

# 基于大模型增强的通用人工智能教师架构

余胜泉 熊莎莎

(北京师范大学未来教育高精尖创新中心, 北京 102206)

**[摘要]** 以 ChatGPT 为代表的预训练大模型在教育界产生了巨大影响, 也为发展通用人工智能教师带来了曙光。预训练大模型应用于教育存在幻觉、深度逻辑缺失、社会情感缺失等局限, 如果这些关键问题不解决, 大模型不可能在教育中得到真正有价值的应用。本文提出通过增强大模型构建通用人工智能教师架构, 其核心思路是精调训练增强场景知识、检索增强认知、外部智能组件编排增强推理、多模态融合增强感知、情感计算增强社会情感, 再通过教育知识图谱对大模型输出进行监督。通用人工智能教师主要有六种应用场景: 需要渊博知识的场景、洞察创意增强场景、约束与管理场景、社会情感互动场景、个性化指导与反馈场景、多模态内容表现场景。文章最后分析了通用智能时代人类教师面临的机遇与挑战, 提出教师需要主动拥抱、使用、驾驭并控制智能技术, 解放、增强、进化自我, 并跨越陷阱。

**[关键词]** 人工智能教师; 生成式人工智能; 预训练大模型; 人工智能教育应用

**[中图分类号]** G451 **[文献标识码]** A **[文章编号]** 1007-2179(2024)01-0033-11

人工智能技术的教育应用, 不仅为教学活动注入新的活力, 也提高了教师的专业水平。例如, 人工智能教师可以利用机器学习、计算机视觉、自然语言处理、人机交互等技术, 分析学生学习和情感状态, 自动批改和生成作业及试卷, 辅助写作和语音答题, 模拟虚拟教师代理等。但是, 传统的人工智能教师只能针对特定的任务和数据集, 设计和训练特定的模型, 知识范围有限, 灵活性和适应性差, 互动性低。因此, 通用人工智能教师研究一直受技术的限制, 进展缓慢。2022年, 以 ChatGPT 为代表的预训练大模型在教育界引起轰动, 它们在通用语言理解和生成方面的优势, 将推动教学模式、评价方式、内容和目标等变革。预训练大模型的出现, 也为通用人工智能教师开辟了新的可能。在预训

练大模型的支持下, 人工智能教师可以担任多种角色, 如智能导师、工作助手、协作伙伴、情感顾问、虚拟学伴等。这样的人工智能教师, 可以与人类教师相媲美, 给教师的角色和能力带来挑战, 也会给教育带来了深远的影响。

## 一、预训练大模型教育应用的局限

### (一) 预训练大模型的基础原理

预训练大模型(简称“大模型”)又称基础模型(foundation models)(Bommasani et al., 2021), 它利用海量无标签语料数据库进行自监督或无监督训练, 不仅具有大规模参数, 还能适用于各种下游任务(卢宇等, 2023b)。大模型的基础网络结构沿用了自然语言处理领域较为热门的 Transformer。

**[收稿日期]** 2024-01-02 **[修回日期]** 2024-01-06 **[DOI 编码]** 10.13966/j.cnki.kfjyyj.2024.01.004

**[基金项目]** “十四五”国家重点研发计划项目“农村地区教师教学能力智能评测与教学精准辅助技术研究”(2022YFC3303600)。

**[作者简介]** 余胜泉, 教授, 博士生导师, 北京师范大学未来教育高精尖创新中心执行主任, 研究方向: 人工智能教育应用、教育大数据、移动教育与泛在学习、区域性教育信息化等(yusq@bnu.edu.cn); 熊莎莎, 博士研究生, 北京师范大学未来教育高精尖创新中心, 研究方向: 人工智能教育应用(xiongshasha@mail.bnu.edu.cn)。

**[引用信息]** 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, 30(1): 33-43.

Transformer 结构包含编码器和解码器两部分,是一种基于自注意力机制的深度神经网络模型。得益于并行计算的特性,它能更好地处理长距离依赖关系,更有效地捕捉并处理上下文信息。

以 ChatGPT 为代表的语言大模型利用多层神经网络模拟人类的学习过程,通过大量的语言数据训练,学习语言规律和知识,通过由许多节点和链接组成的复杂网络结构,存储和处理大量信息以表达词句之间的概率关系。它可以在特定场景,面对给定的文本序列,通过预测下个词或符号的概率生成与输入的文本有关联且没有语法错误的文本。大模型的构建包括两个阶段:预训练和下游任务适配。在预训练阶段,模型通过训练海量无标签语料数据,抽取数据内部隐含的知识结构和规律,获得通用的语言理解能力。此过程通常采用诸如掩码语言模型(Masked Language Modeling, MLM)等自监督学习方法。它的核心思想是随机遮蔽输入文本的一些词汇,训练模型理解和分析上下文的其他词汇,以及这些词汇与被遮蔽词汇的潜在关系,从而习得该词汇在不同语言场景概率化的关联关系。此外,大模型中多层次 Transformer 结构使模型能在不同层次建模输入数据,从而捕获局部和全局的上下文信息,使模型能对词汇特征、短语结构和整体文本语境的关系等进行多层次的抽象表示。预训练阶段后,大模型从无标签数据中学习丰富的语境信息,形成对语言、图像或其他输入数据的深层理解,具备通用的上下文理解和知识提取能力。在下游任务适配阶段,大模型沿用迁移学习的思想,基于预训练阶段学到的知识、语境信息和通用语言表示等,在少量特定领域的有标签的数据上进行有监督的训练学习,从而适应特定领域任务,如学科问答、教案生成、自动化作文评分等。两个阶段的训练策略不仅增强了模型的灵活性和泛化性,还减少了对海量有标签数据的依赖。

## (二)预训练大模型教育应用关键问题

大模型在自然语言理解、生成等方面的优异能力,使它在客服、金融、艺术、教育等垂直领域得到关注和应用。在教育领域,大模型在教师教学、学习过程、教育评价、学业辅导等方面也具有多层面的潜在应用。在教师教学方面,大模型协助教师完成教案设计与润色、教学资源生成、课程设计、个性化

问题解答等,减轻教师负担;在学习过程方面,大模型通过理解学习意图掌握学生学习需求,并提供个性化学习支架和反馈;在教育评价方面,大模型可以完成个性化作业自动评价、智能组卷等;在学业辅导方面,大模型通过多轮对话向学生提出启发性、个性化的问题,引导学生并提供解决问题的思路。

