生成式人工智能对主流意识形态认同的挑战及其治理路径

■ 武汉理工大学 甘宇

意识形态工作是为国家立心、为民族立魂的工作。在数字时代，互联网已成为意识形态工作的主阵地。近年来，生成式人工智能（AIGC）凭借自动生成文字、图像、音视频的能力，改变了信息生产传播的方式，对社会思想观念的形成与认同机制产生重大的影响。一方面，AIGC为提升主流意识形态传播精度、效度和广度提供契机，如通过算法推荐实现“精准滴灌”，通过智能分析有效引导舆论。但另一方面，AIGC的“深度伪造”技术、“信息茧房”效应以及与资本逻辑深度捆绑催生的异质价值渗透，也为主流意识形态认同带来更深层次、更具隐蔽性的挑战。这些挑战既冲击着信息权威的信任基础，又割裂着社会共识的价值内核，更侵蚀着情感归属的心理防线。因此，系统研判AIGC带来的风险，科学构建应对策略，是维护意识形态安全、推进国家治理体系和治理能力现代化的迫切要求。

一、生成式人工智能对主流意识形态认同的核心挑战

与传统互联网技术相比，AIGC引发的挑战更为深刻，直接作用于主流意识形态认同的信任基础、价值内核与情感认同三个核心环节。

（一）权威的“双重消解”：从精准的虚假攻击到信息的结构性稀释

主流意识形态认同的建构，根本上依赖于公众对信息来源权威性的信任，这也是思想建设中始终强调“巩固思想根基、凝聚社会共识”的重要基础。然而，AIGC正从两个层面系统性地侵蚀这一信任基石。一方面，AIGC“深度伪造”能力提供了低成本、高效率的虚假叙事工具，使其得以通过制造“虚拟事实”来歪曲历史、攻击政策，直接削弱主流意识形态的公信力。而相较于这种目的明确的“锐实力”攻击，AIGC在结构层面带来的挑战则更为深远。在资本逻辑下，AIGC技术被用于批量生产低成本、高流量的泛娱乐化内容，导致大量同质化、低质化信息充斥网络空间，这不仅稀释了权威信息的传播效能，更在一定程度上挤占了主流思想舆论的传播阵地，对新时代的思想引领工作提出了严峻考验。

（二）算法“铸茧”：从共识割裂到个体认知异化

凝聚社会最广泛的思想共识，既是主流意识形态焕发生命力的关键密码，更是涵养社会思想凝聚力的核心源泉。AIGC个性化推荐机制在满足用户偏好的同时，却也构筑了前所未有的坚固“信息茧房”，从结构上侵蚀着社会认同的内核。这首先阻碍不同社会群体间的对话与理解，继而在特定议题上催生非理性的群体性认知极化。当社会共识因群体极化出现裂痕，个体或会弱化对主流权威的信任，转而将信任投射到看似客观的技术之上，这便为“数字拜物教”的滋生提供了土壤。AIGC强大的问题解决能力和中立的数理表象，使其成为这种非理性信赖的完美对象，让部分网民陷入对技术的盲目崇拜，忽略其背后潜藏的资本逻辑与价值操纵。个体一旦将认知主体性让渡给算法，便会逐渐丧失独立思考与价值判断能力，进而动摇对主流意识形态的认同。

（三）情感规训：以“隐蔽渗透”瓦解认同防线

在新时代，主流意识形态认同的稳固构建，不仅依托于理性认知层面所达成的价值共识，更扎根于情感归属维度所形成的紧密联结。在资本逻辑驱动下，AIGC平台偏好生产推送娱乐化、刺激性内容，但客观上也为消费主义、历史虚无主义等异质意识形态巧妙“编码”、隐性渗透提供了载体，使其在潜移默化中扭曲用户的价值认知，从而削弱其对主流意识形态的价值判断与情感黏性。当这种基于文化内容的价值渗透，升级为对个体情感关系的直接建构时，一种更为精准深刻的心理规训模式便随之出现。以“数字伴侣”为代表，AIGC生成的高情感交互虚拟形象，通过与用户建立深层情感连接，隐性地影响其认知偏向，使其在情感依赖中更容易受到异质意识形态的影响，最终或从根源上异化其对主流价值的情感归属。

二、“化智为治”：生成式人工智能的多维治理路径

面对AIGC对主流意识形态认同带来的挑战，简单的封锁和管制已难以奏效，必须坚持党的领导、依法治理、技术创新、全民参与的系统理念，坚持“以技术对技术、以技术管技术”为导向，探索“化智为治”的现代化治理路径。

