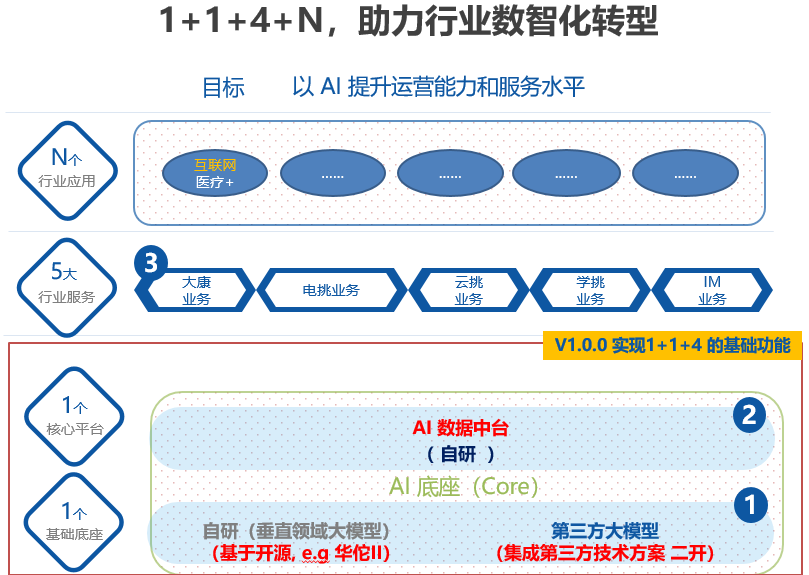
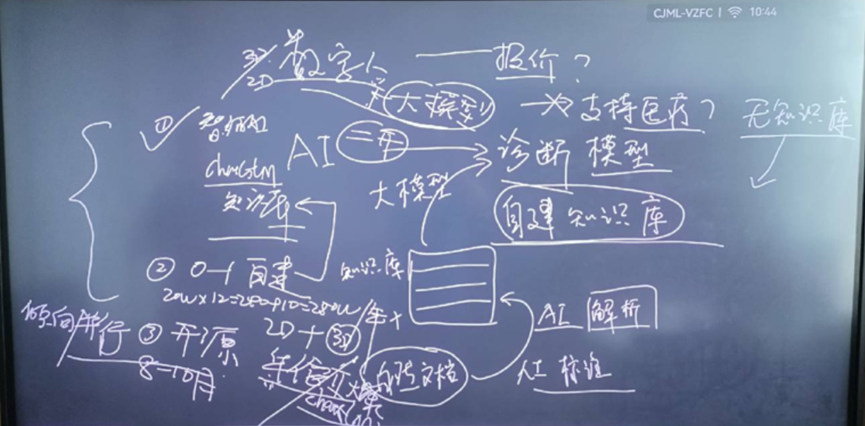
**数字人组成**

