

生成式人工智能教育：关键争议、促进方法与未来议题

——UNESCO《生成式人工智能教育和研究应用指南》报告要点与思考

兰国帅 杜水莲 宋帆 肖琪 钟琦 齐春燕

【摘要】目前我国生成式人工智能教育应用研究理论层面存在顶层设计缺乏、监管框架缺位、政策框架缺失等问题,实践层面缺少相对科学完善的应用指南、监管认证工具、创新实施方法和行动措施,严重削弱了生成式人工智能技术全方位赋能教育数字化转型的能力,制约着生成式人工智能与教育和研究融合创新发展的进程。文章采用文本分析和内容分析法,介绍了联合国教科文组织《生成式人工智能教育和研究应用指南》的发布背景和研究目的,从人本主义批判视角探究生成式人工智能引发的争议和伦理风险,构建了生成式人工智能教育和研究应用的政策框架,提出了生成式人工智能教育和研究创新应用的促进方法和未来关注议题。文章最后结合我国生成式人工智能教育和研究政策规划制定进行了思考,以期为我国加速推进教育数字化,推进生成式人工智能与教育和研究深度融合提供参考。

【关键词】GPT; ChatGPT; 生成式人工智能; 教育生成式人工智能; 教育 GPT; 人工神经网络

【作者简介】兰国帅,博士后,河南大学教育学部副教授,研究方向:信息技术教育应用、教育数字化转型、教育人工智能、教师信息化能力建设等, E-mail: cqdxlgs@163.com; 杜水莲、宋帆、肖琪,河南大学教育学部硕士研究生,研究方向:信息技术教育应用(河南 开封 475004); 钟琦,硕士,赣南师范大学数学与计算机科学学院副教授,研究方向:信息技术教育应用(江西 赣州 341000); 齐春燕(通讯作者),博士,岭南师范学院数学与统计学院副教授,研究方向:信息技术与数学教育整合(广东 湛江 524048)。

【原文出处】《开放教育研究》(沪),2023.6.15~26

【基金项目】2021年度河南省哲学社会科学规划项目“乡村振兴战略下河南省乡村教师 ICT 能力提升路径与对策研究”(2021BJY007); 2023年河南省软科学研究计划项目“‘互联网+教育’推进河南县域义务教育优质均衡发展的机制与路径研究”(232400410019); 2024年度河南省高等学校重点科研项目资助计划“河南省高校教师数字化转型能力框架构建及提升路径研究”(24A880003); 2023年度河南大学研究生教育教学改革研究与实践项目“双一流建设背景下河南省专业学位研究生 TMACK 培养模式改革研究与创新实践”(YJSJG2023XJ061)。

一、问题提出

通用人工智能给教育教学带来机遇,也存在风险。例如,多数国家缺乏关于生成式人工智能教育的国家法规,生成式人工智能发展速度远超国家监管框架的制定速度,使得用户数据隐私缺少法律保护,而教育机构也基本没有准备好验证这些生成式人工智能教育工具。因此,学界亟须探明生成式人工智能会引发哪些争议和伦理风险? 这些争议和

伦理风险会给教育带来何种影响? 如何构建生成式人工智能教育和研究应用的政策框架与创新方法? 未来应重点关注哪些议题? 中国生成式人工智能教育应用研究目前仍处初级阶段,主要侧重赋能教育变革、发展机遇、伦理风险、规避路径等(张黎等,2023; 尚智丛,2023; 林国浒等,2023; 王立群等,2022)。理论层面存在顶层设计缺乏、监管框架缺位、政策框架缺失等问题,实践层面缺少科学完

善的应用指南、监管认证工具、创新实施方法和行动措施(卢宇等,2023),严重削弱了生成式人工智能技术全方位赋能教育数字化转型的能力。

在此关键节点,联合国教科文组织2023年9月发布的《生成式人工智能教育和研究应用指南》(简称《指南》)(UNESCO,2023),旨在支持各国实施即时行动、规划长期政策和发展人的人工智能,以确保生成式人工智能技术以人为本。《指南》可为我国制定生成式人工智能教育政策总体规划提供参考。本研究介绍了《指南》的发布背景和目的,从人本主义批判视角探究了生成式人工智能引发的关键争议和伦理风险,提出了生成式人工智能教育和研究应用的政策框架、创新方法和未来关注议题,探讨我国生成式人工智能教育和研究政策规划的创新路径。

二、争议与伦理风险

(一)数字贫困加剧

人工智能的发展依赖于大数据和计算能力,但这些创新大多只适用于少数国际科技公司巨头和少数经济体(主要是美国、中国和欧洲)。因此,大多数科技公司和国家(尤其是南半球的公司与国家)无法创造与控制生成式人工智能。随着获取数据对国家经济发展和个人数字发展变得日益重要,那些无法获取足够数据的国家和人民将陷入“数据贫困”。生成式人工智能在技术先进国家和地区的快速普及,加速了数据生成和处理,也加剧了人工智能财富向北半球集中,将导致数据贫困地区在生成式人工智能应用领域更加落后,并面临被GPT模型嵌入的标准“殖民化”的长期风险。目前的ChatGPT模型是根据在线用户数据训练而成的,这些数据反映了北半球的价值观和规范,并不适用于南半球数据匮乏地区或北半球弱势地区。

(二)超越国家法规的适应性

科技公司生成式人工智能的基础技术往往作为企业知识产权受到保护。尽管人工智能行业呼吁监管生成式人工智能,但关于生成式人工智能创建和使用的立法落后于其快速发展的步伐,这在一定程度上解释了国家或地方机构在理解和管理法律、道德问题方面面临的挑战。虽然生成式人工智能可以增强人类完成某些任务的能力,但国家对推广生成式人工智能公司的数据监管有限。这就提

出了监管问题,特别是在获取和使用国内数据方面(包括当地机构和个人的数据及在某一国家产生的数据)。因此,适当立法,以便地方政府机构能够监管生成式人工智能,确保对其作为公共产品开展治理,变得非常必要。

(三)未经同意使用内容

生成式人工智能模型是由大量数据构建而成的,这些数据通常来自互联网,没有获得所有者的许可。许多图像生成式人工智能系统和代码生成式人工智能系统因此被指控侵犯知识产权。此外,GPT可能违反欧盟的《通用数据保护条例》等法律,因为某一数据一旦被GPT模型纳入训练,就不可能从该模型中删除。

