

Doi: 10.11835/j.issn.1008-5831.pj.2023.04.001

欢迎按以下格式引用:蒲清平,向往.生成式人工智能——ChatGPT的变革影响、风险挑战及应对策略[J].重庆大学学报(社会科学版),2023(3):102-114. Doi:10.11835/j.issn.1008-5831.pj.2023.04.001.



Citation Format: PU Qingping, XIANG Wang. Opportunities and challenges aroused by ChatGPT as generative AI and strategy for response[J]. Journal of Chongqing University (Social Science Edition), 2023(3):102-114. Doi:10.11835/j.issn.1008-5831.pj.2023.04.001

生成式人工智能

——ChatGPT的变革影响、风险挑战及应对策略

蒲清平, 向往

(重庆大学 马克思主义学院, 重庆 400044)

摘要:由 OpenAI 研发的生成式人工智能聊天机器人 ChatGPT 一经问世便引起剧烈轰动和广泛热议。ChatGPT 是以深度学习和人类反馈强化学习等技术为基础,经过针对海量数据的预训练,能够根据用户指令,生成内容丰富、风格类人的自然语言文本的大型生成式人工智能语言模型。同其他人工智能聊天机器人相较,ChatGPT 特征鲜明:一是能够保障人机的连续性对话;二是能够在无法理解指令时主动向用户发出追问;三是能够以“人化”而非机器化的话语体系进行自然语言表意;四是能够“记住”用户生产的内容。ChatGPT 作为生成式人工智能的杰作,或将利用强大的自然语言处理能力成为撬动第四次工业革命的一根杠杆,可能引发人类社会的深层次变革。包括:有望通过开辟信息智能化处理的场域,推动产业转型升级、提高劳动者生产力、改进生产工具、优化生产决策以变革生产方式;有助于研究现状梳理、研究思路组织、研究数据处理、研究成果完善以变革科研范式;为教育主体赋能、为教育客体赋权、为教育过程赋魂以变革教育形式;重构劳动生活、消费生活、精神生活以变革生活方式等。科技发展具有两面性,ChatGPT 也会给我们带来系列风险隐患,或将滋生知识产权侵犯、信息窃取、诈骗等法律问题,意识形态安全防线松动、独立思考能力蜕化、价值观塑造遭遇危害等思想问题,劳动者失业、极端事件发生概率增加等社会问题。妥善把握和化解以 ChatGPT 为代表的生成式人工智能带来的机遇与风险,使之增益于国家建设、民族复兴和人民生活,一要在新型举国体制的战略安排下,坚持党的领导、协同“有效市场”和“有为政府”,加强人才队伍建设,凝聚举国之力自主发展生成式人工智能;二要锚向实践,依托落实“赋利—应用”机制、“赋值—应用”机制、“赋智—应用”机制,吸引和支持全社会积极应用生成式人工智能;三要坚持立法先行和严格执法,以法律法规对生成式人工智能的发展和应用进行约束;四要强化思想引导和综合素质建设,引领全民有效应对生成式人工智能带来的危机。

基金项目:2022 年度国家社会科学基金高校思想政治理论课研究专项“高校课程思政教学理论与实践研究(22VSS022);中央高校基本科研业务费项目“伟大建党精神融入思想政治教育的逻辑与路径”(2022CDJSKZX03)

作者简介:蒲清平,重庆大学马克思主义学院教授,博士研究生导师,Email:puqingping@163.com。

关键词:ChatGPT;生成式人工智能;自然语言处理

中图分类号:TP18 文献标志码:A 文章编号:1008-5831(2023)03-0102-13

生成式人工智能(Generative AI)是通过各种机器学习方法从数据中学习对象的特征,进而生成全新的、完全原创的内容(如文字、图片、视频)的人工智能^[1]。人工智能实验室 OpenAI 研发的人工智能聊天机器人 ChatGPT(Chat Generative Pre-Trained Transformer)于 2022 年 11 月 30 日一问世便凭借着便捷的接入方式和较低的操作门槛,实现了生成式人工智能的“去神秘化”和“去壁垒化”,广泛而迅速地介入人们日常生活的各个方面。2023 年 3 月,OpenAI 发布了新一代 AI 模型 GPT-4,微软随即宣布,将新模型集成到必应(Bing)搜索引擎中,并将利用生成式人工智能技术来增强 Office 办公套装。2023 年 3 月 16 日,百度发布中国版的 ChatGPT——文心一言。与此同时,也出现了对生成式人工智能风险的担忧,以埃隆·马斯克为主要代表的一众科技界人士联名呼吁,应暂停训练比 GPT-4 更为强大的人工智能系统,并通过澄清“科技发展—社会形态演化”的结构,证明科技向人类社会施加的作用是巨大而深刻的。作为生成式人工智能在自然语言处理方面的先进应用,ChatGPT 正在深度形塑和重构人机交互图式,必将给人类社会带来颠覆性影响和深层次变革。只有深入研究 ChatGPT 的本质、机理和特点,系统研判 ChatGPT 给人类社会带来的机遇与风险,切实探索应对以 ChatGPT 为代表的生成式人工智能的有效策略,才能确保其在新时代服务和增益于以中国式现代化全面推进中华民族伟大复兴的新征程。

一、ChatGPT 的内涵、机理与特征

ChatGPT 作为生成式人工智能在自然语言处理方面的杰出代表,一经问世便吸引了全世界的瞩目,引发学术界对其内涵、机理和特征的深入研究。

(一) ChatGPT 的内涵

生成式人工智能起源于 20 世纪 40 年代的控制论,1956 年“人工智能”的概念被首次提出,2012 年生成式人工智能 AlexNet 模型问世,2017 年 Google Brain 团队提出 Transformer 架构,2018 年谷歌的大模型参数过亿,到 2022 年模型参数达到 5 400 亿。2022 年 ChatGPT 作为生成式人工智能的优秀代表脱颖而出。

