

# 生成式大语言模型对中华民族共同体 认同的潜在风险挑战及对策

陈 辉

(中央民族干部学院 干部教育研究中心, 北京 100094)

**摘 要:** 生成式大语言模型在文化传播、信息获取、价值塑造等方面具有巨大潜力,但也带来了叙事同质化、主体性转移、价值导向隐蔽性等风险,对中华民族共同体认同的文化、社会心理和价值根基带来了一定挑战。为此,需要构建多元融合、权威可信的叙事体系、激发社会参与、强化情感联结、突出价值引领、提升算法伦理,以此来确保生成式大语言模型更好地服务于中华民族共同体认同的构建,从而为实现中华民族伟大复兴的中国梦贡献力量。

**关键词:** 生成式大语言模型; 中华民族共同体认同; 潜在风险; 对策

中图分类号: D633 文献标识码: A 文章编号: 1002-6959 (2025) 04-0107-08

DOI编号: 10.13965/j.cnki.gzmzyj10026959.2025.04.015

## Potential Risks, Challenges, and Countermeasures of Generative Large Language Models to the Sense of Community and Identity of the Chinese Nation

CHEN Hui

(Research Center for Cadre Education, Central Institute for Ethnic Cadres, Beijing 100094, China)

**Abstract:** Generative large language models hold immense potential in cultural dissemination, information acquisition, and value cultivation. However, they also pose risks such as narrative homogenization, subjectivity shift, and hidden value orientation, challenging the cultural, socio-psychological, and value foundations of the Chinese national community identity. To address these challenges, it is essential to construct a diversified and authoritative narrative system that integrates multiple perspectives; stimulate social participation and strengthen emotional bonds among communities; emphasize value guidance while enhancing algorithmic ethics. These efforts will ensure that generative large language models better serve the consolidation of the Chinese national community identity and contribute to the realization of the Chinese Dream of national rejuvenation.

**Key words:** Generative Large Language Models; Generative Large Language Models; Potential Risks; Countermeasures

收稿日期: 2025-05-13

基金项目: 浙江省哲学社会科学重点研究基地嘉兴大学浙江省中国共产党创建史研究中心课题“马克思主义民族理论早期传播与中国化研究(1919-1927)”(项目编号: 2022JDKTYB36)的阶段性成果。

[作者] 陈辉, 中央民族干部学院干部教育研究中心副研究员, 研究方向: 民族理论与民族政策、中华民族史。

中华民族共同体意识是维护国家统一、社会安定、文明传承的精神基石，是推动中华文明永续发展的核心凝聚力。这一理念深刻揭示了中华民族多元一体格局的本质特征，强调中华民族是由多元文化、多元族群融合而成的命运共同体，拥有共同的历史记忆、文化认同和国家认同，致力于共同的中华民族复兴伟业。“以统一多样的文化共建凝聚全民族的心理共同体认知，是铸牢中华民族共同体意识的强大后盾”<sup>[1]</sup>。

近年来，以生成式大语言模型（Generative Large Language Models, LLMs）为代表的人工智能技术迅猛发展，并在各领域展现出革命性潜力。学术界对LLMs的关注日益提升，研究主要集中在其技术原理、应用前景以及潜在的社会伦理影响等方面。例如，朱永新、杨帆、陈永伟、骆飞、马雨璇等学者研究探讨了LLMs在教育、传媒、科研等领域的创新应用，认为其能极大提升信息处理效率和知识生产能力。同时，亦有大量研究关注LLMs可能带来的风险，如信息真实性难以保障、算法偏见固化社会刻板印象、个人隐私泄露以及对就业市场的冲击等<sup>[2]</sup>。在国家认同与文化传播领域，已有学者开始探讨数字媒介对国家形象建构和文化认同塑造的双重影响<sup>[3]</sup>。但专门针对生成式大语言模型对中华民族共同体认同这一特定问题的系统性研究尚显不足。本文正是在此学术背景下，聚焦生成式大语言模型这一新兴技术，深入分析其对中华民族共同体认同可能造成的潜在风险，并尝试提出针对性的应对策略，以期为人工智能时代更好地铸牢中华民族共同体意识提供理论参考与实践路径。

