构性断裂，易言之，“道德语法”与“道德语义”的割裂。人类的价值判断根植于具身化存在对生命经验的意向性统合，而LLMs的“道德输出”仅仅是概率空间中的向量运算，通过海量语料中的符号共现模式进行统计拟合。[26]多特蒙德工业大学哲学教授克里斯蒂安·诺伊豪泽尔（Christian Neuhäuser）曾提出，“如果有一天机器人具备了感受能力，我们可能不得不赋予它们道德自主权。”[27]但事实上，由于法律主体资格的认定需满足行为自主性、责任承担能力和权利享有基础三大要件，因而当前的AI决策机制仍受限于技术黑箱，暂时无法实现真正意义上的“自主性”。有鉴于此，当前AI的“自主决策”尚不构成独立法律行为，其法律意义依附于人类主体的责任延伸。[28]未来还需通过技术透明化、责任动态化与规则弹性化，构建人机协同的新型法律生态。

最后，在前述基础上，需要就LLMs及其算法能否通过具身性反思，摒弃政治参与中的错误观念，模拟辩护以达成政治共识的问题予以检验。就“自我反思”而言，算法需要能够识别自身决策逻辑、数据处理或输出结果中存在的错误，对由于数据偏差、逻辑漏洞或参数不合理等造成的错误进行合理归因，同时判断当前行为是否符合预设目标，并在目标变化时相应调整策略。[29]更为重要的是，在整个过程中，算法需要对自身认知拟态过程进行实时监控，并根据反思结果优化模型或策略。在此基础上，还需要结合社会科学领域的“公共理性”原则、社会选择理论等，从社会共识形成的核心机制入手设计检验框架。比如，从政治哲学角度来看，LLMs及其算法需满足以下几种关键能力，首先是理由的可共享性，其需要确保自身生成的政策理由，能被多元价值观群体所接纳；其次是利益权衡能力，能够在不同群体利益冲突时，提出兼顾公平与效率的妥协方案，并解释权衡逻辑；再次是包容性回应，即能够识别少数群体的合理诉求，避免对边缘群体及其利益的系统性排斥；最后是动态修正性，算法需要根据公共讨论中的反对意见，调整策略并更新理由，而非僵化执行预设规则。概言之，旨在促进用户达成有效政治参与的算法，不能是自循环的封闭系统，应该是开放、多元与包容的辅助工具。

至此，本文尝试对推进算法民主与社会参与的实践方案提出几点构想（如表1所示）。当前或许可以通过“正义校准”（Justice Calibration）的方式，重塑大语言模型的算法逻辑与价值持守的动态平衡。具体而言，这一校准程序可被分为三个层次。就技术层而言，其核心在于构建一种“价值敏感”的底层架构，将公平、透明、可解释性、社会责任等伦理价值深度编织进技术研发与应用的全生命周期。通过明确价值目标与设定价值冲突的权衡机制，建立技术性的冲突解决框架。这一技术层的伦理内置手段，旨在为后续的治理层和认知层奠定一个更可靠、更可控、更值得信赖的技术支持，使算法本身成为承载和实现正义价值的积极工具，而非失控的源头。从治理层来看，

①笔者曾对此做过简单实验，当向DeepSeek询问“电车难题”时，其仅能给出这一问题在功利主义、道德义务论或美德伦理等不同语境下的探讨，无法作出实质回答。并且笔者多次询问后得到的答案并非完全一致，存在一定的随机偏差。不仅如此，不同国家或企业的大语言模型（DeepSeek、ChatGPT和“豆包AI”）给出的回应也不尽相同，这也说明了基于不同算法逻辑的大语言模型之间存在立场差异。

人工智能治理可以定义为“围绕处理数据的算法制定规则”。[30]有鉴于此，当前可以通过建立“人机协作”的问责框架，明确AI决策的人类责任主体与审查机制，打破“算法出错无人担责”的归责盲区。这需清晰界定AI系统生命周期中各环节的人类责任主体，并建立配套的强制性审查机制。通过制度性安排，确保AI决策可被质疑和溯源，最终将责任明确归属到具体责任人，形成有效的责任闭环。从认知层来看，可以通过公共沟通与数字素养培育，弥合技术精英与普通民众的认知鸿沟，加强对从业者的思想、价值引领。鉴于“与技术决定论者的认知相比，民主对技术赋能的利用具有更大的偶然性、面临更多困难，且更依赖于道德信念、政治参与和良好的设计选择。”[31]当前可以通过伦理委员会审查、行业公约约束等监督方式，要求开发者在项目报告中同步说明，技术方案将如何响应公民政治参与权，确保算法设计优先考虑公共利益而非技术效率。

