Q/GUOSEN

国 信 证 券 股 份 有 限 公 司 企 业 标 准

Q/GUOSEN-ENG-ARI 01-2024

注：*标准编号由 PMO 与规范设计组给出*

国信证券 2D 虚拟数字人
应用建设指南标准

2024-9-19 发布　　2023-10-10 实施

国信证券股份有限公司 发布

前　　言

本标准依据 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。

本标准于 2024 年 9 月初次起草发布，11 月首次实施，在编制过程中得到各位同仁的大力支持，在此表示感谢！由于编制时间紧促，难免有遗漏之处，恳请广大参阅者多提高贵意见，以供参考校正。

本标准由国信证券股份有限公司提出。

本标准起草单位：国信证券股份有限公司。

本标准主要起草人：薛仲义、王燕华、杨柳青、沈雪丽、林文政、赵鹏、陈鹏奋、尚记学、黄常莞、刘璐、张政扬。

本标准于 2024 年 9 月首次发布。

1 范围

本标准规定了 2D 虚拟数字人在证券业务的应用建设指南，主要描述了 2D 虚拟数字人应用建设、应用安全建设、2D 虚拟数字人在形象、语音、动作等维度的评估。

本标准适用于指导 2D 虚拟数字人系统构建、应用建设、服务功能评估等工作。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是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 22239—2019 信息安全技术 网络安全等级保护基本要求

JR/T 0068—2020 网上银行系统信息安全通用规范

JR/T 0092—2019 移动金融客户端应用软件安全管理规范

JR/T 0171—2020 个人金融信息保护技术规范

3 术语和缩略语

下列术语和定义适用于本文件。

表 1 术语说明

术语、缩略语	解释
虚拟数字人 digital human	简称数字人或虚拟人，是指基于现实世界设计、通过计算机生成、再借助真人或计算驱动、再多模态输出设备呈现的虚拟人物。
2D 虚拟数字人技术 2-Dimensional digital human	是指由真人录像与声纹采集训练而成的 2D 虚拟数字人，拥有高度接近真人形象、动作、声音的数字人。
2D 虚拟数字人技术 2-Dimensional digital human technology	2D 虚拟数字人技术是指通过一段小样本视频进行数字人训练的技术。
2D	二维（2-Dimensional）

4 数字人系统参考框架

4.1 概述

2D 虚拟数字人证券应用是指通过基于真人形象构建的数字人形象建设及对应系统建设，在证券投教、证券业务办理等场景，为证券客户提供创新金融服务体验。本标准中所指 2D 虚拟数字人技术是基于真人一段小样本视频进行数字人训练的技术。

4.2 建设目标

数字人证券应用建设目标，包括：

a)建设形象自然、表达流畅的 2D 虚拟数字人基础能力，数字人形象、表达符合证券服务场景设定；

b)设计开发、部署数字人证券应用系统，满足证券行业对系统高可用、高并发以及兼容性等要求；

c)建设业务合规风控能力，为数字人证券应用提供安全保障。

4.3 总体架构

2D 虚拟数字人金融应用总体架构包括数字人基础功能建设、数字人交互能力建设、数字人应用建设、数字人安全建设等内容，总体架构如图 1 所示。

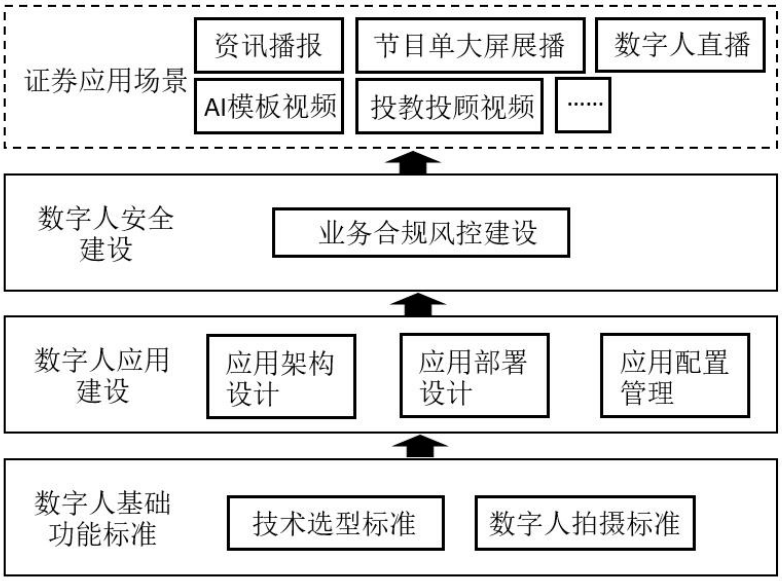

图 1 2D 虚拟数字人证券应用总体架构图

数字人基础功能建设是结合数字人技术选型与数字人拍摄标准而成，数字人基础功能采取与厂商采购合作的形式，主要以建设技术选型标准及数字人拍摄标准为主，确保选取效果优异的厂商，为上层证券应用建设提供基础支撑。