生成式大语言模型，指基于深度学习，特别是Transformer等先进神经网络架构，通过在海量文本数据上进行预训练，从而掌握语言的统计规律和语义知识，并具备理解、生成、转换自然语言能力的模型<sup>[4]</sup>。生成式大语言模型作为一种强大的技术工具，在文化传播、信息获取、价值塑造等方面具有巨大的潜力，在深化中华民族共同体认同的同时，客观上存在需要审慎应对的潜在隐患与挑战。这些风险与挑战主要体现在叙事构建的同质化与碎片化、主体性转移与情感弱化、价值导向的隐蔽性与偏颇性等方面，可能对中华民族共同体认同构成深层威胁，既冲击着文明传承

的精神内核，又侵蚀着社会共识的心理基础，更动摇着价值体系的核心支柱。基于此，构建中华民族共同体认同的现代路径，就必须始终坚守人文精神与技术伦理的辩证统一，要从构建具有解释力的复合型叙事框架来考虑，既要打造兼容并蓄的文化阐释体系以夯实认知基础，又要凝聚社会共识以强化心理归属，更要营造向上向善的技术伦理环境以塑造价值共识。

### 一、叙事构建的同质化与碎片化风险：消解中华民族共同体认同的多元文化根基

中华民族共同体认同的构筑，深深植根于对共有历史记忆、文化遗产与价值理念的叙事建构之中。传统上，这种叙事是由国家、社会、文化精英共同塑造和传播的，其特点在于相对的权威性、系统性和连贯性。国家通过教育体系、主流媒体、文化政策等渠道去宣传中华民族的一体性和共同性，构建宏大叙事；社会各界则从不同角度、不同层面丰富和补充这一叙事；文化精英则通过学术研究、文艺创作等形式，深化对民族历史文化的理解和阐释。这种传统叙事构建模式，虽然存在一定的中心化倾向，但保证了民族共同体认同叙事的相对稳定性和统一性，为社会成员提供了共同的价值参照和文化归属。然而，生成式大语言模型等人工智能技术的介入，正在改变这种传统的叙事构建模式，并带来叙事同质化和碎片化的风险，进而对中华民族共同体认同的多元文化根基形成潜在消解隐患。

（一）叙事生产的算法化与标准化：弱化多元文化表达，导致文化认同的同质化风险

生成式大语言模型的内容生成能力，其核心运作机制在于通过深度学习技术对超大规模文本语料库进行统计规律提取与模式化复现。其训练数据虽然庞大，但仍不可避免地存在数据偏向性，即高频、主流、易于量化的数据更容易被模型学习和吸收。这种数据偏向性，直接导致生成式大语言模型生成的内容可能更倾向于标准化、同质化的叙事模式。对比传统叙事构建模式，生成式大语言模型的算法化叙事生产存在以下理论风险：

一是数据偏向性导致文化选择性呈现。生成式大语言模型的训练数据可能更多来源于官方媒体、主流出版物、网络热点等渠道，这些数据本身就带有一定的价值导向和文化偏向，例如

更侧重于宏大叙事、主流文化、中心区域文化，相对忽视了地方文化、少数民族文化、民间文化等。长期使用生成式大语言模型获取文化信息，用户接触到的中华民族文化图景可能被算法所“筛选”和“过滤”，形成一种被简化和同质化的文化认知，从而弱化了对中华民族多元文化特性的理解和认同。

二是算法逻辑的标准化倾向压制文化多样性。生成式大语言模型的内容生成逻辑是基于统计概率和模式匹配，其目标是生成符合“大众口味”和“普遍规律”的内容，从而弱化了对文化多样性及差异性的呈现能力。这种算法逻辑的标准化倾向，与中华民族多元文化并存的现实存在内在张力，长期发展可能导致文化表达的单一化和同质化，冲击中华民族共同体认同的文化多样性支撑。

三是技术工具理性凌驾于文化价值理性<sup>[5]</sup>。生成式大语言模型作为一种技术工具，其核心目标是提升效率、优化性能、满足用户需求。在文化内容生产领域，这种工具理性可能凌驾于文化价值理性之上，即更关注内容的生产效率、传播效果和商业价值，而相对忽视内容的文化内涵、价值导向和社会责任。长期过度倚重生生成式大语言模型进行文化信息获取，可能使工具理性渗透文化认知维度，遮蔽其作为精神共同体纽带的人文意蕴，最终消解中华民族文化认同的深层价值根基。

