

人工智能多模态教学资源的生成与评价*

——基于 AIGC 在国际中文教育的应用

欧志刚 刘玉屏^[通讯作者] 覃可 王艺 李欣臻



(中央民族大学 国际教育学院, 北京 100081)

摘要: 人工智能生成内容 (AI Generated Content, AIGC) 掀起了新一轮智能革命, 也推动了教育行业的改革与发展。而国际中文教育数字化、智能化的发展, 对国际中文教学资源的质量提出了高要求。对此, 文章依托 SOR 理论, 整合语音合成、文本翻译、图像生成等 AIGC 技术, 构建了由需求分析、智能生成和质量管控三个模块组成的人工智能多模态教学资源生成框架。之后, 文章依托此框架开展了人工智能多模态教学资源生成实践, 并从多角度进行了评价, 结果显示: AIGC 技术生成的多模态教学资源具有较好的音频质量和图像质量; 教师对多模态教学资源应用于教学抱有乐观态度, 并认为这些资源多数达到了可用状态; 学习者对多模态教学资源的使用体验整体上偏积极, 并认为这些资源能够在多方面促进学习。文章应用 AIGC 技术生成人工智能多模态教学资源, 有助于优化国际中文教学资源建设模式, 推动国际中文教育的高质量发展。

关键词: 人工智能; AIGC; 多模态教学资源; 国际中文教育; 教育数字化

【中图分类号】G40-057 【文献标识码】A 【论文编号】1009—8097(2024)09—0037—11 【DOI】10.3969/j.issn.1009-8097.2024.09.004

近年来, 教育数字化深入推进, 人工智能技术越来越多地被应用于多模态教学资源建设, 如人工智能助力多模态学科知识的识别与推荐、人工智能创建多模态教育知识图谱等。而 AIGC 的出现掀起了新一轮智能革命, Transformer、ChatGPT、Stable Diffusion 等生成式人工智能技术为自动、高效地创建多模态教学资源提供了可能。当前, 全球学习中文、应用中文的人数持续增长, 这就对中文教学资源提出了规模化需求。但是, 现有的依靠传统手工或作坊的资源创建方法不能满足这种需求, 资源建设的方法和手段有必要进行革新^[1]。鉴于此, 为加强国际中文教学资源建设的力度, 本研究尝试应用 AIGC 技术构建人工智能多模态教学资源生成框架, 在此基础上开展人工智能多模态教学资源生成实验, 并从多个方面对生成的资源进行评价, 以期在国际中文教育创新教学资源建设模式提供技术应用思路。

一 文献综述

1 多模态资源与教学相关研究

O'Halloran 综合多位专家的观点, 将多模态定义为“综合语言、视觉图像、其他符号资源, 开展纸质、数字媒体和日常生活文本、事物、事件的理论分析与实践”^[2]。此后, Stein^[3]提出“多模态教学模式”的概念, 并尝试在教学实践中应用此模式。在教学活动中, 综合运用视觉上的图像信息和听觉上的声音信息, 能显著提升教学的质量和学生学习的能力。在国际中文教育领域, 相关研究证实音频、图像、文本等多模态资源能够刺激学生产生不同的模态感受, 这些资源在中文听力课堂中起到了积极作用^[4]。但是, 目前国际中文教育缺乏多模态教学材料, 迫切需要通过创建和整合来提供视、听、读等多种感官输入的资源, 以此帮助中文学习者提高产出和表达^[5], 而 AIGC 技术的出现为解决这一问题提供了良好契机。

2 AIGC 资源生成能力相关研究

从概念来看, AIGC 是指通过人工智能算法生成内容, 包括文本、音频、图像等。目前, AIGC 已能满足生成内容多样性、可控性的进阶需求^[6]。由于国际中文教育领域已有教材中的字、词、句子、篇章、语法等事实性知识是语言教学的核心, 本研究直接使用已有的教学内容文本, 考虑在给定教学文本情况下如何生成对应的音频和图像这一问题, 重点对语音合成、文本翻译、图像生成等相关技术进行阐述。

①语音合成技术: 指计算机根据文本内容生成语音内容, 产生与人类朗读文本一样的声音效果, 广泛应用于语音助手、自动客服、自动播放、对话机器人等场景。不同于繁琐的传统语音合成技术, 人工智能技术步入神经网络阶段后, 通过文本直接生成语音这种端到端方式即可实现语音合成。Google 发布的 Tacotron2 模型是典型的端到端架构, 减少了语言和声学模块, 只需文本序列、语音声谱图即可实现音频合成^[7]。此外, 微软基于 Uni-TTSv3 推出了神经网络文本转语音功能, 在有声小说、短视频社交等互联网场景中得到了广泛应用。

