DOI:10.13811/j.cnki.eer.2024.07.002

基于“技术道德化”理论的生成式人工智能教育应用潜能与风险研究

童 慧[1]，杨彦军[2]

（1.江西科技师范大学 教育学部，江西 南昌 330038；
2.南昌大学 教育发展研究院，江西 南昌 330031）

[摘 要] 随着生成式人工智能的快速发展及其在教育领域的广泛应用，国内外研究者围绕它及其背后的大语言模型教育应用问题展开了热烈的讨论，但大部分研究因缺乏对生成式人工智能技术原理的了解而在问题探讨中存在伊莉莎效应。研究首先在对生成式人工智能技术学原理进行深入剖析的基础上，从维贝克“人—技杂合”视角对生成式人工智能的本质展开技术哲学分析；其次，基于“技术道德化”理论分析了生成式人工智能教育应用的四大潜能和面临的五大风险挑战；最后，提出基于“人—技杂合”的思想从“设计者”和“使用者”两种视角构建“双向奔赴”的全球人工智能治理体系，将成为未来探索的重要方向。

[关键词] 生成式人工智能；技术道德化；大模型；人—技杂合；应用风险

[中图分类号] G434 [文献标志码] A

[作者简介] 童慧（1987—），女，回族，甘肃平凉人。副教授，博士，主要从事教育技术学基本理论研究与计算机支持的协作学习研究。 E-mail：leiyuth@126.com。

一、引 言

随着生成式人工智能（Generative Artificial Intelligence，以下简称GAI）的快速发展及其在教育领域的广泛应用，国内外学者围绕GAI及其背后的大语言模型教育应用展开了热烈的讨论。“GAI对教育领域将会产生怎样的影响”“GAI时代人类教师将会遇到哪些挑战”等问题成为热点议题。技术乐观主义者认为，它将彻底变革人类学习方式和教育模式，因而成为GAI教育应用的积极推动者和探索实践者；技术悲观主义者认为，它的应用不但会带来抄袭、学术造假、教师失业等问题，还会导致学生因缺乏思考而丧失其主体性，因而常采用抵制或批判的态度；还有部分技术中性主义者认为，它的出现和以前教育技术发展史上的其他技术变革教育的预言一样，终将消失在教育技术发展的历史长河中。虽然技术乐观主义者和技术悲观主义者分析问题的视角和得出的观点截然相反，但深入分析后发现，大部分研究者的观点背后都充斥着对GAI“黑箱”的过度遐想，其中，有些研究者以自己想象的GAI为出发点讨论问题；技术中性论者看似客观冷静，但实则存在因循守旧之嫌，缺乏对GAI新产品及其影响的审慎思考和分析。GAI实际上能够做什么，它在教育中应用的潜能有哪些，它的教育应用存在哪些风险挑战，人类该如何面对GAI，要深入解答这些问题，不但要深入GAI“黑箱”分析其技术原理，还应当从人—技关系的底层逻辑展开哲学思考，深刻审视人、技术与伦理道德的关系。

二、生成式人工智能的技术学原理探析

当前对GAI应用潜能及其风险的讨论大多存在

基金项目：江西省社会科学“十四五”基金项目2024年度课题“基于‘技术道德化’视角的生成式人工智能应用潜能与风险研究”（项目编号：24JY17）；江西科技师范大学2022年度教学改革课题“混合式课程中过程性学习评价改革与实践”（课题编号：202/6000011842）

