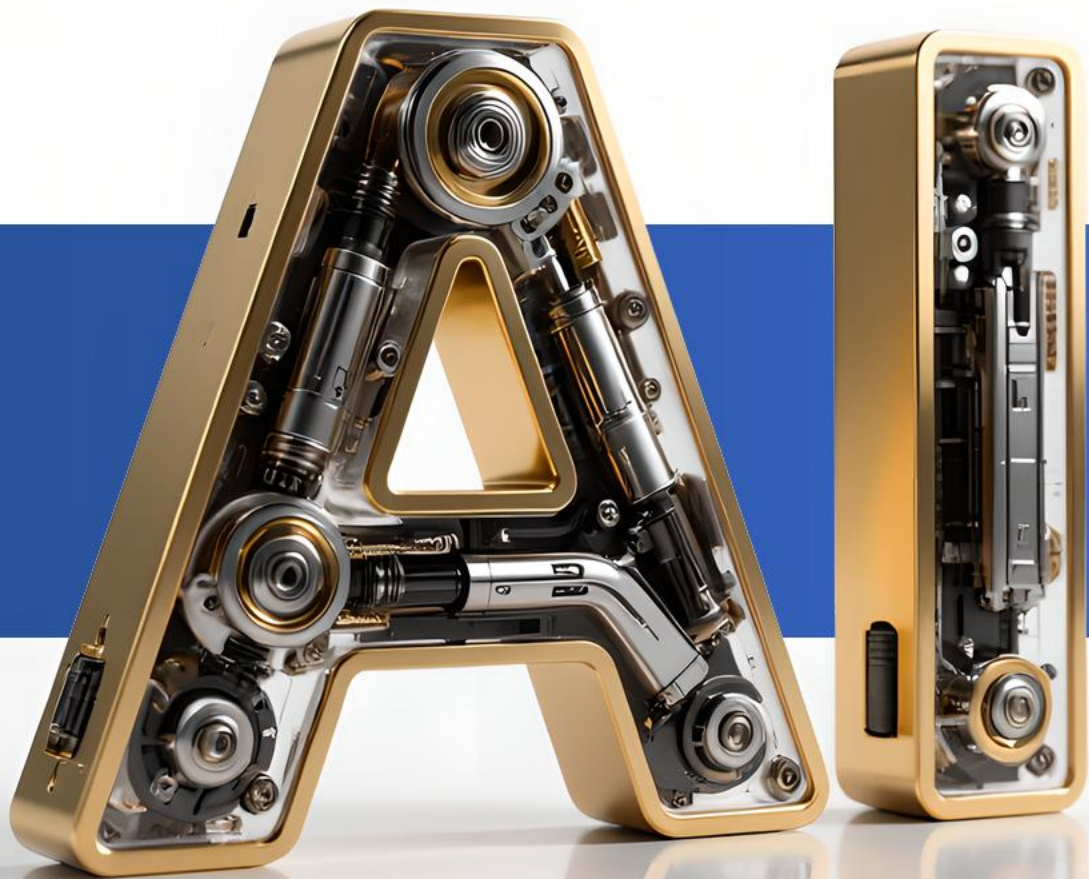




深圳市人工智能产业协会
Shenzhen Artificial Intelligence Industry Association



数字创意项目制培训

Digital Creative Project Training

AI数字人制作全流程



目录

CONTENTS



01

数字人制作要素

02

AI生图到视频

03

免费对口型工具

04

4K修复工具实操

01

数字人制作要素

生成您的专属数字人

数字人制作要素



数字人制作四要素

首帧图

图生视频的首帧图，数字人的颜值取决于首帧图的形象是否好看。

AI语音声音

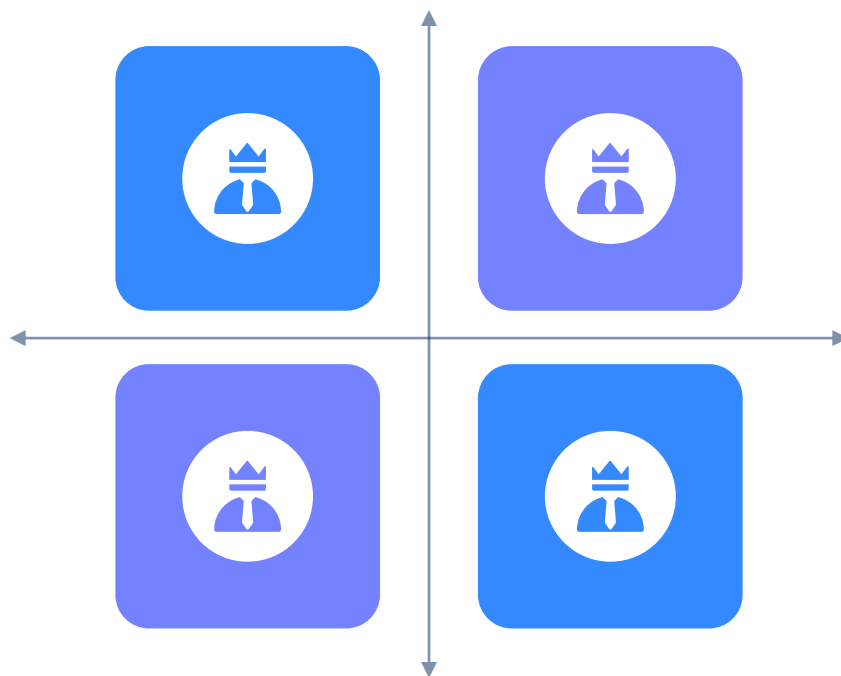
包括数字音效、环境音效、MP3音效等不同类型的，满足不同场景需求。

语音文案

视频语音的文案是视频的灵魂，决定了观众对整个视频有没有兴趣，决定了视频的完播率。

4K修复成片

使用软件4k修复成片。



02

AI生图到视频

生成您的专属数字人

AI形象生成

输入自己的照片：

- 1.输入提示词：我想AI生图，生成一个一模一样的图片，帮我反推生成的提示词。
- 2.输入参考图，并输入豆包生成的提示词，生成满意的图片。



我想AI生图 生成一个一模一样的图片 帮我反推生成的提示词

要生成与这张图片几乎“一模一样”的 AI 生图提示词，需涵盖人物外观、着装、姿态、环境等关键信息，以下是较为细致的反推示例（不同 AI 绘图模型对提示词理解有差异，需根据实际微调）：

人物主体：一位亚洲男性，黑色中长发，自然垂落，佩戴耳饰，身穿纯黑色简约短袖上衣（领口为小 V 型设计）、黑色下装，左手手腕戴红色珠串手链，端坐在黑色扶手椅上，双手呈特定手势（掌心向上、手指自然舒展），面部表情平静、自然

环境背景：中式风格屏风（带有浅蓝花卉、莲蓬图案），室内明亮光线，背景呈现简洁的室内空间氛围

你可以基于此，结合使用的 AI 绘图工具（如 Midjourney 需遵循其提示词格式，添加风格、比例等参数；Stable Diffusion 可补充模型、采样方法等信息）进一步调整，比如用更精准的描述替换括号内容，让生成效果更贴近原图。要是用特定模型，还得补充像“realism style（写实风格）”“8k resolution（8K 分辨率）”这类适配的风格、画质词，来提升还原度。



