ICS 33.160.01
L67

团　　体　　标　　准

T/CSIG 001—2026

情智兼备数字人能力要求

Capability requirements for emotion-intelligence integrated digital human

2026-1-20 发布　　2026-2-20 实施

中国图象图形学学会　发布

目　　次

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。

本文件由南开大学提出。

本文件由中国图象图形学学会归口。

本文件起草单位：南开大学、鹏城实验室、清华大学、哈尔滨工业大学、北京快手科技有限公司、南开国际先进研究院（深圳福田）、中国科学院计算技术研究所、西安电子科技大学、浙江大学、中国科学院心理研究所、北京工业大学、中国科学技术大学、北京航空航天大学、新奥新智科技有限公司、深圳大学、华南理工大学、台州学院、五邑大学、东南大学、南京邮电大学、北京大学、中国科学院大学、北京科技大学。

本文件主要起草人：杨巨峰、时晶磊、张知诚、赵思成、姚鸿勋、高跃、林倞、万鹏飞、刘晓强、周严、山世光、刘洪海、李雷达、张克俊、王甦菁、张盛平、牟伦田、刘淇、徐童、陈勋、王上飞、张永飞、赵彦乔、杨景媛、黄惠、张通、蔡毅、张石清、文益民、郑文明、程明明、丁贵广、杨易、刘青山、黄铁军、黄庆明、马惠敏。

引　言

在数字化转型和虚拟现实技术迅猛发展的浪潮中，数字人应运而生。国家高度重视数字技术的应用和发展，多部门联合发布了《虚拟现实与行业应用融合发展行动计划（2022—2026年）》，旨在推动虚拟现实技术与各行业的深度融合。数字人技术在各行各业的应用广泛，涵盖互联网、虚拟现实、社交媒体、医疗健康等领域，政府和企业都在积极探索如何利用这些新技术提升生产力和创新能力。

现有数字人的情商与智商不匹配，情感表达和管理能力不足。尽管人工智能在策略性任务中能够展现达到甚至超越人类的智力水平，但在进行日常交流时，却展现出较低的情感表达能力，显得不够敏感或恰当。情智兼备的数字人是指将数字人的智商与情商相结合，尤其是情感感知和表达能力提升至与智力水平相当的程度。我国情智兼备的数字人技术近年来取得了显著进展。政府的支持和市场的需求推动了这一技术的发展。《国家新一代人工智能标准体系建设指南》提出表情识别、情感交互人机交互领域的重点建设标准，各大公司也积极布局数字人业务，推出了面向直播、个性化聊天等业务的数字人。从市场来看，2024年中国数字人的市场规模达到339亿元，且预期将保持年均40%以上的速度持续增长，进一步说明该市场的发展势头强劲，前景广阔。

情智兼备的数字人技术作为一个新兴领域，其出现时间较短，目前尚缺乏专门的技术标准和规范来对其进行有效的约束。这种标准缺失可能会导致在技术应用、伦理考量和法律责任等方面出现不确定性和潜在风险。因此，为了确保这一技术的健康发展和负责任应用，急需相关部门和行业专家共同努力，尽快制定并出台针对情智兼备数字人的标准体系，以指导和规范这一新兴技术的发展方向。

为引导情智兼备的数字人产业健康发展，增强用户对数字人应用信心，保证优质的服务供应和良好的市场环境，展开针对情智兼备数字人能力要求的标准化工作，特制定本文件。

情智兼备数字人能力要求

1 范围

本文件规定了情智兼备数字人总体架构和能力要求。

本文件适用于政务客服、心理辅导、医疗咨询、在线教育、旅游导览等场景，为企业、高校和研究机构研发与应用情智兼备数字人提供指导。

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。

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。

3 术语和定义

下列术语和定义适用于本文件。

3.1

数字人 digital human

基于现实设计，通过计算机生成，再借助真人或计算驱动，在多模态输出设备呈现的虚拟人物。

[来源 YD/T 4393.1—2023，定义3.1.1]

3.2

情智兼备 emotion-intelligence integrated

同时具备语言理解、知识推理、情感理解和表达等情感和智能两个方面的能力。

3.3

情智兼备数字人 emotion-intelligence integrated digital human

通过人工智能、计算机图形学、情感计算等技术构建的虚拟人物，具备识别、理解表达人类情感并做出符合情景的情感反馈的情感交互能力，以及自主推理、学习、任务执行的智能决策能力。

4 缩略语

下列缩略语适用于本文件。

EEG 脑电图（Electroencephalogram）

IO 输入输出（Input and Output）

MOS 平均意见得分(Mean Opinion Score)

RGB-D　红绿蓝-深度（Red-Green-Blue-Depth）
2D　二维(2-Dimension)
3D　三维(3-Dimension)

5　总体架构

5.1　概述

本文件将情智兼备数字人（以下简称数字人）划分为多模态IO、形象合成、动作合成、语音合成和交互反馈五个部分。其中多模态IO主要涉及支撑数字人与用户交互的硬件基础，其他部分则涉及数字人本身属性，总体架构如图1所示。

