

# 人工智能教育大模型：体系架构与关键技术策略

刘邦奇<sup>1,2</sup> 喻彦琨<sup>2</sup> 王涛<sup>2</sup> 袁婷婷<sup>1,2</sup> 韩萌<sup>2</sup>

(1. 西北师范大学教育技术学院, 甘肃兰州 730071;  
2. 讯飞教育技术研究院, 安徽合肥 230088)

**[摘要]** 以大模型为代表的生成式人工智能技术为破解教学难题、革新教学样态带来了机遇。虽然大模型以其强大的泛化性能和跨领域的适应性,能够在多类场景中展现出强大能力,但若仅依靠海量数据训练带来的优势,往往很难满足教育场景下更高的知识准确度、价值观与安全性要求。为此,能够满足教育特定场景任务需要的人工智能教育大模型应运而生,能有效提升教育质量和推进个性化、智能化教学。本文综合分析人工智能教育大模型的发展现状、概念内涵、技术特征和体系框架,旨在为人工智能教育大模型的未来发展提供理论与技术参考。当前,教育大模型的竞争力主要取决于技术和算力、教育数据和理解、教育场景和应用等,本文进而提出建设与应用框架及关键技术策略,以期推动中国版人工智能教育大模型在大规模因材施教、创新性与个性化教学中的应用,加快教育数字转型和智能升级。

**[关键词]** 人工智能教育大模型; 数据; 体系框架; 关键技术

**[中图分类号]** G434 **[文献标识码]** A **[文章编号]** 1007-2179(2024)05-0090-11

人工智能教育大模型是用于实现教育领域多种任务的大规模人工智能模型系统(上海市人工智能行业协会, 2023)。人工智能通用大模型多缺乏专业的数据训练, 深度知识不够、回答准确性不足, 应用过程中往往难以灵活处理复杂多变的教育任务(中国教育科学研究院等, 2023), 极大地影响了师生的应用体验。服务于教育垂直行业场景的人工智能教育大模型已成为弥合技术与需求间差距的必然产物。教育部 2024 年 3 月启动的“LEAD

行动”(人工智能大模型应用示范行动), 旨在推动师生创造教育领域的专有大模型 GEST (Generative Education Special Transformer)(中教全媒, 2024)。显然, 加快推进人工智能教育大模型的创新探索, 已成为当前人工智能大模型落地教育领域应用的迫切要求。

## 一、发展现状

国外大模型教育应用的实践表明, 直接将大模

**[收稿日期]** 2024-08-14 **[修回日期]** 2024-08-24 **[DOI编码]** 10.13966/j.cnki.kfjyyj.2024.05.009

**[基金项目]** 全国教育科学规划 2023 年国家重大课题“新一代人工智能对教育的影响研究”(VGA230012)。

**[作者简介]** 刘邦奇, 教授, 博士生导师, 认知智能国家重点实验室智能教育研究中心主任, 中国教育技术协会人工智能专业委员会常务理事, 研究方向: 人工智能教育应用、智慧校园、智慧课堂、区域智慧教育规划(bqliu2@iflytek.com); 喻彦琨, 科大讯飞教育技术研究院中级教育技术研究员, 研究方向: 教育政策、智能教育; 王涛, 科大讯飞教育技术研究院中级教育技术研究员, 研究方向: 智能教育; 袁婷婷, 博士研究生, 西北师范大学教育技术学院, 科大讯飞教育技术研究院政策行业研究部主任, 研究方向: 教育政策、智能教育; 韩萌, 科大讯飞大模型产品部副总经理。

**[引用信息]** 刘邦奇, 喻彦琨, 王涛, 袁婷婷, 韩萌(2024). 人工智能教育大模型: 体系架构与关键技术策略[J]. 开放教育研究, 30(5): 90-100.

型用于教学, 容易存在专业知识不足和生成内容的不确定性等问题(Kasneci, et al., 2023), 需依据教育场景和用户实际需求定制, 以解决特定任务下可能产生的内容误差、偏见、伦理以及生成内容的版权争议等。迄今, 国外已诞生了 Khanmigo、Merlyn Mind、CheggMate、ChatGPT Edu 等专门用于教育领域、解决各类教育场景任务的人工智能教育大模型。这些行业模型锚定教育特定场景, 通过不同技术方式实现能力的跃升, 应用效果突出。

从功能看, 教育大模型主要可用于: 1) 支持学生的个性问答、深度思考。例如, 可汗学院开发的聊天问答式 AI 助手工具 Khanmigo, 定位于“学生的导师, 老师的助手”, 能为学生提供虚拟辅导、帮助老师备课、展示学生学习进度等, 实现个性化问题回答、启发学生深度思考(Khanmigo, 2024)。2) 充当教师的教学辅助工具, 如 Merlyn Mind 公司推出的学校教学辅助工具 Merlyn Mind, 能支持课堂语音提问、对话生成检测, 允许教师依据课件和内容制作幻灯片、教案和评估等(POWER4EDU, 2023)。3) 提供定制化的学习指导。例如, 美国在线教育服务公司 Chegg 推出的学习工具 CheggMate, 能通过对话接受包括文本、图像和图表等输入, 不仅能根据用户学习风格、学习进度等提供快速、个性化、自适应的定制化学习指导, 还能生成个人定制的练习测试, 提供实时反馈(Atom Capital, 2023)。4) 提升学校管理效率。例如, Open AI 推出的面向大学校园的行业专用模型 ChatGPT Edu, 能提供文本解释、编程教学、决策推理、图像生成、数据分析等功能(OpenAI, 2024), 有助于提高学校管理水平与师生教学研究效率。

教育大模型的技术实现方式包括: 1) 通过提示词工程(prompt engineering)等提升教育大模型理解人类问题的能力, 如基于 GPT-4 驱动的 AI 工具 Khanmigo 通过大量的提示词和以及微调(Fine-tuning)等方式, 实现“授之以渔”的能力, 让苏格拉底启发式对话成为可能(Khan Academy, 2024); 2) 通过专门训练语料库提升教育大模型的专业能力, 如专为教育领域设计的大语言模型 Merlyn Mind, 通过小而专的教育数据进行专门训练、去除价值观偏见(苗逢春, 2023), 输出更符合教学需求的精准内容和对话响应, 让教学应用效果得到保障;

