人工智能影响政治安全的逻辑及其风险治理

——基于传播政治经济学的分析

张彦华　徐　帆

（中国矿业大学 网络风险治理研究中心，江苏 徐州 221116）

摘要： 人工智能因具有精准识别和筛选需求信息、整合议题并优化治理情景、智能模拟和辅助治理主体决策、实时跟踪评估和反馈治理效能等方面的优势，深度嵌入公共治理领域，在为政治安全提供技术支撑的同时，也可能带来国家网络主权、意识形态、政治制度等维度的政治安全风险，增加不同国家网络主权冲突的可能性，提升部分用户群体对主流意识形态的疏离度，削弱民众对政治制度的认同度。人工智能诱发政治安全风险的原因，可从技术嵌入逻辑、把关失灵逻辑、结构嵌入逻辑等维度进行分析。以人工智能的技术效能提升国家政治安全效能并强化政治体系对政治安全风险的动态适应能力，强化对人工智能算法推荐的人工把关效能并提升“人机”双重把关体系的风险防控水平，加强人工智能向善的制度设计并优化国家政治安全制度保障体系，是智能社会语境中政治安全风险的善治之道。

关键词： 人工智能；政治安全风险；国家网络主权；政治制度安全；意识形态安全

中图分类号： D63　**文献标识码：** A　**文章编号：** 0257-0246（2022）12-0196-10

中国共产党第二十次全国代表大会报告指出，以政治安全为根本推进国家安全体系和能力现代化，坚决维护国家安全和社会稳定。① 政治安全关乎国家长治久安和人民安居乐业，在国家安全体系中占据根本性地位，是保障其他领域安全的前提。近代以来，技术作为一种公共治理工具备受青睐。作为现代科技的前沿技术，人工智能因具有精准识别和筛选需求信息、整合议题并优化治理情景、智能模拟和辅助治理主体决策、实时跟踪评估和反馈治理效能等方面的优势，深度嵌入公共治理领域，丰富、拓展政治安全内容，为政治安全提供技术支撑。同时，人工智能对公共治理领域的嵌入，可能带来国家网络主权、意识形态、政治制度等维度的政治安全风险。因此，审视人工智能诱发政治安全风险的逻辑，优化人工智能自身“善”的设计，发挥人工智能“善治”效能，探索人工智能推进国家安全体系和能力现代化的路径，就成为当务之急。

基金项目：国家社会科学基金项目（20BZZ034）；江苏省教育厅社会科学重大项目（2022SJZD019）。

作者简介：张彦华，中国矿业大学网络风险治理研究中心、公共管理学院副教授，厦门大学两岸关系和平发展协同创新中心（国家级“2011 计划”）副教授，研究方向：传播政治经济学、新媒体技术与政治安全；徐帆，中国矿业大学网络风险治理研究中心研究员，研究方向：行政管理、网络政治安全。

① 习近平《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》，http://www.gov.cn/xinwen/2022-10/25/content_5721685.htm。

一、人工智能诱发的政治安全风险

政治安全是指一个国家的主权、领土、政权和政治制度以及意识形态不受别国干涉和破坏，社会政治稳定，政权巩固，拥有自主性和独立性。① 随着人工智能技术的快速发展及其对政治领域介入进程的加快，国家政治安全的内涵及影响因素均发生了深刻变化。风险社会理论认为，现代社会是一个高度现代化和高度风险化的社会，人类不合理的实践活动和发展方式等人为因素诱发了现代社会风险，并导致整个世界逐渐演变成了一个风险社会。② 作为人类实践活动的最新成果，人工智能为风险社会带来新的变量和表现形式，尤其是对政治领域的深层嵌入，诱发了诸多政治安全风险。

1. 人工智能容易增加不同国家网络主权冲突的可能性

主权是指国家独立自主地处理其内外事务的权力，保持主权独立、完整和安全是国家生存和发展的基本前提。随着多元化、去中心化的网络社会的迅猛发展，国家主权边界不断变化、调整，逐渐从现实空间向网络空间延伸。作为国家主权在互联网领域的延伸，网络主权指能够确保一个国家在互联网领域独立自主处理各项事务且免受外部网络侵犯和攻击的权利，是网络政治安全的基础要素。在不确定因素日益增多的时代背景下，占据网络技术优势的国家容易滋生技术霸权冲动，凭借其先进技术资源优势在网络空间开疆拓土，给他国网络主权安全带来安全风险。如“美国坚持其网络霸权主义，不仅无视他国的网络主权，丝毫未放松对世界网络信息的监控，而且不断利用其先进的网络信息技术，将一些以破坏政治稳定、引发社会动荡为目的的政治议程强加于他国”③。美国中央情报局前雇员爱德华·斯诺登曝光的美国政府“长期试图取用并暗中拥有所有数字通信记录”④ 的“棱镜门”项目，表明部分发达国家将人工智能技术政治化、意识形态化，利用技术优势对他国网络主权进行胁迫。这些国家掌握着绝大多数互联网根服务器，对其人工智能产品设置有“后门”，此种不良行为已经在实质上严重侵害了他国网络主权的自主性，且相关侵害网络主权的行为更为多元、隐蔽，治理难度更大。曼瑟·奥尔森认为，“国家自主性指国家作为一个组织拥有自己的利益、想法和追求，并拥有排除其他国内和国际社会力量干扰和压力而去实现自己目标的能力”⑤。部分国家以技术霸权行为对他国的干扰，不仅严重影响了后者作为一个独立国家依据自身意愿自由地制定或执行相关政策的能力，而且会长远地削弱该国民众的政治福祉。