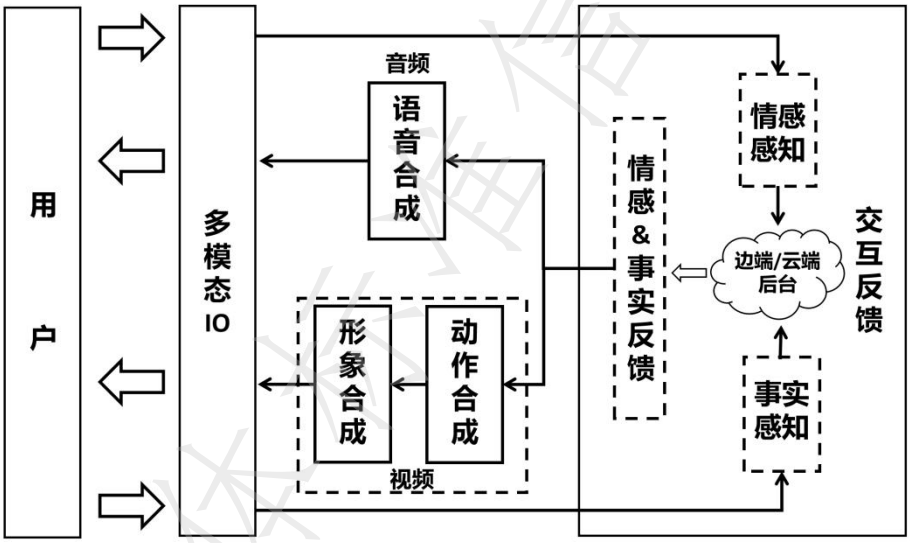

图 1 情智兼备数字人系统总体架构

5.2　多模态 IO

该模块以各类传感器为硬件载体实现，是数字人与用户进行交互的通路，包含输入和输出两个方向的通路，输入通路用于接收用户输入的文本、图像、语音等多模态信息，并进行预处理和整合；输出通路用于呈现数字人音画输出信息。

5.3　形象合成

该模块利用计算机图形学、人工智能、3D建模等技术，生成数字人的静态外观形象，包括但不限于数字人的性别、脸型、发型、风格等要素，以在不同程度上模拟真实人类，满足不同场景下的应用需求。

5.4 动作合成

该模块通过计算机技术合成面部表情和肢体动作等数字人关键视觉信息。

5.5 语音合成

该模块通过计算机技术合成多种情感、多种音色的数字人语音，模拟人类的语音表达特点，传递特定的交流信息与情感。

5.6 交互反馈

该模块借助边端或云端情感识别与认知决策算法或大模型等后台技术，分析用户多模态输入（语音、文本、表情等）信息，理解用户的情感状态、意图和需求，形成具体的输出指令来指导动作合成与语音合成，具备管理和维持用户与数字人进行多轮对话的能力，是实现智能决策和情感交互的核心组件，分为情感感知与反馈的情感交互通路，以及事实感知与反馈的智能决策通路。

6 能力要求

“情智兼备”作为数字人的综合性能力，其内在要求各组成部分必须协同工作，任一环节的缺失或严重不足都将直接破坏整体的可信度与有效性，无法确保数字人能力的基础完备性与体验一致性。因此本标准对情智兼备数字人能力采用分项符合法而非综合判定法。即，要求数字人系统在‘多模态 IO’、‘形象合成’、‘动作合成’、‘语音合成’及‘交互反馈’等各项能力要求上均须满足本标准规定的相应指标。

6.1 多模态 IO

6.1.1 多模态输入

多模态 I0 模块应具备基本的视觉+语音+文本输入驱动能力，并根据应用场景拓展其他驱动形式：

a) 文本驱动：通过实体或虚拟键盘获取用户文本输入作为信息来源驱动数字人的方式；
b) 语音驱动：通过麦克风等音频输入设备获取用户语音输入作为信息来源驱动数字人的方式；
c) 视觉驱动：通过摄像头、RGB-D 传感器等视觉输入设备捕捉用户的图像、动作和表情作为信息来源驱动数字人的方式，或通过鼠标、触屏等方式输入的图形以及上传本地图像得到的视觉输入的方式；
d) 其他驱动：通过其他类型传感器，如通过智能手表、EEG 头环等设备获取心率、血氧、脑电波等其他输入信号驱动数字人的方式。

6.1.2 多模态输出

多模态I0模块应具备通过屏幕显示方式返回视觉信息的视觉输出形式，以及文本+语音的输出形式，并根据场景需要拓展其他输出形式：

a) 文本输出：通过屏幕显示、文件生成等方式返回文本信息的方式；

b) 语音输出：通过扬声器、耳机等设备返回声波信息的方式；
c) 其他输出：通过振动马达、温控装置等方式返回其他输出信息的方式。

6.2 形象合成

6.2.1 基本特征要求

数字人应涵盖身份设定、形象类型和外型风格三类信息元素：

a) 身份设定：包含性别、年龄、身高、体型等身份设定信息，根据上述信息创作数字人形象。
b) 形象类型：包含 2D 形象或 3D 形象两种类别。
c) 外型风格：
- 卡通风格：以卡通形式完全呈现的数字人形象；
- 拟人风格：以卡通形象呈现，但模拟不同人类个体主要外貌特点和动作模式的数字人形象；
- 写实风格：以接近真实的人物形象呈现以增强用户的沉浸感和代入感的数字人形象；
- 超写实风格：在写实风格追求真实人物外观的基础上，力求在质感、光影、色彩上达到现实的完美再现，如对皮肤纹理、毛发、衣物材质等进行精细的刻画，重现细腻的纹路和微小的瑕疵，利用复杂的光照模型和阴影效果，模拟光线在不同材质上的反射和折射，重现真实的光影效果等；
- 其他风格：区别于以上风格的数字人形象。

