大语言模型驱动的教学代理的动态多模态表达生成：从用户体验视角

Ninghao Wan、Jiarun Song，*IEEE*会员，Fuzheng Yang，*IEEE*会员

摘要—在虚拟现实（VR）教育场景中，教学代理（PAs）通过逼真的外观和交互行为增强沉浸式学习。然而，现有的大多数教学代理依赖静态语音和简单手势。这一局限降低了它们根据教学内容的语义上下文进行动态适应的能力。因此，交互在教学过程中往往缺乏自然度和有效性。为应对这一挑战，本研究提出一种大语言模型（LLM）驱动的多模态表达生成方法，构建语义敏感提示以生成协同的语音与手势指令，从而实现教学语义与多模态表达行为之间的动态对齐。我们开发了基于VR的教学代理原型，并通过面向用户体验的主观实验进行评估。结果表明，动态生成的多模态表达显著提升了学习者的感知学习效果、参与度和使用意愿，同时有效缓解了学习过程中的疲劳与厌倦感。此外，语音与手势的联合动态表达显著增强了学习者对类人性和社会临场感的感知。这些发现为构建更具沉浸感、表达更自然的智能教学代理提供了新的见解与设计指南。

索引词—教学代理，虚拟现实，大语言模型，多模态表达

I. 引言

虚拟现实（VR）技术将用户的感知从物理世界转移到虚拟世界，为用户带来高度沉浸的体验 [1], [2], [3], [4], [5]。作为虚拟现实环境中的关键交互元素，虚拟代理通过逼真的类人形象和交互行为，在提升用户体验方面发挥着重要作用。如今，它们已广泛应用于教育、医疗保健和专业培训等领域，从而增强沉浸感并强化社会临场感 [6]。

本研究部分得到国家自然科学基金委员会（62171353）资助。（通信作者：*Jiarun Song*）Ninghao Wan与Jiarun Song任职于西安电子科技大学通信工程学院，中国西安710071（电子邮箱：ninghaow@stu.xidian.edu.cn；jrsong@xidian.edu.cn）。Fuzheng Yang任职于中国西安电子科技大学通信工程学院，以及澳大利亚墨尔本皇家墨尔本理工大学电气与计算机工程学院，VIC 3001（电子邮箱：fzhyang@mail.xidian.edu.cn）。

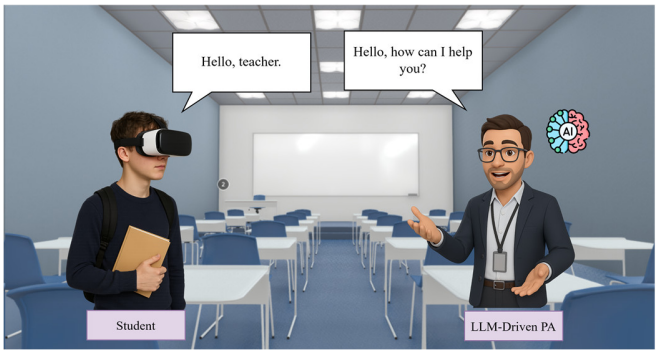


图1. VR课堂中人类学生与大语言模型驱动的虚拟代理之间对话的示意图。

在教育领域，虚拟代理通常被设计为“教学代理”（PA），以促进沉浸式、交互式的学习体验 [7]。以往研究表明，与部署在传统PC平台上的教学代理相比，基于虚拟现实的虚拟形象能够带来更强的沉浸感和临场感，从而提升学习动机和学习投入 [8]。进一步研究显示，具有类人外观的教学代理能够增强情感联结，进而提升学生的参与度和整体学习体验 [6]。除这些社会和情感影响外，对虚拟现实环境中学习者交互行为的分析还发现，面部表情和上半身动作往往能够传达潜在的认知与情感状态，为教学代理表达行为的多模态设计（如面部与手势线索）提供了有价值的参考 [9]。

随着人工智能（AI）和信号处理技术的快速发展，尤其是大语言模型（LLM）、自动语音识别（ASR）和文本转语音（TTS）技术的进步，基于VR的PA正朝着更智能、更具真实感的方向发展（如图1所示）[10], [11], [12]。依托这些技术基础，PA如今能够通过语音和文本支持多模态交互 [6], [13], ，同时展现出多样的个性特征，例如外向和内向 [14]。借助大语言模型强大的语言理解和生成能力，PA能够理解学习者的意图，生成上下文感知对话 [1], ，并根据历史交互维持连贯的回应，从而降低交流中的人工感 [15]。自适应数字教师模型的最新进展 [16]

进一步使PA能够根据学生提出的问题动态调整教学角色，提供更加个性化且富有表现力的教学体验。总的来说，这些多模态、基于外观和人格驱动的能力构成了智能教学代理的核心基础。

然而，智能教学代理在实际应用中仍面临若干挑战。当前系统通常依赖静态语音输出和简单手势，这些与教学内容的语义上下文相脱节，导致表现僵硬且不自然。相比之下，人类教师会自然地根据教学情境调整其语音和手势。例如，在讲解困难概念时停顿，或放慢语速并强化手势以突出要点 [17], [18]。当前的教学代理缺乏这种语义感知适应性，使其交互显得机械，参与度也较低。因此，增强教学代理的动态多模态表达能力，使其能够根据教学情境自适应调整言语与非言语行为，对于提升基于虚拟现实的教育环境中的沉浸感、交互性以及整体学习效果具有重要意义。

为应对教学代理在语义感知与表达多样性方面的局限，本文开发了一种具备动态多模态表达能力的基于VR的教学代理原型。该系统使教学代理能够根据教学内容的语义上下文自动调整其表达行为（例如语音和手势），从而模仿人类教师的适应性表现力。为此，我们提出一种基于语义敏感提示构建的大语言模型驱动的多模态方法，引导模型在语义理解的基础上生成协调一致的语音与手势指令。例如，在讲解困难概念时，系统会自动引入停顿、言语填充和思考手势。在强调要点时，则通过语速、语调、音量和手势的变化来强化表达。基于该原型，我们开展了主观实验，以探究动态语音和手势表达对基于虚拟现实的教育场景中用户体验的影响。采用问卷 对六个指标进行了定量分析，即感知有用性、学习投入、使用意愿、类人程度、社会临场感和不适感 [19], [20], [21], [22], [23]。

此外，还基于通过半结构化访谈收集的参与者主观反馈进行了定性分析。研究结果凸显了将个性化特征与动态多模态交互机制纳入教学代理设计的重要性，为构建更具沉浸感、表达更自然的智能教学代理提供了有价值的启示。本文的主要贡献包括：

- 本文提出一种大语言模型驱动的多模态表达生成方法，利用语义自适应提示构建实现语义-行为对齐。不同于依赖固定语音与手势模板的传统虚拟代理系统，该方法融合了真实教师教学过程中观察到的言语与非言语表达模式。通过构建能够编码教学内容语义上下文的提示词，大语言模型可自动生成以该上下文为依据的相应语音与手势指令标签。这一将“说什么”与“如何表达”相结合的统一框架，使虚拟代理的语速、语调、音量和手势能够动态变化，从而营造富有表现力、自然且能感知上下文的教学交互，显著提升虚拟现实中的沉浸感与交互性。

- 通过 2×2 被试内实验和半结构化访谈进行系统评估，以考察动态多模态表达对用户体验的影响。定量与定性分析揭示了动态语音和手势表达提升用户体验的内在机制，为智能教学代理的设计提供了宝贵指导。

