

生成式人工智能赋能高等教育形成性评价的价值、挑战及路径

卢国庆, 杨沁, 贺相春

(西北师范大学 教育技术学院, 甘肃 兰州 730030)

[摘要] 生成式人工智能已经具备与人类进行有意义多轮对话的重要能力。传统高等教育形成性评价在及时性、个性化等方面存在不足。如何利用生成式人工智能优化高等教育评价过程, 已经成为教育界共同关注的话题。研究旨在通过生成式人工智能, 提供智能反馈, 优化形成性评价, 提高评价的效率与个性化水平, 进而促进学生有效学习的发生。研究利用文献法与归纳法, 聚焦形成性评价, 在探讨生成式人工智能赋能高等教育形成性评价的基础上, 从形成性评价的理念、对象、过程、结果和伦理等方面, 探讨生成式人工智能赋能形成性评价的挑战, 并提出了生成式人工智能赋能形成性评价的路径以期生成式人工智能更好地赋能高等教育形成性评价提供借鉴和参考。

[关键词] 生成式人工智能; 智能反馈; 教育评价; 形成性评价; 技术赋能

[中图分类号] G434 **[文献标志码]** A

[作者简介] 卢国庆(1988—), 男, 河南周口人。讲师, 博士, 主要从事学习分析研究。E-mail: luguqing@nwnu.edu.cn。贺相春为通信作者, E-mail: hxc@nwnu.edu.cn。

一、引言

教育评价是高等教育中的重要环节, 关乎国家人才培养的方向与质量。2020年, 中共中央 国务院印发《深化新时代教育评价改革总体方案》提出“创新评价工具, 利用人工智能、大数据等现代信息技术, 探索开展学生各年级学习情况全过程纵向评价”^[1]。作为教育评价的重要形式, 形成性评价与反馈有助于提升学习绩效, 在高等教育中的作用越来越重要^[2]。形成性评价旨在帮助学生成为自主学习者, 是师生在教与学过程中持续使用的、有计划性的过程, 以增加学生对特定学科学习成果的理解^[3]。目前, 有关人工智能在教育评价中的运用多以总结性评价为主, 而人工智能在形成性评价中的应用潜力有待进一步挖掘^[4]。

生成式人工智能 (Generative Artificial Intelligence, 简称 GAI) 是一种能生成新内容的新型人

工智能。美国人工智能研究实验室 OpenAI 于 2022 年 11 月发布的 GAI 产品 ChatGPT, 引发了全球教育界的关注。GAI 是由大语言模型驱动的可以生成文本、代码等技术, 其底层基于 Transformer 模型, 可以根据用户输入的提示词, 生成相应的内容, 并进行多轮智能对话, 它不仅生成多样化的内容资源, 而且能够实现智能问答, 为学生提供个性化、及时的智能反馈。国内类似的 GAI 工具如文心一言、讯飞星火等纷纷涌现, 逐渐改变了人类的学习和思维方式。GAI 具备为常见教育任务提供高质量反馈的能力, 引发了全球教育工作者思考应该教会学生什么及如何开展评价的基本问题。

GAI 赋能教育评价已经得到了研究者的关注^[5], 目前, GAI 在学习与评价中的应用还存在着诸多争议。第一, 赋能还是限制? 有学者认为, GAI 让老师有更多的时间专注于教学, 是学习与评价的未来^[6]。然

基金项目: 2024 年度中央引导地方科技发展资金项目“教师专业能力智能测评与人机协同教学关键技术研究及系统研发”(项目编号: 24ZYQA038); 2024 年度国家自然科学基金资助项目“基于多模态数据的混合学习情境投入分析模型与干预研究”(项目编号: 62467007)

而,大模型算法存在偏见,学生在面对 GAI 流畅的表达时缺乏批判性思考能力^[7]。因此,应当禁止或限制高等教育中使用 GAI。第二,剽窃还是合作? GAI 在形成性评价中的运用可能使得人机合作与学业作弊之间的界限变得愈发模糊和迷离。第三,以机器还是以人为中心?以机器为中心的导向会让学生对机器产生高度依赖,这可能陷入认知外包的陷阱。如何构建以人为中心的形成性评价路径是亟须探讨的重要方向。本研究围绕形成性评价这一主题,在分析 GAI 赋能形成性评价的基础上,从形成性评价的理念、对象、过程、结果和伦理等方面,探索 GAI 对形成性评价的潜在挑战,并提出 GAI 赋能形成性评价的路径。

