

T/CASME

团 体 标 准

T/CASME XXX—2025

智慧课程 AI 数字人教师技术要求

Technical requirements for AI digital teachers in smart courses

（征求意见稿）

2025 - XX - XX 发布

2025 - XX - XX 实施

中国中小商业企业协会 发 布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 系统框架 1

5 技术要求 2

 5.1 形象 2

 5.2 语音 2

 5.3 动画 3

 5.4 交互处理 3

 5.5 多模态输出 5

 5.6 整体逼真度 5

6 评估方法 5

 6.1 形象 5

 6.2 语音 5

 6.3 动画 6

 6.4 交互处理 6

 6.5 多模态输出 6

 6.6 整体逼真度 7

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由重庆威爱云科技有限公司提出。

本文件由中国中小商业企业协会归口。

本文件起草单位：重庆威爱云科技有限公司、重庆大学、西南大学、中国科学院重庆绿色智能技术研究院、重庆邮电大学、重庆理工大学、重庆工商大学、重庆师范大学、重庆三峡学院、重庆移通学院、重庆电子科技职业大学、重庆商务职业学院、北京神州绿盟科技有限公司、重庆西信天元数据资讯有限公司、成都农业科技职业学院。

本文件主要起草人：黎岩松、龚玉霞、陆艳军、胡春强、陈芋文、武春岭、刘革平、贾韬、冯玉明、王宁、罗吉、祝清意、何春林、朱钦候、桑春燕、宋群、闫东方、何奕颖、蒲誉文、周宇、冯骥、叶坤、陈红艳、周小艳、孙启龙、郑慧娟、甘臣权、陈佳俊、张扬、冉靖、孙德福、冉银、杨山金、彭春蓉。

引 言

本文件的发布机构提请注意，声明符合本文件时，可能涉及到5.3.1.1与一种语音驱动视频人脸唇形重建的方法、装置及设备（专利号：ZL118351222B）相关的专利的使用。

本文件的发布机构对于该专利的真实性、有效性和范围无任何立场。

该专利持有人已向本文件的发布机构承诺，他愿意同任何申请人在合理且无歧视的条款和条件下，就专利授权许可进行谈判。该专利持有人的声明已在本文件的发布机构备案。相关信息可以通过以下联系方式获得：

专利持有人姓名：重庆威爱云科技有限公司。

地址：重庆市渝北区华山南路16号1号楼502号房、504号房、506号房。

请注意除上述专利外，本文件的某些内容仍可能涉及专利。本文件的发布机构不承担识别专利的责任。

智慧课程 AI 数字人教师技术要求

1 范围

本文件规定了智慧课程AI数字人教师的系统框架、技术要求、评估方法。
本文件适用于智慧课程中AI数字人教师及其系统的研发、设计。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 21023—2007 中文语音识别系统通用技术规范

GB/T 21024 中文语音合成系统通用技术规范

YD/T 4393.1—2023 虚拟数字人指标要求和评估方法 第1部分：参考框架

YD/T 4393.2—2023 虚拟数字人指标要求和评估方法 第2部分：2D真人形象类合成技术

3 术语和定义

GB/T 21024界定的以及下列术语和定义适用于本文件。

3.1

AI 数字人教师 AI digital teachers

利用人工智能技术创建，在智慧课程中模拟人类教师进行教学活动的虚拟形象。

4 系统框架

AI数字人教师系统分为多模态输入、交互处理、形象、语音、动画、多模态输出六个部分，系统框架见图1，具体内容如下：

- 多模态输入：接收用户的输入信息，支持文字、语音、图像、触控等输入方式；
- 交互处理：使 AI 数字人教师能够识别感知、和理解决策的能力，基于感知与识别、分析与决策等智能技术识别用户输入的信息，理解用户意图，驱动 AI 数字人教师后续的语音和动作；
- 形象：提供 AI 数字人教师的 2D 写实风格人物形象；
- 语音：呈现 AI 数字人教师的人物语音；
- 动画：呈现 AI 数字人教师的人物动画，包括唇形、面部表情、肢体动作；
- 多模态输出：将输出结果呈现给用户，支持手机、电视、计算机、投影等输出方式。