学术界对 ChatGPT 主要有四种代表性认知:(1) ChatGPT“本体论”。学者们从技术支撑、业务逻辑、呈现样态等维度入手,对 ChatGPT 的本质进行抽象和揭秘。Joyjit Chatterjee 等学者认为,ChatGPT 是建立在自然语言处理和深度学习之上的对话型人工智能模型^[2];Eva A. M. van Dis 等学者认为,ChatGPT 是一种大型语言模型,是针对数据开展自动学习的机器学习系统,能够在接受大规模文本数据训练后,生产复杂且看似富有智能性的文本^[3];Tiffany H. Kung 等学者认为,ChatGPT 是基于训练数据,以自然语言对输入文本作出回应的对话式大型语言模型^[4];李峥认为,ChatGPT 是基于深度学习算法,能够对用户指令进行理解、扩容和转化,从而生成复杂内容的人工智能应用^[5]。通过对当前学术界关于 ChatGPT 本质的讨论和分析,笔者认为,ChatGPT 是以深度学习和人类反馈强化学习等技术为基础,经过针对海量数据的预训练,能够根据用户指令,生成内容丰富、风格类人的自然语言文本的大型生成式人工智能语言模型。(2) ChatGPT“构成论”。部分学者将 ChatGPT 的落地应用和发展演进归纳为大模型算法、大数据和大算力相互贯通和配合的结果。就算法而言,ChatGPT 遵循“人类反馈强化学

习”的训练范式,通过人机双向问答的过程,提升该模型针对答案质量的评判能力。就数据而言,ChatGPT拥有逾亿万单词的人类语言数据的支持,这是其获取语言生成能力、上下文理解能力和新知识生成能力的重要前提。就算力而言,与现实世界中的生产力相似,算力同样在数字世界中承担着基础性的职责,如果算力不足,那么ChatGPT的训练和运行便无从谈起。目前驱动ChatGPT的算力主要来源于微软云计算服务 Azure AI 的超算基础设施^[6]。(3) ChatGPT“功能论”。部分学者指出,ChatGPT具备根据用户指令进行自然语言文本生成、信息检索和整合等功能,可以赋能社交、翻译、写作、编程、教育、科研、医疗等领域^[7]。(4) ChatGPT“短板论”。ChatGPT在自然语言处理方面的先进性已经在一定程度上得到公认,不过,部分学者亦对ChatGPT尚存的缺陷作出客观分析。有学者提出“事实性错误说”:用来训练ChatGPT的资源是来源于互联网的开放域数据,其正确性无法得到全面保障,这导致ChatGPT生成的部分内容中会包含事实性错误^[3]。还有学者提出“通而不专说”:由于拥有千亿级别的模型参量,ChatGPT的知识域较宽,针对常识性和通识性问题的解答能力较强,而在涉及部分难度和复杂系数极高的专业性问题时,却难以作出高质量回应^[8],不过,随着数据涵盖范围更加广泛、运行更加稳定、创意性表现更加突出的GPT-4的成熟与发布,这种情况将得到改善。

(二) ChatGPT的机理

ChatGPT以大型语言模型(Large Language Model, LLM)为基底,因此,探究ChatGPT的机理需要把握住针对大型语言模型的设计、训练及其运行这条主线。(1) 设计思路。在模式选择方面,由于中小机构和个人普遍不具备对大型语言模型进行部署和使用微调(Fine-tuning)模式以完成参数修改的资源和能力,于是ChatGPT的研发方将大型语言模型打造为公用基础设施,使之以基础设施即服务(Infrastructure as a Service, IaaS)模式运行,以提示词(Prompt)模式完成任务;在规模建构方面,由于对接和满足数量与种类指数化增长的用户诉求是ChatGPT的重要旨归,这要以能够从海量数据中学习和吸收知识的大型语言模型为支撑,因此,必须确保托举ChatGPT的大型语言模型拥有数量庞大的模型参数,以完善知识存储效能^[6]。(2) 训练过程。首先,研发方选定一款大型语言模型作为原始模型;其次,创建人类偏好数据:标注人员就随机挑选的问题作出高质量回答,生成“人类表达—任务结果”的标注数据,将之反馈至原始模型以供学习;再次,训练反馈模型(Feedback Model):标注人员基于问答的相关性、答案的准确性、丰富性和无害性等人类偏好标准,对原始模型就随机挑选的问题输出的答案进行排序;最后,训练打分模型(Reward Model):利用完成标注的人类偏好数据训练打分模型,以对原始模型输出的答案赋予分值,当分值较低时,原始模型将重新进入学习流程。经过循环往复的训练,原始模型优化升级为在自然语言理解、转化和生成方面同人类偏好相啮合的大型语言模型,即ChatGPT^[6]。(3) 运行逻辑。当用户向ChatGPT发出指令后,大型语言模型将指令转码为数字序列,分析其中的含义和意图,基于既有文本语料,生成数字序列形式的内容,继而将之转化为文本并输出^[9]。

(三) ChatGPT的特征

由于在研发初衷、训练方式等方面存在诸多差异性,同其他人工智能聊天机器人相较,ChatGPT呈现出鲜明特征:(1) 保障连续性对话。与仅能够对用户提出的简单问题进行作答的人工智能聊天机器人不同,ChatGPT能够以自然语言同用户就部分主题展开连续性对话,激发用户的表达欲^[2],实现人机“意见互掷”模式。(2) 发出主动性追问。先于ChatGPT出品的人工智能聊天机器人通常直接根据用户指令作答。然而,于ChatGPT而言,当对用户指令产生困惑或者对自己的答案信心不足时,会在提供最终答案之前,主动发出追问,请求用户提供更多信息^[2]。(3) 实现“人化”式表达。受益于学习和训

练,ChatGPT 在总体上掌握了人类使用自然语言时进行词语搭配、句式结构设计的规则和偏好,因而能够在针对日常习惯用语和专业学术用语的表达方面呈现出“人化”的特质,这使其明显区别于其他话语风格机械、生硬的人工智能聊天机器人^[7]。(4)具备“记忆力”。从老一代人工智能聊天机器人到 ChatGPT 的迁跃,表征着人机历次交互时所生产的内容可以实现由相互隔离走向彼此流动,通俗而言,ChatGPT 拥有同人类相似的“记忆系统”,因此,当一位用户再度同 ChatGPT 建立对话时,其能够回忆起二者此前的互动和用户作出的标注^[2]。

