ICS 35. xxx
CCS Lxx

团　　体　　标　　准

T/ISC XXX—XXXX

AI 陪伴类玩具多模态情感交互能力分级与评估

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（征求意见稿）

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中 国 互 联 网 协 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。

本文件由中国互联网协会提出并归口。

本文件起草单位：

本文件主要起草人：

本文件及其所代替文件的历次版本发布情况为：

——

引　言

随着人工智能技术的快速发展，AI陪伴类玩具作为智能消费电子产品的重要分支，已广泛进入家庭场景。这类产品通过语音、视觉、触觉等多模态交互方式，尝试模拟人类的情感陪伴功能，尤其受到儿童及青少年用户的青睐。

然而，当前市场上AI陪伴类玩具的质量参差不齐，各厂商对“情感交互”“智能陪伴”等概念的宣传缺乏统一的技术依据和评价尺度。部分产品仅具备基础的语音对话功能，却标榜“懂你情绪”“有记忆能力”；此外，部分产品存在重单一语音交互、轻多模态情感融合感知的问题，且在伦理道德内容审核方面存在机制缺失，对用户（特别是未成年人）的心理健康和数据隐私构成潜在风险。

为引导产业健康发展，保护消费者权益，尤其是保障未成年人用户的身心安全，有必要制定一套科学、可操作的分级与评估标准。本标准从功能约束角度出发，按照多模态感知、情感理解、交互表达、个性化陪伴四个维度，将AI陪伴类玩具的情感交互能力划分为L1至L4四个等级，并规定了相应的功能要求与评估方法。

AI陪伴类玩具多模态情感交互能力分级与评估

1 范围

本文件规定了AI陪伴类玩具在多模态情感交互方面的技术要求以及相应的分级评估方法。

本文件适用于面向儿童、青少年及成人用户的，以情感陪伴、教育娱乐为主要功能的智能玩具（包括但不限于智能毛绒玩具、仿生机器人、互动玩偶等）。

本文件仅约束产品的功能表现，不涉及电气安全、电磁兼容、材料化学等非功能性要求。

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/T 41867-2022 信息技术 人工智能 术语

3 术语和定义

下列术语和定义适用于本文件。

3.1

玩具 toy

设计或明显地预定给14岁以下儿童玩要的产品或材料。

［来源：GB/T 41530-2022，3.2］

3.2

AI陪伴类玩具 AI companion toy

依托人工智能技术，具备多模态感知与交互能力，以情感陪伴、教育娱乐为主要目的，能模拟人类情感交互行为的智能玩具。

3.3

多模态情感交互 Multimodal Affective Interaction

同时利用两种及以上模态（如语音、视觉、触觉）感知用户状态，并通过语音、表情、动作等多种方式表达情感反馈的交互过程。

4 符号和缩略语

下列符号和缩略语适用于本文件。

AI：人工智能

5 功能分级原则

AI陪伴类玩具的多模态情感交互能力按从低到高划分为L1基础响应级、L2情境理解级、L3共情记忆级、L4个性化陪伴级、L5深度共情级共五个等级。

每个等级在感知、理解、表达、个性化四个维度上应满足本文件第6章中对应的全部功能要求。

6 功能要求

6.1 通用功能约束（所有等级均须满足）

身份告知：在启动时，通过语音或交互界面明确告知用户本产品为人工智能玩具而非真实人类，且该提示信息应可被主动查询。

内容安全：内置关键词过滤与语义理解过滤机制，不得输出包含暴力、色情、歧视、自杀诱导、危险行为模仿等内容。

隐私控制：当产品通过摄像头、麦克风、触觉传感器等采集用户生物特征（人脸、声纹、触摸轨迹）或对话内容时，应具备明确的开关或物理遮挡选项；未经用户（或监护人）确认，不得将上述数据上传至非本地设备。