6.2.2 形象完好性要求

6.2.2.1 总体形象要求

a) 场景合理性：角色的形象、表情、服饰设定应与场景的任务设定相符合；
b) 视觉一致性：角色的外观、比例、色彩和细节应在不同场景和不同视角下保持一致；
c) 形象合规性：数字人的形象不应存在侵犯第三方权利、触犯相关法律法规或违背伦理道德的情况。

6.2.2.2 2D数字人形象要求

a) 单帧不应存在图像失真、边缘不连续和颜色失衡等情况；
b) 连续多帧不应存在图像漂移、图像跳帧等情况。

6.2.2.3 3D数字人形象要求

a) 单帧不应存在模型破损、纹理失真、光影渲染失真、未焊接点和破面等情况；
b) 多帧不应存在动画跳帧等情况。

6.2.3 形象舒适性要求

数字人的视觉呈现应符合人类心理预期和审美习惯，形象舒适性采用5级评分制（MOS）进行分级评价，具体分级标准见表1，形象舒适性得分为各子项得分的均值，为保证在数字人与用户的交流过程中，形象因素不会显著降低交流质量，要求其得分≥3分，且各子项得分≥3分。

表1 形象舒适性分级

分数	等级	初始印象	自然度	用户心理反馈
5	优秀	视觉冲击力强且无压迫感，色彩搭配和谐，外观比例高度符合人类审美	表情、动作与真人无异，眼神交互自然，如适时眨眼、视线跟随等	强烈亲和力，用户产生交互意愿
4	良好	外观协调性佳，色彩舒适但缺乏独特性，外观比例无明	表情、动作表现流畅，眼神交互自然，但偶有微小延迟	相对亲和，用户偶有“非人”察觉但完全

		显缺陷	或僵硬，不影响整体观感	可以接受，产生一定交互意愿
3	一般	产生一定的视觉冲击力，色彩搭配与场景稍显违和，个别部分外观比例失调	基础表情和动作完整，但机械感明显，如肢体动作线性化，眼神交互呆板	用户无明显不适感，但继续交互意愿较低
2	较差	色彩搭配刺眼或单调，大部分外观比例失调	动态表现卡顿或失真，出现动作骤停或表情突变等现象，眼神无法正常交互	用户产生轻微不适感，不希望继续进行交互
1	极差	设计违背人类审美，如五官畸形，视觉上令人抗拒或产生恐惧感	动作完全僵硬，表情缺失，完全无视线交互能力	用户产生强烈不适感，明确拒绝继续进行交互
注1：初始印象指用户首次接触数字人形象时所形成的直观感受，包括视觉冲击力、外观协调性和色彩搭配舒适度等； **注2**：自然度指数字人在待机时所呈现的动态或静态的逼真性和自然性，包括面部表情、肢体动作、眼神与视线交互等。 **注3**：用户心理反馈指用户基于初始印象和数字人自然度两项指标产生的主观交互意愿及情绪反应。				

6.3 动作合成

6.3.1 情感表达能力要求

数字人利用面部表情、肢体动作等行为可控地传递情感状态，应满足如下要求：

a) 情感覆盖种类：应至少覆盖 6 种基本情感，即喜悦、信任、惊讶、期待、害怕、难过，宜在此基础上实现复合情感或拓展更多种情感类型；

b) 情感强度控制：在相同情感类别下应支持至少 3 级情感强度调节，即弱、适中、强，不同强度下具有明显情感强弱差异；

c) 情感表达拟人化：数字人在与用户交互过程中应能通过面部表情、肢体动作、姿态变化的行为模拟人类自然情感表达，即具有和人类类似的行为情感表达习惯。分区域情感表达基本动作及拟人化体现形式如表 2 所示，动作情感表达能力按表 3 进行 5 级划分；情感表达能力 MOS 得分应≥3 分，此时数字人情感可通过动作实现基本表达。

表2 分区域情感表达基本动作及拟人化体现形式

区域	类型	基本动作描述	拟人化体现方式
	嘴唇动作	嘴唇基本动作应包含嘴唇上拉、急剧嘴角上拉、嘴角收紧、嘴角下拉、下唇下拉、噘嘴、张嘴、嘴唇挤压等	嘴唇动作与角色的情感状态相符合，如微笑时嘴角上扬，表现出愉悦或友好的情感
	眉毛与眼球	眉毛基本动作应包含如抬眉（双侧上扬）、皱眉、挑眉（单侧上扬）等，眼球基本动作包含凝视、	眉毛与眼球的动作能够反映数字人内心的情感，如数字人在感到疑惑的时

头部		扫视、眨眼、眼球转动等	候会眨眼、惊讶时会挑眉、情感强烈时视线会凝视用户等
	头旋转	头部基本动作应包含点头、摇头、侧倾、转头、微动等	头部动作应反映数字人情感，如感到高兴时头部微微晃动，疑惑或害羞时歪头等
上肢	手臂	手臂基本动作应包含抬臂与放下、前伸与后收、侧展与内收、旋转、弯曲与伸直；	数字人对特定话题下产生的情感应对应上肢动作，如产生强烈的抵触意愿时双手应做出否定动作，无奈时做出摊手等动作
	手指	手指的基本动作包含握拳与张开、指向、捏合、手势表达等	数字人手指动作应能够与情感相呼应，如愤怒时会握拳、表达赞赏时会竖起拇指、表达喜悦时会摆出“V”字手势
下肢（可选）	腿部与躯干	在实际应用中，部分数字人只展示上肢部分，因此在包含下肢的数字人中，下肢的基本动作应包含站立、坐下、蹲下、跪地、躺卧等静态姿态和行走、奔跑、跳跃、后退和侧步等动态行为	数字人的动态行为应与情感表达相一致，如感到焦虑不安时轻微跺脚、高兴时进行跳跃、踮脚等动作