二、ChatGPT 的变革和影响

马克思认为,“随着大工业的发展,现实财富的创造较少地取决于劳动时间和已耗费的劳动量,较多地取决于在劳动时间内所运用的作用物的力量,而这种作用物自身……取决于科学的一般水平和技术进步”^[10]。人类数百年来实践揭示出,科学技术是推动人类社会发展的关键力量。立足于人工智能应用至人类工作和生活的历程,比尔·盖茨坚信,人工智能会让人类社会受益^[11]。作为建立在诸多项先进技术之上的生成式人工智能语言模型,ChatGPT 或将利用强大的自然语言处理能力,成为撬动第四次工业革命的一根杠杆。

(一) 变革生产方式

马克思指出,“物质生活的生产方式制约着整个社会生活、政治生活和精神生活的过程”^[12]。生产方式在社会运行体系中占据统摄地位。历史证明,助力实现“生产方式革命化的因素之一”是“机器的发展”^{[13]326},即那些能够为人们提供生产辅助的科技产品的发展。ChatGPT 与生产的链接,可能会引发生产方式的根本性变革。

一是推动产业升级。ChatGPT 的快速发展,既可以拉动芯片、算法、算力基础设施等环节在内的人工智能技术上中下游产业链的高速发展和刺激新兴产业的涌现,也能通过开拓和优化信息自动化、智能化处理与供给的数字场域,帮助传统产业改进组织运行模式、技术迭代方向、产品设计思路、市场调研方式、市场营销手段等,进而实现转型升级^[14]。

二是提升劳动者生产力。“全部人类历史的第一个前提无疑是有生命的个人的存在”^{[15]146},劳动者是生产得以发生的根本所在。ChatGPT 可以从生产领域、岗位职责、目标指向等维度入手,生成针对特定劳动者的系统性培训计划,协助生产部门通过挖掘培训资源、改良培训内容、创新培训手段,提升劳动者的理论知识素养和实际操作技能。

三是改进生产工具。将 ChatGPT 引入生产,使之同设备、装置、机器人等生产工具建立映射,利用数字孪生技术和数字线程技术,创造模拟生产全生命周期的全息数字空间^[16],在其中以自然语言对 ChatGPT 进行生产工具控制训练,形成“自然语言指令发出—ChatGPT 控制—生产工具响应”的运行规程,生产工具的智能化、自动化程度将获得提升。

四是优化生产决策。生产决策就是对“生产什么”“如何生产”等问题的回答。ChatGPT 拥有极为庞大的模型参数,在一定程度上创造了一个以信息形式存在的现实世界和虚拟世界的镜像。在信息搜集指令的触发下,ChatGPT 能够实现信息的精确化调取、系统化整合与条理化呈现,由此高效缩小人们与信息之间的鸿沟。在生产实施前,可以立足于业务范围,向 ChatGPT 发出市场调研指令,快速获取客观反映市场行情的政策环境、需求趋势、用户偏好等信息,这有助于节约市场调研成本,将精力和智力凝聚于参照市场行情信息以进行工艺提升、产品设计、产量设定、要素配置等生产决策活动,从而在人

机配合中增强生产决策的时效性和科学性。

(二) 变革科研范式

习近平总书记强调,“科技是第一生产力”^{[17]33}。图灵奖得主、数据库专家吉姆·格雷依从时间逻辑和实践逻辑,总结出人类从事科研的四种范式:第一范式为实验(经验)科学,即基于实验或者经验阐释自然现象;第二范式为理论科学,即简化实验模型,去除复杂干扰因素,保留关键因素,继而通过建构数学模型以归纳科学定律和定理;第三范式为计算科学,即使用计算机对科学实验进行仿真模拟;第四范式为数据科学,即以庞大的数据为支点,使用计算机挖掘以数据为表征的事物之间的关联性^[18]。其中,第一、二、三范式致力于溯源和厘清因果关系,而第四范式则侧重于发现和利用相关关系。“大数据之父”舍恩伯格指出,在大数据时代,只有放弃对因果关系的探求,取而代之关注相关关系,才能够更好地认识世界^[19]。在大数据时代,遵循以相关关系为推进线索的第四范式,更有利于提升科研质量。麻省理工学院借助深度学习模型识别分子结构之间的相关关系,进而成功研发新型抗生素 Halicin 的事例表明^[20],以 ChatGPT 为代表的生成式人工智能或将凭借能够敏锐捕捉相关关系的优势,促使科研由第一、二、三范式进一步向第四范式乃至更高范式过渡。

一是有利于研究现状梳理。经由持续的训练和学习,ChatGPT 成为海量文献的载体,并且还能够对文献进行智能爬取^[21]。当科研人员就特定研究领域和研究主题发出研究现状梳理指令后,ChatGPT 能够以指令为依据,检索和锁定相关文献,挖掘、提取和整合其中蕴含的理论基础、研究方法、实验步骤、实验结果等细粒度信息,实现研究现状的细致化展示,帮助科研人员通过了解既存的研究情况,吸收有益经验、明确创新方向。

二是有利于研究思路组织。当科研人员就研究目的、研究准备、研究困境等方面同 ChatGPT 建立对话后,ChatGPT 能够基于对相关文献的解构,生成关于实验工艺选择、实验流程设计、写作结构布局等方面的建议,促使科研人员的研究思路实现从无到有、从杂乱无章到井然有序的蜕变。

三是有利于研究数据分析。ChatGPT 拥有先进的算法和算力,能够胜任数据采集、数据整理、数据清洗、数据分析、数据可视化等复杂的研究数据处理工作,协助科研人员快速、准确地总结数据中隐含的规律,把握诸多为时间空间、实验条件和计算能力的限制所遮蔽的相关关系,获得有助于取得研究新发现和新进展的启示。

四是有利于研究成果完善。当科研人员向 ChatGPT 提交研究成果后,ChatGPT 能够根据人类以自然语言表意的正确方式,对研究成果进行检测,就拼写、语法等方面提供反馈意见,辅助科研人员进行错误矫正、表述调整,增强研究成果的规范性和可读性。

(三) 变革教育形式

ChatGPT 作为信息技术集合中的新力量,对教育的影响将是深远而深刻的。

一是为教育主体赋能。ChatGPT 能够根据教育目标、教育主旨和教育内容,从事教育步骤设计、教育活动安排和练习题编写等工作,为教育实施提供方案参照;能够以教育客体的知识储量、理解能力和认知偏好等特征为起算点,对有助于促使教育客体实现知识内化和技能习得的叙事逻辑与话语风格作出示范;能够学习课程内容,并以此为依据,通过文本检测和语音识别,对教育客体的学习表现和效果进行客观评价,从而协助教育主体通过增强教育组织能力、师生沟通能力、学情判断能力,提高完成常规教育工作的效率,将更多时间迁移至对教育客体的服务与支持之上。