全球化时空场域中的政治安全问题较以往更加复杂，人工智能深度伪造技术的出现，推动大量虚假内容、敏感资讯或误导性信息在全球范围快速传播，导致部分国家的政治精英赖以作出重大决策的信息环境被干扰，降低不同国家之间战略互信的强度和不同主体之间的合作水平，给国际安全形势带来微妙变化。比如，“2019 年，印度和巴基斯坦之间的紧张局势升级期间，路透社发现了 30 起关于该事件的深度伪造视频——这些视频大部分是 2017 年印巴冲突期间，发生在印巴两国一些城市的民众向对方抗议示威的场景。‘有心人士’利用人工智能对这些视频作了处理，被有意错误描述为正在印度和巴基斯坦发生的视频片段。这些深度伪造的合成视频极大地刺激了印巴两国的民族主义者，导致事态升级”⑥。利用人工智能深度伪造技术对某些敏感问题的挑衅性行为或煽动性言论，不仅会影响人们对客观现实与虚拟现实的正确认知，而且会影响部分国家良性信息生态的发育进程，降低多元

① 刘跃进 《国家安全学》，北京：中国政法大学出版社，2004 年，第 110 页。

② 刘玮 《人工智能与社会风险管理：通往现代社会安全之路》，长春：吉林大学出版社，2020 年，第 86 页。

③ 罗俊 《网络信息传播安全的核心议题：互联网时代的认知操纵及应对策略》，《学术论坛》2021 年第 2 期。

④ 爱德华·斯诺登 《永久记录：美国政府监控全世界网络信息的真相》，萧美惠、郑胜得译，北京：民主与建设出版社，2019 年，第 159 页。

⑤ 曼瑟·奥尔森 《国家的兴衰：经济增长、滞胀和社会僵化》，李增刚译，上海：上海人民出版社，2007 年，第 3 页。

⑥ 刘国柱 《深度伪造与国家安全：基于总体国家安全观的视角》，《国际安全研究》2022 年第 3 期。

政治主体的态势感知能力，增加相关政治精英发生误判的可能性，并在日益高涨的民族主义思潮裹挟下加剧对抗的烈度。在此情形下，政治决策者对相关深度伪造信息的管理会面临更大挑战，要求其必须具备快速、科学的决策能力。如果在关键时刻决策失误、处理不当，就会引发或加剧不同国家之间的主权冲突。

2. 人工智能容易加深部分用户群体对主流意识形态的疏离程度

意识形态安全是指一个国家占主导地位的价值体系、思想观念能够稳定、持久，得到民众的广泛认同、接受和支持，有效抵御外来思潮的威胁。意识形态安全在政治安全中发挥着思想指导和思想引领作用，关乎国家的道路旗帜、社会的凝聚力和民众理想信念的建构。人工智能可以通过对民众的价值观念、社会的价值趋向等施加作用，进而影响民众对主流意识形态的认知和情感。

首先，人工智能精准化内容推送诱发的“信息茧房”效应，容易增加部分用户对社会主流意识形态的疏离程度，弱化主流意识形态的思想价值引领效能。人工智能对网络社会的深度和广泛嵌入，极大地丰富了民众政治参与的方式和途径，但民众在资讯消费过程中生成的信息常会被人工智能收集并形成系统反映该用户偏好的画像，相关智能媒体便可以通过工业化生产途径向该用户持续推送成本更低且更受青睐的标准化、同质化信息，同时屏蔽其他类型的信息，形成“信息茧房”。如此一来，不仅处于不同“信息茧房”的民众之间会因获取的信息不同而容易出现分化，而且在单个“信息茧房”内部具有相似价值偏好的民众也会因群体极化而出现群体情绪、群体行为等极端化现象。① 当民众的观念认知被此类极化情绪、极化行为“圈子”包裹时，主流意识形态对其形成有效浸润并发挥价值引领的难度便会大幅增加。

其次，人工智能精准化内容推送方式容易被某些政治组织或利益集团利用，作为对他者发动意识形态攻击的工具。例如，部分发达国家凭借人工智能技术优势，通过大数据挖掘、分析、精准推送等方式，对他国某些具有特殊偏好的网络社群进行历史虚无主义、无政府主义等不良价值观的渗透，进行变相政治营销，或者对他国意识形态进行诋毁。我国意识形态安全同样会受到人工智能及其背后隐藏的不良利益团体的恶意冲击，“海外社交媒体上中国议题有被自动化操纵的影子，社交机器人可以成功渗入社交网络，改变既有信息交互结构”②，其危害不容小觑。此种行为的负外部性容易被群体极化效应放大，加剧民族凝聚力削弱、社会主流价值观失序等问题的严重性。