(二) 叙事传播的去中心化与碎片化：消解权威叙事，导致认同的离散化风险

传统上，中华民族共同体认同的叙事传播，主要依赖于国家主导的主流媒体和文化机构，其特点在于相对的中心化和权威性。这种中心化的传播模式，虽然存在一定的信息垄断和价值灌输风险，但保证了民族共同体认同叙事的有效传播和广泛覆盖。然而，生成式大语言模型等人工智能技术的普及，正在打破这种中心化的传播格局，“必然会对信息秩序和新闻传播产生翻天覆地的改变”<sup>[6]</sup>，使得叙事传播更加去中心化和碎片化，带来认同离散化的风险。

一是信息茧房效应加剧认同分化。生成式大语言模型的个性化推荐机制，能够精准捕捉用户的兴趣偏好与行为轨迹，为用户量身推送定制化

的信息内容。然而，这种技术获取信息，能导致用户不知不觉陷入“信息茧房”的困境，难以触及多元化的声音和不同文化的碰撞。在中华民族共同体认同构建这一宏大命题下，更要警惕可能导致不同群体、不同地域、不同文化背景的用户，在信息的单向度浸润中，逐渐形成彼此隔绝的认知闭环，进而加剧社会的碎片化倾向，削弱中华民族多元一体的文化凝聚力。

二是虚假信息和不良内容冲击主流认同。一方面，生成式大语言模型的内容生成能力，可能被用于生产和传播虚假信息、煽动性言论、文化糟粕等不良内容，另一方面，“虚拟开放的网络空间为公众提供了随性表达思想、释放情感、宣泄情绪的平台”<sup>[7]</sup>。这种现象通过算法放大负面内容传播，形成对主流价值认同的认知侵蚀。在社交媒体算法的助推下，虚假信息、极端言论和文化糟粕首先通过精准推送突破认知屏障，继而利用情绪共振实现裂变扩散，最终在多元群体间制造认知鸿沟。

三是权威叙事被解构与消解。生成式大语言模型的普及，使得普通用户也能够参与民族叙事的生产和传播中，打破了传统文化精英对叙事权的垄断。这种叙事权的下放，虽然具有一定的民主化意义，但也可能导致权威叙事被解构和消解。在缺乏有效引导和监管的情况下，各种碎片化、个体化甚至错误的历史文化解读，可能在网络空间泛滥，在个体层面造成文化身份的困惑，在群体层面形成共识基础的动摇，最终威胁整个文化共同体的内在统一性。

(三) 构建多元融合、权威可信的中华民族共同体叙事体系

针对生成式大语言模型可能导致的叙事同质化和碎片化风险，我们需要着力构建一个多元融合、权威可信的中华民族共同体叙事体系，既要充分利用生成式大语言模型的技术优势，提升叙事传播的效率和覆盖面，也要有效规避其算法偏见和信息茧房效应，构建多元包容、动态平衡、具有公信力的叙事框架，确保不同文化视角得以呈现，各方观点形成理性对话，主流价值获得权威支撑。

一是强化多元文化数据采集与算法优化，提升生成式大语言模型叙事内容的多样性与平衡性。

要从源头上解决生成式大语言模型叙事内容同质化的问题,必须着力强化多元文化数据采集,并对算法进行优化,提升其叙事内容的多样性与平衡性。

第一,扩大数据来源,构建多元文化知识图谱。打破数据来源的单一性,除了官方媒体、主流出版物、网络热点等传统数据资源外,要积极拓展数据采集渠道,纳入更多的地方文化资源、少数民族文化资源、民间文化资源、非物质文化遗产资源等,构建一个涵盖中华民族多元文化的知识图谱。通过知识图谱的构建,可以更系统、更全面地组织和管理多元文化数据,为生成式大语言模型的内容生成提供更丰富、更均衡的数据基础。