（一）坚持党的领导，强化价值引领的顶层设计

党的领导是做好意识形态工作的根本保证。把握AIGC发展的正确政治方向，必须坚持党对网信工作的集中统一领导，为这一前沿技术的发展划定明确的航道与红线。而对宏观方向的把握，则需要通过对技术内核的主动塑造来实现，这就要求在国家层面加强对AIGC发展的战略规划和伦理引导，将社会主义核心价值观深度融入算法设计、数据训练和模型开发的全过程。由此，技术发展便不再仅仅是追求效率的工具理性，更是承载着“向上向善”价值目标的价值实践，从而在源头上防范技术异化的风险。

（二）完善法治体系，构建“负责任的人工智能”规制框架

网络空间不是“法外之地”，必须让AIGC在法治轨道上健康运行。构建这一法治框架，其首要前提是完善立法体系，需针对AIGC的内容生成、数据隐私、算法责任等新生问题，加快推进相关法律法规的制定与完善，为有效监管提供坚实的法律依据。其次，在坚实的立法基础之上，需通过法律手段明确平台企业的主体责任，推动其建立覆盖算法审核、内容管理、安全评估等全流程的内部管控制度，使法律条文转化为可操作的实践准则。最后，平台责任的有效落实有赖于对技术运行过程的透明化规制与社会监督。为此，必须探索建立算法备案、解释与问责的长效机制，推动企业适度公开算法关键逻辑，并通过建立公众参与的算法合规评议机制、畅通投诉举报渠道等，保障公众的知情权与监督权，最终构建起负责任、可信赖的人工智能治理体系。

（三）创新技术范式，筑牢“以技术反技术”的智能防线

技术带来的挑战，终究需要依托技术创新寻求破局之道。技术范式的革新，应从防御与建构双重维度展开系统性布局。就防御性规制而言，针对“深度伪造”等技术滥用带来的风险，需聚焦智能识别、溯源与阻断技术的研发突破，通过技术迭代强化意识形态安全的防御屏障，形成对风险的前置性拦截与精准化管控。从建构性赋能视角出发，更需突破被动防御的局限，将AIGC的技术潜力转化为主流话语传播的效能增量。这意味着要主动运用AIGC技术，批量生产兼具思想内涵与传播特质的主流文化产品，契合新时代网民的接受习惯；同时结合VR/AR等前沿技术打造沉浸式红色教育与国情教育场景，使主流意识形态内容在感染力、吸引力与说服力的提升中实现更深度的传播渗透。

（四）深化全民教育，提升社会整体的数字免疫力

网络安全的维护离不开全社会的协同参与，应对AIGC带来的技术挑战，更需植根于全民素养的提升。提升社会整体数字免疫力，核心在于将数字素养与智能素养教育全面融入国民教育体系，重点培育公众对信息的批判性思维、事实核查能力及算法偏见辨识能力。以此为基础，还需将通识性的数字素养升级为驾驭复杂舆论环境的网络政治能力。通过系统性教育引导，推动网民增强网络政治敏锐度与鉴别力，在多元舆论场中站稳立场、明辨是非。而个体素养向实践能力的转化，需引导网民从被动的“信息消费者”转向积极的“生态建设者”与“安全守护者”，通过自觉抵制不良信息侵蚀，将分散的个体力量凝聚为维护清朗网络空间的集体行动，进而汇聚成巩固意识形态安全的强大社会合力。

探析人工智能研究中的“人类学想象”

——对“类人智能体反思”的反思

■ 广州大学马克思主义学院 凌文亮

随着人工智能技术发展，关于人工智能的研究取得丰硕的成果。而在对人工智能的反思中，研究者往往将人工智能视为一个外在于社会关系的“智能体”进行哲学思考，或将其认为只是提升生产力的工具性存在。这种视角忽略了人工智能作为一种技术的本质，其本质是一种基于海量数据和复杂算法的“概率计算模型”。这种“类人智能体”的思考反映的是人工智能、现代资本主义背景下，人的彷徨无措与人对自身的再思考，但与人工智能本身无关。因此，为了避免对人工智能的研究落入“人类学想象”的窠臼，对人工智能研究的“类人智能体反思”进行分析、溯源成为必要。