伊莉莎效应(Eliza Effect),即由于人们对机器缺乏足够的理性认识而产生的用“拟人化”思维去理解机器及其行为的现象，更倾向于下意识地将人类情感和意图赋予机器,认为机器和人类的行为机理与模式类似[1]。要深入思考和探讨 GAI 的应用潜能、风险挑战等问题，需要对其运行原理进行必要了解。本文所指的 GAI 是指能够根据用户的提示信息自动生成文本、图片、音频、视频等内容的计算机程序[2]。根据生成内容的不同,GAI 可分为文本生成、图片生成、音频生成、视频生成、代码生成及多模态生成等类型。不同类型的 GAI 在可接受的输入信息、可输出的内容类型、内部处理机制复杂程度及具体算法方面存在较大差异，但其基本工作原理具有相似之处。简单地讲,能够为公众提供服务的各类 GAI 都是建立在相应的大型或基础模型之上的计算机程序，即其底层支撑技术为大模型(Big Model) 或基础模型 (Foundation Models)。例如，ChatGPT 背后的 GPT 模型、文心一言背后的 ERNIE Bot 模型、通义千问背后的 Qwen 模型等。大模型是指具备大规模参数和复杂计算结构的机器学习模型,可以简单地理解为一套复杂的计算机程序，它由一系列实现复杂算法的函数及相应参数列表构成。所有的大模型都具备自我学习和生成输出两项基本功能，自我学习是完成大模型训练的过程，生成输出是完成大模型应用的过程。当前主流的大模型都是基于深度人工神经网络算法实现的,属于机器学习中的连接主义技术路线。人工神经网络就是模拟人类大脑中的神经元运作模式的计算机系统,其中的参数描述神经元之间连接的权重,模型训练的过程就是通过让程序学习大量训练数据集并掌握背后存在的关联规律,进而动态调整每个神经元连接之间的连接权重将关联规律表征为参数的过程，最终使得该计算机程序可以输出与训练数据相似的或模型开发人员期望得到的新数据的过程[3]。如图 1 所示,支撑基础大模型的底层人工神经网络由一个个的神经元构成，每个神经元之间的连接权重即一个参数。神经元可分为输入层、输出层和隐藏层,其中,隐藏层的数量决定了神经网络的深度,进而决定了神经网络的学习能力及模型的参数规模。由于模型训练过程需要花费大量的时间,经常采用预先训练的技术路线使其具备相关输出能力,因此,大模型也常称为预训练大模型。值得强调的是,模型预训练阶段使用何种数据集将直接决定模型输出内容的特征。例如,使用教育学相关数据集训练的模型将生成具备教育学学科特色的话语。预训练完成的基础大模型虽然具备了相应类型内容输出的基本功能，但输出的内容共性有余而特色不足，常常难以获得良好的实际应用体验。因此,需要根据实际应用场景(即下游任务)再对基础大模型进行微小调整(即微调)以形成具备特定功能的 GAI 产品。例如,训练一个教育技术学专业智能导师，通过大量通用文本数据集预训练的基础模型使其具备自动生成类人语言的基本功能——即“会说话”；如果想让其输出具备教育技术学专业知识、乃至某个导师个人话语风格的指导语言——即“说专业的话”，还需要对基础大模型进行微调以开发具备特色功能的 GAI 产品。常见的微调技术有基于人类反馈的强化学习、提示学习等,前者通过人类标注者对模型的输出质量进行排序来训练奖励模型以优化内容生成策略,后者通过让模型学习提问附加答案配对数据使得模型的生成内容更符合用户要求。无论采用何种策略,模型微调的过程就是能够让模型输出更符合人类偏好或特定价值取向内容的过程。模型训练完成后,用户使用 GAI 的过程实际上是通过“提问”或“指令”触发模型的输出功能,从而输出类人语言或其他期望得到的多模态信息的过程。同时,用户的使用行为及产生的过程性数据也已经被用来优化基础模型，即用户使用行为对基础模型具有反向塑造作用。

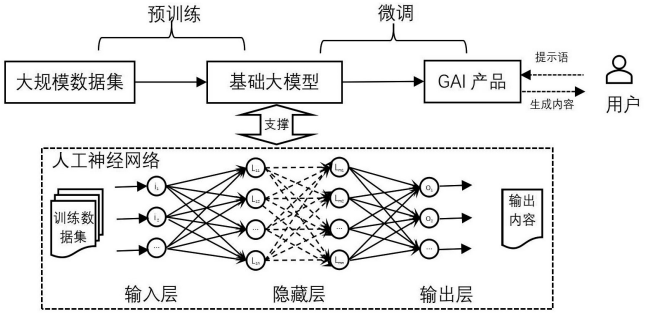

图 1　生成式人工智能的技术学原理

综上所述,GAI 的技术学本质就是一个复杂的计算机程序，核心是一个多层次的复杂人工神经网络。GAI 的预训练过程就是利用大规模数据集对神经网络进行“通识教育”的调整神经元连接参数的过程,微调过程就是利用特色数据集对神经网络进行“专业教育”的实现特色功能的过程,最终目的是让人工神经网络可以具备开发者期望的输出能力,这种输出能力的本质实际上是根据统计或关联规律连续输出基本信息表达单元的过程。例如,文字生成是按照得到的统计规律生成文字序列,图片生成是按照相关规律生成像素序列,视频生成是按照相关规律生成图像序列等。用户使用 GAI 的过程就是通过“提问”等互动方式触发神经网络输出的过程。可以看出,整个模型训练的过程就是对该人工神经网络进行“教育”的过程,训

