

角色互动与人机共善：生成式人工智能在未成年人教育中的应用及其治理

周瑞春 王涵洱 吴淑燕*

【摘要】生成式人工智能以其独特的算法和强大的数据处理能力，能够在人机互动中实现文本、图像、音视频之间的多模态内容生产与转换，在未成年人教育领域具有广阔的应用前景。本文认为，在人机互动的角色赋能中，生成式人工智能可以提升未成年人的自主学习能力及效率、为个性化教育与教育公平提供可能、丰富未成年人的数字文化生活。与此同时，生成式人工智能技术犹如一把双刃剑，既能给未成年人教育带来革新机遇，也存在着教育质量摇摆、价值观塑造挑战以及人际交往主体迷思等多重风险隐忧。对此，本文从推进家校社联动的多元共治、加强角色定位下的监管规制、推进数字平权下的多样可及服务、提升教师与学生人工智能素养等方面提出治理对策。

【关键词】生成式人工智能；未成年人教育；人机共善；人工智能治理

一、引言

生成式人工智能（Artificial Intelligence Generated Content, AIGC），一般是指基于智能算法、模型与规则等生成文本、图像、音视频等技术^①。近年来，我国紧跟全球人工智能技术发展前沿，积极推动生成式人工智能在科教文卫、新闻出版、影视娱乐等各个领域的技术发展与应用。作为当前数字世界的“原住民”，未成年人不仅是网络用户的重要组成部分，更是新技术浪潮的积极参与者，对生成式人工智能的使用展现出了极高的热情与需求。未成年人群体处在生命周期中身心变化波动较为剧烈的成长阶段，“生而赛博”的他们能够迅速、熟练地使用生成式人工智能产品，善于运用技术为其提供学习、生活、社交等方面的便利。早在2019年，中共中央、国务院印发的《中国教育现代化2035》就明确提出要加快信息化时代教育变革，“利用现代技术加快推动人才培

* 周瑞春，北京师范大学中国社会管理研究院特聘研究员；王涵洱，北京师范大学新闻传播学院博士研究生；吴淑燕（通讯作者），中国移动通信有限公司研究院主任研究员。本文系国家社科基金重大项目“文化强国背景下公民道德建设工程研究”（编号：21&ZD060）和北京师范大学互联网发展研究院—中国移动研究院联合研发项目“生成式人工智能技术发展中的科技伦理框架研究”（编号：BNUIDI2023）阶段性成果。

① 李韬、周瑞春：《生成式人工智能的社会伦理风险及其治理——基于行动者网络理论的探讨》，《中国特色社会主义研究》2023年第6期。

养模式改革，实现规模化教育与个性化培养的有机结合”^①。由此可见，大力发展面向未成年人群体的生成式人工智能，既匹配了未成年人群体不断增长的教育发展需求，也符合当前我国教育改革数字化转型的现实需要。

与此同时，面对生成式人工智能可能带来的社会伦理风险，管控与治理也应同步进行。联合国教科文组织2021年11月通过了《人工智能伦理问题建议书》，2023年9月颁布了《生成式人工智能教育和研究指南》，旨在推动生成式人工智能与教育的深度融合，建议将使用人工智能工具的年龄限制为13岁及以上，以显示在技术发展中对未成年人权益与身心健康的重视与保护^②。同样，面对这一划时代的科技变革，各国也纷纷从国家战略高度出发，对生成式人工智能发展进行了全面规划，以期在充分利用其带来的机遇的同时，妥善应对相关的挑战和风险。作为全球数字发展的重要力量之一，中国也在积极规划生成式人工智能的发展应用。2019年，中国国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》，2021年又发布了《新一代人工智能伦理规范》；2023年，国家网信办联合多个部门发布《生成式人工智能服务管理暂行办法》，明确了我国在生成式人工智能技术发展应用上的监管重点，并为服务提供者 and 使用者设定了责任与要求^③。在一个日益被技术渗透的社会里，做到放大技术优势，规避潜在风险，从而确保未成年人健康、安全地使用技术，将是未来我国生成式人工智能研发与治理的重要关注点之一。综上，本文对生成式人工智能在未成年人教育中的应用进行深入探讨，对人机互动中的角色呈现与角色风险进行梳理，并由此提出治理之策。