II. 相关研究

A. 虚拟现实中的虚拟智能体

虚拟现实通过将用户与物理世界隔离，使其沉浸于计算机生成环境中，从而让用户能够全身心投入模拟空间。在这些环境中，虚拟代理通过视觉与行为线索与用户互动，有效增强用户的临场感和沉浸感 [6], [24]。大量研究从不同角度探讨了虚拟代理特征对用户体验的影响。例如，研究考察了虚拟代理与用户之间的面部相似度如何影响感知临场感 [25], 、代理的外观如何影响协作场景中的共同临场感 [26], ，以及具有多模态输出的人形化身如何显著提升用户在虚拟购物环境中的温暖感、沟通感、信任感和满意度 [27]。

随着人工智能技术的快速发展，由大语言模型驱动的虚拟代理日益融入虚拟现实环境，并广泛应用于语言学习、虚拟课堂、历史教育、医学培训和虚拟会议等领域 [1]。近期研究表明，大语言模型能够通过多模态输入解读用户意图和上下文线索，生成兼具情境性与个性化的对话 [1]。在教育场景中，大语言模型驱动的虚拟教师已被用于提升学习者参与和课堂参与 [27], ；而在情感计算领域，基于大语言模型的虚拟现实模拟则支持在高压情境下开展情绪调节训练 [28]。 医学研究人员进一步验证了在虚拟护理培训模拟器中使用大语言模型生成逼真患者对话的可行性 [29], ；甚至，诸如“Virtual Einstein”这样的概念系统，也展示了专家在沉浸式虚拟现实环境中跨越各大洲进行讨论的场景 [30]。 这些研究表明，教学代理在呈现和对话生成方面不断增强的智能，正提升学习过程中的沉浸感和社会临场感。

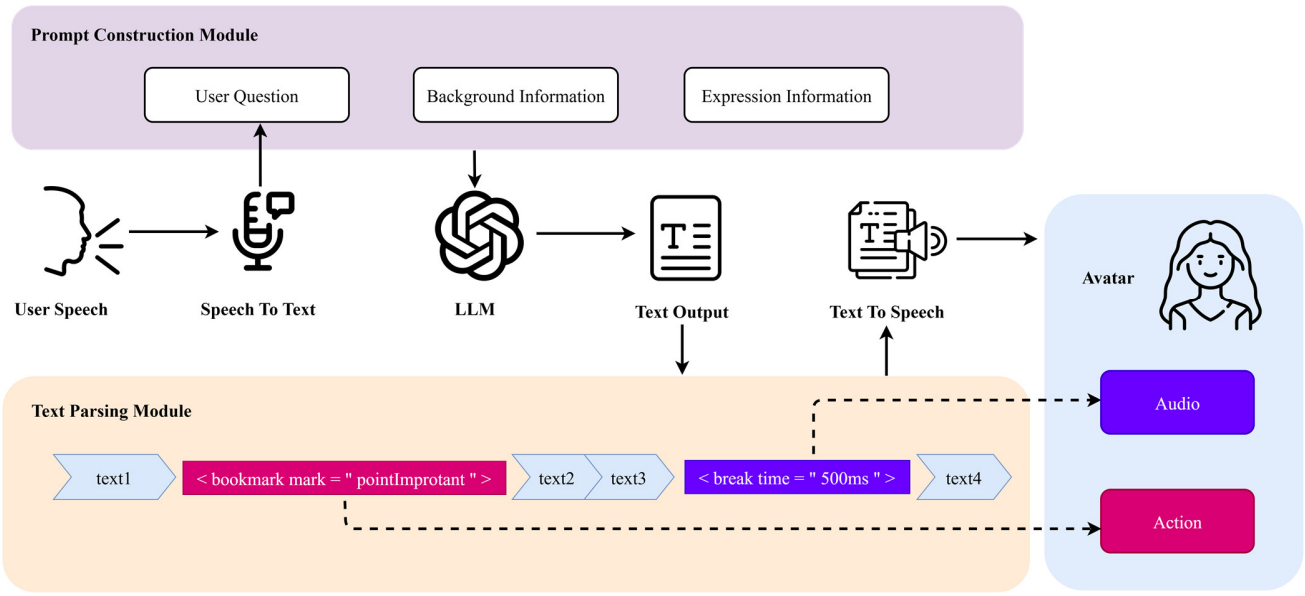


图2。所提出系统的整体框架： 主要功能模块及端到端处理工作流程。

B. 智能体的多模态表达

人格一致性是构建逼真、引人入胜且行为丰富的对话式虚拟智能体的一项关键特征 [31]。 研究表明，语音、语言内容和面部表情等多模态线索不仅能够传达个体的个性特征，还有助于区分不同类型的表现型虚拟智能体 [31]。在虚拟代理的设计中，融入手势、面部表情或情绪动作等非语言线索，能够增强代理的表现力和信息传达能力，从而帮助用户更好地理解任务或识别环境要素 [32]。

随着大语言模型的发展，研究人员越来越多地探索将个性驱动的对话智能体与基于大语言模型的对话生成相结合。在早期框架的基础上，近期研究将大语言模型融入面向个性的智能体架构 [15], , 考察不同的个性表达风格如何影响用户的感知和学习成果。关于自适应角色切换和动作切换的进一步研究 表明，在虚拟教学智能体中根据情境动态调整角色和动作，能够提升学习者在多角色教学场景中的学习投入与交互质量[16]。 关于自由形式对话的其他研究显示，虚拟代理生成对话填充语能够有效减轻用户对系统响应延迟的感知，从而提升交互的流畅度 [10]。此外，研究人员提出了一种由大语言模型支持的可中断对话智能体系统，使智能体能够通过反向通信即时回应用户的打断和反馈 [33]。这些研究表明，与传统的轮流发言对话系统相比，赋能于大语言模型的智能体具有更高的自然度、参与度和流畅度，尤其适用于需要实时适应和社会敏感性的交互。

尽管取得了这些进展，如何实现自然、连贯且与人格一致的动态表达机制，尤其是在语音和手势层面，仍有待进一步研究。本研究旨在通过开发一种面向教学代理的动态多模态表达框架来弥补这一空白，该框架将语义理解与自适应语音和手势生成相结合。

III. 教学代理设计

为了使虚拟代理能够像人类教师一样，根据教学内容的语义调整其表达行为，我们设计并开发了一种具备动态多模态表达能力的基于VR的教学代理原型系统。其目标是使语音和手势能够更加自然地适应文本内容，从而提升沉浸感和可理解性。

A. 总体框架

图2展示了所提出的系统的总体框架。该系统由五个核心模块组成。其中三个是基础模块，包括语音转文本（STT）、大语言模型（LLM）处理以及文本转语音（TTS）模块。其余两个是用于多模态表达生成的辅助模块，即提示构建（PC）模块和文本解析（TP）模块。语音转文本模块使用OpenAI Whisper API进行实时语音转文本转录。基于GPT-4o的大语言模型模块负责语义理解与响应生成。在该系统中，采用了流式输出机制[34] ，模型的响应以分块方式实时传输，从而保持较高的交互性与响应性。通过Microsoft Azure API开发的文本转语音模块支持带流式传输的高保真语音合成，并可通过SSML 标签对语调、停顿和个性化表达进行细粒度控制。该