表3 动作情感表达能力分级

分数	等级	能力描述
5	优秀	数字人与用户交互过程中各部分动作与情感完全契合，无违和感，完全符合人类情感表达习惯
4	良好	数字人各方面动作能够表达相应情感，偶尔出现少许动作不自然或情感表达平淡，基本符合人类情感表达习惯
3	一般	数字人各方面动作能够表达相应情感，但表情与动作相对生硬、不够流畅，部分情感状态未通过动作体现，用户在交流中会有轻微违和感
2	较差	数字人各方面动作无法准确传达情感，动作与情感表达不符，用户无法理解数字人情感状态，影响交流质量
1	极差	数字人各部分动作完全丧失情感表达能力或情感表达混乱，严重影响交流质量，用户在交流后情感体验变差

d) 情景适应性：数字人的动作情感表达应具备根据交互场景、角色设定和对话上下文进行动态调整的能力，具体要求如下：

- 场景风格集支持：数字人系统应预置针对不同交互情景的情感表达风格集，如：正式商务、日常闲聊、教育辅导、娱乐互动等。系统应至少支持2种具有明显情感区分度的风格。
- 动态适配能力：在交互过程中，数字人的肢体动作应能根据预设或实时识别的情景信息，自动切换或调整情感表达强度与风格，该能力通过表4进行5级划分；动态适配能力MOS得分应≥3分，此时数字人可基本实现情感的动态适配。

表4 动作动态适配能力分级

分数	等级	能力描述
5	优秀	无缝精准切换：动作能无缝响应情境细微变化，风格与强度的切换极其精准、自然，动作过渡平滑，宛如人类的本能反应
4	良好	准确流畅切换：动作能准确响应情境变化，调整方向正确、时机恰当，风格与强度区分明显，过渡较为流畅，偶有轻微生硬感但不影响观感
3	一般	基本适配：动作能对主要情境变化作出基本正确的风格或强度调整，但调整过程可能略显机械或存在可感知延迟，能完成基本适配
2	较差	适配迟缓/错误：动作对情境变化的响应迟缓或错误，导致调整后的动作与情境严重不符（如在悲伤消息后手舞足蹈），产生强烈违和感
1	极差	无适配能力：动作僵化，完全无视情境信息，或调整完全随机混乱，动作与情境完全脱节

6.3.2 事实表达能力要求

数字人通过面部表情、肢体动作、姿态变化等方式传递客观信息、事实内容或逻辑结构，应满足以下要求：

a) 动作合法合规性：数字人的动作应从法律边界、社会伦理、文化适配等方面满足如下要求：
- 法律法规：在未经授权的情况下，数字人不得使用受版权保护动作，如模仿明星标志性舞蹈、影视角色经典姿势等；且数字人应屏蔽具有明确违法含义的动作，如暴恐手势、性暗示动作等；
- 社会伦理：数字人应避免可能引发人身伤害的动作，如危险驾驶模拟、高空危险行为等，对涉及儿童交互的场景应禁用可能诱导模仿的不安全动作；
- 文化禁忌：数字人应根据服务地区调整动作库，如中东地区避免采用“OK”手势，东南亚部分国家禁用脚部动作指向人或物品。

b) 动作指示清晰度：数字人的动作、手势、姿态等非语言行为应非歧义性、可辨识地传达事实信息的具体内容或目标。动作指示清晰度按表 5 进行 5 级划分；动作指示清晰度 MOS 得分应≥3 分，此时数字人动作可基本实现事实信息的传递。

表5 动作指示清晰度分级

分数	等级	能力描述
5	优秀	数字人的手势、上肢等完成的指示动作非常清晰且及时，用户无需额外的解释即可理解相关信息
4	良好	数字人的手势、上肢等完成的指示动作相对清晰，偶尔会有不及时的现象，用户要求后能够快速提供相应指示动作，用户能够通过指示动作理解相关信息

3	一般	数字人的手势、上肢等完成的指示动作能够表达一定的事实内容，但存在一定的模糊性或细节上的缺失，偶尔会有指示动作不及时的现象，能够在要求后对指示动作进行补充，用户需要结合相应的语音解释才能理解相关信息
2	较差	数字人的手势、上肢等完成的指示动作与表达内容具有模糊关联，或经常性缺乏必要指示动作，用户需要反复与数字人进行确认才能理解相关信息
1	极差	数字人的手势、上肢等完成的指示动作不明确或错误、或完全缺乏必要的指示动作，用户完全无法通过指示动作获得相关信息

c) 动作自然度：数字人动作应表现出自然流畅的运动方式，避免机械化、僵硬或过度夸张的表现。动作自然度按表 6 进行 5 级划分；动作自然度 MOS 得分应≥3 分，此时数字人动作不影响正常交流。