AI形象生成



输入自己的照片：

- 1.输入图片
- 2.输入提示词：把双手的位置调整到胸前，做讲解视频手势状，人物轻微侧身，保持人物五官特征不变。

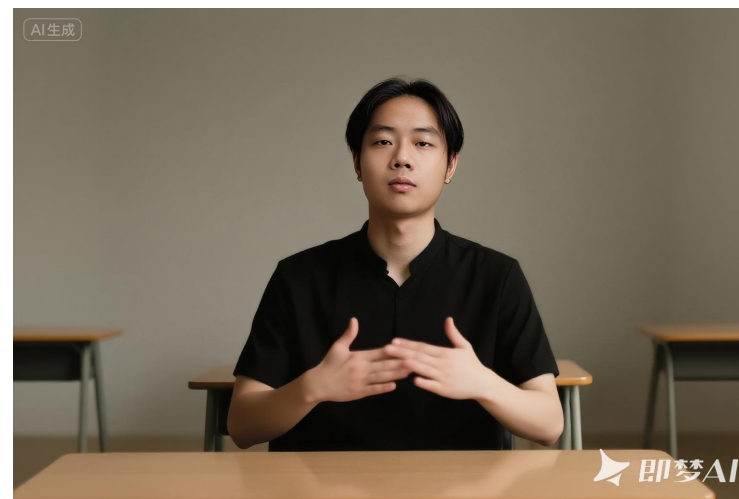
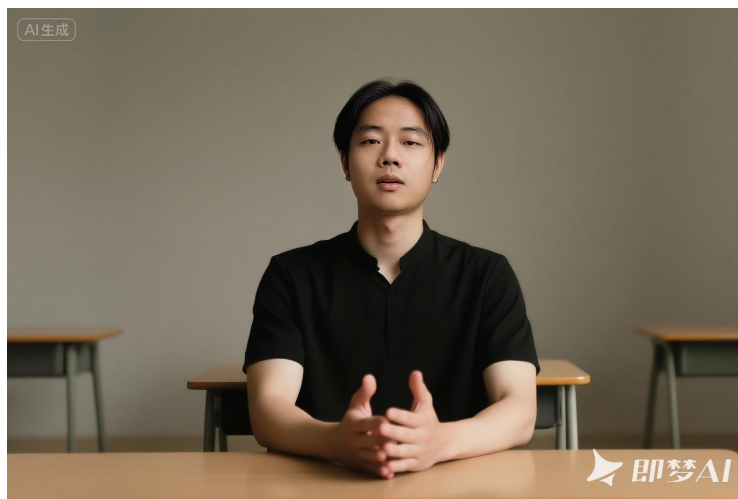
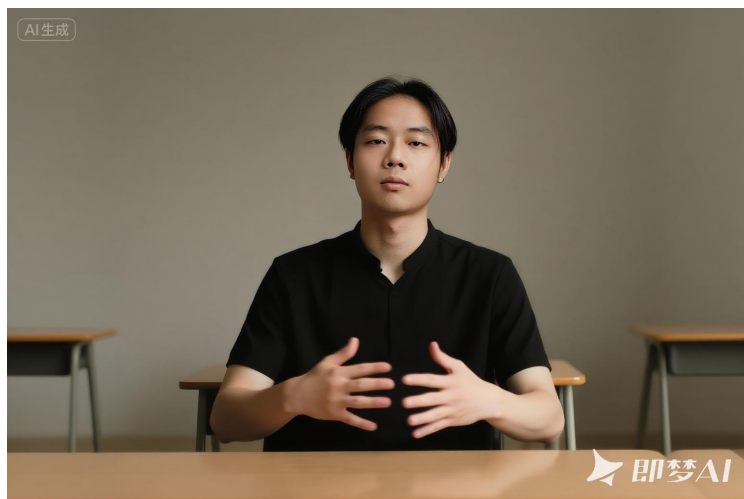
AI形象生成



输入自己的照片：

- 1.输入图片
- 2.输入提示词：把双手的位置调整到胸前，做讲解视频手势状，人物25度轻微侧身，保持人物五官特征不变

使用连续帧生成连贯视频



输入连续帧图片：

- 1.输入连续帧图片；
- 2.输入提示词：男主播在微笑着讲解视频，眼睛看着前方，手部随着视频讲解随机做动作，不要做拜拜的动作，镜头锁定；
- 3.生成一段10s的视频。

口播文案生成



输入自己的口播文案需求：

输入提示词：帮我生成一段口播文案 用于AI数字人制作的教程

帮我生成一段口播文案 用于AI数字人制作的教程

📄 🔗 分享 ...