最后，人工智能信息推送方式的碎片化、无序性等特征，容易削弱用户群体对主流意识形态认知的深刻和全面程度。人工智能可以根据用户的偏好来持续推送某种个性化内容，但网络风潮的瞬时变化性容易导致用户群体的喜好、行为飘忽不定。因此，从长期来看，智能媒体推送的内容也会呈现出分散性、碎片化、去中心化、无序性等特征，破坏用户对相关政治信息和政治知识体系的整体性认知，削弱用户对主流意识形态认知的深刻性、全面性和系统性。同时，人工智能精准化的信息推送过于强调用户的个体感受及个体利益的达成，此种相对“狭隘”的关注不利于用户对公共利益的感知和公共精神的培养，由此孕育而成的个性化话语体系和碎片化利益加持下的狭隘的多元话语体系，也可能对主流话语体系构成冲击。

3. 人工智能对治理场景的革新及其对政治制度的冲击

政治制度安全是指国家的根本政治制度等各项制度得以正常稳定运转，并能够根据社会发展进行有序协调。政治制度安全是政治安全的核心要素和其他领域安全的重要保障。在人工智能驱动下，社会运转模式向智能化、精细化方向发展，但传统科层制治理结构以及长久以来建立在自然人类行为基础上的法律规制体系难以适用于以算法、数据为主体的应用环境。③ 目前法律与技术之间存在“自说

① 张彦华、崔小燕《网络社群行为规范对公共政策的影响及其风险治理》，《青海社会科学》2021 年第 6 期。

② 张爱军、王芳《“大数据杀熟”的政治安全风险》，《未来传播》2021 年第 4 期。

③ 贾开、蒋余浩《人工智能治理的三个基本问题：技术逻辑、风险挑战与公共政策选择》，《中国行政管理》2017 年第 10 期。

自话”“貌合神离”等现象，相关改革建议虽然在法律上逻辑自恰，但却未能触及技术的实质与本源。① 人工智能对社会的影响广泛而深刻，但政府相关政策的制定与人工智能的发展不相匹配。② 上述问题的存在，一定程度上呈现了相关法律制度与人工智能表层“耦合”但深层彼此冲突的尴尬境况。

人工智能可以通过对他国大规模用户数据进行深度挖掘和系统分析，来洞悉其政治偏好等群体特征，进而针对其心理弱点来供给带有政治目的的内容产品。例如，人工智能深度伪造技术不仅可以用来生成有关持有某一政治主张的政治候选人的暴力、色情等不良视频，通过抹黑等标签化运作来改变某些国家的政治进程，而且可以通过对部分国家政治事件的不当描绘、刻意诱导来加剧群体心理失衡、激化社会矛盾、诱发大规模群体性事件，推动该国政权发生更迭。换言之，人工智能可以隐秘地影响他国民众心理，微妙地改变其对事物的普遍看法，进而颠覆“有图有真相”的传统认知，动摇政治信任基础，侵蚀政治制度安全基石。人工智能的此种破坏力，容易强化人们对普遍事物的怀疑态度和对现实问题的冷漠态度，大幅增加政治信任赤字。在此种社会生态中，人工智能深度伪造技术不仅能够依靠掩盖真相、侵蚀客观事实而获利，而且会帮助该不当获利者隐藏得更深，增加其逃脱惩罚的概率。

人工智能可以凭借社交机器人等的应用，通过对相关用户数据资源进行深度加工，更为隐秘地改变用户群体的政治认知，解构其政治价值偏好、政治认同倾向，进而直接或间接地参与到公共政策制定、执行等国家政治制度建设的进程中。美国 Unisys 公司 2018 年的安全指数调查发现，将近 19%的美国人不愿意参加中期选举投票，因为他们担心外部行为者会损害该国的选举投票系统；近 86%的受访者对美国选举投票系统可能受到局外人影响的前景表示担忧。③ 由此可知，人工智能可以通过智能媒体的信息涵化效应来影响用户的政治观点、政治态度和政治参与方式。人工智能对政治舆论场的直接、深度介入，不仅会对个体的政治社会化过程施加影响，而且可以通过相关操作来对某个政治制度品牌进行增值或减值的标签化建构，削弱部分民众对其国家政治制度的认同度，给国家政治制度安全埋下隐患。

二、人工智能诱发政治安全风险的逻辑

在人工智能技术的研发和应用过程中，人类秉持的某种政治偏好会嵌入其中并通过技术互动、政治互动等表现出来，故人工智能也具备政治属性。同时，人工智能对人们的生活、生产和思想的嵌入过程，也是其与政治安全的互动过程。在此过程中，人工智能以技术嵌入、认知嵌入、关系嵌入、结构嵌入等方式来施加影响，驱动政治安全状态发生新变化。英国技术哲学家大卫·科林格里奇发现：“一项技术的社会后果不能在技术生命的早期被预料到，然而，当不希望的后果被发现时，技术却往往已经成为整个经济和社会结构的一部分，以至于对它的控制十分困难。”④ 这就是所谓的“科林格里奇困境”。当前人工智能嵌入政治领域引发的政治安全问题距离“科林格里奇困境”尚有较大距离，但部分问题的严重性、紧迫性仍不容小觑。因此，有必要深入探究并把握人工智能诱发政治安全风险的内在逻辑，从而为把握政治安全风险规律、制定风险治理目标、明确治理权限、分配治理任务提供参考依据。

---

① 周佑勇：《论智能时代的技术逻辑与法律变革》，《东南大学学报》（哲学社会科学版）2019 年第 5 期。

② Seth D. Baum, “On the Promotion of Safe and Socially Beneficial Artificial Intelligence,” *AI & Society*, Vol. 31, 2017, pp. 543-551.