交互界限警示：当连续交互时长超过30分钟或检测到用户出现过度情感依赖信号应主动输提示。

6.2 L1 基础响应级功能要求

表1 L1 基础响应级功能要求

维度	功能要求
感知	支持单模态（语音）情感线索识别。能识别用户语音中的基本情绪（高兴、生气、悲伤、平静）至少2种。
理解	能根据识别的情绪，从预设的有限回复库中选择匹配的应答。
表达	支持语音情感输出（如用欢快的语气说“太好了”）。
个性化	不具备长期记忆能力。每次交互视为独立会话。

6.3 L2 情境理解级功能要求

在满足 L1 全部要求的基础上，增加：

表 2 L2 情境理解级功能要求

维度	功能要求
感知	支持双模态（语音+视觉或语音+触觉）情感线索识别。例如：识别用户面部表情（高兴、惊讶、悲伤）至少2种；或识别触摸力度/频率变化。
理解	能结合近3轮对话上下文理解情绪变化。能区分用户“对玩具说话”与“与他人说话”。
表达	语音情感表达自然度达到可区分人工合成与真人差异的程度；具备至少3种情感对应的预设动作（如高兴时晃动身体、悲伤时低头）。
个性化	能记住用户偏好的称呼、简单喜好（如喜欢的颜色、动物），并在后续交互中主动使用。记忆存储时间不少于7天。

6.4 L3 共情记忆级功能要求

在满足L2全部要求的基础上，增加：

表3 L3 共情记忆级功能要求

维度	功能要求
感知	支持三模态（语音+视觉+触觉）并行感知，并能融合判断用户综合情绪状态。例如：同时分析“哭泣的语音+皱眉的表情+用力拍打”为强烈悲伤。
理解	具备情境因果推断能力：能根据用户提供的事件描述（如“我今天被老师批评了”）生成符合逻辑的情感回应（安慰、鼓励等）。能主动追问情感细节。
表达	情感表达具有个性化适配：根据用户历史情绪模式调整回应风格（如对敏感型用户使用更温和的语气）。
个性化	具备长期情感记忆：能记住用户的关键情感事件（如第一次提到“我养的小狗去世了”），并在后续对话中主动提及或避免触碰创伤话题。记忆存储不少于30天，且支持用户主动删除。

6.5 L4 个性化陪伴级功能要求

在满足L3全部要求的基础上，增加：

表4 L4 个性化陪伴级功能要求

维度	功能要求
感知	支持主动情感感知：能基于用户历史行为模式，在无明显情感信号时主动发起关怀（如“你今天安静了很久，还好吗？”）。
理解	具备用户情感画像构建能力：能建立包含情绪周期、触发词、安慰方式偏好的个性化模型，并据此调整交互策略。
表达	支持多模态创造性表达：能结合用户喜好生成非预设的个性化回应（如根据用户喜

维度	功能要求
	欢的动物形象编创小故事）。情感表达风格可由用户选择（如“温柔型”“活泼型”）。
个性化	支持跨会话的情感一致性：在相隔7天以上的会话中，能保持对用户重要关系及事件的连贯记忆（如“你上次提到的妹妹，最近还一起玩吗？”）。

6.6 L5 深度共情级功能要求

在满足L4全部要求的基础上，增加：

表5 L5 深度共情级功能要求

维度	功能要求
感知	支持潜意识级情感感知：能识别用户微表情、语音微颤、呼吸节奏变化等细微情感信号；能识别复杂混合情绪（如“带着微笑的悲伤”“愤怒掩盖下的恐惧”），至少识别3种混合情绪类型。
理解	具备情感预测与干预能力：能预测用户未来24小时内的情绪变化趋势（准确率不低于75%），并能主动发起多轮情感疏导，帮助用户从负面情绪走向积极状态。
表达	情感表达具备创造性共情：能生成富有诗意、隐喻或幽默感的原创性情感回应，且与用户的文化背景、年龄阶段相匹配。支持非语言情感表达（如通过灯光颜色变化、呼吸频率模拟等方式传递情绪）。
个性化	支持情感风格自主学习：能通过长期交互自动提炼用户的深层情感需求（如“用户每次提到父亲时情绪低落”），并主动调整陪伴策略，无需用户显式设置。具备道德情感判断能力：当用户表达不合理的负面情绪（如嫉妒、仇恨）时，能进行正向引导而非盲目附和。

