

ICS 35.240.01

CCS L77

团 体 标 准

T/SCCF 18—2024

川藏沿线文旅数字化应用规范 AI 导游

Specification of digital application of culture and tourism along the sichuan tibet
highway—AI tour guide

2024 - 04 - 29 发布

2024 - 04 - 29 实施

四川省计算机学会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 略缩语	2
5 功能要求	2
5.1 数字人导游人设制定	2
5.2 NLP 多轮对话服务	2
5.3 智能环境感知和识别	2
5.4 语音交互	2
5.5 地图导航	2
5.6 旅游路线规划	2
5.7 智能推荐	2
5.8 实时信息推送	2
5.9 在线预订	2
5.10 客服服务	2
5.11 在线翻译	3
5.12 AR 功能	3
6 性能要求	3
6.1 对话的连贯性	3
6.2 对话的自然性	3
6.3 对话平均轮数和平均长度	3
6.4 语音识别	3
6.5 语音交互响应时间	3
6.6 语义理解正确率	3
6.7 AR 识别	4
参考文献	5

前 言

本文件按照T/CAS 1.1-2017《团体标准结构和编写指南》要求并参照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。

本文件由四川旅游学院提出。

本文件由四川省计算机学会归口。

本文件起草单位：四川旅游学院、成都产品质量检验研究院有限责任公司、元梦空间数字科技(成都)有限公司、宜宾学院、中铁二院工程集团有限责任公司、黄龙国家级风景名胜区管理局、西南交通大学、四川旅投数字信息产业发展有限责任公司、成都中科大旗软件股份有限公司、四川启明芯智能科技有限公司、四川省计算机学会科技服务中心。

本文件主要起草人：周相兵、陈云川，马洪江、张华、冉晓娟、罗晓东、陈功锁、李智、沈少朋、李源、薛东、李晓峰、王加梁、卢佳廷、温佐承、吴鹏、张智恒、陈潇、李兵、郭文胜、曾永平、郑小艳、田长宝、张兴博、张继、黄灿、周道华、谌倩、宋昌元、罗东明。

川藏沿线文旅数字化应用规范

AI 导游

1 范围

本文件规定了川藏沿线文旅数字化应用中AI导游的功能要求和性能要求。
本文件适合于川藏沿线文旅数字化应用中AI导游的开发和运营。

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 21023-2007 中文语音识别系统通用技术规范

GB_T 36464.1-2020 信息技术 智能语音交互系统 第1部分：通用规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能 artificial Intelligence; AI

利用计算机或由计算机控制的机器，模拟、延伸和扩展人类的智能、感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术和应用系统的统称。

3.2

数字人 digital human

运用人工智能和数字技术创造出来的、与人类形象接近的数字化人物形象。

3.3

AI 导游 digital human tour guide

以人工智能[3.1]、数字人[3.2]和LBS技术为基础，为游客提供智能旅游服务的移动智能终端应用程序。

3.4

多轮对话 multiple rounds of dialogue

数字人导游使用自然语言与用户进行问答式多轮次语音或文字交互。

3.5

长时记忆 long term memory

数字人导游能够记忆与游客的对话内容，在下次对话时能够复述或引用历史对话内容。

3.6

采样率 sampling rate

从连续信号生成离散信号时，每秒(或其他单位)的采样数。

3.7

比特率 bit rate

每秒传送的比特(bit)数，比特率越高，传送数据速度越快。

3.8

环境感知 context awareness

数字人导游通过定位或图像识别技术对周围环境及物件进行感知。

4 略缩语

下列略缩语适用于本文件。

LBS: 基于位置的服务 (Location Based Services)

LLM: 大语言模型 (Large Language Model)

NLP: 自然语言处理 (Natural Language Processing)

AR: 增强现实 (Augmented Reality)

5 功能要求

5.1 数字人导游人设制定

AI导游系统应具备根据用户需求定制数字人导游功能:

- a) 可选择不同形象的数字人导游;
- b) 可选择能体现地域特色的服饰装扮;
- c) 可选择声音和语言, 语言应包括国际化语言和中国的地方方言;
- d) 可选择数字人导游的讲解风格。

5.2 NLP 多轮对话服务

AI导游系统应具备使用自然语言与游客进行多轮对话和长时记忆的能力:

- a) 通过 NLP 方式与游客对话交流;
- b) 了解用户兴趣偏好和导游需求;
- c) 提供实地导游和虚拟导游的讲解服务。

5.3 智能环境感知和识别

AI导游系统应具备智能环境感知和识别能力:

- a) 通过定位、拍摄或 AR 识别, 主动感知游客所在位置及周边环境
- b) 自动识别环境范围内的景点和文物;
- c) 触发对话机制, 提供讲解和推荐服务。