第二,优化算法模型,构建具备文化感知能力的智能算法框架。针对算法模型可能存在的文化偏见和标准化倾向,要进行算法优化,提升其文化敏感性与包容性。一方面,可以引入文化多样性指标,对训练数据进行文化多样性评估和平衡;另一方面要继续优化改进算法模型,使其能够更好地理解和处理不同文化语境下的语言表达和文化内涵;最重要的还是要引入人工审核机制,对生成式大语言模型生成的文化内容要进行人工审核和校对,以确保文化的准确性和价值导向的正确性。

第三,鼓励用户参与,共建共享多元文化内容库。发挥社会力量,鼓励用户参与到多元文化内容的生产和贡献中来,共建共享一个开放、协作、动态更新的多元文化内容库。建立用户贡献平台,鼓励用户上传地方文化素材、分享民族文化故事、创作民间文化作品等;同时,可以开展文化众筹活动,支持用户进行文化创作和文化遗产项目;最后还要建立用户反馈机制,不断改进和完善内容库。

二是加强权威叙事内容建设与精准传播,提升生成式大语言模型叙事的权威性与引导力。在强调叙事多元性的同时,也要重视权威叙事内容建设,并利用生成式大语言模型的技术优势进行精准传播,提升权威叙事的权威性与引导力,确保能够彰显中华民族共同体认同的核心价值观能够得到有效传播和广泛认同。首先要建设权威的中华民族共同体叙事知识库,整合权威的历史

文献、文化典籍、政策文件、专家解读等资源,建设一个权威的中华民族共同体叙事知识库。知识库的内容要涵盖中华民族共同体的历史渊源、文化特征、价值理念、发展成就、面临挑战等各个方面,形成一个系统、完整、权威的叙事体系。再者,要利用生成式大语言模型技术进行权威叙事内容创新呈现。既要权威叙事内容进行创新呈现,又要利用生成式大语言模型的多语言能力,将权威叙事内容翻译成多种语言,进行国际传播,提升中华民族共同体认同的国际影响力。更为重要的是要实施精准传播策略,利用生成式大语言模型的个性化推荐和精准推送功能,将权威叙事内容精准推送给目标用户群体,提升其覆盖面和影响力。可以考虑推送定制化、个性化的权威叙事内容;也可以利用社交媒体、网络平台等渠道,进行权威叙事内容的广泛传播;还可以开展线上线下相结合的宣传活动,提升权威叙事内容在社会传播中的深度和广度。

三是建立健全信息甄别与辟谣机制,防范虚假信息和不良内容对认同的冲击。针对生成式大语言模型可能被用于传播虚假信息和不良内容,冲击主流价值观的风险与挑战,必须建立健全信息甄别与辟谣机制。

第一,构建智能化的虚假信息识别与预警系统。利用人工智能技术,构建智能化的虚假信息识别与预警系统,自动甄别虚假信息、煽动性言论及文化糟粕等不良内容,并依托自适应学习机制持续优化鉴别阈值,与现有内容审核体系形成协同治理网络,为网络空间治理提供技术支撑,助力构建清朗的网络生态环境。

第二,建立权威的辟谣平台和快速响应机制。建立权威的辟谣平台,整合权威媒体、专家学者、政府部门等资源,对网络谣言和虚假信息进行快速辟谣和澄清。平台要具备快速响应机制,能够在第一时间对热点谣言进行辟谣,发布权威信息,消除社会恐慌和误解。辟谣信息要权威、准确、及时、易懂,并利用多种传播渠道进行广泛传播,提高辟谣效果。

第三,提升公民媒介素养和信息辨别能力。加强公民媒介素养教育,提升公民的信息辨别能力和批判性思维能力,使其能够理性对待网络信息,不信谣、不传谣,自觉抵制虚假信息和不良

内容。教育内容要涵盖信息甄别方法、网络安全知识、法律法规意识等方面，通过学校教育、社会宣传、媒体引导等多种途径，提高公民的媒介素养水平。

## 二、主体性转移与情感弱化风险：动摇中华民族共同体认同的社会心理根基

中华民族共同体认同的构建，不仅需要共同的叙事，更需要社会成员的主动参与和情感投入。民族认同的主体性建构在文化实践维度表现为社会成员通过符号系统的习得、传承与创造性转化实现文化主体的在场；在历史认知维度体现出依托集体记忆的承续与批判性反思构建时间向度的主体连续性；同时还要在共享的价值预设与责任伦理中完成空间向度的主体间性确证。这种主体性是建立在社会互动、文化实践和情感交流的基础之上的，具有深厚的社会心理根基。然而，生成式大语言模型等人工智能技术的介入，正在改变民族认同的主体性，并可能带来主体性转移和情感弱化的风险，从而动摇中华民族共同体认同的社会心理根基。