二、形成性评价与智能反馈的内涵

(一)形成性评价:促进学生学习

形成性评价起源于 20 世纪 60 年代,最早由布鲁姆将形成性评价概念引入到教学领域^[8]。布鲁姆认为,形成性评价是为教与学提供反馈与纠正的评价。理解形成性评价的概念,需要辨析形成性评价与总结性评价的区别。大多数学者认为,形成性评价和总结性评价分别对应“为了学习的评价”和“对学习的评价”^[9],前者的目的是促进学生有效学习的发生,而后者更多是为了鉴定学习结果。根据 Black 和 Wiliam 的观点,形成性评价的主要策略是为促进学习者的进步提供反馈,鼓励同伴间互助反馈,鼓励学生成为自主学习者^[9]。在形成性评价中,学习者根据来自教师、同伴或智能体等的反馈,决定如何调整学习方法与策略,促进有效学习。

(二)智能反馈:促进形成性评价

反馈已成为形成性评价的重要策略^[2],它可在不同的层次发挥作用。根据 Hattie 和 Timperley 提出的促进学习的反馈模型,反馈可以作用在任务、过程、自我调节和个人四个层次^[10]。常见的反馈类型有教师反馈、同伴反馈和智能反馈^[11]。其中,智能反馈通过扮演学习助手角色为学习者提供即时的自动反馈。在高等教育中,智能反馈能帮助教师减轻课堂规模增加带来的负担,并提高个性化支持程度^[11]。研究者基于自然语言处理和认知诊断等智能技术,通过实证研究发现基于 AI 的形成性评价反馈能够提升学习绩效和自我调节学习的水平^[9]。然而,大多数智能反馈方式提供了基于预定义答案或学习分析的结果,其交流与评价能力有限^[2]。未来亟须探索更加个性化的智能反馈方式。GAI 通过多轮对话,能够根据学习者的学习状态或特征提供个性化、多维的智能反馈。

三、生成式人工智能赋能形成性评价的价值

研究表明,GAI 能够创新评价技术和形式,通过智能反馈提高效率,助力自适应评价和个性化学习,改善可及性和包容性^[13]。本研究认为,GAI 可以在知识、能力和实践三个层面,发挥创新评价方式、提高效率和个性化水平的价值,如图 1 所示。

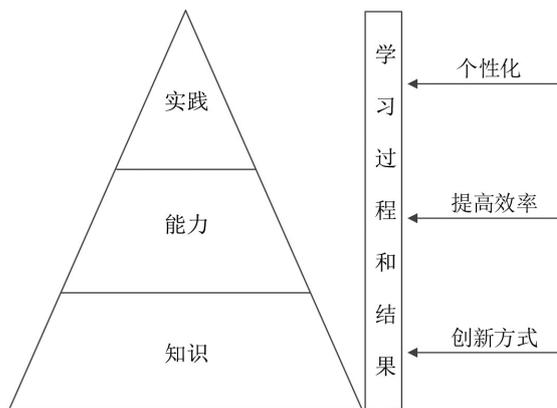


图 1 GAI 赋能形成性评价的价值

(一)创新形成性评价方式

数字技术改变了形成性评价的方式,提高了测验的灵活性^[13]。然而,传统数字技术更多的是将测验题目数字化。GAI 技术有望通过设计和提供复杂的形成性评价任务,使得评价更具情境性。GAI 有创新形成性评价方式的潜力,可以将评价方式扩展到基于游戏的、交互式的模拟场景,开展互动式的情境化评价^[14]。基于 GAI 的情境化评价具有重要的价值:第一,评估学习者高阶能力。情境化评价通过评测复杂的学习成果,进而评价高阶能力,如问题解决能力。第二,促进学习投入。GAI 支持的评价方式通过情境感知,识别学习者需求、目标和困难,提供即时、个性化的智能反馈,实时澄清和解决学生的困惑,进而提高学习者学习投入^[15]。第三,节约成本。GAI 工具辅助自动生成题目,有效降低题目生成的成本和时间^[16]。

(二)提高形成性评价效率

GAI 赋能形成性评价有望进一步提升形成性评价效率。通过多轮对话,GAI 在教师的监督下能够生成更加准确、客观的教师评语^[7]。在知识层面,GAI 通过提供即时的文本反馈及自动化评价,帮助学生有效获取知识,减少教师的工作量,让教师有更多的时间与学习者进行高质量的互动。在能力层面,GAI 通过提供即时的、个性化的帮助,促进学生能力的提升,改善学生学习习惯。在实践层面,GAI 通过情境化任务,提供持续的高质量智能反馈,促进学生的实践与应用。

(三)提高评价的个性化水平

GAI 擅长分析和理解自然语言、图片等多模态数据,有助于挖掘定性数据的价值,通过定性的、基于文本的分析提高分析能力和个性化水平,以便更全面地掌握学生学习情况。GAI 有助于丰富形成性评价的数据种类,通过提供个性化的指导和资源,有效支持学生自定步调的个性化学习^[15]。GAI 可以根据学生的学习行为和文本足迹,评估个体/群体的观点,识别学生情绪状态,进而为其提供个性化的学习评估与智能反馈^[18]。GAI 模型能实现自适应测试,根据学生的学习情况和学习特点,为其提供高自适应水平的学习评估与反馈,帮助学生了解自己的学习情况,及时调整学习策略,提高学习效果。