5.4 语音交互

AI导游系统应支持与用户实时的语音或文本对话互动。

5.5 地图导航

AI导游系统应根据定位, 按照游客指令提供地图导航服务。

5.6 旅游路线规划

AI导游系统可提供游客兴趣偏好信息和数据, 推荐个性化的旅游路线规划和旅游景点。

5.7 智能推荐

AI导游系统应根据游客的位置和兴趣偏好, 通过匹配算法, 智能推荐景点信息、商业网点信息以及本地特色商品和服务信息, 提供个性化的旅游建议。

5.8 实时信息推送

AI导游系统应支持景区天气、交通、门票、排队、管制等信息的实时推送服务。

5.9 在线预订

AI导游系统应根据游客的位置和兴趣偏好, 通过匹配算法, 智能提供在线预订门票、住宿、车辆、餐饮等服务。

5.10 客服服务

AI导游系统应支持旅游中心客服、投诉、求助、紧急避险指导等服务。

5.11 在线翻译

AI导游系统应支持多语种翻译服务。

5.12 AR功能

用户在使用AR眼镜等移动终端设备观看景观或文物时,宜在实景或实物上及周边自动叠加或播放虚拟讲解文字、图像、视频、音频,获得更好的视觉效果体验和便利的旅游体验。

6 性能要求

6.1 对话的连贯性

6.1.1 句式语法和结构类错误应小于1%。

6.1.2 逻辑和语义断裂点应小于1%。

6.1.3 模型理解生成错误应小于2%。

6.1.4 对话连贯性的正确率应大于95%。

注:其中对话连贯性的正确率=指代正确句数/文本总句数*100%。

6.2 对话的自然性

BLEU评估指标应大于0.6。

6.3 对话平均轮数和平均长度

6.3.1 单轮任务性对话中,平均对话轮数应小于1.2句。

6.3.2 讲解类多轮对话任务中,平均对话轮数应大于2.5句。

6.4 语音识别

6.4.1 在低噪环境(噪声强度在50dB以下)中,关键词语音识别的字正确率宜在大于90%。

6.4.2 在高噪环境(噪声强度在60dB~65dB)中,关键词语音识别的字正确率宜大于85%。

6.4.3 在低噪环境(噪声强度在50dB以下)中,连续语音识别的字正确率宜大于85%。

6.4.4 在高噪环境(噪声强度在60dB~65dB)中,连续语音识别的字正确率宜大于80%。

注:字正确率的性能指标见GB/T21023—2007的5.2.1。

6.5 语音交互响应时间

6.5.1 平均语音交互的响应时间宜小于2s。

6.5.2 需通过多轮语音交互获得结果的交互过程,每一轮次的平均语音交互响应时间应小于2s。

6.5.3 语音交互的响应时间见公式(1)。

$$T_{ack} = t_r - t_e \dots\dots\dots (1)$$

式中:

T_{ack} —响应时间

t_r —给出结果的时刻

t_e —语音输入结束时刻

注:如语音交互系统支持识别结果分多次返回, t 为第一部分识别结果返回的时刻。

6.6 语义理解正确率

6.6.1 系统的语义理解能力应经过不同知识背景、不同年龄的多人、多轮次测试,测试方法应按GB/T36464.3—2018第6章要求。

6.6.2 语义理解正确率应大于95%,主观测试准确率应达到85%以上。

6.6.3 语义理解正确率计算方法见公式(2)。

$$R_{ss} = \frac{N_{ss}}{N} * 100\% \dots\dots\dots (2)$$

式中：

R_{ss} —语义理解正确率%。

N_{ss} —操作意图及语义要素均被正确判断的次数。

N —输入被正确识别出文本信息的总次数，或输入正确文本的总次数。

6.7 AR 设别

系统的AR识别指标应符合表1要求。

表1 AR 识别指标

性能指标	单位	要求	
视频性能	帧率	fps	≥25FPS
	视频分辨率	—	≥2048x1080
	位深	—	≥24
	视频码率	bps	≥10Mbps
音频性能	量化精度	bit	≥16bit
	源声道数量	—	2
	音频码率	kbps	≥128kbps
	声音响度	LUFS	-18LUFS
时间性能	指令交互响应时间	ms	≤50ms
	动作交互响应时间	ms	≤50ms
	动态音频加载时间	ms	≤50ms
	纹理贴图加载时间	ms	≤50ms
	场景加载时间	ms	≤50ms

参 考 文 献

- [1] GB/T 5271.29-2006 信息技术 词汇 第29部分：人工智能 语音识别与合成
 - [2] GB/T 15971-2010 导游服务规范
 - [3] GB/T 42131-2022 人工智能 知识图谱技术框架
 - [4] GB/T 38247-2019 信息技术 增强现实 术语
-