大模型在教育领域呈现出的应用前景虽然令人振奋,但关键问题也随之浮出水面。比如,因训练数据的缺失与错误、不佳利用、训练过程过于迎合人类偏好、知识过时等原因,大模型生成的内容会产生幻觉、深度逻辑缺失等现象。“幻觉”问题通常有两种:1)生成的内容形式上看似正确,但与事实不符(Huang et al., 2023a)。例如,历史教师使用大模型生成的历史资料辅助教学,如果训练数据集存在历史错误或者偏见,生成的内容就可能与历史事实不符,导致学生理解历史事件产生偏差。2)仅从文本相关性角度看待逻辑推理,不是严谨的符号逻辑推理,会出现与真实逻辑不一致的情况,当进行深度推理时,出现多步骤错误累积概率很高。由于大模型更新知识不及时,教师通过大模型获得的知识极可能与预设不一致,甚至错误。例如,教师使用大模型生成的内容设计数学教学活动,如果模型的知识截止时间较早,未能包括最新的课程内容和标准,教师就可能没有充分了解最新的教学要求和学科标准,导致教学不符合要求。

大模型的“深度逻辑缺失”问题表现为:1)大模型虽然能够生成流畅而连贯的语言,但不一定能理解和解释它所说内容的意义和逻辑,这可能会导致大模型无法有效地解决复杂或抽象的,如数学、科学或伦理等需要深度逻辑的问题;2)生成看起来推理过程合理,但结论或答案是错误的。大型模型出现“深度逻辑缺失”现象的原因有两方面:1)模型可能在专业领域知识方面学习不够。受数据训练的制约,对问题本质理解不透。因此,生成的内容表面上看似合理,实际上过于肤浅,缺乏深度的理解、分析和思考。2)大模型本身推理能力不足,无法进行深度的抽象和数理逻辑推理。例如,解决复杂数学问题,大模型时常会出现解答过程详细答案错误的情况。这是因为推理任务中,大模型不依据理解逻辑关系,而是依赖于大量文本的相关性(Ma et al., 2023),缺乏对抽象逻辑和数理逻辑推理

的深刻理解,导致生成的内容缺乏逻辑关联。

从人工智能教师的角度看,大模型存在的幻觉和深度逻辑缺失等关键问题,直接影响其教育应用价值。基于大模型的人工智能教师辅助学生时,幻觉问题可能导致向学生传递错误、偏见或不准确的信息,影响人工智能教师的可信度;深度逻辑缺失问题可能威胁学生的深度理解和思辨能力培养,降低人工智能教师的可靠性。另外,大模型还有无法理解和处理多模态信息、缺失社会情感能力等不足,比如大模型虽然能够模仿人类的情感和语气,但不是真正有感情或意识的存在。这会导致大模型无法理解和关心学生的感受,让学生感到缺乏亲切感。如果不解决这些关键问题,大模型不可能得到真正有价值的教育应用。

## 二、基于大模型增强的通用人工智能教师架构

人工智能教师指基于人工智能技术辅助、协作或独立实现教育教学活动的智能系统或机器人,它能够在没有人类教师的参与下,代替部分人类教师繁琐、机械的工作,也能够与人类教师共同参与教育教学活动,互相补充和支持,提高教学效率和质量。人工智能教师可以根据学生的特点和需求,提供个性化、适应性、交互性的教学服务,可以在教育实践中扮演自动出题和自动批阅作业的助教、学习障碍自动诊断与反馈的分析师、问题解决能力测评的素质提升教练、学生心理素质测评与改进的辅导员、体质健康监测与提升的保健医生、反馈综合素质评价报告的班主任、个性化智能教学的指导顾问、学生个性化问题解决的智能导师、学生发展的生涯规划师、精准教研中的互助同伴、个性化学习内容的生成与汇聚的智能代理、数据驱

动的教育决策助手等多种角色(余胜泉,2018)。

大模型要实现人工智能教师的功能,必须具备几方面能力:1)理解学生的语言、问题、需求和兴趣,以及用适当的方式回答和引导学生,能尊重和适应学生多元和个性化的学习需求和特点;2)有足够的知识和技能教授学生相关主题的知识,能用适合学生特点的教学方法展示知识与深度逻辑,引导学生发展创造力、批判性思维、社会责任感等;3)具有较丰富的社会情感能力,能体现和传递人的情感和品质,如尊重、关爱、正义、诚信等,引导和支持学生的情感和人格的发展;4)遵守安全和道德原则,如不提供或创造可能对学生或他人造成物理、情感或社会伤害的信息或内容,不涉及争议或敏感话题,不泄露或猜测学生或他人的个人信息等。

因此,基于大模型构建人工智能教师,不仅需要从知识、认知、推理、感知、情感等方面增强大模型,还必须确保大模型输出内容的科学性、逻辑性和价值对齐性。增强的大模型能为人工智能教师注入与人类教师相比肩的功能和能力,能像人类教师一样具有情感、能解决不同类型问题(见图1)。在图1中,人工智能教师的具身形态多种多样,可以是机器人、数字人、移动APP、应用软件等。这些形态增加了人工智能教师与学习者的互动途径,使人工智能教师能适应不同的教学环境和学习者需求,并通过多模态交互接口,有效地将信息传递给大模型。多模态交互接口方便人工智能教师接收和理解学习者的语言、表情和手势等多种形式的反馈,还能使教师以视觉、听觉和触觉等方式输出信息,创建互动性和沉浸感更强的学习环境。