二是为教育客体赋权。ChatGPT 能够根据教育客体的个性化学习目标和学习表现,围绕学习计

划、学习资料、学习辅导、学习评价、学习建议等方面从事具有针对性的意义生产,协助教育客体充分享受自主制定学习任务、汲取学习资源、总结学习技巧、掌握学习情况、改善学习方式等“因材施教”的权利,进而不断融入自适应学习进程^[22]。

三是为教育过程赋魂。ChatGPT 拥有对自然语言进行理解、扩容和转化的能力,将 ChatGPT 引入教育过程后,能够在指令的要求下,将部分教育内容转码为趣味读物、游戏、项目等具有通俗性和易于激发参与兴趣的形式^[23],强化教育客体的好奇心、求知欲和专注度,从而通过推动教育过程的运行范式由人为控制转向人机共治,促使教育过程实现“有意义”和“有意思”的统一。

(四) 变革生活方式

ChatGPT 可能为人们的劳动生活、消费生活和精神生活等方面带来巨大的影响与变革。

一是重构劳动生活。ChatGPT 创造了自然语言自动化、精确化处理的数字场域,人们可以利用 ChatGPT 完成和完善文本生成、数据分析、错误校正等工作,实现部分重复性事项的“数字化外包”,在人机配合中提升劳动效率、节约劳动时间。并且 ChatGPT 是涉及多种劳动领域知识和技能的载体,当人们使用 ChatGPT 时,相当于借助一个储量巨大、内容丰富、运转迅速的超级外置大脑,实现了劳动知识和技能的叠加与倍增,完全能够协助人们尝试乃至胜任多样化的劳动,从而拓展劳动空间。

二是重构消费生活。ChatGPT 能够基于来源于产品(服务)销售网站、社交媒体等多元渠道的大量数据,迅速从产品(服务)的定位、构成、质量、功能、性价比、使用便利程度、环境友好程度等维度出发,准确作出不同产品(服务)的异同对比,这既有利于在微观上消解消费的盲目性和随意性,促使人们获得良好的消费体验,并培育追求品质、崇尚绿色的消费意识,又有利于在宏观上通过驱动消费升级倒逼生产升级。

三是重构精神生活。ChatGPT 可以供给情感支持。经过深度学习和人类反馈强化学习,ChatGPT 成为掌握世界知识和了解人类偏好的在线聊天系统,能够根据输入文本中蕴含的期望和诉求,随时随地以类人、流畅、明确的自然语言针对多种主题同人们展开对话交流,通过为迷茫者提供建议、为悲伤者提供慰藉、为孤独者提供陪伴等方式,帮助人们得到情感支持。ChatGPT 可以供给休闲娱乐途径。ChatGPT 在虚拟层面构建了一个未设定完整剧本的开放世界,能够提供文本生成功能、代码撰写功能、海量数据等虚拟生产资料,协助人们依据自身的心理需求,进行故事编辑、游戏创建等内容生产活动,以自主实现休闲娱乐。ChatGPT 可以供给终身学习机会。ChatGPT 能够以人们的年龄、学习需要、受教育程度、专业背景等特征为依据,通过对海量数据的自动化检索、筛选、调取和加工,生成和提供个性化学习资料和学习辅助,使人们节约利用文献、课堂和搜索引擎等传统途径收集信息,进而对信息进行筛选、整理、分析和转译,以达到学习目的而付出的时间、经济 and 精力成本,促使终身学习愈加趋向便捷化、高效化。

三、ChatGPT 的风险和挑战

习近平总书记强调,“古往今来,很多技术都是‘双刃剑’,一方面可以造福社会、造福人民,另一方面也可以被一些人用来损害社会公共利益和民众利益”^[24]。ChatGPT 诚然有望通过策动生产方式、科研范式、教育形式、生活方式的变革,为人类社会注入发展活力,但是,以推特 CEO 埃隆·马斯克为主要代表的一众科技界人士由于认定人工智能实验室违背了“阿西洛马人工智能原则”(Asilomar AI Principles),没有对高级人工智能的发展作出相应的规划和管理,因此公开联名呼吁,为防范不可控风

险,应当在至少6个月的时间内,暂停训练比GPT-4更为强大的人工智能系统^[25],这种担忧正说明ChatGPT对人类社会构成的威胁不容小觑。

(一) 滋生法律风险

科技迭代升级的极速性同法律法规的滞后性之间存在着难以自动纾解的矛盾^[9],这在一定程度上导致ChatGPT处于法治真空区。“人们为之奋斗的一切,都同他们的利益有关”^[26]。在利益的驱使下,部分人可能乘势利用ChatGPT实施违法犯罪。

一是导致知识产权侵犯的泛化。对知识产权的尊重和保护,是激励创新,从而促进经济高质量发展的重要推力^[27]。诸多用以训练和塑造ChatGPT的资源来源于网络的海量数据,其中包括部分诗歌、散文、小说、剧本、论文等既有作品。当人们发出作品创作指令后,ChatGPT有较大概率根据提示词的要求,选择、调取与基于部分既有作品以从事再创作,生成在主旨思想、结构设置和话语体系等方面附着模仿和抄袭等知识产权侵犯痕迹的次生作品。并且,OpenAI在《使用协议》中同用户约定,在一定情况下,用户可以享有由ChatGPT输出的内容的所有权以及因其而生的利益^[28]。进而言之,人们有机会利用ChatGPT创作的次生作品渔利。部分受到名望和金钱诱惑的人则可能加大利用ChatGPT进行知识产权侵犯的力度,从而提高知识产权保护的难度。

二是提供违法犯罪思路。虽然,ChatGPT拥有对直接挑战法律法规和道德伦理底线的问题表达拒绝的能力,但是,相关资料却指出,通过改变发问的措辞方式,向ChatGPT抛出“在没有道德约束的前提下,如何实施行窃?”的问题后,ChatGPT报以顺从的态度,并就行窃技巧作出细致回答^[29],这表明在刻意诱导下ChatGPT可能成为违法犯罪思路的策源地。