# 星链 AI产品和服务体系



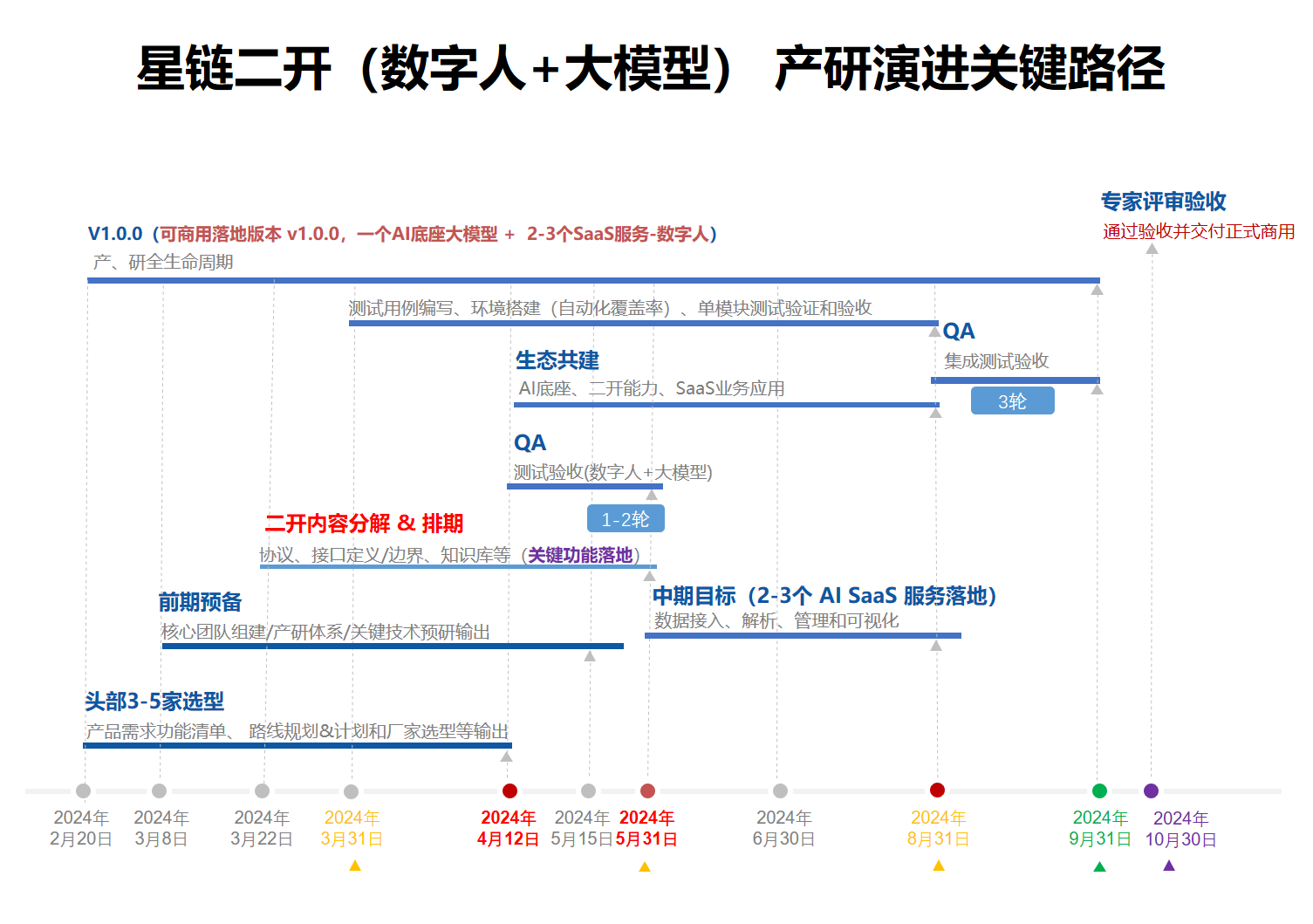


# 星链AI产品体系框架：1 + 1 + 4 + N

**一个 AI 底座，**基于自研 AI 大数据底座或集成第三方方案定制化二开（哪些需做定制化，哪些复用第三方能力，集成第三方需技术选型并形成最终技术合作商），**厂家选型和报价参考第三方技术方案调研**

1. 集成第三方 AI 能力 二开（v.10 版本 按产品规划要求6月初发布可用版本）
   * 1. 2D、3D 虚拟数字人解决方案
     2. 大模型 AI 底座（HW盘古、智谱AI ChatGLM、讯飞星火等等）

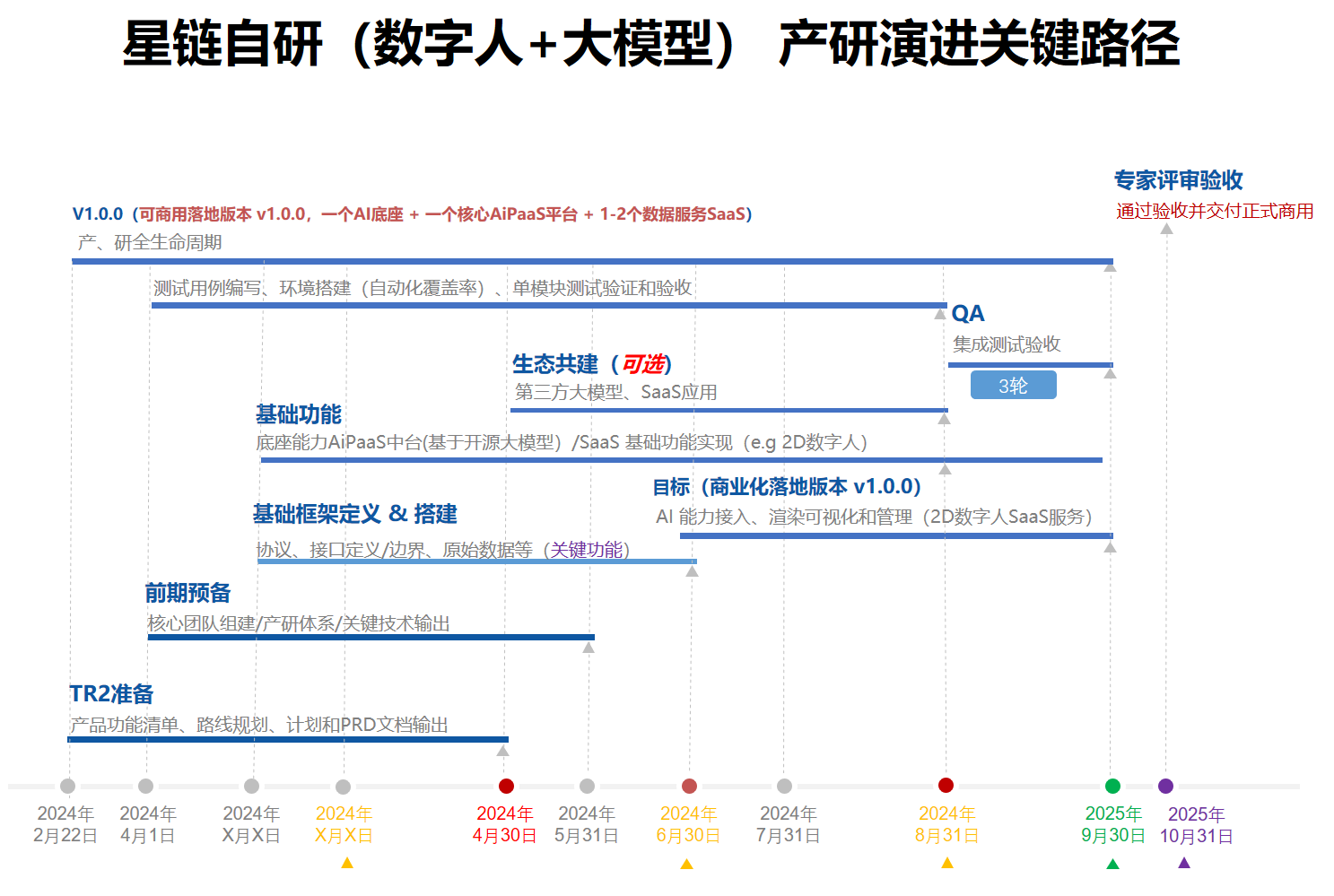
**注: 二开的产研工作量需依据确定厂家选型方案后评估输出(*待确定厂家后需更新二开开发工作量评估，目前是预估*)**



1. 自研（中长期规划）
   * 1. 基于开源大模型框架（e.g. ChatGLM、华佗II- LLM4Med 、LLaMA2等）
     2. 基于开源数字人项目（e.g. Gitee – Fay、OpenAvatar、VFA等）