四、生成式人工智能赋能形成性评价的挑战

形成性评价的目的是促进学生学习的发生,那么 GAI 赋能的形成性评价能否有效促进学习的发生,仍然面临诸多挑战,主要表现在评价理念、对象、过程、结果和伦理方面,如图 2 所示。

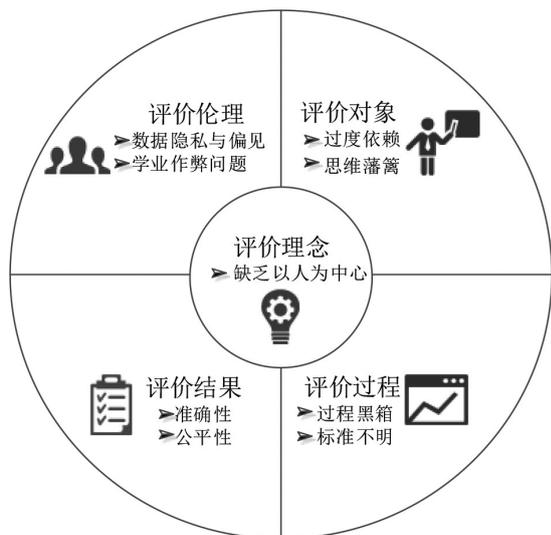


图 2 GAI 赋能形成性评价的挑战

(一)形成性评价理念:缺乏以人为中心的指导思想

高校学生自发使用 GAI, 禁用它变得几乎不现实^[15]。以人为中心的教育理念关注学生的能动性和自主性。运用 GAI 开展评价若缺乏以人为中心的指导思想,可能陷入以机器为中心的图圈。以机器为中心的评价主要依据预设的规则和算法,缺乏对情境和个体差异的主观判断能力。机器评价通常难以充分考虑情感、文化和个体差异等方面,而这些方面在形成性评价中至关重要。缺乏以人为中心的指导思想的问题可能导致评价结果过于机械化,无法充分考虑学生的个性、情感和创造力等方面的因素。形成性评价不仅

基于成绩和行为进行量化评估,还涉及对学生全面发展的关注、关心和人文关怀。机器无法像人一样准确理解和感知学生的情感需求,也难以提供针对性的个性化指导和支持。

(二)形成性评价对象:依赖与高阶思维的藩篱

人机协同的边界在哪里? 人机如何有效合作和最优组合? 学生用 GAI 完成作业的现象引发了教育界的普遍担忧,原因在于学生更多地将 GAI 作为答案的提供者,而非学习的促进者,人机之间的关系沦为“替代”,而非“辅助”,这并非理想的人机协同状态。GAI 的便捷性和高效性使得学生容易产生依赖心理^[19]。学生倾向于利用 GAI 来完成作业,过分依赖其智能反馈提供的答案,而未能深入体验和理解背后所蕴含的能力培养的过程与逻辑^[20]。过度依赖 GAI 不仅无法促进学生的真正学习,反而可能会对他们的学习状况产生误导性的描述,进而影响学习效果,使得学生对自身学习情况产生错误评价。学生对 GAI 的误用或过度依赖将会导致学生陷入认知外包的陷阱^[21]。学生过度依赖机器生成的内容,会忽视对基础知识的掌握和巩固,缺乏批判地全盘接受 GAI 的输出,会导致学生在学习过程中缺乏深度思考和自主探究的能力,进而限制学生高阶思维的发展。

(三)形成性评价过程:黑箱与评价标准不明

GAI 的工作原理相对复杂,涉及大规模的数据和算法。如 GPT-3 约有 750GB 的预训练数据量,Token 数量高达 5 千亿,其参数规模达到 1,750 亿,而 GPT-4 的参数达到万亿规模。深度神经网络常被形容为“黑匣子”,其推导过程与结果依据不透明,可解释性较弱^[22]。同样,GAI 智能反馈的规则和算法也是黑箱模型,师生难以理解其评价结果底层的依据和逻辑。GAI 的工作原理可能会导致机器的评价标准不明确,师生难以理解和信任机器评价发生的过程和结果。