(四)无法解释的模型

人工神经网络通常是“黑匣子”,不是“透明”或“可解释”的,也不可能知晓其输出是如何确定的。虽然人工神经网络的总体方法(包括所使用的算法)通常是可解释的,但特定模型及其参数(包括模型的权重)是不可检查的,这就是人工神经网络生成的特定输出无法解释的原因。例如,ChatGPT-4模型有数十亿个参数或权重,由于参数或权重的不透明,人们无法解释这些模型的特定输出方式。随着生成式人工智能变得越来越复杂,其缺乏透明度和可解释性问题日益严重,经常产生意想不到或不符合用户预期的结果。另外,生成式人工智能模型继承并延续了训练数据存在的偏见。鉴于生成式人工智能模型的不透明性,这些偏见很难被发现和解决,这也是造成生成式人工智能信任问题的关键原因。如果用户不了解生成式人工智能系统是如何获得特定输出的,他们就不太可能愿意使用生成式人工智能。

(五)人工智能生成的内容污染互联网

GPT训练数据多来自互联网,经常包含歧视性内容和语言偏见,开发者不得不设置“护栏”,以防止GPT输出具有攻击性或不道德的内容,但由于缺乏严格的法规和有效的监督机制,生成式人工智能生成的有偏见的内容在互联网上传播,污染了大多数学习者的学习内容或知识来源,这对没有扎实先验知识的年轻学习者来说风险更高。这也给未来的GPT模型带来了递归风险,这些模型将根据GPT模型创建的从互联网上抓取的文本进行训练,其中

也将包含偏见和错误。

(六) 对现实世界缺乏理解

文本 GPT 有时被称为“随机鹦鹉”,它可以生成令人信服的文本,但文本通常包含错误或有害的陈述。这是因为它只重复数据训练中发现的语言模式,从“随机”模式开始,但不理解其含义,这就像鹦鹉可以模仿声音而不真正理解所说的内容。生成式人工智能模型“看起来”理解所使用和生成的文本,但“现实”是它们不理解语言和现实世界。这两者间的脱节可能给未来教育带来严重风险。事实上,生成式人工智能既不基于对现实世界的观察,也不基于科学方法,更不符合人类或社会价值观。它无法生成关于现实世界、人与社会关系、人与物体关系或人与技术关系的真正新颖的内容。生成式人工智能模型生成的看似新颖的内容是否可以被视为科学知识有争议。

(七) 减少意见的多样性

ChatGPT 往往只输出标准答案。如果某些单词频繁出现在训练数据中,ChatGPT 很可能会在其输出中重复它。这可能限制和破坏多元观点与多元表达思想的发展。数据贫乏人群(包括北半球的边缘化社区)的网上数据很少或有限。因此,考虑到基于互联网网页和社交媒体对话数据的预训练方法,GPT 模型可能会进一步边缘化弱势群体。

(八) 产生更深层次的深度造假

生成式人工智能还可以改变或操纵现有的图像或视频,生成难以与真实图像区分的假图像或假视频。生成式人工智能使得制造这些“深度造假”和“假新闻”变得越来越容易。换句话说,生成式人工智能容易做出不道德和犯罪行为,比如传播虚假信息、宣传仇恨言论,以及在人们不知情或未经同意的情况下,将他们的面部融入完全虚假甚至有损形象的电影中。

三、政策框架

(一) 促进包容、公平、语言多样性和文化多元化

生成式人工智能的整个生命周期都存在包容性问题。具体而言,如果生成式人工智能工具不能通过设计促进公平、语言多样性和文化多元化,它们将无法帮助解决教育的根本挑战或实现可持续发展目标 4,除非它具有包容性(无论性别、种族、特

殊教育需求、社会经济地位、地理位置、流离失所状况等)。为此,《指南》建议采取以下三项政策措施:

(1) 确定哪些是没有或无法负担互联网连接或数据的人,并采取行动促进连接和提升其数字能力,以减少人们获取人工智能应用(程序)所面临的障碍;建立可持续的资助机制,为残疾或有特殊需求的学习者开发和提供支持人工智能的工具;推广使用生成式人工智能支持所有年龄、地点和背景的终身学习者。(2) 制定生成式人工智能系统验证标准,确保数据或算法不存在性别偏见、对边缘群体的歧视或仇恨言论。(3) 制定和实施生成式人工智能系统的包容性规范,并实施制度措施,能保护语言和文化多样性;相关规范应要求生成式人工智能提供者在 GPT 模型训练中包括多种语言数据(特别是当地或土著语言数据),以提高生成式人工智能响应和生成多语言文本的能力;规范和制度措施应严格防止人工智能提供者有意或无意地删除少数民族语言或歧视土著语言使用者,并要求生成式人工智能提供者停止推广只涉及主流语言或文化规范的系统。

(二) 保护人类能动性

生成式人工智能日益复杂,关键的危险是它有可能削弱人类的能动性。越来越多的个人用户使用生成式人工智能支持写作或其他创造性活动,这会造成无意中的依赖,影响人类智力技能的发展。虽然生成式人工智能可以用来挑战和扩展人类思维,但它不应篡夺人类思维。设计和采用生成式人工智能时,保护和增强人类能动性始终是核心。《指南》建议:(1) 告知学习者生成式人工智能可能从他们那里收集的数据类型和数据使用方式,及其可能对教育和生活产生的影响。(2) 保护学习者作为个体成长和学习的内在动机。在使用日益复杂的生成式人工智能系统的背景下,加强人类研究、教学和学习方法的自主权。(3) 防止生成式人工智能剥夺学习者通过观察现实世界和实验、与他人讨论、独立逻辑推理等发展认知能力和社交技能的机会。(4) 确保充分的社交互动和适当接触人类产生的创造性产出,防止学习者沉迷于或依赖生成式人工智能。(5) 使用生成式人工智能工具最大限度地减少作业和考试的压力,而不是加剧压力。(6) 咨询研究人员、教师和学习者对生成式人工智能的看