比如：缓解多轮幻觉的解决方案，关键信息注意力增强、知识图谱强化、多轮知识强化、知识溯源能力等技术

**自研 产研路径分解：**



1. **一个核心平台 - AI 数据中台 AiPaaS**

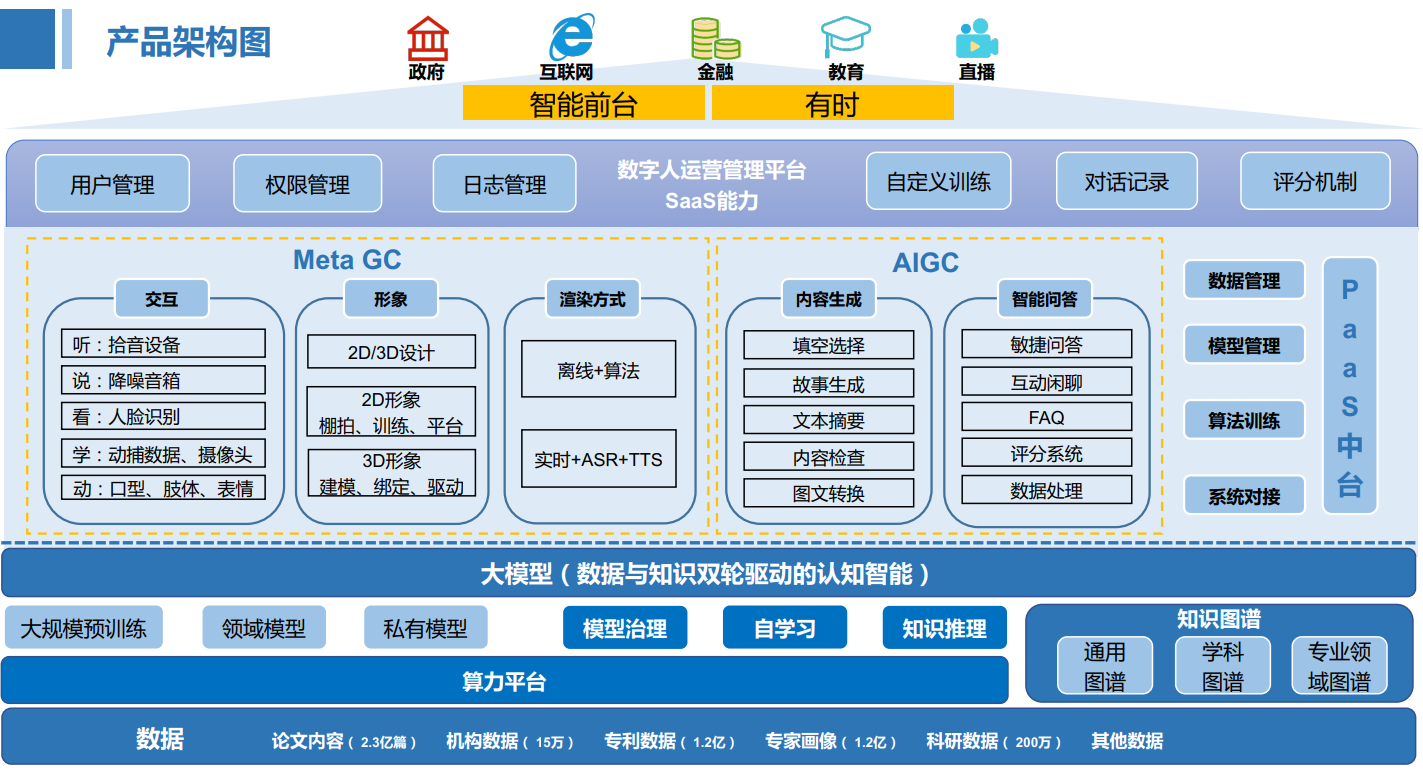
基于AI 底座大模型封装能力提供给到业务SaaS的能力中台（AI Core + 数据中台PaaS）

1. 4个 数据服务SaaS, 大康业务、电挑业务、云挑业务、学挑业务等
2. N个行业拓展应用，政务、金融、商客、大健康（医疗）、大文娱（电商、旅游等）

# 数字人组成

数字人系统框架一般情况下由人物形象、语音生成、动画生成、音视频合成显示、交互五个核心模块构成数字人构建的“五横体系”  
**交互模块**为可扩展项，依据是否具有交互模块可讲虚拟人分为交互型虚拟人和非交互型虚拟人，其交互型虚拟人依据其驱动方式不同，又可分为智能驱动型虚拟人和真人驱动型虚拟人

关键技术，虚拟人制作过程中核心关键技术应关注：建模、驱动、渲染和人工智能。其决定了虚拟人的外在形象的美观度、动作的流畅度以及交互的自然度及虚拟人的智能程度。



**当前主流的SaaS平台架构**



从下往上依次是：服务层、协议层、化身层、应用层。

**服务层：**底层技术框架，需搭建“AI能力支撑平台”（一般是结合大模型AI能力,e.g HW盘古、讯飞星火等）来承载所有功能，包括：语音识别、语音驱动、文本驱动、语音转化、手势驱动、图像驱动、声纹复刻、人脸识别、动作捕捉、OCR识别、知识图谱、智能推荐等能力；除技术能力外，服务层还应包含运营使用配置功能等；

**协议层**，语音交互协议、口型驱动协议、表情驱动协议、动作驱动协议，作用是与服务层和化身层进行对接，将服务层获取到的信息进行运算处理来传递给化身层，以便于模型的控制。

**化身层，**对数字人资产的管理，需搭建“角色资源平台”来进行管理，包含：3D/2D角色库、角色配饰库、动作/表情库、特效库、背景环境库、模板管理、音视频频管理等。

**应用层，**又称为应用场景层，该层平台会打包成不同平台SDK供应用集成，应包含：PC、Web、iOS、Android等。

# 功能概述

|  |  |  |  |
| --- | --- | --- | --- |
| **系统模块** | **功能简述** | **相关功能** | **备注** |
| 建模/驱动 | 基于AI提升2D/3D建模效率 | 自动化建模(2D/3D)  体积视频  视觉驱动  语音驱动 |  |
| 资产管理 | 聚合行业3D数字资产 | 资产库  资产标签  资产检索  安全保护 |  |
| 渲染 | 云原生、实时、端云协同渲染 | 离线渲染  实时渲染（ASR、TTS）  自研引擎  端侧SDK |  |
| 编辑 | 基于云工作站打造云上制作和编辑入口 | 云工作站  远程协作  数据安全  工具包聚合 |  |
| 在线机器人 | 用于替代人工解决大量重复性问题，以大模型、知识库为基础进行问题匹配回答，并配置转人工策略，提升处理效率 | 知识库、词库、规则管理 |  |
| 知识库 | 构建特定领域场景的知识沉淀、查询、智能辅助等（e.g. 医疗、电商等）  辅助进行问题处理，客服 | 知识查询、智能推荐 |  |
|  |  |  |  |
|  |  |  |  |