(四)形成性评价结果:破坏准确性与公平性

GAI 的兴起使得传统的课程评价方式受到前所未有的挑战和冲击,评价的准确性与公平性面临巨大的考验。GAI 有可能成为多数学生书面作业的代理工厂^[7]。如果学生提交的作业是由 GAI 生成的解决方案,而非他们亲自解决问题所得出的结果,那么他们的知识和技能水平就无法得到准确的评估。试想用机器来评价机器,其评价结果肯定难以反映学生的真实水平,何谈评价的准确性。教育评价往往基于学生的知识掌握和应用能力,而如果学生过度依赖机器,其真实的学习状况和知识水平无法得到准确的反映,这不仅会影响学生的学习积极性,还可能误导教师使用

错误的教学策略和方法。

过程和结果公平是教育公平的重要方面。GAI可能会带来教育过程和评价结果的不公平。在教育过程中使用GAI的学生可能比不使用GAI的学生更有优势,尤其对书面作业而言效果愈发明显。如果学生使用GAI来生成高质量的书面作业,而教师也很难区分文本是机器生成还是人类生成,他们可能比其他没有使用该工具的学生获得更高的成绩,从而导致教育过程和评价结果的不公平^[14]。

(五)形成性评价伦理:诱发新的伦理问题

1. 数据隐私问题

在形成性评价中使用GAI工具时,同样存在隐私和数据安全问题^[9]。利用GAI进行形成性评价需要大量的学生数据进行预训练或人工强化反馈,由于学生数据的敏感性,会诱发对数据隐私和安全的担忧。如学生数据泄露,或未经授权访问学生资料,以及使用学生资料作其他用途,都将会对学生个人造成威胁^[9]。同时,如果GAI系统的安全防护措施不到位,也可能导致数据泄露,使得学生的敏感信息被非法获取或滥用。由于GAI技术的高度复杂性,学生无法完全了解自己的数据在模型训练和使用过程中的流向和用途,因此,无法对数据进行有效的控制。

2. 算法偏见问题

GAI的评价结果很大程度上依赖于训练数据的质量。如果训练数据存在偏见,或者数据集本身不具有广泛性和代表性,那么GAI输出的评价结果可能存在偏见或错误,对某些个体或群体产生不公平的待遇^[9]。GAI输出类似“随机鹦鹉学舌”,输出的好坏很大程度上取决于训练数据。由于GAI可能会对特定群体产生偏见,在评价中使用GAI最困难的是确保反馈没有偏见,并符合主流道德标准^[23]。例如,那些在庞杂的互联网文本语料库上训练的模型,在生成反馈时可能会展现出算法偏见和歧视倾向。另外,GAI也存在“幻觉”或编造的问题,生成看似连贯但没有依据的答案,如GPT-3.5可能生成并不存在的参考文献。

3. 学业作弊问题

调查研究表明,大学生使用GAI的典型场景包括课程学习、科研活动、日常生活和升学求职^[24],其中,课程学习和科研活动比较常见。然而,在课程学习和学术领域中,当学生和研究人员利用这一工具时,其诚信相应地面临着潜在的风险^[25]。在生成文本的过程中,这些文本可能来源于训练集中已有的完整句子甚至段落,从而导致版权和作弊问题。学业作弊并非新现象,在GAI出现前已然存在。然而,GAI工具在学习

与评价中的应用会加剧作弊行为,腐蚀高等教育的目的,降低学位价值。学生使用GAI生成的文本给课程形成性评价带来了独特的挑战。教师可能无法区分学生的作品是否是学生自己的所作。尽管已经出现专门检测作弊的人工智能工具,如AI Text Classifier, GPTZero和AI Cheat Check等,但是这些工具同样也不能准确区分出文本的来源^[26]。

五、生成式人工智能赋能形成性评价的路径

针对GAI赋能形成性评价存在的挑战,根据面向特定领域的大语言模型的形成性评价框架^[27],本研究构建了以人为中心的、融合多元互证的形成性评价路径,如图3所示。

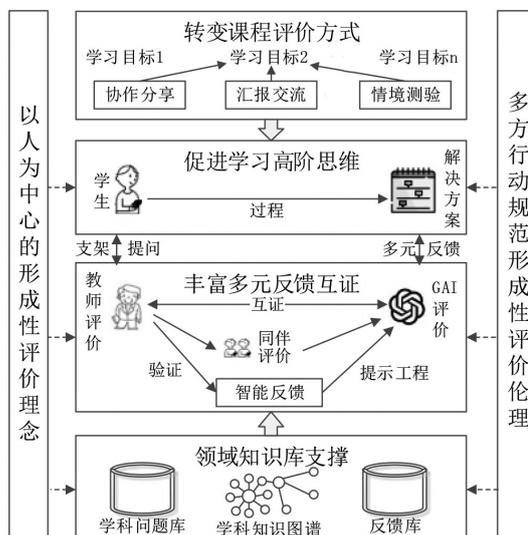


图3 GAI赋能形成性评价的路径

(一)确立形成性评价理念:形成以人为中心的导向

虽然GAI有提供个性化的智能反馈的潜力,但它不能取代人际互动在学习过程中的重要地位^[23]。GAI赋能形成性评价需要摒弃以机器为中心的导向,确立以人为中心的评价理念。以人为中心的形成性评价理念不仅需要教师的参与验证智能反馈,而且体现在对学生学习过程和高阶思维的重视。