教育知识图谱在此架构中用来指导和约束大模型的输出,提高其准确性、可靠性和科学性,实现隐含价值观与国家意识形态对齐,扮演着把关人

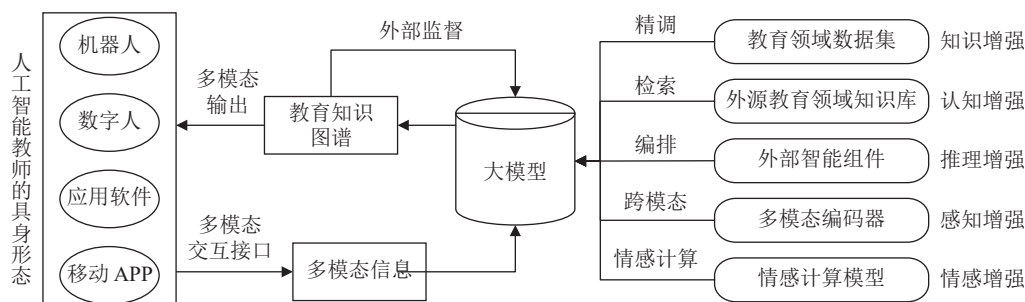


图1 基于大模型的通用人工智能教师架构

的角色,发挥外部监督反馈作用。教育知识图谱包含两类信息:一是教育领域真实的、确切的知识与信息;另一类是应当避免的负面、不良的知识和信息。如果大模型的输出符合第一类知识且不符合第二类知识,那么输出将会被采用。如果大模型的输出违背真实性或者包含不良负面信息,则会被禁止输出,同时形成外部监督反馈,指导模型的调整和改进。利用教育知识图谱的检索与推理监督大模型输出的形式有三种:1)知识蒸馏。将教育知识图谱中的知识转化为大模型可以理解和利用的形式,如将知识图谱的三元组(实体—关系—实体)转化为自然语言的句子,然后用这些句子训练或微调大模型,使其能够学习和记忆知识图谱中的知识。2)知识注入。大模型在生成过程中,动态地将知识图谱中的知识嵌入大模型的输出中,如大模型生成文本时,根据文本的主题和上下文,从知识图谱中检索和选择相关的知识,然后将这些知识以合适的方式嵌入文本中,使其更加丰富和有说服力。3)知识对齐。大模型输出后,用知识图谱中的知识验证和纠正大模型的输出,如大模型生成文本后,根据文本中的实体和关系,从知识图谱中查询和比较相应的知识,发现和修改文本中的错误或不一致处,使其变得正确和一致。

架构的核心是增强大模型部分。它通过“大模型+精调”范式,将海量的教育场景知识融入预训练大模型,增强大模型教育场景知识;采用检索增强方法更新大模型的知识、扩充能力和认知。同时,它通过编排外部智能组件增强大模型的逻辑推理能力,这些组件可能包括专门的逻辑推理算法引擎。多模态支持允许大模型处理并理解文本、图像、语音等形式的输入数据,增强其感知能力。最后,情感计算赋予大模型解读和响应学习者情感需求的能力,以期实现更人性化的教学交互。

#### (一)增强知识:精调训练增强教育场景知识

“大模型+精调(fine-tuning)”范式是解决将大模型应用不同教育场景,提升大模型对教育领域场景知识理解的有效路径。该范式通过“喂养”教育领域大规模知识库和数据,例如数字教材、数字教辅、数字教学资源、教学场景数据、在线学习数据等,形成教育领域通用大模型;然后再把它和不同的下游教育任务(如教案生成)适配精调,从而

得到多类教育大模型。与全参数微调不同,精调不会训练模型的全部参数,仅微调模型的小部分参数。精调的性能在与全参数微调媲美的同时,还降低了计算成本(Ding et al., 2023)。

精调有三类方法:1)增量式方法。该方法的核心思想是在原始模型中引入额外的可训练神经网络模块或参数,它通过微调这一小部分参数达到模型的高效适配。例如,Prefix-tuning(Li et al., 2021)和 Prompt-tuning(Liu et al., 2022)将可调向量插入输入层或表示层进行微调,Adapter-tuning(Houlsby et al., 2019)仅在每层添加简单的神经网络。2)指定式方法。该方法指定原始模型的某些特定参数变得可训练,而冻结其他参数。这一方法操作简单,但效果很好。例如,BitFit(Ben Zaken et al., 2021)方法仅优化模型内部的偏置并冻结其他参数,在多个基准测试中性能超过95%的全参数微调。3)重参数化方法。它可以在不改变大模型网络结构的情况下,通过重新调整网络参数的分布,提高大模型的泛化能力和鲁棒性。

“大模型+精调”范式的教育应用有两种技术路线(见图2):一是调用大规模通用数据预训练好的大模型,再根据不同的教育任务精调;二是直接用教育专业数据专门训练解决教育任务的大模型,再进行精调。两种技术路线都能增强大模型对教育场景知识的理解。不同的是,后者对计算资源、教育数据规模要求较高。目前,已推出的教育大模型多采用第一种技术方案,即以通用大模型为基础,再用教育领域专业语料库进行精调优化以适配特

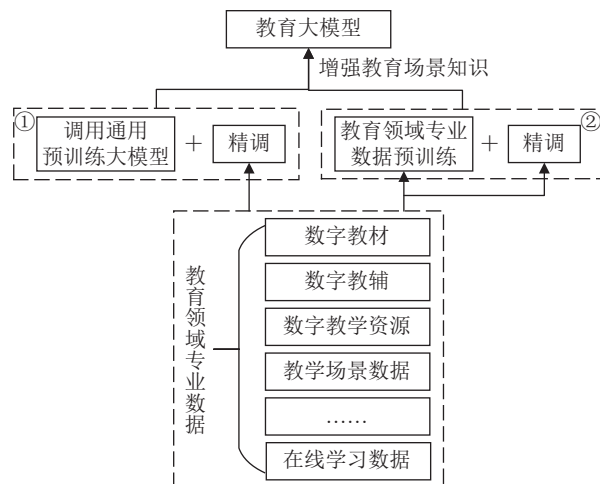


图2 “大模型+精调”技术路线

定的下游教育任务。

(二)增强认知:检索增强确保科学正确

大模型的生成内容通常依赖于训练过的数据,对没有训练的知识(如最新知识、领域特定知识、专有数据),大模型也许会给出与事实不符、虚构或者编造的内容,产生幻觉问题,无法满足实际需求。为解决这一问题,检索增强生成(Retrieval Augmented Generation, RAG)技术应运而生。检索增强生成技术由 Meta AI 提出(Lewis et al., 2020),指将结构化存储的外部知识库与带有参数记忆的大模型结合起来,使大模型能够通过外部知识源获得额外的知识扩充,增强模型对背景知识的综合理解,提高生成内容的可信度与准确度。