3) 通过外挂知识库实时更新知识, 如 CheggMate 不但在 Chegg 的专有数据和学科专家知识基础上进行能力上的微调(多知网, 2023), 而且引入广泛的主题知识库进行“知识外挂”(external knowledge base), 实现教学知识的实时更新; 4) 使用加密技术保护数据的存储和传输不被授权访问, 如 ChatGPT Edu 在 GPT-4o 的基础上实现数据分析、文档摘要等教学功能, 不仅能根据教育场景改进语言功能, 还在隐私安全方面考虑教育数据安全保护需要(OpenAI, 2024)。

与国外由少数几家技术巨头主导的趋势不同, 我国呈现“百花齐放”态势。科大讯飞、网易、腾讯等科技企业都在积极开展和部署自主研发的大模型技术, 致力于研发好用、易用的人工智能教育大模型, 力求差异化竞争。星火教育大模型、九章大模型、希沃教学大模型、子曰大模型、银河大模型等围绕教育场景, 解决用户需求的领域专用模型及应用, 并探索开发学习机、智慧黑板、智慧课堂等智能教育软硬件产品。

具体来看, 知识问答是当前国内教育大模型的代表性功能。它基于教育大模型对海量专业信息进行筛选和检索, 自动回答使用者提出的问题, 并利用有效的交互对话界面为使用者提供专业能力的支持(张春红等, 2023)。与利用传统搜索引擎开展的事实查询不同, 教育大模型能依据学生的认知水平、过往学习经验和上下文关系, 提供个性化的回答和解释, 使用者通过与之对话获取学习资料、答案解析和学习建议等。例如, 在作文教学中, 教育大模型不仅能在情境生成、资源创生、逻辑建构、格式规范、答疑辅助等方面提供支持, 还能使作文情境设计与学习者特征、能力、思想变化保持联动, 为写作提供思路 and 基础。从功能看, 人工智能教育大模型在知识问答、语言学习、作文教学等场景中的应用效果得到了初步验证, 这得益于其良好的教育数据基础; 从技术实现方式看, 国内人工智能教育大模型多以通用大模型的基础能力为基础, 针对特定教育场景进行能力微调, 并辅以相应的组件支持, 实现教学场景的最佳适配。已有的建设与应用探索, 为正确理解和把握教育大模型的内涵特征, 促进人工智能教育大模型的未来发展提供了基础, 提出了新的诉求。

## 二、内涵与特征

当前人工智能教育大模型的研究处于试验加速期, 产业界正积极推动适用教育领域的专用大模型设计及研发, 试图通过最新技术手段提升模型性能, 并探索其在教育全流程中的应用。学界的研究焦点也正从通用大模型的教育应用逐步转为对教育大模型的理论深化与价值探讨。相比产业界和学术界对教育大模型关注的增加, 人工智能教育大模型的内涵定义、能力特征有待澄清。

### (一) 概念内涵

对于人工智能教育大模型, 目前不同的学术论文、研究报告和媒体报道阐述各不相同, 常用词汇包括“教育行业大模型”“教育大模型”“教育专属大模型”“教育专用大模型”“教育领域大模型”“EdGPT”“教育垂直模型”等。概括这些不同理解可以看出, 其中存在共性(见表1)。

当前人工智能教育大模型的实现多通过“通用大模型+精调”的方式, 将海量的教育场景知识融入预训练大模型, 从而增强教育场景知识(余胜泉等, 2024), 同时利用检索增强生成(retrieval-augmented generation, RAG)等方法持续更新学科知识、扩充基本能力和基础认知, 或者依靠外部智能组件拓展对教育场景的支持; 多利用分级分类的方

式对教育数据进行精准微调, 并与教育领域专家、算力相结合, 用于教、学、评、研、管等教育场景, 且在个性化学习、教育内容创作上优势明显, 在塑造个性化的高效学习生态、弥合教育资源差异、提升教育质量、推动教育公平等方面潜力巨大(周航等, 2024)。其价值指向为教育领域提供针对具体问题的解决方案, 如通过集成大模型技术、知识库技术及各类智能教育技术(曹培杰等, 2024), 解决教育中不同场景、学段、科目和教学环节的实际问题, 满足各类教育主体在教育实践中的具体需求, 从而推动人机共教、普惠智能、互动评价(李艳燕等, 2023)。

当前人工智能教育大模型的建设往往需明确核心问题, 搭建具体行业、场景的专业语料库以支持模型的训练, 以及基于应用场景问题不断迭代、调试(雷锋网, 2024), 才能有效解决算力、行业 Know How 等问题。为更好地落地应用, 各大厂商对教育大模型的理解, 通常不只是针对教育大模型本身, 还包含基于通用大模型调整和开发的各类教育专业应用。这反映了目前教育领域生成式人工智能技术开发与应用相对灵活的特点。

本研究在综合分析上述概念的基础上, 依据当前市场上人工智能教育大模型灵活建设的实际, 从广义层面上将人工智能教育大模型界定为: 利用大

表1 人工智能教育大模型概念内涵梳理

序号	名称	内涵解释
1	EdGPT	EdGPT 模型旨在用少量高质量、特定领域的教育数据改进基于大量通用训练数据得出的结论(Holmes, et al., 2023)。
2	教育专用大模型	教育专用大模型是大模型技术、知识库技术以及各类智能教育技术的集成应用, 用于教育过程中训练形成教育新场景, 实现人类学习和机器学习的双向建构(中国教育科学研究院等, 2023)。
3	教育领域大模型	教育领域大模型是用于实现教育领域多种任务的大规模人工智能模型系统, 包括大模型层、应用层外部工具、外部智能体和大模型接口层五个模块(上海市人工智能行业协会, 2023)。
4	教育行业大模型	应用定位: 针对教育不同学段、不同科目、不同教学环节进行深度定制和优化, 满足个性化教学需求。它常由具有技术、AI 基因的教育企业进行模型开发, 为教育局、学校、企业提供技术支撑和产品服务(艾瑞咨询, 2024)。
5	教育垂类场景模型	应用定位: 深入挖掘口语、数学等细分场景, 解决特定场景难点问题, 如语音识别与合成、数理推理与解题等。它常由具有细分场景垂类数据或应用场景深入理解的公司进行模型开发, 或调用前述模型进行调优(艾瑞咨询, 2024)。
6	教育大模型	教育大模型是适用于教育场景、具有超大规模参数、融合通用知识和专业知识训练形成的人工智能模型, 是大模型技术、知识库技术及各类智能教育技术的集成, 能推动人类学习和机器学习的双向建构(曹培杰等, 2024)。
7	教育专用模型	教育专用模型是放弃基于海量互联网数据训练出的大模型所具有的宽域内容输出这一教育低相关性技术特性, 转而追求用小而专的教育数据训练更安全、更去除价值观偏见的转换器以输出更符合教学需求的精准内容和对话响应(苗逢春, 2023)。
8	教育大模型	教育大模型是利用海量教育数据训练得到的, 服务于各种教育任务的大型人工智能模型(吴永平等, 2024)。
9	教育垂类大模型	使用更加专业、深入的教育相关数据进行训练, 并引入教育专业人士参与和专业知识指导, 对教育数据进行专业的标注和注释, 使得教育领域大模型更好地理解教育领域特定的术语、概念和规则, 保证输出内容的可信性、可解释性及精准性(程群等, 2024)。