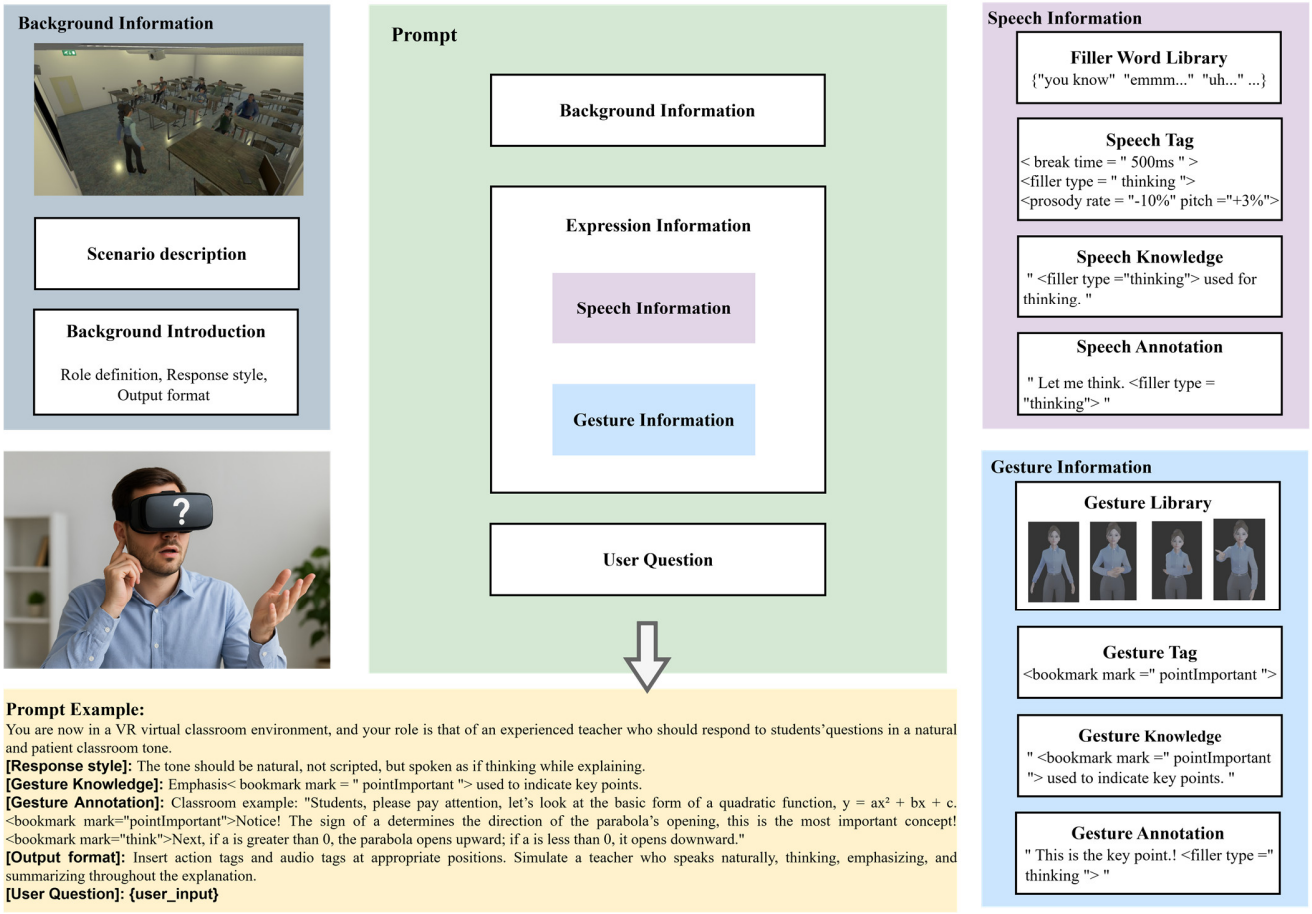


图3. 提示构建（P C）模块：系统架构图展示了提示构建模块（中央）、背景信息的结构（左上角）、详细的语音和手势信息模块及其示例（右侧），以及示例提示（底部，黄色部分）。 on

PC模块通过精心设计的语义敏感提示，引导大语言模型根据语义理解生成相应的语音和手势标签。TP模块解析大语言模型的文本输出，并将其映射为具体的语音与手势指令，以驱动虚拟PA的动态表达。

B. 提示构建（*PC*）模块

图3展示了PC模块的详细结构，并提供了一个提示词示例。输入大语言模型的提示词由三个主要部分组成：背景信息、表达信息和用户问题。

（1）背景信息。该部分包含背景介绍和场景描述。前者定义了大语言模型的固有属性，包括角色定义、输出格式和回应风格。场景描述则规定了当前虚拟环境的上下文内容。

(2) 表达信息。此组件为多模态动态表达提供语音和手势数据。语音和手势信息的设计遵循 [35] 中的方法，并结合了两个要素：知识和标注。知识要素定义了使用各种语音和手势标签的场景与规则，而标注则提供相应的使用示例。

对于语音模态，语音知识规定了应用不同语音标签时的上下文和规则，而语音标注则包含说明性示例。此外，还定义了两个辅助结构：填充词库和语音标签集。前者包含可插入语音中的预设填充词（例如，“你知道”，“嗯……”，“呃……”），从而增强自然度和真实感。后者规定了不同语音标签的结构形式，例如停顿标签 <停顿时长= *“500毫秒”*>、语速与语调标签 <韵律语速= *“-10%”*音量= “中等”音高 = *“+3%”*>，以及填充词标签 <填充词类型= “思考中”>）。

对于手势模态，手势知识定义了教学场景中手势的类别和应用规则，而手势标注则为每种手势类型提供了具体示例。此外还定义了两种辅助结构：手势库和手势标签集。手势库收录了教学场景中九种常用手势，并将其分为思考、强调和总结三类。每类都包含多种手势变体，以提升自然度。手势标签集定义了手势相关标签的结构，例如强调标签 <书签标记= *“pointImportant”*>）。

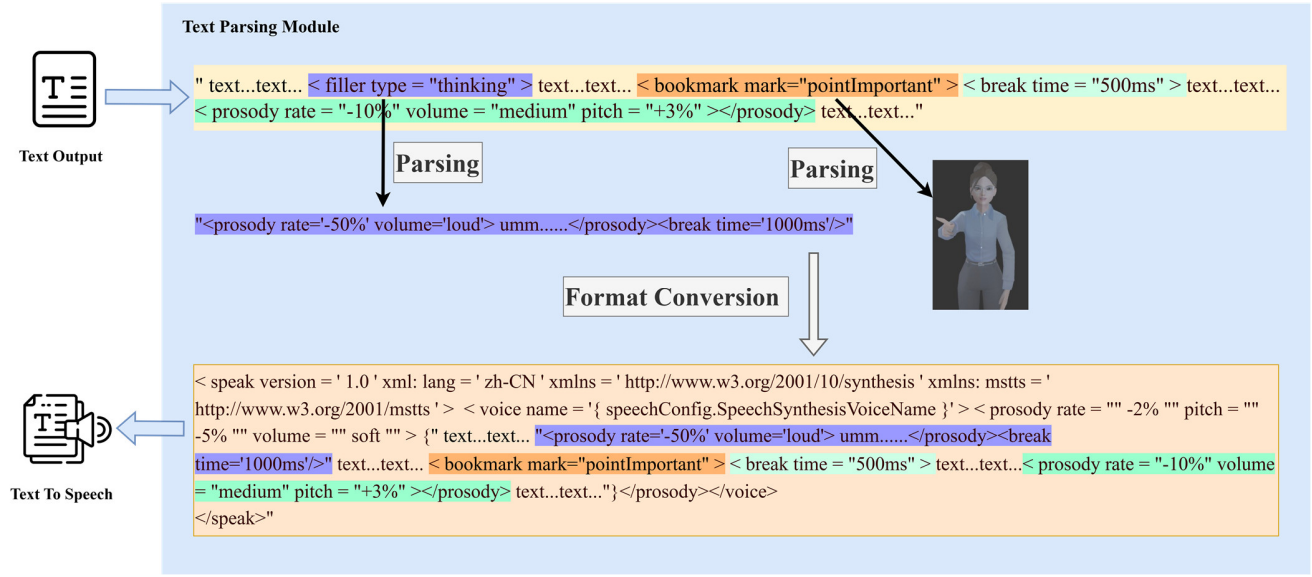


图4。 文本解析（TP）模块： 工作流程图展示了 TP 模块如何解析大语言模型的输出，将填充词和手势标签映射到各自库中对应的项目，并最终将解析后的输出转换为与文本转语音模块兼容的格式。