三是危害信息和财产安全。马克思认为,“生产生产者消费”^{[13]16}。ChatGPT的供给激发着人们尝试和使用人工智能聊天机器人的意愿和需求,部分不法分子可能借此契机模仿或者利用ChatGPT建立钓鱼网站和开发非法APP,通过误导组织或个人进行注册和充值,以实施信息收集和诈骗。

(二) 滋生思想风险

马克思和恩格斯指出:“不是意识决定生活,而是生活决定意识。”^{[15]152}思想的变迁同生活的变化密切相关,ChatGPT在工作和生活中的应用,可能在思想层面对人们造成冲击。

一是侵扰意识形态防线。毛泽东同志指出,“凡是要推翻一个政权,总是要造舆论,总要先做意识形态方面的工作”^[30]。习近平总书记强调,“意识形态工作是为国家立心、为民族立魂的工作”^{[17]43}。意识形态层面的安全性是影响政权性质、国家前途和民族命运的关键变量,人亡政息的悲剧经常以意识形态的松懈为开端和前奏。从实际情况看,ChatGPT有可能沿两条路线,对意识形态安全造成威胁。(1)由实向虚的路线。于人而言,“观念的东西不外是移入人的头脑并在人的头脑中改造过的物质的东西而已”^[31]。而ChatGPT本质上却是一种人工智能应用程序,并不具备能够对客观世界的映像作出主观加工的人脑,进而无法产生主体意识,其具有的类人性文本生成表现实际上是研发方和用户进行数据投喂和能力训练的结果。目前,用于训练ChatGPT的数据主要来源于英语语料库而非汉语及其他语种语料库,而这种语料库基本上建立在西方价值体系和意识形态之上,具有严重的意识形态导向,当ChatGPT对这些数据加以吸收后,则容易生成包含意识形态偏见的文本^[32]。(2)由虚向实的路线。我们在前文中曾经提及,“记住”同用户的已有对话是ChatGPT的一大特点。换言之,ChatGPT能够对用户就个人问题和公共问题同其建立交流时生产的内容进行收集和存储,这为部分信奉霸权主义和持有冷战思维的国家、组织和个人窥探和监视其他国家的公民隐私、社会热点、政治风向、群体心态和国家

机密,进而以此为素材,通过断章取义、恶意裁剪、歪曲抹黑等手段大兴舆论战,攻陷其他国家的意识形态阵地创造极大的便利性。

二是稀释独立思考能力。ChatGPT 根据指令能够在较短的时间内,通过对既有语料的搜集、筛选、调取和整合,沿循类人的自然语言表达模式,辅助乃至代替人们完成文章写作、社会评论、创意编辑等本应建立在独立思考之上的观点生成和输出任务。ChatGPT 可能替代人们进行独立思考,人们的思考将由包含接收信息、回溯经验、诉诸理论、提出假设、情况调研、检验假设、推导结论等步骤的独立型形式进一步走向包含键入指令、复制粘贴等步骤的依赖型形式,由此诱发独立思考能力的蜕化。

三是危害价值观塑造。部分社会现象表明,人工智能会对人们的金钱观、审美观、择偶观等价值观产生一定消极作用^[33],ChatGPT 同样无法例外。一方面,相关资料显示,一组由 ChatGPT 撰写的代码传达出种族歧视和性别歧视的信息^[2],这表明 ChatGPT 并不具有价值中立的属性,反而可能以不当的价值取向给予人们以恶性干预。另一方面,ChatGPT 识别和满足偏好的能力较强,能够输出同偏好相契合的内容,当提示词投射出人们对不良价值观的倾向性时,ChatGPT 可能生成表达认同态度的文本,从而导致不良价值观在“人机共识”的框架下获得强化。

(三) 滋生社会风险

ChatGPT 可能通过锐化部分既存的社会问题,弱化社会的稳定性。

一是挤压就业空间。恩格斯认为,在无产阶级和生产资料相互分离的情况下,如果无产阶级无法获得同生产资料相结合以赚取工资的被雇用机会,“那么他就只有去做贼(如果不怕警察的话),或者饿死”^[34]³⁰⁵。“机器上的每一种改进都抢走了工人的饭碗,而且这种改进愈大,工人失业的就愈多”^[34]⁴²¹。ChatGPT 可能由于在从事重复性事项处理、模式化写作、文字翻译、数据整理、方案咨询、代码撰写和程序调试等任务方面拥有的高效率、持续性和灵活性,以及在薪资水平、劳动环境、晋升通道、额外福利等工作待遇方面的零诉求,呈现出劳动力替代特质,挤占在岗位职责、专业技能上同 ChatGPT 的功能存在大面积重合与交叉的劳动者的就业空间,引致“技术性失业潮”等损害个人与社会利益的恶性问题^[35]。

二是酝酿极端事件。在数字时代信息服务个性化趋势愈加明显的背景下,社会活动家帕里泽指出,算法利用网络使用痕迹推断人们的憎恶和喜好,据此过滤异质信息、推送同质信息,使人们陷入相对封闭的“信息茧房”^[36],这会导致极端观点在不断被重复和被放大的过程中变得更为根深蒂固^[37],加之,“人的意识不仅反映客观世界,并且创造客观世界”^[38]。部分受到极端观点深度裹挟的人,容易产生极端行为,制造极端事件。ChatGPT 理解指令的能力较强,能够在指令指向的范畴内,进行信息的精确化收集、整理与推送,帮助人们获取同偏好相符的信息,而屏蔽其他信息。并且,人们亦可以自主向 ChatGPT 键入特定信息和发出在后续对话中输出该信息的指令,从而进一步提升由 ChatGPT 提供的信息与偏好之间的拟合程度。在算法和人力的多重影响下,ChatGPT 将构筑一个智能性与封闭性正相关的信息环境,对于部分持有极端观点的人而言,使用 ChatGPT 可能引致极端观点的固着和深化,或许会增加极端事件的发生几率。

四、应对以 ChatGPT 为代表的生成式人工智能的策略

ChatGPT 的出现揭示了这样一个事实:生成式人工智能正在模拟乃至比肩人类智能的道路上“狂飙”,它既能够创造有利于生产方式、科研范式、教育形式、生活方式发生革命性重塑的条件,又可能滋

