《学术界》(月刊)
总第303期,2023.8
ACADEMICS
No.8 Aug. 2023

ChatGPT技术视野下政治秩序重构及其风险治理〔*〕

孙会岩

(上海大学　政党治理与社会发展研究中心,上海　200444)

〔摘　要〕随着技术不断迭代升级,作为人工智能技术最新表现的ChatGPT已经进入到政治生活领域,其核心功能与已有政治秩序要素日益深度链接,为新一代人工智能视野下政治安全议题构建提供了新的载体。然而,ChatGPT使用过程中的数据流动也会冲击虚拟空间主权独立性,资本逻辑下的技术霸权会威胁网络政治参与过程,主流意识形态安全也面临生成式大模型的渗透风险。针对ChatGPT技术"人—机—物"高度融合带来的政治秩序挑战,应在总体国家安全观指导下,加快推进ChatGPT向善发展的顶层设计,打造ChatGPT风险防范的技术治理系统,构建ChatGPT协同监管的制度保障体系,合力形成智能技术、组织设计、伦理系统与制度机制等多元协同发展的政治秩序治理模式,进而实现新一代人工智能视野下政治安全风险有效治理。

〔关键词〕ChatGPT;政治安全;风险;协同治理
DOI:10.3969/j.issn.1002-1698.2023.08.007

政治安全事关人民安居乐业和国家长治久安,在国家安全治理中居于根本性地位。党的二十大报告明确指出,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]时至今日,良好的政治秩序需要以政治安全为根本性依托,而技术的不断进步正在成为影响政治安全的重要因素,以智能技术为代表的新技术革命影响社会发展,日益成为人们的共识。在多重技术加速叠加升级背景下,智能传感、深度学习、人机协同、自主操控等技术与日常生活深度融合的趋势愈加明显,以ChatGPT为代表的新一代人工智能再次掀起新一轮数字智能化浪潮,逐步引发社会、经济、政治、外交等领域的划时代变革。ChatGPT以无与伦比的聊天互动能力和知识学习能力已然把人类带入人机软融合的快车道,这不仅改变了人类生产生活方式,也引发了人们对传统政治秩序稳定的担忧。

目前学界对ChatGPT已经给予高度关注,研究成果涉及教育方式、数字经济、社会民生,以及人机关系深度融合等诸多领域。然而,鲜有从政治安全的视角分析其带来的机遇与挑战,以及思考政治秩序在新一代人工智能背景下的重构方

作者简介:孙会岩,政治学博士,上海大学政党治理与社会发展研究中心研究员、副主任,研究方向为技术进步与政党政治。
〔*〕本文系国家社会科学基金青年项目"互联网时代中国共产党的执政安全研究"(18CDJ004)的成果之一。

向等。然而,如何评估 ChatGPT 对已有政治秩序带来的影响?如何稳定新一代人工智能技术视野下政治安全新局面?这既是新一代人工智能技术发展需要审视的问题,也是以新安全格局保障新发展格局的现实需要。为此,本文聚焦以 ChatGPT 为代表的新一代人工智能技术在政治领域的议题转向,厘清 ChatGPT 诱发政治安全风险的深层逻辑,进而探讨 ChatGPT 视阈下推进国家安全体系和能力现代化的路径,以期为推进"数字中国"和践行总体国家安全观提供价值启示。

一、以 ChatGPT 为代表的新一代人工智能及其政治安全向度

早在智能化之始,就有学者认为"机器可以全面、综合地复现人类的所有思维能力,且聪明程度能够达到或超过人类"。[2] 在多重智能技术叠加影响下,以 ChatGPT 为代表的新一代人工智能快速发展中的政治特征正在显现,并成为影响政治秩序的重要变量。

(一)ChatGPT 技术快速发展的工作机理

发展至今,依托大规模神经网络支撑的新一代人工智能技术正在加速迭代升级,其典型代表 OpenAI 公司去年底推出的 ChatGPT,凭借模拟神经网络架构实现了拟人思考和交流,使用者可以用它完成撰写学术论文、生成资讯、创作诗歌等工作,以及能从多场景中助力工作和生活。其短时间内凭借先进的算法和硬件支撑,成为用户增长最快的应用程序。简单而言,ChatGPT 可理解为基于人工智能技术生成性预训练转换器的大型语言机器学习系统,[3] 并能将视频、语音、图片等信息都整合到这一大模型内。它的工作机理主要包括训练和推理两部分,一是通过对大量数据进行训练学习,理解语言结构、理解语义和上下文,以及如何生成自然流畅的智能回复;二是训练完成后使用其模型生成器来处理用户的智能需求,并尝试理解用户的意图和需要,在此基础上完成用户的指令。究其本源,ChatGPT 是一种经强化的人工智能深度学习算法,[4] 是基于人类反馈的深度学习范式,具备强大的数字化内容推理和创作能力,能够采集、加工、升华人类知识与人工内容,产出可以匹配与满足用户需求的各种数字化内容。随着模型的参数量、参与训练的数据量以及训练过程累积的计算量的不断增大,ChatGPT 日益具有较强的综合性类人智能,对语言文字的"理解"和互动能力已经超越普通人,甚至有专家预测它可以依靠强大的模拟能力复盘通过"图灵测试",以至于把世界推向强人工智能时代,甚至有研究者认为,人类社会主动或被动地进入第四次工业革命阶段已不可逆转。[5]