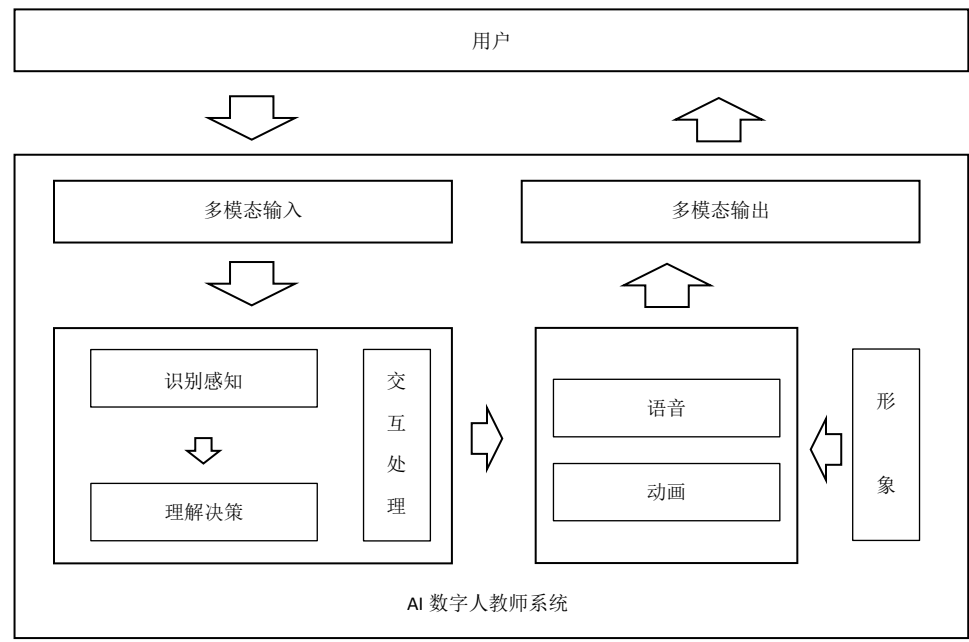


图1 AI 数字人教师系统框架

5 技术要求

5.1 形象

5.1.1 人物特性

非真实形象的AI数字人教师，其面部特征、发型、服饰应美观且符合教育场景；基于真实教师形象创建的AI数字人教师，其面部特征、发型、服饰等应与原型相似。

5.1.2 完好性

AI数字人教师形象应无任何以下破损情况：

- a) 严重的扭曲；
- b) 严重的马赛克；
- c) 明显跳帧；
- d) 其他破损情况。

5.1.3 舒适性

AI数字人教师形象应让用户生理上感到舒适，好感度、自然度、使用愿景三个维度的主观评分得分均不低于4.5分。

5.2 语音

5.2.1 发音准确性

AI数字人教师的发音字准确率应不低于98%，发音句准确率应不低于95%。

5.2.2 韵律准确性

AI数字人教师的韵律准确率应不低于90%。

5.2.3 语音舒适性

AI数字人教师的语音舒适性主观评分得分应不低于表1的规定。

表1 语音舒适性主观评分得分

评测维度	描述	得分
语音语调	整体发音是否标准	4.8
	发音吐字是否清晰	4.8
	断词断句、停顿是否恰当	4.5
	语气语调是否自然	4.5
	重读发音是否得当	4.5
	语速表达是否恰当	4.5
流畅连贯度	语音表达是否流利	4.8
情绪饱满度	按照文本语义和内容，情绪表达是否恰当	4.5
拟人舒适度	声音拟人程度是否和真人一样	4.5
	聆听该声音时，感受是否愉悦	4.8
	是否愿意使用该声音进行服务	4.8

5.3 动画

5.3.1 动作契合度

5.3.1.1 唇形

AI数字人教师应能通过语音驱动技术建立高清的视频人脸唇形，满足单帧时刻同步音素发音的需求，满足前后若干相邻帧时刻音素对当前时刻嘴形的影响。

5.3.1.2 其余动作

AI数字人教师的眉毛与眼皮动作、眼球动作、头旋转动作、上身肢体动作、下身肢体动作、全身动作应符合YD/T 4393.1—2023中5.3.1提供的表3中序号2～序号7的契合度要求。

5.3.2 动作舒适性

AI数字人教师的动作舒适性主观评分得分应不低于表2的规定。

表2 动作舒适性主观评分得分

评测维度	描述	得分
口型匹配度	口型与发音是否匹配	4.8
面部表情自然度	面部表情是否自然	4.5
肢体动作自然度	肢体动作是否自然	4.5

5.4 交互处理

5.4.1 教学功能

5.4.1.1 课程内容生成

AI数字人教师应能依据给定的教材章节和知识点，智能生成完整的教案大纲和讲稿，并符合以下要求：

- a) 教案大纲应涵盖教学目标、教学重难点、教学方法、教学过程：
 - 1) 教学目标：清晰阐述学生在知识、技能、情感态度等方面预期达成的结果，且与课程标准及教材内容相符；
 - 2) 教学重难点：提炼准确，不超过3个重点和3个难点；
 - 3) 教学方法：至少列举2种，并说明其适用性；
 - 4) 教学过程：导入、新授、巩固练习、课堂小结、作业布置等环节安排合理，时间分配明确。
- b) 讲稿内容应准确阐释知识点，语言表达符合学生认知水平，且知识点引用出处标注清晰。