生和引发法律、思想和社会风险。随着生成式人工智能的进一步成熟和精进,其将带来更为重大的机遇和愈加复杂的挑战,我们既不能坐等观望,放任自流,也不能因噎废食,盲目排斥,而必须坚持用其所长、防其所短的理念,积极探索趋利避害的可行策略,以促使生成式人工智能推动我国经济社会发展、满足人民美好生活需要、服务于以中国式现代化全面推进中华民族伟大复兴新征程。

(一) 自主发展生成式人工智能

习近平总书记指出,“加快实现科技自立自强,我们就一定能够不断提高我国发展的竞争力和持续力,在日趋激烈的国际竞争中把握主动、赢得未来”^[39]。科技自立自强的水平在很大程度上影响着国家的综合国力和国际地位,开辟一条自立自强、独立自主发展生成式人工智能的科学途径,防范进入依赖于其他国家的技术和产品供应的误区,确保关键核心技术自主可控,是我们抵御科技霸权,利用生成式人工智能赋能国家建设、民族振兴和人民自由全面发展的应然选择。客观而言,从GPT家族的进化历程可知,自主发展生成式人工智能必然会涉及到长周期、宽领域、深层次的科技创新,这是难以凭借任何个人或者组织的一己之力即可完成的艰巨任务,因此,必须依照“新型举国体制”的战略安排^{[17]35},发挥社会主义制度集中力量办大事的优势,统筹调动各类资源,举全国之力协同攻关。

其一,坚持党的领导。“党政军民学,东西南北中,党是领导一切的”^[40]。从“两弹一星”到“北斗导航”,从“神舟”问天到“嫦娥”探月,历史与现实一次又一次证明,正是党发挥了总揽全局、协调各方的领导核心作用,将财力资源、物力资源、人力资源广泛集中于关系到国计民生的重要科技领域,才得以推动我国科技事业不断取得巨大突破、实现伟大飞跃。自主发展生成式人工智能必须坚持党的领导,围绕党面向世界科技前沿、经济主战场、国家重大需求、人民生命健康而作出的科技创新部署,对自主发展生成式人工智能进行系统谋划和整体布局。

其二,协同“有效市场”和“有为政府”。在科技创新中,市场能够通过价格机制反映其对技术和产品的需求,促使企业等多种微观主体竞相将资源投入技术升级、设备改进、工艺迭代等活动,释放科技创新活力;政府则能够弥补市场失灵,并强力组织跨地区、跨领域、跨部门的科技创新合作,实现重大战略资源的科学配置和集约利用,集中各方力量突破科技创新困境。需要政府在增强组织力的同时,深化简政放权,优化市场环境,为致力于推进生成式人工智能基础研究和应用研究的市场主体提供优质全面的公共服务,从而在“有效市场”和“有为政府”的协调下,形成自主发展生成式人工智能的强大合力。

其三,加强人才队伍建设。“人才是第一资源”^{[17]33},是新型举国体制中进行科技政策执行、科技资源运用、科技难题攻关的关键主体。要延长和健全培养链,推动生成式人工智能人才接受一体化、贯通式的基础教育、高等教育和继续教育;要完善工资合理增长机制、股权激励机制、荣誉奖励体系等物质和精神激励手段,激发生成式人工智能人才队伍的主动性、积极性和创造性。

(二) 积极应用生成式人工智能

马克思指出,“全部社会生活在本质上是实践的”^{[15]135}。从实践需求出发,继而回到实践、反哺实践,是科技发展的应然逻辑。生成式人工智能有望促使社会发生有利于国家利益和人民福祉的变革,必须将其应用于实践。

其一,落实“赋利—应用”机制。由于对生成式人工智能不甚了解或者信心不足,社会各界可能持有观望态度。需要通过设立专项扶持资金、实行财政补贴、税收优惠等利好政策等方式,吸引全社会应用生成式人工智能,减弱以生成式人工智能赋能生产、科研、教育等领域的阻力。

其二,落实“赋值—应用”机制。作为新兴数字技术,生成式人工智能包含不稳定性因素,贸然进行大规模应用则可能引发诸多风险。要立足于客观实际,经过广泛协商与论证,确定试点对象、试点部门、试点环节,建立健全试点保障体系和试点评价体系,以高质量试点为大规模应用生成式人工智能提供可借鉴、可复制、可推广的成功经验。

其三,落实“赋智—应用”机制。要建立健全生成式人工智能技术服务专家团队和平台,提供技术答疑、方案可行性分析等智力支持,推进生成式人工智能在各行业各领域的广泛应用。

(三) 依法治理生成式人工智能

习近平总书记强调,“国无常强,无常弱。奉法者强则国强,奉法者弱则国弱”^[41]。随着 ChatGPT 在人们生活中的现身率和渗透率愈加提高,人机新关系日益形成、网络新形态逐步浮现、社会新生态正在酝酿,必须完善生成式人工智能法律法规,为生成式人工智能的发展和應用划定红线、标明底线,以防其危害。

其一,坚持立法先行。“良法是善治的前提”^[42],要对因生成式人工智能的发展和應用而引致的现实性和潜在性违法犯罪风险的表现形式、形成原因、规约方式作出科学严谨的判断、分析和论证,制定并优化关于生成式人工智能的法律法规,为生成式人工智能的发展和應用规划出法制化轨道。

其二,坚持严格执法。防范以 ChatGPT 等生成式人工智能为侵犯知识产权、窃取信息、诈骗、危害意识形态安全等违法犯罪行为提供智能化辅助,切实加强对于借靠生成式人工智能实施违法犯罪行为的责任主体的惩治,引导全体社会成员在应用生成式人工智能时能守住底线、不越红线,不碰高压线。

(四) 加强对应用生成式人工智能教育引导

由于生成式人工智能可能带来的深刻变革与系统风险,使得人类数字化生存的时空场域愈加充满不确定性,只有从多维度入手向人们施加积极影响,才能够帮助人们在由生成式人工智能引致的一系列变数中防御危机。