我将从准备工作入手，逐步讲解 AI 数字人口播文案制作的关键步骤，包括设定风格、撰写内容、调整细节等，帮助你顺利完成制作。

📄 AI数字人口播文案制作教程
创建时间：17:42

这个教程文案涵盖了制作的主要环节。如果你觉得某些步骤需要更详细的讲解，或者想换一种风格，都可以告诉我。

每一句分一段

我会按照你的要求，将每一句都分成一段，让文案的结构更清晰，方便查看和使用。

📄 AI数字人口播文案制作教程
创建时间：17:42

这样分段后，每句的独立性更强了。你看看是否符合预期，要是还有其他修改想法，比如对某些句子进行调整，都可以告诉我。

🔊 📄 ↺ 📄 🔗 分享 ... | 👍 🗑️

发消息、输入 @ 选择技能或 / 选择文件

大家好，今天来教大家用 AI 数字人制作口播内容，新手也能快速上手，跟着步骤做就行啦。

首先，咱们得明确口播的场景和风格。

是想做知识科普、产品介绍，还是日常分享？

不同场景的语气差别很大。

比如卖货就得热情有感染力，讲知识就要沉稳清晰。

确定好后，打开你常用的 AI 数字人工具，先选一个贴合风格的数字人形象 —— 比如职场风选西装造型，生活类选休闲穿搭，这一步能让观众更快代入。

接下来是写文案，这可是核心。

记住三个要点：**口语化、有节奏、信息明确**。

别用太复杂的句子，就像平时聊天一样，比如不说“此产品具备高效能特质”，而说“这东西用起来特别顺手，效率超高”。

段落别太长，每说两三句话就换行，这样数字人读的时候会有自然的停顿，听着不费劲。

如果是带货，开头可以加个钩子，比如“你是不是也遇到过这种麻烦？今天这个方法能帮你解决”，先抓住注意力。

然后把文案导入工具，调整数字人的语速和语气。

一般来说，正常语速就好，每分钟 200 到 220 个字，太快听不清，太慢容易走神。

如果文案里有重点词，比如“限时优惠”“关键步骤”，可以让工具在这些地方加重语气，或者配上抬手、点头的小动作，增强表现力。

最后检查一遍整体效果。

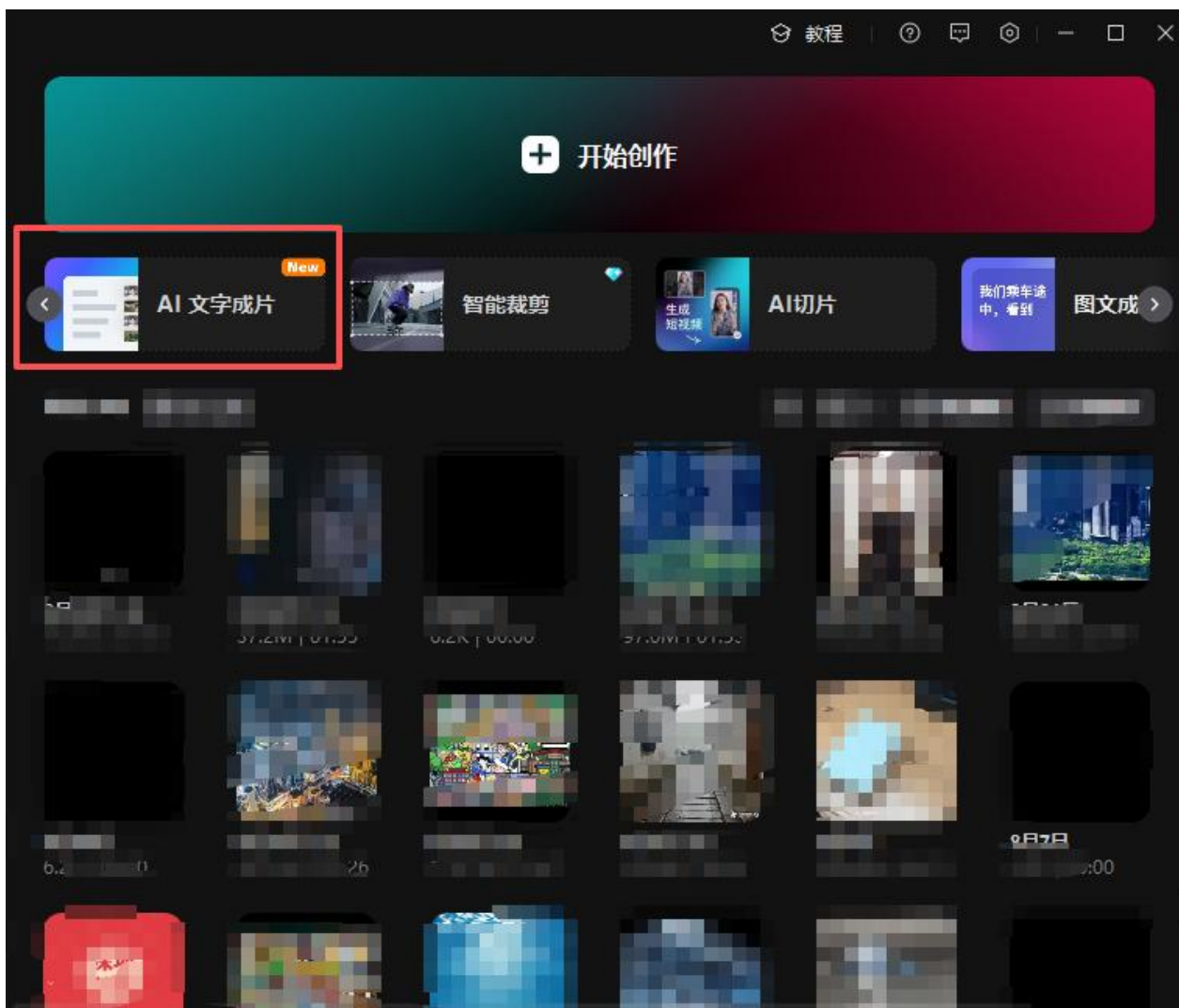
播放预览时，留意数字人的口型是不是和文字对上了，表情有没有违和感。

如果有不自然的地方，微调一下文案的断句，或者换个更匹配的数字人形象。

搞定这些，一条流畅自然的 AI 数字人口播就做好啦，是不是很简单？

赶紧试试吧！

口播语音生成



口播文案语音生成：

输入提示词：帮我生成一段口播文案 用于AI数字人制作的教程

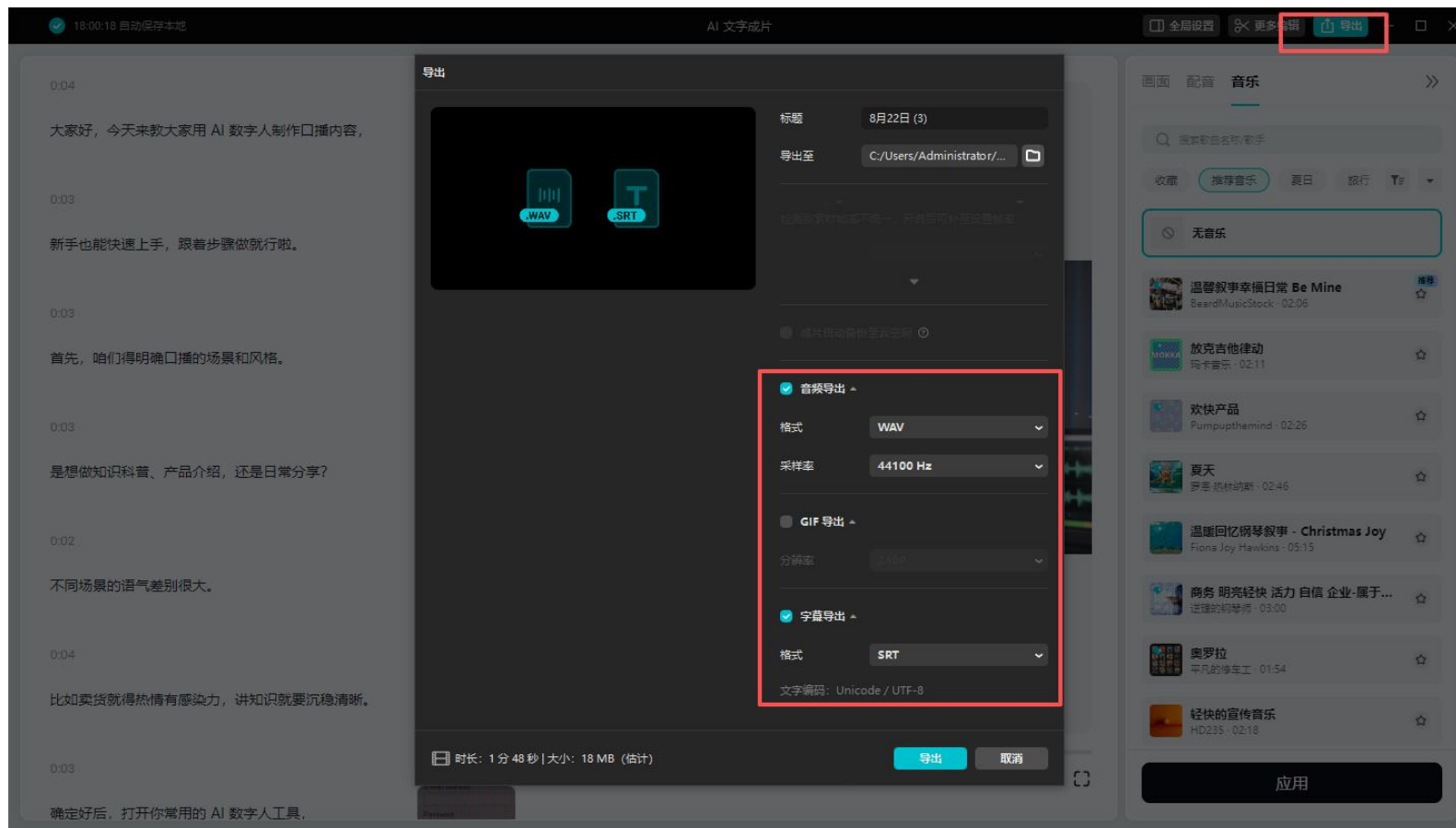
口播语音生成



口播文案语音生成：

输入文案后，选择好配音，点击视频生成。

口播语音生成



口播文案语音生成：
点击导出，只勾选字幕和音频

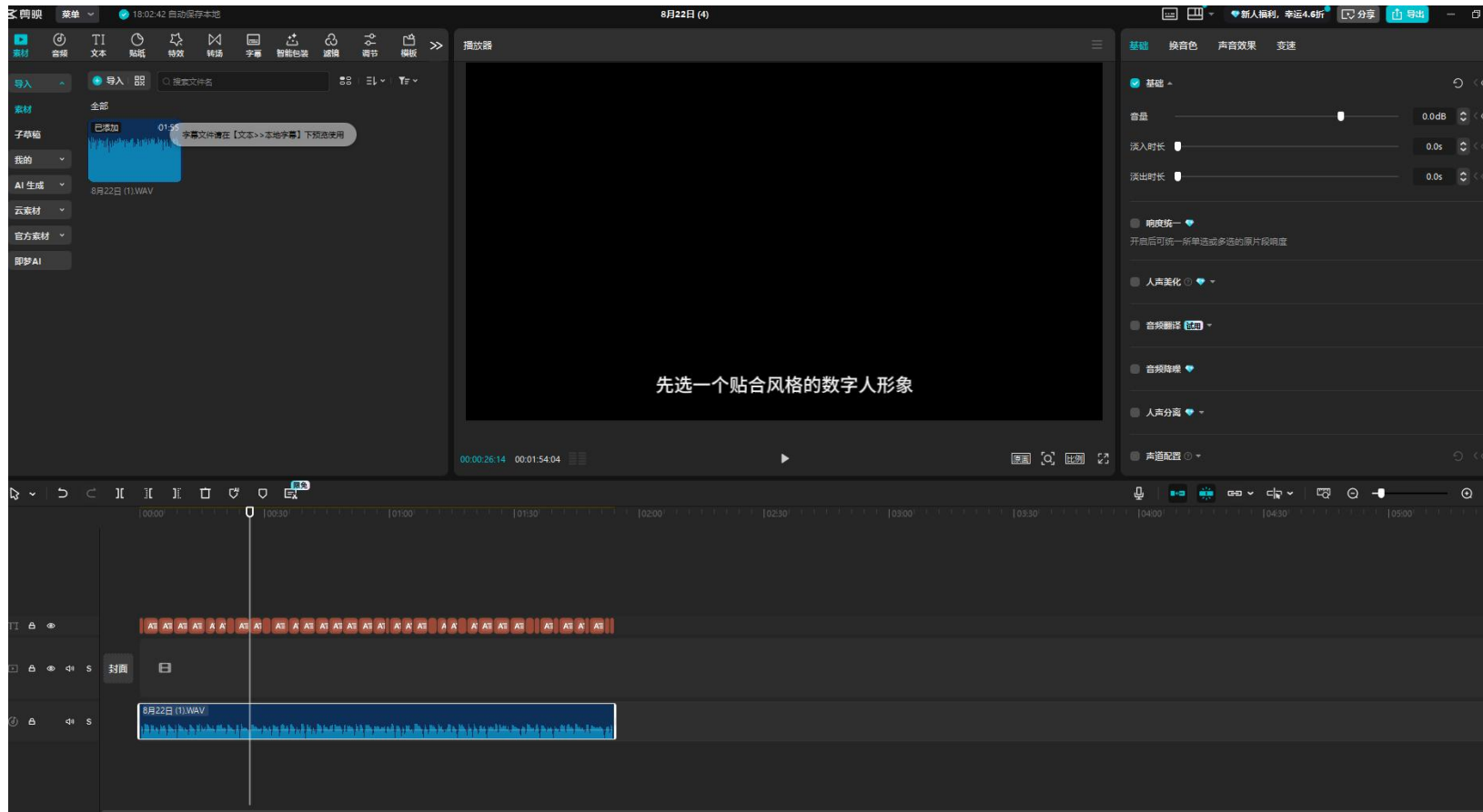
口播语音生成



口播文案语音生成：

点击开始创作，直接拖拽音频和字幕文件进入剪映即可。

口播语音生成



口播文案语音生成：

点击开始创作，直接拖拽音频和字幕文件进入剪映即可。

03

免费对口型工具

生成您的专属数字人

免费对口型工具——HeyGem



什么是HeyGem：

HeyGem 是硅基智能推出的一款开源数字人模型，专为 Windows 系统设计。

它是一个离线 AI 头像生成工具，用户可以通过它利用自己的外貌和声音生成逼真的 AI 化身，无需连接云端，也无需分享数据，保证了用户的隐私。

HeyGem 仅需 1 秒视频或 1 张照片，就能在 30 秒内完成数字人形象和声音克隆，在 60 秒内合成 4K 超高清视频。它支持多语言输出、多表情动作，具备 100% 口型匹配能力，在复杂光影或遮挡场景下也能保持高度逼真的效果。此外，HeyGem 还支持低配置硬件部署，极大地降低了使用门槛，为内容创作、直播、教育等场景提供了高效、低成本的数字人解决方案。