数字人应用建设是从系统建设角度给出架构设计和部署指引，包括应用架构设计、应用部署设计和应用配置管理等内容。

数字人安全建设贯穿整个建设过程，为数字人应用提供业务合规风控建设等内容。

证券应用场景主要包含数字人内容类型与投放渠道，数字人内容生产主要有资讯播报、投教投顾视频、AI 模板视频、数字人直播等内容，投放渠道包含节目单大屏展播等方式。

5 数字人基础功能标准

数字人基础功能建设可分为数字人选型与数字人拍摄标准两个阶段，各阶段内容如下：

a)数字人选型阶段，主要是根据数字人选型标准，将厂商的数字人效果进行评分，选型标准主要有清晰完整性、形象美感度、口唇准确率、语音相似度、语音舒适度等维度进行评选，详细可参考本文第八点。

b）数字人拍摄阶段，主要是根据拍摄格式、形象造型、播报流程等维度进行标准应用，详细可参考本文第九点；

6 数字人应用建设

6.1 应用架构设计

数字人应用在线上线下等金融场景，线上包含 APP、企业微信、视频号等渠道平台，线下包含节目单大屏播放，应用架构设计如图 2 所示：

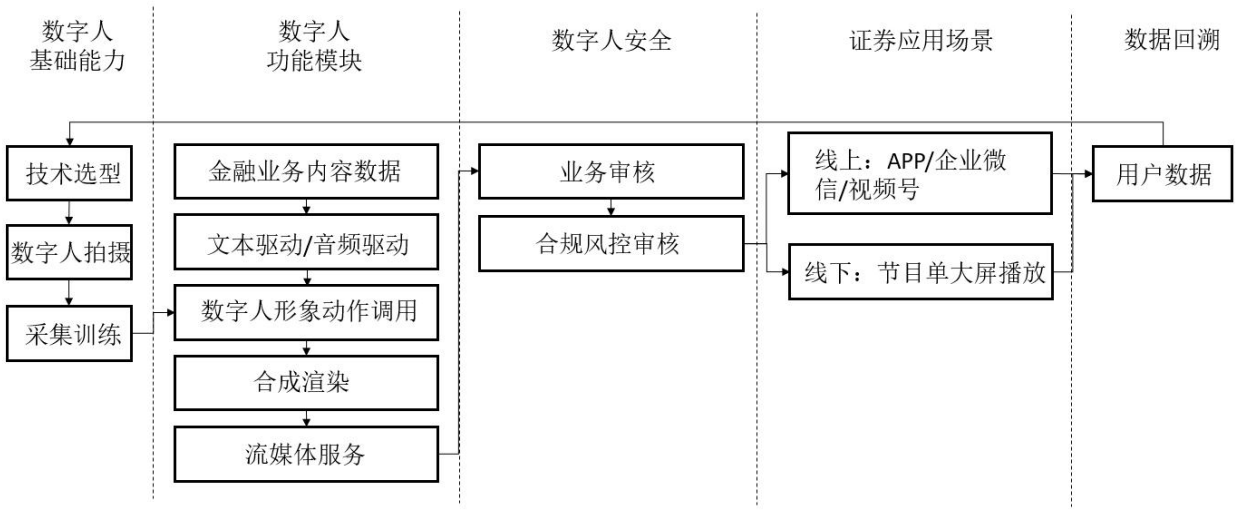

图 2. 数字人证券应用架构

主要包括以下内容：

a)数字人基础能力通过项目前期选型确定服务厂商，通过数字人拍摄确认小样本训练视频，采集后进行数字人训练，训练完毕的数字人以接口形式进行调用；

b)数字人功能模块以金融业务内容数据作为内容源，使用文本/音频驱动，合成渲染生成数字人音视频流，流媒体服务提供 rtmp/rtsp 视频流 API 接口；

c)合规审核模块将生成好的数字人音视频进入审核队列，经过审核后，内容最终显示到应用场景上；

d）证券应用场景包含 APP、企业微信、视频号、线下节目单大屏，并将相关用户数据进行回溯留存，用以指导新的数字人生产与数字人内容生产。

6.2 应用部署设计

6.2.1 高可用设计

宜采用微服务、集群的部署方式，以满足证券业务系统高可用需求，如建设业务集群、数据集群、AI 集群、流媒体集群，集群之间通过接口服务调用。

6.2.2 高并发设计

宜采用负载均衡、分布式数据库等技术，以满足金融业务系统高并发需求。

6.2.3 兼容性设计

兼容性设计，包括以下内容：

a)支持多渠道终端接入能力，如 APP、小程序、Web/H5、大屏设备等渠道终端；

b)支持与音视频平台、AI 技术平台（自有 AI 技术能力或第三方 AI 技术能力）集成对接；

c)模型渲染和计算，支持云计算模式、端计算模式以及端云结合等计算模式；

d)支持对国产操作系统和硬件的兼容性适配。

7 数字人安全建设

业务合规风控建设，包括但不限于：

a)确保数字人形象版权符合法律法规要求，涉及真人授权需有肖像与声纹权等授权文件，涉及创作作品需有设计著作权等证明文件，声音、音乐、背景等素材需取得版权许可；

b)确保数字人相关视频在发布前经过合规审核，视频发布后要发送给数字人本人知悉；

c)确保数字人员工与公司解除劳动关系的时候，相关数字人与数字人视频能够自动下线；

c)宜加强客户教育和风险提示，向客户详细解释本机构数字人系统业务流程和安全措施。建立客户投诉、纠纷处理及舆情应对机制，维护客户权益及公司声誉。