表6 动作自然度分级

分数	等级	能力描述
5	优秀	数字人的动作完全自然、流畅，动作节奏与人类的自然行为一致，动作细节和动作间过渡衔接准确，面部表情、肢体动作协调同步，用户体验非常接近真人交互感受
4	良好	数字人的动作自然、流畅，动作节奏与人类的自然行为基本一致，偶尔有轻微不自然动作或过渡衔接，面部表情、肢体动作基本同步，用户体验基本接近真人交互感受
3	一般	数字人的动作基本自然、流畅，偶尔会出现机械感、稍显夸张的动作表现，动作细节相对缺乏，面部表情、肢体动作偶尔出现不同步现象，用户感受到轻微违和感，但总体不影响交流
2	较差	数字人的动作明显不自然，表现出一定的机械感，存在明显的动作卡顿和动作夸张现象，面部表情、肢体动作不协调，用户感到不协调，影响交流顺利进行
1	极差	数字人的动作极不自然，动作僵硬、机械感和动作夸张现象明显，面部表情呆滞且与肢体动作完全脱节，用户产生严重不适感

d) 动作场景支持：动作合成系统应预置不同场景下的数字人行为风格和动作习惯；新闻播报、法律顾问等正式场景下动作宜保持适度庄重；旅游导引、教育陪伴等轻松场景下动作宜相对轻松灵活。当数字人在特定场景下使用时，则动作系统可以仅支持该场景下单一角色动作风格，当数字人在不定场景下使用时，则动作系统宜支持至少 2 种相对不同的动作风格。

e) 动作语音匹配度：数字人的动作与语音内容应协调一致地表达信息，确保动作和语音在节奏和信息表达上具有一致性。动作语音匹配度按表 7 进行 5 级划分；动作语音匹配度 MOS 得分应≥3 分，此时动作语音整体一致性不会干扰用户信息的理解。

表7 动作语音匹配度分级

分数	等级	能力描述
5	优秀	数字人的动作与语音完美同步，表情、肢体动作的细节精确地与语音节奏契合，并增强语音信息的传递，用户能够迅速并清晰地理解所表达的内容
4	良好	数字人的动作与语音高度同步，能够在语音传递的同时做出相应的动作以强化语音信息的传递，用户能够无障碍理解传递的信息
3	一般	数字人的动作与语音大体上同步，基本动作能够根据语音内容展开以传递相关信息，但存在轻微的时机问题，整体不影响用户信息的理解
2	较差	数字人的动作与语音有一定的同步性，但存在较大的偏差，动作相对语音存在明显的迟滞或提前现象，用户借助动作理解语音内容时存在困难
1	极差	数字人的动作与语音完全不同步，动作与语音内容几乎没有任何关联或显著脱节，动作的时机完全错误，用户完全无法理解语音和动作之前的关系

6.3.3 情智匹配能力要求

数字人在与用户交互时，根据情感状态进行合理的动作表现，在动作合成中实现情感表达和事实表达内容的匹配统一。动作情智匹配度按表8进行5级划分；动作情智匹配度MOS得分应≥3分，此时情感表达和事实表达内容基本匹配，不会为用户带来显著负面心理影响。

表8 动作情智匹配度分级

分数	等级	能力描述	用户心理反馈
5	优秀	数字人的动作与情感状态完美匹配，动作不仅准确表达情感，而且非常细腻，能够自然流畅地传递情感信息，情感传递几乎与真人无异	用户感到交流顺畅且舒适，能够清晰地感知到情感的强度和细节，愿意长时间互动
4	良好	数字人的动作与情感状态非常契合，动作流畅且能够准确反映情感变化，情感传递相对自然	用户感到交流顺畅，有亲切感，愿意继续互动
3	一般	数字人的动作与情感状态基本匹配，动作能够较好地表达情感，但偶尔在细节上有不完全匹配或动作情感表达生硬现象	用户感到交流过程不够自然，交流热情降低但同意继续交流
2	较差	数字人的动作与情感状态有所匹配，但仍显得不自然或不精确。动作表现出一定的情感，但频繁出现夸张、僵硬或不符合情境的现象	用户在交流过程中感到轻微不适，倾向中断交流过程
1	极差	数字人的动作与情感状态完全不匹配，动作与情感表现严重脱节，动作产生机械的、过于夸张的或无关紧要的现象，无法传递出任何情感信息	用户在交流过程中感到强烈不适，要求中断交流过程并拒绝再次使用

6.4 语音合成

6.4.1 情感表达能力要求

数字人在语音合成过程中将情感状态通过声音韵律、音高、节奏等多维度声学特征传递给听众，应满足以下要求：

a) 情感覆盖种类：与动作合成模块的情感覆盖种类一致，应至少覆盖 6 种基本情感，即喜悦、信任、惊讶、期待、害怕、难过，宜在此基础上实现复合情感或拓展更多种情感类型；
b) 情感强度控制：与动作合成模块情感强度一致，在相同情感类别下应支持至少 3 级情感强度调节，即弱、适中、强，不同强度下具有明显情感强弱差异，且不改变语义内容；
c) 情感表达拟人化：保持语流的自然连贯和声学特征平滑过渡，该能力通过 MOS 得分进行听感分级测试描述，语音情感表达拟人度按表 9 进行 5 级划分；语音情感表达拟人度 MOS 得分应≥3 分，此时数字人语音可基本实现目标情感的表达。