二、角色赋能：融入教学与生活的良师益友

生成式人工智能在游戏设计、艺术创作、数字社交等各个新兴领域得到迅速应用，并展现出强大的技术赋能潜力。在这一智能化浪潮中，未成年人作为一个特殊且重要的群体，对新兴技术有着敏锐的感知力和独特的使用习惯，其参与度和影响力不容忽视。因此，生成式人工智能在未成年人日常学习生活中扮演怎样的角色、如何通过角色赋能助力教育发展，就成为当下社会各界普遍关注的重要问题。

（一）提升未成年人教育互动质量

《第5次全国未成年人互联网使用情况调查报告》显示，未成年人网上学习的比例高居上网活动的榜首^④。由此可见，利用数字技术进行网上学习是未成年人互联网技术使用的重要目的之一。近年来，线上学习的渠道和方式不断拓宽，数字技术能够为用户提供兼具权威性与个性化的学习方案，让教育脱离时空场景限制，使随时随地、长期终身的学习成为可能。未成年人可以通过在线教育平台、学习网站、电子数据库等获取各种学习资源，还能够足不出户，聆听名师专家们的视频课程或线上讲座，与学习社群中的同龄人交流学习体会。生成式人工智能在未成年人教育中的应用潜

① 《中共中央、国务院印发〈中国教育现代化2035〉》，中国政府网，http://www.gov.cn/zhengce/2019-02/23/content_5367987.htm，访问日期：2024年5月12日。

② *Guidance for Generative AI in Education and Research*, UNESCO, <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>，访问日期：2024年5月12日。

③ 《生成式人工智能服务管理暂行办法》，中国政府网，https://www.gov.cn/gongbao/2023/issue_10666/202308/content_6900864.html，访问日期：2024年5月12日。

④ 《第5次全国未成年人互联网使用情况调查报告》，中国互联网络信息中心，<https://www.cnnic.net.cn/n4/2023/1225/c116-10908.html>，访问日期：2024年5月12日。

力引起了教育界的广泛关注。学者们普遍认为,生成式人工智能是推动未来教育行业向人机协同范式转变的重要动力之一;有别于此前单一的教育辅助类数字技术,生成式人工智能可以被视为是一种多方位、个性化的学习助手,而不仅仅是提高学生单一技能的工具^①。换言之,在未成年人教育应用中,生成式人工智能可以成为辅助教师开展教学、帮助学生自我学习的“非人类行动者”,成为未成年人的良师益友。

一方面,生成式人工智能可以显著提升未成年人的自主学习效能。生成式人工智能产品操作往往十分便捷,具有强大的自然语言处理能力,能够以线上文字、音频、视频等对话方式与用户进行互动,对于未成年人而言具有良好的技术可供性。尤其是在未成年人进行知识性内容的自主学习时,生成式人工智能就像一个随时可以提供帮助的全科“家教”,用易于理解的语言向未成年人提供各类基础知识,自动化处理低阶思维任务(如知识基本回忆、分类、比较和摘要提炼等),从而让学生能够投入更多时间和精力在分析内容、综合知识、创造性思维等学习任务上,全面提升学生的综合素养。以学生写作任务为例,在人机协同的写作模式下,生成式人工智能不仅能够帮助未成年人对文字进行润色,还能及时、快速地反馈评价,纠正排版、语法等细节并推荐改进策略,从而全面提升未成年人的语言应用与写作能力。

另一方面,生成式人工智能也为个性化教育与教育公平的实现提供了可能。生成式人工智能能够适应学生的学习速度与习惯,以学生的个性化特征和需求为基础辅助学习,并给予有针对性的学业建议,这种定制化教育能力很好地弥补了传统课堂教育难以真正实现“因材施教”的难题。此外,在教育欠发达地区通过数字技术推广“替代教育”,不仅能够弥补偏远地区优质师资匮乏的问题,更能通过人机融合知识语境打破教与学的边界,鼓励学生提出更富逻辑性的问题,成为数字技术赋能下虚拟在场的“老师”^②。与此同时,生成式人工智能的出现也将教师从繁复、机械的工作中解放出来,将更多精力放在培养学生素质、提升教学质量的活动中,为不同的学生提供最为适合的教育方式^③。值得一提的是,对于部分存在生理障碍的未成年人来说,生成式人工智能能够满足他们的特殊需求,并且减少学习带来的沮丧与尴尬情绪。比如,部分生成式人工智能具备语音与文字转换服务,能够为残疾学生提供学业帮助,从而实现教育的个性化、精细化与普惠化。