法,并利用反馈决定是否以及如何根据机构规模部署特定的生成式人工智能工具,并鼓励学习者、教师和研究者批评和质疑人工智能系统背后的方法、输出内容的准确性以及它们可能强加的规范。(7)在作出高风险决策时,防止将人类应承担的责任转嫁给生成式人工智能系统。

(三) 监控和验证生成式人工智能教育系统

开发和部署生成式人工智能应符合设计伦理。一旦生成式人工智能投入使用,人类需要对其整个生命周期进行监控,包括考虑其伦理风险、教学的适当性和严谨性,以及对学习者、教师和课堂或学校关系的影响。《指南》建议采取以下五项行动:(1)建立验证机制,测试教育和研究使用的生成式人工智能系统有无偏见(特别是性别偏见),以及它们是否接受代表多样性(性别、身体状况、社会和经济地位、种族和文化背景以及地理位置)的数据训练;(2)解决知情同意问题,尤其是在儿童或其他弱势学习者无法真正知情同意的情况下;(3)审核生成式人工智能输出是否包括深度造假图像、虚假新闻或仇恨言论,如果发现生成式人工智能生成了不适当的内容,教育机构和教育者愿意并能够采取迅速而有力的行动加以缓解或消除;(4)教育或研究机构正式采用生成式人工智能应用程序前,应对其进行严格的伦理验证(即采用设计伦理的方法);(5)机构应确保生成式人工智能应用程序不会对学习者造成可预测的伤害,对目标学习者的能力具有教育效果,并符合教学原则(即符合相关知识领域和预期的学习成果、价值观的发展)。

(四) 培养学习者生成式人工智能相关技能

培养学习者人工智能能力是在教育及其他领域安全、合乎道德和有意义地使用人工智能的关键。然而,联合国教科文组织的调查数据显示,2022年初,只有约15个国家制定并实施或正在制定政府认可的学校人工智能课程。生成式人工智能最新发展表明,每个人都需要在人工智能技术方面达到适当的素养水平,了解生成式人工智能在广义上是如何工作以及可能产生的影响。为了做到这一点,《指南》提出了迫切需要采取的五项行动:(1)致力于为学校教育、职业技术教育与培训以及终身学习提供政府批准的人工智能课程,内容涵盖人工智能对人类生活的影响,以及正确和创造性地

使用人工智能工具(包括生成式人工智能应用程序)的技能;(2)支持高等教育和研究机构加强培养所需的人工智能人才计划;(3)在发展先进的人工智能能力方面促进性别平等,并建立性别平衡的专业人才库;(4)跨部门预测最新的生成式人工智能自动化导致的国家和全球就业变化,并根据需求的预期变化,在各级教育和终身学习系统提高学习者面向未来的技能;(5)为可能需要学习新技能和适应新环境的公民提供特别方案。

(五) 培养教师和研究者生成式人工智能能力

根据联合国教科文组织2023年关于政府将人工智能用于教育的调查数据显示,只有约7个国家(中国、芬兰、格鲁吉亚、卡塔尔、西班牙、泰国和土耳其)已经或正在制定针对教师的人工智能框架或培训计划。只有新加坡教育部声称,正在建立以ChatGPT教学应用为中心的在线知识库。这表明,大多数国家的教师没有机会获得在教育中使用人工智能的结构化培训,尤其是生成式人工智能方面的培训。因此,为了让教师做好责任和有效地使用生成式人工智能的准备,各国需采取以下行动:(1)根据当地情况制定或调整指导方针,帮助研究者和教师使用广泛可用的生成式人工智能工具,并指导新的特定领域人工智能应用程序的设计;(2)保护教师和研究者使用生成式人工智能的权利及其实践价值,分析教师在促进高阶思维、组织人际互动和培养人类价值观方面的独特作用;(3)定义教师所需的价值取向、知识和技能,以便有效、合乎道德地理解和使用生成式人工智能系统,使教师能够创建特定的基于生成式人工智能的工具,促进学生课堂学习及教师专业发展;(4)动态评估教师在教学和专业学习中理解和使用人工智能所需的能力,将新兴的人工智能价值观、人工智能理解和人工智能技能整合到在职及职前教师培训的能力框架与计划方案中。

(六) 提倡多元观点和思想的多元表达

生成式人工智能既不理解提示,也不理解响应。其响应是基于训练模型从互联网数据中发现的语言模式的概率。为解决其输出的一些基本问题,新方法包括将生成式人工智能与知识库和推理引擎连接起来。然而,由于工作方式、原始材料和开发者隐含的观点,从根本上讲,生成式人工智能

在其输出中再现了主流世界观,忽略了多元观点。因此,如果人类文明要蓬勃发展,人们必须认识到,生成式人工智能永远不可能成为权威的知识来源。由此,《指南》强调,用户需要批判性地看待生成式人工智能输出,尤其是:(1)理解生成式人工智能是一种快速但依赖的是不可靠的信息来源。虽然基于大型语言模型的工具和插件是为了支持访问且利用经过验证的最新信息而设计的,但目前还没有强有力的证据表明这些工具是有效的。(2)鼓励学习者和研究者批判性认识生成式人工智能提供的回答,认识到生成式人工智能通常只重复既定或标准的意见,忽略了多元意见以及思想的多元表达。(3)为学习者提供从试错、实证实验和对现实世界的观察中学习的机会。

(七)测试当地生成式人工智能应用程序模型并建立累积证据库

目前,生成式人工智能模型主要基于北半球的信息,南半球和土著社区的声音代表性不足。利用综合数据才能使生成式人工智能工具对当地社区,特别是对南半球社区背景和需求敏感。《指南》指出:(1)确保生成式人工智能的设计和采用是战略性规划的;(2)激励生成式人工智能设计者瞄准开放式、探索性和多样化的学习;(3)根据教育优先事项,扩大人工智能教育和研究应用的循证案例;(4)指导使用生成式人工智能来触发研究创新,包括利用计算能力、大规模数据和生成式人工智能输出,为改进研究方法提供启发;(5)评估将生成式人工智能纳入研究的社会和伦理影响;(6)建立基于实证教学研究和方法的具体标准,验证生成式人工智能是否有助于提供包容性学习机会、达成学习和研究目标以及促进语言和文化的多样性;(7)采取迭代方式,搜集生成式人工智能的社会和伦理影响的证据;(8)分析大规模利用人工智能技术的环境成本(如训练 GPT 模型所需的能源和资源),并制定人工智能提供商要实现的可持续目标。

