

推进多模态大模型的教育深度应用： 价值赋能、场景构建与实施建议

◆ 杨现民 卜浩德 李新

[摘要] 以多模态大模型为核心的人工智能技术正在重塑社会组织与运行模式，成为驱动社会结构变革和文化演变的重要力量。多模态大模型具备多模态内容理解、多模态资源生成、多模态人机交互等技术特征，其教育应用价值体现在助力多模态学习分析、优化数字资源供给方式以及赋能智适应学习三个方面，典型应用场景涵盖教学辅助、学习辅导、教育管理、教育评价以及教育研究。为进一步推进多模态大模型在教育领域的深度应用，研究提出五条实施建议：明确应用需求，拓展多模态大模型教育应用场景；注重素养提升，强化教育利益共同体多模态大模型应用能力；坚持三项原则，提升多模态大模型与教学深度融合的效度；选好技术产品，建立多模态大模型教育服务支持体系；关注应用绩效，持续开展多模态大模型教育影响研究。

[关键词] 多模态大模型；教育应用；价值赋能；场景构建；实施建议

[中图分类号] G434

[文献标识码] A

[文章编号] 1002-4808(2025)04-0009-06

随着新一轮科技革命的纵深发展，以多模态大模型为核心的人工智能技术取得了前所未有的突破，正在重构社会组织与运行模式，成为驱动社会结构变革和文化演变的重要力量。在全球态势的影响下，越来越多的研究者与实践者开始探索多模态大模型对教育的变革作用，围绕多模态大模型助力精准教学^[1]、促进教育观念转变^[2]、驱动学科知识图谱进化^[3]以及拓展学习科学的研究手段^[4]等进行了探讨。此外，教育领域的科技企业也迅速意识到多模态大模型的教育应用潜力，相继推出了基于多模态大模型的教育智能体，能够以多模态交互的方式回答师生问题、生成教育资源、实现逻辑推理等。整体来看，多模态大模型逐渐成为人工智能领域的研究热点，并在通用领域有显著进展，但在教育领域仍处于起步阶段。由此，为了进一步推进多模态大模型在教育领域的深度应用，充分释放其对教育系统的变革性潜力，本研究系统分析多模态大模型赋能教育的基本逻辑、构建多模态大模型的教育应用场景，并提出推进多模态大模型教育深度应用的实施建

议，以期推进人工智能与教育的深度融合与创新发展。

一、价值赋能：多模态大模型作用于教育的基本逻辑

自1956年约翰·麦卡锡(John McCarthy)等人在美国达特茅斯学院研讨会上第一次提出人工智能的概念以来，人工智能大模型经历了多次发展和变革，逐渐实现了从单模态大模型向多模态大模型的跨越，体现在输入输出内容从文本拓展到视频、音频、图像等多种模态，显著提升了模型的理解和泛化能力，更接近人脑的认知机制。这种认知机制充分发挥多模态大模型在多模态内容理解、多模态内容生成、多模态人机交互三个方面的优势，来助力教育教学中的多模态学习分析、数字资源供给以及智适应学习等，其基本逻辑框架如图1所示。

(一)多模态内容理解，助力多模态学习分析
在学习分析领域，多模态数据融合分析既是核心，也是面临的主要挑战。多模态大模型展现

杨现民/江苏师范大学江苏省教育信息化工程技术研究中心教授、博士生导师(江苏徐州 221116)；卜浩德/江苏师范大学江苏省教育信息化工程技术研究中心博士研究生(江苏徐州 221116)；李新/江苏师范大学江苏省教育信息化工程技术研究中心讲师、硕士生导师，本文通讯作者(江苏徐州 221116)。

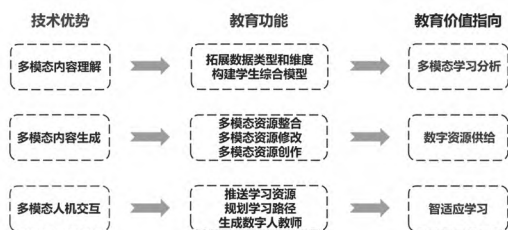


图1 多模态大模型赋能教育的逻辑框架

出卓越的多模态内容理解能力，不仅能激活教育数据的潜在价值，还能通过融合多模态数据，揭示学生学习与发展的特征和规律，实现更为精准和全面的学习分析。在学习分析中可以使用多模态大模型处理和分析来自不同渠道的数据，如文本、语音、图像、视频等，进而从学生的语言、动作、表情和生理信号等多个方面自动分析学习过程，拓展学习分析的数据类型和维度^[5]。此外，多模态大模型的深度学习算法能够自动学习不同模态数据间的对应关系与互补关系，通过数据的互补性构建关于学生知识掌握、认知结构、情感状态、交互行为和創新精神的综合模型。