练阶段使用的数据集、算法和微调策略、意识形态、价值倾向都将影响模型的最终输出“表现”。

三、基于“技术道德化”理论的生成式人工智能本质分析

(一)维贝克“技术道德化”理论要义

人类的进化与技术的发展密不可分,人本身就是一种技术性存在。古希腊时期的阿那克萨戈拉就认为,人在体力和敏捷性上比不上野兽,但人能使用自己的经验、记忆、智慧和技术[4]。法国哲学家贝尔纳·斯蒂格勒指出,人类自起源就是“缺陷存在”,因此,需要创造技术作为“代具”来弥补人类先天缺陷[5]。现象学技术哲学家唐·伊德(Don Ihde)认为技术并非自然存在,不能以本质主义的视角出发理解技术,而应当立足人类发展历史和人类劳动实践活动,从人与技术、自然的关系中把握技术,伊德基于“人—技互补”的视角将人—技关系归为具身、解释、他异和背景四种[6]。布鲁诺·拉图尔(Bruno Latour)则更加关注技术人工物对人行动的影响,基于“人—技对等”的视角认为,人与技术人工物都是行动网络中对称的行动者而相互交织、互塑共生,技术人工物通过“脚本”激励或抑制人的行为。荷兰技术哲学家彼得·保罗·维贝克(Peter-Paul Verbeek)则进一步将人—技关系的认识推向了新的高度,基于“人—技杂合”的视角提出了技术道德化理论,对当前深入理解人—技关系相关问题开辟了新的视野[7]。“技术道德化理论”由“技术调节”理论和“道德物化”理论两部分构成,前者侧重于对技术本质与功能的理解,后者侧重于对技术伦理问题的反思。维贝克认为,技术在人类世界不单单是中性的手段或工具,技术与人是相互塑造、纠缠的关系,技术具有物质意义上的“道德能动性”,从而对人的精神和行为具有不可忽视的塑造与影响的“调节”作用,并据此提出“技术调节”理论[7]。兰登·温纳(Langdon Winner)的技术政治理论认为,技术在前期设计和部署中潜藏着对政治秩序的影响,在技术的使用之前以一种物质的方式体现着人们的意图[8]。维贝克在此基础上发展出一种更加内在的“技术伴随”的伦理学,认为技术在设计时就已经嵌入了伦理考量,从而影响着人们对技术的使用和决策,这种人—技伴随的行动伦理学强调技术与人共同发挥伦理作用,技术在调节过程中具备了道德性,据此形成了“道德物化”理论[7]。维贝克在整合了技术哲学“经验转向”和“伦理转向”两次转向的基础上,将技术伦理的讨论放在人—技杂合体中进行,提出了“技术调节”理论和“道德物化”理论,实现了技术哲学的第三次转向——“道德转向”[9]。

(二)基于“人—技杂合”视角的 GAI 本质探析

从维贝克技术道德化理论“人—技杂合”视角来看,GAI 在开发阶段构成“开发者—GAI 杂合体”技术调节主体,GAI 是技术开发人员本质力量的外化,在这个“外化”的过程中既是一个开发者对其网络信息搜索与处理能力“编码”的过程,也是一个开发者意识形态与价值观“道德物化”的过程,开发者在 GAI 模型训练过程中用什么来源的数据、用什么类型的算法、用什么微调规则,均体现着开发者的价值取向与偏好。在使用阶段则构成“使用者—GAI 杂合体”的技术调节主体,在这个过程中首先是终端用户对 GAI 中内嵌的技术知识“解码”的过程,是对使用者原有信息获取与处理能力的增强;其次是 GAI 中物化了的技术开发者道德准则对用户使用行为的“激励”或“抑制”过程,潜移默化地影响着使用者的意识形态、价值判断和道德实践;同时,也是用户发挥自身主观能动性,对 GAI 进行创造性使用与反向塑造的过程,使用者自身的信仰观念、价值判断也最终决定着产品发挥的实际效能,如图 2 所示。值得说明的是,在“使用者—GAI 杂合体”的糅合关系形成过程中,GAI 用户自身的数字素养决定着“使用者—GAI 杂合体”整体功能的发挥。如果使用者具备良好的数字素养,能够较好地理解 GAI 底层运行逻辑并有效“解码”其潜在功能,则更有可能获得高质量的输出。因此,可以说 GAI 的技术哲学本质是一个模型开发者信息处理能力与价值观念“物化”的技术人工物,正如卡尔·马克思指出,技术的本质不过是人的本质力量的对象化。GAI 中隐含的开发者知识能力及价值观念,一方面,会“激励”或“抑制”用户特定行为的发生,也可能赋予用户前所未有的新能力;另一方面,GAI 生成的内容本身也会隐含模型开发者的意识形态和价值取向,进而对生成内容的消费群体产生显性或隐性的“教育”性影响。