(二)新一代人工智能技术的功能定位

与传统执行任务的弱人工智能不同,以 ChatGPT 为代表的新一代人工智能技术会对输入产生自主性反应。随着高性能芯片、神经语言模型、深度算法技术的进一步融合,ChatGPT 的虚拟体验、算法沉浸等与人类社会发展相关联的功能日益显现,以至有学者指出,"现代信息技术和由人工智能驱动的认知机器的出现,有力地改变了人们的生活和工作",[6] 新一代人工智能技术蕴涵的功能特征和应用场景日益成为影响人类社会的重要力量。其一,多重人工智能技术叠加深化了内容的自动生成。得益于机器学习、区块链、云计算等多重技术叠加,ChatGPT 通过理解和学习人类对话,在生成文字、图片、视频等内容方面,按照逻辑关系进行重新梳理和整合,不仅能够根据语言模型进行互动对话,还能根据自主意愿完成创作,同时也能全方位进行算法分析,有效拓展、放大了人类的意图和能力。其二,基于人类反馈的强化学习提升了体验效果。新一代人工智能已经训练出拟人化认知和价值并衍生至网络空间中,ChatGPT 通过对现实世界内容的深度学习,将大量结构化和非结构化数据转化为可视化知识图谱,从而更好地实现了传播资源、环境与受众的深度融合。其三,实时交互与多重数字化逻辑增强了社会价值。新一代人工智能通过全息数据追踪式分析,推动了 ChatGPT 的自然语言处理能力向增强现实、扩展现实快速

发展,为受众提供了解决复杂意图、因果关系、相关关系等方案,并在此基础上产生出一种智能式的认知理解的伦理道德模式。总之,在算法技术不断迭代过程中,企业、学校、个人会继续丰富和深化 ChatGPT 的功能特征,不断推动 ChatGPT 应用场景在理论和实践层面的发展。尽管 ChatGPT 还未正式引入中国,但国内很多知名的互联网企业正在开发和测试类似产品,这为新一代人工智能技术在中国的发展提供了有利契机。

(三)新一代人工智能技术的政治安全转向

技术与政治具有重要的二元关联性,寻求安全的社会越来越重视将技术进步转化为维护安全的载体。当安全形态与智能技术的互动模式发生转变后,智能技术范式兴起和全球化发展为政治领域奠定了宽阔的平台与深厚的基础。[7]就政治安全而言,世界正进入一个由大数据驱动的人工智能系统为政治秩序提供信息的时代,新一代人工智能日益对民众的日常政治生活产生深远影响,甚至会颠覆我们对政治过程和权力关系的认识,政治安全中的意识形态安全、政权安全、制度安全、秩序安全等构成要素也随之不断延展。[8]前不久上线的 GPT-4 更是以海量的文本训练打破了虚拟与现实的边界,重新打开了人工智能技术可能在安全领域产生影响的巨大想象空间。[9] ChatGPT 对政治生活的介入将对国家的政治体系运转产生深刻影响,也让与新一代人工智能相关的安全议题进入学术研究视野。越来越多的研究者认为智能技术及技术力量是政治秩序重构的重要推动者,强调智能技术对于政治安全要素在实践方向上具有重要影响。同时,新一代人工智能技术的嵌入给政党政治及其社会治理带来诸多新挑战,如在意识形态安全、网络安全治理、隐私保护、智能技术能力培育等方面。掌握技术的行为体甚至可以通过 ChatGPT 相关议程来对某个政治制度进行增值或减值的标签化建构,削弱部分民众对其所在国主流意识形态与政治制度的认同,从而给政治安全添加众多不可控性和不确定性因素。基于此,本文拟在学界已有研究的基础上,考察以 ChatGPT 为代表的新一代人工智能技术进入政治生活后,对政治秩序的运行逻辑、主体行为偏好与互动方式、议题范畴等产生怎样的影响,在实然层面给主权国家及其政治行为体带来怎样的风险,进而从组织、技术、社会、制度多元协同上提出 ChatGPT 快速发展下政治安全治理策略,以期勾勒出新一代人工智能视野下政治安全治理的全新图景。

二、ChatGPT 技术视野下政治安全的主要议题

以 ChatGPT 为代表的新一代人工智能技术进入到日常政治领域后,日益推动政府将其运用到意识形态、利益表达、政治整合、主权安全等议题中,甚至直指政治安全的深层本质,促使人和机器多元交互、虚拟与现实融合的政治安全模式迅速形成,这对国家安全体系和能力现代化将产生重要影响。