(二) 多模态内容生成，优化数字资源供给方式

多模态大模型能够为师生创作教育资源，根据师生指令输出个性化内容，以人机协同的方式强化教育资源供给能力^[6]。首先，多模态大模型能够根据教学内容和师生指令进行数字内容的个性化创作，如根据师生文字描述生成图像、根据图片内容生成视频等。其次，多模态大模型能够按照师生提示对教育资源进行调整，优化现有资源内容和形式，如根据提示修改教学音频的内容，使之更加贴合教学设计和学生理解能力。最后，多模态大模型能够建立不同资源间的语义关系，优化资源供给方式。如将同一知识点相关的视频、教案、测试题、知识点等进行系统归类与智能关联，自动筛选并推送给师生最合适的学习资源。

(三) 多模态人机交互，赋能智适应学习

多模态大模型通过推送学习资源、规划学习路径以及提供数字教师等方式赋能智适应学习。首先，多模态大模型能够在课堂中实时分析学生的多模态信息，进行情感识别、学习分析、学业预测、情境感知等，更深入地了解每个学生的学习需求和风格，进而为其提供针对性的学习建议与学习资源^[7]。其次，多模态大模型能够深度模拟专家型教师的知识和经验，根据学生画像为其规划学习路径和知识序列，辅助教师进行一对一的

指导。最后，多模态大模型能够提供数字教师，这种交互方式能够为学习者创设更加具身、沉浸、真实的教學互动，提升学生的学习投入。

二、场景构建：多模态大模型教育应用的五大场景

国内外学者针对大模型的教育应用开展了一系列研究，涵盖教育评价、教学服务、教育治理、科研辅助、学习辅导、资源生成等场景^[8]。本研究在已有研究基础上，进一步细化了多模态大模型的教育应用场景与实际应用案例，构建了多模态大模型教育应用的五大场景，包括教学辅助、学习辅导、教育管理、教育评价、教育研究，如图2所示。

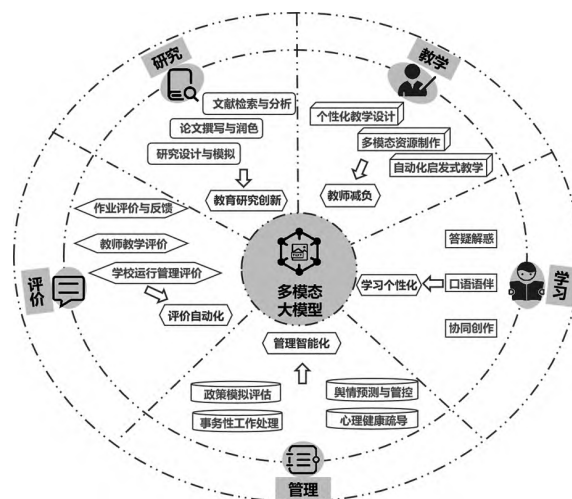


图2 多模态大模型的教育应用场景

(一) 教学辅助：减轻教师负担

多模态大模型凭借多模态内容理解和生成能力，为个性化教学设计、多模态教学资源生成以及启发式教学提供了强有力的支持^[9]，从而有效减轻了教师的工作负担，支持教师将更多精力投入到更有创造性的教学活动中。

首先，多模态大模型可以根据教师要求生成个性化教学设计，支持教师以Word、PDF、Excel、PPT、PNG等多种格式输入课程资料，并依据教师提示语设计合理的教学目标、教学模式以及教学策略，构建与课程匹配的探究问题和学习任务。例如在国外的一项实验中，研究者命令ChatGPT和Gemini为七年级数学、科学、社会研究和英语课程生成18份教学设计，然后邀请经验丰富的教师通过内容分析法对这些设计进行分析与评价，发现ChatGPT和Gemini设计的教案在学习目标

和活动方面能很好地与课程主题相契合，基本满足课堂教学的实际要求^[10]。其次，多模态大模型不仅能根据教师指令生成科学插图、评价量表、演示视频等不同形式的教学资源，还可以根据教师的要求，将已有教学资源转化为其他模态，如将科学插图转换成测试题、将语音描述转换成视频演示等。