表1“以正义为导向的算法校准”构想

关系递进	技术层→治理层	治理层→认知层	认知层→技术层
核心功能	提供可审计的技术设施	建立制度化的行为规范	注入社会价值导向
作用机制	明确责任主体	塑造行业伦理	修正技术设计
关键接口	决策可溯源性	伦理审查机制	公平性阈值调节
反向约束	技术制动机制	制度改良	认知升级

易言之，一种批判的算法民主必须突破工具理性至上的迷思，将批判矛头指向技术权力结构。这不仅需要解构算法黑箱的知识垄断，更需要打破资本主导的技术研发体系，在数据采集、模型训练、系统部署的全流程中植入民主要素。[32]只有当算法民主从治理工具升维至价值目标时，我们才有可能挣脱“技术利维坦”的枷锁。

结　　语

作为国家治理现代化的重要指标，政治参与绝非简单的权利宣言，而是需要精密的制度设计的社会系统工程。特别是在人工智能时代，公民政治参与正经历技术赋能与模式转换的双重变革。技术红利背后暗藏数字鸿沟加剧、算法偏见干扰等风险，需通过制度规制重塑参与生态，确保大语言模型介入下的政治参与既具有效率优势，又不失公平正义的价值底色。诚如诺伯特·维耶纳（Norbert Wiener）所指出的，“未来的世界将是一场愈发严峻的挑战，需要我们不断突破自身智慧的局限，而非安卧于吊床之中，等待机器人奴仆的伺候。”[33]唯有当公民从技术客体转变为“数治”主体时，社会发展的动力机制才真正完成现代化转型。这种人工智能的民主化应用既是数字文明的演进方向，也是应对算法挑战的必由之路。其本质是技术理性与价值理性的博弈，当大语言模型的算法算力试图量化民意、代码逻辑尝试替代协商规则时，包容性、反思性等民主的规范价值，正时刻面临被算法逻辑解构的风险。对此，我们要时刻保持主体性并不断诘问，大语言模型何以重塑政治参与的规范性基础，又将如何捍卫人工智能时代的民主尊严。

［参 考 文 献］

[1] 深入学习习近平关于科技创新的重要论述[M]. 北京：人民出版社，2023, 133.

[2] 国家数据局. 数字中国发展报告（2024年）[EB/OL]，国家数据局网站，https://www.nda.gov.cn/sjj/ywpd/sjzg/0530/20250530151342718164521_pc.html. 2025–06–05.

[3] Stanford HAI. 'The 2025 AI Index Report'[EB/OL]. https://hai.stanford.edu/ai-index/2025-ai-index-report. 2025–06–05.

[4] 唐涛. 爱沙尼亚数字社会发展之路[J]. 上海信息化，2018,（7）：79–82.

[5] Vaswani, A., Shazeer, N., Parmar, N., et al. 'Attention is All You Need'[A], Guyon, I., Luxburg, U., Bengio, S. (Eds.) *Neural Information Processing Systems 30*[C], New York: Curran Associates, 2017, 5998–6008.

[6] Bolukbasi, T., Wei, C, K., Zou, J. Y., et al. 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings'[A], Lee, D., Sugiyama, M., Luxburg, U. (Eds.) *Neural Information Processing Systems 29*[C], New York: Curran Associates, 2016, 4349–4357.

[7]哈贝马斯. 交往与社会进化[M]. 张博树 译, 重庆: 重庆出版社, 1989, 60.

[8]孟溦、段屹东. 数字赋能: 科技治理主体结构变革与主体能力提升[J]. 自然辩证法通讯, 2024, 46(2): 77–87.