HeyGem 的项目代码在 GitHub 上开源，开发者可以基于其框架进行二次开发，拓展更多应用场景。

在云服务器中使用HeyGem



服务器链接：

<https://www.xiangongyun.com/image/detail/45071494-26e9-45ed-96b0-03be05ee6ddc?r=16E6MH>




视频Ai对口型+HeyGem=艾橘溪

上传视频+语音=Ai对口型=20250316

镜像作者：艾橘溪 最后更新：2025-03-16  仙宫云认证创作者

47.06 GB

¥ 0.10 / 镜像每小时费用

 立即部署

主要功能：上传视频对口型，720P+分辨率，支持最长30分钟语音！

新用户福利：新用户注册赠送8元体验金，大概4小时免费！

视频教学：<https://www.bilibili.com/video/BV1X9X7Y5EPi/>

=== 对口型基本要求 ===

1.上传的视频要有声音（随便搞一首背景音乐）同时视频必需是h264编码，才能提交定制成功（可以用剪映导入原视频，然后发布的时候要选择H264编码）。

2.分辨率最低720P，最高不要超过1920x1080，时长不要超过30分钟。一开始先用10~20秒的视频和语音测试。

3.视频最短不能低于8秒。全程出现人脸，人脸不能太小，也不能太侧脸，人脸不能转身消失。

在云服务器中使用HeyGem



48 G RTX 4090 D
显存 ¥2.59 / 小时

24 G RTX 4090 D
显存 ¥1.59 / 小时

24 G RTX 4090
显存 ¥1.89 / 小时



请选择部署 GPU 型号

已选择 公开镜像

 ✓ 视频Ai对口型+HeyGem...
@艾楠溪 · ¥ 0.10

8月份新卡预告：新增 800 张 RTX 4090D 48G

← 请选择镜像

 ✓ 视频Ai对口型+HeyGem...
@艾楠溪 · ¥ 0.10

48 G RTX 4090 D
显存 ¥2.59 / 小时

24 G RTX 4090 D
显存 ¥1.59 / 小时

24 G RTX 4090
显存 ¥1.89 / 小时

GPU RTX 4090 D × 1 2 3 4

配置信息

24 GB × 1
显存

56 GB
内存

14核
CPU

3900Mbps
公网速率

磁盘配置

镜像尺寸
100 GB + 47.06 G 免费容量

系统盘 扩容

扩展功能

云储存 初始化

SSH 密钥 选择

计费方式

按小时计费 按天计费 按周计费 按月计费

按量计费

$$\text{¥1.69} = \text{¥1.59} + \text{¥0.1} \times 1$$

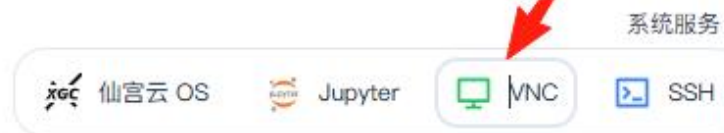
每小时 基础 镜像 GPU

确认部署

账户余额 ¥ 6.666307 充值

在云服务器中使用HeyGem

AI 模型生成违法违规、涉黄涉暴、违背公序良俗的内容以及进行虚拟货币相关业务！
或收到检举投诉，将直接封禁账号留存取证，并将相关数据、IP记录等信息提供至有关部门。



VNC

启动 VNC 后，实例会自动安装并运行 VNC 服务，运行时会占用少量的显存来保障 GUI 的流畅运行。

启动 VNC

停止 VNC

按量计费 ¥ 1.69 /小时

系统服务



仙宫云 OS



Jupyter



VNC



SSH

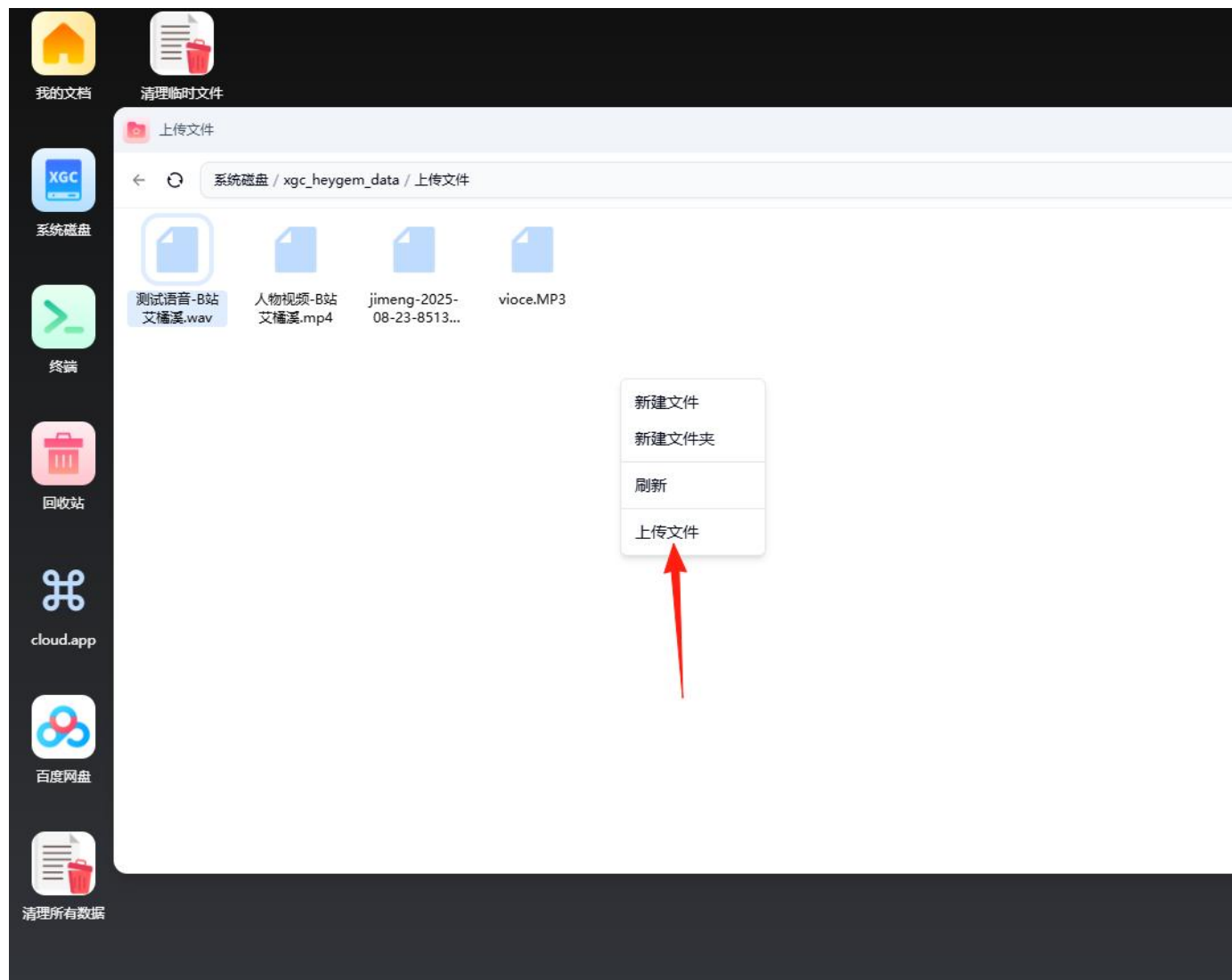
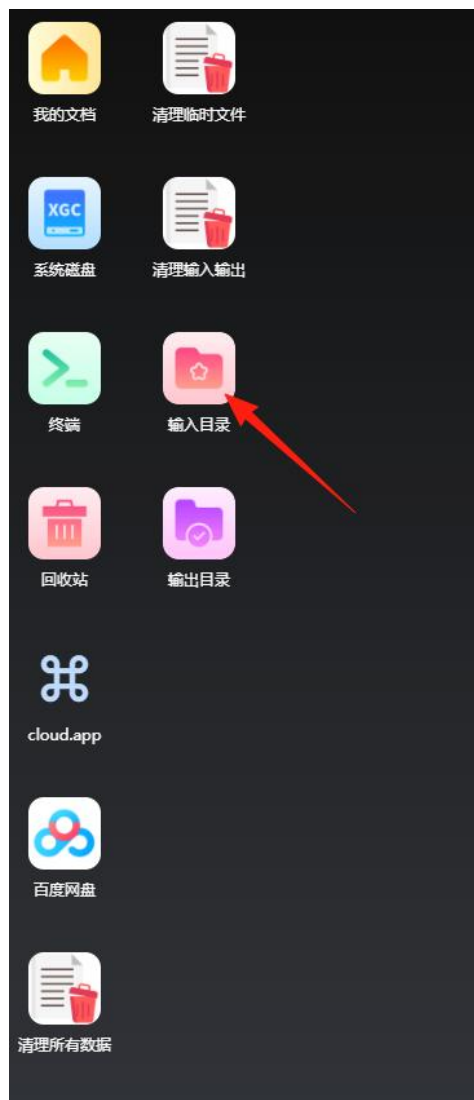
镜像自定义服务