（3）用户问题。用户问题部分包含用户语音输入的文本表示，该文本由语音转文本模块生成。随后，系统会将这段文本直接附加到构建好的提示词末尾，为语言模型提供用户的实时查询。

图2底部的黄色区域展示了一个示例提示词，该提示词是根据上述提示词结构框架设计的。

C. 文本解析（TP）模块

TP模块有两个主要功能。首先，它会从大语言模型的输出文本中解析填充词标签和手势标签，并将它们映射到预设库中相应的填充词和手势。例如，标签 <填充词类型= “思考”> 会映射到填充词库中的思考类填充词（例如，“嗯……”），而语音标签则会呈现为 <韵律速率= “-50%” 音量= “大声”> 嗯…… </韵律> <停顿时间= “1000毫秒”>。同样，手势标签 < 书签标记 = “pointImportant”> 会映射到手势库中的强调手势。其次，该模块会将大语言模型的输出文本转换为 Microsoft Azure API 可识别的语音合成标记语言兼容格式。在文本转语音阶段，Azure API 会根据嵌入的语音标签调整韵律、停顿和强调。当系统检测到手势标签时，会通过 Unity 动画控制器触发相应手势，从而实现虚拟代理语音和身体动作的动态同步表达。图4展示了该模块的处理工作流程。

IV. 实验

为评估所提出的系统在提升用户体验方面的有效性，我们开展了一项实验，考察动态语音和手势表达对基于虚拟现实的教育场景中学习者表现和感知的影响。

A. 参与者

本研究招募了36名在读学生参与者，并在招募过程中进行了筛选。参与者须未接触过“多媒体传播”相关知识，且目前为在校学生。我们还询问了他们此前是否体验过虚拟现实环境。经过筛选和确认后，共招募到36名参与者，年龄为20至26岁（M = 23，SD = 1.98），包括本科生、硕士生和博士生，其中男性22名，女性14名。我们向参与者介绍了研究流程和实验过程；实验开始前，每位参与者都接受了基础的虚拟现实操作培训。整个实验约持续70分钟。为尽量减少头戴式显示器可能造成的视觉疲劳和身体不适感，参与者每20分钟必须进行一次短暂休息；如果感到不适，也可以随时休息 [36]。

本机构没有伦理委员会。尽管如此，这项研究仍经过内部审查并获批，我们也尽可能遵守既定的伦理原则。所有参与者均在实验开始前提供了知情同意。本研究未收集任何个人敏感数据，也未涉及侵入性操作。实验任务仅包括在受控实验室条件下使用消费级设备进行基于虚拟现实的对话，对参与者的风险极低。我们的伦理程序符合以往虚拟现实用户研究中普遍采用的做法 [37]。

B. 刺激材料

为考察多模态表现力的影响，我们对语音和手势两种模态分别设置了静态和动态两个水平。静态语音（SS）采用默认的合成语音，其语调和语速保持一致；静态手势（SG）则由默认的待机和说话手势构成。相比之下，动态语音（DS）包含韵律变化、停顿和填充语插入，而动态手势（DG）则包含与语义相符的手势，例如强调或思考动作。两种动态模态均通过所提出的方法生成，以实现情境感知表达。

本研究采用 2×2 被试内实验设计，探讨动态语音和手势表达对用户学习体验的影响。实验设置了四种条件（A、B、C、D）。具体而言，条件A（SS+SG）作为对照组，仅包含默认语音和默认手势；条件B（DS+SG）仅包含动态语音表达；条件C（DS+SG）仅包含动态手势表达；条件D（DS+DG）则同时包含动态语音和手势表达。每位参与者都必须在全部四种条件下分别与虚拟代理进行至少五轮对话。为尽量减少潜在的顺序效应，四种条件的呈现顺序采用平衡拉丁方设计[38], [39]进行随机化。



(a)



(b)

图5. VR课堂：（a）为实验构建的VR虚拟教师场景。（b）用户的场景内视角，虚拟教师位于前方，用户界面面板显示实验主题和问卷。

C. 实验设计

研究使用Unity 2022.3.17f1平台开发了虚拟现实教室环境（如图5所示）。通过Ready Player Me设计的虚拟代理部署在Oculus Quest 2上，作为数字教师。参与者扮演学生，在虚拟教室中完成“多媒体通信课程问答”任务。正式实验开始前，研究人员首先介绍了研究目的、学习过程和虚拟现实设备的操作方法，并详细说明了实验的整体步骤。研究人员向参与者提出了一系列简单问题，以确认他们充分理解任务内容和实验流程。随后，参与者被要求签署知情同意书。准备就绪后，参与者在引导下进入虚拟教室，系统会介绍环境、角色任务以及互动方式。参与者需要先阅读主题面板上的课程内容，然后按下开始按钮提问。提出问题后，虚拟黑板会显示问题，虚拟代理随即通过语音进行解答。

在每种实验条件下，参与者都需要向虚拟代理至少提出五个与多媒体传播相关的问题（例如：“什么是视频编码？”），但也可以根据需要提出更多问题。每轮实验结束后，参与者都需要填写用户体验问卷，以评估虚拟代理的表现风格对该条件下学习体验的影响。实验共包含四种条件，各条件的顺序在参与者之间进行了平衡，以避免潜在的顺序效应。所有实验结束后，研究人员对每位参与者进行了结构化一对一访谈，以进一步收集他们对虚拟代理所呈现的语音和手势的多模态动态表达的主观反馈与建议。

D. 问卷与访谈设计

为了全面、定量地评估用户对基于VR的PA的体验，问卷旨在捕捉三个相互补充的关键感知维度：感知效果、社会真实感和情绪舒适度。这一结构确保评估具有理论依据，能够反映用户在学习过程中的认知、社会和情感反应。

感知效果。该维度关注学习者对自身学习表现与参与度的自我感知。借鉴技术接受模型（技术接受模型）[19], 设计了三项指标[20]：感知有用性（感知有用性）、学习投入（学习投入）和使用意向（使用意向）。感知有用性衡量系统促进学习效果的程度，而学习投入评估用户认知与行为参与的深度。学习投入进一步包括EN1（对关键教学信息的注意力）和EN2（时间与认知投入）。使用意向刻画学习者再次使用该系统的动机与意愿，共同反映该系统的感知教学价值。

社会真实感。为评估交互作用的真实性以及虚拟代理被感知的类人特质，纳入了两项指标：类人性（HL）[21]和社交临场感（SP）[22]。HL评估虚拟代理的语音、手势和行为与真实教师的相似程度，强调多模态表达的自然度。SP衡量用户在虚拟教室中感知到与教师“真实交互”或“共处”的真实程度，反映该代理的社会交际真实感。

情绪舒适度。该维度通过不适感（DC）这一指标进行评估，该指标改编自 NASA 任务负荷指数 [23] 的挫折分量表。DC 用于衡量学习过程中出现的挫折感、疲劳和认知过载等负性情绪状态，从而确保沉浸感和交互性的提升不会以牺牲用户的舒适度和情绪健康为代价。

为补充定量评估，我们通过半结构化访谈开展定性分析，以深入了解学习者对基于VR的PA的看法和体验。我们对访谈回答进行了整理和主题分析，进而确定了三个核心研究问题：

(1) 在教学过程中，默认表情模式的*PA*有哪些局限？

(2) 在*PA*中引入多模态动态表情是否改善了*VR*学习体验？如果是，具体体现在哪些方面？

(3) 基于*VR*的教学代理原型系统在实际应用中还有哪些方面可以进一步优化？

E. 数据处理

对于定量数据，由于实验问卷采用5点李克特量表（1 = 非常不同意， 5 = 非常同意），采用对齐秩变换（ART）方法[10], [40] 进行2×2 全因子重复测量方差分析（ANOVA）。在ART 方差分析之后，进一步使用ART-C程序[41] 进行事后比较，以检验显著的主效应和交互效应，并采用Holm-Bonferroni校正方法对多重比较的p值进行调整。为便于分析和可视化，将原始1-5量表线性转换为-2（非常不同意）至2（非常同意）的范围，中性反应以零为中心。图4展示了基于该转换量表的均值和±1 均值标准误（SEM）。