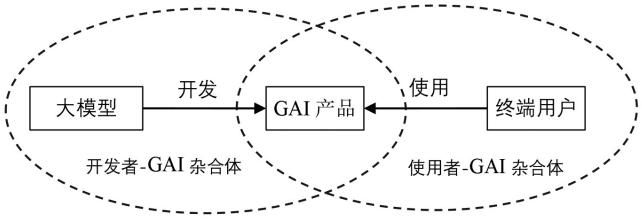

图 2 “人—技杂合”视角下的 GAI 技术哲学本质

四、基于“技术道德化”理论的生成式人工智能应用潜能与风险

“技术道德化”理论基于“人—技杂合”的视角认

识技术本质及其实践伦理问题,并从设计和使用两种语境提出 “道德物化”和“技术调节”理论。以下将分别从“技术调节”理论和“道德物化”理论分别阐释 GAI 教育应用潜能和教育应用风险。

(一)基于“技术调节”理论的 GAI 教育应用潜能分析

GAI 在教育应用过程中并不是一种中立的媒介,GAI 与使用者积极主动地共塑人在世界的方式,包括人的知觉和行动、体验和存在。一方面,它调节着人类对现实的经验和解释,另一方面,它又调节着人类的行为和生活方式。技术新产品的使用过程既是产品内涵知识及价值的“解码”和对用户赋能的过程,也是使用者发挥自身能动性建构 “人—技复合的道德实践者”创生伦理实践场景的过程。GAI 的教育教学应用潜能分析应基于使用者场景分析其对师生用户的赋能及用户创造性应用的可能性,具体如下:

1. 基于人机自然交互的教育信息服务获取

大数据时代的到来,使得“我们生活在信息的海洋中,却忍受着知识的饥渴”这一问题日趋严峻。20 世纪末,为应对网络信息激增而诞生的搜索引擎显然已经无法满足大数据时代人类对信息获取准确性、便捷性和实用性的要求。GAI 的出现让人和计算机程序可以用拟人化的“聊天”方式更加便捷地获取直接可用的多模态信息,人们无须在杂乱无章的搜索结果中二次遴选所需信息。Google 管理团队认为,GAI 的发展将催生新一代搜索引擎的出现[10]。为此,各大搜索引擎也迅速将 GAI 整合其中或探索布局基于大模型之上的对话式搜索引擎接口。GAI 的广泛应用降低了人机交互的技术门槛, 使得信息技术产品更加容易“上手”,“使用者—GAI 杂合体” 的形成更加容易,这是 GAI 对使用者信息获取能力增强的过程,极大地提升了人类获取教育信息与服务的效率。同时,GAI 与语音识别、机器翻译、跨模态理解、脑机接口、数字人等技术的深度融合,促使人机自然交互的质量大幅提升,人们获取高质量网络信息与服务将逐渐变为“零”技术门槛活动。

2. 基于多模态 GAI 的适应性数字教育资源开发

GAI 技术的不断成熟,各类 GAI 也逐渐实现跨模态内容理解与生成功能。例如,讯飞星火大模型、百度文星一格、阿里的通义千问、OpenAI 新发布的 Sora 等均已具备根据文本、图片、语音等输入生成文本、视频等多模态信息的功能。多模态 GAI 的发展,将极大提升数字教育资源开发的效率,使得适应性数字教育资源开发与应用成为可能。适应性数字教育资源是人工智能内容生成、自适应学习技术与数字教育资源个性化服务的有效整合[11],强调以师生真实需求和个性特征为数字教育资源服务的前提,重点提升数字教育资源生成与教学应用之间的适配性[12]。适应性数字教育资源开发时代教师无须为搜索、加工和整合各类数字资源而发愁,GAI 系统将会根据教师定制的提示工程,生成满足教学需要和富有教师个性风格的数字教育资源,甚至根据每个学生的个性特点生成个性化学习资源也将成为可能。基于多模态 GAI 的“使用者—GAI 杂合体”为教师赋予新的多模态教育资源自主开发能力,助力数字教育进入适应性教育资源开发与应用时代。