(八)以跨部门和跨学科方式探究生成式人工智能的长期影响

跨部门和跨学科方法对在教育和研究中有效和合乎道德地使用生成式人工智能至关重要。采取跨部门和跨学科方式,才能及时发现并有效应对生成式人工智能的关键挑战,在持续利用和累积收

益的同时,最大限度地减少生成式人工智能的长期负面影响。这包括:(1)人工智能提供者、教育者、研究者、家长和学生代表合作,对课程框架和评估方法进行全系统调整,充分利用生成式人工智能在教育和研究方面的潜力并降低风险。(2)汇集跨部门和跨学科专业知识,即教育者、研究者、学习科学家、人工智能工程师和其他利益攸关方的代表共同研究生成式人工智能对学习和知识生产、研究和版权、课程和评估以及人类合作和社会动态的长期影响。(3)及时提供建议,为法规和政策的迭代更新提供信息。

四、促进方法

(一)促进生成式人工智能教育和研究创新应用的战略

教育和研究机构应制定、实施和验证适当的生成式人工智能战略和道德框架,以指导使用者责任和合乎道德地使用生成式人工智能系统,以满足教学、学习和研究的需求。《指南》建议采取以下策略:

(1)道德原则的制度实施:确保研究者、教师和学习者负责任且合乎道德地使用生成式人工智能工具,并严格评估生成式人工智能输出内容的准确性和有效性。(2)指导和培训:为研究者、教师和学习者提供有关生成式人工智能工具的指导与培训,确保他们理解数据标签和算法的偏见等伦理问题,遵守数据隐私和知识产权的法律法规。(3)发展生成式人工智能“提示工程”能力:除特定学科知识外,研究者和教师需要学习提示工程方面的专业知识,并能批判性地评估生成式人工智能生成的提示。鉴于生成式人工智能带来的挑战是复杂的,研究者和教师必须接受高质量的培训和支持。(4)在书面作业中检测基于生成式人工智能的抄袭:学生可能将不是自己写的文章冒充为自己的作品,这是一种新型“抄袭”。生成式人工智能提供商应在其输出上标注“由人工智能生成”的水印,开发识别人工智能生成文件的工具。当下可以通过人工严格检测维护学术诚信并加强问责制。长期战略应是各教育机构和教育者重新思考书面作业的设计,布置生成式人工智能和其他人工智能工具不能做但人类能做的任务,包括将同情心和创造力等人类价值观应用于复杂现实的挑战。

(二) 以人为本、适合教学的互动方法

研究者和教育者在决定是否以及如何使用生成式人工智能工具时,应优先考虑人类能动性,以及人类与人工智能工具间负责任、适合教学的互动。这包括:(1)使用生成式人工智能工具应有助于满足人类需求,并使学习或研究更有效;(2)教育者和学习者对生成式人工智能工具的使用应基于他们的内在动机;(3)使用生成式人工智能工具的过程应由人类教育者、学习者或研究者控制;(4)生成式人工智能工具及其生成内容的选择和组织应与学习者年龄、预期结果以及目标知识(如事实性知识、概念性知识、程序性知识或元认知性知识)或目标问题(如结构良好问题或结构不良问题)的类型相称;(5)使用过程应确保人类与生成式人工智能的有效互动以及促进人类高阶思维的发展,明确人工智能生成的内容、教学或研究策略的准确性及其对人类行为的影响。

(三) 共同设计生成式人工智能教育和研究应用

教育和研究中使用生成式人工智能不应使用自上而下的方式,也不应受商业驱动。相反,它的安全和有效使用应由教师、学习者和研究者共同设计。《指南》从生成式人工智能用于研究、促进教学、指导自学语言和艺术基本技能、促进探究学习和项目学习、支持有特殊需求的学习者五个方面,提出了由六个维度组成的设计框架(见下页表1):适当的知识或问题领域、预期结果、适当的生成式人工智能工具和比较优势、对用户的要求、人性化教学方法和实例提示及伦理风险。

五、未来议题

(一) 未知的伦理问题

日益复杂的生成式人工智能工具将引发更多的伦理问题:

(1)机会和公平:可能加剧现有技术和教育资源获取的差距和不平等;(2)人际关系:可能会减少人与人之间的互动以及学习中存在的社会情感;(3)人类智力发展:可能会通过提供预定的解决方案或缩小学习体验范围来限制学习者的自主性和能动性,这就需要调查生成式人工智能系统对年轻学习者智力发展的长期影响;(4)心理影响:模仿人类互动的生成式人工智能系统可能会对学习者产

生未知的心理影响,引发人们被操纵的可能性的担忧;(5)隐藏的偏见和歧视:随着更复杂的生成式人工智能系统被开发并应用于教育,它们可能会基于训练数据和方法产生新的偏见和歧视,从而导致未知的和潜在有害的输出。

(二) 版权及知识产权问题

生成式人工智能的出现正在迅速改变科学、艺术和文学作品的创作、传播和消费。未经版权持有人许可,擅自复制、传播或使用受版权保护的作品,侵犯了版权持有人的专有权利,可能导致相关违法后果。生成式人工智能模型的训练可能会被指控侵犯版权。最近的案例是,由人工智能生成的模仿加拿大歌手德雷克和威肯创作的新歌“心在我袖上”(Heart On My Sleeve)吸引了数百万听众,但由于版权纠纷而下线。虽然生成式人工智能监管框架要求提供商承认和保护模型使用内容所有者的知识产权,但生成作品的所有权和原创性正变得越来越具有挑战性。这种可追溯性的缺乏引发人们对保护创作者权利和确保其智力贡献得到公平补偿的担忧,教育领域需要负责任地使用生成式人工智能工具。

(三) 内容和学习来源问题

生成式人工智能工具正在改变教学内容的生成和提供方式。未来,通过人工智能对话产生的内容可能成为知识生产的主要来源之一。这可能会进一步削弱学习者对基于人类创建和验证的资源、教科书和课程教育内容的直接参与。生成式人工智能文本的权威外观可能误导没有足够先验知识的年轻学习者。学习者参与未经验证的内容是否应该被视为“学习”也是有争议的。由此产生的对二手信息的关注也可能减少学习者通过直接感知和体验真实世界、从试验和错误中学习、开展实验和发展常识等行之有效的方法构建知识的机会。