(一)深度机器学习助推意识形态传播效果

随着人们越来越依赖数字增强技术,人类的理性思维正在放弃其作为世界现象的唯一发现者、认识者和分类者的首要地位。ChatGPT 大模型不仅具备对人类语言的仿真输出能力,而且还拥有对人类复杂语言的理解能力,能够在庞杂的人类语言中把握关键逻辑,对人的信息输入和思想输出都产生重要影响,这远超以往的弱人工智能模型,形成了人与机器实时互动的全新接口,这对于机器辅助人类意识形态传播具有重要影响。细言之,ChatGPT 以机器学习的方式全面融入人类信息生产领域,不仅能够实现内容创作,还能实现高效剔除干扰信息提升信息获取效率、高效处理原创问题提升信息处理质量、快速搭建智能框架提升内容创作效率,以此变革人类获取信息方式、人与机器交互方式、数字内容生产方式,为意识形态传播提供新的载体。如国内头部数字人企业已经推出冬奥手语数字人、央视数字主持人等应用,这些算法加持下的人工智能载体,为意识形态的实时传播提供了新的可能。在此基础上,新一代人工智能可以通过阅读和分析

受众的个性特征,实现意识形态的精准传播,通过提供一种快速、便捷、高效的语言交流方式,使人们更容易学习和理解不同意识形态的观点,有助于消除不同意识形态之间的误解和隔阂。此外,人工智能大模型本身带有某种竞争倾向和政治倾向,[10]ChatGPT 能够赋予使用多种语言的国家中被边缘化的人们更多的权力,能够为组织与个人提供更丰富的信息与更好的交流方式,且在很多领域都将具有广泛的应用空间,信息的力量被无限放大,进而赋能意识形态传播。

(二)数字智能叠加提升利益整合能力

智能技术在当下的发展速度已经突破摩尔定律,这使得新一代人工智能作为生产要素日益成为理解政治数字化转型的重要力量。这有助于衍生出符合大众政治利益诉求和表达逻辑的 ChatGPT 平台,解决传统政治表达中单向灌输、数据孤岛、整合效果差等问题,进而提升其政治利益整合与表达能力。其一,多重技术叠加下的信息聚合能更加清晰地呈现政治表达效果。当前人工智能技术已经把握了数字交流的主要渠道,[11]ChatGPT 作为“通用语言模型”的表现已经超过专用语言模型,新技术新算法能够把政府、普通民众、政治内容深度数字化,在此基础上进行数据挖掘和分析,进而实现对受众进行内容推送、过程监督、效果反馈以及整合效果评估等。其二,ChatGPT 可以融入个性化内容,推动政治整合精准化。随着算法不断升级为具有更大计算能力的虚拟集成器,ChatGPT 拥有海量的数据信息以及强大的学习能力与逻辑演绎能力,虚拟现实能够不限量地提供个性化政治内容,突破受众的时空阻隔,进而实现基于受众数字画像和性格特征的定向的精准化政治整合。其三,ChatGPT 促使广大民众更加接近公共管理决策,通过海量已知数据的训练可以判断那些未知数据,[12]使民众自主解决那些相对容易收集大量数据来分析的稳定且可预测的问题,能够更好地应用到其他工作内容与场景中,为政治决策提供科学依据。总之,新一代人工智能技术的演化与迭代正在直接改变政治诉求方式,相应地也会改变人的政治活动方式,进而提升利益表达和政治整合效果。

(三)人机即时交互增强政治参与效能

现代政治强调民众参与对民主过程的重要性,智能化进程中的受众参与度已然成为政治研究的重要议题,ChatGPT 以超高速度、模拟智能和千亿级连接能力建立起高度拟人化的场景互动模型,能有效保障政府数字化转型的需求,并成为促进政治、经济、社会与文化发展的技术工具,政治参与场景的虚实界限和时空界限进一步被打破,广大受众的沉浸式体验和立体式交互日益增强。一方面,ChatGPT 的即时交互为网络政治参与提供新载体。“人工智能技术带有更多的传感器,模拟人类行为的能力将变得更精细”,[13]更多微观个体数据都能保存在 ChatGPT 中,算法升级也降低了受众线上参与的时空成本,使个体更加容易地表达自己的观点,有助于整个政治过程循环有序进行。ChatGPT 推动政治关系日益多元开放,这种开放性深度影响着政治过程的智能反馈、智能评价和体验,激发了受众的自主参与热情。另一方面,ChatGPT 的智能创造有助于提升政府数字化能力。即时算法交互下的 ChatGPT 缩短了信息获取与政府决策之间的距离,这可以赋能数字政府、降低行政负担、提升政务效率,并体现在工业系统、交通管理、能源管理、环境监控、供应链物流等公共管理和公共服务的情境中,使政府决策形态从传统的人际结合向人机结合转变。总之,ChatGPT 的实时交互与沉浸特性能够充分释放数字化发展红利,推动公民政治参与并实现公共政策的可持续发展目标,这意味着人机即时交互已成为一种新的政治参与范式。

(四)数据去中心化流动重塑虚拟空间主权

如今的人工智能技术已全面进入虚拟空间主权问题研究视野,并推动虚拟空间主权日益成为大国博弈的焦点,数字化治理的精准性为维护虚拟空间主权安全提供了重要指向。ChatGPT