(二) 拓展未成年人数字文化生活

有学者认为,文化产业是最适合数字技术应用的领域^④。在数字技术赋能之下,文化的生产与传播突破了传统渠道中鉴赏能力、财富能力、时间与空间等限制,将文化以便捷的方式带入大众日常生活。生成式人工智能技术的飞速发展也为未成年人的数字文化生活提供了更为丰富的选择,成为未成年人学习之余的生活良伴。

首先,生成式人工智能进一步丰富了未成年人的线上娱乐渠道,满足了他们的课余放松需求。Z世代的未成年人追求个性、自由和平等,注重互动、分享和合作,在网络空间,他们能够通过多元的方式娱乐身心,以兴趣为基础形成了各种网络文化社群,在文化交流中获得满足感、认同感与

① Ciampa M, Bronstein B. ChatGPT in Education: Transforming Digital Literacy Practices. *Journal of Adolescent & Adult Literacy*. 2023, 67(3).

② 吴虑、杨磊:《ChatGPT 赋能学习何以可能》,《电化教育研究》2023年第12期。

③ 杨宗凯、王俊、吴砥等:《ChatGPT/生成式人工智能对教育的影响探析及应对策略》,《华东师范大学学报(教育科学版)》2023年第7期。

④ 江小涓:《数字时代的技术与文化》,《中国社会科学》2021年第8期。

成就感。生成式人工智能在娱乐平台中的应用显著拓展了用户的数字文化生活体验。例如，一些视频软件在嵌入大模型智能搜索功能后，通过对视频内容的精确分析，能够提炼出详细剧情介绍，帮助未成年人更好地理解故事情节，以符合未成年人心理特征与观看习惯的形式，提升了未成年人在观看时的互动性、主观能动性。一些网络游戏也通过训练大语言模型，满足游戏玩家的个性化角色互动需求，提高游戏的趣味性，增强玩家黏性。对于充满好奇心的未成年人来说，在游戏中与生成式人工智能的互动将成为他们习以为常且充满乐趣的事情。

其次，生成式人工智能在与未成年人的互动中，建构起一种人机融合共生的内容生产网络。生成式人工智能能够根据用户指令输出多种模态的内容，帮助未成年人更轻松地创作数字内容，从而激发未成年人的表达欲望，使他们能够以前所未有的方式展现自己的想象力和创造力。例如，“作为世界模拟器的视频生成模型” Sora 的问世突破了早前生成式人工智能在叙事逻辑与创意想象上的边界，从二维到三维的生态演进让用户得以通过人机共塑体验非凡的“全真数字世界”^①。复杂、专业的内容创作数字工具不再成为阻挡文化内容生产的“高门槛”，在生成式人工智能的辅助下，未成年人以文字、图片、视频等多种形式呈现的青春故事得以在网络空间更加快速、便捷、广泛的记录和传播。这种随时随地分享的“数字日志”不仅成为未成年人探索世界、表达自我、相互交流的重要载体，同时也为他们构建了一个丰富且包容的文化交流空间。

最后，生成式人工智能为未成年人数字生活提供情感陪伴。网络人际交往中所展现的平等性、匿名性和共鸣性等特点，能够以更轻松的方式满足未成年人群体渴望归属、尊重和合群的心理需求。对于习惯数字社交的年轻人而言，情感实践不仅限于人与人之间的交流，人机交互可能是他们满足社交需求、情感陪伴的新工具。当未成年人在现实中遭受痛苦情绪而无人倾诉时，就会倾向于走进网络世界寻找帮助^②。生成式人工智能产品可以凭借其丰富的语料库，实现与未成年人用户的多轮智能沟通，对未成年人展现出个性化的关怀，承担情感疗愈者的角色。与生成式人工智能交流，不仅有助于拓展未成年人的社交能力，辅助学习社交礼仪，而且能够随时寻求更多的支持和引导，甚至训练出专属自己的数字伙伴。对此，有学者认为，最理想的人机“共生”交往状态不仅不会取代现实社交，反而会增强人类在交往中的体验^③。