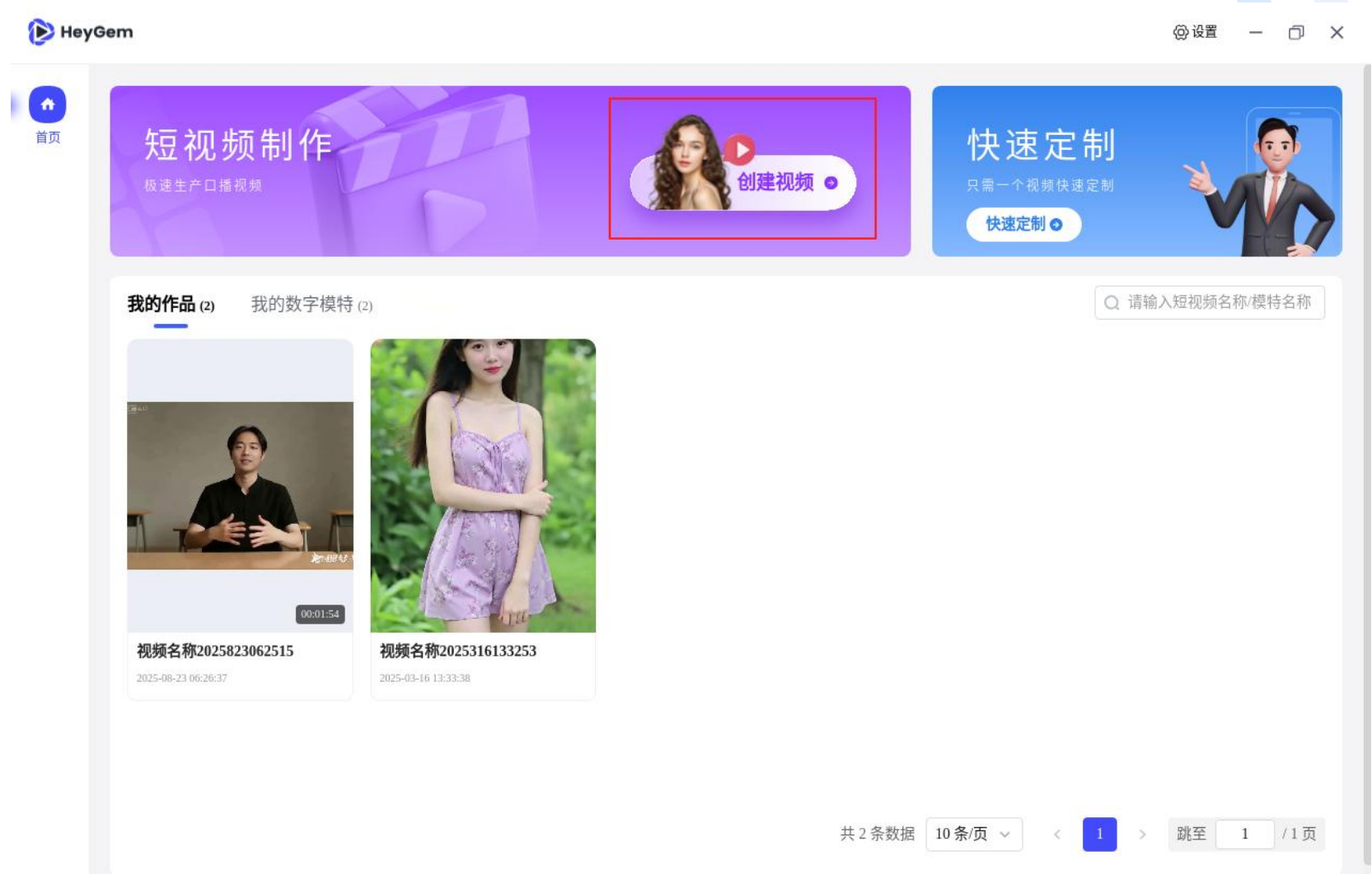
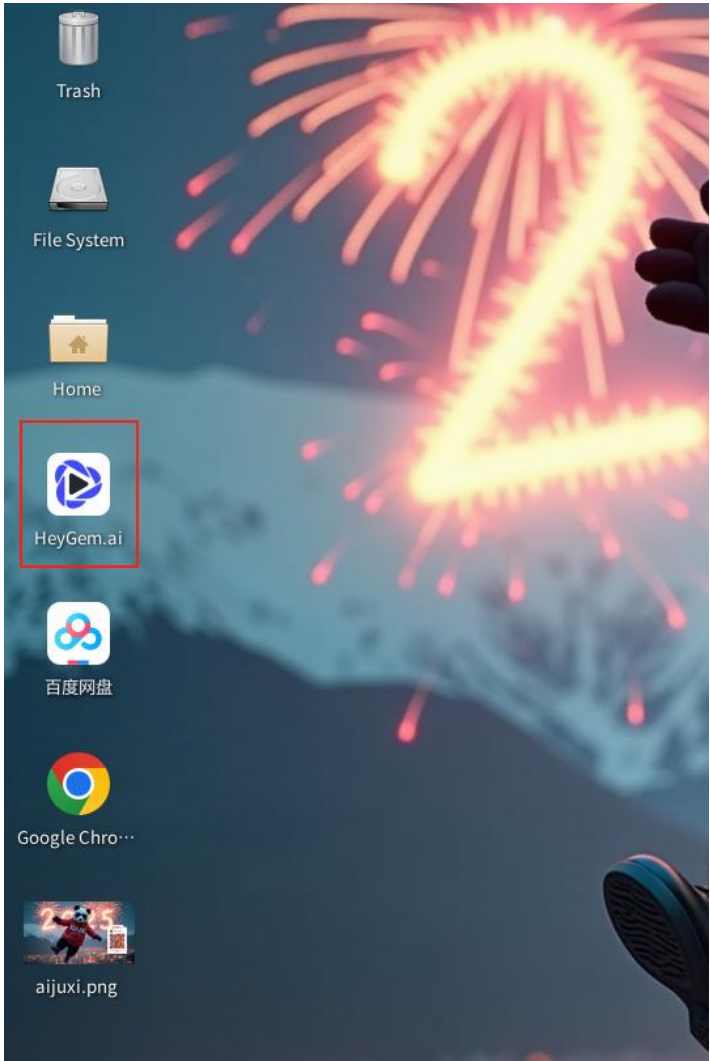
WebUI