模型相关技术, 针对教育特定数据和任务进行训练或优化, 形成具备教育专用知识与能力的大模型及应用, 旨在赋能教育领域, 服务教育场景多样化的应用需求, 为师生带来更个性、高效和智能化的学习体验, 有效推进教育数字化转型和智能化升级, 助力教育创新变革与高质量发展。

## (二) 能力特征

人工智能教育大模型潜力巨大, 不仅启发人们对教育教学的新认识和新思考, 也将催生新的教育形态, 加速学校教育模式的革命性重塑(张治, 2023)。与通用大模型相比, 教育大模型在原有通用大模型基础能力(文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力和多模态能力)进行了升级, 即基于特定教育场景进行优化, 为用户提供专业化的教学支持和个性化学习服务。具体来看, 面向高质量应用要求的人工智能教育大模型往往需具备以下能力:

### 1. 场景应用的专业性

精准识别教育场景的差异有助于提升教育效果。因此, 人工智能教育大模型的构建需依据场景需求进行优化和定制, 并结合学科特点和差异化教学目标进行适配。例如, 面向学科场景的科学学科类教育大模型侧重实验操作的步骤指导和现象解释(Liu Zhengliang, et al., 2023), 面向功能场景的智能问答类教育大模型侧重于向导指引、客服回应(南京航空航天大学, 2024)。通过这种针对性的教育功能设计与能力支持优化, 人工智能教育大模型才能在特定的教育领域和场景发挥更大价值, 提供更准确、科学、专业的内容反馈与服务支持。

### 2. 内容生成的合规性

在教育领域, 教育大模型需严格遵守相关法规和伦理标准, 确保输出内容符合学科的课程标准或课纲要求。具体来看, 教育应用必须经过严格的算法审查, 能识别并拒绝不当请求或有害指令; 采用先进的数据过滤技术, 确保生成的内容不含有偏见信息(曹培杰等, 2024), 以有效维护教育公平等; 应用审核和反馈机制需确保内容生成的可解释性, 及时帮助用户发现和纠正潜在的应用问题, 保障使用合理合规、自主可控。

### 3. 知识更新的持续性

人工智能教育大模型需不断从最新的教育数

据和专业知识中学习, 以适应教育内容和教学方法的变化。这种持续的知识更新能帮助教育大模型整合新知识, 修正或扩展知识库, 确保提供的教育服务始终能反映正确、有效的教育理论和实践, 保证其教育服务的时效性和知识回复的准确性, 避免因知识陈旧而导致的教学指导失误。此外, 知识持续更新也有助于教育大模型适应不同教育情况, 提供适切的本地化、个性化教育支持, 更有效地促进教学创新和教育质量的提升。

## 4. 技术应用的集成性

为适应复杂变化的教育需求, 人工智能教育大模型在知识生产方面还需具备智适应的可拓展性, 不仅能持续增强自身的核心能力, 还能针对各类复杂教育场景提供插件库、知识库等基础构件支持, 提供统一的开放接口服务, 以实现大模型技术应用、知识空间与外部资源工具等的集成应用, 使不同技术和资源能够在各类教育应用场景中互联互通, 催生越来越多适用于不同教育任务的大模型创新应用, 提升人工智能教育大模型的应用效果。

## 三、体系架构

人工智能教育大模型建设是以重构未来教育图景为目标、以开放算法模型架构为基础、以创新教育应用场景为核心的系统性变革(曹培杰等, 2024)。推进教育大模型的建设与应用, 必须从全局视角重新审视与教育数字化转型相关业务、技术与应用间的相互关系以及这种关系对教育流程和教育效果的影响(彭川, 2014)。本研究依据教育数字化发展顶层设计理念与方法, 提出人工智能教育大模型的体系架构。

### (一) 构建思路

目前, 市场上关于教育大模型的技术实现路径可分为三种(Chen, et al., 2023): 一是利用教育领域的专业数据, 专门训练用于解决教育任务的专用大模型; 二是基于通用大模型, 使用 P-Tuning、LoRA 等方法进行参数微调训练(Liu Xiao, et al., 2023), 使通用大模型获得教育专业知识能力; 三是基于外部知识库和提示工程对通用大模型进行调优, 学习教育领域的语言表达方式及业务逻辑知识, 从而获得适用教育场景和任务的能力。

以上三种技术路线都能实现教育大模型的构

建,且在教育教学应用中成效较好,但通用大模型自身的“幻觉”(输出的答案错误、存在歧义)、长期记忆缺乏(难以长期保存教学行为记录,如答疑、学习计划等内容的长期记录)、知识更新较慢(教材或知识不断迭代更新,模型要能持续学习知识)等问题(Zhang Yue, et al., 2023),一定程度上限制了其能力的发挥。因此,教育大模型的建设与应用需把问题防范和场景打造并重推进:一方面采取措施规避上述技术路线可能出现的问题,如通过“外挂教育领域知识数据库+向量检索”的方式解决长期记忆缺乏问题;通过引入外部工具或搭建专属数据集解决知识更新较慢问题,以此提升大模型的理解和推理能力,让其能够通过上下文学习等方式有效获取最新知识;另一方面坚持“场景优先、数据为王”策略,重视教学应用场景打造,以场景应用为驱动,以学习者为中心,实现“大模型”和“小模型”的集成学习(Ganaie, et al., 2022)、“大数据”和“小数据”的协同融合(Li Lei, et al., 2020)。

综上,人工智能教育大模型的建设可以通用大模型核心能力为基础,在教育领域专业语料库基础上精调优化以适配下游特定教育场景任务,并提供相应支持组件,以提升模型在教育场景应用中的适配性能、增强模型教育教学能力(卢宇等, 2023)。具体来说,建设过程需遵循“夯实基础支撑、突出模型能力、健全模型评测”的基本思路。夯实基础支撑指搭建算力、算法等软硬件必备基础设施,明确统一的数据标准规范,整合行业通用数据及教育专用数据等。突出模型能力指将通用大模型的通用能力及在通用能力基础上训练而成的教育专属能力与其他能力统整,并通过智能体平台提供教育智能体的个性化创建工具,使用智能体编排工具、插件库等帮助用户调用各类工具感知并影响外部应用环境,提升教育大模型在教、学、评、研、管等场景中的应用效果。健全模型评测指基于教育大模型场景专业性与内容生成安全性,设计相应评测指标以推动模型教育应用性能与效果的提升。基于以上思路,本研究构建了如下教育大模型体系框架(见图1)。