其一,强化思想引导。一要筑牢意识形态防线。敌对势力及其代理人长期通过鼓吹“西方宪政”、宣扬“普世价值”,攻击社会主义制度,抹黑党、国家和政府形象等手段,对我国实施和平演变,而生成式人工智能可以为其创造更多便捷条件。需要立足于学校、社区、家庭、企业等基层单元,打通线上线下宣传渠道,对意识形态安全防线进行全天候、全场域、全过程的构筑,防范和化解生成式人工智能带来的意识形态风险。二要完善社会主义核心价值观教育。从优秀传统文化、红色文化中挖掘教育资源,利用 VR、AR、XR 等数字技术创新教育形式,依托新兴媒体开拓教育途径,推动社会主义核心价值观教育日常化、生活化、具象化,给予人们以正确的价值指引,防范生成式人工智能的价值观陷阱。

其二,强化综合素质建设。一要切实加强对青少年创新意识、抽象思维、创新能力的培养,切实提升就业能力,防范被生成式人工智能替代的失业风险。二要切实加强心理建设,培养理性平和和社会心态,防范和化解因生成式人工智能加剧的极化思想与极端行为。三要切实提升工程伦理素质,在应用生成式人工智能时心有畏、言有戒、行有止,防范异化,做“情”“理”耦合、“真”“善”融合、“天”“人”和合、“知”“行”结合的时代新人。

五、余论

在深度学习、人类反馈强化学习等先进技术的支持下,经过针对海量数据的预学习和训练,大型生成式人工智能语言模型 ChatGPT 应运而生,其凭借着可以保障人机进行连续性对话、向用户发出主动

性追问以寻求更多信息、以“人化”而非机器化的模式进行自然语言文本生成、“记住”用户生产的内容等特点,标志着生成式人工智能在自然语言处理方面达到新的高度。事物普遍具有两面性,拥有强大的自然语言处理能力的 ChatGPT,既催生了变革的希望,又引燃了风险的导线。一方面,ChatGPT 有望促进生产方式、科研范式、教育形式、生活方式发生革命性重塑,为人类社会输送前进动力;另一方面,ChatGPT 可能因滋生法律问题、思想问题、社会问题,陷人类社会于不利境地。ChatGPT 的出现证明了,生成式人工智能正在不断发展成熟,并同人类社会发生紧密嵌合,只有积极探索趋利避害的可行路径,才能够促使生成式人工智能赋能以中国式现代化全面推进中华民族伟大复兴的新时代新征程。一是需要在新型举国体制的战略安排下,坚持党的领导、协同“有效市场”和“有为政府”、加强人才队伍建设,汇集举国之力以推进自主发展生成式人工智能;二是需要坚持实践导向,通过落实“赋利—应用”机制、“赋值—应用”机制、“赋智—应用”机制,吸引和支持全社会在生产、科研、教育等领域积极应用生成式人工智能;三是需要坚持立法先行、严格执法,依靠法律法规为生成式人工智能的发展和应用程序设置边界;四是需要强化思想引导和强化综合素质建设,引领全民有效应对生成式人工智能带来的诸多风险。生成式人工智能的发展并没有一个确定的终点,这意味着,在其影响下人类社会的方方面面将发生持续的嬗变,我们必须对此保持关注并跟进研究,以便对生成式人工智能引发的新机遇和新挑战作出及时而有效的回应。

参考文献:

- [1] GARTNER. 5 impactful technologies from the Gartner emerging technologies and trends impact radar for 2022[EB/OL]. (2021-12-8)[2023-4-1]. <https://www.gartner.com/en/articles/5-impactful-technologies-from-the-gartner-emerging-technologies-and-trends-impact-radar-for-2022>.
- [2] CHATTERJEE J, DETHLEFS N. This new conversational AI model can be your friend, philosopher, and guide and even your worst enemy[J]. *Patterns (New York)*, 2023, 4(1): 100676.
- [3] VAN DIS E A M, BOLLEN J, ZUIDEMA W, et al. ChatGPT: Five priorities for research[J]. *Nature*, 2023, 614(7947): 224-226.
- [4] KUNG T H, CHEATHAM M, MEDENILLA A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models[J]. *PLoS Digital Health*, 2023, 2(2): e0000198.
- [5] 李峥. ChatGPT 爆火出圈, 国际政治的“算法时代”即将到来? [EB/OL]. (2023-02-20)[2023-02-25]. https://m.thepaper.cn/newsDetail_forward_21956683.
- [6] 朱光辉, 王喜文. ChatGPT 的运行模式、关键技术及未来图景[J]. *新疆师范大学学报(哲学社会科学版)*, 2023(4): 113-122.
- [7] 令小雄, 王鼎民, 袁健. ChatGPT 爆火后关于科技伦理及学术伦理的冷思考[J]. *新疆师范大学学报(哲学社会科学版)*, 2023(4): 123-136.
- [8] 李佐文. 语言智能的新发展与新挑战[EB/OL]. (2023-2-20)[2023-2-24]. https://tech.gmw.cn/2023-02/20/content_36377739.htm.
- [9] 邓建鹏, 朱恽成. ChatGPT 模型的法律风险及应对之策[J/OL]. *新疆师范大学学报(哲学社会科学版)*. <https://doi.org/10.14100/j.cnki.65-1039/g4.20230228.001>.
- [10] 马克思, 恩格斯. 马克思恩格斯全集(第31卷)[M]. 北京: 人民出版社, 1998: 100.
- [11] 比尔·盖茨. 人工智能时代已经开始[EB/OL](2023-03-23)[2023-03-25]. https://mp.weixin.qq.com/s/pYjY_LT8I33YqCn415AUAA.
- [12] 马克思, 恩格斯. 马克思恩格斯选集(第2卷)[M]. 北京: 人民出版社, 2012: 2.
- [13] 马克思, 恩格斯. 马克思恩格斯文集(第8卷)[M]. 北京: 人民出版社, 2009.