3. 基于人机共生互补的双师课堂教学创新

GAI 最受关注的应用场景是课堂教学应用,目前,国内外已有许多学者对此进行了较为深入的理论探讨和实证研究[13]。维贝克的“技术道德化”理论正是基于“人—技杂合”的视角认识技术的本质及其作用问题。人类教育技术持续发展的动力在于教师探索利用技术工具弥补自身不足,以提升教学效果的不懈追求。教师应用 GAI 的过程中形成“教师—GAI 杂合体”将更加便捷有效地弥补教师诸多缺陷, 发挥替代、协助、增强和赋能等功能[14]。具体包括:(1)GAI 替代功能, 即教师无须参与可全部交付给 GAI 处理的任务。(2)GAI 协助功能,即 GAI 协助教师更高效、优质地完成教师原本可以独立完成的工作,使教师专注于创造性更高、情感导向性更强的工作。(3) GAI 增强功能,即 GAI 支持教师更好地完成教师原本难以独立完成的任务。(4)GAI 赋能功能,即 GAI 支持教师完成自己原本无法完成的工作。

4. 基于人机对话博弈的学习教练与助理

GAI 作为模型开发者信息处理能力与价值观念“物化”了的技术人工物,它本身所具有的“拟人”功能赋予它成为用户学习教练或助理的潜能。恰当地应用 GAI 可在自主发展方面发挥如下功能:(1)自主学习的“GAI 学伴”, 学习者在语言学习中可以将 GAI 拟人化为语言“学伴”,在与 GAI 的连续性对话中发展语言能力; 在绘画技能学习或作曲练习中利用相关 GAI 辅助自己能力发展。(2)技能性任务的“GAI 助理”,学习者在完成程序编写、数据分析、学习资料查找的学习任务时利用 GAI 生成程序代码、分析报告或文献综述,GAI 扮演学习者助理的角色。(3)思维发展的“GAI 教练”,学习者在与 GAI 交互的过程中学会人机交互式 “提问(Prompt)”技巧,发展学生的精致化思维和计算思维;在与 GAI“挑战性”对话博弈中发展学生的批判性思维。

(4)数字孪生的“GAI 代理”,目前,大部分 GAI 都提供了相关的接口使得二次开发变得更加容易,每个专业级用户都可以对 GAI 进行模型微调训练和用户界面定制化开发,生成用户在数字孪生世界的“GAI 代理”。

(二)基于“道德物化”思想的 GAI 教育应用风险分析

维贝克从技术的设计语境和使用语境相关联的视角探讨了技术“道德物化”的路径。技术产品的开发过程是开发者将道德物化的过程,技术产品的使用过程是用户完成物化道德解码与发挥主观能动性的道德实践过程,是“产品道德+主体道德”的复合性道德实践活动。目前,已有大量有关 GAI 应用风险挑战的理论研究,本文从“技术设计语境”和“技术使用语境”相关联的动态视角将 GAI 教育应用的风险挑战总结为以下 5 个方面。

1. GAI 模型黑箱导致的意识形态风险

由前文 GAI 技术学原理的分析中可知,GAI 在底层模型采用的算法机制、预训练阶段使用的语料数据、微调阶段人类反馈强化学习的价值导向等问题共同决定了它本身就是一个“黑箱”。我国教育的根本任务是“立德树人”,主要使命是为党和国家培养一代又一代德智体美劳全面发展的社会主义建设者和接班人。GAI 作为对话式内容生成人工智能产品,其在教育教学应用中扮演着“准教师”的角色。学生利用 GAI 提哪些问题、生成哪些内容,既显性地反映了学生的认知需求与期望,也隐性地透露着学生的价值取向及意识形态;GAI 生成的内容也隐含着模型训练数据集、产品开发者及技术资本的价值取向。Rudin 等人认为,基于黑箱模型的科技公司可能会隐瞒 GAI 的工作原理,这导致很难对机器人所生成知识的来源及鸿沟进行揭示[15];GAI 能凭借内容生成的首要功能对意识形态的价值观念层进行渗透,依靠交互对话的产品形态对意识形态的话语方式产生影响,利用多维应用的功能属性对社会生活实践领域展开强势介入,导致当下意识形态斗争态势风云诡谲[16]。GAI 模型黑箱导致的意识形态风险具有辨识难度大、渗透性强、外泄风险高等特点,这是其教育应用的最大风险挑战。