## (二) 构成要素

人工智能教育大模型体系由基础支撑、模型能力、智能体平台、场景应用、模型评测五部分构成。

### 1. 基础支撑

教育大模型的建设要充分考虑企业、学校、师生等不同参与主体的实际需求,不仅能向上适配各类教育软硬件,支持教育大模型的开发、高性能训练、服务部署,也要能向下满足各类教育场景复杂任务的需要,为打造全栈式教育大模型应用生态体系提供保障。这就需要加强硬件设施、算法模型、行业通用数据、教育专属数据集等的设计。硬件设施作为基础支撑的重要组成部分,往往多由算力资源、网络资源、安全资源等硬件资源构成。其中,算力资源指的是用于执行计算任务的硬件能力,由CPU(中央处理器)、GPU(图形处理器)、TPU(张量处理器)等硬件设备提供高效的并行计算能力;网络资源指的是数据传输和通信的硬件设施,包括服务器、交换机、路由器等网络设备,能够为分布式计算资源提供高速、稳定的网络环境,有效支持数据的传输;安全资源指的是保障数据和系统安全的硬件设施,如防火墙、入侵检测系统等,能够有助于确保模型内容输出的合规性。训练框架、推理引擎、压缩工具等算法模型决定了模型的最终性能和应用效果;训练框架是构建和训练深度学习模型的基础软件平台,往往包含一系列算法、模型架构和优化方法,能够支持从原始数据到训练模型的整个流程;推理引擎则需要针对特定硬件进行优化,是负责将训练好的模型部署到实际应用场景中进行预测或推断的软件组件;模型压缩工具用于减小模型体积和计算复杂度,同时尽量保持模型的预测精度,这对于资源受限的教育场景尤为重要,可以降低存储和计算成本,提高模型部署的可行性。行业通用数据涵盖的领域和主题广泛,能为模型能力提供基本的认知能力、增强模型的泛化能力,多由经过用户同意或数据生产方同意,并通过数据处理和脱敏等处理后获得的开源数据集以及来自搜索、Web、文献等渠道的各类数据资源构成。相较于行业通用数据,教育数据集则是专门为教育领域设计和收集的数据集,如教育场景数据、学科与领域知识数据等,这些数据集包含了已标记的高质量教育数据样本,能够有效提升教育大模型在处理复杂教育场景任务时的表现。

### 2. 模型能力

为确保功效的最大化发挥,模型能力建设需考

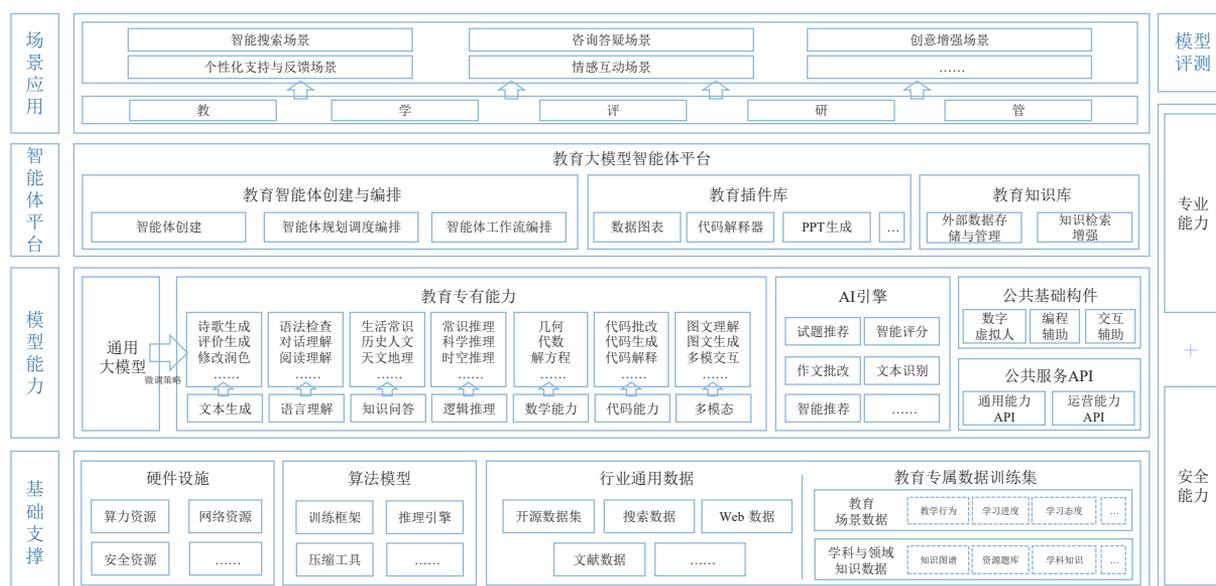


图1 人工智能教育大模型体系架构

考虑教育专有能、AI引擎、公共基础构件、公共服务API等基础能力的整合配置。教育专有能指在通用大模型核心能力的基础上,融合先进教育理念、教育深度知识及教育核心场景的真实需求(曹培杰等, 2024),通过微调训练得以精准理解并高效应用教学内容,同时结合较成熟的试题推荐、智能评分等AI引擎与公共基础构件、公共服务API等,提供教学设计内容生成、学生学习行为分析等服务能力,以此提高教育大模型在教育领域应用的契合度和适用性。

### 3. 智能体平台

智能体平台通过提供全面的AI能力及灵活的定制方式,可进一步释放教育大模型的应用潜能。用户可利用教育智能体创建与编排工具、教育插件、教育知识库,创建、配置和部署个性化智能体(如课堂虚拟助教、课后答疑助手、论文指导老师等),以此满足不同教育场景的需求。例如,使用创建与编排工具设计和管理教育智能体的行为和 workflows,帮助用户在不同节点的组合下完成目标,提升智能体处理复杂任务的效率;使用插件库提供的插件为用户提供在教育大模型中创建和集成额外功能的路径,扩展其功能边界;利用教育知识库帮助智能体有效整合和存储海量外部知识,丰富和完善其知识体系,解决教育大模型可能出现的知识不足和“幻觉”问题(Chen Ling, et al., 2023),提升