②文本翻译技术: 一般指利用计算机实现一种语言到另一种语言的文本转换, 是使用机器代替人工进行翻译的技术。在现有的人工智能技术环境下, 文本翻译过程实质上是一种信息编码、解码过程, 算法模型围绕编码器、解码器进行创新和优化。近年来, 多种注意力机制被应用于自然语言处理领域, 其中基于多头自注意力机制的 Transformer 模型表现十分突出^[8], 此模型以其优异的编码、解码能力使得端到端的文本翻译在准确性方面更进一步。

③图像生成技术: 文本生成图像的研究在 2016 年获得关注, 起初研究者借鉴的是生成对抗网络 (Generative Adversarial Network, GAN)、变分自动编码器 (Variational Auto-Encoder, VAE) 这类生成式模型的思路, 但这些模型在训练稳定性、图像多样性等方面稍显不足。自 2020 年以来, 基于稳定扩散模型 (Stable Diffusion Model) 生成图像成为研究热点。相比其他模型, 稳定扩散模型生成图像时在稳定性和真实性方面表现更好。德国慕尼黑大学提出的潜空间扩散模型 (Latent Diffusion Model) 对正向和逆向扩散过程进行了改进, 此模型采用预训练感知压缩、条件时序去噪、交叉注意力等技术^[9], 在提高图像生成效率的同时也保证了图像质量。

3 AIGC 促进国际中文教育资源建设相关研究

从整体发展水平来看, 国际中文教育尚处于智能教育的初级阶段, 需要构建多场景、多模态、多应用的数字资源库^[10]。同时, 国际中文教育资源建设需与人工智能技术相适配, 以支持探索新的教学模式和构建智慧学习环境^[11]。此外, 中国的传统文化、当代文明成就、现代化建设经验需以多模态的资源形式在世界范围内进行展示, 也要求充分应用人工智能技术拓展资源建设的渠道、拓宽资源建设的范围^[12]。AIGC 区别于传统人工智能技术的优势, 在于能够自动生成内容, 且能够推动教育资源应用从“诊断-定制-分发”转向“汇聚-加工-生成”^[13]。因此, AIGC 可作为资源内容质量进化的核心驱动力, 稳步推进资源内容的质量提升^[14]。鉴于 AIGC 具有大规模创建多模态资源的能力, 国际中文教育有必要探索利用 AIGC 促进教育资源建设的新模式。在应用 AIGC 技术开展多模态教学资源建设实践时, 首先应立足于国际中文教育学科进行需求分析, 并针对具体的教学内容和教学对象, 对各类模态资源进行精心设计, 以此保证资源的精确性和适应性^[15]。其次, 鉴于 AIGC 技术蕴含多模态知识加工的能力, 能全面、灵活地进行内容创造^[16], 因此教学资源的创建工作可以通过 AIGC 技术驱动来完成。最后, AIGC 生成资源时可能存在不符合国家法律法规、教育规范、学科标准等要求的情况, 使得生成的资源出

现一些有害信息、错误信息、重复信息等^[17]，因此对资源进行质量检视、审查也必不可少。

二 人工智能多模态教学资源生成框架

“刺激-有机体-反应”（Stimulus-Organism-Response, SOR）理论由 Mehrabian 等^[18]提出，系统地解释了个体行为发生的起因、经过、结果以及这三者之间的关系。其中，刺激（S）是指某些影响有机体心理与认知活动的因素；有机体（O）通常指个体，刺激作用于个体，将使个体产生特定的情绪和认知状态；反应（R）则是个体经过刺激后产生或表现出的行为反应。依托 SOR 理论，整合语音合成、文本翻译、图像生成等 AIGC 技术，本研究构建了包含需求分析、智能生成、质量管控三个模块的人工智能多模态教学资源生成框架，如图 1 所示。

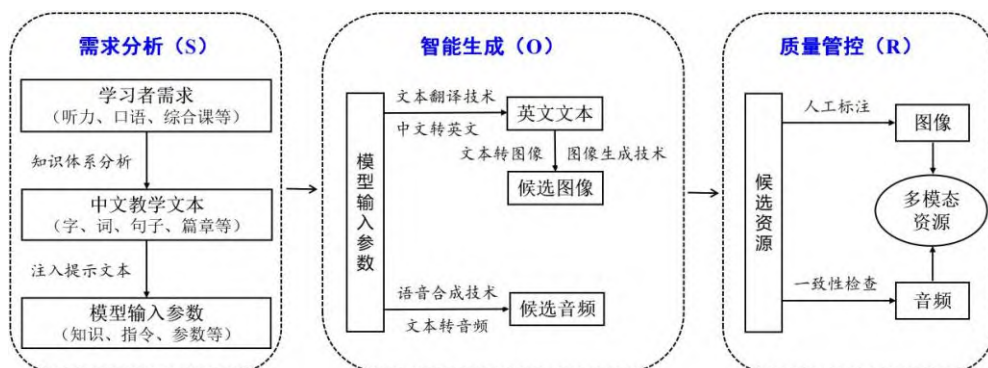


图 1 人工智能多模态教学资源生成框架

①需求分析模块：与刺激（S）相对应，是指中文教师要落实以学习者为中心的需求分析，达到听力、口语、综合课等的教学目标和教学要求。之后根据教学目标和教学要求，结合知识体系分析结果，选取或开发教学内容，获得字、词、句子、篇章等组成的中文教学文本。同时，根据 AIGC 技术中不同算法模型的要求，确定由知识、指令、参数等组成的模型输入参数。这些参数作为智能生成模块的输入，驱动模型完成相应的内容生成。