③ 刘国柱：《深度伪造与国家安全：基于总体国家安全观的视角》，《国际安全研究》2022 年第 3 期。

④ 刁生富、冯利茹：《重塑大数据与数字经济》，北京：北京邮电大学出版社，2020 年，第 202 页。

1. 人工智能诱发政治安全风险的技术嵌入逻辑

“根据自然法则，实力界定统治的边界。”① 作为引领新一代科技革命和产业革命的战略性技术，人工智能的发展程度直接彰显国家的综合国力，并影响着国家的国际政治地位和话语权。由于历史等方面的原因，发展中国家人工智能的理论基础、核心技术、实践水平与发达国家尚有较大差距。不同国家之间的技术差距容易诱发部分技术优势国家的霸权冲动，增加技术优势国家以人工智能技术不断试探他国的网络主权边界并对后者构成实质威胁的可能。当然，部分后发优势国家在深度学习、知识图谱产品、智能机器人等领域的快速发展，会给发达国家带来危机感，促使部分发达国家利用现有技术优势、话语优势进行阻挠和挑衅，威胁发展中国家的政治安全。

智能感知、智能分析等人工智能技术的发明和应用，极大地拓展了人们获取资讯、表达利益诉求和参与政治的渠道。通过此种技术赋权、赋能的方式，越来越多普通用户的知情权、选择权等正当权利得到更大程度的保障和实现，而多元利益主体的话语权等资源也得以重新调适。然而，人工智能嵌入政治参与的过程，也是其以某种数据化、智能化的政治参与方式对国家政治安全态势施加影响的过程。在此过程中，某些客观因素和主观因素的交互作用，可能会加剧相关政治安全风险发生的概率。例如，在数字化语境中，人类在消费相关数据的同时，自身生活、生产的诸多行为也源源不断地转化为数据资源，通过人工智能技术便可以以相对低廉、便捷的方式获得这些数据集，为深度仿造作品的生产提供资源。来自现实生活的诸多数据常会沾染用户个体或社会文化的某种偏见，使得人工智能赖以运作的原始数据也存在偏见，这可能会诱发算法歧视等风险。数据采集的不真实性、数据运行过程中出现的遗漏、数据的片面化、数据规模不足、使用被篡改的数据等，均会降低数据资源的真实性、客观性、可信度，削弱人工智能辅助政治决策的科学性、合理性、可行性。因此，数据来源的真实可靠程度直接关乎人工智能运行的输出结果，而相关技术漏洞则会削弱数据价值，为政治安全埋下隐患。

相关研究表明，人工智能虽然可以凭借深度学习等功能来构建相应的模型并试图达到某种目的，但模型在进行训练时所做的假设越来越多，复杂程度也越来越高，构建的模型数据验证上所显示的误差也就会增大，从而模型更多地只能拟合训练数据，而无法充分拟合训练外的其他数据。② 人工智能凭借机器学习等功能进行深度学习并依靠专家系统等技术来辅助相关公共决策时，容易导致过拟合问题。③ 过拟合问题的存在，意味着人工智能决策并不具有绝对的准确性，可能会因技术失灵而使得政策或制度缺乏有效性和可行性，由此引发的政治决策失败或政治失灵，常会导致社会秩序紊乱并诱发政治流言、政治谣言，加剧政治安全风险。

2. 人工智能诱发政治安全风险的把关失灵逻辑

在人工智能和政治安全体系共同建构的活动网络、关系网络中，所牵涉的各个主体均可能会通过技术嵌入、关系嵌入等途径形成某种正式或非正式的合作网络，而该网络承载的技术价值、情感认同、利益博弈和专业信任等均会塑造相关主体的情感倾向和行为选择。把关人理论认为 “在群体传播过程中存在着一些把关人，只有符合群体规范或把关人价值标准的信息内容才能进入传播的渠道。”④ 可见，人们常处于经过筛选后的信息语境中，对多数政治问题的认知也常受其接触的信息内容制约。传统媒体时代把关人多为专业的媒体人，但人工智能的赋权和赋能效应使得传统把关人的权力被分散。在泛智能化的社会生态中，“5G 网络、网络 IP 技术等的发展、衍变，常会呈现向该领域

① 格雷厄姆·艾利森 《注定一战：中美能避免修昔底德陷阱吗?》，陈定定、傅强译，上海：上海人民出版社，2019 年，第 135 页。

② 魏斌 《符号主义与联结主义人工智能的融合路径分析》，《自然辩证法研究》2022 年第 2 期。

③ 所谓过拟合，即“每当算法在数据中找到现实世界中不存在的模型时，我们说它与数据过于拟合”。参见佩德罗·多明戈斯 《终极算法：机器学习和人工智能如何重塑世界》，黄芳萍译，北京：中信出版社，2017 年，第 91 页。