(四) 同质化反应与多样化和创造性输出问题

生成式人工智能缩小了多元叙事的范围,因为其输出倾向于代表和强化主流观点。由此产生的知识同质化限制了多元化和创造性思维。教师和学生越来越依赖生成式人工智能工具寻求建议,这可能导致回应的标准化和一致性,削弱了独立思考和自主探究的价值。书面和艺术作品中潜在的同质化表达会限制学习者的想象力、创造力和表达的替代视角。生成式人工智能供应商和教育者需要

表1 共同设计生成式人工智能教育和研究应用

应用类型	潜在但未经证实的用途	适用的知识或问题领域	预期结果	适用的生成式人工智能工具和比较优势	对用户的要求	需要人性化教学方法和实例提示	伦理风险
生成式人工智能用于研究	人工智能研究大纲顾问	结构良好的研究问题	提出和回答研究问题, 建议适当方法。潜在转变: 研究规划一对一教练。	评估生成式人工智能工具是否可在当地访问、开源、经过严格测试或权威机构验证; 进一步考虑任何特定生成式人工智能工具的优势和挑战, 并确保其正确满足特定的人类需求。	研究者须对主题有基本了解, 应该培养验证信息的能力, 尤其是能够检测不存在的研究论文、文本的引用和问题回答。	研究问题定义的基本思路(如目标受众、问题、背景), 以及方法、预期结果和格式。实例提示: 为[主题x]写下10个潜在研究问题, 并根据它们对[研究领域y]的重要性排序。	警惕生成式人工智能编造信息(例如, 不存在的研究出版物)以及用户复制和粘贴生成研究大纲的高风险, 这可能会减少初级研究者从试错中学习的机会。
	生成式数据探索者和文献审稿人	结构不良的研究问题	自动收集信息, 探索广泛数据, 提出文献综述草案, 并自动化解释部分数据。潜在转变: 用于数据探索和文献综述的人工智能培训师。		研究者须在数据分析的方法和技术方面有扎实知识。	问题的渐进定义、数据范围和文献来源、用于数据探索和文献综述的方法以及预期结果及其格式。	警惕人工智能生成虚假信息、数据处理不当、侵犯隐私、未经授权的分析、性别偏见、主导规范的传播及其对替代规范和多元意见的威胁。
生成式人工智能促进教学	课程共同设计者	教学主题的概念性知识和教学方法的程序性知识	协助课程设计过程, 确定课程结构; 通过提供问题和评估准则的例子帮助教师准备测试与考试。潜在转变: 人工智能生成的课程。		教师须理解并说明希望课程或测试涵盖和实现什么内容、解决程序性知识还是概念性知识, 以及应用何种教学理论。	向生成式人工智能提问有关主题事实知识结构和示例、教学方法和过程, 或基于主题和格式创建课程包或课程计划。人类课程设计师需验证事实知识, 并检查课程包的适当性。	生成式人工智能强加主导规范和教学方法的风险很高。警惕生成式人工智能延续有利于数据丰富群体的排他性做法, 加剧数据贫乏群体获得高质量教育机会的不平等。
	作为助教的生成式聊天机器人	结构良好问题中跨多个领域的概念知识	提供个性化支持, 回答问题并确定资源。潜在转变: 生成式助教双胞胎。		需要学习者有足够先验知识、能力和元认知技能来验证生成式人工智能的输出并注意错误信息。因此, 它可能更适合高等教育中的学习者。	教师能清楚地理解问题, 监控对话, 帮助学习者验证生成式人工智能提供的可疑答案。	警惕错误信息的风险和限制学习者获得人类指导和支持的机会, 阻碍牢固的师生关系发展, 这对儿童来说尤其令人担忧。
生成式人工智能作为一对一教练	一对一语言技能教练	语言学习(包括会话练习)	通过提供母语或外语的反馈、纠正和建模, 让学习者参与会话练习, 帮助其提高听、说、写技能。潜在转变: 一对一初学者语言教程。		学习者须达到与生成式人工智能系统独立对话的年龄, 具有与人工智能系统进行对话的内在动机, 能够对生成式人工智能的建议采取批判性方法, 并检查它们是否准确。	人类教师可以引导学习者使用生成式人工智能工具来请求反馈, 以改进、纠正发音或写作示例。实例提示: 让我用某语言进行对话, 帮助我不断进步。提出一些想法帮助我写某主题。	警惕对文化不敏感或语境不准确的语言, 以及刻板印象或文化偏见的无意延续。如果没有适当的策略模拟学习者的内在动机, 可能会限制孩子的创造力和独创性, 导致公式化写作。它也可能限制现实生活中的互动、多元观点、多元表达和批判性思维的机会。
	一对一艺术教练	音乐和绘画等艺术领域技术技能	提供个性化支持, 回答问题并确定资源。潜在转变: 一对一初级艺术教师。		学习者须有创作艺术或音乐的初始目标, 对艺术或音乐领域的关键要素有基本了解, 以及分析艺术作品或音乐作品的基本能力。	人类教师应要求学习者比较人工智能工具生成的艺术作品与自己的艺术作品, 鼓励学习者发展和应用其想象力与创造力。实例提示: 给我提一些建议启发我创作关于某主题/想法的图像。	可能会使儿童接触不当或冒犯性的内容, 从而可能侵犯儿童获得保护和幸福的权利。
	一对一编程或算术教练	初级概念编程知识和技能以及基础数学的学习	支持自主学习基本编码知识和技能, 发现学习者编码中的错误并提供即时反馈, 以及定制问题答案。潜在转变: 一对一初级编程教师。		学习者须有使用编码的内在动机, 以及使用编程语言的一些基本知识和技能。	教师应教授基本知识和技能, 激励学习者使用计算思维和编程解决问题。实例提示: 提出一些不寻常的编码想法。	警惕生成式人工智能的反馈和建议的准确性、阻止学习者发展计算思维技能和能力, 以发现和定义有意义的编码问题。