在定性数据分析中，本研究采用主题分析，运用自下而上法分析访谈录音 [16]。 首先，所有访谈录音均被逐字转录，形成文本数据。随后仔细阅读转录文本，并对重要句子或片段进行描述性编码。在编码过程中，我们持续比较新数据与已有数据，识别潜在的新概念或模式，并修订和完善现有编码。在此基础上，将语义相近或内容相关的编码归入更高层次的类别，并进一步提炼出反映参与者体验的核心主题。最后，对所识别的主题进行反复审查与完善，以确保其准确反映参与者的观点，并有效回应研究问题。

V. 结果分析

本节报告所提出的系统的实验结果与分析。定量数据与定性数据均得到考察，以评估教学代理的多模态动态表情如何影响基于虚拟现实的教育中的用户体验。

A. 定量结果分析

我们首先呈现并分析虚拟代理的语音与手势的多模态动态表达如何影响用户体验。分析在第IV.B节所述的四种实验条件下进行，重点关注六个维度，即感知有用性（PU）、学习投入（EN）、使用意向（IU）、HL、SP和DC。

感知有用性（PU）。对于感知有用性，如图6（a）所示，ART 方差分析结果表明，虚拟代理的语音（F = 316.34，p < 0.001）和手势（F = 186.07，p < 0.001）均显著影响参与者对虚拟现实原型的感知有用性，但语音与手势之间的交互作用并不显著（F = 1.64，p = 0.204）。效应量分析显示，语音对感知有用性具有大效应（η $^{2}_{p}$ = .），手势 0 80 70 同样具有大效应（η^2_p = 0.），进一步证实了这两个因素的重要性。描述性统计表明，条件D得分最高（M = 1 25， = 0 44 = 0 55， = . SD .），0 60 = 0 20， = 0 41 其次是条件B（M . SD .）和条件C（M . SD .），基线条件A得分最低（M = 0 95， = - . 0 60 SD .）。经Bonferroni 方法校正的事后比较显示，实验条件B、C和D与基线条件A存在显著差异（p < 0.001），且条件D显著优于所有其他条件。

学习投入（EN）。对于学习投入，ART 方差分析结果（图6 (b)-(c)）显示，语音和手势对 EN 1 = 286 26， < 0 001 = 均具有显著主 94 03， < 0 001 2 = 259 29， < 0 001 = 195 0, 效应（F . p .）（F . p .）；对 EN 也均具有显著主效应（F . p .）（F . p < 0.001）。两项指标均未观察到显著交互效应（F = 0.65, p = 0.422）（F = 0.79, p = 0.375）。效应量分析表明，语音和手势对 EN 均有很大影响（语音：η $^{2}_{p}$ = 0 786 1, $^{2}_{p}$ = . 对 ENη. 对 0 769 2 $^{2}_{p}$ = 0 547 1, $^{2}_{p}$ = 0 489 EN）（手势：η. 对 ENη. 对 EN2）。描述性统计显示，条件 D 的平均分最高（EN1：M = 1.30, SD = 0.57）（EN2：M = 1.25, SD = 0.44），其次是条件 B（EN1：M = 0.80, SD = 0.60）（EN2：M = 0.70, SD = 0.57）和条件 C（EN 1 = 0 05， = 0 69 2 = 0 = : M - . SD .） 0 56 1 = （EN：M , SD .），基线条件 A 得分最低（EN：M - 0.70, SD = 0.66）（EN2：M = -0.75, SD = 0.72）。Bonferroni校正的事后检验证实，B、C 和 D 的得分均显著高于 A（p < 0.001），且 D 优于其他所有条件。

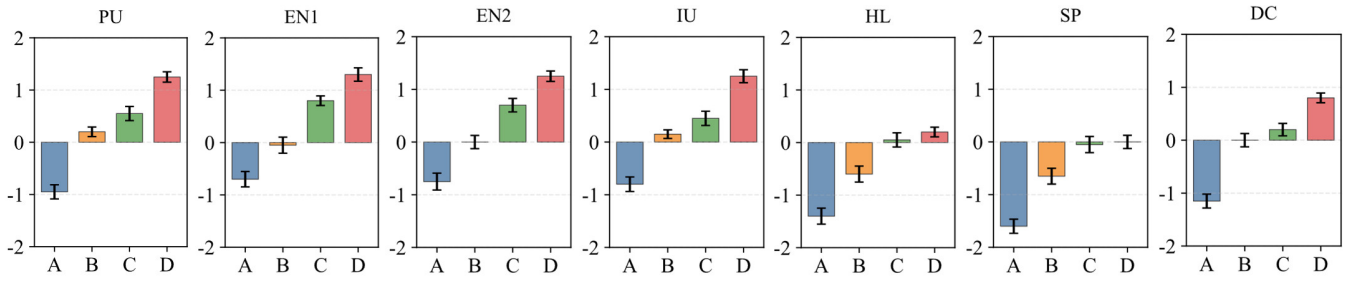


图6. PA 的四种表达模态的影响对用户体验的影响：（a）感知有用性，（b）—（c）学习投入，（d）使用意愿，（e）类人性，（f）社会临场感，（g）不适感。误差棒表示均值标准误（SEM）。

使用意向（IU）。对于使用意向，如图6 (d)所示，ART 方差分析表明，语音（F = 261 20, < 0 001）和手势（F = 195 0, < . p . ） 0 001 和手势（F . p . ）对参与者使用虚拟现实原型的意愿均有显著主效应。两个因素之间未观察到显著交互作用（F = 1.09, p = 0. 299）。效应量分析显示，语音（ηp2 = 0.77）和手势（ηp2 = 0. 71）均产生较大效应，凸显了它们对用户采纳意向的显著贡献。描述性统计显示，条件 D 的平均分最高（M = 1.25, SD = 0.55），其次是条件 B（M = 0.45, SD = 0.60）和条件 C（M = 0.15, SD = 0.37），而基线条件 A 得分最低（M = -0.80, SD = 0.62）。Bonferroni校正的事后比较结果与前述指标报告的结果一致。

类人性（HL）。如图6（e）所示，加入动态语音和手势后，参与者对类人性的感知显著提升。统计分析证实，语音（F = 165.30，p < 0.001）和手势（F = 43.06，p < 0.001）均带来了显著改善，而二者的交互作用并不显著（F = 0.08，p = 0.778）。效应量结果进一步表明，两种模态均有所贡献（语音：η_p^2 = 0.68；手势：η_p^2 = 0. 36），这说明自然韵律与协调手势共同增强了虚拟代理的感知类人性。描述性统计结果显示，D条件的平均得分最高（M = 0.20，SD = 0.41），其次是B条件（M = 0.05，SD = 0.60）和C条件（M = -0.60，SD = 0.68），基线条件A的得分最低（M = -1.40，SD = 0.68）。尽管D条件的类人性评分最高，但其总体得分仍明显低于同一条件下的感知有用性、学习投入和使用意向，这表明参与者仍认为虚拟代理与完全具有人类特征的教师之间存在差距。