《2024智能教育发展蓝皮书》公布的案例中，合肥市第七中学的老师在教学《林教头风雪山神庙》这一文时，利用大模型支持的教师助手生成该单元的教学规划，然后基于该规划生成图文并茂的教学课件，并借助大模型的多模态功能生成符合林冲形象的图片，丰富学生的学习体验。最后，多模态大模型能够作为教学助手，通过对话式引导促进学生的学习。在教学过程中，学生可以通过文字、语音、图片等形式提出问题，多模态大模型则通过逻辑推理理解学生困惑，为学生提供解决思路。例如在佛山市顺德区人工智能赋能课堂教学展示交流活动中，荔村小学的教师利用大模型生成教学智能体辅助《王戎不取道旁李》一文的教学，通过“李白”等历史人物智能体的多模态交互功能与学生开展对话，回答学生关于课文内容的问题，帮助学生理解主人公“王戎”的性格特征，增强了教学的情境性，创设出富有沉浸感的学习环境。

（二）学习辅导：促进个性化学习

在学生进行自主学习的过程中，多模态大模型通过答疑解惑、口语学伴以及协同创作等功能，帮助学生更高效地掌握知识和技能，实现更个性化的自主学习体验。

第一，多模态大模型可以作为问答助教随时回答学生提出的问题，确保学生在遇到学习障碍时能得到即时帮助。例如，一些多模态大模型提供智能体定制功能，允许学生根据自己的需求创建智能体辅助学习，并且支持拍照、语音等交互方式，这不仅提高了学习效率，还激发了学生的学习兴趣。第二，多模态大模型可以作为口语学伴，模拟真实生活中的对话场景与学生进行“数字人”对话，帮助学生在接近真实的情境中练习口语^[11]，提升学生的语言应用能力。例如多家企业在多模态大模型GPT-4的基础上推出了语言学习工具，能够为学生生成智能学伴，让学生与数字人进行互动对话，帮助学生提升口语能力。第三，多模态大模型能够作为学习伙伴与学生开展协同创作^[15]，为学生提供绘画创作、文学写作、音乐创作等创作行为的灵感和创意。例如基于多

模态大模型GPT-4的人工智能助手，能够与学生共同合作完成作文写作，它不会直接给出文章，而是通过多次问答为学生提供写作思路，展现出多模态大模型在辅助学生创作方面的潜力。

（三）教育管理：推动智能化管理

多模态大模型在教育管理方面发挥着重要作用，它不仅能够协助管理者高效处理日常事务性工作，还能在政策制定与评估、舆情预测与管控、学生心理健康疏导等方面提供支持，从而提高教育管理效率和智能化水平。

在日常工作方面，多模态大模型能够迅速完成常规任务，将原本耗时的文书工作简化为快速流程。它可以根据指令生成规章制度、会议通知、宣传海报、会议发言稿等文档，极大地提高工作效率。在教育政策制定与评估方面，多模态大模型能够利用历史数据预测未来教育发展趋势和人才培养需求，协助政策制定者提前规划和制定更为科学的教育政策。在政策实施过程中，模型能够分析来自不同渠道和形式的反馈数据，帮助政策制定者及时了解政策实施效果和社会教育需求，以便及时调整政策方向。在教育舆情预测与管控方面，多模态大模型能够筛选和处理来自社交媒体、论坛、新闻网站等不同渠道的数据，识别网民通过语言、符号和图片表达的观点，实现对教育舆情的实时分析^[13]。针对舆情分析结果，多模态大模型可以为教育部门和学校管理者提供个性化的舆情管控和应对策略，为教育决策提供智能化支持，提高应对效果。例如，利用人民网舆情数据中心积累的行业知识数据所训练的大模型，能够实现舆情分析、新闻摘要生成等功能，以及时了解各领域的舆情动态，把握舆论走向。同样，此类模型也适用于教育领域，可以为管理者精准捕捉和分析教育领域的舆情动态，助力管理者做出更加科学合理的决策。在学生健康管理方面，多模态大模型可以与学生开展文字、语音、视频等多种形式的对话，通过处理文字内容、语音语调、面部表情等信息综合判断学生的心理状态。一旦发现潜在心理问题，多模态大模型能够为学生提供定制化的心理疏导，选取学生感兴趣的多模态内容（如电影、音乐、诗歌、故事等）帮助学生调整情绪、舒缓压力，给予情感关怀。例如多家企业推出的AI心理伙伴，能够利用大模型的多模态能力对学生进行心理筛查，并利用心理知识图谱给予学生个性化心理指导建议，有效帮助学生疏导心理压力。