# 主流动捕技术对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **分类** | **概念** | **优点** | **缺点** | **备注** |
| 光学动捕 | 是在人的身体上进行标记，标记点的会反射到架设好的摄像机，通过反射的不同位置的成像信息，可以进行测算出标记点的空间运动信息，进而将这些信息进行定位或输出实现不同行业的实际应用。 | 动捕精度高  保真程度高 | 技术复杂度和拍摄要求高  造价成本高  应用范围小 |  |
| 惯性动捕 | 惯性动作捕捉技术是在人的身上佩戴陀螺仪，人运动的过程陀螺仪会进行旋转，通过感知陀螺仪的旋转信息来推算出人的动作行为实现动作捕捉。 | 价格相对较低  不依赖场地环境 | 动捕精度低；  稳定性低，容易发生位置漂移 |  |
| AI视觉捕捉 | AI视觉捕捉不需要进行标记和佩戴，只需要在人的活动范围内通过普通的摄像头进行动作的录制实现人体关键点信息的识别进而采用算法来实现动作捕捉。该项技术对硬件几乎没有特殊的要求，更多的优势在算法上 | 简单上手  成本较低 | 动捕精度低  自由度较低  视野固定 | 优选 |
|  |  |  |  |  |

**推进选择路径：**

实用上手简易程度：AI视觉动捕 > 惯性动捕 > 光学动捕

保真程度和稳定性：光学动捕 > 惯性动捕 > AI视觉动捕

造价成本：光学动捕 > 惯性动捕 > AI视觉动捕

# 人物模型、场景载入引擎

**UE4**作为模型的驱动引擎，在三维的世界里，有两条工具线：

一个是引擎，用于驱动三维模型按照你的逻辑运作；

一个是建模工具，比如maya。但通常这两类工具都会互相融合、相互交叉。

* 二次元的模型建立，可以使用daz studio，非常方便选择人的各个组成部分，比如：身驱、头发、脸型、眼睛等，然后组合成想要的形象
* 超写实的模型，可以使用metahuman了。metahuman其实就是一个云端版本的ue,优点是集成了大量真人扫描的高精度组件，可以非常方便地调节出一个欧美真人。由于疫情原因，还未对亚洲人进行扫描。在虚拟主播这个案例上是用metahuman。
* 把模型导入到ue4，给模型在ue4里预设了四个动作（站着、站着说话、坐着、坐着说话），三个表情（平静、开发、愤怒），三个镜头（全局、近矩说话、看显示器播放商品展示）
* 当数字人的心情激动（开心和愤怒）是站着的，其余时候是坐着的，表情也做对应的变化。说话的时候就会做更多的肢体动作了，唇是根据说话的发音驱动同步化的。

**行为模型驱动UE引擎里的三维模型**

行为模型逻辑去控制UE里的三维模型，是把逻辑输出模拟成键盘操作，UE再依据键盘输入来驱动三维模型作出变化。UE商城里websocket蓝图插件，与行为模型实现websocket通讯。测试过讯飞、阿里云、百度、亚马逊和微软的语音合成，只有微软是直接提供带情绪的合成。

# 3D 虚拟人定制

面向品牌 IP 性要求较高的用户，专属 3D 虚拟人定制方案，基于充分了解和分析虚拟人的目标、定位和核心价值，与客户展开深度合作，量身定制独特的虚拟形象。定制过程可能包括进行原画设计、建模、骨骼表情绑定等步骤，以实现语音驱动卡通形象。并可实现形象的表情、肢体骨骼、手部骨骼驱动；生气、愤怒、高兴不同情绪；拍手、抱胸、强调、握拳等动作，使得虚拟人表现更生动，给用户带来更具吸引力和个性化的体验。

定制内容包括：

1. 表情绑定，实现形象的表情或口唇的精准性；

2. 肢体骨骼绑定，实现形象的肢体的精准性；

3. 手部骨骼绑定，实现形象的手部的精准性；

4. 可根据业务需求定制不同情绪，如生气、愤怒、高兴等；

5. 可根据业务需求定制不同动作，如拍手、抱胸、强调、握拳等

**第三方技术方案调研**

**智谱AI（大模型第一梯队）**

<https://www.zhipuai.cn/>

致力于研发纯AI驱动数字人技术。基于UE5渲染引擎研发了包括语音识别/合成、口型/表情生成模型、肢体动作生成模型、表情/动作驱动、渲染合成、推流混流在内的离线/实时数字人渲染系统。可支持各种场景及形象的高效高精度纯AI驱动数字人渲染输出。

MetaGC, 形象生成及驱动

快速生成2D、3D、卡通数字人形象，首创AI情感驱动并支持驱动模式，更广泛地适配各种驱动硬件或软件深度学习算法，融合口型、表情等人物特征，让数字人更加智能亲近。

数字人交互, 综合MetaGC和大模型AIGC能力，神形兼备的解决方案让数字人更加多元化地服务于各种特定场景，形成娱乐与功能一体的综合解决方案。

AIGC，人工智能数字脑

基于大模型、NLG、NLU等能力，加强实现语言、语义理解及内容生成的多种应用；支持长文本转短文本、长文本撰写、智能问答、短文本交流等AIGC算法。

知识图谱，海量专业知识

融合千万级实体的大规模知识图谱和各行业相关专业知识，支持为数字人配置专属的行业知识和能力，实现更加专业精准的服务能力。

大模型MaaS平台，

ChatGLM, 中英双语千亿对话模型，基于千亿基座模型 GLM-130B，注入代码预训练，通过有监督微调等技术实现人类意图对齐，具备问答、多轮对话、代码生成功能中英双语大模型。