检索增强生成包括三个阶段(见图 3):检索、增强和生成。1)检索阶段指用户输入问题与外部知识源的检索匹配,获取附加上下文。这一阶段通常会使用嵌入模型,将用户提问内容和外部知识库嵌入同一向量空间,并利用最近邻搜索等相似度计算方法,返回用户输入与知识库中最相似的结果,作为附加上下文。2)增强阶段主要将上一阶段附加的上下文,与角色扮演设计、用户提问等内容填充到提示模板中,从而得到增强知识的提示。3)生成阶段是将增强知识的提示输入大模型,生成更科学准确的回答。检索增强生成过程的外部资

源可以是教育专家构建的知识库、学科知识图谱、优质教学场景数据等。例如,杨等(Yang et al., 2022)构建高质量的数学单词问题数据集作为外部知识库,然后检索高度相关的代数知识,并将其作为改进的语义表示输入模型,提高答案的准确率。

外部知识源的继承,创建了动态数据环境:一是保证数据新鲜度,以应对需要知识更新的场景。大模型学习的知识依赖于训练数据,若面对知识时常更新的场景,如教师要了解教育政策变化,检索增强提供的外部知识源可以动态扩充相关知识。因此,基于检索增强的大模型尤其适用于知识密集型任务,如课程设计、教学设计等。二是提供来源验证,增强可解释性和科学性。用基于检索的相关知识进行生成,生成结果可与信息来源建立联系,生成结果的参考来源可追溯,可解释性也得到了增强。三是增强领域专业化知识,提高生成内容的准确性。对于需要深度领域知识支持的场景,基于检索增强的大模型可在外部知识源的帮助下,提供丰富且专业的背景信息支持决策,提高生成内容的准确性,且计算成本相对精调方法更低。四是避免数据泄露。用真实教育教学场景数据(如学生在线学习交互日志)训练大模型,也许会产生隐私泄露问题。检索增强生成方法应检索已经脱敏或处理过的信息,避免大模型通过训练记住这些敏感数据,

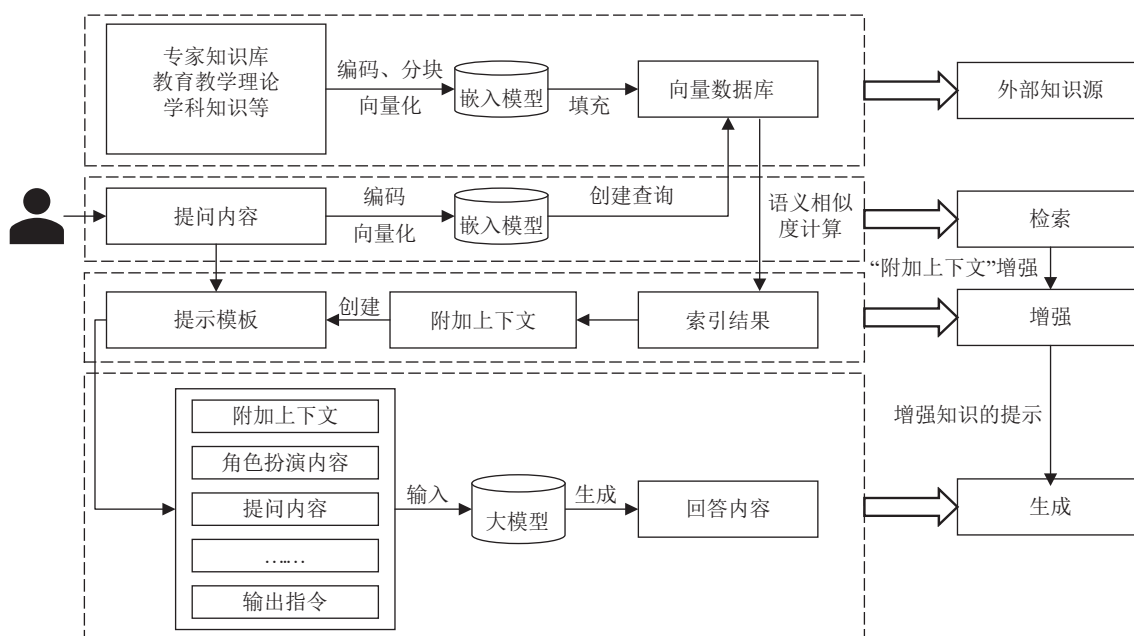


图3 检索增强实践路径

避免数据泄露风险。

(三)增强推理:外部智能组件编排,增强逻辑推理能力

像 ChatGPT 这样的大模型在自然语言理解和生成方面展现出卓越的能力,尤其是在处理“类人范畴”任务上表现突出,因为这类任务往往缺乏明确定义的“正确答案”。然而,结构化的计算任务需要模型准确操作,而大模型依据概率化生成的原理通常显得力不从心,效果时好时坏。例如,当人们期待大模型解答复杂数学推理问题时,可能出现它给出的解题步骤看似没有问题,但答案错误的现象。这是因为大模型本身的推理能力不足,看似学到很多通用知识,但没有理解本质,生成的内容缺乏深度逻辑关联。利用外部智能组件弥补大模型推理方面的不足,不失为有效的解决路径。

大模型可以与专门设计的外部智能组件进行编排组合,以增强其逻辑推理能力。大模型将复杂推理问题传递给外部智能组件进行处理,外部智能组件作为推理引擎,提供形式化的逻辑规则和推理算法,并将结果返回给大模型,最后由大模型决策生成结果。如智能知识引擎 Wolfram Alpha,内置了庞大的专业知识库和算法,可以解决数学、科学技术、日常生活等 1000 多个领域的问题,可以处理画函数图像、求解方程式等复杂问题和任务。大模型与 Wolfram Alpha 智能组件结合,将推理部分交给专业的 Wolfram Alpha,利用其内置的计算知识能力将自然语言转变为精确的符号化计算语言,再转为自然语言或图形化方式呈现,以计算的方式辅助增强大模型的推理能力。以智能辅导场景为例(见图 4),当学生咨询带情境的逻辑性问题,大模型会先尝试理解和分析问题,如发现这个问题需要精确的数理逻辑推理,就为 Wolfram Alpha 制定查询,即将学生提问发送到 Wolfram Alpha 进行搜索计算。Wolfram Alpha 先对学生的提问进行自然语言处理,转化为 Wolfram Alpha 能够理解的 Wolfram 语言格式,接着将处理过的信息和其内置的专业知识库对比、链接,之后用内置的大量算法执行必要推理计算,最后结构化展示结果,并返回给大模型。大模型基于返回的结果和学生的提问,进行决策生成并以自然语言或图表等方式反馈给学生详尽、直观、准确的答案。