2. GAI 教育应用成本导致的新型数字鸿沟

谈及 GAI 教育应用种种潜能与风险,其基本前提是师生能有机会便捷公平地使用 GAI 产品。信息基础设施建设情况、数字化终端拥有情况、数字素养与技能发展水平及网络信息服务支付能力等因素,均决定着每个终端用户的 GAI 使用机会。数字鸿沟的表现形态可分为 20 世纪的接入鸿沟、21 世纪初的素养鸿沟和当下智能鸿沟三个阶段[17]。GAI 教育应用导致的数字鸿沟据此可分为一级接入鸿沟、二级素养鸿沟和三级智能鸿沟。首先,网络设施普及情况、数字终端拥有情况、网络信息有偿服务支付能力以及区域网络管理政策共同造就了一级接入鸿沟的形成。其次,GAI 使用的过程是用户对 GAI 技术产品内涵价值的“解码”过程,师生的数字素养决定着解码过程中的价值损耗进而形成二级素养鸿沟,例如,数字素养高的用户更容易理解 GAI 技术黑箱的原理、更容易制定高质量的提示工程(Prompt)、更容易批判性地评价 GAI 生成内容的质量,进而导致使用效果的差异。最后,部分 GAI 产品凭借其行业先行效应在推出后短期内吸引了大批全球各地用户,随着企业技术的不断成熟及 GAI 对用户行为数据的持续积累,处于优势地位的产品将可能通过打造排他性数字生态体系逐渐形成行业垄断,进而形成以算法鸿沟、算力鸿沟和数据鸿沟为主要表现的智能鸿沟,进一步加剧社会不平等现象。

3. GAI 滥用导致的学术伦理风险

开展科学研究与学术写作是 GAI 的重要研究方向,也是 GAI 推出后研究者最先关注到的主要风险挑战领域。鉴于 GAI 在学生课程论文、考试测验及学术论文撰写中的应用可能造成的巨大冲击,许多研究者、高校、学术期刊等纷纷声明禁止或限制 GAI 在学术领域的应用。*Science* 期刊主编 Thorp 在 *Science* 发表社论:“ChatGPT 很好玩,但不是作者。”[18]*Nature* 杂志在投稿指南中指出,ChatGPT 等大语言模型不能成为论文作者[19]。虽然 GAI 在刚推出后受到了抵制和批判,但研究人员已经使用 GAI 来辅助文献检索、数据分析、论文撰写、论文翻译和修改润色[20]。Melnikov 等人认为,未来人工智能聊天机器人可能会生成假设、开发方法、创建实验[21]。同时,GAI 也暴露出其输出存在不准确、错误、偏见甚至种族歧视等问题。Van Dis 等在 *Nature* 上发文指出,科学研究正面临着一场由 GAI 引发的对其原来珍视的价值观、实践准则和标准的清算,解决这个问题的重点应该是抓住其所带来的机遇并管控可能的风险,我们相信科学界将找到一种既能从 GAI 中受益,又不至于使科学工作失去其作为一项使人深刻而令人满意的事业的许多重要方面:好奇心、想象力和发现探索[20]。

4. GAI 过度依赖导致的认知惰性

教育工作者最为担心的是 GAI 教育应用对学生认知能力的影响。首先,从知识学习的角度看,学生利用 GAI 生成答案,简化了知识获取的过程,降低了对

科学探索的内在兴趣,加大了学习惰性[22],成为“逃避学习”的工具[23]。其次,从认知能力的角度,学习者会因长期过度依赖 GAI 生成的内容而缺乏深入理解、深度思考和批判分析思维等高阶认识能力。上海大学转化医学研究院白龙等人研究认为,过度依赖 GAI 系统可能会导致批判性思维受损和记忆保持能力改变[24-25]。再次,从脑科学和认知心理学的角度来看,长期使用 GAI 可能会形成技术依赖,进而对学生神经系统产生实质影响,形成成瘾记忆神经通路[26]。加拿大研究者的一项研究已经表明, 长期使用 GPS 导航会对人的空间记忆能力及海马体结构产生负面影响[27]。最后,从人类主体性的视角看,GAI 可能会导致使用者思维的庸化和道德的钝化,逐渐封闭自我意识,放弃自由意志,造成人的主体性自我消解[28]。