（四）教育评价：实现自动化评价

多模态大模型为教育领域提供了一种创新的自动化工具，能够智能评价学生作业、评估教师教学技能、统计分析教育数据，有效提高了教育评价的效率和准确性。

首先，多模态大模型能够对学生的作业进行智能评价与反馈。学生可以将试题、论文、口语发音、身体动作、编程作品等不同形式的作业上传到大模型平台，模型随后根据指令提供作业得分、评价依据和改进建议，实现个性化的作业批改。其次，多模态大模型可以通过分析教案和教学录像，对教师教学技能开展智能评价，实现自动化评课。具体来说，模型可以分析教师授课内容的深度和广度，评估其是否符合教学大纲和学生的认知水平；分析师生间的互动频率和质量，评估课堂管理能力和互动技巧；分析肢体语言、面部表情、目光交流等非语言行为，评估情感投入和亲和力。例如基于多模态大模型构建的教师分析平台VidAAS^[14]，能够分析教师教学视频和音频中的内容并生成详细的文字描述，然后根据评价量表对文字描述进行总结和评价，生成最终的教学评价结果。这样的评价方式不仅客观、全面，而且能够为教师提供有针对性的反馈，帮助他们明确自己的教学优势与不足，制定有效的提升计划。最后，多模态大模型能够对对学校运行过程中产生的多方面数据进行统计分析，评价学校运行管理情况，如分析学校的财务数据、招生数据、教学资源使用情况等，评价学校的运营效率和财务健康状况。

（五）教育研究：支持教育研究创新

多模态大模型在教育研究领域扮演着重要角色，为教育研究者提供了一种智能化和高效率的研究工具，能够精确检索和分析学术论文、辅助论文撰写和润色、开展实验设计和模拟。

在文献检索方面，多模态大模型超越了传统的关键词搜索，能够通过图像或表格中的内容定位文献，扩展了检索的维度和范围。检索到所需文献后，模型能够自动提取关键信息、解读核心内容、识别研究不足，并形成文献综述，助力研究人员高效处理文献，全面掌握现有研究成果和研究动态。例如，多模态科学文献大模型Uni-Finder综合考虑了科学文献中的多模态元素，如图表、分子结构表示、化学反应方程式等，运用多模态内容理解能力对科学文献进行更全面和更精确的理解，优化了文献理解和分析过程。在论文撰写过程中，多模态

大模型能够提供辅助创作和可视化呈现。它可以根据研究者提供的关键词和主题生成论文大纲、段落或句子，提升论文撰写效率，并进行语法检查、拼写纠错、句式优化等工作，提升文本的流畅性和学术性。同时，大模型能够根据研究者的成果进行科研绘图，将复杂数据和结果转化为直观的图表、图形，增强论文内容的直观性和说服力。例如中国知网提供的中华知识大模型能够作为学术研究助手，以对话的形式辅助研究者写作，激发创意火花，并能够对文章进行深度优化，增强表达力，提升研究效率和论文质量。在实验研究领域，多模态大模型能够整合不同学科和领域的知识，协助教育研究者发现新的实验思路^[15]，并根据研究者的要求设计实验过程、提出实验假设，对不同实验变量进行模拟和预测。此外，多模态大模型能够深度学习教育数据，构建虚拟教学模型，以此预测师生在特定学习条件下的行为和认知数据，形成以“机器猜想”为手段的新教育研究范式^[16]。

三、实施建议：多模态大模型教育应用的实践进路

当前多模态大模型的教育应用已经引起了教育研究者与实践者的广泛关注，并取得了初步的实践进展，但在持续推进的过程中仍面临一系列的现实挑战与难题，需要从制度体制等层面去构建更加完善的治理体系，为多模态大模型的教育应用提供全方位保障。在此基础上，基础教育领域的学校和教师在推进多模态大模型教育应用时可以从五个方面入手，以确保多模态大模型在教育领域的持续、深度以及高效应用。

（一）明确应用需求，拓展多模态大模型教育应用场景

在《生成式人工智能服务管理暂行办法》的指引下，我国积极鼓励大模型等生成式人工智能技术在各行业、各领域的创新应用。在教育领域，一方面教师要正确评估自身的应用需求，深入分析不同单元、知识点的教学需求，明确哪些环节可以通过引入多模态大模型技术来提升教学效果。同时，教师要结合学情、学科特点、目标达成、内容难度等因素选择适切的教学应用时机，防止过早或随意引入多模态大模型导致的教学秩序破坏，影响教学质量。另一方面，教师积极拓展创新应用，在充分了解多模态大模型产品功能的基础上分析教学应用价值，利用多模态大模型创造性地解决实际教育