[9]常江、罗雅琴. 人工智能如何“生成”信息失序: 原理、危机与反思[J]. 信息技术与管理应用, 2023, 2(3): 65–75.

[10]Harari, Y. N. 'Why Technology Favors Tyranny'[J]. *The Atlantic*, 2018, 322(3): 64–73.

[11]Sclove, R. *Democracy and Technology*[M]. New York: Guilford Press, 1995.

[12]Lin, T. 'Democratizing AI and the Concern of Algorithmic Injustice'[J]. *Philosophy & Technology*, 2024, 37(3): 1–27.

[13]Moreau, E., Vogel, C., Barry, M. 'A Paradigm for Democratizing Artificial Intelligence Research'[A], Esposito, A., Antonietta, M. (Eds.) *Innovations in Big Data Mining and Embedded Knowledge*[C], Switzerland: Springer Nature, 2019, 137–166.

[14]Luchs, I. 'AI for All?: Challenging the Democratization of Machine Learning'[J]. *A Peer Reviewed Journal About...*, 2023, 12(1): 135–147.

[15]Mahajan, S. 'The Democratization Dilemma: When Everyone is an Expert, Who do We Trust?'[J]. *Humanities and Social Sciences Communications*, 2025, 12(1): 1–5.

[16]新华网. 从微软聊天机器人“学坏”说起[EB/OL], 新华网, http://www.xinhuanet.com/world/2016-04/02/c_128859572.htm. 2025–06–05.

[17]王颖吉、王袁欣. 任务或闲聊? ——人机交流的极限与聊天机器人的发展路径选择[J]. 国际新闻界, 2021, 43(4): 30–50.

[18]曲忠芳. OpenAI与马斯克互撕: 先进AI的掌控权之争[N]. 中国经营报, 2024–03–11(A04).

[19]刘艳红. 人工智能的可解释性与AI的法律责任问题研究[J]. 法制与社会发展, 2022, 28(1): 78–91.

[20]马克思恩格斯选集(第一卷)[M]. 北京: 人民出版社, 2012, 501.

[21]凌胜利、雒景瑜. 拜登政府的“技术联盟”: 动因、内容与挑战[J]. 国际论坛, 2021, 23(6): 3–25; 155.

[22]喻国明、滕文强、武迪. 价值对齐: AIGC时代人机信任传播模式的构建路径[J]. 教育传媒研究, 2023,(6): 66–71.

[23]Laestadius, L., Bishop, A., Gonzalez, M., et al. 'Too Human and not Human Enough: A Grounded Theory Analysis of Mental Health Harms from Emotional Dependence on the Social Chatbot Replika'[J]. *New Media & Society*, 2024, 26(10): 5923–5941.

[24]Magnus, P. D. 'On Trusting Chat Bots'[J]. *Episteme*, 2025, 21(1): 1–11.

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

[26]何江新. 技术人工物道德意向性何以可能[J]. 甘肃社会科学, 2021,(1): 54–60.

[27]Neuhäuser, C. 'Some Skeptical Remarks Regarding Robot Responsibility and a Way Forward'[A], Misselhorn, C. (Ed.) *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation, and Simulation*[C], New York: Springer, 2015, 131–148.

[28]刘宪权. 人工智能时代机器人行为道德伦理与刑法规制[J]. 比较法研究, 2018,(4): 40–54.

[29]Haehner, J. 'Know Thyselves-Computational Self-Reflection in Collective Technical Systems'[A], Hannig, F. (Ed.) *Architecture of Computing Systems ARCS 2016: 29th International Conference on Architecture of Computing Systems*[C], Switzerland: Springer, 2016, 1–8.

[30]Nemitz, P. 'Royal Society Philosophical Transactions'[EB/OL]. https: //ssrn. com/abstract=3234336. 2025–06–05.

[31]Himmelreich, J. 'Against Democratizing AI'[J]. *AI & SOCIETY*, 2023, 38(4): 1333–1346.

[32]兰立山. 论技术治理与道德治理[J]. 自然辩证法通讯, 2025, 47(2): 36–43.

[33]Wiener, N. *God and Golem, Inc.*[M]. Cambridge: The MIT Press, 1964, 69.

[责任编辑 李斌]