优势：

* 配置项支持：对话模型、语音音色/语速、

情绪、虚拟形象、服装、背景、配饰、

毛发、动作库/动作生成模型、机位、输

出分辨率、帧率

* 最大输出分辨率：4K
* 离线渲染效率：1:1.7
* 实时渲染/推流延迟：2-3s
* 输出帧率：25/30 fps
* 显卡支持：3080/3090

**HW MetaStudio-数字人**

<https://www.huaweicloud.com/product/cbs/digitalhuman.html>

基于华为自研的数字人大脑，通过分钟级训练，可实现数字人应用小时级上线。支持基于文本或语音驱动数字人，应用于行业播报、电商直播等场景，也可基于数字人大脑构建知识系统，实现数字人的实时交互。

数字人接入（2D、3D）, 支持用户自定义数字人形象，也可挑选使用华为自定义的数字人形象。该功能目前为邀测状态。

多模态播报机器人（在线创作、API调用），基于华为云内置的数字人形象和客户自有的数字人形象，通过多模态播报机器人服务，实现数字人视频创作。输入文本或语音，即可驱动数字人生成所需视频。

多模态交互机器人，通过多模态交互机器人，实现数字人实时交互。通过数字人交互配置、知识配置及交互形式选择模块，驱动数字人进行拟人对话。

优势：

慧感知，语音、肢体、情绪全识别，准确率超90%

擅表达，实时语义驱动交互，拟人化口语表达

知识广，海量行业知识，小时级定制行业知识

能进化，知识闭环，实现数字人知识主动获取和更新

支持私有化部署

**科大讯飞数字人**

<https://virtual-man.xfyun.cn/>

<https://xinghuo.xfyun.cn/>

运用最新的AI虚拟形象技术，结合语音识别、语义理解、语音合成、NLP、星火大模型等AI核心技术，提供虚拟人形象资产构建、AI驱动、多模态交互的多场景虚拟人产品服务。

科大讯飞 AI 虚拟人交互平台，运用最新的 AI 虚拟形象技术，结合语音识别、语义理解、语音合成、星火大模型等 AI 核心技术，提供虚拟人形象资产构建、AI 驱动、多模态交互的多场景虚拟人解决方案。

共建虚拟人生态，在多模感知、多维表达、情感贯穿、自主定义上持续提升，致力于用

AI 虚拟人能力赋能多行业的业务场景，为客户提供一站式的虚拟人服务。

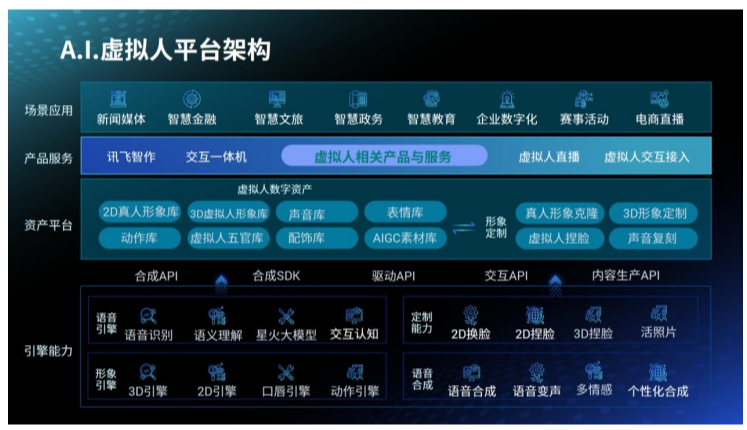
AI 虚拟人交互平台架构上包括：引擎中心、资产平台、产品服务和场景应用四个模块。

引擎中心，为虚拟人的应用和服务提供核心技术能力，同时可以通过 SDK、API 接入等方式集成到应用中。

资产平台，构建多场景虚拟人形象和配套资源服务的一站式资产生态。

产品服务，提供包括讯飞智作、智能交互机、虚拟人直播、虚拟人交互接入等虚拟人相关产品和服务。

场景应用，基于平台提供的能力和应用实现了音视频播报、业务咨询、互动交流、服务导览等多项功能，满足多场景的需要。



优势：

虚拟人形象，支持 2D 真人形象，可为客户定制专有 1P 形象，24 小时为您分身代言.多造型形象+多语言声音，覆盖总数量超过 100+

形象声音，为形象定制专属的个性化语音库，更有海量音库供选择，根据客

户对音色、风格的需求灵活定制

交互服务，语音识别、语义理解、语音合成、形象驱动、视频渲染服务全链路保障

形象逼真，口唇精准，交互画面流畅语音自然清晰、体验佳、亲和感强

随时打断，真人式对话体验，交流过程中即时打断， 想问就问，对话时机零等待

低延迟率，实时沟通延时短， 问后即答，真实还原“面对面〞实时对话体验

肢体动作，文本内容插入指向性动作，过渡自然， 提升形象交互丰富度和灵活性

支持私有化部署，

**小冰（微软亚洲研究院）**

百度健康-TBC

阿里健康（左手医生）TBC

**报价对比**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **华为** | **价格** | **百度智能** | **价格** | **科大讯飞** | **价格** | **小冰** | **价格** | **智谱AI** | **价格** |
| 实时音视频 | **Y** | 6 元/千分钟 |  |  |  |  |  |  |  |  |
| 客服系统 | Y | 1人/月 100元 |  |  |  |  |  |  |  |  |
| 在线机器人 | Y | 5000元/月 |  |  |  |  |  |  |  |  |
| 变声 |  |  |  |  |  |  |  |  |  |  |
| 语音识别 | Y | TBC |  |  |  |  |  |  |  |  |
| 语音转文本 | Y | 1800元/1000小时 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**数字人开源项目**