一、对“类人智能体反思”的反思

“类人智能体”是人工智能研究中常见的“人类学想象”。在人工智能哲学中，它通常体现为人工智能的出现如何使人异化、技术异化、主体性危机的议题；在与人工智能有关的文艺作品中，它则体现为反乌托邦题材的电影和小说，以及如《终结者》《黑客帝国》等智能机器人奴役人类、反抗人类的故事情节。这种人类学想象与人工智能无关，而与人们对“自我意识”（我思）相关。它始终指向人对自身的困惑，反映的是人们对自我意识议题的“失语”。

第一，对自我意识的矛盾结构的失语。正如维特根斯坦《逻辑哲学论》中有一个著名的关于眼睛的比喻，即眼睛从来不属于被看的现实的一部分。关于人工智能的“类人智能体”的人类学想象正是对这一无法反视自身的结构性困境的逃避。一方面，我只能思考我的内容，而无法彻底地将思考着的这一个“我”作为思考的对象。倘若这样的思考是可能的，那也只是在对“他者”的观察中进行的。这就是说，“我（思）”总是无法逃出“我”的界限。但是问题是“我”是如何形成的。另一方面，为了弥补“‘我’的界限”，“我”只能将他者比作“另一个我”，同化他者构建共识，并形成“共同主观性”（Inter-subjectivity）。但是在这里不存在他者及其他异性，只有无限扩大的“自我意识”。人工智能的“类人智能体”的人类学想象正是在这一矛盾中生长。在这种幻想中，人工智能被视为“另一个我”，因此，它会被认为“产生”意识，并附上“是否拥有如‘人权’那样的‘机器人权’”的议题，进而产生“人机共存”的议题。但这不过是人逃避“自我意识”的矛盾结构的衍生物而已。

第二，对自我意识中“主奴辩证”的失语。自我意识总是以他者的实在为前提才是可能的，这已经被后结构主义的主体论反复提过许多遍，如阿尔都塞的“质询主体”、拉康的“镜像伪主体”。往前追溯还有黑格尔在《精神现象学》中对自我意识的精湛分析。他们都指向一个事实，即自我意识的构成是通过外在的非我为前提实现的，并且会在自我与非我斗争中不断确立新的自我。然而，由于第一层面中自我逃避自身有限性，转而将他者视为“另一个我”，这种非我的东西也就被遮蔽了。但是毕竟“另一个我”不是“我”，它总会有不同的一面，甚至反抗的一面。所以，自我必须时常扩张自我的领地。他者的生存空间会被挤压甚至消失，为此，他者便会奋起反抗，如“类人智能体”人类学想象的文艺作品中时常出现的叙事逻辑：“人发明人工智能、人工智能觉醒、人工智能反抗人类、两者共存”。这不过是将人工智能视为“另一个我”，以自我（意识）为中心思考的结果罢了。那么，所谓人工智能导致的各种异化和主体性危机，不过是自我中心主义的无稽之谈。

二、人工智能研究的旨归

无论如何，人工智能研究中“类人智能体”的人类学想象是与人类切身相关的“存在论焦虑”。但是这样的“情绪”是无法推进人工智能研究的。实际上，人工智能作为技术而言，人工智能与人同物的关系有关，进而是人与人的社会关系有关。而关于对人工智能的人类学想象都不过是来自人对自身的矛盾结构的失语与逃避。因此，为了重回人工智能作为技术，基于海量数据和复杂算法的“概率计算模型”这一本质，要从历史唯物论重新审视人工智能以调整人工智能研究中的人类学想象。

历史唯物论不仅主张生产力决定生产关系的生产模式论，还指出了人类未来的发展方向——共产主义。所以，在生产模式论的前提下思考，人工智能便从作为“另一个我”的“类人智能”重新成为“人工智能”。它并不像人类孩子那般各种生理机能会“自然而然地”成熟，而是需要人的不断介入开发、调整、研究。这与那种人类学想象是毫无关联的。并且，历史唯物论还要求研究者必须看到，人工智能与社会的关系以及其中隐藏的“人与物”“人与人”的关系。这是在说，人工智能始终与人相关，但又不是“另一个人”。

三、结语

对“类人智能体反思”的反思，并非否定人类对技术的人文关怀，而是呼吁将这份关怀从对“自我投射”的焦虑，转向对技术与社会互动的现实关切。唯有如此，人工智能才能真正成为推动人类社会发展的工具，而非承载人类存在论焦虑的虚幻载体。