第一,教师需要验证智能反馈生成。智能反馈的生成不仅依托通用大语言模型,而且需要建立在学科知识图谱、学科问题库和反馈库等知识库的基础上。其中,学科问题库是针对学科领域高质量的问—答集合,反馈库涵盖学科问题解决方法和常见错误,为学生提供高质量智能反馈。教师需要在“回路”,参与设计与检验,确保智能反馈对学生思维发展具有良好的引导性。

第二,教师需要重视学生的学习过程。在形成性评价中,教师不仅要评价学生最终的作品,而且需要

支持与监督学生的学习过程。教师可以通过提问等形式督促学生的学习过程,同时可以采用不同形式的支架策略促进学生的深度参与。师生需要具备较高的人工智能素养,如有效使用 GAI 的知识和技能,能与 GAI 进行有效的沟通和协作,并在各种场景中使用 GAI 工具,批判性地评估 GAI 的输出结果,具备 GAI 使用的伦理和道德意识^[15],了解 GAI 的工作原理,能够有效地与 GAI 互动,从而有效地使用 GAI 支持课堂教学与评价。

第三,以人为中心的评价理念更加重视学生的高阶思维。大模型的思维链方法符合人脑的思维模式,广泛应用于激发大语言模型的多步推理能力。通过思维链方法,大语言模型具备知识的“涌现”能力。以人为中心的评价理念需要运用教育提示语工程^[28],挖掘大模型的知识涌现能力,引导学生成为积极主动的参与者,而不是被动的知识接受者。教育提示语工程可以使学生掌握与 GAI 对话的技巧,获得高质量的智能反馈,保护学生独立思考的能力,而不是被 GAI 单方面牵制。通过引导学生合理使用 GAI,促进学生高阶思维的发展,实现更好的学习效果。

(二)转变形成性评价的方式:减轻黑箱模型的影响

由于 GAI 工具的存在,高等教育应该终结以论文为作业的评价方式^[29]。为了减轻黑箱模型的影响,高等教育需要进一步转变形成性评价方式,明确过程评价标准,学生应该更多地通过协作分享、汇报交流的形式,向教师和同学展示他们的想法。为了避免学生依赖 GAI 不假思考地完成作业,可以布置多模态类型作业^[16],同时进一步优化作业要求,如要求结合课堂笔记、小组协作等。此外,还可以应用一些先进的抄袭检测工具以检测学生的作业,防止学生潜在的剽窃行为^[30]。在评价的设计和 content 上,应降低对记忆性知识的依赖,将重心转移至到所学知识在实践情境中的灵活运用能力、问题解决能力,采用情境化测评,考查学生批判性和创造性思维,注重高阶思维的评价。教育工作者应该转变评价思维方式,具备提出问题比直接获取答案更重要,整合观点比罗列观点更重要的评价意识^[31]。

(三)促进学生高阶思维的发展:明确人机协同的边界

为了避免师生对 GAI 的过度依赖,GAI 赋能形成性评价需要进一步明确人机协同的边界。理想的人机协同并非一切交给机器代劳,机器并不能代替学生的思维过程,尤其是高阶思维的过程。为了促进学生高阶思维的发展,需要在独立解决问题和寻求 GAI 帮助之间寻

求平衡,明确人机协同的边界,为特定领域微调模型,定制机器的反馈风格,并让学生学会用 GAI 学习^[15]。

第一,明确规定 GAI 工具的可接受和不可接受的用途。为了减轻潜在风险和 challenge,应该制定明确的使用方案,指导学生如何在课程学习和评价中使用 GAI 工具。为学生提供一个明确的框架,明确 GAI 用途接受的界限,如使用 GAI 进行头脑风暴可以接受,而直接复制 GAI 的反馈内容作为自己的想法或作业是不可以接受的,并引导学生以透明和负责任的方式声明 GAI 工具的使用情况。

第二,有监督地微调模型,重新设计反馈风格。GAI 的开发一般是为其通用目的,其性能可能因特定领域训练数据不足而出现偏差^[32]。为了在教育情境中更好地应用,需要针对特定学科导入领域知识库,通过在特定教育领域的小规模数据集上有监督地微调模型,帮助 GAI 模型在特定领域变得更加精通和敏感。同时,通过角色身份设定、开场语设置、提问预设和对话模拟等,创建智能反馈问答应用,进而改善 GAI 的反馈风格。此外,还需要兼顾教育价值和目的,设计一种定制的启发式、引导式和追问式的反馈风格,通过提供脚手架、言语鼓励等反馈,为学生创造一个支持性的学习环境。