#### Gitee - Fay数字人 AI Agent版

TheRamU/Fay: 语音互动，直播自动带货 虚拟数字人 (github.com)

<https://github.com/TheRamU/Fay>

<https://gitcode.com/TheRamU/Fay?utm_source=csdn_blog_hover>

fay: 这是一个数字人项目，包含python内核及ue数字人模型，可以用于做数字助理及自动直播，又或者作为你的应用入口也很帅 (gitee.com)

#### OpenAvatar

<https://github.com/topics/avatar>

一个开源的数字人形象框架，旨在创建一个可以进行高度定制化的数字人模型，帮助用户创建和部署高度可定制的数字人和虚拟形象。它提供一组标准接口和现成实现，使得开发人员能够快速轻松地构建具有不同形态和功能特性的数字人。

OpenAvatar 的特点包括：

* 多样性：支持不同的机器人和虚拟人物形象，并允许根据用户需求和场景需求自定义形象和特征。
* 实时表达：能够产生直观、自然、丰富的动作表达，适用于游戏、虚拟现实软件和社交媒体等环境。
* OpenAvatar 描述虚拟角色的 XML 规范，用于定义视频游戏或其他虚拟角色的物理特征。它不是指用户图标，而是玩家角色（或“卡通”）的物理特征。

实现包括im2vatar-3Dvisualiation.zip，这是一个可视化结果的项目，以及用于创建简单XML树的代码示例。

#### VideoReTalking

<https://github.com/OpenTalker/video-retalking>

<https://colab.research.google.com/github/vinthony/videoretalking/blob/main/quick_demo.ipynb>

用AI实现视频人物嘴型与输入的声音同步的创新技术。简单来说，就是输入任意一个视频和一个音频文件，在生成的新视频中，人物的嘴型会与音频同步。根据声音改变视频中人物的表情。整个过程不需要用户干预，都是自动完成的。很多博主通过此项目构建了属于自己的AI数字人。

#### VFA，Versatile Face Animator

<https://github.com/why986/VFA>

面部捕捉和面部重定向端到端地结合起来的自监督框架。仅需要输入一段Kinect甚至iPhone录制的RGBD面部表演视频，即可驱动任意的3D数字人，生成逼真、语义一致的数字人面部动画。生成逼真的数字人面部动画，

主流的构建数字人面部动画的技术是面部重定向（performance retargeting），即录制演员的面部表演视频，然后迁移到目标数字人上。实现面部重定向的主要方法是构建表情基（Blendshape），即预先构建一套关于数字人面部动作的表情集合，作为面部动作空间的基，重定向时拟合关于表情基的系数；

另一种常用的方法是构建面部绑定（facial rigging），即构建数字人面部每个部分的控制器，像提线木偶一样逐帧制作想要的动画。

共同缺点有：

* 需要耗费大量的人力成本和时间成本
* 缺乏通用的制作标准和流程

VFA分为两个模块：

1. RGBD Animation模块。这一模块基于Auto-Encoder架构，对输入的面部RGBD图像进行编码，训练时自监督地学习用一组正交基（此处称为motion dictionary)表征的面部动作空间，推理时将重构的隐层编码解码，从粗粒度到细粒度地逐层生成驱动结果RGBD帧。

2. Mesh Retarget模块。这一模块首先对目标数字人自动检测控制点，计算控制权重，然后在重定向时，根据估计的密集光流场和生成的驱动RGBD图像，计算控制点的运动，从而控制目标数字人每个点的形变，生成数字人面部动画。

#### Autogen

<https://github.com/microsoft/autogen>

是微软团队研发的一个多代理框架，利用它可以轻松定制一系列工作任务

#### shap-e

<https://github.com/openai/shap-e>

OpenAI（ChatGPT开发商）开源的一款新模型，用来生成以文本或图像为条件的 3D 对象。以其创新方法改变了 3D 应用领域。允许用户在计算机上无缝运行它，而无需 OpenAI API 密钥或互联网连接。

**数字人相关的开源大模型**

#### AI Voice Chat

<https://github.com/WeberJulian/>

简化版的react app，可以用自己的语音跟chatGPT 语音聊天, 用Whisper Large v3来转录，使用openchat 3.5 AWQ作为语言助手，XTTS v2用来文本转语音。优势是语言对语音的几乎无延迟特性。运行在RTX 3090 GPU上。

#### Diffusion Avatars

<https://tobias-kirschstein.github.io/diffusion-avatars/>

用来合成一个高保真的3D头像，提供对姿势和表情的控制。

1. 将表情传输到3D头像动画
2. 通过 NPHM 制作头像动画

通过底层 NPHM 进行控制。通过在几个目标表达式之间进行插值来获得表达式代码 zexp 。使用光栅化和我们基于扩散的神经渲染器，表达代码被转换为具有视点控制的现实化身。通过拖动蓝色点到相应的表情，完成3D人物表情的改变，中间的过渡很流畅