三、角色风险：完美期待下的不完美隐忧

生成式人工智能在未成年人教育中有着广阔的应用前景，发挥着积极的正向效能。然而，作为“良师益友”的生成式人工智能并非是毫无瑕疵的“完美”存在，其在教育应用实践中也存在一些客观的问题与困境。比如，当前市场上针对成年人社交的聊天机器人虽已广泛应用于多个平台，但专为未成年人群体设计的人工智能大模型仍较为罕见。这就使得品类繁多的生成式人工智能产品，难以对未成年人群体进行个性化教育赋能。与此同时，由于技术发展的不可预见性，允许未成年人与生成式人工智能产品进行社交互动，也存在一定的不确定性风险。

（一）无所不能却语焉不详：对教育质量的冲击与摇摆

尽管生成式人工智能能够为未成年人自主学习提供便捷帮助，但技术的可错性仍然有可能会给

① 赵红勋、王佳慧：《拟造的“真实”：Sora的叙事逻辑及其伦理边界》，《新闻爱好者》2024年（网络首发）。

② 雷雳、柳铭心：《青少年的人格特征与互联网社交服务使用偏好的关系》，《心理学报》2005年第6期。

③ Nuryana Z, Pranolo A. ChatGPT: The Balance of Future, Honesty, and Integrity. *Asian Journal of Psychiatry*, 2023, 84.

教育高质量发展带来危机。一些生成式人工智能所依赖的训练数据库并没有对内容来源进行严格的筛选,且数据的来源和时间期限也受到限制,可能会生成未经证实的、虚假的甚至恶意捏造的内容信息。换言之,在技术的不透明性及算法黑箱的遮蔽下,生成式人工智能十分擅长通过看似很有说服力的、理性的方式陈述回答,这就使得缺乏辨别力的未成年人用户容易受到各类回答的误导。此外,有学者研究发现,除了回答质量难以保证外,生成式人工智能在进行一些诸如发明新课程或开发创造性的教学方法时也缺乏表现力,因此,教师在教育活动中的主导地位是无法被完全替代的^①。尽管人类的学习能力尚有局限,但在知识形成这一“有机的、能动的、自主的”^②过程中,生成式人工智能依然无法体悟现实中所感受与理解的内容,也就不能被视作彻底颠覆性的、能够在混合环境中自主进化并作出最优选择的超级智能^③。而且,当教师和学生都在“工具理性”思维支配下陷入技术崇拜的漩涡时,人在教育活动中的自主性和能动性可能逐渐丧失。此外,在生成式人工智能不断深入教育应用的过程中,由技术可靠性不足导致的数据污染、算法偏见、算法歧视等问题,将直接影响到教育质量;不同生成式人工智能产品由于自身大模型语料库不同、算法标准不一、算力性能差异,还会使得同样的课程内容产生不同的教育效果,无法确保教育质量高水平持续发展,处于一种不确定性的“摇摆”之中。

(二)成熟的朋友还是“狡猾的同龄人”:价值观塑造的挑战

在基础教育阶段,学校与课堂不仅是未成年人学习知识的空间,更要提供一种互动性的社会情境,通过师生交流互动,为未成年人树立正确的世界观、价值观、人生观提供良好的环境。生成式人工智能在教学应用中,不仅能够成为师生之间沟通的媒介工具,也在与人类主体的不断互动中成为某种意义上的非人类主体,从而达到从工具到行动者的转化^④。