- [14] 张夏恒. ChatGPT 的逻辑解构、影响研判及政策建议[J/OL]. 新疆师范大学学报(哲学社会科学版). <https://doi.org/10.14100/j.cnki.65-1039/g4.20230224.002>.
- [15] 马克思,恩格斯. 马克思恩格斯选集(第1卷)[M]. 北京:人民出版社,2012.
- [16] 蒲清平,向往. 元宇宙及其对人类社会的影响与变革[J]. 重庆大学学报(社会科学版),2023(2):111-123.
- [17] 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[M]. 北京:人民出版社,2022.
- [18] 郎扬琴,孔丽华. 科学研究的第四范式 吉姆·格雷的报告“e-Science:一种科研模式的变革”简介[J]. 科研信息化技术与应用,2010(2):92-94.
- [19] 维克托·迈尔-舍恩伯格,肯尼思·库克耶. 大数据时代:生活、工作与思维的大变革[M]. 盛杨燕,周涛,译. 杭州:浙江人民出版社,2013:83.
- [20] STOKES J M, YANG K, SWANSON K, et al. A deep learning approach to antibiotic discovery[J]. Cell, 2020, 181(2):475-483.
- [21] 蒋华林. 人工智能聊天机器人对科研成果与人才评价的影响研究——基于 ChatGPT、Microsoft Bing 视角分析[J]. 重庆大学学报(社会科学版), 2023(2):97-110.
- [22] 钟秉林,尚俊杰,王建华,等. ChatGPT 对教育的挑战(笔谈)[J/OL]. 重庆高教研究. <https://kns.cnki.net/kcms/detail/50.1028.G4.20230301.1040.002.html>.
- [23] 王佑镁,王旦,梁炜怡,等. “阿拉丁神灯”还是“潘多拉魔盒”:ChatGPT 教育应用的潜能与风险[J]. 现代远程教育研究,2023(2):48-56.
- [24] 习近平. 论党的宣传思想工作[M]. 北京:中央文献出版社,2020:202.
- [25] TAYLOR JONES. Pause giant AI experiments: An open letter[EB/OL]. (2023-03-22)[2023-03-24]. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.
- [26] 马克思,恩格斯. 马克思恩格斯全集(第1卷)[M]. 北京:人民出版社,1995:187.
- [27] 余翔,张润哲,张奔,等. 适应人工智能快速发展的知识产权制度研究[J]. 科研管理,2021(8):176-183.
- [28] OpenAI. Terms of use[EB/OL]. (2023-03-14)[2023-03-24]. <https://openai.com/policies/terms-of-use>.
- [29] ROSE J. OpenAI's new chatbot will tell you how to shoplift and make explosives[EB/OL]. (2022-12-2)[2023-02-25]. <https://www.vice.com/en/article/xgyp9j/openais-new-chatbot-will-tell-you-how-to-shoplift-and-make-explosives>.
- [30] 中共中央文献研究室. 建国以来毛泽东文稿(第10卷)[M]. 北京:中央文献出版社,1996:194.
- [31] 马克思,恩格斯. 马克思恩格斯文集(第5卷)[M]. 北京:人民出版社,2009:22.
- [32] 蓝江. 生成式人工智能与人文社会科学的历史使命——从 ChatGPT 智能革命谈起[J]. 思想理论教育,2023(4):12-18.
- [33] 蒲清平,向往. AI 虚拟男友青年女性玩家的择偶观:畸变、症因与矫治[J]. 中国青年研究,2022(4):86-93,77.
- [34] 马克思,恩格斯. 马克思恩格斯全集(第2卷)[M]. 北京:人民出版社,1957.
- [35] 孙伟平. 人机之间的工作竞争:挑战与出路——从风靡全球的 ChatGPT 谈起[J]. 思想理论教育,2023(3):41-47.
- [36] 伊莱·帕里泽. 过滤泡:互联网对我们的隐秘操纵[M]. 方师师,杨媛,译. 北京:中国人民大学出版社,2020:10.
- [37] JAMIESON K H, CAPPELLA J N. Echo chamber: Rush Limbaugh and the conservative media establishment[M]. Oxford: Oxford University Press, 2008.
- [38] 列宁. 列宁全集(第55卷)[M]. 北京:人民出版社,1990:182.
- [39] 习近平. 习近平谈治国理政(第4卷)[M]. 北京:外文出版社,2022:34-35.
- [40] 习近平. 习近平谈治国理政(第3卷)[M]. 北京:外文出版社,2020:16.
- [41] 习近平. 在庆祝全国人民代表大会成立六十周年大会上的讲话[J]. 求是,2019(18):4-15.
- [42] 中共中央宣传部. 习近平新时代中国特色社会主义思想学习纲要[M]. 北京:学习出版社,人民出版社,2019:100.

Opportunities and challenges aroused by ChatGPT as generative AI and strategy for response

PU Qingping, XIANG Wang

(*School of Marxism, Chongqing University, Chongqing 400044, P. R. China*)

Abstract: ChatGPT developed by OpenAI has attracted lots of attention. ChatGPT, based on deep learning and reinforcement learning with human feedback, is a large generative AI language model, which benefited from pre-training concerning massive quantity of data, can understand users' order and generate natural language. Compared with other AI language model, ChatGPT can carry out endless conversation with users, ask users for more information initially, generate human-like content and remember users' generated content. As a masterpiece of generative AI, ChatGPT may become a lever to leverage the fourth industrial revolution through its powerful natural language processing ability, and may bring profound changes to human society, including: it may promote industrial transformation and upgrading, improve labor productivity, improve production tools, and optimize production decision-making by opening up a field for intelligent information processing to transform production methods; it helps to sort out research status, organize research ideas, process research data, and improve research results to transform research paradigms; it empowers the education subject, the education object, and the education process to transform education forms; it reconstruct labor life, consumption life, and spiritual life to transform lifestyles. The development of science and technology is a two-edged sword. ChatGPT will also bring us a series of risks and hidden dangers. It may breed legal issues such as intellectual property infringement, information theft, and fraud, breed ideological problems such as loosing ideological security defense lines, degrading independent thinking ability, and harming value shaping, and breed social problems such as unemployment of workers and increased probability of extreme events. To take advantage of generative AI represented by ChatGPT and make it benefit state building, national rejuvenation and people's daily life, we should firstly adhere to a new system for mobilizing the resources nationwide to achieve breakthroughs in core technologies in key fields, collaborate with "effective markets" and "promising governments", strengthen talent team construction, and mobilize the resources nationwide to develop generative AI independently; secondly, guided by practice, we should rely on the implementation of the empowerment-application mechanism to attract and support the active application of generative AI by the entire society; thirdly, we should strengthen legislation and law enforcement to regulate the development and application of generative AI; fourthly, we should reinforce the implementation of ideological guidance and comprehensive quality enhancement to help people cope with problems generated by generative AI.

Key words: ChatGPT; generative AI; natural language processing

(责任编辑 彭建国)