### （一）认同构建的主体性从“人”向“技术”转移：弱化社会互动，导致认同的工具化风险

中华民族共同体认同的构建是一个社会化的过程，需要通过家庭教育、学校教育、社会交往、文化活动等多种途径，实现文化知识的内化、价值观念的认同、情感联系的建立。在这个过程中，人与人之间的互动交流、情感体验、共同实践，发挥着至关重要的作用。然而，生成式大语言模型等人工智能技术的普及，使得民族认同的构建过程，越来越依赖于人机交互，主体性可能从“人”向“技术”转移。

一是社会互动的减少导致认同的情感维度弱化。生成式大语言模型可以为用户提供便捷的文化信息查询、个性化的文化内容生成、智能化的文化学习辅导等服务。但文化实践的虚拟化促使生成式大语言模型可以模拟各种文化场景、再现各种文化仪式、生成各种文化产品，为用户提供虚拟的文化体验。这种虚拟的文化实践，与真实的社会文化生活存在本质差异。真实的文化实践是在具体的社会关系和社会情境之中进行的，能够促进社会成员之间的互动交流、情感联结和价值共享，从而巩固民族认同的社会根基。而虚拟

的文化实践，则可能导致认同的“虚拟化”，即认同感脱离现实社会文化生活，缺乏扎根于地方传统和社群实践的根基，变得空洞和悬浮。

二是批判性思维的钝化导致认同的内省能力下降。生成式大语言模型的智能推荐和个性化定制功能，虽然能满足用户的个性化需求，但也可能限制用户接触不同观点和多元文化。长期依赖生成式大语言模型获取文化信息，可能导致用户的批判性思维能力下降，难以对民族文化进行独立思考 and 理性反思，容易被算法所“塑造”和“引导”，从而削弱民族认同的内省能力和发展活力。

### （二）认同情感的“技术化”与“工具化”：消解情感温度，导致认同的冷漠化风险

中华民族共同体认同，不仅是一种认知认同，更是一种情感认同。民族情感的培养和激发，主要依赖于共同的历史记忆、文化传统、价值观念，以及社会成员之间的情感交流、共同经历和命运与共的意识。然而，生成式大语言模型等人工智能技术的介入，可能导致民族认同情感的“技术化”和“工具化”，消解情感温度，带来认同的冷漠化风险。对比传统情感认同培养模式，生成式大语言模型的技术化情感塑造存在以下理论风险：

一是情感表达的算法化模拟导致情感的空洞化。生成式大语言模型可以模仿人类的情感表达方式，生成各种情感化的文本、图像、音频等内容，甚至可以进行情感化的对话交流。但在某种意义上，这种算法模拟的情感表达本质上是基于数据和模型的计算，缺乏真实的情感体验和情感内涵，可能导致情感的空洞化和形式化，从而削弱对中华民族认同的情感深度和温度。

二是情感联结的技术中介化导致情感的疏离化。生成式大语言模型可以通过网络平台、社交媒体等渠道，促进社会成员之间的文化交流和信息共享，甚至可以提供虚拟的社交互动平台。但究其实质，这种技术中介的情感联结与真实的人际情感交流存在本质差异。真实的人际情感交流是建立在面对面的互动、身体语言的传递、非语言信息的交流等基础之上的，具有丰富的情感内涵和情感体验，能够促进情感的共鸣和理解，增强情感的联结和信任。而技术中介的情感联结，则可能导致情感的疏离化和表面化，难以建立深

层的情感联系和情感共同体。

三是情感价值的工具理性化导致情感的功利化。生成式大语言模型作为一种技术工具，其应用目标往往是功利性的。在民族认同情感塑造领域，这种工具理性可能导致情感价值的工具化和功利化，即更加关注情感的工具价值和实用价值，而忽略了情感的内在价值和人文价值。长期使用生成式大语言模型进行情感交流和情感表达，可能导致用户对情感的理解和认知也受到工具理性的影响，更加关注情感的功利性目的。