的去中心化属性不断突破传统民族国家的政治属性,为人类在虚拟空间内拓展主权、抢夺话语权提供可能,促使国家间的竞争走向硬件支撑、算力算法等技术层面的博弈和较量。一方面,ChatGPT 模拟人类行为为虚拟空间主权提供新的场域,新一代人工智能技术致力于实现开放共享、便捷灵活、个性智能、及时有效、以用户为中心的立体化服务,为人类文明的发展与传播提供新载体,正在改变着人们的思维方式和行为走向。因此,当前国家间对于智能技术安全的考量,不仅仅是对科技安全等硬实力的维护,更是对人工智能所隐藏着的智能文化、智能价值等在政治方向上的维护。[14]另一方面,国家间通过自主学习、交融合作以及人机协同等方式不断延展虚拟空间主权,如北约国家援助的人工智能军事应用使乌克兰在情报、舆论等多个领域占据优势,成为影响俄乌军事力量对比的重要因素。此外,ChatGPT 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,全球化时空场域中的政治安全问题较以往更加复杂,降低了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。可见,ChatGPT 技术发展不仅不会消灭国家间的竞争,其还会在各国努力开拓技术新边疆的过程中进一步被强化。

三、新一代人工智能技术潜在的政治秩序风险

海德格尔和马尔库塞曾经强调,新技术的嵌入与运用对于政治秩序而言往往是一把“双刃剑”。ChatGPT 的信息传播对价值观、政治态度、政治心理、政治决策、政党竞争以及国际政治等都会产生深远影响。[15]新一代人工智能的走向及其所带来的影响如果不能受到有效控制,就会导致政治关系不可预测、政治空间趋向碎片化以及政治决策不足和无效等问题。

(一)算法沉浸催生主流意识形态认同困境

政治价值普遍存在于算法当中,国家间的技术竞争深层次而言涉及意识形态之争,ChatGPT 的算法依托大数据会根据受众使用习惯推送同质化信息,使受众身处被建构的舒适空间中,形成了艾里·帕雷瑟所说的“过滤气泡”现象,即接触媒介程度越高,对信息“质量”的关注便越低,以避免不适之感。[16]这就从无形中减少了受众判断信息与关注主流媒体的可能性,对传统大众媒介构建的主流价值传播模式的稳定性带来很大冲击。一方面,ChatGPT 带来的信息爆炸,助长了西方意识形态在虚拟空间中的传播。算法沉浸使得超级链接把 ChatGPT 指向到任何地方,这为数字世界的野蛮行为开辟了新空间。ChatGPT 在全球范围内驱动了数据与信息的跨境流动,进而使数据流转至美国等西方国家的服务器中,而技术设计国是什么样的意识形态,就决定了 ChatGPT 遵守什么样的价值判定规则。如 ChatGPT 对左翼政党或思潮就表现出明显的政治偏见,对一些国家和特定人群也有偏见,如不合理规范其价值趋向,就会影响这些国家的社会稳定和主流意识形态安全。另一方面,ChatGPT 的智能化程度较高,很大程度降低了信息生产的技术门槛,海量信息传播也会带来政治叙事极化。ChatGPT 能够根据使用者的输入推算出他的价值取向、政治观点、行为特征,进而分析并引导使用者的价值取向。当意识形态传播被 ChatGPT 不良使用或滥用时,智能算法生成内容的正确率便无法保证,会加快虚假内容生产速度,导致传统的网络意识形态认同“失灵”,进而带来民粹化和政治叙事极化。此外,ChatGPT 创造了海量似是而非的虚假信息,其产生的文本、图像、视频等可轻而易举地以假乱真,当其被境内外敌对势力操纵后,会以数字人的形式在虚拟空间发声,通过制造舆论引导用户成为其不法声音的帮凶,为数字民粹主义意识形态提供新的活动空间。

(二)资本逻辑下审查逃逸带来社会治理难题

新一代人工智能技术开发需要丰富的技术储备、数据基础与庞大的资金支持,目前处于领先地位的均为西方大型互联网企业,这些大公司的企业行为多于政府行为,使技术的工具理性日

益超越其价值理性,资本逻辑下的信息把关、认知操纵功能决定着受众接受的信息内容,算法将传统数字资本主义再升级为“自动化新自由主义”,而且影响着用户的时间、精力、注意力的配置倾向,甚至会带来技术失控风险。一方面,资本的趋利性决定了 ChatGPT 的审查逃逸特征。“资本的力量正在逐渐支配市场,商家和企业依靠技术加持来获得更大利润,很大程度上会加剧恶性竞争。”[17]智能技术嵌入到政府部门,虽然会改变传统的办公形式、简化相应的行政程序,但是获取这些便捷的代价却是,以更为繁缛的方式转移到其他环节。企业推动 ChatGPT 开发的根本目的在于获取利润,而不是单纯为了凸显其社会功能。以往政府可以通过关键词过滤、实名制认证等手段来治理网络言论,但 ChatGPT 技术的出现打破了这种传统的治理方式,使管理权力被各个互联网巨头瓜分,甚至治理主体最终将陷入资本操控下的算法依赖困境。另一方面,新一代人工智能技术的资本逻辑易造成“数字形式主义”。新一代人工智能技术的高度不确定性意味着政府部门往往缺乏预见性,政府既有的监管模式难以对其发展态势进行把握和提前研判。智能技术并没有全面重塑政府,而只是改变了政府的运作方式,诸如维护数字办公系统、更高水平的技术操作技能与项目管理能力培育等,这些努力并无益于政府专注公共管理与公共服务的职能,“无数据不决策”现象容易演化成“数字形式主义”。加之政府部门的人力、技术和制度等资源配备和调度力不从心,监管不得不通盘考虑技术应用的广泛性和深入性。总之,新一代人工智能技术很难归结为某个政府部门的监管职责,部门之间往往难以就监管达成共识并步调一致,甚至对政府部门现有的监管架构带来很大挑战。