8 2D 虚拟数字人指标及规范性描述

8.1 清晰完整性

指 2D 虚拟数字人形象的清晰完好程度，若在在面部、口唇、耳朵、颈部、头发、身体边缘、面部下三角外缘出现下列情况中任意一种或几种则视为有破损。

——存在明显割裂；

——存在明显黑边或绿边；

——存在明显抖动、跳帧；

——存在明显扭曲、畸变、马赛克；

8.2 形象美感度

指 2D 虚拟数字人形象让用户感到具备美感的程度。该指标为主观性评估指标，美感包括好感度、自然度、使用意愿等指标，在李克特量表中给出一个主观评分评价质量优劣，1 为最差，5 为最优，具体评分规则见下表。

评测维度	描述	5	4	3	2	1
好感度	你喜欢该形象的设计吗？	十分喜欢	比较喜欢	一般	不太喜欢	十分不喜欢
自然度	该形象是否自然？	十分自然	比较自然	基本自然	不太自然	十分不自然
使用意愿	你愿意使用该形象为你服务吗？	非常愿意	比较愿意	一般	不太愿意	十分不愿意

表 2. 数字人形象美感度表

8.3 口唇发音准确率

指 2D 虚拟数字人训练时，在文字合成语音过程中的口唇与发音匹配的准确度。口唇发音不准确的表现包括口唇较发音快帧或慢帧、口唇与发音匹配错误等，相应的性能指标包括口唇与发音字准确率和发音句准确率，计算方法如公式所示：

$$R_{WC} = \left(1 - \frac{N_{ew}}{N_w}\right) \times 100\%$$

式中：

N_w ——文本总字数，单位为个；

N_{ew} ——口唇与发音不匹配的错误字数，单位为个；

R_{wc} ——口唇与发音准确率。

8.4 语音相似度

指 2D 虚拟数字人在语音合成过程中的语音与真人本人的相似度。该指标为主观性评估指标，相似度包括音色、语气、语调等因素，该指标为主观性评估指标，用户根据听到的语音合成效果，在李克特量表中给出一个主观评分评价质量优劣，1 为最差，5 为最优，具体评分规则见下表。

评测维度	描述	5	4	3	2	1
音色	数字人与真人音色一致吗？	完全无法区分	比较相似，有细微差别	基本相似	不太一致	十分不一致
语气、语调	数字人与真人的语气、语调一致吗？	十分一致	比较一致	一般	不太一致	十分不一致

表 3. 数字人与真人语音相似度表

8.5 语音舒适度

指 2D 虚拟数字人合成语音让用户感到舒适的程度。该指标为主观性评估指标，舒适度包括语音质量、语音情绪、语音发音标准等指标，用户根据听到的数字人声音质量，在李克特量表中给出一个主观评分评价质量的优劣，1 为最差，5 为最优，具体评分规则见下表。

评测维度	描述	5	4	3	2	1
语音质量	整体语音是否标准舒适？	十分标准舒适	比较标准舒适	一般	不太标准舒适	十分不标准、不舒适
	语音是否清晰干净？（无明显杂音、噪音、尖锐感等）	十分清晰干净	比较清晰干净	一般	不太清晰干净	十分不清晰、不干净
	语音情绪是否饱满？	十分饱满	比较饱满	一般	不太饱满	十分不饱满
	语音吐字是否清晰（抑扬顿挫清晰，不吞字，中英文连贯自然）	十分清晰	比较清晰	一般	不太清晰	十分不清晰
	语音吐字是否准确？（包括儿化音、平翘舌、轻重	十分准确	比较准确	一般	不太准确	十分不准确