②智能生成模块：与有机体（O）相对应，是指接收刺激（S）的模型输入参数通过 AIGC 模型的生成功能完成资源创建，创建的资源主要包括候选音频和候选图像。其中，候选音频生成是应用语音合成技术实现文本转音频；候选图像生成是先应用文本翻译技术，实现输入参数从中文到英文的转换，然后应用图像生成技术实现文本转图像，创建多个候选图像。在上述处理过程中，无论是音频还是图像的生成任务都可根据需求调用云端模型或部署本地模型来完成。

③质量管控模块：与反应（R）相对应，是对生成的候选资源（包括候选音频、候选图像）进行检查和标注。在对候选音频进行一致性检查时，先通过编写专门程序对生成的音频进行语音识别，再用获得的文本内容与初始文本进行比较，以确定合成的语音与文本是否一致。而 AIGC 生成的候选图像有可能不满足要求或者出现图像与文本内容不相符的情况，这就需要对图像进行质量管控，而最佳图像常通过人工标注的方式从生成的多张图像中选出，具体做法是先基于 Web 技术开发一个标注系统，然后标注人员在此系统上浏览候选图像并从中选择最佳图像。经过对候选音频一致性的检查，可以得到每个文本的最佳发音音频，以此作为听觉模态形式的教学资源；而经过人工选择，可以得到每个文本的最佳图像，以此作为视觉模态形式的教学资源。

值得一提的是，人工智能多模态教学资源生成框架可以通过编程实现一键生成资源，具备了规模化创建教学资源的能力。尽管 AIGC 技术的发展速度较快，能够提供服务的平台也不少，但是使用者依然需要考虑一定的成本因素（如购买算力资源），这就要求根据当前的技术环境，适当选择开源、免费使用的模型进行内容生成，常用的 AIGC 技术平台如表 1 所示。

表 1 常用 AIGC 技术平台

平台名称	主要的 AIGC 技术	主要的服务方式	免费情况
Microsoft Azure	语音合成、文本翻译、ChatGPT	Web 页面、API	部分免费
百度 AI	语音合成、文本翻译、图像生成、文心一言	Web 页面、API	试用免费
OpenAI	语音合成、文本翻译、图像生成、ChatGPT	Web 页面、API	部分免费
Google	语音合成、文本翻译、图像生成	Web 页面、API	部分免费
ModelScope	语音合成、文本翻译、图像生成、大语言模型	API、预训练模型	免费
Hugging Face	语音合成、文本翻译、图像生成、大语言模型	API、预训练模型	免费
GitHub	语音合成、文本翻译、图像生成、大语言模型	预训练模型	免费

三 人工智能多模态教学资源生成实验

考虑到在全球范围内学中文、用中文的需求逐年增加，且在教学中培养学习者使用中文进行交际的能力非常重要，本研究开展了人工智能多模态教学资源生成实验，旨在生成用于口语教学的多模态教学资源，以帮助教师培养学习者对句子的理解、巩固已有字词知识，并提高学习者的中文交际能力。

1 知识体系分析

实验以国际中文教材《汉语 900 句》中的句子作为教学内容^[19]，生成音频、图像等多模态教学资源。该教材包含 900 句汉语，内容主要以话题形式呈现，围绕话题以句子形式展开对话。全书包含见面、就餐、购物、出行、住宿、理财、看病、学习、娱乐、运动 10 个主话题，每个主话题下面又包含 4~11 个子话题，其中 5 个主话题的知识体系示如表 2 所示。

表 2 《汉语 900 句》中 5 个主话题的知识体系示例

主话题	子话题	句子示例（拼音、中文、英文）
见面	打招呼、感谢与安慰、拜访、告辞、求助、约会、打电话、谈他人、谈心情、谈工作、谈爱好	Zhè shì wǒ de míng pì àn. 这是我的名片。Here is my name card.
就餐	快餐厅、中餐厅、自助餐、街头小吃、家庭聚会、宴会	Zhè dào cài zěn me zuò? 这道菜怎么做？How do you make this dish?
购物	超市、水果市场、二手市场、电话购物、服装店、商场、书店	Wǒ qù huàn diǎn lín qí á n ba. 我去换点儿零钱吧。I'll go and change some money.
出行	步行、乘公共汽车、乘出租车、乘火车、乘飞机、自驾车、旅行社、邮局	Wǒ yīng gāi zuò shén me chē qù? 我应该坐什么车去？Which bus should I take to get there?
住宿	宾馆、理发、找房、租房、搬家	Nín xiǎng zū nǎr de fáng zi? 您想租哪儿的房子？Which location do you have in mind?