(三)新型数字鸿沟下的不平等加剧民主政治分化

新一代人工智能算法推荐会通过收集用户网络行为习惯,进而实现对社会进行控制的目的,这会导致不同群体之间的观念差异越来越大。一方面,不同国家和地区间技术迭代能力与政治发展水平的差异不断增大。新一代人工智能技术需要越来越多的财力和技术支撑,技术富国的垄断发展,拉大了其同落后国家间的数字鸿沟。由于网络设施、通信技术在空间上的不均衡分配,不同国家和地区间存在着数字化差异,发达国家和地区拥有技术研发和规则设定的主导权,基于全网数据训练的 ChatGPT 很容易突破一国的法律限制,一旦其作为政府的官方代表进行对外交流,很容易出现严重的技术强国制约技术弱国的问题。以美国为例,其近年来科技政治化倾向越来越明显,[18]加之西方国家坚持遏制中国的数字经济发展,目前,我们很难在短时间内彻底改变这一现状,国家和地区间的分裂和对立局面将持续加剧。另一方面,群体间技术差异会带来“数字化移民”的技术剥夺感。由于新一代人工智能的技术门槛相对较高,目前仅少数用户意识到其重要性,并且能够顺畅使用 ChatGPT。多数算法素养较低的用户很难顺畅使用 ChatGPT 的高级功能,这会导致产生 ChatGPT 原住民与数字化移民两个群体,依靠数字化和知识创造的财富成为“富人现象”,大多数群体依赖免费服务,处于金字塔的底层,其所获得的信息有限,不利于拓展自身知识,内心就会产生严重的数字剥夺感,而这种技术带来的鸿沟会进一步演化成政治分化。总之,新型数字鸿沟把不同国家、地区、人群的视野局限在特定的圈子里,用户容易被笼罩在新一代人工智能技术织造的信息茧房之中,民意的选择、投向和屏蔽等变得越来越容易被操纵和“计算”,[19]进而造成民主政治分化难题。

(四)制度滞后影响个人隐私和虚拟空间主权安全

“技术改变政治”的逻辑肯定了技术的潜力及其影响,但其忽视了组织变革具有艰巨性。由于缺乏开发机器人技术相关软件的道德、伦理和法律基础,新技术经常处于监管真空状态而得不到应有的规制。一方面,广大民众的个人隐私安全日益受到威胁。智能技术的经济价值驱动导

致个人隐私安全面临着多方面挑战，OpenAI 公司在未经数据权人同意的前提下，已经收集了上千亿条的数据资源，且其中包含大量个人敏感数据。无论是在国家立法还是在行业制度层面，目前，还没有明确的数据使用边界和范围，亦没有形成统一的标准，容易诱发网络“钓鱼”事件。此外，由于缺乏开发虚拟技术和相关软件的道德、伦理标准，新一代人工智能的自主和独立逻辑能力可能会存在一定程度的不可预测性。譬如，在 ChatGPT 大规模推广应用时仍存在性别歧视、种族歧视等缺陷。更有甚者，ChatGPT 不仅极力维护美国利益，而且还秉持美国现行的政治立场，在用户使用时会出现针对同一问题的双标现象。另一方面，新一代人工智能技术会对国家虚拟空间主权形成冲击。海量数据作为训练集合及信息导入与内容生成的源泉，导致数字世界的信息流动无法受到国界限制，使得人工智能模型生成或运行场景中的数据主权无法如传统主权一样清晰，会威胁到国家第五空间主权安全。一些西方国家为了巩固其在虚拟空间的垄断地位，纷纷出台各种鼓励技术研发的制度，如美国 2021 年专门出台鼓励领先技术研发的《无尽前沿法案》，以确保美国在智能技术领域的优势和霸权地位。目前，国内在开源框架、算法模型、编辑器等方面的基础性技术仍落后于发达国家，当 ChatGPT 垄断企业被其背后的国家政治力量影响时，ChatGPT 有可能操纵用户做出违背其最佳利益的事情，甚至是操纵选举、破坏民主进程和传播虚假信息，也可能会进一步分化社会，制造分裂和不信任。