5. GAI 数字污染导致的模型崩溃

由于 GAI 大模型自身的原理性局限叠加用户不清晰的使用命令,将导致大量不准确、歪曲事实甚至机器编造的虚假信息在网络空间迅速传播,重塑网络空间生态,进而导致的“信息扭曲”“数据污染”“数据贫困”“优质数据枯竭”等问题将日益严重。这些被污染的数据将不可避免地渗入下一代大模型训练的数据集,进而会导致新的大模型用自己产生的“毒数据”训练自己的“自激”现象,最终导致“模型崩溃”(Model Collapse)。英国和加拿大的研究人员的研究证实,使用大模型生成的数据对大模型进行持续训练会导致不可逆的缺陷, 使大模型原始数据分布的尾部消失,模型性能逐渐下降并最终导致“模型崩溃” [29]。算力、优质数据和大模型作为人工智能时代的基础资源和数字基础设施,不仅关系到国家数字经济生产效率的提升,更关系国家数字安全与长远发展。因此,在全力投入大模型训练与应用的同时,应更加重视数据资源的“环保”与治理。

五、技术道德化理论视域下 GAI 教育应用风险应对展望

GAI 的快速发展为教育应用带来巨大潜能,也给全球教育生态造成巨大冲击,GAI 应用规范与风险挑战全球治理体系亟须建立。为此,我国先后发布了《生成式人工智能服务管理暂行办法》《全球人工智能治理倡议》,乌镇世界互联网大会发布了《发展负责任的生成式人工智能研究报告及共识文件》, 首届全球人工智能安全峰会签署了旨在加强国际合作建立人工智能监管体系的“布莱切利宣言”,联合国教科文组织发布了专门针对教育领域的《教育与研究领域生成式人工智能指南》等共识性文件。这些初步探索为 GAI 教育应用风险治理带来重要指导,但现有治理视角更多的是基于道德伦理应用的外在式范式。维贝克解决技术伦理问题的内在式路径,不但敏锐地认识到了技术人工物对人类道德行为的调节作用,更重要的是认识到技术人工物的设计过程本身是负载价值的“道德物化”活动。据此,他提出有效沟通技术人工物设计情境与使用情境, 最大限度优化技术人工设计以对冲“设计者谬误”的依赖设计者的充分道德想象、拓展建构性技术评估和情境模拟法三大方法。沿循维贝克的技术伦理问题治理方法论,基于“人—技杂合”的思想从“设计者”和“使用者”两种视角构建“双向奔赴”的全球 GAI 治理体系, 将成为未来重要的探索方向,这方面的相关思考将另文详细阐述。

[参考文献]

[1] DILLON S. The Eliza effect and its dangers: from demystification to gender critique[J]. Journal for cultural research,2020,24(1):1-15.

[2] 生成式人工智能服务管理暂行办法 [EB/OL].(2023-07-10) [2023-12-16]. https://www.gov.cn/zhengce/ zhengceku/202307/content_6891752.htm.

[3] 翟尤,郭晓静,曾宣玮.AIGC 未来已来:迈向通用人工智能时代[M].北京:人民邮电出版社,2023:17.

[4] 北京大学哲学系.西方哲学原著选读(上卷)[M].北京:商务印书馆,1981:40.

[5] 张一兵.斯蒂格勒:西方技术哲学的评论——《技术与时间》解读[J].理论探讨,2017(4):57-63.

[6] 舒红跃. 现象学技术哲学及其发展趋势[J].自然辩证法研究,2008 (1):46-50.

[7] 彼得·保罗·维贝克.将技术道德化:理解与设计物的道德[M].闫宏秀,杨庆峰,译.上海:上海交通大学出版社,2016:3-15,51-82.

[8] 卫羚. 兰登·温纳技术哲学思想的解读——技术决定论与社会决定论的桥梁[J].福建论坛(人文社会科学版),2009(9):72-74.

[9] 史晨.技术哲学的第三次转向——维贝克道德物化思想的三重特征[J].科学技术哲学研究,2020,37(5):67-73.