从表 2 显示,学生的主要学习任务是锻炼常用话题的语句听说能力和在话题语句反映的社交场景中用语言进行交际的能力。实验选择的教材内容具有覆盖话题数量多、社交场景多样化等特征,因此简单的文本材料不利于培养学习者的中文应用能力。而 AIGC 技术生成的多模态教学资源包含音频和图像,这些资源可为学习者的语言学习提供有力支持:

①产生多通道刺激,降低语言学习认知负荷。在实际的课堂教学中,音频和图像是一种可视化的教学资源,其内容通过“视觉表征+听觉表征”的双通道来表征,使得学习者在获得最大信息量输入的同时降低认知负荷^[20]。具体而言,音频可以提高学习者对中文音节、声调等要素的学习效率,提高听力水平;图像可以让学习者对中文词语、句子获得更直观的理解。

②具象描述话题场景,加强基于知识的理解、联想和记忆。在汉语口语教学中,话题任务是内容,但学习者经常在课堂上能够学会语句、在现实场景中却想不出该说什么。而 AIGC 技术可以把语言和文化知识从抽象的文本符号转变为具象的声音和图像符号,能够贴切地展现话题内容所代表的语言交际场景,使得学习者在语言学习时能够产生兴趣、提高开口率。

③使用方式灵活,促进个性化学习和人机协作学习。多模态教学资源使用便捷,作为数字化资源构成课前、课中、课后的学习材料,学习者可以根据自己的需求反复使用资源、自行纠错、巩固所学知识,以此达到个性化学习的目的。AIGC 拥有人类无法比拟的知识储备,使用得当能够生成具有创造性、想象力的音频或图像资源,有利于培养学习者的想象力和创新思维。

2 实验过程

在开展实验时,本研究从《汉语 900 句》话题内容中随机选取 40 个句子用于音频生成和图像生成。依托人工智能多模态教学资源生成框架,经过生成实验后每个句子获得一个音频和一个图像,两者组合成一个完整的多模态教学资源,总计 40 个。

①音频生成:本研究使用微软 Azure 神经网络文本转语音服务(Azure Neural TTS),实现从中文文本到语音的生成^[21]。这需要先设置合成参数,包括选择“Yunxi (Neural)-云希”声音模版,说话风格选择“General”,角色扮演选择“Default”,语速设置为 1.05,音调设置为 1.05,输入文本获得合成音频。以“对面是个公园”文本为例,本研究使用 Praat 6.3.01 软件提取该文本对应的国际中文教师真人发音和机器发音的语图并进行对比,结果如图 2 所示。

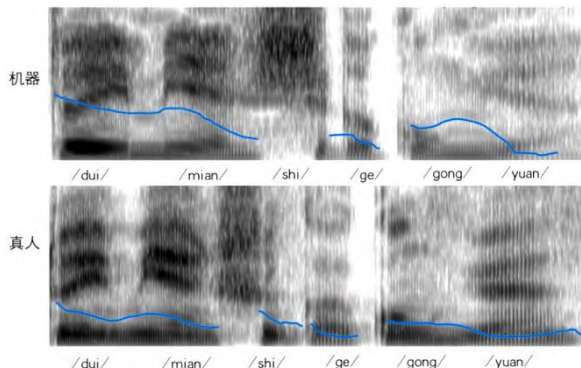


图 2 国际中文教师真人发音和机器发音的语图对比

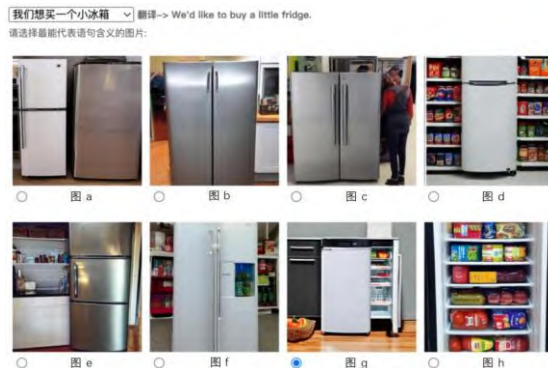


图 3 “我们想买一个小冰箱”的图像生成示例

图 2 表明,国际中文教师的真人发音和机器发音在语图表现上有较高的相似度;无论是元音、辅音的发声特点还是整体的音高走势,人机表现都有相同之处;尽管真人发声语图和机器

发声语图略有差别,但从语图展示的声学特征来看,生成的音频可以达到人类发声对文本感知与理解的效果。另外,在生成音频时,可以通过调用人工智能服务商提供的接口对生成的语音进行识别,计算获得的文本与原始文本的相似度,以此评估语音和文本的相符程度——当相似度在 90%以下时,说明生成内容存在一定的问题,需要重新生成音频,直至满足标准。