按量计费 ¥ 1.69 /小时

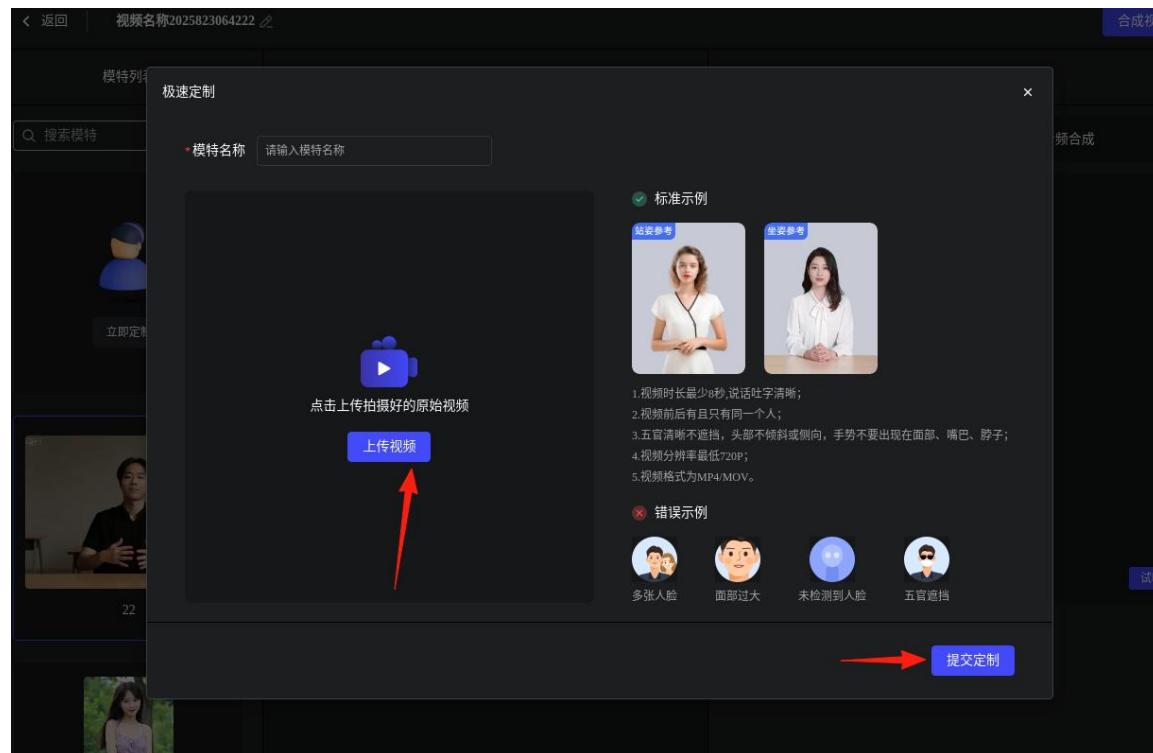
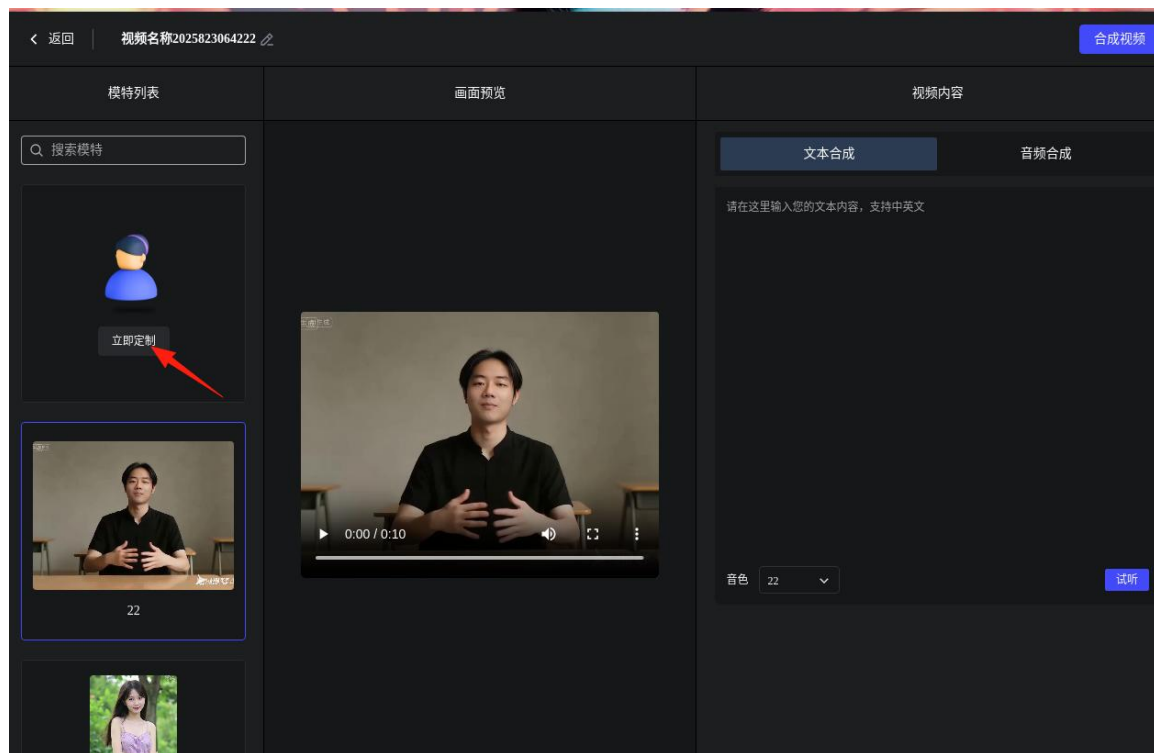
进入仙宫云OS