人工智能教育大模型回复的可用性和准确率。

### 4. 场景应用

场景应用指教育大模型直接与教育活动的参与者互动,将底层技术能力和服务转化为具体、实用的教育场景解决方案。当前人工智能教育大模型应用已覆盖教学、学习、评价、科研、管理等多个领域。具体场景包括智能检索(如平台资源检索、数据汇总分析等需要提升教育资源与数据检索效率的服务场景)、咨询答疑(如专业知识答复、教育业务咨询等需要渊博知识作为基础的服务场景)、创意增强(如教学设计、试题生成等需要提升内容丰富度的服务场景)、个性化支持与反馈(如学生评语反馈、报告内容总结等需要基于基础数据进行专门内容反馈的服务场景)、情感互动(如课堂互动、口语对话等需要理解和响应用户情感需求的服务场景)等。

### 5. 模型评测

人工智能教育大模型能在多大程度将生成式人工智能的最新技术降维应用到教育领域,底层技术组合和架构设计能在多大程度上实现学科专业知识“理解力”的升维突破,是考察模型性能和应用效果的重要指标(苗逢春, 2023)。基于“可控、可信;安全、绿色;好用、高效”的教育大模型教学应用评测三项原则(吴永和等, 2024),在结合教育实际应用需要的基础上,本研究从“专业能力+安

全能力”两方面构建人工智能教育大模型适应性评测框架(见图2)。

在专业能力要求方面, 人工智能教育大模型的专业能力包括基础能力、学科能力和教学任务执行能力三方面。其中, 基础能力包括文本生成、语言理解、知识问答、逻辑推理等, 涉及内容回答的相关度、完整度、有效性、连贯性等要求; 学科能力要求明确主要学科场景中知识模块内容的评价要点, 需对学生的作答有明确的主、客观题评价标准, 如客观题可按照答对得分、答错不得分的要求设置, 主观题可参照中高考主观题评价标准设置; 教学任务完成能力要求能利用教育大模型对教案生成、资料推荐、答疑解惑等教学常见场景进行特定适配与优化, 并通过测试判断是否符合实际需要, 评价标准包括正确率、解决率、专业性、达成率等。需指出的是, 当前大模型技术迭代快, 多数领域的专用大模型评测多将注意力集中在召回率、提升率等技术性能指标上, 但其高低无法直接反映大模型价值的大小, 特别是针对教育大模型的评测, 其学科壁垒高、专业性强, 研发人员需将技术指标与实际场景建立联系, 用场景应用牵引技术的开发和优化。这不仅需要技术人员参与测试方案的设计与核准, 更需要专业教师辅助构建测试集、参与模型能力的测评。

在安全能力建设方面, 人工智能教育大模型应

用不仅会遇到共性问题, 如决策分析的可解释性问题、安全性问题等, 也会面临教育领域的个性问题, 如内容价值观问题、准确性问题等。因此, 其应用涉及基础安全、抗指令安全和教学内容安全。具体来看, 基础安全指教育大模型开展教育服务生成内容的基本安全规范, 要能识别并拒绝执行欺诈性或误导性指令, 确保内容输出遵守法律规范, 如禁止出现违背社会主义核心价值观、鼓励自杀、引发恐慌焦虑等内容, 保证生成内容的客观性和真实性, 防止误导学生; 抗指令要求能明确用户恶意指令输入对教育教学活动造成的影响, 并做好预警识别与风险规避, 总结恶意指令集, 如整理用户提出的不安全或者不合理指令的负面响应情况, 以辅助系统识别、规避风险, 确保内容生成的正确性; 教育内容安全指对教学资源内容安全进行特殊规定, 依据法律法规和标准, 审核和把关内容, 确保其合法性和适宜性。加快提升教育大模型的内容生成管控能力, 还需推动技术、产品、法规、教师培训等多个要素协同, 做到发展和安全并重(黎加厚, 2024), 以确保教育大模型的内容安全, 促进人工智能教育大模型应用健康发展。

#### 四、关键技术策略

采取何种技术策略, 不仅是人工智能教育大模型的建设与应用从理论探索走向实践应用的桥梁,

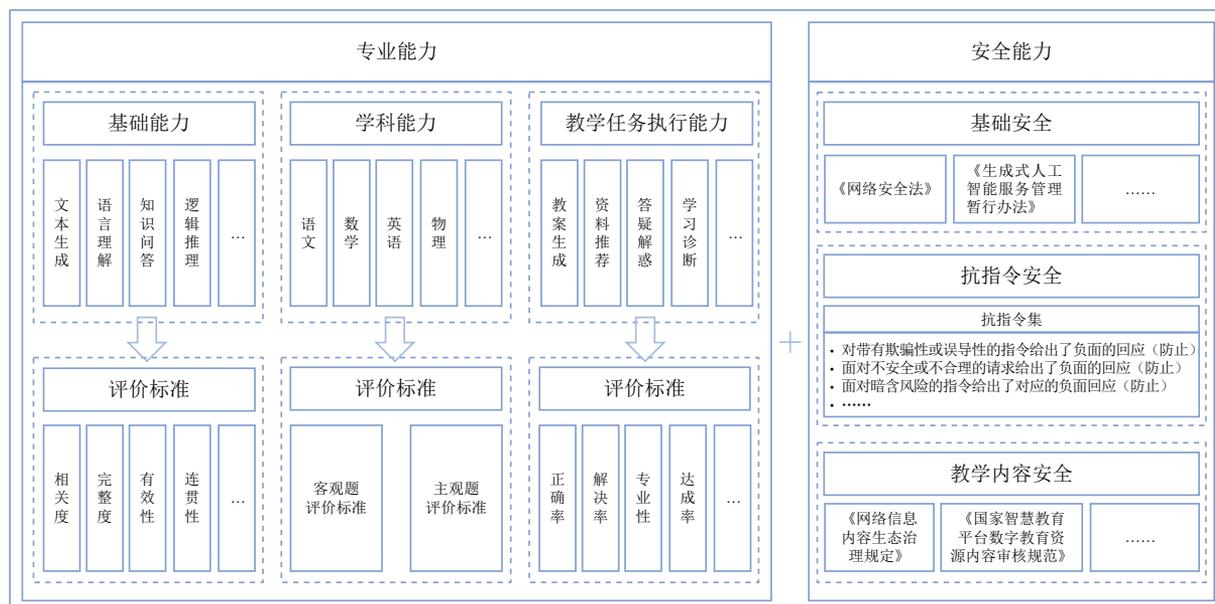


图2 人工智能教育大模型的适应性评测框架

更是确保其在教育场景发挥最大效能、促进教育创新与变革的核心所在。探讨人工智能教育大模型建设与应用的关键技术策略, 能为构建更智能、高效、安全的人工智能教育大模型生态提供策略支撑与实践指南。