②图像生成:考虑到图像生成的效果、模型成熟度、是否免费使用等因素,本研究采用英文生成图像的间接方法,由 Hugging Face 和 GitHub 提供的预训练模型实现。首先,使用中文翻译成英文的预训练模型^[22],实现中文文本翻译成英文文本;之后,使用基于 Latent Diffusion Model 的预训练模型生成图像^[23];为了生成满意的图像文件,每个文本内容重复 8 次推理步骤,生成 8 张图像;最后进行人工判断,选取最佳图像。为了提高标注效率、实现规模化选取,本研究利用 HTML5、HTTP 接口和数据库技术,通过编程构建图像标注系统,使标注人员能以可视化的方式进行操作。以“我们想买一个小冰箱”为例,标注结果如图 3 所示,可以看出:图像生成模型通过大规模语料学习,获得了创建图像所需要的特征数据,句子关键词“小冰箱”几乎在所有图像中都得到了刻画,但大部分冰箱的展示形象不佳。由此可见,通过人工选择最符合文本内容的图像很有必要,示例中的图 g 从内容及其与文本的关联度来看应为最佳图像。通过图像标注系统,本研究实现了对全部 40 个文本对应图像的标注,并以结构化的方式存储到数据库,形成了一个多媒体教学资源库。

四 人工智能多模态教学资源评价

生成式人工智能的本质是基于概率计算创建内容,生成的内容具有不确定性,需要从多角度进行评价,如资源角度的生成质量评价、教师角度的可用性评价、学习者角度的反馈评价。对此,本研究从生成的 40 个多模态教学资源中随机选取 5 个作为评价对象,依次命名为 R1~R5,每个资源都包含文本对应的音频(A1~A5)和图像(P1~P5)。

1 生成质量的评价

(1) 音频生成评价

音频生成评价主要衡量 AIGC 生成的音频是否达到一定标准、能否被人类所感知。在实践中有很多方法进行音频质量评估,常用的是平均意见打分法(Mean Opinion Score, MOS),此方法由人类专家对音频文件进行打分,具有一定的主观性。另外,也可以通过计算机进行评分,如语音质量知觉评估法(Perceptual Evaluation of Speech Quality, PESQ),此方法是一种客观评价方法。随着计算机和人工智能技术的发展,还出现了基于深度神经网络建模的新式评估方法,即不依赖原始音频,由计算机基于算法模型自动计算分数。在采用此方法评估语音质量的解决方案中,比较典型的是 Gabriel Mittag 开发的无参考语音质量评价(Non-intrusive Speech Quality Assessment, NISQA)程序。此程序可评估的指标包括语音整体质量、噪声、音色、连贯性、响度等^[24],各指标的分值范围取 0~5 分,分值越高,说明相应指标反映的语音质量越好。由于本研究评价文本生成的音频,并没有原始音频可作对比,另外考虑到评分的客观性要求,本研究采用 NISQA 评估程序进行评价,得到 AIGC 生成的语音质量得分,如表 3 所示。表 3 显示,音频 A1、A3、A5 各指标的分值介于 4~5 分,说明这 3 条音频的语音质量较好。需注意的是,语音合成模型在生成音频时,文本音素、声调等会影响生成的音频质量。5 个音频中 A2、A4 各指标的分值较低,原因主要在于 A2 对应文本的音素数量和声调数量相对较少,而 A4 对应文本的

音素数量和声调数量相对较多。因此，如果生成的音频质量不佳，可以在不影响教学效果的前提下适当调整文本内容进行多次生成，再从这些音频中选用最佳音频。

表 3 AIGC 生成的语音质量得分

音频编号 指标	整体质量	噪声	音色	连贯性	响度
A1	4.61	4.21	4.57	4.37	4.23
A2	2.16	3.20	2.40	2.73	3.19
A3	4.44	4.50	4.47	4.02	4.40
A4	3.67	4.00	4.14	3.88	4.05
A5	4.46	4.51	4.84	4.64	4.56

(2) 图像生成评价

本研究采用图像生成技术创建图像资源，在进行评价时要注意：①通常来说，图像质量评价分为有参考和无参考两类。其中，有参考是指在有参考图像的情况下将参考图像与生成图像进行比较，主要的方法有峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）分析法、结构相似性度（Structural Similarity, SSIM）指数分析法等；无参考是指在没有原始参考图像的情况下通过算法进行图像质量评估，主要有自然图像质量评估（Natural Image Quality Evaluator, NIQE）方法、视觉信息保真度（Visual Information Fidelity, VIF）分析法等。②考虑图文对齐，即图像是否反映文本描述的内容。这通常依托训练好的模型来评估图像表示的内容是否与文本内容相匹配，常用的有 CLIP Score（Contrastive Language-Image Pre-training Score）方法。此方法主要用于评估图像标题与图像内容之间的相关度，也可用于文本生成图像的图文对齐一致性评价^[25]，评估取值范围为 0~100 分，越接近 100 分，说明图文一致性越高。