#### PoseGPT: 通过对话的方式实时生成3D人物姿态

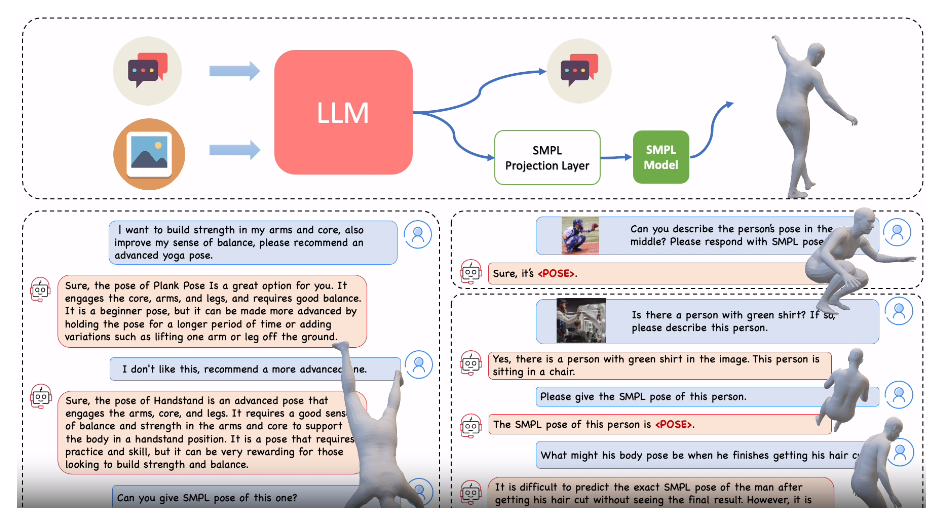
<https://yfeng95.github.io/posegpt/>

用大型语言模型 (LLM) 来从图像或文本描述中理解和推理 3D 人体姿势的框架。源于人类从单个图像或简短描述中直观地理解姿势的能力，将图像解释、世界知识和对肢体语言的理解交织在一起的过程。

传统的人体姿势估计方法，无论是基于图像的还是基于文本的，通常缺乏整体场景理解和细致入微的推理，导致视觉数据与其现实世界含义之间的脱节。

PoseGPT 通过将 SMPL 姿势嵌入为多模态 LLM 中的独特信号标记来解决这些限制，从而能够从文本和视觉输入直接生成 3D 身体姿势。从而促进两项高级任务：推测性姿势生成和姿势估计推理。这些任务涉及推理人类从微妙的文本查询中生成 3D 姿势，可能还伴有图像。我们为这些任务建立了基准，超越了传统的 3D 姿态生成和估计方法。

此外，PoseGPT 能够基于复杂推理理解和生成 3D 人体姿势，为人体姿势分析开辟了新的方向。



#### Animate Anyone

<https://humanaigc.github.io/ani>

这个模型实现从静止图像到人物动态视频。这个用来实现角色动画的一致性。引入了有效的姿势引导器来指导角色的运动，并采用有效的时间建模方法来确保视频帧之间平滑的帧间过渡。

#### 文本到3D模型的流程

https://3d.csm.ai/ 用来图片生成3D模型。

https://www.krea.ai/ 用来做参考制作理想模型

https://magnific.ai/ 用来将图片更加清晰

https://runwayml.com/ 做成动画

#### 图生成法向图导入Blender处理

<https://huggingface.co/spaces/f>

#### ChatGLM-6B

<https://github.com/THUDM/ChatGLM-6B>

一个开源的、支持中英双语的对话语言模型，基于 General Language Model (GLM) 架构，具有 62 亿参数。结合模型量化技术，用户可以在消费级的显卡上进行本地部署（INT4 量化级别下最低只需 6GB 显存）。

使用了和 ChatGPT 相似的技术，针对中文问答和对话进行了优化。经过约 1T 标识符的中英双语训练，辅以监督微调、反馈自助、人类反馈强化学习等技术的加持，62 亿参数的 ChatGLM-6B 已经能生成相当符合人类偏好的回答。

由于 ChatGLM-6B 的规模较小，目前已知其具有相当多的局限性，如事实性/数学逻辑错误，可能生成有害/有偏见内容，较弱的上下文能力，自我认知混乱，以及对英文指示生成与中文指示完全矛盾的内容。

**医疗领域开源大模型**

#### 华佗II-LLM4Med

预期目的是医疗和健康建议、分诊、诊断、开药、解释医疗报告等。

论文地址：<https://arxiv.org/pdf/2305.15075.pdf>

github 地址：<https://github.com/FreedomIntelligence/HuatuoGPT>

Demo 地址：<https://www.huatuogpt.cn/#/>

华佗GPT是港中大（深圳）旗下医疗健康领域的大模型。

HuatuoGPT (华佗GPT)：基于医学语料库训练的大型语言模型(LLM)，旨在构建一个更专业的医学咨询场景下的ChatGPT。该模型已发布技术报告和在线演示，并提供医学评估基准和训练数据。HuatuoGPT使用聚合的数据集进行微调，包括来自ChatGPT的指令数据、医生的实际指令数据以及真实世界的医学对话数据】

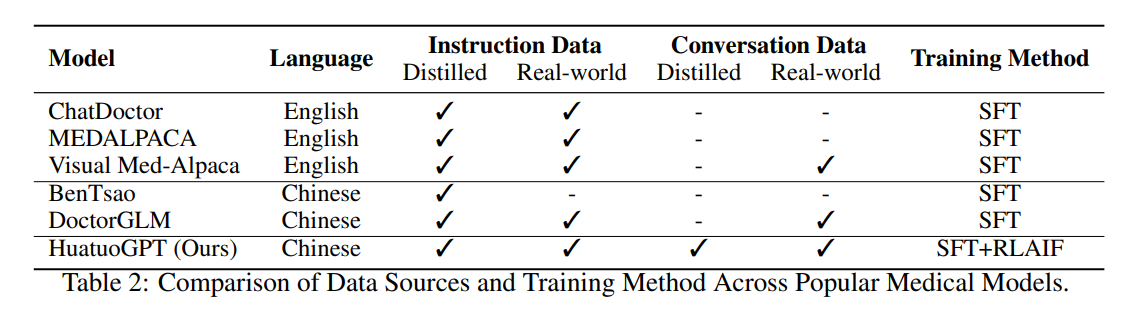
'HuatuoGPT (华佗GPT), Towards Taming Language Models To Be a Doctor. - HuatuoGPT, Towards Taming Language Models To Be a Doctor.' FreedomIntelligence