④ 吴小君 《舆论应对危机传播》，北京：中国传媒大学出版社，2015 年，第 38 页。

主导技术靠拢或集聚的态势”①，故在科技系统中占据核心位置的人工智能在某种程度上对信息的传播方式、传播渠道等具有部分决定权，扮演着资源交换“强制通道点”的角色。同时，在人工智能技术加持下的信息传播过程中，常会依据流量、变现等标准而对超级平台、用户个体等不同利益主体赋予差异化的权重，进而影响合作网络的公平性。基于公平原则的社会交换过程，可以保证不同利益主体资源的平等交换，推动信息传播网络体系的良性发育，建构场域内不同主体之间的良性关系；反之，则可能会诱发诸多风险。由相关利益主体和人工智能共同构成的“人机”把关体系存在一些缺陷，会在某种程度上导致把关失灵，进而降低社会主流价值的引领力并诱发意识形态风险。

首先，从“人机”把关体系中的人为层面来看，相关技术或管理人员的价值偏见容易嵌入部分把关环节，削弱把关体系理应具备的中立、客观属性。作为利益相关者重要组成部分的企业及其研发人员掌握着智能算法等人工智能核心技术并对其运行有较大的控制权，其自我约束强度直接影响着相关政治安全风险发生的概率。在算法黑箱效应下，普通民众难以对技术精英进行有效监督，而技术精英可能在高额利润的诱惑下放松对自身的约束，并将自身利益嵌入智能算法的编码、译码过程中，从而削弱人工智能产品、服务的公共产品属性。部分研发人员的情感、价值偏见对人工智能编码的嵌入，也会导致算法过滤程序本身带有一定的倾向，将富含歧视性的态度、情感以算法译码的方式传递给用户群体，削弱其在把关过程中的有效作用及该把关体系的公平、公正程度。

其次，从“人机”把关体系中的技术层面来看，部分人工智能的迎合式把关模式容易诱发把关失灵风险。在万物互联的泛智能化社会中，以智能媒体平台为代表的人工智能技术平台之间的数据、信息流动，常会为民众之间的信息沟通、关系建构提供丰富的途径，而相关人工智能技术则发挥着技术把关的作用。在人工智能的技术把关过程中，用户通过信息选择行为向智能算法传递某种兴趣信号，而后者会依据这个信号来进行信息把关、筛选并向用户进行精准化内容推送。由于用户常对暴力、色情、刺激的“膻色腥”内容感兴趣，故较为严肃的主流政治观点和艰涩的专业知识便容易被忽略或抛弃。相关研究表明，‘娱乐产品是令人着迷的，并且能够让人保持一种固定的习惯，进而让使用者和生产者乃至整个社会产生联系和反应，同时，在这一过程中，产品能够发挥思想灌输和操纵的作用”②。超额利润需求导向下的人工智能，常以用户对信息的偏好和反应为依据设定算法推荐程序和信息过滤机制，呈现的多为用户感兴趣的内容。在此种迎合式把关模式下，体现主流意识形态的信息逐渐让位于娱乐化等“三俗”信息，主流价值观逐渐淹没在过度娱乐化浪潮中，主流意识形态的引领能力逐渐降低并诱发意识形态安全风险。

3. 人工智能诱发政治安全风险的结构嵌入逻辑

政府机构会通过公共行政途径对技术企业进行约束，而技术企业在某种制度环境中规范自身行为并发挥相应的功能，进而反向影响制度的调整，故人工智能和政治安全的相关行动主体常受到其所处的政治体制、权力结构和政策环境影响。通过此种结构嵌入过程，人工智能虽然可以用数字治理等方式对公共事务进行流程设计、权力再造并使相关制度发生可以预见的变化，但同时也会加剧政治安全风险的外溢程度。

首先，部分用户群体对人工智能的依赖，容易加剧技术理性对人的价值理性的驯化风险，并为相关政治安全埋下隐患。具体而言，为了获得更具效能的技术支持，民众愿意部分让渡隐私权、平等权、自由权等权利，此种权利让渡容易诱发人工智能技术权力的扩张风险，并以数据隐私泄露等方式侵蚀民众的正当权利。人工智能在为用户群体推送便捷化、精确化、个性化信息内容的同时，也在不断从心理、情感、行为上强化用户对此智能产品或服务的依赖，并利用人们对其的依赖而获得各种隐私、敏感信息，在数据分析、系统观察、深度挖掘、体系化评估的基础上，了解用户兴趣爱好、判断

① 张彦华 《网络视频行业的利益分配风险及其治理策略——基于传播政治经济学的分析视角》，《编辑之友》2022 年第 9 期。

② 赫伯特·马尔库塞 《单向度的人：发达工业社会意识形态研究》，刘继译，上海：上海译文出版社，2014 年，第 11 页。

用户动机倾向、预测用户行为偏好，为用户贴上某种政治身份标签，进而向用户推送契合其需求的内容信息。在此过程中，人工智能通过算法歧视等数据使用的偏差行为，将部分价值不高的“非优质”用户群体标签化、污名化，从而降低其政治参与的机会和质量，并在实质上构成对其政治地位平等性等正当权利的侵犯，部分消解了相关用户群体的政治能力、作用，导致其政治发展的无奈。人工智能还可以通过大数据挖掘、用户画像等方式来掌握富有地域或民族特色的民众的政治认知、情感、价值和行为偏好，因势利导地对其进行思想渗透。