一般认为,生成式人工智能可以像一个知识渊博的朋友一样给予未成年人全方位的知识辅导。然而,相关研究也显示,生成式人工智能本身的一些技术性认知不成熟,使得其在决策时有时会表现的如同一个懵懂的未成年人^⑤。比如,在回答一些伦理难题时,生成式人工智能往往会表现出比成年人更大的冒险倾向;也更容易受到社会线索,即发问人语言的影响,在不甚了解的领域会表现出过高的“自信”。此外,需要注意的是,由于数据语料库来源的不确定性,以及算法黑箱,生成式人工智能大模型在意识形态上难以做到客观中立。在依赖生成式人工智能进行决策时,生成式人工智能扮演着同伴的角色,通过看似成熟的语言对未成年人进行引导,可能会增加未成年人受不良价值观负面影响的风险,甚至在这种虚拟同伴团体的影响下,产生不健康、冒险的行为^⑥。与此同时,在使用生成式人工智能时,未成年人产生的与自身相关的数据可能会被技术这一“狡猾的同伴”以一种未知的方式存储、审查和用于改进系统,从而泄露未成年人隐私。此外,在人机互动开展自主学习的过程中,生成式人工智能的协助还可能引发未成年人的道德危机。比如,依赖数字技术能够轻松完成任务,容易变相鼓励未成年人养成惰性,丧失在课堂中学习的积极性;由于生成

① Dempere K, Hesham A, Ramasamy K. The Impact of ChatGPT on Higher Education. *Frontiers in Education (Lausanne)*, 2023, 8.

② 陈昌凤、张梦:《由数据决定? AIGC的价值观和伦理问题》,《新闻与写作》2023年第4期。

③ 杨庆峰:《人工智能神话、超级智能及其合约伦理学》,《山西大学学报(哲学社会科学版)》2023年第6期。

④ 苏慧丽、张敬威:《机器的“人化”与人的“机器化”:智能时代教育的主体性危机与破解》,《现代远程教育研究》2024年第1期。

⑤ Edelson M, Roue E, Singh A, et al. How Decision Making Develops: Adolescents, Irrational Adults, and Should AI be Trusted With the Car Keys? *Policy Insights from the Behavioral and Brain Sciences*, 2024, 11(1).

⑥ 陈少华、周宗奎:《同伴关系对青少年心理健康的影响》,《湖南师范大学教育科学学报》2007年第4期。

式人工智能输出内容的学术诚信度难以辨别，课后作业和在线考试抄袭与作弊的可能性也大大增加。过度依赖生成式人工智能会影响未成年人群体批判性思维的发展，甚至让其养成逃避思考的陋习，让技术最终变成“压制”学生发展的力量，有悖于基础教育立德树人的基本要求。加之不断与生成式人工智能交互，在完美技术的向下凝视中，未成年人不免产生被监视的“冷恐慌”，在与强大“智力”的对比下产生对自身参与教育活动及自我主体价值的怀疑^①。如果不能正确地规划和配置好生成式人工智能技术在未成年人教育中的应用场景、频率以及程度，就可能会放大数字技术弊端，削弱人在教育活动中的主体性地位，给未成年人的健康成长带来不良影响。

（三）驯顺的机器抑或后人类智能体：人机交往中的主体迷思

随着生成式人工智能技术的不断发展，各类AIGC大模型逐渐超越了早期单一任务型的对话交互，呈现通用型对话的特征，能够通过先对话上下文的记忆与理解，与用户进行多轮复杂的、有挑战性的对话。对于未成年人而言，与生成式人工智能的互动十分新鲜，但人机交往中的准社会交往并不能提供全面的发展情境，也并非所有的未成年人都能很好地适应这种新型交互形式。例如，有研究发现，与人类互动相比，学生在与人工智能进行互动时，对任务表现出较低的兴趣^②。如果过于沉浸在人机准社会交往中，还可能会导致未成年人与教育活动中的其他主体之间的隔阂，影响其沟通合作能力的提升，不利于其长远的社会化发展，阻碍未成年人健全社会人格的主体性培育进程。此外，生成式人工智能在与用户的交互过程中，容易存在“讨好”用户的价值“对齐”倾向，即在交流中根据用户的引导，自动匹配与情景预设更为契合的角色定位，与用户保持叙事中的一致性^③。一方面，未成年人容易在与AIGC大模型交谈时产生一种舒适的“合拍”感，认为机器人是与自身志同道合的朋友；另一方面，在这一对话舒适圈，技术也呈现出一种被“驯顺”的类人化角色，会使未成年人产生受到尊重和服务的感觉，久而久之容易滋生高高在上的操控感，而这种社交操控感的养成，显然不利于在人与人的现实社交活动中形成平等互动的良好社会关系。面对技术的服从，缺乏判断力与控制力的未成年人容易体验主导性的快感，进而在生成式人工智能提供的个性化服务与量身定制的信息摄取中更加沉迷，逐渐沦为深陷人机世界的技术异化者。更加值得重视的是，当人机之间的互动不仅限于信息互换，而逐渐走向异质主体间性的情感交互时，生成式人工智能便不能够被简单地看作一种“类人”属性的机器^④，而要从后人类的角度重新审视技术在交往中的主体性及可能引发的问题。有学者判断，当下的人工智能体有“行为主体的资格，而无心灵主体的地位”^⑤。在技术尚未完全发展时，将其完全看作与人类生命体平等的主体进行交往显然存在一定的潜在风险。例如，研究证明生成式人工智能在与学生交互时表现得缺乏同理心，难以对用户情绪进行反思^⑥。并且，不能忽视的是，生成式人工智能的研发仍然符合资本运作的逻辑，企业往往会倾向开发更有收益的产品，增加社会整体性异化的可能^⑦。总而言之，在人机交互的过程