5.4.1.2 智能习题生成

AI数字人教师应能依据教材章节智能生成习题，并符合以下要求：

- a) 知识点相关性：生成的习题应与所指定教材章节的知识点紧密相关；
- b) 题目类型多样性：生成的题目类型应包含选择题、填空题、简答题、论述题等；
- c) 题目难度合理性：按照设定的难度等级生成题目，容易、中等、困难题目数量比例合理。

5.4.2 交互性能

5.4.2.1 语音识别性能

AI数字人教师的对话正确响应率应不低于95%。

5.4.2.2 平均交互响应时间

AI数字人教师的平均交互响应时间应不超过2.0 s。

5.4.2.3 用户交互体验

AI数字人教师的用户交互体验主观评分得分应不低于表3的规定。

表3 用户交互体验主观评分得分

评测维度	描述	得分
交互准确度	在学科知识讲解、习题解答等教学问答场景下，是否能够准确解答学生问题，提供的答案是否符合教学大纲要求	4.8
	在布置作业、指导实验操作等教学任务执行场景中，是否能准确联系上下文，理解学生当前的学习进度，帮助学生顺利完成学习任务	4.5
	是否能准确识别打断	4.5
风格一致性	在整个交互过程中，语音语调、表情动作、语言风格等是否保持前后一致，且符合设定的教师人物特性	4.8

5.5 多模态输出

5.5.1 视频合成实时率

AI数字人教师的视频合成实时率应不超过1.2。

5.5.2 流畅度

AI数字人教师的视频帧率应不低于30帧/秒。

5.5.3 画面准确率

AI数字人教师的画面准确率应不低于98%。

5.5.4 音视频匹配度

AI数字人教师的音视频匹配度应不低于98%。

5.6 整体逼真度

AI数字人教师应能表现出与真人教师等价或无法区分的逼真度，误识率不低于90%。

6 评估方法

6.1 形象

6.1.1 人物特性

通过技术测试进行检查。

6.1.2 完好性

按YD/T 4393.2—2023中4.2.1的规定进行。

6.1.3 舒适性

招募教育学专家及学生代表各10人作为评分者，向评分者展示AI数字人教师形象不低于60 s，按YD/T 4393.1—2023中5.1.4提供的评分量表进行评分后，统计好感度、自然度、使用愿景三个维度的主观评分均值。

6.2 语音

6.2.1 发音准确性

按YD/T 4393.2—2023中4.3.1的规定进行。

6.2.2 韵律准确性

按YD/T 4393.2—2023中4.3.2的规定进行。

6.2.3 语音舒适性

按YD/T 4393.2—2023中4.3.3的规定进行。

6.3 动画

6.3.1 动作契合度

通过技术测试进行检查。

6.3.2 动作舒适性

按YD/T 4393.2—2023中4.4.1的规定进行。

6.4 交互处理

6.4.1 教学功能

6.4.1.1 课程内容生成

招募教育学专家5人对AI数字人教师智能生成的教案大纲和讲稿按5.4.1.1的要求进行通过性审核，5人均投票通过视为满足要求。

6.4.1.2 智能习题生成

招募对应学科的高级教师5人对AI数字人教师智能生成的习题按5.4.1.2的要求进行通过性审核，5人均投票通过视为满足要求。

6.4.2 交互性能

6.4.2.1 语音识别性能

按GB 21023—2007中第7章的规定进行。

6.4.2.2 平均交互响应时间

通过技术测试进行检查，按YD/T 4393.1—2023中5.4.2.2的规定进行计算。

6.4.2.3 用户交互体验

招募教育学专家10人作为评分者，为每位评分者提供不少于1 h的交互体验，按表4进行评分后，统计各项描述的主观评分均值。

表4 用户交互体验主观评分规则

评测维度	评分				
	5	4	3	2	1
交互准确度	完全准确	比较准确	基本准确	不太准确	完全不准确
风格一致性	十分一致	比较一致	基本一致	不太一致	完全不一致

6.5 多模态输出

6.5.1 视频合成实时率

按YD/T 4393.2—2023中4.5.1的规定进行。

6.5.2 流畅度

按YD/T 4393.2—2023中4.5.2的规定进行。

6.5.3 画面准确率

按YD/T 4393.2—2023中4.5.3的规定进行。

6.5.4 音视频匹配度

按YD/T 4393.2—2023中4.5.4的规定进行。

6.6 整体逼真度

按YD/T 4393.2—2023中4.6.1的规定进行。