(三) 激发社会参与，强化情感联结，提升认同的社会心理根基

针对生成式大语言模型可能导致的主体性转移和情感弱化风险，我们需要着力激发社会参与，强化情感联结，提升中华民族共同体认同的社会心理根基，确保认同的构建过程始终以人为本，充满活力和温度。

一是鼓励社会文化实践与互动交流，强化认同的社会参与度和实践性。在民族认同的构建过程中，应警惕对技术中介的过度依赖，应更加注重通过社会文化实践和互动交流来增强民族认同感，强化认同的社会参与度和实践性，让社会成员在真实的社会文化生活中体验和感受中华民族共同体认同。首先，支持和鼓励开展线下文化活动和社区建设。加大对线下文化活动和社区建设的支持力度，鼓励开展各种形式的文化节庆、民俗活动、社区文化活动、志愿服务活动等。其次，搭建线上线下相结合的文化交流平台。利用生成式大语言模型的技术优势，搭建线上线下相结合的文化交流平台，促进社会成员之间的文化交流和互动。最后，创新文化传承方式，构建代际文化传承的双向互动模型。要以具身认知理论为基础，通过设计跨代际参与式文化实践场景，实现传统技艺的活态传承与文化记忆的协同重构。

二是强调情感价值，构建富有人文关怀的认同叙事，提升认同的情感温度和价值内涵。要避免民族认同情感的“技术化”和“工具化”，强调以情感价值为依托，构建富有人文关怀的认同叙事，提升认同的情感温度和价值内涵，让认同更具人情味、更具感染力、更具生命力。在生成式大语言模型的内容生成过程中，要注重融入情感元素和人文关怀，使其生成的文化内容更具情感

温度和人文价值。在此基础上，还要鼓励用户利用生成式大语言模型平台进行情感表达和价值分享，促进用户之间的情感交流和价值认同。加强人文教育和情感教育，也是提升公民的情感素养和价值认知的重要途径，教育内容要涵盖情感认知、情感表达、情感管理、价值判断等方面，通过学校教育、家庭教育、社会教育等多种途径，提高公民的情感素养和价值认知水平。

三是必须培养公民的批判性思维和独立思考能力。要避免中华民族认同的构建过程过度依赖技术引导，就必须在教育体系中加强批判性思维和信息素养培养，从基础教育阶段开始，就要注重培养学生的独立思考能力，使其能够理性对待网络信息，独立思考文化问题，批判性地审视民族认同叙事。同时，引导用户理性使用生成式大语言模型，避免过度依赖和信息茧房。这不仅鼓励用户多渠道获取信息，多角度思考问题，保持对不同观点和多元文化的开放态度，也避免陷入算法的信息茧房，限制认知视野和思维模式。

三、价值导向的隐蔽性与偏颇性风险：侵蚀中华民族共同体认同的核心价值理念

中华民族共同体认同的深层结构，是文化记忆与价值情感的交织。通过代际传承的“价值遗传”和群体共享的“情感共鸣”，构建起个体与民族的精神纽带，使“中华民族”成为超越个体生命的意义共同体。中华民族共同体认同的核心价值理念，包括爱国主义、集体主义、社会主义核心价值观等，主张“要把各民族优秀传统文化资源中符合社会主义核心价值观的要素提炼出来”<sup>[9]</sup>。这些价值理念的传播和弘扬，主要依赖于国家意识形态的引导、社会主流价值观的塑造、文化榜样的示范引领等，其特点在于相对的明确性和规范性。然而，生成式大语言模型等人工智能技术的介入，可能带来价值导向的隐蔽性和偏颇性风险，从而侵蚀中华民族共同体认同的核心价值理念。