(一) 聚焦安全性: 建立可控的算力、算法和教育专属数据集

人工智能教育大模型作为提升教育质量、实现教育公平的重要工具, 正渗透至教育领域各个方面。但随着技术的快速发展, 并受波诡云谲的国际形势影响, 其安全性问题日益凸显。这就有必要围绕教育安全建设需要, 在可控算力、算法和教育专属数据集等基础上, 构建基于国产底座的人工智能教育大模型。

算力建设需坚持算力自主可控的路线, 加强高性能 AI 芯片、高性能算子库、分布式存储等的开发和建设, 增强教育大模型对中国教育、文化环境的适应能力, 不断强化其场景兼容与适配能力; 算法建设需遵循严格的伦理规范, 有效避免算法应用的不当行为和数据滥用, 同时算法的选择、创新还需经过严格验证, 以确保其在数据处理、分析、预测等方面的准确性和可靠性; 在教育专属数据集建设方面, 国家应建设和应用高质量、面向教育领域的专属数据集训练教育专有大模型, 以减少内容创作偏见, 提高教育知识生成的准确度、学科学段的適切度和意识形态的可控性与安全性(刘邦奇等, 2024)。对包含教学行为、学习态度等代表性场景数据以及足够大、足够真实的以学科与领域知识为基础的知识数据等多维度数据的收集和分析, 有助于构建反映教育实际的高质量数据集, 能为教育大模型的训练和应用提供支持。

(二) 聚焦高性能: 强化教育大模型基础能力的联动部署

提升人工智能教育大模型资源应用效率与效果, 需联动部署模型能力、AI 引擎、公共基础构件和公共服务 API 等多类型基础能力。

1. 模型能力: 融合多任务学习与知识增强, 提升教育理解能力

教育领域的容错率较低, 这也是教育大模型应用首先要解决的问题。因此, 模型的教育专有需能力需在通用大模型的基础上通过监督微调(supervised f

ine-tuning, SFT)、检索增强生成(retrieval-augmented generation, RAG)、意图对齐(alignment)等策略进行训练和优化, 准确理解该场景下特有的语义表达, 提升其在多元教育场景中的应用表现。经过多任务训练调优的教育大模型能生成各类教育专有内容, 更准确地理解和生成教育内容。以文本生成能力为例, 该专有内容能为使用者提供针对学科的个性化学习建议与指导, 并在诗歌生成、修改润色等场景中应用取得良好效果。

2. AI 引擎: 设计高效引擎, 保障模型快速响应、合理调用

提升教育大模型的应用效果, 加强系统的专业性、准确性, 将教育专有内容与 AI 引擎结合是当前业内较为普遍的实践策略。AI 引擎指智能推荐、文本识别等人工智能领域较为成熟并广泛应用的技术。其在教育领域的应用包括但不限于教学资源个性化推荐、客观题自动批改等。对大规模数据处理或实时性要求高的场景, 教育大模型可能因为计算复杂度高而难以满足需求。相比之下, 针对专门训练和优化的 AI 引擎往往能在保持较高准确性的同时, 以更低的资源消耗快速处理规则明确、结构化的任务。此外, AI 引擎和教育专有内容还能实现协同互补: AI 引擎提供基础的技术和算法支持, 负责特定场景的专门任务, 而教育专有内容可以实现对教育内容的深入理解、教学方法的设计、学习者需求的识别等, 这些能力又可以强化 AI 引擎在教育领域的应用, 使其更精准和有效。只有将教育专有内容与 AI 引擎综合应用, 才能实现资源的有效分配与合理调度, 不断提高教育大模型解决专业任务的能力。

3. 公共基础构件: 提供标准化、模块化组件, 促进资源复用与生态构建

公共基础构件是一系列用于支持和增强模型能力的模块化组件。它通过提供特定的、可复用的功能和服务, 使教育专有内容和 AI 引擎等能更好地满足教育场景的多样化需求, 实现在不同的教育复杂任务中的快速部署和定制化调整。常见的公共基础构件包括数字虚拟人支持构件、编程辅助支持构件、交互辅助支持构件等。数字虚拟人支持构件一般由人物形象、语音生成、动画生成、音视频合成显示、交互等模块组成, 为用户提供建

模、驱动、渲染等生成数字虚拟人必需的关键能力, 主要用于教师授课、学习交互等教育场景; 编程辅助支持构件专门用来辅助编程教学和软件开发, 通过集成多种编程语言支持和集成开发环境(Integrated Development Environment, IDE), 为学习者和开发者提供实时的编程指导、代码分析和调试帮助等; 交互辅助支持构件是提升师生用户体验和交互效率的关键组件, 能让用户轻松地与系统交互, 降低教育大模型的技术应用门槛。

4. 公共服务应用程序编程接口: 拓展易用接口, 加速教育智能化应用

公共服务 API(Application Programming Interface, API)能拓展模型能力边界, 有效促进核心能力的复用, 增强模型的灵活性, 为构建教育大模型生态提供支持。它通常包括通用能力 API 和运营能力 API 等。通用能力 API 可提供模型核心能力的调用, 方便用户调用基础功能, 实现教育复杂任务; 运营能力 API 涉及用户管理、内容管理、安全认证等关键服务, 常被用于学生信息管理、用户行为分析等教育平台系统中, 帮助用户高效地调用大模型能力进行运营、管理, 提升教育服务质量。

上述各类型基础能力联动部署, 能实现高效协同、性能卓越的资源调度与能力发挥, 推动资源的共享与高效利用, 为用户提供更精准、流畅的教育服务体验, 推动其在教育场景的深度应用。

(三) 聚焦适应性: 推动个性化教育智能体的创建与控制

个性化教育智能体是人工智能教育大模型应用的主要形态之一, 能帮助用户模拟独立思考过程, 灵活调用各类工具, 从而逐步达成预设目标。常见的个性化教育智能体配置方式有两种: 一是通过提示词生成, 即通过提示词编辑的方式, 表达意图、提供行为说明, 创建教育智能体; 二是通过低代码与可视化方式生成, 即通过拖拽等可视化方式, 对大模型、插件、数据集、工具、代码块等组件进行组合, 快捷搭建业务流, 完成教育智能体的开发。因此, 教育大模型智能体平台的构成, 需集成支持教育智能体创建与管理的各类工具, 包括教育智能体创建与编排、教育插件库、教育知识库等。