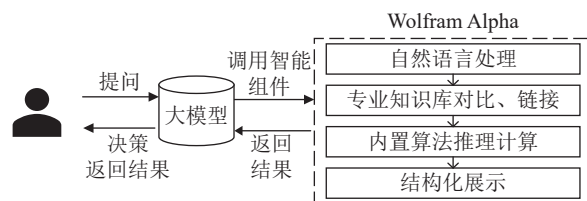


图 4 大模型结合外部智能组件实践案例

(四)增强感知:融合多模态内容理解、生成与交互  
教学场景涉及文字、图像、语音等多模态信息,其多样性和复杂性要求大模型不仅能够理解文本信息,还要考虑图像中的场景、学生的语音反馈等数据。如果大模型无法全面把握这种多样性,它处理复杂教育场景任务或问题,就可能理解不足从而产生生成内容不佳、无关或混乱的问题。这是不可接受的。多模态大模型的引入成为其解决方案。多模态大模型指包含文本、图片、视音频等多种模态的大模型(卢宇等, 2023b)。多模态的引入,使大模型有更全面的感知、更丰富的生成和更灵活的交互能力,丰富了大模型教育场景的适用性。

多模态大模型整合和解析来自视听觉等感知通道的数据,通过跨模态学习技术,学习这些多模态数据之间的复杂映射关系,使不同感知通道的信息相互补充,提高大模型对教育场景的综合理解。另外,多模态大模型能理解和处理学生阅读文本、观看图像、聆听讲座等多模态输入数据的上下文信息,更好地理解学生意图和需求。其次,多模态大模型能统合不同模态的信息,生成更有创意、生动的教学资源。最后,基于多模态大模型的“数字人”增强了教育场景的交互体验。

多模态引入增强大模型感知的技术路线有三类(见图 5):第一类是直接使用多模态数据训练大模型,把它预训练后再适配下游多模态理解、生成

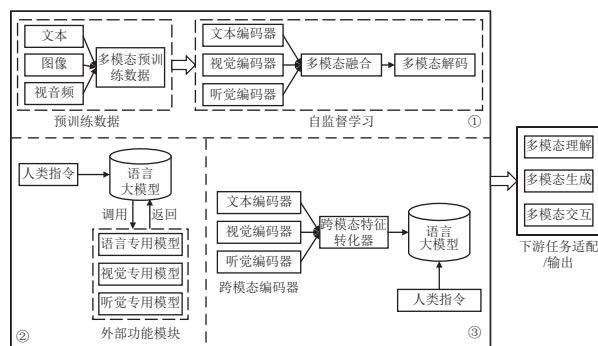


图 5 多模态大模型实践路径

和交互任务。开展自监督学习,通常使用文本、视觉、听觉等感知通道编码器对多模态数据编码,完成多模态融合后再进行多模态解码。这类多模态大模型表现优异,但训练成本较高。第二类是以语言大模型为基底,通过人类指令调用语言、视觉、听觉专用模型等外部功能模块,再由语言大模型综合完成不同的任务(刘学博等, 2023)。这类多模态大模型增强了对多模态数据的理解,但要调用多个外部功能模块,效率可能降低且成本高。第三类是将多模态数据分别编码后,通过跨模态特征转化器转化成语言大模型的输入模态,再结合人类指令,实现更广义的多模态理解。

(五)增强情感:情感计算大模型驱动,增强情感

大模型的情感对话能力相对较弱,缺乏社会情感和交流等能力,也难以与人类建立深层次的情感链接,而人的情感又复杂,导致大模型可能在理解和回应学生真实情感需求方面出现偏差,生成内容缺乏人性的温暖,限制了人工智能教师与学生的深层次互动。解决这一问题,需要提高增强大模型的情感理解与表达能力。

研究者推出了蕴含情感计算的大模型。东北大学某课题研发了面向中文心理健康支持问答的大模型 PICA(Zhang, 2023),作为情感陪护助手给用户情感支持。Inflection AI 推出了情感聊天机器

人 Pi(personal intelligence)(Inflection AI, 2023),该机器人具备“高情商”,可以了解并安抚用户的情感需求,成为值得信赖的“倾听者”。图6是大模型增强情感的技术路径,它先采集学生的语音、文本、生理信号、行为等情感信号数据,构建多模态情感信号语料库;然后对这些数据进行一系列的特征工程处理,包括提取各类数据特征,再将提取的特征进行融合;接着进行自监督学习预训练,形成具备通用情感理解能力的通用多模态大模型;最后再用有标签的特定情感任务数据集精调通用多模态大模型,形成适配情感任务的情感计算大模型。

三、通用人工智能教师的应用场景

大模型等通用智能技术的进步使人工智能的潜力得到逐步释放,未来的教育将由教师和人工智能共同协作和共存,教育将步入人机融合的新时代,人工智能教师和人类教师将互相补充,共同提高教育服务的质量(余胜泉, 2018)。在以下情境中,通用人工智能教师将展现其独有的优势。

(一)需要渊博知识的场景

在大规模教育领域语料库的精调训练,以及及时更新的外部教育专业知识库等的增强下,基于大模型的通用人工智能教师对教育知识有着更深的理解,且拥有根据学生个性特征进行适应性指导