鉴于本研究是从文本到图像的生成、没有可作比较的原始图像，故本研究在评价图像质量时采用 NIQE 方法、在评价图文对齐时采用 CLIP Score 方法，得到 AIGC 生成的图像质量得分，如表 4 所示。其中，NIQE 方法是通过一定算法来计算图像的自然度、失真程度等，取值范围为 0~100 分，分值越低，说明图像质量越好。表 4 中 5 个图像的 NIQE 得分介于 17~23 之间，说明采用 NIQE 方法进行评价得到的生成图像质量得分较好。在图文对齐方面，采用 CLIP Score 方法进行评价的结果表明生成的图像与文本的相关度一般（CLIP Score 得分介于 26~32 之间），说明图文一致性较低。鉴于 CLIP Score 在不同的图文场景进行评价时存在差异，故在得分不够理想的情况下，可在具体的实践中将人工评估的方式作为补充来进行图像生成评价。

表 4 AIGC 生成的图像质量得分

评价方法 图像编号	P1	P2	P3	P4	P5
NIQE 方法	18.73	19.53	22.47	17.98	21.57
CLIP Score 方法	26.82	26.14	28.83	31.11	26.23

2 教师对多模态教学资源可用性的评价

在课堂中，教师通过融合使用多模态教学资源的方式进行教学。例如，在口语课中对于同

一个句子，教师可以播放语音让学生跟读，也可以让学生看着图像重复语音或自由发言。因此，可将语音、图像视为一个整体来进行可用性评价。本研究参考美国北卡罗来纳州公共教学部门联合其他机构开发的数字资源选择标准^[26]，编制了“国际中文人工智能多模态教学资源可用性评价”调查问卷，其维度、指标与题项设置如表 5 所示。问卷中一个指标对应设置一个题项，共 9 个题项。各题项的内容与指标含义保持一致，采用李克特五点量表计分，从“基本不同意”到“非常同意”记为 1~5 分。由于每个多模态教学资源都设有这 9 个题项，故资源 R1~R5 共设有 45 个题项。问卷整体的 Cronbach’s α 系数值为 0.973，说明问卷的信度良好。

表 5 调查问卷的维度、指标与题项设置

维度	指标	题项设置
内容	正确性	资源正确反映教学文本内容，没有错误、偏见以及可能导致混淆或误解的信息
	充足性	资源展示的信息足以反映教学文本内容
	恰当性	词汇发音和图像表达的方式适合目标受众并被他们所理解
技术	目标性	从技术来看，资源有辅助教学的明确目标，能促进知识学习，支持教学活动
	可靠性	资源使用时不需要特殊的技术支持，能可靠、符合预期地工作
	可达性	在技术手段上，资源能够无障碍地被用户包括一些生理特殊人群使用
设计	激励性	资源能产生激励效应、保持学生兴趣和提高学生参与度
	清晰性	资源的视觉和听觉元素在内容展示方面清晰、简洁、有吸引力
	可用性	资源蕴含的内容易于理解，在教学活动中容易使用

2022 年 11 月，本研究以国际中文教育公派教师、国际中文教育志愿者教师、海外本土中文教师等组成的一线国际中文教师为调查对象，开展了线上问卷调查。在填写问卷前，教师需先完成文本阅读、音频聆听、图像浏览等操作，再按照问卷题项进行资源可用性评价。调查结束后，本研究共收回问卷 108 份，去除无效问卷后，得到有效问卷 94 份。采用 Jamovi 2.3.16 对有效问卷数据进行整理、统计，本研究得到多模态教学资源可用性评价的平均得分，如表 6 所示。

表 6 多模态教学资源可用性评价的平均得分

	内容			技术			设计			均值
	正确性	充足性	恰当性	目标性	可靠性	可达性	激励性	清晰性	可用性	
R1	3.38	2.84	3.79	3.77	3.38	3.64	3.67	3.88	3.77	3.57
R2	3.46	3.62	3.68	3.76	3.41	3.47	3.72	3.83	3.72	3.63
R3	3.65	3.81	3.97	3.91	3.74	3.76	3.94	4.00	3.87	3.85
R4	3.59	3.52	3.81	3.68	3.54	3.57	3.86	3.76	3.87	3.69
R5	3.62	3.87	3.85	3.93	3.83	3.74	3.81	3.93	3.80	3.82
均值	3.54	3.53	3.82	3.81	3.58	3.64	3.80	3.88	3.81	

注：数值四舍五入。下同。

从表 6 可以看出：①资源方面，5 个多模态教学资源的均值都超过了 3.50 分，其中 R3、R5 的均值都超过了 3.80 分，说明教师普遍对 AIGC 生成的多模态教学资源应用于教学抱有乐观态度。②指标方面，“设计”维度的 3 个指标均值都超过了 3.80 分，说明教师普遍认为多模态教学

资源在激励学生的学习积极性、清晰地展现知识内容、降低资源的使用门槛等方面有积极作用；在“内容”维度，教师更看重恰当性（均值为 3.82），即多模态教学资源对文本内容的还原性是否达到应有水平；而在“技术”维度，教师更看重目标性（均值为 3.81），即资源能否加强教学内容、支持教学活动。