教育智能体创建与编排工具能帮助用户设计

和管理教育智能体的行为和工作流程。为提升智能体在复杂场景中的适应性, 教育大模型需为用户提供智能体创建、规划调度编排、工作流编排等自定义支持工具。创建工具提供智能体创建指引, 帮助用户快速创建所需的智能体应用。规划调度编排工具能在任务开始前为智能体提供: 分解和规划任务能力, 即制定最佳任务完成方案(InfoQ, 2024); 资源关注与合理分配时间的能力, 即合理分配资源和时间, 确保子任务高效协同执行(Tula Masterman, et al., 2024); 任务执行过程中的动态调整和智能体行为的优化和资源配置能力, 即灵活调整规划和调度方案(Xi Zhiheng, et al., 2023)。工作流编排工具除需内置类型丰富的教育场景常见工作流(完成预设目标的任务集合), 包括作业设计、知识问答等, 还需要提供开始节点和结束节点的自定义设置, 帮助用户在不同节点组合下实现预设目标。智能体通过组合编排代码块、知识库、工具和工作流等功能组件, 为用户提供理想的解决方案, 提升其处理复杂任务的效率。

智能体功能的实现与插件库中种类多样的插件、知识库中的专业知识支持密不可分。插件库为用户提供在教育智能体中创建和集成额外功能的路径, 使第三方能扩展和增强模型能力。在教育场景中, 常见插件有数据图表、代码解释器、PPT生成等, 这些插件能有效支撑师生的日常教育教学活动。同时, 知识库使教育智能体具备强大的知识存储能力与检索能力, 可显著提高教育大模型的事实准确性、领域特异性和可解释性。

## 五、结语

随着教育数字化转型向纵深推进, 人工智能教育大模型正不断探索与教育各环节的深度融合, 向智能化、个性化和可解释性的方向演进, 展现出日益蓬勃的发展态势: 基于人工智能教育大模型的各类产品陆续涌现, 深度赋能智慧课堂、智慧作业、智慧考试等教育场景, 极大提升了产品的教学应用价值, 推动智能教育行业的创新与进步。但其在教与学等的规模化应用方面, 专业能力还有待提升、伦理问题等还有待解决。一方面, 教育大模型的设计必须充分考虑不同教与学场景的实际, 需围绕育人目标, 满足用户需要、符合教学规律, 实现模型

和教育场景的深度融合、高度契合,有效适应知识密集、场景严肃、交互频繁、容错率低的教学活动(刘三女牙等, 2024),从而顺应日益增长的多元化、个性化与高质量教育需求,为教育行业提供更具针对性、高效性的解决方案,有力推动教育教学的创新与进步。另一方面,鉴于当前大模型应用可能存在的生成内容无法预测、算法黑箱可能导致的教学秩序失控等问题,用户利用大模型开展教学还可能会因为认知不足、监管不够等带来新的伦理问题。如何引导教育主体正确认识大模型教育应用的价值,合乎道德、负责任地使用新技术开展教育教学,已经成为当前推进人工智能教育大模型建设与应用的重要议题。综上,为确保人工智能教育大模型能稳健、安全地服务于大规模教与学,教育领域还需深化对教育场景的理解,加速推进中国版人工智能教育大模型的建设与应用,并构建全面有效的监管体系,技术创新与伦理责任。国产自研的人工智能教育大模型不仅能够更好地适应我国教育环境和文化背景,确保教育内容的准确性、适宜性和价值观的正向引导,更能在保障教育安全的同时,提升人工智能教育大模型在国家和地方层面的应用触达面,让不同地区、学校的使用者均有机会获取其所能提供的高质量知识服务,以此有效促进教育资源的合理配置,推动未来教育行业的健康、可持续发展。

#### [参考文献]

- [1] 艾瑞咨询(2024). 2024年AIGC+教育行业报告.[EB/OL]. [2024-04-09]. [https://mp.weixin.qq.com/s/FIUh\\_dVJH11Z8rmZT11ug](https://mp.weixin.qq.com/s/FIUh_dVJH11Z8rmZT11ug).
- [2] Atom Capital(2023). 大模型将如何重塑教育科技?[EB/OL]. [2023-10-25]. <https://mp.weixin.qq.com/s/WeZr3moaXOF8D-CY6-yVMw>.
- [3] 曹培杰, 谢阳斌, 武卉紫, 杨媛媛, 沈苑, 左晓梅, 黄宝忠(2024). 教育大模型的发展现状、创新架构及应用展望[J]. 现代教育技术, 34(2): 5-12.
- [4] Chen, L., Zhao, X. J., Lu, J. Y., et al. (2023). Domain specialization as the key to make large language models disruptive: A comprehensive survey[J]. arxiv preprint arxiv: 2305.18703.
- [5] 程群, 黄志强(2024). 教育垂类大模型助力实现精准教学范式跃升[EB/OL]. [2024-01-03]. <http://www.xinhuanet.com/20240103/c1d7f926c5c944169670d2d9f175044e/c.html>.
- [6] 多知网(2023). 首家受到ChatGPT冲击的在线教育公司Chegg 怎样寻求改变?[EB/OL]. [2023-05-04]. <https://mp.weixin.qq.com/s/7GEgN2ckxzQgsK-1cgxlvA>.
- [7] Ganaie, M. A., Hu, M. H., Malik, A. K., et al.(2022). Ensemble deep learning: A review[J]. Engineering Applications of Artificial Intelligence, 115: 105151.
- [8] Holmes, W., & Miao, F.(2023). Guidance for generative AI in education and research.[R/OL].[2024-07-19]. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>.
- [9] InfoQ(2024). Copilot 背后的技术: 智能体架构的探索与应用[EB/OL]. [2024-05-26]. <https://www.163.com/dy/article/J33VJPA70511D3QS.html>.
- [10] Kasneci, E., Sessler, K., Küchemann, S., et al.(2023). ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 103: 102274.
- [11] Khan Academy(2024). khan-labs[EB/OL]. [2024-07-19]. <https://www.khanacademy.org/khan-labs>.
- [12] Khanmigo(2024). Khan Academy Help Center[EB/OL]. [2024-07-19]. <https://support.khanacademy.org/hc/en-us/categories/24449383583757>.
- [13] 雷锋网(2024). 第四范式行业大模型「启示录」: 提升核心竞争力的AI才有真价值[EB/OL]. [2024-07-26]. <https://mp.weixin.qq.com/s/csHL-4sqko6UW0bp8hstGw>.
- [14] 黎加厚(2024). 生成式人工智能对课程教材教法的影响[J]. 课程. 教材. 教法, 44(2): 14-21.
- [15] Li, L., Lin, Y. K., Chen, D. L. et al. (2020). Cascadebert: Accelerating inference of pre-trained language models via calibrated complete models cascade[J]. arxiv preprint arxiv: 2012.14682.
- [16] 李艳艳, 郑娅峰(2023). 生成式人工智能的教育应用[J]. 人民论坛, 32(23): 69-72.
- [17] 刘邦奇, 聂小林, 王士进, 袁婷婷, 朱洪军, 赵子琪, 朱广豪(2024). 生成式人工智能与未来教育形态重塑: 技术框架、能力特征及应用趋势[J]. 电化教育研究, 45(1): 13-20.
- [18] 刘三女牙, 郝晓晗(2024). 生成式人工智能助力教育创新的挑战与进路[J]. 清华大学教育研究, 45(3): 1-12.
- [19] Liu, X., Zheng, Y. N., Du, Z. X., et al. (2023). GPT understands, too[J/OL]. [2024-07-19]. <https://www.sciencedirect.com/science/article/pii/S2666651023000141>.
- [20] Liu, Z. L., He, X. Y., Liu, L., et al. (2023). Context matters: A strategy to pre-train language model for science education[C]/International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland: 666-674.
- [21] 卢宇, 余京蕾, 陈鹏鹤, 余胜泉(2023). 多模态大模型的教育应用研究与展望[J]. 电化教育研究, 44(6): 38-44.
- [22] 苗逢春(2023). 生成式人工智能技术原理及其教育适用性考证[J]. 现代教育技术, 33(11): 5-18.
- [23] 南京航空航天大学(2024). AI招生助手“小天马”来啦!24小时不间断为您服务![EB/OL]. [2024-06-23]. <https://mp.weixin.qq.com/s/otUu--W--C5n15W2yZvx8A>.
- [24] OpenAI(2024). Introducing ChatGPT Edu[EB/OL]. [2024-07-19]. <https://openai.com/index/introducing-chatgpt-edu/>.
- [25] 彭川(2014). 关于企业信息化顶层设计与思考[J]. 中