进入VNC



创建你的数字人模板

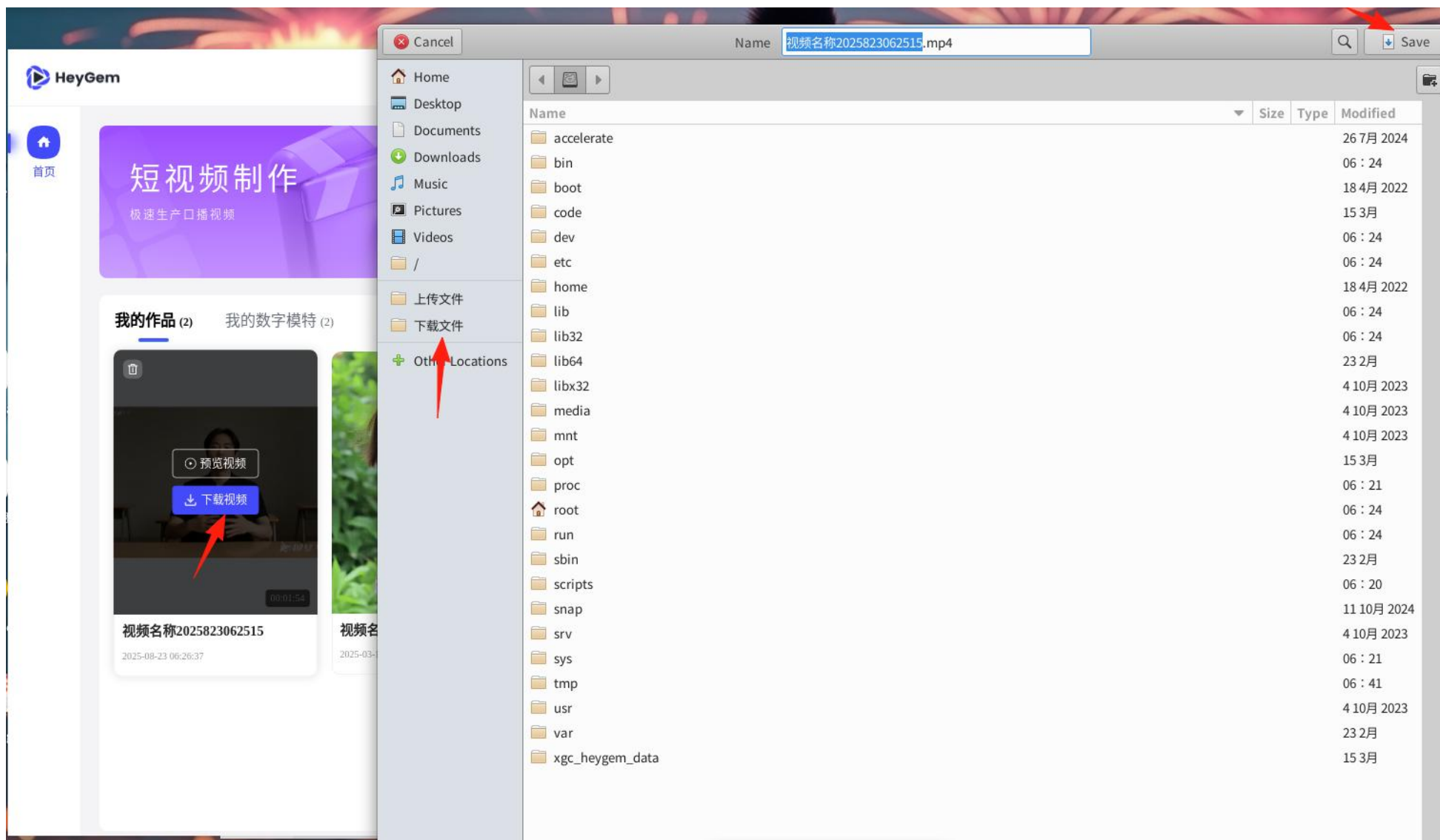


注意事项：

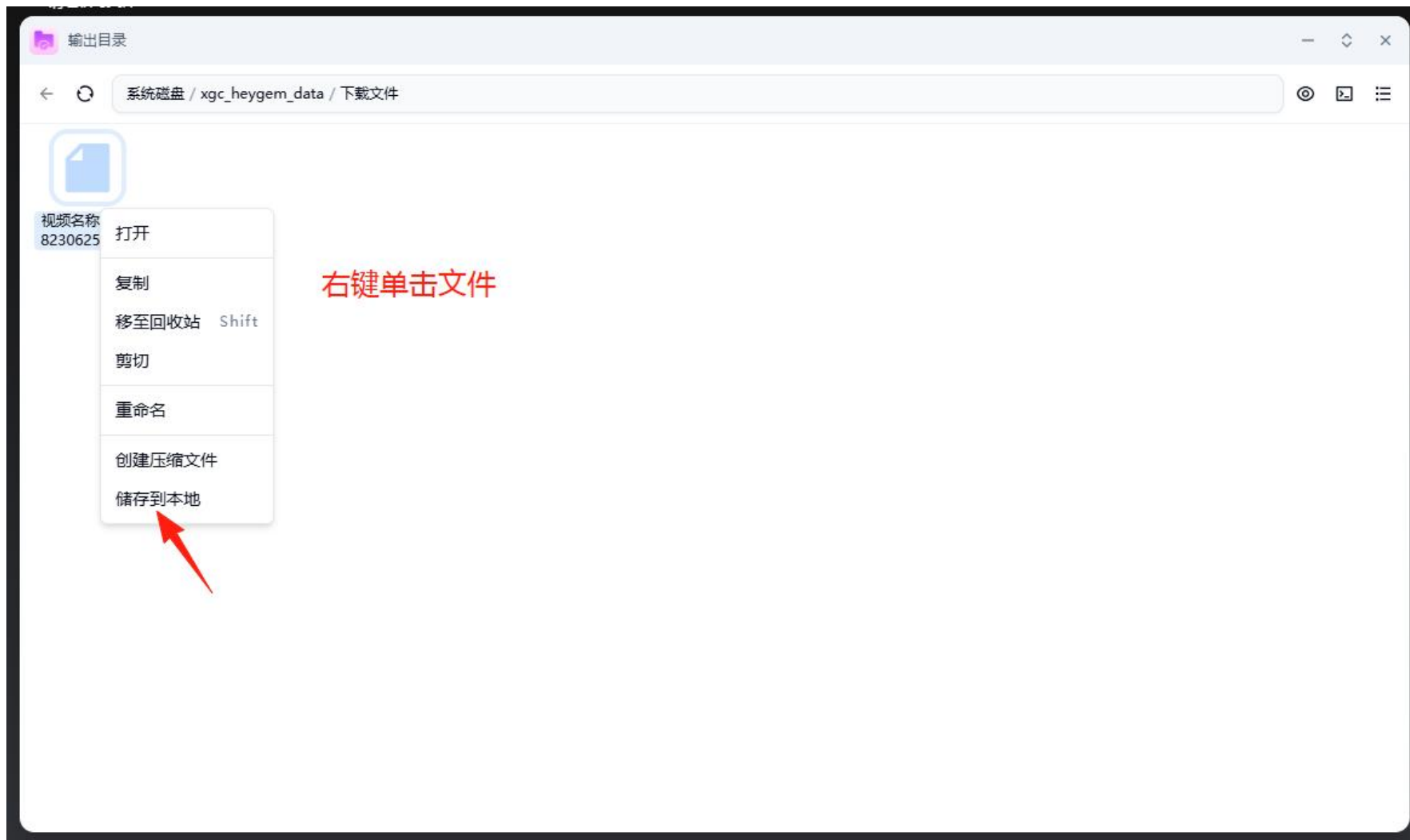
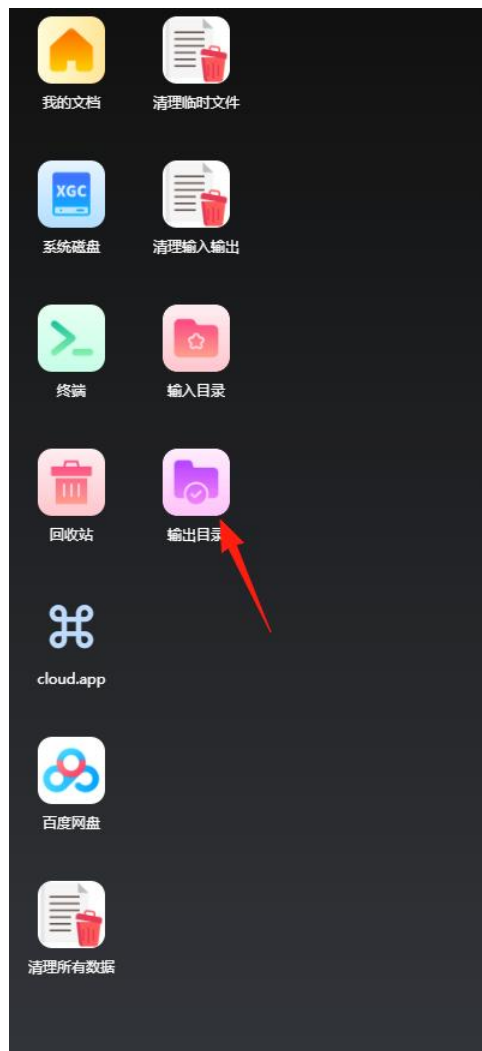
- 1.上传的视频要有声音；
- 2.视频的编码格式需要是H.264；
- 3.人脸大小适中；
- 4.画面中只能有一张人脸；
- 5.不能遮挡五官。