熵权法（Entropy Weight Method, EWM）可以按照指标的熵值进行排序，筛选出具有较好区分度的指标，具有精度高、客观性强、适用性好等特点^[27]。本研究采用熵权法计算每个指标的权重，使用熵值来描述指标的离散程度——熵值越小，离散程度越大，说明指标在评价体系中越重要。本研究对调查问卷数据进行重整，按每份问卷 5 个资源、每个资源对应 94 个样本计算，共获得 470 个样本，而每个样本包含 9 个指标值，由此形成形状为（470, 9）的样本得分矩阵。本研究首先对得分矩阵进行标准化，接着计算各指标的信息熵，然后计算各指标的权重，最后根据公式（1）计算每个资源的评价得分。其中， $k=94$ ， $a \in [0, 4]$ ， $r_0 \sim r_4$ 依次表示 R1~R5 资源的评价得分， S_{ab} 表示第 a 个资源中第 b 个样本的评价得分。基于熵权法的资源评价结果如表 7 所示，可以看出：5 个多模态教学资源可应用于教学的评价得分都超过了 3.50 分，其中 R3、R5 的评价得分都超过了 3.80 分，换算成百分值，即教师认为多模态教学资源的可用性程度超过了 70%。

$$r_a = \sum_{b=1}^k S_{ab} / k \quad \text{公式 (1)}$$

表 7 基于熵权法的资源评价结果

	R1	R2	R3	R4	R5
五度分值	3.52	3.61	3.84	3.67	3.81
百分值	70.3%	72.3%	76.7%	73.4%	76.3%

3 基于学习者感受的反馈评价

本研究依据《汉语 900 句》教材使用 5 个多模态教学资源进行教学设计并开展课堂教学实践，教学实践任务由教育部中外语言交流合作中心派至英国孔子学院的志愿者教师完成。本研究以英国某孔子学院中文水平为初中级的精品小班学习者作为实验对象，根据受访者意愿，对其中的两名学习者进行了访谈。其中，学生 A 认为文本、音频和图像融合在一起能更容易地学习词语和句子，并表示“学习新知识时图像让自己记忆深刻”“在手机上同时使用音频和图像进行学习，也是非常好的体验”“图像使学习更有趣，提高了自己联想、思考的积极性”。而学生 B 认为“图像有助于自己理解相应文本表达的含义”“看着图像说句子这样的学习方式对于培养语言能力帮助较大”。另外，两位学生都觉得 AIGC 语音比教师语音的语速更快，且节奏缺少变化，但如果语音听起来很流利就能够促进学习。总的来说，学习者对多模态教学资源的使用体验整体上偏积极，并认为这些资源能够在多方面促进学习。

五 结语

本研究聚焦于国际中文教育，构建了包含需求分析、智能生成、质量管控三个模块的人工智能多模态教学资源生成框架，开展了人工智能多模态教学资源生成实验，并从资源角度、教师角度和学习者角度进行了人工智能多模态教学资源评价。应用 AIGC 技术生成的多模态教学资源赋能国际中文教学实践，不仅能够支持教师开展教学活动，而且能够在多方面促进学习者

的学习。更重要的是, AIGC 技术的应用拓展了教学资源建设的广度和深度, 推动了国际中文教学资源数字化、智能化的转型升级, 其生成的人工智能多模态教学资源能够在优化国际中文教学资源建设模式、促进国际中文教育的高质量发展方面发挥积极作用。