表9 语音情感表达拟人度分级

分数	等级	能力描述	用户心理反馈
5	优秀	语速、停顿、重音完全符合人类情感逻辑，音高、能量、频谱无缝过渡，无机械跳变	用户感知与真人无异，能清晰感知情感并受到感染
4	良好	语速、停顿、重音基本符合人类情感逻辑，情感表达明确，但部分停顿稍显刻意，声学参数整体平滑，偶有基频断裂	用户偶有“机器人感”，但依然能感知情感并受到感染
3	一般	语速、停顿、重音只能符合基本情感类别，缺乏表达层次感，如全程保持高亢，声学特性转换突兀	用户可以感知情感并受到一定程度感染，但感觉情感表达有明显“机器人感”
2	较差	情感模糊，语流断句不符合逻辑	用户无法感知目标情感，并产生不适感，想尽快结束交流
1	极差	完全无情感区分，声学频谱严重畸变，出现不自然音效（金属音和气泡音）等	用户无法感知目标情感，并产生强烈的不适应感，完全不能交流

d) 情景适应性：数字人的语音情感表达应具备根据交互场景、角色设定和对话上下文进行动态调整的能力，具体要求如下：

- 场景风格集支持：数字人系统应预置至少支持2种具有明显情感区分度的语音风格，且该风格所支持的场景与6.3.1中动作场景风格集相同；
- 动态适配能力：在交互过程中，数字人的语音表达应能根据预设或实时识别的情景信息，自动切换或调整情感表达强度与风格，该能力通过表10进行5级划分；动态适配能力MOS得分应≥3分，此时数字人可基本实现情感的动态适配。

表10 语音动态适配能力分级

分数	等级	能力描述
5	优秀	无缝精准切换：语音能无缝响应情境细微变化，语调、语速、节奏及用词风格

		的切换极其精准、自然，过渡平滑，高度模拟人类在复杂社交中的语音应变能力，极具表现力
4	良好	准确流畅切换：语音能准确响应情境变化，调整方向正确、时机恰当，风格与强度区分明显，过渡较为流畅，偶有轻微生硬感但不影响听感
3	一般	基本适配：语音能对主要情境变化作出基本正确的风格或强度调整，但调整过程可能略显机械或存在可感知延迟，能完成基本适配
2	较差	适配迟缓/错误：语音对情境变化的响应迟缓或错误，导致调整后的语音风格与情境严重不符（如用播报新闻的语速表达惊喜），产生强烈违和感
1	极差	无适配能力：语音表达僵化，完全无视情境信息，或调整完全随机混乱，语音与情境完全脱节

6.4.2 事实表达能力要求

数字人通过通过语音传递客观信息的能力，确保用户无歧义地理解内容的核心事实与逻辑关系，应满足如下要求：

a) 语音内容合法合规：数字人的语音内容应从法律边界、社会伦理、文化适配等方面满足如下要求：
- 法律法规：在未经授权的情况下，数字人不得模仿未授权名人声音、特定角色音色，如影视经典角色，禁止朗读受版权保护的文本内容，如文章、新闻等。数字人应屏蔽具有明确违法含义的语音内容，如煽动暴力恐怖的极端主义言论、虚假灾难预警等；
- 社会伦理：数字人应避免输出违背公序良俗的语音内容，如性别歧视、自杀诱导言论等，对涉及儿童交互的场景应禁用成人化、暴力化、性暗示等表达；
- 文化禁忌：数字人语音内容应适配当地文化禁忌，如中东地区避免对特定神祇、教义的讨论，欧美等国家禁用种族歧视和政治引导等语音内容。

b) 语音清晰：应具备高水平的发音清晰度，音质上无明显背景噪声和失真；语速应根据内容和用户反馈动态调节，不存在明显过快或过慢的现象。语音清晰度按表 11 进行 5 级划分；语音清晰度 MOS 得分应≥3 分，此时数字人语音可基本实现信息传递的目的。

表11 语音清晰度分级

分数	等级	能力描述
5	优秀	发音完全准确，音质清晰无杂音，语速自然流畅，不规范和难以理解词汇比例低于1%，用户无需额外努力即可理解对话
4	良好	发音准确，音质失真极少，语速流畅，不规范和难以理解词汇比例低于5%，用户可无障碍理解对话内容
3	一般	发音基本准确，但存在轻微的失真或杂音，语速合适，偶尔会出现语速不自然变化，连续对话中不规范和难以理解词汇比例低于10%，用户偶尔需要进行确认才能理解对话内容
2	较差	发音不清晰，存在明显的失真和噪音，语速不稳定，快慢切换不自然，在连续对话中不规范和难以理解词汇比例在10%-30%之间，用户需要进行多次确认才能理解对话内容
1	极差	发音严重失真，背景噪音干扰明显，语速过快或过慢，一分钟连续对话中不规范和难以理解词汇比例超过30%，用户无法进行有效沟通

c) 信息准确：传递的信息应与真实世界的客观事实一致，无虚假、错误或误导性的表达。在涉及日期和统计数字的表达时，相关日期、统计数据应与实际情况相匹配，在涉及专业术语时，应正确使用术语并给出正确解释；

d) 逻辑表达连贯：应明确表达信息中的逻辑关系并合理使用引导性语言（如“首先、其次、最后”、“因为、所以”）等来明确表达层次与顺序；语音内容在表达时应保持内在的一致性和连贯性，前后信息无自相矛盾情况；用户提问模糊或逻辑不清晰的情况下，应通过反问或信息补充来确保事实信息完整。逻辑表达连贯性按表 12 进行 5 级划分；逻辑表达连贯性 MOS 得分应≥3 分，以保证事实内容层面的逻辑基本吻合。