续表 1

应用类型	潜在但未经证实的用途	适用的知识或问题领域	预期结果	适用的生成式人工智能工具和比较优势	对用户的要求	需要人性化教学方法和实例提示	伦理风险
生成式人工智能促进探究学习或项目学习	苏格拉底的挑战者	结构不良的研究问题	让学习者参与对话,联想到苏格拉底对先验知识的质疑,从而发现或加深对新知识的理解。潜在转变:一对一苏格拉底对手。	评估生成式人工智能工具是否可在当地访问、开源、经过严格测试或权威机构验证;进一步考虑任何特定生成式人工智能工具的优势和挑战,并确保其正确满足特定的人类需求。	学习者须达到与生成式人工智能工具独立对话的年龄、具备先验知识和能力,检查呈现的论点和信息是否准确。	教师可以准备一份逐渐深入的问题清单,作为学习者适应提示的例子。学习者可以从宽泛的提示开始。实例提示:“让我参与苏格拉底式对话,帮助我从批判性角度看待某主题”,然后通过越来越精细的提示逐渐加深对话。	生成式人工智能工具可能会产生类似或标准的答案,限制学习者接触不同观点和替代观点,导致“回音室效应”,阻碍学习者独立思考。
	项目学习顾问	科学或社会研究中结构不良的研究问题	通过帮助学习者开展项目学习支持知识创造(扮演人工智能研究大纲顾问)。潜在转变:一对一项目学习教练。		学习者可作为初级研究者,规划和实施项目学习,须达到能够独立使用生成式人工智能平台的年龄、有动机和能力参与自主的项目式学习。	教师引导学习者要求生成式人工智能为教育系统的监控和验证问题提出基本的研究思路,使用生成式人工智能工具开展文献综述,收集和处理数据,并创建报告。	没有扎实先验知识和验证答案准确性所需能力的学习者,可能会被生成式人工智能工具提供的信息误导。它还可能限制学习者与同伴的讨论和互动,减少合作学习机会,从而损害其社会发展。
生成式人工智能支持有特殊需求学习者	学习困难的会话诊断	因心理、社会或情感问题而面临学习困难的学习者	利用自然语言参与确定有心理、社会或情感问题或学习困难的学习者的需求,以便为其提供相关支持或指导。潜在转变:为有心理、社会或情感问题或学习困难的学习者提供一对一的初级顾问。		教师或专家需确保生成式人工智能提出的建议是准确的。	教师或辅导员需提供舒适环境,让学习者参与对话,以诊断心理、社会或情感问题或学习困难。	可能会在不经意间误诊学习者面临的挑战,导致提供错误支持。
	人工智能辅助工具	有听力或视觉障碍的学习者	通过为音频或视频内容提供支持生成式人工智能的字幕或手语翻译,以及为文本或其他视觉材料提供音频描述,满足学习者的获取需求,并支持其获取特定学科知识。潜在转变:一对一个性化人工智能语言辅助工具。		教育者须帮助学习者访问并学习如何操作生成式人工智能工具,确保生成式人工智能工具的输出真正支持学习者,而不会加剧他们面临的挑战和偏见。	需测试生成式人工智能平台或工具的可访问性,使用前识别并解决可访问性问题。生成式人工智能工具只能提供对内容的访问,教育者和辅导员应专注提高学习者的学习质量和社会福祉。教育者和辅导员需教会学习者根据自己的能力创建语音或文本提示。	生成式人工智能平台产生的字幕或音频描述不是专门为支持视觉或听觉而设计的,通常是不准确的,可能会误导有特殊需求的学习者,无意中强化已有的偏见。
	面向边缘化学习者的生成式放大器	帮助少数民族的边缘化学习者发出和扩大他们的声音、在线参与和开展合作社会研究	提供实时翻译、改写和自动纠错,支持边缘群体学习者使用自己的语言与来自不同语言背景的同龄人交流。潜在转变:面向边缘化学习者的包容性大语言模型。		学习者应对对话或合作学习的主题有独立见解,能够作出负责任和非歧视性的贡献,避免仇恨言论。	教育者应为学习者设计有关社会或文化主题的研究和写作任务,或组织在线研讨会或跨文化合作,以激发学习者产生想法和分享观点。	需识别和纠正人工智能翻译和释义中可能导致跨文化误解的错误。这种使用可为边缘化学习者提供放大其声音的机会,但不会触及数据贫困的根本原因,因此无法使人工智能工具去殖民化。

考虑教育 GPT 在多大程度上可以被开发和用于培养创造力、协作能力、批判性思维和其他高阶思维技能。

(五) 反思评估和学习成果问题

生成式人工智能对评估的影响远远超出了人

们对学习者在书面作业中作弊的担忧。它不仅可以生成相对完善的论文、散文以及令人印象深刻的艺术作品,还可以通过某些学科领域的考试。因此,我们需要重新思考到底应该学习什么,学习的

目的是什么,以及如何评估和验证学习,教育者、政策制定者、学习者和其他利益相关者的批判性讨论需要考虑以下学习成果:

(1) 价值观:确保以人为本的设计和和技术使用所需的价值观是数字时代重新思考学习成果及其评估的核心。教育目的确定需明确技术与教育之间关系的价值。正是通过这种规范视角,学习成果及其评估和验证需要不断更新,以应对包括人工智能在内的技术在社会中日益普及的使用。(2) 基础知识和技能:即使在生成式人工智能工具比人类做得更好的能力领域,学习者仍然需要良好的基础知识和技能。基础识字、算术和基本科学素养仍将是未来教育的关键。教育工作者需定期重审这些基本技能的范围和性质,以反映日益丰富的人工智能环境。(3) 高阶思维技能:学习成果需要包括支持人类与人工智能协作和使用人工智能生成输出的高阶思维与问题解决所需的技能。这些技能可能包括理解事实和概念知识在高阶思维中的作用,以及对人工智能生成内容的批判性评估。(4) 与人工智能合作所需的职业技能:在人工智能可以比人类做得更好并且日益自动化的领域,人类学习者需要培养能开发、操作和使用生成式人工智能工具的新技能。总之,要重新设计学习成果和教育评估以反映与人工智能有关的职业技能。