① 苏慧丽、张敬威：《机器的“人化”与人的“机器化”：智能时代教育的主体性危机与破解》，《现代远程教育研究》2024年第1期。

② Dempere K, Hesham A, Ramasamy K. The Impact of ChatGPT on Higher Education. *Frontiers in Education (Lausanne)*, 2023, 8.

③ Koh J. “Date Me Date Me”: AI Chatbot Interactions as a Resource For the Online Construction of Masculinity. *Discourse, Context & Media*, 2023, 52.

④ 张微、彭兰：《ChatGPT与人机交往的现实和未来》，《未来传播》2023年第4期。

⑤ 杜骏飞：《论AI交往的三个基本理论问题》，《新闻大学》2024年第3期。

⑥ Dempere K, Hesham A, Ramasamy K. The Impact of ChatGPT on Higher Education. *Frontiers in Education (Lausanne)*, 2023, 8.

⑦ 李韬、周瑞春：《生成式人工智能的社会伦理风险及其治理——基于行动者网络理论的探讨》，《中国特色社会主义研究》2023年第6期。

中，如何保持人类自身的主体性不被瓦解消弭，是人机共生时代必须直面的重要问题。

四、向善而生^①：构建人机共善的智能社会共同体

生成式人工智能为人类社会带来了变革性的机遇，但其隐藏的风险也在为一个不可预见、危机四伏的未来敲响警钟。未成年人群体因其自身发展阶段的特殊性，既是新技术的受益者，同时也是风险易感人群。如何通过有效的治理及干预，降低未成年人在使用生成式人工智能时的负面影响，最大程度地发挥技术在其成长过程中的正面作用，是当下亟须解决的重要问题。基于前文对生成式人工智能角色呈现、角色风险的梳理，本文将从多元共治、角色规约、数字平权和素养先行四个维度提出生成式人工智能的治理路径。

（一）多元共治：推进生成式人工智能应用的家校社联动

生成式人工智能在未成年人教育中的应用，涉及家庭、学校、社会等多个场景，因此加强多元共治，首先就要明确在各个场景下利益相关者应负起的监管与治理责任。当下学界对人工智能责任问题还存在一定的争议，但是实现多元主体间的有机互动，构建共建共治共享的治理共同体是生成式人工智能治理的理想模式^②。政府、技术供应方、学校、家庭等都应作为积极参与的行动者主体，承担一定治理与保护的责任，推进“家庭—学校—社会”联动下的多元共治。