其次，人工智能具有高度的专业性，其运行过程具有较高的复杂性，由此导致的算法黑箱效应容易诱发政治信任风险。人工智能算法主要由技术基础层（输入端）、技术程序层和技术结果层（输出端）三部分组成，其呈现方式为计算机代码，而并非大多数人所能理解的自然语言，① 故多数非专业技术人员对于人工智能主要运行区域的隐含层缺乏了解。此种算法黑箱拉开了民众和人工智能之间的距离，增强了人工智能技术的神秘感和不可解释性。在此黑箱效应下，民众不了解数据采集、算法运行的程序，故其知情权、监督权存在一定程度的缺失；同时，政府决策过程中使用的数据信息的真实性或有效性并不能得到很好的审核、监督，而一旦接入错误信息、虚假信息，便会直接影响最终输出的结果和政府的公共决策能力。因此，人工智能的缺陷在“黑箱”下得以隐藏，而此种不透明性引发的密室政治忧思容易放大政府信任危机，诱发政治安全风险。

最后，政治安全风险的人为建构因素，也使该风险领域充满复杂性和不确定性。在泛政治化的社会生态中，不同网络社群及其代表的利益集团，不仅可能因为偏见而夸大或缩小某种风险，而且可能因为圈层对话效能的低下而形成偏见固化等现象，进而导致政治矛盾扩大化。因此，人工智能诱发相关政治安全风险的过程在某种程度上具有社会建构的色彩，且部分政治组织或个体等行动者对该领域风险的选择性认知、选择性接受或选择性拒绝等人为因素使得该政治风险的来源充满了复杂性。例如，为了迎合用户群体的娱乐化需求并获取超额经济利润，以及为了规避因敏感政治问题诱发的诸多不确定性风险，部分智能媒体平台不仅以竞价排名制度、热搜排行榜来高价售卖用户的注意力等可见性资源，而且通过议程设置等方式将民众的注意力从主流意识形态、公共事务领域引导至娱乐消费资讯领域，从而削弱了部分用户群体对公共事务正常的参与程度，进而引发政治情感疏离、政治冷漠、不良政治生态发育等政治安全问题。

三、人工智能嵌入政治安全风险的智慧治理策略

人工智能嵌入政治安全的过程，是一个逐步递进的过程——双方先从技术工具嵌入过程认识到彼此的重要性，并通过关系嵌入建立起某种合作关系网络，进而推动制度建设等结构嵌入进程并实现某种政治嵌入效果。当然，此种嵌入过程是一个螺旋式的递进过程，会通过不断的嵌入循环来持续提升相关效果。由此可见，政治安全问题不仅是人工智能技术负外部性的某种体现，也是泛智能化社会关系演化和相关制度体系共同作用的产物。因此，有必要在以技术工具视角、权力把关的关系视角和制度结构视角对人工智能与政治安全运作的立体图景进行系统勾勒的基础上，将人工智能的技术优势转化为相关政治安全产品和服务的供给效能优势，进而以此长效机制来提升社会福祉并保障国家的政治安全。

1. 以人工智能的技术效能提升国家政治安全效能，强化政治体系对政治安全风险的动态适应能力

在由数据信息建构而成的关系网络中，不同类型的用户会基于某种共同目的或行动旨趣形成网络社群，并以数据、信息、关系和能量为介质与政治安全体系建立联系，进而构成人工智能嵌入政治体系的关联性基础。在由不同数据流、信息流、关系流、能量流共同建构而起的关系网络中，部分关键

① 谭九生、范晓韵 《算法“黑箱”的成因、风险及其治理》，《湖南科技大学学报》（社会科学版）2020 年第 6 期。

节点功能的紊乱或失效所诱发的政治风险会对整个社会网络系统构成冲击。因此，有必要强化作为该关系网络关键技术支撑的人工智能的技术工具效能，持续提升人工智能对政治安全风险的治理效能。

国际传媒大亨默多克曾言：“谁控制了传播的入口，谁就控制了整个世界。”① 此论虽有夸张成分，却在某种程度上揭示了泛智能化社会中人工智能对信息传播、话语建构和政治资源分配等的强大影响力。换言之，在国际信息传播场域中具有信息把关、认知操纵功能的智能媒体的议程设置能力、媒介框架效应等功能，不仅决定着相关信息的可见性和普通民众在海量信息中实质上能够接受的信息范围，而且影响着用户的时间、精力、注意力的配置倾向。特别是在竞争激烈的信息传播场域中，用户注意力呈稀缺状态，故人工智能生产、传播的相关内容产品和服务的质量，以及在政治话语议程中的设置能力和反认知操纵能力，将直接影响政治安全相关维度的风险状态。因此，针对人工智能引发的政治安全风险，相关公共管理机构应根据不同维度、不同类型风险的特征，进行精准而有效的治理。以人工智能深度伪造技术诱发的认知操纵风险为例，相关政府部门应在认知操纵的社会知觉操纵阶段，针对虚假资讯、政治谣言或流言的选择性传播等风险，予以充分揭露和反击，从而规避因首因效应而令自身陷入公信力受损的被动境地；在认知操纵的社会判断阶段，应主要以客观的事实呈现策略来揭露部分智能媒体资讯报道的选择性视角及其判断事务的双重标准，引导用户群体跳出其设置的议程框架，帮助用户正确认知、判断相关政治现象的性质、类型，并形成正确的政治认知和政治判断；在认知操纵的社会解释阶段，应以有效的媒介框架与部分不良媒体展开话语竞争，解构后者精心设置的意识形态框架，进而以涵化效应来影响用户框架的形成进程并在话语竞争中赢得舆论的主导权。