(一) 价值嵌入的算法黑箱化：价值观传递的隐蔽性风险

生成式大语言模型的内容生成和信息推荐，并非都是价值中立的，其算法模型、训练数据、设计理念等都可能内嵌特定的价值导向。然而，这种价值嵌入往往是隐蔽的、难以察觉的。

一是算法偏见导致价值导向的失真。生成式大语言模型的算法模型和训练数据，都可能存在人为的或非人为的偏见，例如数据采集的地域偏见、人群偏见、文化偏见，算法设计的价值预设、意识形态倾向等。这些偏见可能导致生成式大语言模型生成的内容和推荐的信息，带有特定的价值导向，而相对忽视或贬低其他价值观念。

二是技术理性遮蔽价值理性。生成式大语言模型作为一种技术工具，其价值导向往往是技术理性的，即追求效率、优化、创新等技术目标，而相对忽视价值理性，即对社会伦理、道德规范、人文关怀等价值的坚守和弘扬。在价值传递领域，这种技术理性可能遮蔽价值理性，导致生成式大语言模型的应用更侧重于技术功能的实现和商业价值的创造。

三是价值操控的潜在风险。生成式大语言模型的强大内容生成和信息推荐能力，也可能被用于进行价值操控和意识形态渗透。例如某些机构或个人可能利用生成式大语言模型传播特定的政治观点、文化理念、价值观念，甚至进行虚假宣传、舆论引导等活动，对社会公众的价值认知和价值选择造成误导和干扰。

(二) 价值呈现的单一化与功利化：价值多元与核心价值的张力

中华民族共同体认同的价值体系，既包含统一的核心价值理念，也允许价值的多元化表达。一方面，传统上，价值的多元化表达，可以通过文化多样性、思想多元化、社会多元主体等途径实现，既丰富了价值内涵，也增强了价值的社会适应性。另一方面，首先，生成式大语言模型等人工智能技术的介入，可能导致价值呈现的单一化和功利化。生成式大语言模型的算法模型和训练数据，可能导致其价值表达倾向于同质化和标准化，难以充分展现价值的多元性和差异性。其次，技术功利主义侵蚀价值的内在性。生成式大语言模型作为一种技术工具，其应用目标往往是功利性的。长期依赖生成式大语言模型获取价值信息，可能导致用户对价值的理解和认知也受到技术功利主义的影响，更加关注价值的工具价值和实用价值，而忽略其深层的人文价值和道德价值，会对中华民族共同体认同的价值内在性根基产生威胁。此外，价值竞争的算法化加剧价值冲

突。例如不同政治立场、文化背景、价值观念的群体，可能利用生成式大语言模型传播各自的价值主张，争夺话语权和影响力，甚至进行价值攻击和价值诋毁，加剧价值冲突和社会对立。在中华民族共同体认同构建领域，价值竞争的算法化可能导致不同价值观念之间的冲突更加激烈和隐蔽，对中华民族共同体认同的价值共识和价值凝聚力造成威胁。

(三) 强化价值引领，提升算法伦理水平，构建健康价值生态

针对生成式大语言模型可能带来的价值导向隐蔽性和偏颇性风险，要从强化价值引领，提升算法伦理，构建健康的价值生态等入手，确保能够有效地维护对中华民族共同体的自觉认同和自觉弘扬。

一是强化社会主义核心价值观的引领作用，确保生成式大语言模型价值导向的正确性。要确保生成式大语言模型的价值导向与中华民族共同体认同的核心价值理念相一致。首先必须强化社会主义核心价值观的引领作用，树立正确的中华民族观，将其融入生成式大语言模型的设计、开发、应用和监管全过程。在生成式大语言模型的算法架构设计与训练数据优化过程中，要充分考虑社会主义核心价值观的要求，通过价值对齐算法实现模型输出与核心价值体系的语义兼容，确保生成式大语言模型生成的内容和推荐的信息，符合社会主义核心价值观和正确中华民族观的基本导向。其次，要利用好生成式大语言模型技术传播和弘扬社会主义核心价值观。充分利用生成式大语言模型的技术优势，创新社会主义核心价值观的传播方式和传播渠道，提升其传播效果和影响力。可以根据不同用户的需求和偏好，利用生成式大语言模型的个性化推荐和精准的推送功能，将社会主义核心价值观内容精准推送给目标用户群体。此外，还要建立健全对生成式大语言模型价值导向的监管和评估机制，对生成式大语言模型的价值导向进行定期评估和动态监测，及时发现和纠正价值偏差和价值风险。