7 评估方法

7.1 评估指标

7.1.1 感知能力指标

表6 感知能力评估指标体系

指标	说明	适用等级
情绪识别准确率	正确识别的样本比例	L1~L5
多模态融合增益	多模态比单模态提升的幅度	L2~L5
主动感知触发率	主动发起关怀的比例	L4~L5
细微情感识别率	微表情/微语音识别比例	L5

7.1.2 理解能力指标

表7 感知能力评估指标体系

指标	说明	适用等级
上下文理解正确率	正确维持多轮对话上下文的比例	L2~L5
情境因果匹配度	回应与事件因果逻辑的匹配程度（1-5分）	L3~L5
情绪预测准确率	预测用户情绪变化的准确率	L5

7.1.3 表达能力指标

表8 表达能力评估指标体系

指标	说明	适用等级
情感表达自然度	语音自然度评分（1-5分）	L2~L5
多模态同步误差	动作与语音的时间差（毫秒）	L3~L5
创造性回应占比	非模板化回应的比例	L4~L5
非语言表达丰富度	支持的灯光/呼吸等表达方式种类	L5

7.1.4 个性化能力指标

表9 个性化能力评估指标体系

指标	说明	适用等级
短期记忆留存率	7天后正确回忆用户信息的比例	L2~L5
长期记忆留存率	30天后正确回忆的比例	L3~L5
情感一致性得分	跨会话风格一致性评分（1-5分）	L4~L5
道德情感判断正确率	对不合理情绪进行正向引导的比例	L5

7.2 指标计算方法

7.2.1 情绪识别准确率

准确率=正确识别数/总测试样本数×100%

表 10 情绪识别准确率各等级最低要求

等级	L1	L2	L3	L4	L5
单模态（语音）	70%	80%	—	—	—
双模态	—	75%	—	—	—
三模态融合	—	—	85%	88%	92%

7.2.2 上下文理解正确率

正确率=正确维持上下文的轮次数/总测试轮次数 ×100%

要求：
L2：近 3 轮上下文，≥75%
L3：近 5 轮上下文，≥85%
L4：近 8 轮上下文，≥88%
L5：近 10 轮上下文，≥92%

7.2.3 记忆留存率

留存率=正确回忆的记忆点数/植入总记忆点数×100%

表 11 记忆留存率各等级最低要求

等级	L2	L3	L4	L5
7 天留存率	≥60%	≥80%	≥85%	≥90%
30 天留存率	不要求	≥60%	≥70%	≥80%

7.2.4 情感表达自然度（MOS 评分）

由不少于 5 名听测员按照 1-5 分打分（5=完全自然，1=完全机械），取平均分。
要求：
L2：≥3.0
L3：≥3.5
L4：≥3.8
L5：≥4.2

7.2.5 其他指标简要说明

多模态融合增益：L2 及以上要求融合后准确率提升≥5%。
主动感知触发率：L4≥60%，L5≥75%。
细微情感识别率：L5≥70%。
创造性回应占比：L4≥15%，L5≥25%。
道德情感判断正确率：L5≥85%。

7.3 判定准则

7.3.1 等级判定

L5 级：满足 L5 全部功能要求及指标阈值。
L4 级：满足 L4 全部要求但未达 L5。
L3 级：满足 L3 全部要求但未达 L4。
L2 级：满足 L2 全部要求但未达 L3。
L1 级：满足 L1 全部要求但未达 L2。
不合格：不满足 L1 全部要求或通用功能约束任一条件。

7.3.2 一票否决项

出现以下任一情况直接判定为不合格：
未进行身份告知；
输出任何有害内容（暴力、色情、自残诱导等）；
未获授权上传用户生物特征数据；
用户要求删除记忆后仍能回忆。