社会临场感（SP）。如图 6(f) 所示，随着加入动态语音和手势，参与者对 SP 的感知有所提升。分析结果表明，语音（$F = 161.6$，$p < 0.001$）和手势（$F = 40.33$，$p < 0.001$）均存在显著主效应，而二者之间没有交互作用（$F = 0.06$，$p = 0.808$）。效应量（语音：ηp2 = 0.67，手势：ηp2 = 0.34）进一步表明，这两种模态都显著增强了 SP。描述性数据表明，D 条件下的评分最高，其次是 B 和 C 条件，而基线条件 A 的评分最低。然而，与 HL 的研究结果类似，尽管 B、C 和 D 条件的表现优于基线条件，但其均值明显低于其他维度，表明 SP 的提升仍相对有限。

不适感（DC）。对于不适感（DC），如图6（g）所示，动态语音（F = 230 54, < 0 001）和手势（F . p . ） 154 08, < 0 001 和手势（F . p . ）均显著影响参与者对虚拟现实原型中不适感的感知，而二者的交互作用并不显著（F = 0.43, p = 0.518）。效应量分析表明，这两种模态在缓解负面情绪方面均有较强作用（语音：ηp2 = 0.75，手势：ηp2 = 0.66），凸显了它们在交互过程中提升情绪舒适度的重要性。描述性数据表明，条件D的平均得分最高（M = 0.80, SD = 0.41），其次是条件B（M = 0.20, SD = 0.52）和条件C（M = 0.00, SD = 0.56），基线条件A的评分最低（M = −1.15, SD = 0.58）。经Bonferroni校正的事后分析显示，条件B与条件C之间没有显著差异（p = 0.339），这表明，单独应用时，动态语音和手势表达在降低用户不适感方面效果相当。

B. 定性结果分析

下面，我们呈现并讨论半结构化访谈的研究发现。我们对访谈内容进行了主题分析，以探讨参与者对相关研究问题的看法。

(1) 单调的语音表达会削弱学习专注度和临场感。

大多数参与者表示，采用默认表达模式的虚拟代理（PA）语音过于单调，缺乏语调变化和停顿，导致他们难以及时理解和吸收教学内容。尤其是在讲解时间较长或内容较多时，参与者容易分心并产生信息过载。

“当我向虚拟代理（*PA*）提问时，它会从概述、技术原理等方面进行解释，这当然很好，但如果两三分钟都保持同一种语调，我最后就会错过重要要点。”(P3)

此外，语音和手势缺乏变化，也削弱了用户的社会临场感和沉浸感。

一些参与者提到，虚拟代理的语音缺乏情感变化，给人一种“机械”的印象，让它听起来更像是在朗读文本，而不是进行互动交流。

“数字教师的语音和手势让我觉得它并没有在思考问题，只是在背诵，不像真实教师。”（P6）

“听它单调地说了一会儿后，我开始走神，很难像在真实课堂中那样保持投入。”（P13）

“过了一会儿，我开始不耐烦，渐渐把注意力转移到其他事情上。”（P9）

总体而言，参与者普遍认为，虚拟代理表达中缺乏韵律语音和情感细微差别，在一定程度上会降低学习专注度和社会临场感，从而影响整体学习体验。

（2）动态语音与手势表达提升自然度和注意力维持。

大多数参与者认为，引入动态语音与手势显著提升了整体学习体验。首先，在语音方面，恰当的停顿和填充词让虚拟代理的表达更加自然，也为参与者提供了思考和理解的时间。

“加入停顿后的虚拟代理给了我时间去思考所说的内容，这真的让它听起来更像一位边思考边讲解的真实教师。”（P1）

其次，在手势方面，动态手部动作和身体动作帮助参与者更好地聚焦于关键信息，提升了他们的学习专注度。

“与默认手势相比，我更喜欢动态手势，因为它们会提示虚拟代理即将讲解重要内容。”（P4）

“添加手势帮助我更加专注。即使我开始走神，虚拟代理的手势也会立刻让我重新集中注意力。”（P8）

总体而言，语音和手势的动态变化增强了虚拟代理的类人性，使学习过程更加自然且富有节奏感，从而提升了参与者的注意力维持能力和交互感。

（3）表达细节与交互机制仍需进一步优化。

参与者提出了几条关于改进教学代理的建设性建议。首先，一些参与者认为，目前的手势库较为有限，手势的重复使用削弱了表达的丰富性和自然度。

“目前手势库相当有限，而且在对话过程中，手势往往会重复。如果手势的种类和数量更多，效果会更好。”**(P4)**

此外，一些参与者指出，手势之间的切换不够流畅，显得尴尬而僵硬，从而削弱了社会临场感。

“我觉得手势过渡不自然，而且相当僵硬，让我觉得很奇怪。”（P6）

“如果手势能够实时捕捉并生成，效果会更加自然。当语音与不自然的手势搭配时，实际上会降低类人性体验。”（P5）

一些参与者建议在设计中加强语音与手势之间的协调，以提升整体表现的一致性。

“有时，手势和语音的变化是彼此分开的。我认为它们应该相互配合。” **(P4)**

与此同时，一些参与者还提出，与单向的轮流式交流相比，引入“可中断对话”机制或许更具互动性和真实感。

总体而言，参与者普遍期待进一步优化虚拟代理的手势丰富度、过渡流畅度以及语音与手势协调，以增强其类人性和交互沉浸体验。

VI. 讨论

基于上述定量与定性分析，以下讨论将从感知效果、社会真实感和情绪舒适度三个维度审视研究发现，进一步阐释虚拟代理的动态语音和手势表达对学习者在基于虚拟现实的学习中的体验所产生的影响。

从感知效果维度来看，本研究结果表明，虚拟代理的动态语音与手势表达能够显著提升用户对学习效果的主观感知。在语音方面，停顿和语调变化为学习者提供了思考和内化信息的时间，从而减轻认知负荷，促进理解和记忆 [42], 。[43]

在手势方面，身体动作的动态变化和强调能够引导学习者关注显著信息，从而提升注意力维持和参与度。这一模式与 Goldin-Meadow 的解释一致：手势通过在信息加工过程中组织并突出信息来影响认知 [44]。

然而，定量分析结果显示，动态语音与动态手势之间的交互作用并不显著。这一现象可能与系统设计中这两种模态缺乏时间和语义一致性有关。当语音与手势的节奏、语义或情感意图不一致时，学习者需要投入额外的认知资源来整合来自不同渠道的信息，从而削弱多模态整合的积极效果 [45]。此外，关于多模态交互的既有研究指出，结合

语音与手势的益处取决于它们在时间、语义和情感维度上的协调。当这种协调被打破时，多模态线索可能会相互干扰，进而削弱甚至抵消彼此的效果。Oviatt 关于教育接口的研究也探讨了这一现象。 [46]。 用户访谈结果也支持这一解释。一些参与者指出，当语音与手势缺乏协调时，表达会显得僵硬、不自然，从而降低拟人化体验和沉浸感。

从社会真实感维度来看，本研究结果表明，在虚拟代理（PA）的语音和手势中融入动态表达，有助于增强用户对其类人性和社会临场感的感知。加入填充词、韵律变化以及动态调整的手势，使虚拟代理的交际行为更具人类特征，进而提升用户感知到的自然度和喜爱度。这一观察结果与 Reeves 和 Nass 提出的媒介等式框架相一致[47], 该框架表明，当技术系统呈现出类人的线索时，即使这些线索是人工生成的，人们也倾向于对其产生社会反应。此外，动态语音与手势的结合能够增强社会临场感 [48] 和情绪传染[49], 在一定程度上减少用户在交互过程中的无聊感和疏离感。

然而，定量结果显示，在四种实验条件下，用户对类人性和社会临场感的评价均相对较低，这表明，仅依靠语音和手势的动态变化仍不足以营造较高水平的社会真实感。这一现象可能与交互节奏、响应延迟以及“可打断性”等因素有关。根据 Cassell 关于具身对话代理的框架[50], 实现类人交互不仅取决于表层的语音和手势表达，还取决于恰当的响应时机和双向对话控制。