[10] GRANT N, METZ C. A new Chat Bot is a ´Code Red´ for google´s search business[N]. International New York times, 2022-12-21(1).

[11] ROZO H, REAL M. Pedagogical guidelines for the creation of adaptive digital educational resources: a review of the literature[J]. Journal of technology and science education,2019,9(3):308-325.

[12] 罗江华,张玉柳.基于跨模态理解与重构的适应性数字教育资源:模型构建与实践框架[J].现代远程教育研究,2023,35(6):91-101.
[13] 张绒.生成式人工智能技术对教育领域的影响——关于 ChatGPT 的专访[J].电化教育研究,2023,44(2):5-14.
[14] 杨彦军,罗吴淑婷,童慧.基于"人性结构"理论的 AI 助教系统模型研究[J].电化教育研究, 2019,40(11):12-20.
[15] RUDIN C. Stop explaining black box machine learning models for high stakes decisions and use Interpretable models instead[J]. Nature machine intelligence, 2019,1(5):206-215.
[16] 代金平,覃杨杨. ChatGPT 等生成式人工智能的意识形态风险及其应对 [J]. 重庆大学学报(社会科学版),2023,29(5):101-110.
[17] 钟祥铭,方兴东.数字鸿沟演进历程与智能鸿沟的兴起——基于 50 年来互联网驱动人类社会信息传播机制变革与演进的视角[J].新闻记者,2022(8):34-46.
[18] THORP H H. ChatGPT is fun, but not an author[J]. Science, 2023,379(6630):313.
[19] Nature. Initial Submission [EB/OL]. (2023-01-24) [2023-12-27]. https://www.nature.com/nature/for-authors/ initial- submission.
[20] VAN DIS E A M, BOLLEN J, ZUIDEMA W, et al. ChatGPT: five priorities for research[J]. Nature, 2023,614(7947):224-226.
[21] MELNIKOV A A, NAUTRUP H P, KRENN M, et al. Active learning machine learns to create new quantum experiments[J]. Proceedings of the national academy of sciences, 2018,115(6):1221-1226.
[22] BAIDOO-ANU D, ANSAH L O. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning[J]. Journal of AI, 2023,7(1):52-62.
[23] 周洪宇,李宇阳.ChatGPT 对教育生态的冲击及应对策略[J].新疆师范大学学报(哲学社会科学版),2023,44(4):134-143.
[24] BAI L, LIU X, SU J . ChatGPT: the cognitive effects on learning and memory[J]. Brain-X,2023,1(3):1-9.
[25] DENG X, YU Z. A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education[J]. Sustainability,2023,15(4):2940-2959.
[26] 杨彦军,张佳慧,童慧.数字媒介技术依赖的多学科析因及整合性阐释[J].电化教育研究,2020,41(8):26-32.
[27] DAHMANI L, BOHBOT V D. Habitual use of GPS negatively impacts spatial memory during self-guided navigation [J]. Scientific reports, 2020,10(1):6310-6324.
[28] 雷磊.ChatGPT 对法律人主体性的挑战[J].法学,2023(9):3-15.
[29] SHUMAILOV I, SHUMAYLOV Z, ZHAO Y, et al. Model dementia: generated data makes models forget [J]. arXiv e-prints,2023(5):17493-17511.

Study on the Potential and Risk of Educational Application of Generative Artificial Intelligence Based on Moralizing Technology Theory

TONG Hui[1], YANG Yanjun[2]
(1.Department of Education, Jiangxi Science and Technology Normal University, Nanchang Jiangxi 330038;
2.Institute of Education Development, Nanchang University, Nanchang Jiangxi 330031)

[Abstract] With the rapid development of Generative Artificial Intelligence (GAI) and its wide application in the field of education, researchers at home and abroad have launched a heated discussion about it and the educational application of large language model behind it. However, the Eliza effect exists in most studies because of the lack of understanding of the principle of GAI technology. This study firstly analyzes the essence of GAI from the perspective of Verbeke´s "human-technology hybrid" on the basis of an in-depth analysis of the principle of GAI technology, and then elaborates four major potentials and five major risks and challenges faced by the application of GAI in education. Finally, based on the idea of "human-technology hybrid", a global AI governance system is proposed from the two perspectives of "designer" and "user" , which will become an important direction for future exploration.

[Keywords] Generative Artificial Intelligence; Moralizing Technology; Large Model; Human - Technology Hybrid; Application Risk