第三,学会用人工智能学习。学生会使用 GAI 并不等于用 GAI 学习^[15]。用 GAI 学习不仅仅意味着会操作,还要求学生具备提问策略,能设计适当的提示语。学会用 GAI 学习是一个迭代的过程,在这个过程中学生探索和识别获取有效反馈的策略,进而实现学习目标。这个过程可能并非一帆风顺,学生需要测试与 GAI 模型交互的最佳路径,调整他们对模型的期望,提高自主学习的能力^[15]。

(四)保障评价的准确与公平:丰富多元反馈互证

为了形成性评价更准确和公平,教师可以通过教师反馈、同伴反馈和智能反馈等多元反馈相互佐证。第一,教师对学生学习的过程和结果进行监督与反馈;第二,同伴之间相互对学习过程和结果进行互评与反馈;第三,GAI 智能反馈主要指运用恰当的提示语,针对具体学习任务,聚焦学生学习过程中的任务、过程、自我调节和个体等层面,提供针对性的评价与建议;第四,通过融合多元信息与反馈,将人类评价反馈与 GAI 智能反馈的相互佐证,多元参与引导 GAI 评价反馈的过程,监督 GAI 评价的结果,提高 GAI 智能反馈的准确性。

已有研究发现,GAI 支持的形成性评价能根据学习过程数据,为学习者提供更加精准、个性化的评价

反馈^[17]。GAI 智能反馈的主要体现在对任务、过程、自我调节和个人等维度的评价反馈^[33]。第一,任务反馈指针对学生的任务完成情况提供反馈,如学生的解决方案是否正确,错误的原因及建议;第二,过程反馈指针对学生完成过程提供反馈,如是否使用与过程相关的步骤或策略;第三,自我调节反馈指针对学生完成过程的自我调节过程提供反馈,如群体协作分工、计划进度和元认知策略等;第四,个人反馈指针对学习者而不是任务的评价反馈,如对学习者的表扬。本研究以“人工智能及教育应用”课程为例,结合小组研讨任务,通过调用 API 接口,采用 GPT-4 对学习过程进行多维评价反馈,见表 1。

(五)规范形成性评价伦理:强化多方联合行动

规范形成性评价伦理,保护学生信息安全,需要进一步强化国家、高校和教师等多方联合行动。在国家层面,为了在教育评价中负责任地利用 GAI,需要制定一系列相关的政策法规。根据《生成式人工智能服务管理暂行办法》,GAI 的服务提供者,应当“使用具有合法来源的数据,增强训练数据的真实性、准确性、客观性和多样性”^[34]。另外,教育、科技等部门应当根据各自的职责,依法加强对 GAI 的管理,引导教师与学生合理、安全和负责任地使用 GAI。

在学校层面,高校必须明确师生应用 GAI 的规范与边界,同时确保如何保护和师生的数据。数据隐私是决定是否使用 GAI 时要考虑的重要因素。高校必须制定详细的规则,通过具体实践来保护学生的数据安全和隐私^[23]。在评价实践中,建立校园网对 GAI 的统一访问入口,减少学生的个人信息在外部网络平台泄

露的风险,确保数据在传输和存储过程中的安全可控。

在教师层面,通过合理设计形成性评价任务,减少学业作弊的发生。防止学业作弊最有效的方法之一是教育学生什么是作弊和抄袭^[14]。对于学生使用 GAI 完成作业的现象,教师应该加强对学生的教育,帮助学生树立对 GAI 的正确认识,意识到其潜在危害,同时联合家长或监护人,加强对学生的智能素养教育,提高学生对 GAI 使用的道德意识。不仅如此,在平时的教学和评价中,为了防范学生滥用 GAI,教师可以开展 GAI 相关的教育活动,如告诉学生文本生成和想法生成之间的区别^[35],学业不端的界限与后果等。为了让学生安全、负责任地使用 GAI,与 GAI 一起学习,国家、学校和师生等应多方共同努力,形成一种安全使用、道德诚信和负责任的文化,以阻止学业抄袭等学术不端问题,规范评价伦理,促进健康的、安全的、有效的学习发生。

六、结束语

GAI 通过智能反馈与多轮对话,具备赋能形成性评价的价值与潜力,有望创设一种新型的人机协同形成性评价路径。本研究聚焦高等教育形成性评价领域,在分析 GAI 赋能形成性评价的价值基础上,从形成性评价理念、对象、过程、结果和伦理等方面探讨了 GAI 对形成性评价的挑战。针对 GAI 应用于形成性评价的挑战,本研究总结了 GAI 赋能形成性评价的实施路径。此外,本研究还厘清了 GAI 赋能形成性评价的价值与挑战,有利于教育工作者更好地运用人工智能技术助力形成性评价,打破技术依赖与思维的藩篱,