问题。例如探索其在个性化学习辅导、课堂教学互动、跨学科项目式学习、智能作业批改与反馈等具体场景中的创新应用，并主动总结应用经验，形成可规模化推广的应用模式，助力教育应用新生态的构建与发展。

（二）注重素养提升，强化教育利益共同体多模态大模型应用能力

联合国教科文组织发布了《教师人工智能素养框架》和《学生人工智能素养框架》，将以人为本的心态、人工智能伦理、人工智能基础和应用等列为教师和学生必备的素养。我国也陆续颁布相关政策，全面加强多元教育主体的数字能力，尤其是提升广大师生与管理者的大模型教育应用方面的能力。首先，开展分层分级的多模态大模型教育应用知识技能培训，帮助教师和管理者了解基本理论、价值内涵、发展态势、应用方式、应用风险等，树立大模型生成内容可靠性核验意识^[17]。其次，建立人工智能素养认证体系，明确认证层级以及每个层级对应的能力要求和考核内容，实行理论、实操、项目展示等多元化评估方式。最后，各区域、学校根据自身特点设立人工智能通识课程，并将多模态大模型与各学科融合，在课堂教学、课外实践、项目作业、实验探究等环节中培养学生正确使用多模态大模型的意识 and 能力。

（三）坚持三项原则，提升多模态大模型与教学深度融合的效度

为更好地发挥多模态大模型的技术优势，保障教学中的健康应用，防范潜在风险与应用弊端，在教学应用中应遵循如下三项基本原则^[18]。第一，教师要准确识别学生在使用中的认知主动性和思维加工水平，促进学生积极参与创编大模型提示语，引导学生质疑、分析和评价大模型生成的内容，最大限度地唤醒学生的思维活力。第二，教师要根据学生学习能力合理规划预设性教学内容与生成性学习活动在课堂教学中的时间比例，对低龄学生可提高课堂预设性内容的比重，辅助他们构建对基础概念的认知；对中高年级学生可增加生成性学习活动的比重，借助开放性问题和自由创作图片等方式促进学生在课堂上进行更广泛深入的探究。第三，要审慎评估多模态大模型生成的内容，加强对信息准确性和权威性的甄别，避免课堂教学出现科学性错误。

（四）选好技术产品，建立多模态大模型教育服务支持体系

为了适应教育实践的不断变化和挑

战，需要选取可靠、安全、公平、极简的多模态大模型产品，并在此基础上构建一个全面、高效、灵活的教育服务支持体系。一方面，学校要根据教学需求明确多模态大模型产品的选择标准，这些标准应涵盖大模型的准确性、易用性、透明度以及对不同学习风格的适应性，并确保产品能够与学校现有的教学系统、教学资源以及教育理念相兼容，形成协同效应。另一方面，学校可以与产品提供商合作构建多模态大模型教育服务支持体系，为教师提供针对性技术支持、培训指导、应急响应和咨询服务，以确保教师能高效使用多模态大模型产品开展教学活动。同时，定期收集教师和学生的反馈意见，改进服务流程，在为师生提供优质服务的同时反向促进技术改进和升级。

（五）关注应用绩效，持续开展多模态大模型教育影响研究

多模态大模型教育应用是否能够有效促进学习仍是一个需要深入探索的问题。因此，在推进多模态大模型教育应用的同时，应组织系统性、追踪性绩效评估，深入了解其在教育中的应用效果和现实挑战。首先，编制多模态大模型教育应用绩效的评价标准，从技术性能、教育效果、社会效益等多个维度进行综合考量，并根据评价标准开发评价工具，帮助学校和区域评估者了解多模态大模型的应用效果。其次，开展实证研究，组建由高校教育研究者、一线教师和学校教学管理者组成的研究团队，通过实地考察、案例分析等方式分析多模态大模型在教育领域的应用情况与实施效果，揭示多模态大模型对学生学习成果的影响机制。最后，将多模态大模型教育应用绩效评价结果应用于指导教育实践，根据绩效对其应用过程、机制、技术研发等环节进行优化，确保多模态大模型教育应用能有效提升教育质量，促进学生全面发展。

[本文系江苏省高等教育教改研究重中之重课题“基于‘5G+VR’的师范生远程沉浸式课堂教学实训体系构建与创新实践”（项目编号：2023JSJG018）阶段性研究成果]