相关部门还应持续激发人工智能嵌入政治安全体系的正外部性效应，不断推动相关政治体系的优化调整和积极构建进程，持续提升该治理体系对风险社会的动态适应能力，积极应对各种政治安全风险的挑战。虚拟世界、现实社会共同构成的动态变化的风险社会语境，使得承载政治安全的相关政治体系的功能、结构等要素需要不断进行动态调整。同时，在人工智能与政治安全的关系互动场域中，政府应遵循该场域生态系统运作的内在逻辑，增强政治体系适应现实的各种能力，激发并强化其政治活力，使政治安全风险保持在可控水平。例如，人工智能技术可以对相关合作网络中的数据资源进行价值开发，提炼新型政治能量，以弥补传统政治运作中相关资源匮乏的缺陷，进而提升民意在数字政府、智慧治理等公共产品、服务供给中的分量。此种智慧治理方式，不仅可以促进该政治领域生产力方式和生产关系的变革，加快该领域良性政治形态的升级、转型进程，而且可以驱动政治体系不断完善，有效提升应对政治安全风险的自组织能力。

2. 强化对人工智能算法推荐的人工把关效能，提升“人机”双重把关体系的风险防控水平

在泛智能化社会生态中，人与人工智能之间的关系逐渐向共生状态转变。在此“共生”关系状态中，人需要适应机器并保持理性，确保人机关系和谐有序。政府权力的起源是自然状态中的人们愿意将自己的惩罚权交由一人行使，而这个人则必须按照社会规则或授予他权力的这些人一致同意的规则来行使相关权力，正是民众的同意和权力的赋予使政府拥有保护他们财产的能力，而维护政治安全的有效性、合法性也需依赖于被统治者的“同意”。② 其中有效性主要指某个政体及其承载的相关政府职能能够满足多数民众、企业等利益主体的期待，合法性则主要指某个政体拥有提出并长久维持某种信念的能力，以此来证明现行政治制度与其所在社会最为契合。

在当前泛政治化、泛智能化的社会语境中，人工智能不仅冲击、颠覆和重构了传统社会发展形态，而且通过智能辅助决策基本实现了对社会生活、生产场景的深度嵌入。同时，政治体系、普通民众的生活与人工智能的关系愈加密切，而民众的日常生活利益和国家政治安全之间的关联程度也日益

① 王茜 《新媒体概论》，北京：中国传媒大学出版社，2020 年，第 64 页。

② 约翰·洛克 《政府论》，丰俊功译，北京：光明日报出版社，2009 年，第 174 页。

增强。在此过程中，人工智能可以通过影响多数普通民众的政治倾向，直接或间接地对国家的政治安全施加某种影响力。例如，部分用户群体对相关同质化内容的选择性接触、选择性认知和选择性记忆，容易导致“信息茧房”和群体极化风险。相关心理战、信息战、认知战专家正是利用普通人的这种心理特征和认知缺陷来制定有针对性的信息传播、认知作战方案，并在人工智能技术的加持下对用户心理进行窥探，进而利用议程设置、媒介框架等方式来暗示、动摇其政治认同，削弱意识形态安全等政治安全的有效性、合法性。因此，有效提升用户群体的人工智能素养，强化对人工智能的驯化程度，不仅可以提升多数普通民众在认知战场中的抵御能力，以及对深度伪造等相关政治资讯的自我甄别能力，而且可以提升不友好利益集团的认知战成本，有助于提高“人机”双重把关体系的政治安全风险防控水平。

就民众人工智能素养提升而言，政府应建构人工智能素养提升的保障体系，激励民众在实践中培育对信息操纵、心理冲击等信息技术的甄别和批判能力，增强对人工智能技术的认知程度。智能媒体平台在获取用户信息时，不仅应以简单易懂的方式向用户说明相关功能设置需要获取的信息的具体内容、获取方式及其对用户带来的影响，以友好的人机互动界面来赋予用户偏好设置的选择权，进而以良性的政治参与渠道来使民众自觉、自发地成为国家政治安全的维护者。用户应积极增强对人工智能技术的适应力，在与人工智能的协同共进中锻炼、培养和形成独立思考的能力，驯化、掌握人工智能技术而非沦为其附庸，发挥对智能算法推荐的人工把关效能，提升个人因素在“人机”双重把关体系中的权重，防止部分不良利益主体对个人信息的非法或恶意利用。

3. 强化人工智能向善的制度设计，优化国家政治安全的制度保障体系

政治的首要问题就是政治制度化的发展往往落后于社会和经济变革，没有强有力的政治制度，社会便缺乏确定和实现共同利益的手段；创建政治制度的能力就是创建公共利益的能力，政治制度的适应性、复杂性、自主性和开放性是政治制度化水平的主要衡量指标，政治制度化水平越高，政治稳定程度就越高。① 随着泛智能化社会的崛起，完善、稳定、科学且能够夯实政治统治合法性、正当性的政治制度显得尤为必要。换言之，为有效保障智慧社会中的政治安全，不仅应规范人工智能的规制效能，强化该领域技术向善的价值向度，还应不断优化相关政治体系的结构功能，削弱或消泯诱发相关政治安全风险的社会矛盾和民生问题，将人工智能的发展与国家治理体系的现代化进程有机结合起来，以协同共振效应来提升国家政治安全水平。