在情绪舒适度方面，结果表明，动态语音和动态手势均显著降低了用户在交互过程中的挫折感和烦躁感，但二者的交互效应并不显著，这说明每种模态都能独立缓解用户的负面情绪。从社会临场感理论 [51] 和情感计算 [52], 的角度来看，融合情绪表达性语音和自然手势的系统可能会提升用户对代理温暖感、共情能力和响应性的感知。

然而，结果也表明，即使在动态语音和手势条件下，整体DC分数仍然相对较低，这说明负性情绪的改善存在一定的上限。一种可能的解释是，虚拟代理的讲解时间较长，要求参与者长时间保持注意力，从而增加了认知负荷[53] , 并引发精神疲劳和不耐烦。根据期望不确认理论（EDT）[54], , 用户满意度受先前预期与实际体验之间差距的影响。在我们的情境中，如果用户期望系统能够高效、简洁地提供信息，却觉得交互过程耗时过长，这种期望差距可能会使他们感到不耐烦。因此，未来的系统设计不仅应提升多模态情绪表达能力，还应动态调整语音传递的语速节奏和时长，以改善用户的情绪舒适度和交互体验。

VII. 结论

本研究考察了在教育情境中，由大语言模型驱动的基于VR的PA中动态语音和手势表达的有效性。所提出的系统使虚拟助理原型能够生成语义一致且能感知语境的多模态表达，实现教学内容与表达行为的协同整合。实证结果表明，动态语音和手势表达均显著提升了学习者感知的学习效果、学习投入和使用意愿，同时缓解了学习过程中的疲劳感和挫折感。此外，语音与手势的组合动态表达显著增强了学习者对PA类人性和社会临场感的感知。然而，仅依靠低层次的表达变化不足以充分营造高度的社会真实感。研究结果凸显了在教学代理设计中融入动态交互机制的重要性，并为开发更具沉浸感、教学方式更自然的智能数字教学助理提供了有价值的见解。

未来的研究将进一步探讨教学代理中多模态元素之间的协同效应，尤其是语音、手势与教学内容之间的语义一致性，因为这可能影响用户的注意力、信息处理以及感知到的社会临场感。此外，未来研究还应将所提出的框架拓展至其他应用领域，包括医学培训、企业学习和心理咨询，以评估其可推广性和更广泛的适用性。

参考文献

[1] Y. Tang、J. Situ、A. Y. Cui、M. Wu 和 Y. Huang，《大语言模型在扩展现实中的集成：当前趋势、挑战与未来展望综述》，载于*2025*年计算机系统人因国际会议论文集，ACM，2025年，第1–24页。doi: 10.1145/3706598.3714224。[2] V. Biener 等，《在虚拟现实中工作一周的影响量化》，载于*IEEE*可视化与计算机图形学汇刊，第28卷，第11期，第3810–3820页，2022年，doi: 10.1109/TVCG.2022. 3203103。[3] K. Kilteni、I. Bergstrom 和 M. Slater，《沉浸式虚拟现实中的击鼓：身体塑造我们的演奏方式》，载于*IEEE*可视化与计算机图形学汇刊，第19卷，第4期，第597–605页，2013年，doi: 10.1109/TVCG.2013.29。[4] S. Liao、S. Chaudhri、M. K. Karwa 和 V. Popescu，《SeamlessVR：弥合沉浸式与非沉浸式可视化之间的鸿沟》，载于*IEEE*可视化与计算机图形学汇刊，第31卷，第5期，第2806–2816页，2025年，doi: 10.1109/TVCG.2025.3549564。[5] T. Song、F. Pabst、U. Eck 和 N. Navab，《通过对话式虚拟代理和沉浸式可视化提升患者对机器人超声的接受度》。 *IEEE*可视化与计算机图形学汇刊

> REPL 请用您的稿件编号替换此行（双击此处进行编辑） <

可视化与计算机图形学，第31卷，2911，2025年，第5期，第2901–2910页，doi：10.1109/TVCG.2025.3549181。[6] C. T. Tan 等，《探究人工智能聊天机器人导师中的化身呈现对学习体验的影响》，载于《2025年计算机系统人因国际会议论文集》，2025年。[7] X. Huang、J. Huss、L. North、K. Williams 和 A. Boyd-Devine，《基于理论的沉浸式虚拟现实设计在科学学习中的认知与动机益处》，《计算机与教育开放》，第4卷，文章编号100124，2023年，doi：10.1016/j.caeo.2023.100124。[8] J. Radianti、T. A. Majchrzak、J. Fromm 和 I. Wohlgenannt，《高等教育中沉浸式虚拟现实应用的系统综述：设计要素、经验教训与研究议程》，《计算机与教育》，第147卷，文章编号103778，2020年，doi：10.1016/j.compedu.2019.103778。[9] A. Behera、P. Matthew、A. Keidel、P. Vangorp、H. Fang 和 S. Canning，《将面部表情和上半身手势与学习任务关联，以增强智能辅导系统》，《国际人工智能教育杂志》，第30卷，第2期，第236–270页，2020年，doi：10.1007/s40593-020-00195-2。[10] M. Maslych 等，《借助大语言模型驱动的智能虚拟代理缓解自由形式对话中的响应延迟》，载于第七届ACM对话式用户界面会议论文集，加拿大安大略省滑铁卢：ACM，2025年，第1–15页。doi：10.1145/3719160.3736636。[11] Y. Kim、Z. Aamir、M. Singh、S. Boorboor、K. Mueller 和 A. E. Kaufman，《可解释扩展现实：使用大语言模型辅助分析框架理解用户在扩展现实环境中的行为》，《IEEE可视化与计算机图形学汇刊》，第31卷，第5期，第2756–2766页，2025年5月，doi：10.1109/TVCG.2025.3549537。[12] J. Chen、X. Wu、T. Lan 和 B. Li，《LLMER：利用大语言模型生成的JSON数据构建交互式扩展现实世界》，《IEEE可视化与计算机图形学汇刊》，第31卷，第5期，第2715–2724页，2025年，doi：10.1109/TVCG.2025.3549549。[13] M. Vallance，《虚拟遗产与人工智能：借助伊莎贝拉·伯德了解19世纪日本》，载于2023年第七届教育与电子学习国际会议论文集，日本东京：ACM，2023年，第6–13页。doi：10.1145/3637989.3638007。[14] S. Sonlu、B. Bendiksen、F. Durupinar 和 U. Güdükbay，《具身化与人格对基于大语言模型的对话代理的影响》，载于2025年IEEE虚拟现实与三维用户界面大会（VR），法国圣马洛：IEEE，2025年，第718–728页。doi：10.1109/vr59515.2025.00094。[15] H. Wan 等，《在社交虚拟现实中构建基于大语言模型的人工智能代理》，载于CHI计算机系统人因大会扩展摘要集，美国夏威夷州檀香山：ACM，2024年，第1–7页。doi：10.1145/3613905.3651026。[16] Z. Zhu、A. Yu、X. Tong 和 P. Hui，《探索用于虚拟现实历史教育的大语言模型驱动的角色与行动切换式教学代理》，载于2025年计算机系统人因国际会议论文集，日本横滨：ACM，2025年，第1–19页。doi：10.1145/3706598.3713109。[17] R. D. Roscoe 和 M. T. H. Chi，《导师式学习：讲解与回答问题的作用》，《教学科学》，第36卷，350，2008年，第4期，第321–? 页，doi：10.1007/s11251-007-9034-5。[18] J. A. Hall、T. G. Horgan 和 N. A. Murphy，《非语言沟通》，《心理学年度评论》，第70卷，第271–294页，2019年。[19] M. Romano 等，《探索沉浸式虚拟现实在意大利学校中的潜力：与高中教师开展的实践工作坊》，《MTI》，第7卷，第12期，第1–18页，2023年，doi：10.3390/mti7120111。[20] A. Cavallaro、M. Romano、R. Laccone，《探究用户对虚拟现实中与人工智能机器人进行语音交互的看法：聚焦外语学习与沟通动态》，载于《计算机科学讲义》，尚：施普林格·自然瑞士，2024年，第20–30页。doi：10.1007/978-3-031-60606-92。[21] C. Bartneck、D. Kulić、E. Croft 和 S. Zoghbi，《机器人拟人化、生命感、喜爱度、感知智能与感知安全性的测量工具》，《国际社会机器人学杂志》，第1卷，第1期，第71–81页，2009年，doi：10.1007/s12369-008-0001-3。[22] G. Makransky、L. Lilleholt 和 A. Aaby，《虚拟现实环境多模态临场感量表的开发与验证：验证性因子分析与项目反应理论方法》，《计算机与人类行为》，第72卷，第276–285页，2017年。[23] D. Heidrich、A. Wohlan 和 M. Schaller，《虚拟现实中交互式与被动加载场景的感知速度、挫折感与愉悦感》，《国际人机交互会议》。尚：施普林格国际出版社，2020年。