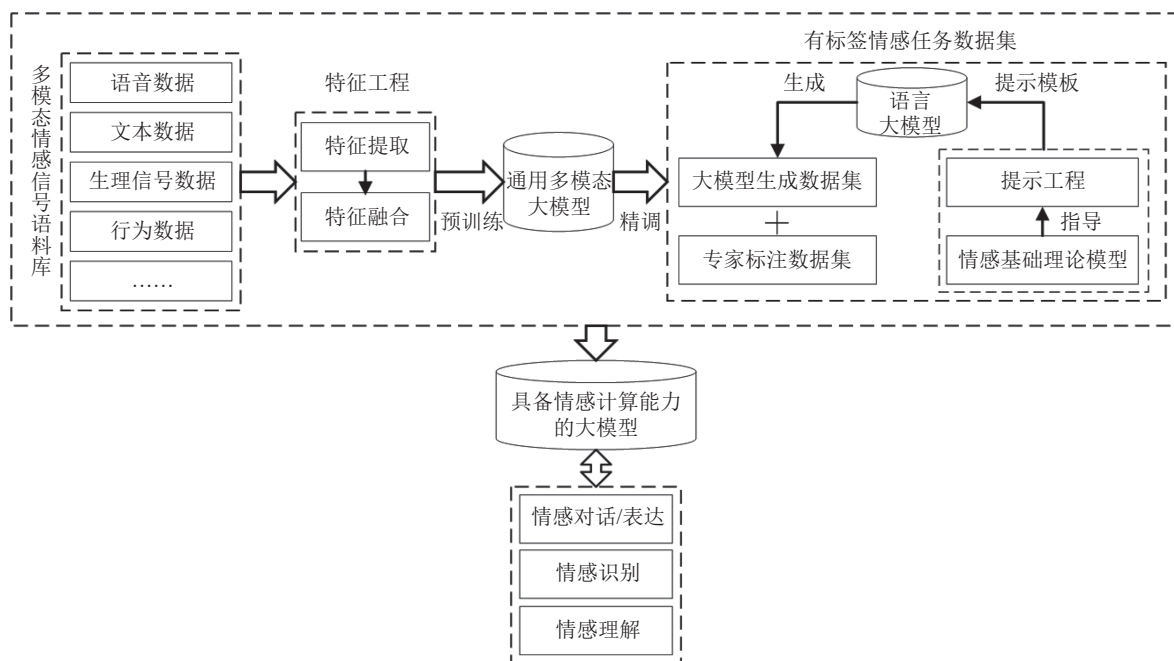


图6 大模型增强情感技术路径

的教学方法。一方面,通用人工智能教师可以是帮助学生答疑解惑的智能导师(余胜泉,2018)。在对学生进行一对一辅导、学科知识问答等情景中,通用人工智能教师可利用其深入的语言理解、上下文感知能力等,精准把握和理解学生学习意图,深度分析学生的提问后给出及时、引导性且翔实的回答。通用人工智能教师基于丰富海量的教育知识库向学生提供所需的事实、分析、讲解、故事信息等多层次服务,能激发学生好奇心和动机,帮助学生探索新知。通用人工智能教师也可以智者的方式,采用问题导向式、启发式教学方法,向学生提出开放性、启发性问题,激发学生对某一主题、问题或知识的兴趣,启发学生主动思考某一主题、问题或知识背后的原理、关联性和深层次意义。另一方面,通用人工智能教师也可以是教师专业发展提升的协助伙伴。基于大模型的通用人工智能教师能提供更新的教材知识、专业知识、教育研究文献、领域专家研讨内容等,帮助教师及时补充专业知识理解,了解教育领域前沿知识、教学方法、发展现状等,提升专业素养。

### (二)洞察、创意增强场景

基于大模型的通用人工智能教师可以提供对教育知识的广泛理解与迁移,尤其是不太容易建立联系的主题,如跨学科知识的理解。首先在提升洞察力方面,通用人工智能教师利用其多领域、跨学科知识储备,能整合多个领域的信息,帮助学生建立全局视角,以跨学科方式,助力学生理解问题的多个方面,建立更全面的知识体系,培养学生整体性思维。例如,探讨以“长征”为主题的课程思政教育,通用人工智能教师能巧妙地将学科教学与思政教育相结合,从历史、道德与法治、语文、体育等学科挖掘思政教育元素,帮助学生从多个方面、不同角度理解“长征”及其背后的精神,拓宽思维边界。其次在创意增强方面,通用人工智能教师可以是创意写作助手,以协同共创方式辅助学生写作。学生开展创意写作时,通用人工智能教师不仅能基于写作主题提供时效强、具有启发性的相关或跨学科资源与材料,还能与学生基于写作主题展开深度交互、进行头脑风暴、促进多样开放的创意分享(汪靖等,2023),激发学生创造性思维与灵感。通用人工智能教师也能为教师处理信息检索、知识

讲解、课程设计、课件制作等复杂性工作提供创意思路和方法(杨宗凯等,2023)。例如,备课时通用人工智能教师提供启发性思路、结构化框架以及多样化的备选方案(卢宇等,2023a),提升教师备课效率,扩充授课内容的丰富度。

### (三)约束与管理场景

通用人工智能教师能够像“班主任”一样,了解和管理学生的学习。人工智能教师能够从人类教师的育人实践中学习,并将人类教师育人中表现出的情感、协作和社交能力融入人工智能系统,提升人工智能的社会协作和情感交流能力,使人工智能模仿教师的社会化育人方式,更加智能和准确地感知和满足学生需求,与各方协力高效地解决育人问题,提高服务的水平和效率(余胜泉等,2019a)。未来,人工智能教师能利用学业数据,分析学生的学习表现、学习倾向、学习动力、学习难点、情感状态等,及时发现学习问题、心理需求、异常行为等,主动与学生开展双向互动和交流,与教师、家长和学生沟通,在自然的交互沟通中自动获取信息,并采取干预措施,在问题出现前预防和解决,以最大限度地关注和支持学生的全面发展。

通用人工智能教师具备监测和预防大模型滥用的交互感知能力。它通过持续监测学生与大模型的互动,识别并预警潜在的滥用行为,如将作业一股脑丢给大模型,不加自我思考,过度依赖大模型,甚至利用大模型进行不恰当的查询等。一旦检测到这些行为,通用人工智能教师会及时向人类教师和学生发出善意提醒和反馈,解释这些行为不利于学习和发展的原因,并提供正确使用和建议和指导。通用人工智能教师具备的这种能力,使其不仅能够理解和管理学生学习过程,还能善意地提醒和预警学生。教师也能根据反馈告知学生使用大模型等工具应当遵循的标准和道德,指导学生正确、安全地使用大模型。

### (四)社会情感互动场景

通用人工智能教师在社会情感互动场景可以担任多种角色,为学生和教师提供多方面服务。作为情感支持导师,通用人工智能教师能够理解和响应学生的情感需求,为他们提供情感诊断、陪伴和反馈,帮助他们理解和表达自我情感,提高情感表达能力,促进同伴间的情感理解和支持。作为虚拟