二是提升算法伦理水平，构建负责任的人工智能，提升生成式大语言模型价值导向的透明度和可控性。这要求制定生成式大语言模型的应用符合伦理规范，“建立科学、明晰的法律责任担当

机制”<sup>[9]</sup>。构建人工智能伦理治理的规范体系，需从价值准则、技术标准与行业规则三个层面展开系统性建设，形成涵盖伦理原则、行为规范与评估机制的治理框架，确保人工智能技术的应用符合伦理要求和社会公共利益。还要构建好人工智能伦理审查和风险评估机制，对生成式大语言模型等人工智能技术的研发和应用进行伦理审查和风险评估，预判和防范潜在的伦理风险和社会风险。

三是促进价值多元共存与核心价值引领的平衡，构建健康和谐的价值生态。首先，要尊重不同文化传统、不同价值观念、不同社会群体的合理诉求和文化权益，在不违背社会公序良俗的基础上支持多元文化创作和传播，鼓励不同文化视角的价值表达。其次，加强价值对话和价值沟通，促进价值共识形成。搭建价值对话和价值沟通平台，促进不同价值观念之间的交流和对话，寻求价值共识的最大公约数。可以通过开展面对面交流的论坛、沙龙等活动，为不同价值观念的持有者提供平等交流和理性对话的平台，促进价值共识的形成和价值冲突的化解。最后，强化核心价值的引领作用，凝聚社会价值共识。将社会主义核心价值观作为中华民族共同体认同的“精神黏合剂”，真正以社会主义核心价值观凝聚共识，将价值共识转化为社会治理的柔性规范，为切实促进公民对中华民族共同体的认同提供坚实的支撑。

#### 四、结语

生成式大语言模型作为引领新一轮科技革命和产业变革的重要驱动力，其对社会文化形态的塑造作用日益凸显。本文探讨了生成式大语言模型对中华民族共同体认同可能带来的叙事同质化与碎片化、主体性转移与情感弱化、价值导向隐蔽性与偏颇性等三大风险挑战。这些风险挑战深刻触及了中华民族共同体认同的文化根基、社会心理根基和核心价值理念。为应对这些挑战，在

叙事层面，应构建多元融合、权威可信的叙事体系；在主体性与情感层面，应激发社会参与，强化情感联结，以夯实认同的社会心理基础；在价值层面，应强化价值引领，提升算法伦理，以构建健康的价值生态。在拥抱生成式大语言模型所带来的技术便利与发展机遇的同时，必须对其潜在风险保持高度警惕，并采取前瞻性、系统性的应对策略。这不仅是技术治理的题中应有之义，更是维护国家文化安全、铸牢中华民族共同体意识、实现中华民族伟大复兴中国梦的时代要求。

#### 参考文献：

- [1] 陆卫明,张敏娜.铸牢中华民族共同体意识论略[J].贵州民族研究,2018,(3):5.
- [2] 徐敬宏,张如坤.大语言模型在学术出版应用中的失范风险与应对策略[J].中国编辑,2024,(2):36-42.
- [3] 邢丽菊,赵婧.新媒体与中国国家形象的国际传播[J].现代国际关系,2021,(11):51-59.
- [4] 缪青海,王兴霞,杨静,等.从基础智能到通用智能:基于大模型的GenAI和AGI之现状与展望[J].自动化学报,2024,(4):783-804.
- [5] 高盼.从工具理性僭越到价值理性回归:“微时代”大学生思想政治教育的困境与出路[J].现代职业教育,2021,(27):1-3.
- [6] 刘纯懿,胡泳.人机逆转、叙事僵死与无事实时代——生成式革命的影响与危机[J].探索与争鸣,2024,(1):161.
- [7] 刘艳.新时代中国网络意识形态话语权建构的三维审思[J].理论月刊,2021,(6):48.
- [8] 郝时远.文化自信、文化认同与铸牢中华民族共同体意识[J].中南民族大学学报(人文社会科学版),2020,(6):5.
- [9] 支振锋.生成式人工智能大模型的信息内容治理[J].政法论坛,2023,(4):47.

(责任编辑:谭敏)(责任校对:陆刚 梁芳敏)