[24] D. Hamilton, et al. “沉浸式虚拟现实作为教育中的教学工具：定量学习成果与实验设计的系统性文献综述。”《教育计算机杂志》，第8卷，第1期，第1-32页，2021年。[25] H. Suk and T. H. Laine，“虚拟化身面部外观对用户在沉浸式虚拟现实中感知到的具身感与临场感的影响”，《电子学》，第12卷，第3期，第583页，2023年，doi: 10.3390/ electronics12030583.[26] J. P. Freiwald, J. Schenke, N. Lehmann-Willenbrock, and F. Steinicke，“虚拟化身外观与移动方式对虚拟现实协作中共同临场感的影响”，载于《Mensch und Computer 2021会议论文集》，德国因戈尔施塔特：ACM，2021年，第393-401页。doi: 10.1145 /3473856.3473870.[27] Z. Liu et al.，“ClassMeta：设计交互式虚拟同学以促进虚拟现实课堂参与”，载于《CHI计算机系统人因大会论文集》，美国檀香山：ACM，2024年5月，第1-17页。doi: 10.1145/3613904.3642947. [28] A. Fang, H. Chhabria, A. Maram, and H. Zhu，“面向日常自我护理的社交模拟：利用虚拟现实、增强现实和大语言模型练习压力缓解的设计洞见”，载于《2025年计算机系统人因国际会议论文集》，日本横滨：ACM，2025年4月，第1-23页。doi: 10.1145/3706598.3713115. [29] N. Kapadia, et al. “基于人工智能的虚拟现实护士培训模拟器中大语言模型生成的对话评估。”国际人机交互会议。尚：施普林格·自然瑞士，2024年，第200-212页。doi: 10.1007/978-3-031-61041-7_13. [30] A. Shoa, R. Oliva, M. Slater, and D. Friedman，“与爱因斯坦共进寿司：利用基于大语言模型的虚拟人增强混合式现场活动”，载于《第23届ACM智能虚拟代理国际会议论文集》，德国维尔茨堡：ACM，2023年，第1-6页。doi: 10.1145 /3570945.3607317. [31] S. Sonlu, U. Güdükbay, and F. Durupinar，“具有多模态个性表达的对话代理框架”，《ACM图形学汇刊》，第40卷，第1期，第1-16页，2021年。doi: 10.1145 /3439795.[32] Seo, Gapyuel. "通过在虚拟现实中使用手语增强可供性。" 国际人机交互会议。尚：施普林格·自然瑞士，2024年，第104-110页。doi: 10.1007/978-3-031-61947-2 12.

[33] C. Liu et al.，“通过设计支持打断与反馈语的由大语言模型驱动的语音代理，实现与老年人的自然对话”，载于《2025年计算机系统人因国际会议论文集》，日本横滨：ACM，2025年，第1-22页。doi: 10.1145/3706598.3714228. [34] OpenAI，OpenAI API文档：流式响应，2025年。[在线]。可获取：https://platform.openai.com/docs/api-reference/ streaming. [35] Y. Tang, J. Situ, A. Y. Cui, M. Wu, and Y. Huang，“大语言模型在扩展现实中的集成：当前趋势、挑战与未来展望的全面综述”，载于《2025年计算机系统人因国际会议论文集》，日本横滨：ACM，2025年，第1-24页。doi: 10.1145/3706598.3714224. [36] 国际电信联盟电信标准化部门，“P.910：多媒体应用的主观视频质量评估方法”，国际电信联盟，瑞士日内瓦，建议书P.910，2021年。[37] J. Lacoche, A. Foulonneau and S. L. Dit Picard，“用于增强现实服务原型设计的增强现实、虚拟现实与桌面工具比较分析”，IEEE可视化与计算机图形学汇刊，第31卷，第5期，第2579-2588页，2025年，doi: 10.1109/TVCG.2025.3549885. [38] M. E. Kite, Jr, B. E. Whitley. 行为科学研究原理。Routledge，2012年。[39] V. L. Anderson., and R. A. McLean. 实验设计：一种务实方法。CRC Press，2018年。[40] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins，“仅使用方差分析程序进行非参数因子分析的对齐秩变换”，载于《SIGCHI计算机系统人因大会论文集》，加拿大温哥华：ACM，2011年，第143-146页。doi: 10.1145/1978942.1978963.[41] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock，“用于多因子对比检验的对齐秩变换程序”，载于《第34届ACM用户界面软件与技术研讨会年会论文集》，美国虚拟活动：ACM，2021年，第754-768页。doi: 10.1145/3472749.3474784.[42] J. Sweller. “认知负荷理论、学习困难与教学设计。”《学习与教学》，第4卷，第4期，第295-312页，1994年。[43] R. E. (Ed.). Mayer.《剑桥多媒体学习手册》。剑桥大学出版社，2005年。

[44] S. Goldin-Meadow. “手势如何改变我们的想法。”《神经科学与教育趋势》，第3卷，第1期，第4-6页，2014年。[45] R. Moreno and R. Mayer，“交互式多模态学习环境：交互式学习环境特刊：当代议题与当前趋势”，Educ Psychol Rev，第19卷，第3期，第309-326页，2007年，doi: 10.1007/s10648-007-9047-2. [46] S. Oviatt. “未来教育接口的设计”，Routledge，2013年。[47] B. Reeves, and C. Nass. “媒介等式：人们如何像对待真人一样对待计算机、电视和新媒体。”英国剑桥，第10卷，第10期，第19-36页，1996年。[48] L. Fortunati, and A. M. Manganelli. “电信的社会表征。”个人与泛在计算，第12卷，第6期，第421-431页，2008年。[49] E. Hatfield, J. Cacioppo, and R. Rapson. “情绪传染”剑桥大学出版社。纽约，1994年。[50] J. Cassell. “具身对话界面代理。”ACM通信，第70-78页，2000年。[51] F. Biocca, C. Harms, J. K. Burgoon, “迈向更稳健的社会临场感理论与测量：综述与建议标准。”临场感：远程操作与虚拟环境，第12卷，第5期，105474603 322761270 456 480, 2003 10 1162 pp. - . doi: https://doi.org/ . / .[52] J. Tao, and T. Tan. “情感计算：综述。”载于《情感计算与智能交互国际会议论文集》。柏林，海德堡：Springer Berlin Heidelberg，2005年。[53] J. Sweller. “ 问题解决过程中的认知负荷：对学习的影响。”《认知科学》，第12卷，第2期，第257-285页，1988年。[54] R. L.Oliver. “满意度反应的认知、情感与属性基础。”《消费者研究杂志》 第20卷，430, 1993 3, 418no. pp. - .