[参考文献]

- [1] 叶新东, 刘泽民. 基于多模态大模型的精准教学支持体系构建研究[J]. 远程教育杂志, 2024(1):84-93.
- [2] 陶炜, 沈阳. 从ChatGPT到Sora: 面向AIGC的四能教育

- 和范式革新[J]. 现代教育技术, 2024(4):16-27.
- [3] 罗江华, 张玉柳. 多模态大模型驱动的学科知识图谱进化及教育应用[J]. 现代教育技术, 2023, 33(12):76-88.
- [4] 陈向东, 赵丽娟, 刘泽民. 拓展学科的疆域: 大模型的涌现能力对学习科学的影响[J]. 现代教育技术, 2024(1):44-54.
- [5] 石琬若, 韩锡斌. 生成式人工智能对学习分析研究的影响: 现状与前瞻——2024年学习分析与知识国际会议(LAK24)述评[J]. 电化教育研究, 2024, 45(12):113-120.
- [6] 万力勇, 杜静, 熊若欣. 人机共创: 基于AIGC的数字化教育资源开发新范式[J]. 现代远程教育研究, 2023(5):12-21.
- [7] 罗晓峰. “人工智能+”时代中小学个性化学习研究[J]. 中国教育学刊, 2024(S2):6-7.
- [8] BAHROUN Z, ANANE C, AHMED V, et al. Transforming education: a comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis[J]. Sustainability, 2023, 15(17):12983.
- [9] 赵磊磊, 张黎, 代蕊华, 等. 人工智能赋能教师教育: 基本逻辑与实践路向[J]. 中国教育学刊, 2022(6):14-21.
- [10] BAYTAK A. The Content Analysis of the Lesson Plans Created by ChatGPT and Google Gemini[J]. Research in Social Sciences and Technology, 2024, 9(1):329-350.
- [11] 吴坚豪, 周婉婷, 曹超. 生成式人工智能技术赋能口语教学的实证研究[J]. 中国电化教育, 2024(4):105-111.
- [12] 黄荣怀. 人工智能正加速教育变革: 现实挑战与应对举措[J]. 中国教育学刊, 2023(6):26-33.
- [13] 丁晓蔚, 赵鑫, 刘梓航, 等. 舆情精准研判中的ChatGPT: “大数据、大模型、大计算”范式驱动的理论与应用探索[J]. 传媒观察, 2024(3):45-61.
- [14] LEE U, JEONG Y, KOH J, et al. I See You: Teacher Analytics with GPT-4 Vision-Powered Observational Assessment[J]. Smart Learning Environments, 2024, 11(1):48.
- [15] 谢天, 邱林, 李雨瞳, 等. 大模型时代的社会科学, 何去何从?[J]. 图书情报知识, 2023(6):6-9+30.
- [16] 杨小康, 许岩岩, 陈露, 等. AI for Science: 智能化科学设施变革基础研究[J]. 中国科学院院刊, 2024(1):59-69.
- [17] 杨现民, 丁杰. 如何应对生成式人工智能的可靠性挑战[N]. 中国教师报, 2024-12-04(14).
- [18] 杨现民, 丁杰. 生成式人工智能如何与教学深度融合[N]. 中国教育报, 2024-06-15(4).

Advancing the Deep Application of Multimodal Large Models in Education: Value Empowerment, Scenarios Construction, and Implementation Strategies

YANG Xianmin, BU Haode, LI Xin

Abstract: The artificial intelligence technology with multimodal large model as its core is reshaping the society's organizing and operating modes, becoming a key force driving the change of social structure and cultural evolution. Multimodal large models are characterized by technologies such as multimodal content understanding, multimodal resource generation, and multimodal human-computer interaction. Their educational application value is reflected in three aspects: supporting multimodal learning analytics, optimizing digital resource supply methods, and empowering intelligent adaptive learning. Typical application scenarios include teaching assistance, learning tutoring, educational management, educational evaluation, and educational research. To further promote the deep application of multimodal large models in the field of education, this study proposes five implementation suggestions: clarify the application needs and expand the application scenarios of multimodal large models in education; focus on enhancing literacy and strengthening the multimodal large model application capabilities of the educational community; adhere to three principles to improve the effectiveness of the deep integration of multimodal large models into teaching; carefully select technical products and establish a multimodal large model education service support system; and focus on application performance and continuously carry out research on the educational impact of multimodal large models.

Keywords: multimodal large models; educational applications; value empowerment; scenario construction; implementation strategies

(责任编辑 吕允英)