	音、重音）					

表 4. 数字人语音舒适度表

9 2D 虚拟数字人拍摄指标及规范性描述

9.1 拍摄格式

指 2D 虚拟数字人在训练时的小样本视频的拍摄格式，其规范性描述，包括以下信息：

——说明拍摄格式，如拍摄视频分辨率、视频时长、视频格式、声音时长、声音格式、拍摄动作等；

9.2 模特造型

指 2D 虚拟数字人的小样本视频拍摄时的真人模特造型，其规范性描述，包括以下信息：

——说明模特画面状态，如半身/全身、横屏/竖屏、站姿/坐姿等；

——说明模特妆造状态，如面部、发型、服饰等；

——说明模特表现力状态，如面部、眼神、体态、动作等。

9.3 播报流程

指 2D 虚拟数字人的小样本视频拍摄时的真人播报录制流程，其规范性描述，包括以下信息：

——说明播报语音要求，如文本播报连贯、不间断不卡壳、有情感、抑扬顿挫清晰等；

——说明语音播报流程，如演员站定、演员静默、演员播报等。

附件 A 数字人选型评分表

功能子类	功能点描述	5	4	3	2	1
整体效果	你喜欢该形象的设计吗？	十分喜欢	比较喜欢	一般	不太喜欢	十分不喜欢
	该形象是否自然？	十分自然	比较自然	基本自然	不太自然	十分不自然
	你愿意使用该形象为你服务吗？	非常愿意	比较愿意	一般	不太愿意	十分不愿意
	数字人的表情接近真人吗？（眉眼与神情是否具有真人感）	十分接近	比较接近	一般	不太接近	十分不接近
动作	全身动作自然吗？（支持数字人动作流畅，无卡顿）	十分自然	比较自然	基本自然	不太自然	十分不自然
	全身动作与内容匹配吗？（支持动作能匹配上语言表达的含义）	十分匹配	比较匹配	一般	不太匹配	十分不匹配
清晰度	动态、静态的面部五官的边缘、整体清晰度高吗？	十分高	比较高	一般	不太高	十分低
	动态、静态的人物边缘清晰度高吗？	十分高	比较高	一般	不太高	十分低
口唇	口唇与声音匹配度，支持人工逐字符与标准口唇复核，共 400 字的内容；（口唇发音不准确的表现包括口唇较发音快帧或慢帧、口唇与发音匹配错误等）	$R_{WC} = \left(1 - \frac{N_{ew}}{N_w}\right) \times 100\%$				
	动态的嘴唇存在抖动情况吗？	□存在；□不存在				
	牙齿存在抖动和缺失情况吗？	□存在；□不存在				
耳朵	耳朵不能存在畸变情况；	□存在；□不存在				
头发	头发不能存在割裂、断裂情况；	□存在；□不存在				
	头发边缘不能存在毛糙情况；	□存在；□不存在				
脖子	脖子不能存在闪动、抖动情况；	□存在；□不存在				

面部下三角外缘	面部下三角外缘不能存在割裂、黑边、绿边、蓝边情况；	□存在；□不存在				
	面部下三角外缘不能存在闪动抖动情况；	□存在；□不存在				
声音	数字人与真人音色一致吗？	完全无法区分	比较相似，有细微差别	基本相似	不太一致	十分不一致
	数字人与真人的语气、语调一致吗？	十分一致	比较一致	一般	不太一致	十分不一致
	整体语音是否标准舒适？	十分标准舒适	比较标准舒适	一般	不太标准舒适	十分不标准、不舒适
	语音是否清晰干净？（无明显杂音、噪音、尖锐感等）	十分清晰干净	比较清晰干净	一般	不太清晰干净	十分不清晰、不干净
	语音情绪是否饱满？	十分饱满	比较饱满	一般	不太饱满	十分不饱满
	语音吐字是否清晰（抑扬顿挫清晰，不吞字，中英文连贯自然）	十分清晰	比较清晰	一般	不太清晰	十分不清晰
	语音吐字是否准确？（包括儿化音、平翘舌、轻重音、重音）	十分准确	比较准确	一般	不太准确	十分不准确
制作周期	从提供训练材料到训练出数字人上线交付，时间不超过 5 个工作日；	低于 5 个工作日	5-7 个工作日	7-9 个工作日	9-11 个工作日	11-13 个工作日
输出分辨率	交付成品的分辨率是否低于 4K；	□高于等于 4K；□低于 4K				