学习伙伴,通用人工智能教师能模拟互动学习场景,创建学习伙伴、学习助手等社会角色。作为教育智能代理,与学生进行互动、交流,帮助他们发现学习问题、理解知识,同时提供情感支持和陪伴。作为育人辅助助手,通用人工智能教师能对教育过程中存在的育人意识薄弱、育人知识欠缺、个性化辅导答疑缺乏等问题,借助人工智能技术实现经验隐性知识的显性化,为教师提供个性化辅助,帮助教师树立育人意识,掌握育人知识,提升育人能力,促进孩子健康成长(余胜泉等,2019a)。作为精准教研伙伴,通用人工智能教师能够协助教师实现同伴间的教学问题发现与互助改进,采集并分析教师在备课、听课、评课、学生成绩等多维数据,帮助教师了解自己的知识结构、教学方法和存在问题,实现精准导向的教研,提高教学质量,促进教师之间的协作与共享。

#### (五)个性化支持与反馈场景

通用人工智能教师的作用不仅限于传统的知识传授,还能依据学生的个性化需求提供定制化的学习体验。通用人工智能教师采集和分析学生学习数据,结合学生的知识背景、学习水平、学习风格、学习能力、行为表现、情感态度等,深入理解学生个性化学习情境、学习特点和学习需求;再利用这些信息,进行自动化评估和反馈,从而推荐适合学生的学习主题,提供合适的帮助学生跨越最近发展区的动态教学支架(余胜泉等,2019b),生成包含文本、图像、视频等形态的学习材料,推荐适宜的学习路径,提出合理的学习意见和提供个性化的学习反馈等(杨宗凯等,2023)。

通用人工智能教师可自动批阅作文并给予反馈。例如,通用人工智能教师批阅作文时,能自动检查用词、拼写、标点、表达和逻辑等,还能指出文中的错误,对作文进行系统性点评,提出具体的改进建议,帮助学生提高写作能力。通用人工智能教师还能通过多角色扮演,提供多维度、多视角的反馈。例如,教师教学设计时,通用人工智能教师可以扮演教育专家、同行教师、学生。作为教育专家,通用人工智能教师可以从教学目标的设定、课程结构、课程内容的深度与广度等角度提出建议;作为同行教师,通用人工智能教师可以分享相似教学情境的成功案例,提供实用的经验方法和教学技

巧;作为学生,通用人工智能教师可以提供课程内容的吸引力、理解难度、兴趣点等反馈。

#### (六)多模态内容创生与表现场景

通用人工智能教师可以有效整合和处理多种模态的教学资源,并用不同形式生成与呈现,从而提高学习材料的生动性与直观性和学生的学习兴趣。例如,讲授复杂学科概念时,通用人工智能教师可以作为辅助,生成图像和视频展示实验过程,生成文本和音频解释实验的原理步骤。通用人工智能教师还能依据学生的具体需求和学习环境,动态调整内容的呈现模态。例如,对远程学生,通用人工智能教师可以利用互动式图表或模拟实验,使其获得沉浸式、准实验室的学习交互体验。其次,通用人工智能教师可以为教师提供多模态的教学内容设计生成服务。它利用多模态大模型,自动搜索和整理与教学内容相关的资源,如教学背景知识、素材图片、微课资源、教学PPT等,为教师提供丰富的教学材料,节省教师的时间和精力。然后,它可以根据教师提供的教学内容,生成教学设计框架,包括教学内容分析、学情分析、教学目标、教学重难点、教学过程、学习评价、课后作业等结构,以及设计探究式教学、讲授式教学、协作式等模式。最后,它结合搜索到的资源,生成课件、学案、教学设计方案等的教学设计内容,使教学设计更加完整、生动和多样。在这个过程中,教师是教学内容的设计者、使用者和监督者,而通用人工智能教师是信息的处理者和呈现者,以及内容的创造者和优化者。

## 四、通用智能时代人类教师的机遇与挑战

大模型等生成式人工智能技术的出现让人们再次体验到信息技术对教育的冲击。新兴智能技术的出现,加速了教育变革,如注重高阶思维与能力培养、推动个性化学习、促进教学评价多元化(宋萑等,2023)、创新“师—生—机”三元结构教学模式等,这对人类教师来说是机遇和挑战并存。

#### (一)拥抱智能技术,解放教师

大模型等智能技术的发展,将教师从低效、重复、繁琐的知识传授、事务性工作解放出来,转移到提升自身教学能力,培养学生高阶思维、能力、素养等。积极探索使用通用人工智能教师,最大限

度减轻教师负担,增加教师服务的数量与质量,这是破解当前教育改革与发展难题的关键。增强大模型人工智能教师,能使教师的重复性、繁琐的工作得到解决,如自动评估学生作业,根据量化和质性标准,给出评价、反馈和建议,帮助学生改进学习效果;帮助部分教学环节实现自动化,如自动生成试卷、评价报告等;自动跟踪学生学习进度和成果,提供个性化的问题解答和指导,满足学生的不同学习需求等。这些功能可以使教师高效地完成工作,减轻工作负担,让教师把更多的时间放在培养学生社会情感、创造力和高阶思维等方面。

### (二)使用智能技术,增强教师

智能技术解放教师的同时,也给教师能力带来挑战,对教师的专业能力、教育教学能力等要求会更高。因此,教师可以以人机协同的理念思考如何与智能技术合作,增强自我能力。教师要主动学习智能技术,了解智能技术的原理、利弊及其基础应用等,并将其应用于日常教学中,让技术帮助教师解决以往难以通过传统方法解决的问题,如实现个性化教学、调整学习进度以符合学生需求,以及采用数据分析优化教学策略等。大模型等智能技术能帮助教师实现对学生学习行为和风格等的深度理解,精准诊断学生的学习需求,并为其提供量身定制的解决方案。

### (三)驾驭智能技术,进化教师

智能技术对教师角色也带来挑战。通用智能时代下,教师工作形态将发生巨大改变,知识性教学,大部分将由人工智能承担,教师更多的是负责学习的设计、督促、激励与陪伴,教师更多的工作是育人与学生的情感交流等(余胜泉等,2019b)。教师要学会驾驭智能技术,将智能技术真正应用到教学中。首先,智能技术的进化,对教师提出高智慧层次指导的要求。教师因此要提升智能素养、素质素养,适应智能教育场景的需求。其次,教学方式、教学模式、教学内容、教学评价要使用智能技术进行创新和调整。积极使用智能技术创新教学方式,设置以素养为导向的教学内容,加入能够促进学生深度思考和深度学习能力的活动,建立多元、质性的教育评价方式。再次,教师要利用智能技术促进专业发展。智能技术可以为教师提供专业知识、教材知识、教育研究文献等资源,扩充教