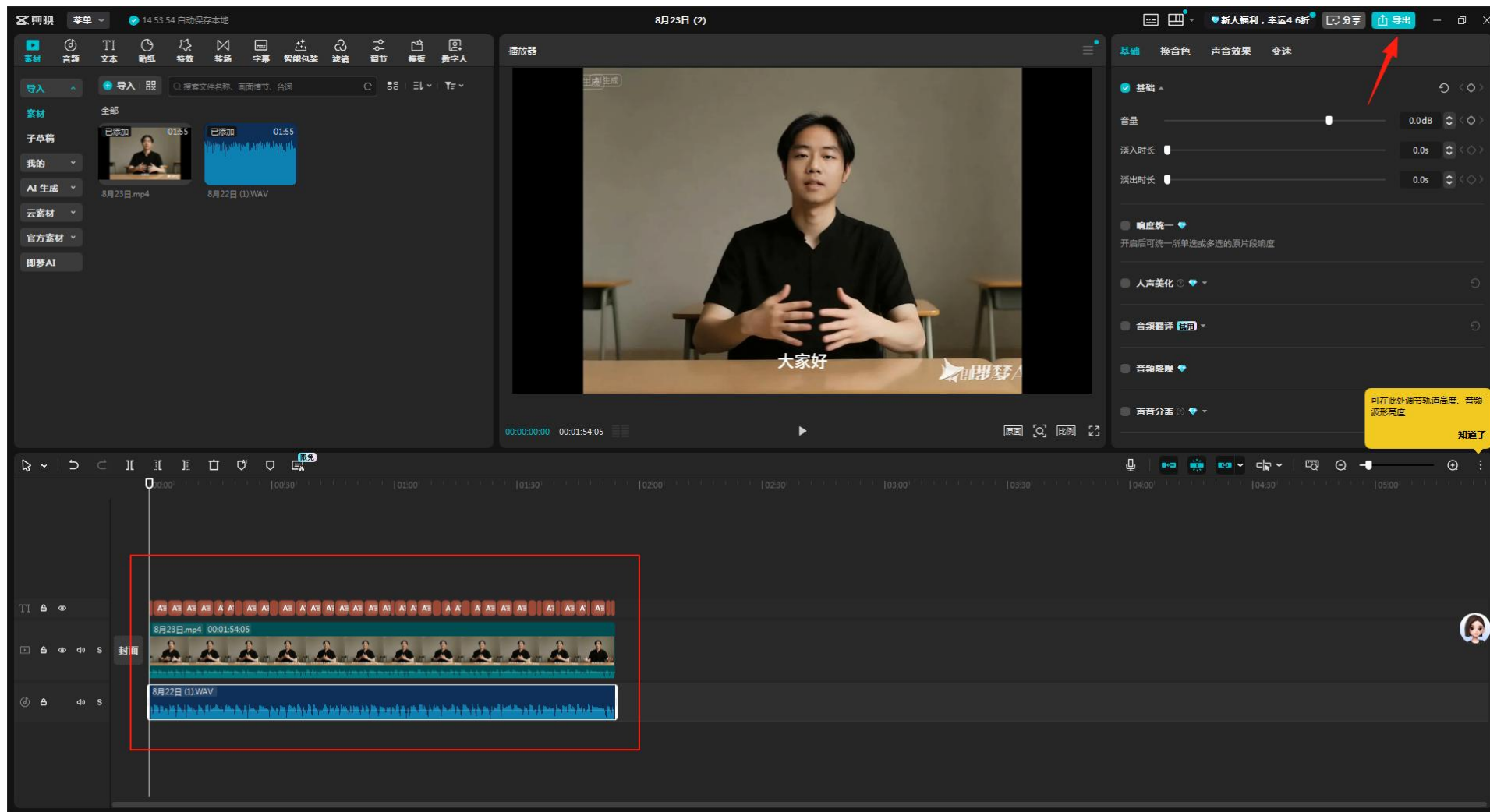
如何保存对好口型的视频



如何保存对好口型的视频



回到剪映合成视频、音频与字幕



04

4K修复工具实操

生成您的专属数字人

什么是Topaz Video Ai



Topaz Video AI 是一款专业的 AI 视频增强和超分辨率工具。

它采用深度学习技术，主要具备以下功能特点：

AI 超分辨率放大：可无损放大视频分辨率，将低清视频转换为高清，最高支持 8K 输出。例如，能把 DVD 画质的视频提升到适合在 4K 大屏电视上观看的蓝光画质。

视频降噪与去模糊：智能分析视频画面，去除噪点，修复因对焦不准、手抖等造成的模糊，重塑清晰细节，让视频更干净、清晰。

智能帧插值：能将 24FPS 或 30FPS 的视频升级至 60FPS 或更高，使运动视频、动画补帧等更加顺滑，制作慢动作时细节也更流畅自然。

老旧视频修复：针对老电影、旧动画等有很好的修复效果，可提升画面质量，还原色彩，甚至能清晰展现一些细微的纹理。

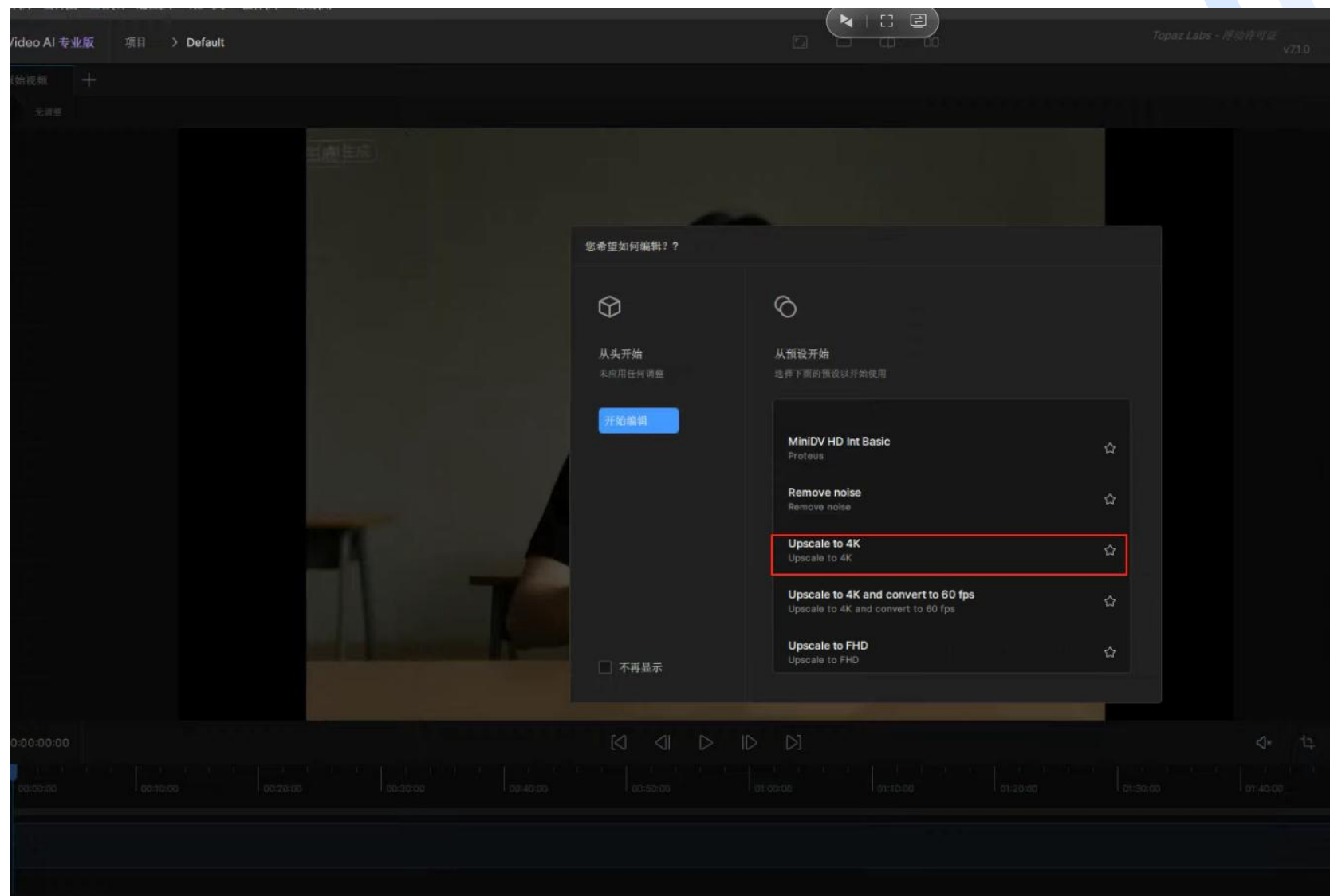
Topaz Video AI 适用于多种场景，如电影修复爱好者可以用它高清修复老电影，动漫收藏者能让旧动画变得更清晰，游戏玩家可优化游戏录制视频，摄影 / 摄像爱好者能提升手持拍摄视频的清晰度，短视频创作者也可以通过它增强视频质量，让作品更专业。

该软件可直接在 Mac 或 PC 上运行，支持独立版按顺序批量处理素材，也可作为插件集成到现有的工作流程中。它经过数百万个视频帧的训练，能够智能地理解噪点和细节之间的区别，还具备时间感知面部识别等功能，可自动、即时地纠正人物和面孔的相关问题。

如何使用Topaz Video Ai

使用AI工具放大视频：

导入视频后，选择放大到4K。



如何使用Topaz Video Ai



使用AI工具放大视频：

选择人脸专用的AI模型



输出视频



谢谢观看