四、多元协同构建 ChatGPT 技术视野下政治安全治理机制

以 ChatGPT 为代表的新一代人工智能技术在给政治发展带来便利的同时，其产生的“数字利维坦”可能会成为颠覆政治秩序的决定性力量。因此，我们应在总体国家安全观的指导下，加快构建新一代人工智能技术向善发展的顶层设计，形成一套完善的用于生成式人工智能风险防范的技术与伦理合力系统，推进生成式人工智能协同监管的制度保障，进而形成组织、技术、伦理、制度多元协同的政治安全治理格局。

（一）顶层引导新一代人工智能技术向善的整体布局

国家性与系统性的技术变革为新一代人工智能技术视野下政治安全提供了顶层设计方向，能够有效应对技术治理路径依赖带来的复杂而动态的挑战。习近平总书记多次强调“要把维护国家政治安全特别是政权安全、制度安全放在第一位”，[20] 不论是国家“十四五规划”还是党的二十大报告对技术变革中的政治安全都有专门论述，在今年的国家机构改革过程中，党中央专门成立了国家大数据局，以期实现跨功能、跨权限、跨层级的整体布局，进而从顶层设计上防范、化解智能技术快速进步带来的潜在风险。

一是制定基于总体国家安全观的新一代智能技术安全规划。基于总体国家安全观规范生成式人工智能技术与产品的合理准入范围与应用边界，关注并重视新一代人工智能技术的发展与应用，投入充足的资金进行技术研发与成果转化，营造良好的发展环境，使其整体方向与国家总体发展规划同向协调。将推动新一代人工智能技术发展上升为一项重要战略，推动技术嵌入—组织变革—制度革新相互串联，明确三者之间的侧重于衔接，整体协调推动政府系统性、全方位、协同式的智能化变革。

二是确立政府数字化转型的内容。首先，将新一代人工智能的结构融入现有行政体系运作框架中，以进一步开发和创新数字政务；其次，完善企业数字化服务平台上性能测试与群众满意度测试的功能和工具，优化数字服务和客户体验流程；最后，成立专家委员会或咨询委员会，注重引进算法领域的高端人才，开发公共数据发布流程，及时更新和维护公共数据集和相关元数据等内容。总之，我们要通过顶层设计打通政企产学研合作通道，尽快出台生成式人工智能领域的团

体标准、行业标准乃至国家标准,顶层设计新一代人工智能行业技术通用标准体系,从而使得新一代人工智能技术的建设具有安全性和可操作性。

(二)协同优化新一代人工智能的技术安全支撑体系

“在一个政治体系中,行政权力覆盖不了的地方,就是政治安全的风险点”,[21]截至2022年底,我国90.5%的省级行政许可事项实现网上受理和“最多跑一次”。随着新一代人工智能在各领域铺开,我们需要把握“系统内”和“系统外”两个布局,通过企业、科研院所、第三方独立机构积极开展新一代人工智能技术研发,多元协同优化生成式人工智能安全治理水平。

一方面,企业和科研院所要加快新一代人工智能自然语言处理技术的研发。结合自身定位及资源禀赋积极布局生成式人工智能相关领域与业务,扩大其应用场景的广度与深度,重视云端算力的硬件这一关键基础设施,加大数字化基础设施的建设,强调针对性和差异化,避免同质化的技术铺设。与此同时,注重安全应用能力的训练培养,从政治安全的高度规范其应用场景和资本逻辑。

另一方面,既要重视“硬件”,也要重视“软件”,齐抓共建负责任算法,加大数据保护技术研发的经费投入。通过资金、项目等资源的投入,将新一代人工智能技术的基础产业牢牢把握在自己手中。通过多方合作消解数字鸿沟,克服经济社会数字化转型过程中产生的数字不平等。充分释放数字化发展红利,以建立一个整体协同、敏捷高效、智能精准、开放透明、公平普惠的数字技术为蓝图。总之,“技术本性”将在智能技术自我迭代和强化中被进一步展现,生成式人工智能不仅是一种数字空间,还是一种生产资源,技术支撑是政治安全治理的重要依托,只有联合各种社会力量,在企业不断完善技术设计,科研单位和社会组织不断发挥作用,共同推动技术发展和应用的基础上充分合作,才能不断打破行业间、国家间的技术壁垒,提升生成式人工智能安全治理水平,实现协同优化治理。[22]

(三)多元共建新一代人工智能安全治理的伦理系统

技术的支撑仅仅代表潜在能力,而能力的运用程度与方式取决于使用者,互动视角注意到了技术应用的“伦理安全”。ChatGPT造就的虚拟空间不像在传统世界中那样,已有一套相对固定的政治秩序,因此需要构建一套维护政治安全的伦理准则,使生成式人工智能更有益于人类世界的发展。

一是不断塑造新一代人工智能技术的社会伦理。由于ChatGPT拥有超强的自然语言处理能力,[23]几何级数式的内容生产不断重塑着人们的思维与认知,会产生许多伦理问题。因此,自主设备和智能系统应该基于人类价值规范和伦理监管体系来运行,这样才能防范其可能带来的算法歧视与偏见;同时还要根植于生成式人工智能的固有特性,积极构建基于虚拟现实的智能技术伦理规范和生成式人工智能道德标准,打造绿色健康的生成式人工智能应用环境。