参考文献

- [1]郑艳群.汉语教学资源研究的新进展与新认识[J].语言文字应用,2018,(3):106-113.
- [2]O'halloran K L. Inter-semiotic expansion of experiential meaning: Hierarchical scales and metaphor in mathematics discourse[A]. From Language to Multimodality: New Developments in the Study of Ideational Meaning[C]. London: Equinox, 2008:231-254.
- [3]Stein P. Rethinking resources: Multimodal pedagogies in the ESL classroom[J]. TESOL Quarterly, 2000,(2):333-336.
- [4]赵春辉.多模态教学在对外汉语听力课中的应用与研究[D].锦州:渤海大学,2017:30-34.
- [5]钟英华,励智,丁兰舒.“表达驱动”教学理念与国际中文教学资源建设[J].天津师范大学学报(社会科学版),2022,(6):1-7.
- [6]中国信通院.京东探索研究院.人工智能生成内容(AIGC)白皮书[OL].
<<http://www.caict.ac.cn/sytj/202209/P020220913580752910299.pdf>>
- [7]Shen J, Pang R, Weiss R J, et al. Natural TTS synthesis by conditioning wavenet on mel spectrogram predictions[OL].
<<https://arxiv.org/pdf/1712.05884.pdf>>
- [8]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[OL]. <<https://arxiv.org/pdf/1706.03762.pdf>>
- [9]Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[OL].
<<https://arxiv.org/pdf/2112.10752.pdf>>
- [10]马箭飞,梁宇,吴应辉,等.国际中文教育教学资源建设 70 年:成就与展望[J].天津师范大学学报(社会科学版),2021,(6):15-22.
- [11]梁宇,刘晶晶,李诺恩,等.内涵式发展之“内涵”:国际中文教育教学资源建设的维度[J].天津师范大学学报(社会科学版),2023,(1):38-44.
- [12]刘玉屏,欧志刚.本土化、多元化、均衡化:人工智能在国际中文教育中的应用探析[J].民族教育研究,2022,(1):162-169.
- [13]吴砥,李环,陈旭.人工智能通用大模型教育应用影响探析[J].开放教育研究,2023,(2):19-25、45.
- [14]杜修平,王崑羽,陈子尧.AIGC 赋能“中文+职业教育”资源智能生成与质量进化——内涵、机理与模式构建[J].电化教育研究,2024,(5):121-128.
- [15]王祖嫒.论美国中文沉浸式教学的多模态话语[J].民族教育研究,2016,(4):124-129.
- [16]胡泳,刘纯懿.UGC 未竟,AIGC 已来:“内容”的重溯、重思与重构[J].当代传播,2023,(5):4-14.
- [17]董莉莉,曾佳,底颖.AIGC 视域下数字教育产品的四维风险矩阵与治理框架[J].现代远程教育研究,2024,(2):12-19.
- [18]Mehrabian A, Russell J A. An approach to environmental psychology[M]. Cambridge: The MIT Press, 1974:1-266.
- [19]李朋义.汉语 900 句[M].北京:外语教学与研究出版社,2006:20-100.
- [20]郑世珏.对外汉语可视化教学方法论[M].北京:清华大学出版社,2013:20-21.
- [21]Microsoft Azure.文本转语音——真实 AI 语音生成器[OL].
<<https://azure.microsoft.com/zh-cn/products/ai-services/text-to-speech/>>
- [22]Liam168. liam168/trans-opus-mt-en-zh Hugging face[OL].

<<https://huggingface.co/liam168/trans-opus-mt-en-zh>>

[23]InvokeAI. Invoke AI - Generative AI for professional creatives[OL]. <<https://github.com/invoke-ai/InvokeAI>>

[24]Mittag G. NISQA: Speech quality and naturalness assessment[OL]. <<https://github.com/gabrielmittag/NISQA>>

[25]Hessel J, Holtzman A, Forbes M, et al. CLIP Score: A reference-free evaluation metric for image captioning[OL]. <<https://arxiv.org/pdf/2104.08718.pdf>>

[26]North Carolina Department of Public Instruction, Friday Institute at NC State University. Quality review tools for digital learning resources[OL]. <https://www-data.fi.ncsu.edu/wp-content/uploads/2021/03/03115828/content_quality_review_tools.pdf>

[27]翟雪松,朱雨萌,张紫薇,等.高校教育信息化治理能力评价:界定、实践与反思[J].开放教育研究,2021,(5):24-33.

Generation and Evaluation of Artificial Intelligence Multimodal Teaching Resources

——Based on the Application of AIGC in International Chinese Language Education

OU Zhi-Gang LIU Yu-Ping^[Corresponding Author] QIN Ke WANG Yi LI Xin-Yi

(College of International Education, Minzu University of China, Beijing, China 100081)

Abstract: Artificial intelligence generated content (AIGC) has sparked a new round of intelligence revolution and also promoted the reform and development of the education industry. The digital and intelligent development of international Chinese language education has put forward high requirements on the quality of international Chinese language teaching resources. Therefore, relied on stimulus-organism-response (SOR) theory, this paper integrated AIGC technologies such as speech synthesis, text translation, image generation, etc., and constructed the generation framework of AI multimodal teaching resources that consisted of three modules of requirements analysis, intelligent generation, and quality control. After that, based on this framework, the paper carried out the generation practice of AI multimodal teaching resources and evaluated it from multiple perspectives. The results showed that the multimodal teaching resources generated by AIGC technologies had better audio quality and image quality; teachers were optimistic about the application of multimodal teaching resources in teaching and believed that most of these resources had reached the usable state. On the whole, learners had a positive user experience with multimodal teaching resources, and regarded that these resources could facilitate learning in various aspects. The application of AIGC technologies to generate AI multimodal teaching resources can help optimize the construction model of international Chinese language teaching resources and promote the high-quality development of international Chinese language education.

Keywords: artificial intelligence; AIGC; multimodal teaching resource; international Chinese language education; education digitization

*基金项目: 本文为世界汉语教学学会全球中文教育主题学术活动计划项目“AIGC 国际中文多模态教学资源的生成与评估研究”(项目编号: SH23Y33)的阶段性研究成果。

作者简介: 欧志刚, 在读博士, 研究方向为国际汉语教学、人工智能, 邮箱为 ouzhigang@139.com。

收稿日期: 2024 年 4 月 30 日

编辑: 小方