(六) 思维过程问题

生成式人工智能对教育和研究影响的长期议题是人类能动性和机器间的互补关系,即人类是否有可能将基本思维和技能交给人工智能,进而专注于基于人工智能输出高阶思维技能。例如,写作通常与思维结构联系在一起,有了生成式人工智能,人类可以从生成式人工智能提供的结构良好的大纲开始,而不是从头开始规划目标、范围和大纲。一些专家将使用生成式人工智能生成的文本描述为“无需思考的写作”。随着新的生成式人工智能辅助工具被广泛采用,现有获取和评估写作技能的方法有待更新。未来,写作学习可能侧重于培养计划和编写提示的技能、对生成式人工智能输出的批判性评估、高阶思维,以及基于生成式人工智能大纲的共同写作。

六、思考和借鉴

(一) 构建生成式人工智能教育和研究的政策框架,动态更新相关法律政策体系与措施

生成式人工智能教育应用将直接影响教育理

念和人才培养模式的重构,也可能引发安全危机和隐私保护等问题。面对新技术带来的冲击,我国需要借鉴国际经验,结合教育实际,构建生成式人工智能教育和研究的政策框架,动态更新相关法律政策体系。

首先,明确法律定位和责任认定。生成式人工智能的复杂性、特殊性和多模态性等特点使国内外对其定位颇具争议,我们应以法律的形式具体分析生成式人工智能的定位(姚志伟等,2023)。另外,责任主体划分不明确造成生成式人工智能的法律监管存在滞后性,难以划分算法设计者、数据集权利者和用户与最终生成内容的关系(朱嘉珺,2023),这就要求在法律制定中,不能采用一刀切模式,要遵从多元化原则。其次,法律制定秉承风险预防原则。生成式人工智能更新迭代迅速,现有的法律法规可能存在滞后性,这就需要采取风险预防原则,提高对技术风险的识别和定位能力,评估可能产生的风险,进而采取有效措施加以预防。最后,法律制定要软硬兼施,兼顾产业健康发展和打击违法犯罪的双重目标。一方面,平衡与鼓励生成式人工智能产业健康发展;另一方面,需要完善现有法律法规,对违法行为加大惩罚力度(姚志伟等,2023)。

(二) 完善生成式人工智能教育和研究的监管框架,促进生成式人工智能的迭代与升级

生成式人工智能赋能教学,可使教师从低效、复杂、烦琐的知识传授中解放出来,回归育人主业;生成式人工智能赋能学习有助于推动学习空间泛在化,满足学生个性化需求。但新技术的引入总会伴随风险,这就需要从多角度完善生成式人工智能教育和研究的监管框架。

首先,建立生成式人工智能技术法律框架。我国已先后颁布《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等,但这些特定措施缺乏连续性和稳定性,需要“建立稳定、持久的生成式人工智能技术法律框架”(张璐,2023)。其次,设立独立监管机构。为有效监管生成式人工智能教育应用,我国应设立独立监管机构,明确其职责范围和任务,包括审查、评估和监督的合规性和质量,处理投诉和违规行为等。为保证机构的权威性和

专业性,监管机构应独立于政府和商业利益,人员可由专业人士、学者、教育从业者、家长等多方代表组成。最后,对掌握前沿技术的大型科技公司进行严格监管(黄宁燕,2023)。这包括建立法律法规,明确大型科技公司在生成式人工智能应用中的权利和义务,规范其行为,防止其滥用权力和侵犯用户权益;加强信息公开,公开算法和数据处理方式,增加透明度和信任度。

(三)构建生成式人工智能教育和研究伦理规避机制,扩大人工智能教育应用可持续性

生成式人工智能技术的迭代发展给教育和研究者带来便利和机遇的同时,也会引发教育公平、人性化互动缺乏、算法歧视、隐私泄露等伦理问题。因此,构建生成式人工智能教育和研究的伦理规避机制迫在眉睫。

首先,政府应该发挥主导作用,面向全体师生提供人工智能公共基础服务,确保技术人人可及,尽可能减少不公平现象(杨宗凯等,2023)。其次,为了规避生成式人工智能教育应用可能会对学习者的人际关系、智力发展和心理健康产生的影响,要坚持生成式人工智能教育应用中人的主体性和以人为本的思想观念,更加重视对教育主体数字素养的培养(张黎等,2023)。最后,针对生成式人工智能存在的算法歧视与偏见,研究者和社会大众应给予正确引导,引领生成式人工智能技术人员审慎地训练模型;学校教育工作者需商讨如何恰当使用生成式人工智能,使学生避免接触偏见性知识,并且在使用前反复测验,确保课堂教学健康持续;教师应提醒学生脱离惯性思维与认知,敢于提出疑问,与社会大众分享观点,提升生成式人工智能教育应用的可持续性(王佑镁等,2023)。

(四)开展生成式人工智能教育和研究的实证研究,助力人工智能教育研究体系的构建

促进生成式人工智能技术发展和规范应用的关键在于积极开展实证研究,深入了解并解决生成式人工智能教育应用的潜在问题。

首先,加快学科建设,推动智能教育发展。这包括深入研究教学、学习、教研等场景的生成式人工智能技术应用边界,围绕教育减负、能力提升、素养培育、个性发展等需求,出台面向不同场景和对象的生成式人工智能应用指南,打造智能教育场

景应用示范,并强化生成式人工智能创作内容的质量审核机制(杨宗凯,2022)。其次,积极开展实证研究,评估生成式人工智能的教育应用效果。高校相关研究项目可引入龙头企业共建共享,实现多元合作和产教映射,将本地教学数据与合作企业的云端数据整合成网络安全人才培育大数据子集,根据实际需求,持续动态调整机制和更新数据(王磊等,2023)。最后,加强跨学科合作,促进智能教育革新。这包括建立跨学科的合作机制,探索智能教育发展、学科建设和人才培养规律,推动信息科学、认知科学、行为科学、管理科学与教育的交叉融合,研究算法黑箱、算法可解释性、深层次学习分析等问题,认识智能教育的学习机理(刘三女牙,2022)。