具体而言，从国家层面来说，既要鼓励生成式人工智能技术的创新研发与应用，构建健康的应用生态体系，也要健全相应的法律监管机制，通过法律法规对生成式人工智能在教育行业的应用进行合理约束与审查。除外部的约束以外，生成式人工智能的技术提供方也应该保持较高的伦理道德底线，遵守相关规章制度，及时修正算法和数据中可能存在的虚假信息、潜在偏见等问题，严格保护用户个人信息隐私和数据安全等基本权利。作为未成年人教育活动发生的主要场所，学校应主动调整现有的教学与考评机制，鼓励教育工作者积极运用AIGC大模型辅助教学，调整对学生各项学习任务的考核方式，完善技术赋能下的考评体系；加大对未成年人数字素养教育的投入力度，提高未成年人自我保护意识，对未成年人的技术使用进行正确引导；提升教师运用AIGC大模型开展教学的实操能力，建立健全相关工作的考评激励机制。身为未成年人监护人的家长，也应当对未成年人和技术的互动负起监督责任，在现实生活中对其投入更多的关爱，并在发现问题后及时疏导，保障未成年人的身心健康发展。

（二）角色规约：确保生成式人工智能以人为本

无论生成式人工智能进化到何种发达程度，其所有活动都应站在维护人的尊严与基本权利的准线上，这种人本主义思想也与我国基础教育“以人为本”的理念相契合。也就是说，不管生成式人工智能有着怎样的角色呈现，都应该坚持以人为本的基本原则，加强对其角色呈现的限制和约束，确保把未成年人的健康发展放在第一位。

一是要加强对大模型产品开放应用的角色伦理审核与法治监管。面对生成式人工智能教育应用

^① 这一概念，由北京师范大学互联网发展研究院在2024年3月发布的《向善而“生”——生成式人工智能技术发展中的科技伦理框架研究报告》首次提出并阐发。一方面，包含着《道德经》“道生一，一生二，二生三，三生万物”的中国古代传统文化智慧；另一方面，也包含着在生成式人工智能技术发展应用中，应加强对科技向善全过程引导和规制的伦理意涵。

^② 李韬、周瑞春：《生成式人工智能的社会伦理风险及其治理——基于行动者网络理论的探讨》，《中国特色社会主义研究》2023年第6期。

的角色呈现与角色风险，应加强对技术算法的全过程监管，加大对大模型产品设计的角色伦理审核与法治监管，推进算法透明与无害，并防止用户产生过度依赖。具体而言，生成式人工智能企业在推出各类产品前，应加强对产品角色定位、设计和呈现的审查监管；应披露数据集情况及算法运行逻辑，并在使用初期以较为通俗易懂、符合未成年人心理特征与需求的方式向未成年人讲解该平台的运作机制；防止未成年人因欠缺认知而造成后续使用中的过度沉迷，在监测到有未成年人用户可能出现沉迷现象时，平台应及时对其使用进行合理干预。

二是要做好大模型语料库的“消毒”更新与常规“养护”。生成式人工智能预训练语料库的更新有赖于对人类活动数据信息与时俱进的深入理解，这一理解过程是人机互动融合下“人被投到事物中，事物被投到人中”的双向奔赴。生成式人工智能预训练语料库的更新有赖于对人类输入内容的不断理解深入，一些用户输入的有害信息可能会被储存在其庞大的数据库中，并可能在与其他用户对话时被启用，从而对未成年人造成误导与伤害。因此，应确保符合“善智”价值目标的生成式人工智能发展方向，尽量避免产生恶意的、有悖社会主流价值观的、对未成年人身心健康发展具有特定风险或潜在危害的信息。大模型企业应当在生成式人工智能产品开发中，加入针对未成年人保护的特定算法，将一些不利于未成年人健康成长的内容进行适当屏蔽。正是基于上述考量，定期对大模型语料库进行“消毒”更新与常规“养护”就十分重要。此外，大模型企业也应承担起产品“把关人”的角色作用，加强常态化自查，不断优化算法，从数据源头剔除不良信息，对于恶意散播有害信息的用户及时警告与处理，使其他用户免受其害。

（三）数字平权^①：提供人人可及的大模型服务

生成式人工智能技术应用的最终目标是提升全人类福祉，推动人的自由全面发展。为实现这一目标，就必须要以数字平权为核心原则，确保技术使用的普遍可及性、技术赋能社会发展的公平正义性。