二是加强国家间的合作,凝聚价值共识。为了有效管控新一代人工智能技术带来的风险,学术界日渐意识到,将其纳入负责任的国际关系体系的重要性。大国间共同合作是虚拟空间安全秩序建构的必由之路,新一代人工智能技术已然更加强化了虚拟空间的广泛联系,技术的去中心化日益促使国家间走向合作,因此需要建立新一代人工智能技术视野下的新型伙伴关系,通过合作共建智能化国家规则,共同应对虚拟空间安全威胁,共同打击网络恐怖主义、网络犯罪。

三是注重新一代人工智能技术安全文化教育。应对“数字利维坦”的一条重要措施是提升民众的思想觉悟,根据民众对新一代人工智能的感知和依赖程度,对算法意识较弱的公众应加强算法知识与应用教育,逐步缩小人工智能意识鸿沟;对ChatGPT强依赖型用户应加强数据干预、数据隐藏、数据阻断等算法抵抗教育。此外,我

们需要重申社会价值高于技术或商业价值,加强技术使用的伦理责任规范,增强对真假信息的甄别力,不盲从 ChatGPT 生成的建议,努力避免算法信息茧房和群体极化现象的出现,促使技术更好地为社会服务。

(四)建立健全新一代人工智能发展的监管制度保障

制度治理与技术治理之间有一种相补充的关系,“社会数字化”不仅会推动现存技术治理的变革,而且也会使社会的法律制度、法律理论与法律文化等发生改变,如果政治系统缺乏足够程度的制度化,突然涌现的政治参与可能会使得政治系统濒临崩溃。[24]新一代人工智能是基于算力、数据和算法三大要素确立和发展起来的,其快速迭代升级会使得原有的制度滞后问题更加凸显,倒逼政府推动现存的法律制度变革。因此,我们要提前着手制定某些资源、某些数据的“防火墙”制度,按照“技术治理—合作治理—制度治理”的思路推行相应的制度约束。

一方面,要建立健全新一代人工智能的监管制度,规避技术滥用。欧盟等监管机构正在努力制定引入审查和问责制的法律,试图减少 ChatGPT 工具造成的与偏见和错误相关的问题。2023 年 5 月 30 日,中共中央审议通过了《加快建设国家安全风险监测预警体系的意见》《关于全面加强国家安全教育的意见》,这对于新一代人工智能技术的合理发展,以及保护公民的隐私和数据安全具有重要意义。诚然,目前有关智能技术发展和应用的法律法规以及规范性文件越来越多,但是这些制度在多大程度上得到了贯彻落实还值得进一步观察。因此,我们还需要进一步细化对新一代人工智能技术发展和应用的备案和审查,从事前审查、事后监督等方面积极制定周详、严密的智能技术应用制度体系。

另一方面,要打破 ChatGPT 的技术鸿沟,推进技术合作与无障碍制度建设。美国等西方国家极力将维护网络霸权作为其核心目标,不遗余力地对中国进行封锁和打压,并坚持以西方发达国家建构的“规范标准”将国家间的对抗延伸至虚拟空间中。为此,我们应坚持将数字化技术与“一带一路”倡议相结合,适当消解信息壁垒、利益壁垒以及体制壁垒,加强新一代人工智能技术“中国标准”建设。同时,不断健全新一代人工智能技术的适老化制度,积极探索建立数字服务地区间、群体间公平机制,打破“信息孤岛”和“数据烟囱”,保障智能技术领域的正义属性,进而实现对民众需求的制度回应以及为民众提供更好的体验环境。

五、结　语

综上所述,我们所处数字时代的重要悖论是一个社会的数字能力越强,这个社会就变得越脆弱。[25]然而,限制人工智能的网络能力相当困难,遏制其扩散也很难。政府、企业和个人都应该着手构建有效的保护体系,以防患于未然。正如技术史学者所言,“技术不是一种天命,而是斗争的舞台。技术是社会的战场,或者用一种更好的隐喻来说,把技术比作一个文明的替代形式相互竞争的‘事态的议会’。”[26]因此,新一代人工智能技术视野下的政治安全治理是一个技术、组织、伦理和制度共同演进的过程。作为新时代维护国家安全的行动指南,总体国家安全观对生成式人工智能应用政治安全治理具有重要指导意义。

我们既要包容审慎地看待新一代人工智能技术对政治安全提供的新动力,ChatGPT 的出现深度验证了强人工智能时代的可能性,技术与社会的深度融合方式为政治安全治理提供更广阔的思考空间。“数字变革”推动政治秩序更加关注意识形态、政治整合、政治表达、主权空间的变化,新一代人工智能技术为政治秩序更高层次的进化提供了基础,人机融合及其方式、途径越来越成为政治安全发展的重要因素。同时,又要打通“技术嵌入—组织运行—数字服务”的链条,以更加全面的资源整合、更加开放的治理格局以及更加明确的标准规范克服技术垄断、各自为