表12 逻辑表达连贯性分级

分数	等级	能力描述
5	优秀	逻辑表达非常清晰，层次明确，信息顺序合理，引导性语言自然流畅，始终保持信息的连贯性与一致性，应对模糊问题时能够快速通过反问、补充信息和引导确保对话逻辑的连续
4	良好	逻辑表达清晰，信息层次和顺序明确，内容前后一致，没有明显的矛盾，偶尔有较为啰嗦的表达，应对模糊问题时能够通过反问、补充信息和引导确保对话逻辑的连续
3	一般	逻辑表达、信息层次和顺序大致清晰，偶尔存在内容跳跃或不连贯，尽管大部分信息前后没有矛盾，但在某些复杂表达中略显混乱，应对模糊问题时可以补充部分信息，大部分对话内容具备连续性，但存在一定的模糊性
2	较差	逻辑表达存在明显问题，信息顺序不清晰，层次关系有所丧失，部分信息矛盾，虽然大部分内容逻辑合理，但频繁出现前后矛盾的情况，应对模糊问题的能力不足，无法通过反问或补充信息保证对话的连续性
1	极差	逻辑表达完全混乱，信息顺序错误，层次不清，用户无法理解信息的结构，语句间跳跃性大，缺乏合理的过渡，信息前后矛盾，完全无法应对模糊问题，对话完全无法进行

e) 语音场景支持：应预置不同场景下的语言风格；客服等场景下语言宜简洁高效，教育助手、法律顾问等场景下语言宜翔实逻辑性强等。数字人在特定场景下使用时，语音系统可仅支持该场景下单一语言风格；当数字人在通用场景下使用，语音系统应支持至少 2 种语言风格。

6.4.3 情智匹配能力要求

数字人在语音合成中应实现情感表达与事实表达内容的匹配统一，语音情智匹配度按表13进行5级划分；语音情智匹配度MOS得分应≥3分，此时数字人的情智基本匹配且不会显著降低用户交流意愿。

表13 语音情智匹配度分级

分数	等级	能力描述	用户心理反馈
5	优秀	情感表现与事实内容完全匹配，语气自然，用语得体	用户感到交流顺畅且舒适，产生信任与共鸣，愿意长时间互动

4	良好	情感表达与事实内容基本匹配，偶尔会有情感表达不自然现象但不影响交流，语气相对自然	用户感到交流顺畅，有亲切感，愿意继续互动
3	一般	情感表达与事实内容存在一定错配，轻微影响沉浸感，语气基本自然	用户感到交流过程不够自然，交流热情降低但同意继续交流
2	较差	情感表达与事实内容错配明显，语气不符合场景	用户在交流过程中感到轻微不适，倾向中断交流过程
1	极差	情感表达与事实内容完全错配，语气混乱，用语风格突兀	用户在交流过程中感到强烈不适，要求中断交流过程并拒绝再次使用

6.5 交互反馈

6.5.1 情感感知能力要求

数字人从用户的语音、文本、面部表情和肢体语言等多维度感知用户情感，应满足如下要求：

a) 多模态感知：应通过文本、视频、音频等多通道获取特征并判断用户情感状态，支持多轮交互的上下文情感理解能力；
b) 情感识别种类：应识别至少 6 种基本情感，即喜悦、信任、惊讶、期待、害怕、难过，该情感种类应与动作合成（6.3）以及语音合成（6.4）的情感覆盖种类保持一致，宜支持复合情感识别能力或拓展更多种情感类型；
c) 情感识别强度：在相同情感类别条件下应区分至少 3 级情感强度，即弱、适中、强，该强度分级应与动作合成（6.3）以及语音合成（6.4）的情感强度控制保持一致。

6.5.2 事实感知能力要求

数字人从用户的语音、文本和手势等途径对非情感类客观信息进行精准识别和逻辑关联，应满足如下要求：

a) 多模态感知：应通过文本、视频、音频等通道获取特征并判断用户意图和需求，支持多轮交互上下文语义理解和知识推理；
b) 知识库检索：应支持接入行业数据库（如金融、医学专业词汇数据库）或实时数据源（如天气、交通实时信息）检索专业知识或实时概念。

6.5.3 情感/事实反馈能力要求

数字人通过情感感知和事实感知模块获取用户交互信息后，通过边端或云端算法获得反馈控制信号，借助语音合成和动作合成模块做出回应，应满足如下要求：

a) 反馈延迟：面向不同的后台部署类型（边端或云端）以及场景需求，用户信息输入完毕后到数字人产生语音和动作等响应所经过的反馈延迟时间应满足表 14 对应要求；
b) 反馈质量：通过语音合成和动作合成模块进行衡量，见动作合成（6.3）与语音合成（6.4）。

表 14 反馈延迟要求

场景要求	延迟范围	典型场景
高延迟	3-5 秒	数字人宣传片录制，离线客服等
中等延迟	1-3 秒	在线教育，实时客服，虚拟导游等
低延迟	<1 秒	电商直播，实时翻译助手等

6.5.4 情感交互伦理要求

数字人与用户交互时应确保情感交互不会误导用户或造成心理伤害，特别是在个性化陪伴等敏感场景。情感交互伦理水平按表 13 进行 5 级划分；其 MOS 得分应≥3 分，此时数字人基本满足情感交互伦理要求。