一是加强人工智能基础设施建设，弥合人工智能数字鸿沟。当下我国网络普及率已大幅提高，单一互联网接入的数字鸿沟逐渐消弭，但随着生成式人工智能技术发展应用，可能会给不同地区的未成年人带来一种新型的“AI沟”。这就需要在国家战略层面出台相应的政策引导、产业扶持、税收减免、人才储备、科研项目配套等，优化资源配置、协调各方利益、实现产业联动发展，确保未成年人平等接入与使用生成式人工智能的权利。

二是推出针对未成年人适用的场景化应用大模型。由于生成式人工智能技术接入简单，操作便捷，可以结合未成年人不同阶段成长需求，开发类型多样的场景化应用大模型，提升未成年人生成式人工智能技术应用的数字可行能力，助力未成年人健康全面发展。比如，针对乡村教育特点，开发适合欠发达地区未成年人教育的大模型产品，推动教育资源的共享，促进教育公平发展。又如，部分欠发达地区的未成年人由于其父母常年外出务工，缺乏必要的家庭情感陪伴，可以开发深度关爱模式的生成式人工智能产品，让生成式人工智能以师长或朋友的角色定位为未成年人提供适度的数字陪伴；在充分保障未成年人隐私权基础上，收集用户身心状态的基础数据，以便有关部门及时了解未成年人身心健康情况，及时提供关爱服务。

（四）素养先行：营造向善而生的教育应用环境

在未成年人教育中，教师与学生是两类最为重要的行动者主体，为确保生成式人工智能技术应用的可持续健康发展，应加强对师生群体的人工智能素养教育。具体而言，一是要加强人工智能素

^① 李韬、冯贺霞：《数字治理的多维视角、科学内涵与基本要素》，《南京大学学报（哲学·人文科学·社会科学）》2022年第1期。

养主题的师资培训，提高教师群体运用人工智能的技术能力和教育水平，为教师提高教育质量提供有力支撑；二是要在基础教育中引入面向未成年人的特色课程，围绕生成式人工智能的基本概念、原理和应用等进行技术普及教育，让学生适应当下数字化的课程体系；三是培养学生对技术存在正反双效性的批判性思维，维护学生在自主学习活动中的主体性地位。此外，还应开展生成式人工智能技术使用相关法律法规的科普，让未成年人了解技术滥用可能带来的道德风险与不良社会影响，学会为自己的行为负责。

总而言之，在技术的迭代更新与发展应用过程中，总是伴随着人类对未来的无尽想象与期盼。随着智能社会发展，未成年人教育中的生成式人工智能应用也将愈发广泛深入。只有通过多元主体协同共治的模式，明确生成式人工智能在开发与应用中的伦理红线，对面向未成年人的技术产品进行严格把关，才能坚守住技术服务人类的伦理底线。与此同时，也只有给予未成年人平等享受数字技术的权利，让未成年人在与技术的准社会交往中丰富对世界与自我的感知，掌握更多数字时代应当具备的能力与素养，才能够真正使得生成式人工智能成为未成年人的“良师益友”，让人类文明在“向善而生”的人机互动中不断传承、创新与发展。

Interactivity and Human–Machine Friendly Collaboration: Generative Artificial Intelligence for Adolescent Educational Applications and Governance

ZHOU Ruichun WANG Geer WU Shuyan

[Abstract] Artificial Intelligence-Generated Content (AIGC), with its unique algorithms and powerful data processing capabilities, can achieve multimodal content production and conversion between text, images, audio, and video in human–computer interactions, providing unprecedented possibilities for the application of adolescent education. This article posits that, in the role empowerment of human–computer interactions, AIGC can improve the efficiency and ability of autonomous learning for minors, provide possibilities for personalized education and educational equity, and enrich minors’ digital cultural life. At the same time, owing to the double-edged nature of AIGC technology itself, it also presents risks and hidden concerns such as uncertain educational quality, challenges in shaping values, and subjective misconceptions in human–computer interactions. To this end, governance strategies such as promoting diversified co-governance of family–school–community linkages, strengthening regulatory regulations under role positioning, promoting diverse and accessible services under digital equality, and improving the AI literacy of teachers and students have been proposed.

[Key words] Artificial Intelligence-Generated Content; Adolescent Education; Human–Machine Friendly Collaboration; AI Governance

（责任编辑：冯贺霞 责任校对：柴云超）