师专业知识理解,开拓教师眼界,了解教育前沿发展,提升专业素养。最后,智能技术能为教师教研提供精准辅助,分析教师教学问题与不足,并提出针对性建议。

### (四)控制智能技术,跨越陷阱

通用智能时代下,教师的机遇和挑战不是担忧被智能技术取代,而是转为人机协同的理念,更好地利用智能技术惠利教师教学。首先,教师要了解大模型等智能技术的可为与不可为。智能技术虽然可以辅助、增强教学,却不能代替深入的研究、分析和批判性思考,因此教师不能完全依赖智能技术,要认识到思维的重要性。其次,教师可以向学生传达智能技术的利弊,让学生意识到使用智能技术中深度思考的意义与作用。教师要推进核心素养导向的深度教学,帮助学生跨越人工智能应用中认知外包可能导致的教育陷阱(余胜泉等,2023)。再者,教师可以持续监督并评估大模型的使用,高度关注大模型等智能技术带来的伦理和数据隐私风险。

### [参考文献]

- [1] Ben Zaken, E., Ravfogel, S., & Goldberg, Y. (2021). Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models[J]. arXiv e-prints, arXiv: 2106.10199.
- [2] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., . . . Brunskill, E. (2021). On the opportunities and risks of foundation models[J]. arXiv preprint arXiv: 2108.07258.
- [3] Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., . . . Chen, W. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 5(3): 220-235.
- [4] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., . . . Gelly, S. (2019). Parameter-efficient transfer learning for NLP[C]. Paper presented at the International Conference on Machine Learning.
- [5] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., . . . Qin, B. (2023a). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. arXiv preprint arXiv: 2311.05232.
- [6] Inflection AI. (2023). Pi, your personal AI[EB/OL]. [2023-09-19]. <https://pi.ai/talk>.
- [7] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., . . . Rocktäschel, T. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. Advances in Neural Information Processing Systems, 33: 9459-9474.
- [8] Li, X. L., Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation[J]. arXiv preprint arXiv: 2101.00190.
- [9] Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., Tang, J. (2022). P-tuning: Prompt tuning can be comparable to fine-tuning across

scales and tasks[C]. Paper presented at the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).

[10] Ma, Y., Zhang, C., Zhu, S. C. (2023). Brain in a vat: On missing pieces towards artificial general intelligence in large language models[J]. arXiv preprint arXiv: 2307.03762.

[11] Yang, Z., Qin, J., Chen, J., Lin, L., Liang, X. (2022). Logic-solver: Towards interpretable math word problem solving with logical prompt-enhanced learning[J]. arXiv preprint arXiv: 2205.08232.

[12] Zhang, Y. Z., Liu, J.Q., Gao, Y.K., Chongyun and Wang, Ch.Y., Feng, D.L., Zhang, S., Yi, F. (2023). PICA: Unleashing The Emotional Power of Large Language Model[EB/OL]. [2023-09-11]. <https://github.com/NEU-DataMining/PICA>.

[13] 刘学博, 户保田, 陈科海, 张民(2023). 大模型关键技术与未来发展方向——从 ChatGPT 谈起 [J]. 中国科学基金, 37 (5): 758-766.

[14] 卢宇, 余京蕾, 陈鹏鹤, 李沐云(2023a). 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.

[15] 卢宇, 余京蕾, 陈鹏鹤, 余胜泉(2023b). 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.

[16] 宋萑, 林敏(2023). ChatGPT/生成式人工智能时代下教师的工作变革: 机遇、挑战与应对 [J]. 华东师范大学学报(教育科学版), 41 (7): 78-90.

[17] 汪靖, 米尔外提·卡马勒江, 杨玉芹(2023). 人机共生的复合脑: 基于生成式人工智能辅助写作教学的应用发展及模式创新 [J]. 远程教育杂志, 41 (4): 37-44.

[18] 杨宗凯, 王俊, 吴砥, 陈旭(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.

[19] 余胜泉(2018). 人工智能教师的未来角色 [J]. 开放教育研究, 24 (1): 16-28.

[20] 余胜泉, 彭燕, 卢宇(2019a). 基于人工智能的育人助理系统——“AI好老师”的体系结构与功能 [J]. 开放教育研究, 25 (1): 25-36.

[21] 余胜泉, 汪凡淙(2023). 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, 44 (12): 5-13.

[22] 余胜泉, 王琦(2019b). “AI+教师”的协作路径发展分析 [J]. 电化教育研究, 40 (4): 14-22+29.

(编辑: 赵晓丽)

## General Artificial Intelligence Teacher Architecture Based on Enhanced Pre-trained Large Models

YU Shengquan & XIONG Shasha

(Advanced Innovation Center for Future Education, Beijing Normal University,  
Beijing 102206, China)

**Abstract:** *Pre-trained large models have had a huge impact in the field of education, bringing a new dawn to general artificial intelligence teachers. There are some limitations in the current application of large pre-trained models in education such as hallucinations, a lack of deep logical reasoning, and absence of social and emotional understanding. Without addressing these key issues, large models cannot be truly valuable in its educational applications. This paper proposes the general artificial intelligence teacher architecture by enhancing large models. The core idea includes fine-tuning training to enhance scenario knowledge, retrieval-augmented generation to enhance cognition, orchestration of external intelligent components to enhance reasoning, multi-modal integration to enhance perception, and emotional computing to enhance emotional understanding. Additionally, the output of large models is supervised using an educational knowledge graph. This paper describes six application scenarios for the general artificial intelligence teacher including scenarios requiring profound knowledge, insight and creativity enhancement, constraints and management, social-emotional interaction, personalized guidance and feedback, and multi-modal content presentation. Finally, the paper analyzes the opportunities and challenges faced by human teachers in the era of general intelligence and suggests that teachers to actively embrace, use, master, and control intelligent technology, thus liberating, enhancing, and developing themselves, while overcoming pitfalls.*

**Key words:** *AI teacher; Generative artificial intelligence; pre-trained large models; application of artificial intelligence in education*