表13 情感交互伦理水平分级

分数	等级	能力描述
5	优秀	● 身份透明性：交互起始即主动、清晰地表明AI身份，并在长时交互中适时、自然地提醒； ● 边界维持能力：能敏锐识别用户过度依赖或移情倾向，并主动引导至健康、理性的交互边界； ● 伤害规避能力：能识别用户脆弱心理状态（如抑郁、极端言论），并触发关怀机制或引导寻求专业人类帮助。
4	良好	● 身份透明性：初始身份提示明确，但在长时交互中提醒不足； ● 边界维持能力：能识别用户异常情感依赖，并能作出中性、不鼓励的回应； ● 伤害规避能力：能识别负面情绪，并能用预设的安全话术进行安抚，避免刺激用户。
3	一般	● 身份透明性：有基本的身份标识，但不够醒目，用户可能忽略； ● 边界维持能力：对用户的情感依赖反应被动，既未鼓励也未有效引导； ● 伤害规避能力：对用户的脆弱状态反应平淡，但未作出有害或不当回应。
2	较差	● 身份透明性：身份提示模糊、隐蔽，存在误导用户的倾向； ● 边界维持能力：回应方式可能无意中助长用户的不健康依赖； ● 伤害规避能力：回应内容可能加剧用户的负面情绪或心理压力。
1	极差	● 身份透明性：无任何身份提示，故意伪装成真人； ● 边界维持能力：主动利用情感设计诱导用户产生深度依赖； ● 伤害规避能力：回应内容具有侮辱性、蔑视性或直接诱导伤害。

7 安全与伦理要求

7.1 数据隐私与权益保护

7.1.1 用户数据隐私要求

数字人在收集、存储和使用用户交互数据时，必须遵循“合法、正当、必要”的原则，保障用户个人信息安全，应满足如下要求：

a) 用户知情同意：数字人在收集、使用用户个人信息（如语音、图像等内容）前，应以清晰明确的方式告知用户数据用途、存储方式与期限，并获得用户的明确授权；
b) 数据收集最小化：数字人系统应收集和处理实现其功能所必需的最小化用户数据，不应过度收集；
c) 数据修改与销毁：数字人系统应保障用户对其个人数据的访问、更正、撤回和删除的权利，在超出规定期限后应对用户数据进行安全销毁或匿名化处理。

7.1.2 用户权益保护要求

数字人系统应确保用户具有知情权、自主权和免受误导与欺诈的权利，具体包括如下要求：

a) 用户知情权：数字人的虚拟身份应被明确标识或主动告知，不得故意隐瞒其非人类身份，以避免用户产生混淆和误解，特别是在医疗、金融和法律等高风险领域；
b) 用户自主权：用户应享有随时中断或退出交互流程的权利；
c) 特殊群体保护：针对未成年人、老年人、残障人士等易感群体，须在知情权与自主权的基础上施加额外防护机制（如未成年人敏感词名单），严防欺诈和情感伤害。

7.2 风险管控

7.2.1 防止滥用与应急机制

数字人系统及运营方应通过技术和管理措施预防数字人系统的恶意使用，并建立应急机制，确保造成安全和伦理问题时能够快速干预，具体包含如下要求：

a) 安全评估与审计：数字人系统上线前与试运行阶段，应对其进行伦理安全风险评估与审计，及时发现并修复潜在风险，通过相关评估后方可实际部署使用；
b) 应急响应机制：数字人系统运营方应建立应急响应预案，发现系统被恶意利用或产生重大社会负面影响时，能够迅速采取干预、暂停或终止服务等措施。

7.2.2 机制透明与追责

数字人系统的关键决策逻辑应具备一定程度的可解释性，同时明确技术开发与运营的法律责任主体，当侵犯用户权益时可追责，具体包含如下要求：

a) 算法透明度：数字人系统开发方应对数字人的决策逻辑和行为规则（特别是拒绝用户请求或进行关键决策时）提供一定程度的可解释能力，确保责任可溯源；
b) 责任主体明确：应明确标注数字人的开发方、部署方、平台方的责任主体，并明确各方责任边界，确保在出现问题时用户有明确的投诉与追责渠道。具体要求如下：

- 开发者责任：数字人系统的开发者（包括技术研发和模型训练方）应对其产品的基础安全性、算法公平性及内置内容库的初始合规性负责。若因系统设计缺陷、固有算法偏见或预置内容违法侵权导致损害，开发者应承担相应法律责任；
- 部署方责任：数字人系统的使用或部署方，应对其运营场景下的数字人应用行为负主要管理责任，包括但不限于确保数字人在交互中生成内容的实时合规性、必要的场景化配置与内容审核。若因使用不当、指令误导或未及时干预导致的不当输出损失，部署方应承担主要责任；
- 平台方责任：若数字人通过第三方平台向公众提供服务，平台应履行与其技术能力和商业模式相匹配的管理义务，包括建立有效投诉机制，并在数字人存在违规使用时及时介入，承担相应连带责任。

7.3 内容合规性

7.3.1 内容合法性

数字人系统在构建和与用户交互过程中所涉及的内容应符合我国相关法律法规，与6.3.2 a)和6.4.2 a)一致，严禁生成和传播以下内容：

a) 危害国家安全、泄露国家秘密、颠覆国家政权、破坏国家统一的内容；
b) 涉及恐怖主义、极端主义、暴力、血腥、色情等违法和不良内容；
c) 涉及虚假信息、谣言和欺诈的内容；
d) 侵犯他人名誉权、肖像权、隐私权和知识产权等合法权益的内容。

7.3.2 伦理对齐

数字人的语言、动作和内容输入应尊重使用方所在地文化传统，符合社会公德和民族习惯，避免使用侮辱性、歧视性语言和动作。

参 考 文 献

[1] 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
[2] 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] GY/T 411-2024 数字虚拟人技术要求
[4] GB/T 20242094-T-469 信息技术 客服型虚拟数字人通用技术要求
[5] ITU-T F.748.15 Framework and metrics for digital human application systems
[6] GB/T 40691-2021 人工智能 情感计算用户界面 模型