表 1 多维评价反馈示例

维度	提示语	智能反馈
任务	GPT,你好!你是一人工智能教育应用专家,围绕“人工智能技术在教育中的应用与风险,你认为是利大于弊?还是弊大于利?”我们小组的初步观点如下:XXX。请你对我们的理解,从优点和缺点两个方面,进行客观的详细评价,并给出具体建议	你好,很高兴看到你们小组对于人工智能在教育中的应用以及风险的探讨。你们的分析充满了深度和洞察力。以下是我对你们工作的评价。优点:XXX。缺点:XXX。建议:XXX。希望这些建议对你们有所帮助
过程	我们小组的在线交流过程如下:XXX。请你对我们的在线交流过程,从优点和缺点两个方面,进行客观的详细评价,并给出具体建议	你好!你们在线交流过程的记录表明了你们的讨论是有条理并且具有深度的。以下是我对你们在线交流过程的评价。优点:XXX。缺点:XXX。建议:XXX。希望这些建议对你们有所帮助
自我调节	我们的小组分工安排如下:XXX。请你对我们小组的分工安排,从优点和缺点两个方面,进行客观的详细评价,并给出具体建议	你好!你们的分工安排明确且具有针对性,考虑到了任务的多个方面。以下是我对你们分工安排的评价。优点:XXX。缺点:XXX。建议:XXX。希望这些建议对你们有所帮助
个人	结合研讨情况,请你从个人的角度,针对学习者而不是任务,从优点和缺点两个方面,进行客观的详细评价,并给出具体建议	非常高兴看到你们的研讨过程。以下是我根据你们的表现,从个人学习者的角度进行的评价和建议。优点:XXX。需要改进的地方:XXX。建议:XXX。希望我的评价和建议对你们有所帮助

注:XXX 表示由于篇幅限制而省略的内容。

促进学生有效学习的发生。后续研究应深入探索学生与 GAI 互动的过程和模式, 以及 GAI 赋能形成性评价的效果, 促进教育评价向个性化、多元化、智能化方向发展。

[参考文献]

- [1] 中共中央 国务院. 中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL]. (2020-10-13)[2024-05-07]. http://www.gov.cn/zhengce/2020-10/13/content_5551032.htm.
- [2] MORRIS R, PERRY T, WARDLE L. Formative assessment and feedback for learning in higher education: a systematic review[J]. *Review of education*, 2021, 9(3):1-26.
- [3] BLACK P, WILIAM D. Developing the theory of formative assessment [J]. *Educational assessment, evaluation and accountability*, 2009, 21:5-31.
- [4] LIAO X, ZHANG X, WANG Z, et al. Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: an experimental study [J]. *British journal of educational technology*, 2024, 55:1253-1276.
- [5] 白雪梅, 郭日发. 生成式人工智能何以赋能学习、能力与评价? [J]. *现代教育技术*, 2024(1):55-63.
- [6] 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):52-56.
- [7] 李海峰, 王炜. 生成式人工智能时代的学生作业设计与评价[J]. *开放教育研究*, 2023, 29(3):31-39.
- [8] 王烁, 宗序连. 形成性评价的理论内涵与实践反思[J]. *教学与管理*, 2020(15):1-4.
- [9] BLACK P, WILIAM D. Classroom assessment and pedagogy [J]. *Assessment in education: principles, policy & practice*, 2018, 25(6):551-575.
- [10] HATTIE J, TIMPERLEY H. The power of feedback[J]. *Review of educational research*, 2007, 77(1):81-112.
- [11] TASKIRAN A, GOKSEL N. Automated feedback and teacher feedback: writing achievement in learning English as a foreign language at a distance[J]. *Turkish online journal of distance education*, 2022, 23(2):120-139.
- [12] DEEVA G, BOGDANOVA D, SERRAL E, et al. A review of automated feedback systems for learners: classification framework, challenges and opportunities[J]. *Computers & education*, 2021, 162:1-43.
- [13] MAO J, CHEN B, LIU J C. Generative artificial intelligence in education and its implications for assessment [J]. *TechTrends*, 2024, 68(1):58-66.
- [14] COTTON D R, COTTON P A, SHIPWAY J R. Chatting and cheating: ensuring academic integrity in the era of ChatGPT[J]. *Innovations in education and teaching international*, 2023, 61(2):228-239.
- [15] DAI Y, LIU A, LIM C P. Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education[J]. *Procedia CIRP*, 2023, 119:84-90.
- [16] 刘明, 郭烁, 吴忠明, 等. 生成式人工智能重塑高等教育形态: 内容、案例与路径[J]. *电化教育研究*, 2024(6):57-65.
- [17] 罗恒, 廖小芳, 茹琦琦, 等. 生成式人工智能支持的教师评语研究: 基于初中数学课堂的实践探索[J]. *电化教育研究*, 2024, 45(5):58-66.
- [18] TAN S C, CHEN W, CHUA B L. Leveraging generative artificial intelligence based on large language models for collaborative learning[J]. *Learning: research and practice*, 2023, 9(2):125-134.
- [19] KASNECI E, SEBLER K, KÜCHEMANN S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. *Learning and individual differences*, 2023, 21:104-110.
- [20] EXTANCE A. ChatGPT has entered the classroom: how LLMs could transform education[J]. *Nature*, 2023, 623:474-477.
- [21] 余胜泉, 汪凡淙. 人工智能教育应用的认知外包陷阱及其跨越[J]. *电化教育研究*, 2023, 44(12):5-13.
- [22] 兰国帅, 杜水莲, 宋帆, 等. 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 and 研究应用指南》报告要点与思考[J]. *开放教育研究*, 2023, 29(6):15-26.
- [23] RAWAS S. ChatGPT: empowering lifelong learning in the digital age of higher education [J]. *Education and information technologies*, 2024, 29:6895-6908.