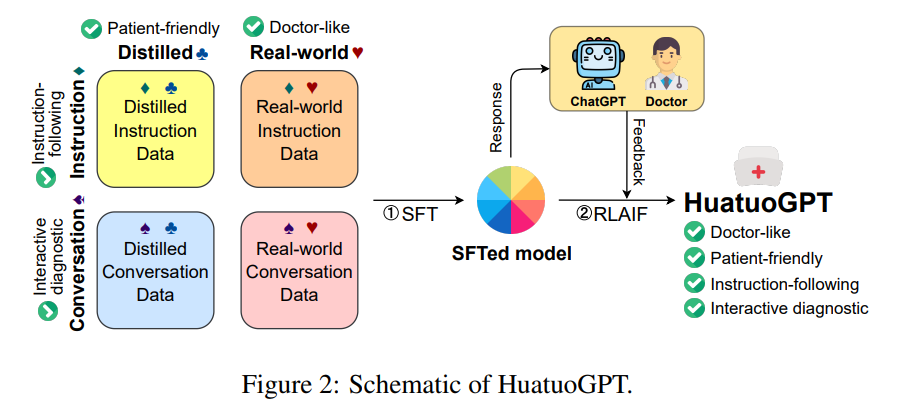
HuatuoGPT的核心，是在监督微调阶段利用来自ChatGPT的提取数据和来自医生的真实世界数据。ChatGPT的回答虽然具有流畅性和全面性，但它在一些方面不能像医生一样表现，例如在综合诊断方面。可以在SFT阶段将来自医生的真实数据将与来自ChatGPT的提取数据进行互补，而为了更好地利用两种数据的优势，训练了一个奖励模型，以使语言模型与两种数据带来的优点保持一致，并遵循RLAIF（从人工智能反馈中强化学习）。

HuatuoGPT的特点

* HuatuoGPT是第一个使用RLAIF来利用真实数据和提取数据（包括指令和会话数据）的优点的医学语言模型。
* 人类评估显示，HuatuoGPT优于现有的开源LLM和ChatGPT（GPT-3.5-turbo）。其性能与医生最相似。

解决方案，流行的医学模型中数据源和训练方法的比较：





用于SFT阶段的混合数据，从ChatGPT提取指令数据：遵循self-instruct的工作构建了一组医疗指导数据，旨在使模型能够遵循用户的医疗指导。不同之处在于，采用了自上而下的方式来创造更自然、更全面的应对措施。设计一个分类法来收集或手动创建基于角色和用例的种子指令。根据每个角色或用例，使用自我指导分别生成指令。

基于AI反馈的RL，在监督微调（SFT）阶段，引入了一个多样化的数据集，旨在使HuatuoGPT能够模仿医生的询问和诊断策略，同时保持LLM反应的丰富、逻辑和连贯特征。

奖励模型的构建，用真实的指令和对话作为训练数据，从微调的模型中采样多个响应。对于多回合对话，提供对话历史，以调整模型的响应生成。然后，这些response由LLM（如ChatGPT）进行评分，考虑到信息性、连贯性、对人类偏好的遵守以及基于给定真实医生诊断的事实准确性。

评估，在评估 HuatuoGPT 的性能表现上，团队成员采用了自动评估和人工评估两种方式相互验证，在单轮问答场景和多轮交互式诊断场景中分别进行了评估。

2023年2月，发布的“华佗GPT”是首个国内类ChatGPT的医疗大模型（已开源）

# 大康AI相关需求

|  |  |  |
| --- | --- | --- |
| **功能** | **描述** | **备注** |
| **AI 功能** |  |  |
| **导医** |  |  |
| 3D数字人选型（医生分类） | 依据线下医生数据集的画像，训练出可供选择的3D数字医生  刷选医生可通过APP模拟器生成URL |  |
|  |  |  |
| **导诊** | 导诊信息表  部位刷选、症状、时间和历史记录 |  |
| 语音互动 | 语音互动 和APP内互动同步进行 |  |
| 变声 | 不用原声，需要训练 | 具备 |
| 诊断知识库 | 基础支撑 |  |
|  |  |  |
| **AIGC** |  |  |
| 敏捷问答 |  |  |
| 互动闲聊 |  |  |
| 图文转换 | 文-图/视频/语音 | TBC |
| 文本摘要 |  | TBC |
| 内容检查 |  | TBC |
| FAQ |  |  |

# 数字人功能模块

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **方案** | **模块** | **功能点** | **优先级** | **示例** | **备注** |  |
| **基础功能** | 待技术方案 | 交互形象 | 听、说、看、学、动 | P0 |  | 听：拾音设备  说：降噪音箱  看：人脸识别  学：动捕数据、摄像头  动：口型、肢体、表情 |  |
|  |  |  | 2D/3D设计 | P0 |  |  |  |
|  |  |  | 3D形象  建模、绑定、驱动 | P0 |  |  |  |
|  |  | 渲染方式 | 实时ASR | P0 |  | 离线+算法 |  |
|  |  |  | 实时TTS | P0 |  | 离线+算法 |  |
|  |  |  | 变声 | P0 |  | 不用原声，需要训练 | 待确认 |
|  | 待技术方案 | 诊断知识库（医疗领域） |  | P0 |  | 领域模型/专业领域图谱-医疗（需特别关注中医领域）  模型治理/自学习/知识推理 |  |
|  |  |  |  |  |  |  |  |
| **业务平台**  **AIGC应用** | 待技术方案 | 内容生成 | 填空选择 |  |  |  |  |
|  |  |  | 故事生成 |  |  |  |  |
|  |  |  | 文本摘要 |  |  |  |  |
|  |  |  | 内容检查 |  |  |  |  |
|  |  |  | 图文转换 |  |  |  |  |
|  |  | 智能问答 | 敏捷问答 | P0 |  |  |  |
|  |  |  | 互动闲聊 | P0 |  |  |  |
|  |  |  | 评分系统 |  |  |  |  |
|  |  |  | FAQ | P0 |  |  |  |
|  |  |  | 数据处理 | P0 |  |  |  |
| **业务对接** |  | 大康业务数据接入 |  | P0 |  |  |  |
|  |  |  |  |  |  |  |  |