政、制度滞后等问题。这不仅意味着技术的升级,而且意味着前所未有的组织方式和认识范式转变,其中涉及越来越复杂的内在机制,个人不再是孤立的决策者,数字化生存让我们更加注重协同治理,我们要保持观念不断更新,也应看到制度设计、社会伦理在新一代人工智能技术中的协调作用,以此来确保机器智能的发展能够为政治安全服务。

总之,只有治理理念、智能技术、制度规范的有机融合,通过健全"跨部门"政治安全协作完善监管机制,主动化解 ChatGPT 带来的政治秩序风险和挑战,才能实现价值、组织、技术与制度协同的政治安全治理能力和治理体系现代化目标。

注释:

〔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》,《人民日报》2022 年 10 月 26 日。

〔2〕王彦雨:《"强人工智能"争论过程中的"态度转换"现象研究》,《科学技术哲学研究》2020 年第 6 期。

〔3〕张爱军:《人与 ChatGPT 交互政治的可能性质化:风险维度与规约路径》,《学术界》2023 年第 4 期。

〔4〕A. M. Uan Dis Eva, J. Bollen, W. Zuidema, et al. , "ChatGPT: Five priorities for research", *Nature*, 2023(9).

〔5〕余南平:《新一代人工智能技术与大国博弈新边疆》,《探索与争鸣》2023 年第 5 期。

〔6〕Nguyen QuocPhu, VoDucHong, "Artificial Intelligence and Unemployment: An International Evidence", *Structural Change and Economic Dynamics*, 2022(63).

〔7〕〔美〕曼纽尔·卡斯特:《网络社会的崛起》,夏铸九译,北京:社会科学文献出版社,2006 年,第 153 页。

〔8〕郑慧等编:《国家治理·依法治国·政治安全》,北京:中国社会科学出版社,2015 年,第 252 – 253 页。

〔9〕参见〔美〕伊恩·古德费洛、〔加〕约书亚·本吉奥、〔加〕亚伦·库维尔:《深度学习》,赵申剑等译,北京:人民邮电出版社,2017 年。

〔10〕〔10〕高奇琦:《GPT 技术与国家治理现代化:基于秩序、赋权与创新的框架》,《山东大学学报(哲学社会科学版)》2023 年 4 月 27 日网络首发,https://kns.cnki.net/kcms/detail/37.1100.C.20230426.1858.002.html。

〔11〕谢波、李晨炜:《人工智能对国家政治安全的影响机理与应对思考》,《国家安全研究》2023 年第 1 期。

〔12〕方滨兴主编:《人工智能安全》,北京:电子工业出版社,2020 年,第 44 页。

〔13〕〔美〕阿莱克斯·彭特兰:《智慧社会:大数据与社会物理学》,汪小帆、汪容译,杭州:浙江人民出版社,2015 年,第 136 页。

〔14〕刘建华:《人工智能的意识形态属性与风险及其应对》,《吉首大学学报(社会科学版)》2022 年第 6 期。

〔15〕汝绪华:《算法政治:风险、发生逻辑与治理》,《厦门大学学报(哲学社会科学版)》2018 年第 6 期。

〔16〕〔美〕丹尼斯·麦奎尔:《受众分析》,刘燕南等译,北京:中国人民大学出版社,2006 年,第 161 页。

〔17〕〔美〕杰瑞·卡普兰:《人工智能时代》,李盼译,杭州:浙江人民出版社,2016 年,第 96 页。

〔18〕Chi Hung Kwan, "The China – US Trade War: Deep – Rooted Causes, Shifting Focus and Uncertain Prospects", *Asian Economic Policy Review*, 2020, 15(1).

〔19〕参见余潇枫主编:《非传统安全概论》下卷,北京:北京大学出版社,2020 年。

〔20〕中共中央宣传部、中央国家安全委员会办公室编:《总体国家安全观学习纲要》,北京:学习出版社、人民出版社,2022 年,第 59 页。

〔21〕樊鹏、郭静、何建宇等:《国家治理与制度安全新视野》,北京:中国社会科学出版社,2019 年,第 4 页。

〔22〕尹振涛、徐秀军:《数字时代的国家治理现代化:理论逻辑、现实向度与中国方案》,《政治学研究》2021 年第 4 期。

〔23〕Deng, J. , & Lin, Y, "The Benefits and Challenges of ChatGPT: An Overview", *Frontiers in Computing and Intelligent Systems*, 2022.2(2).

〔24〕〔美〕塞缪尔·亨廷顿:《变化社会中的政治秩序》,王冠华等译,北京:生活·读书·新知三联书店,1989 年,第 1 – 86 页。

〔25〕〔美〕亨利·基辛格、〔美〕埃里克·施密特、〔美〕丹尼尔·胡滕洛赫尔:《人工智能时代与人类未来》,胡利平、风君译,北京:中信出版社,2023 年,第 189 页。

〔26〕〔美〕安德鲁·芬伯格:《技术批判理论》,韩连庆、曹观法译,北京:北京大学出版社,2005 年,第 16 页。

〔责任编辑:刘 鎏〕