(五)构建教师生成式人工智能教育研究培训机制,赋能教师智能教育素养与技能提升

目前仍有教师缺乏正确的观念,制约技术效能的持续发挥(赵磊磊等,2023)。为更好发挥生成式人工智能的潜能,我国要重视教师培训,建设更灵活的教育体系。

首先,教师要提升利用智能技术协同育人的能力(张惠,2023)。为了有效融合生成式人工智能与教育生态,教师需要提升智能教育素养,认识到生成式人工智能的潜在价值。在制定培训机制的过程中,教师应与专家协作探讨如何有效应用人工智能工具,以有助于教师改善教学。其次,教师要提升人文关怀,将“人”贯穿于教育活动的全过程,打破生成式人工智能的知识传递垄断,将人文、情感和道德教育融入课堂,培养学生人文关怀的能力。相关的培训机制包括建立教师数据素养共同体,提升教师数字技能,同时警惕工具滥用,以免侵犯教师和学生的主体性(宋晔等,2023)。最后,教师要借助智能技术创新教育模式,探究人工智能应用新格局。生成式人工智能可用于优化教师培训,为培训课程提供针对性和实用性的学习材料,助力教师熟练掌握新技术,提高其信息化教学能力。此外,我国还要建立教师考核机制,鼓励教师运用人工智能技术开展信息化教学(张一春等,2023)。

(六)构建学生生成式人工智能教育研究的培训机制,推动学生主体意识的个性化培养

人工智能赋能教育所带来的历史性变革没有

改变教育以人为本的思想(曹培杰,2018),然而,学生对智能技术的甄别和驾驭能力普遍较弱(赵磊磊等,2023)。因此,我国有必要构建相关培训机制,个性化培养学生的主体意识。

首先,在技能培训方面,教育管理部门要鼓励相关科研部门或高校研发人工智能新技术,大力开发以学生为本的人工智能教育培训体系(贾积有等,2023)。其次,在学生学业述评方面,教育机构应结合新一代智能技术,深度分析数据中包含的有关学生认知、学习能力与方式、对教师的喜爱程度等信息,精准掌握学生实际学业水平(周东波等,2021)。当前已被广泛推广并应用的人脸识别、图像分类、语音对话等信息采集和分类技术,可用于拓展信息来源和提高信息准确度,将教育评价系统全面地移植到学情分析中,实现实时和针对性反馈。最后,在智能化学习服务方面,教师要遵循学生为主、智能为辅的原则,与生成式人工智能教育和谐共处(何文涛等,2023),即利用生成式人工智能人机交互的特点,丰富学习者的知识获取途径与应用场景,助力学习者随时随地体验高质量的知识与信息输入(郭炯等,2019),为学生创建安全的学习环境。

注释:

①本文部分内容来自 UNESCO(2023). Guidance for generative AI in education and research[EB/OL]. [2023-10-03]. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>。

参考文献:

- [1]曹培杰(2018).智慧教育:人工智能时代的教育变革[J].教育研究,39(8):121-128.
- [2]郭炯,郝建江(2019).人工智能环境下的学习发生机制[J].现代远程教育研究,(5):32-38.
- [3]贾积有,张誉月(2023).人工智能与教育:机遇、挑战与对策[J].北京大学教育评论,(1):49-61,188-189.
- [4]何文涛,路璐,周跃良等(2023).智能时代人机协同学习的本质特征与一般过程[J].中国远程教育,(3):12-20.
- [5]刘三女牙(2022).人工智能+教育的融合发展之路[J].国家教育行政学院学报,(10):7-10.
- [6]黄宁燕(2023).“ChatGPT之父”提出未来人工智能监管框架[J].科技中国,(8):97-100.
- [7]卢宇,余京蕾,陈鹏鹤等(2023).生成式人工智能的教

育应用与展望——以 ChatGPT 系统为例[J].中国远程教育,43(4):24-31,51.

[8]林国浒,李正福(2023).ChatGPT背景下日本中小学校生成式人工智能技术应用指南的分析与思考[J].中国现代教育装备,(16):1-4,12.

[9]尚智丛,闫禹宏(2023).ChatGPT教育应用及其带来的变革与伦理挑战[J].东北师大学报(哲学社会科学版),(5):44-54.

[10]宋晖,刘博文(2023).生成式人工智能时代教师身份境遇与角色塑建[J].中小学德育,(9):8-13.

[11]UNESCO(2023).Guidance for generative AI in education and research[EB/OL]. [2023-10-03]. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>.

[12]杨宗凯,王俊,吴砥等(2023).ChatGPT/生成式人工智能对教育的影响探析及应对策略[J].华东师范大学学报(教育科学版),41(7):26-35.

[13]杨宗凯(2022).元宇宙推动教育的全面数字化转型[J].教育研究,43(12):23-26.

[14]姚志伟,李卓霖(2023).生成式人工智能内容风险的法律规制[J].西安交通大学学报(社会科学版),(5):147-160.

[15]王立群,杨芸伊(2022).“人工智能+思想政治教育”:生成、风险及应对[J].湖南社会科学,(4):166-172.

[16]王佑镁,王旦,梁炜怡等(2023).ChatGPT教育应用的伦理风险与规避进路[J].开放教育研究,29(2):26-35.

[17]王磊,徐子竞,朱戈等(2023).生成式人工智能赋能网络安全人才培养的探索研究[J].中国电化教育,(9):101-108,116.

[18]赵磊磊,闫志明(2023).生成式人工智能教育应用的生态伦理与风险纾解[J].贵州师范大学学报(社会科学版),(5):151,160.

[19]张惠(2023).固守或超越:人工智能时代高校教师的角色重塑[J].黑龙江高教研究,(9):91-97.

[20]张璐(2023).通用人工智能风险治理与监管初探——ChatGPT 引发的问题与挑战[J].电子政务,(9):14-24.

[21]张一春,汤玲,马春兰(2023).人工智能助推教师发展的路径与对策研究[J].电化教育研究,44(10):104-111.

[22]张黎,周霖,赵磊磊(2023).生成式人工智能教育应用风险及其规避——基于教育主体性视角[J].开放教育研究,(5):47-53.

[23]周东波,徐雯慧,喻宏伟等(2021).大学生日常行为习惯的可发现、可计算与可干预研究框架[J].现代远程教育, (6):77-84.

[24]朱嘉珺(2023).生成式人工智能虚假有害信息规制的挑战与应对——以 ChatGPT 的应用为引[J].比较法研究,(3):34-54.