首先，制度规范的价值导向应充分体现作为权力本质和源泉的人民性的特质，规避政治权力的越位、错位等不良现象，并在此基础上持续优化国家政治安全的制度保障体系。尽管人的作用在快速发展的人工智能面前呈现弱化趋势，但人工智能技术应当是保护而非侵害民众的正当权益。因此，政府应通过充分发挥人工智能正外部性的价值导向作用，在该技术体系中嵌入发展为人类服务的理念，在人工智能技术研发、设计、运用进程中秉持合乎道德价值和伦理规范的标准，使人工智能适应人的发展需求。此种策略，不仅有助于驱动相关政治安全产品、服务在供给侧和需求侧不断优化，推动人工智能与人类协调发展，而且可以加快具备科技向善属性的人工智能成果的接受、应用和普及进程，以人工智能与国家政治安全体系的广泛、深度交融来为国家政治安全提供软件和硬件支撑，努力开辟政治安全产品市场价值和社会价值保值、增值的有利区间，提高国家网络主权的自主性，切实保障国家政治安全。

其次，应将人工智能的升级迭代效能持续转化为制度创新效能，进而实现国家政治安全从应用场景到关系场域的战略定位升级。人工智能相关技术的升级换代及运用过程，同时也是其在政治安全领域不断运用并施加影响力的过程。在此技术赋权、赋能效应下，人工智能虽然能够通过技术集合效应来积累相关政治资源，但其对传统社会结构等领域的冲击常会导致诸多政治安全风险。因此，政府不

① 塞缪尔·P. 亨廷顿 《变化社会中的政治秩序》，王冠华、刘为等译，上海世纪出版集团，2008 年，第 4-19 页。

仅应重视人工智能在新场景、新运用、新传播模式等方面的驱动效应，以更为有效的制度化方式对国内外多重场域、关系进行整合，而且应提升对人工智能和政治安全领域动态要素的有效认知、识别、捕捉能力，更应以组织、制度的弹性变化和对相关风险的韧性治理来强化自身对相关资源的搜集、消化、整合能力，以及以具体、可行、创新的相关措施和有力的支撑机制来强化其实施的能力。例如，东盟国家试图以东盟国防部长会议（ADMM）等合作渠道为契机，建构东盟政治安全共同体（APSC）的制度化框架。① 此种措施，不仅可以为相关国家的国际制度设计策略提供启示，进而建构某种更加自信、开放的运作格局，而且有助于以人工智能赋能等方式来主动进行国内外政治安全保障的创新实验，为政策、组织设置等的适时革新提供保障，为该领域良性发展提供制度化支持。在此正向增强的良性循环之下，更新换代速度相对较快的人工智能，能够将自身对社会的驱动力转化为制度创新的驱动力，进而以系统性、体系化和结构化的优势来提升国家政治安全的保障力度。

最后，应不断优化人工智能技术的外部监管体系，建构政治安全发育的良性生态。人工智能算法推荐等相关产品或服务易引发“信息茧房”和意识形态等风险，而研发企业或相关智能媒体平台也常因追逐流量变现、削减成本投入等商业利益而放松对不良信息的把关，从而导致公共价值受损。因此，政府应有效压实智能媒体平台、技术企业等利益相关者的主体责任，鼓励其以开发虚假信息识别系统等措施来削弱、规避深度伪造等不良信息的传播所导致的政治安全隐患。同时，建立健全监管体系，根据人工智能嵌入政治安全体系的不同维度的运作逻辑，针对人工智能的设计、研发、应用等多维层面，构建专业的监管责任分配机制，并以相关制度、标准或其他激励措施来促进研发企业、技术人员将正外部性的公共价值观嵌入相关内容产品和服务的开发之中。例如，为实现人工智能向善的价值向度，政府不仅应完善算法审计制度，注重对算法所需数据质量的审查，确保数据安全，提升算法设计和研发利用的数据的真实性、有效性，规避或削弱算法歧视或偏见诱发的相关政治安全风险，而且应督促相关技术平台加强常态化的自我审计力度，以职业伦理对相关人工智能应用的设计原理、合法性、风险性等因素进行充分评估和有效把关，适时对反面典型实行约谈惩戒，加大对违法、违规平台的惩罚力度，制定应对算法技术风险的应急预案，构建相应的责任追究机制，以威慑效应来推动相关企业不断完善风险防控体系，为政治安全良性生态的建构夯实制度基础。

综上所述，人工智能的快速发展和社会转型相互交织，由此引发的不确定性大幅增加了政治安全风险，直接考验着国家政治体系保障政治安全的能力。有效防范并化解政治安全风险，是社会转型期必须长期重视和审慎应对的重要问题。当然，政治安全风险的发生也有其现实“病灶”，故政府有必要与时俱进，及时提升人工智能祛除政治安全现实问题的效能，不断优化提升执政党的执政水平、执政能力，通过执政党、政府、人民、社会之间的良性互动来强化社会的公平正义程度，切实维护民众的正当权益，以政治安全外部风险防范体系与内生机制之间的协同共振来持续探索新时期确保政治安全的长效机制。

责任编辑：王永平

① Lee Sookyoun, “Institutional Development for Realization of ASEAN Political Security Community,” *Korean Journal of Political Science*, No. 4, 2021, p. 95.