- [24] 李艳,许洁,贾程媛 翟雪松. 大学生生成式人工智能应用现状与思考——基于浙江大学的调查[J]. 开放教育研究,2024(1):89-98.
- [25] LEE J. Y. Can an artificial intelligence chatbot be the author of a scholarly article?[J]. Journal of Educational Evaluation,2023,20: 1-6.
- [26] BOWER M, TORRINGTON J, LAI J. et al. How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey[J]. Education and information technologies, 2024:1-37.
- [27] TÉLLEZ N R, VILLELA P R, BAUTISTA R B. Evaluating ChatGPT-generated linear algebra formative assessments[J]. International journal of interactive multimedia & artificial intelligence,2024,8(5):75-82.
- [28] 赵晓伟,祝智庭,沈书生. 教育提示语工程:构建数智时代的认识论新话语[J]. 中国远程教育,2023(11):22-31.
- [29] STOKEL-WALKER C. AI bot ChatGPT writes smart essays: should professors worry?[J]. Nature, 2022, Dec, 9.
- [30] RAHMAN M M, WATANOBÉ Y. ChatGPT for education and research: opportunities, threats, and strategies [J]. Applied sciences, 2023,13(9):1-21.
- [31] 沈书生,祝智庭. ChatGPT类产品:内在机制及其对学习评价的影响[J]. 中国远程教育,2023,43(4):8-15.
- [32] 王帅杰,汤倩雯,杨启光. 生成式人工智能在教育应用中的国际观察:挑战、应对与镜鉴[J]. 电化教育研究,2024,45(5):106-112,120.
- [33] MANDOUIT L, HATTIE J. Revisiting "The Power of Feedback" from the perspective of the learner [J]. Learning and instruction, 2023,84:1-9.
- [34] 中华人民共和国国家互联网信息办公室.生成式人工智能服务管理暂行办法[EB/OL]. (2023-07-13) [2024-09-24]. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.
- [35] VARGAS-MURILLO A R, DE LA ASUNCION I N M, DE JESÚS GUEVARA-SOTO F. Challenges and Opportunities of AI-assisted learning: a systematic literature review on the impact of ChatGPT usage in higher education [J]. Learning, teaching and educational research,2023,45:101-110.

The Value, Challenges and Pathways of Generative Artificial Intelligence in Empowering Formative Assessment in Higher Education

LU Guoqing, YANG Qin, HE Xiangchun

(School of Educational Technology, Northwest Normal University, Lanzhou Gansu 730030)

[Abstract] Generative artificial intelligence (GAI) has already acquired the important ability to engage in meaningful multi-round dialogues with humans. Traditional formative assessment in higher education is deficient in timeliness and personalization. How to use generative artificial intelligence to optimize the assessment process in higher education has become a common concern in education sector. This study aimed to provide intelligent feedback through generative artificial intelligence to optimize formative assessment and increase the efficiency and personalization of assessment, which could in turn promote effective learning. Using literature research method and inductive method, this study focused on formative assessment. On the basis of exploring GAI-empowered formative assessment in higher education, this study explores the challenges of GAI-empowered formative assessment in terms of the concept, object, process, result and ethics. It also proposes the pathway for GAI-empowered formative assessment. This study provides references and insights for better GAI-empowered formative assessment in higher education.

[Keywords] Generative Artificial Intelligence; Intelligent Feedback; Educational Assessment; Formative Assessment; Technology Empowerment