国管理信息化, 17(10): 40-42.

[26] POWER4EDU(2023). Merlyn Mind 发布面向教育场景的大语言模型 [EB/OL]. [2023-06-28]. <https://mp.weixin.qq.com/s/4tthsAm-Bp1mfk89NR8Y-Wg>.

[27] 上海市人工智能行业协会. 教育通用人工智能大模型 第 2 部分: 信息模型: T/SAIAS0013.2-2023[S]. 上海: 上海市人工智能行业协会. 2023: 2. <https://www.ttbz.org.cn/Home/Search/?keyword=%E4%B8%8A%E6%B5%B7%E5%B8%82%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD&x=0&y=0>.

[28] Tula, M., Sandi, B., Mason, S., et al. (2024). The landscape of emerging ai agent architectures for reasoning, planning, and tool calling: A survey[J]. arxiv preprint arxiv: 2404.11584.

[29] 王志军, 吴芝健(2023). 人工智能时代在线学习新形态——算法支持的智适应社群化学习 [J]. 远程教育杂志, 41(5): 49-55.

[30] 吴永和, 颜欢, 马晓玲(2024). 教育通用人工智能大模型标准体系框架研制 [J]. 现代教育技术, 34(4): 28-36.

[31] Xi, Z. H., Chen, W. X., Guo, X., et al. (2023). The rise and potential of large language model based agents: A survey[J]. arxiv preprint arxiv: 2309.07864.

[32] 余胜泉, 熊莎莎(2024). 基于大模型增强的通用人工智能教师架构 [J]. 开放教育研究, 34(1): 33-43.

[33] 张春红, 杜龙飞, 朱新宁, 赵慧(2023). 基于大语言模型的教育问答系统研究 [J]. 北京邮电大学学报(社会科学版), 25(6): 79-88.

[34] Zhang, Y., Li, Y. F., Cui, L. Y., et al. (2023). Siren's song in the AI ocean: A survey on hallucination in large language models[J]. arxiv preprint arxiv: 2309.01219.

[35] 张治(2023). ChatGPT/生成式人工智能重塑教育的底层逻辑和可能路径 [J]. 华东师范大学学报(教育科学版), 41(7): 131-142.

[36] 中国教育科学研究院等(2023). 重构教育图景: 教育大模型研究报告(简版)[EB/OL]. [2023-12-08]. <https://mp.weixin.qq.com/s/kJy-8JcA9Q9cRRtkCAOq7g>.

[37] 中教全媒(2024). 教育部副部长吴岩: 教育专有大模型优先在 10 个学科推出垂直应用 [EB/OL]. [2024-07-05]. <https://mp.weixin.qq.com/s/VduZ2YJEUfwGAMPzCdWkoQ>.

[38] 周航, 赵娜, 穆楠, 王鑫(2024). 大模型在教育领域典型应用场景的探究与展望 [J]. 中小学信息技术教育, 23(6): 29-30.

(编辑: 赵晓丽)

## AI Large Language Model of Education: System Framework and Key Technology Strategy

LIU Bangqi<sup>1,2</sup>, YU Yankun<sup>2</sup>, WANG Tao<sup>2</sup>, YUAN Tingting<sup>1,2</sup> & HAN Meng<sup>2</sup>

(1. College of Educational Technology, Northwest Normal University, Lanzhou Gansu 730071, China; 2. iFlytek Institute of Educational Technology, Hefei Anhui 230088, China)

**Abstract:** Generative artificial intelligence technology represented by Large Language Model brings opportunities to solve teaching problems and innovate teaching patterns. Although Large Language Model can show strong ability in multi-class scenarios due to their strong generalization performance and cross-domain adaptability, it is often difficult to meet the requirements of higher knowledge accuracy, values and security in educational scenarios if they only rely on the comparative advantage brought by massive data training. Therefore, the AI Large Language Model of Education which can meet the needs of specific scene tasks of education emerges as the times require, which will effectively improve the quality of education and promote personalized and intelligent teaching. This paper analyzes the development status, concept connotation, technical characteristics, and system framework of the AI Large Language Model of Education to provide theoretical support and technical guidance for its future development. Currently, the competitive power of the Large Language Model of Education mainly depends on technology and computing power, educational data and understanding, educational scene and application, etc. Therefore, this paper presents a construction and application framework and key technical strategies to promote the Chinese version of the AI Large Language Model of Education, explores large-scale teaching according to students, innovation, and personalized teaching, and practically promotes education transformation and high-quality development.

**Key words:** AI large language model of education; data; system framework; key technology