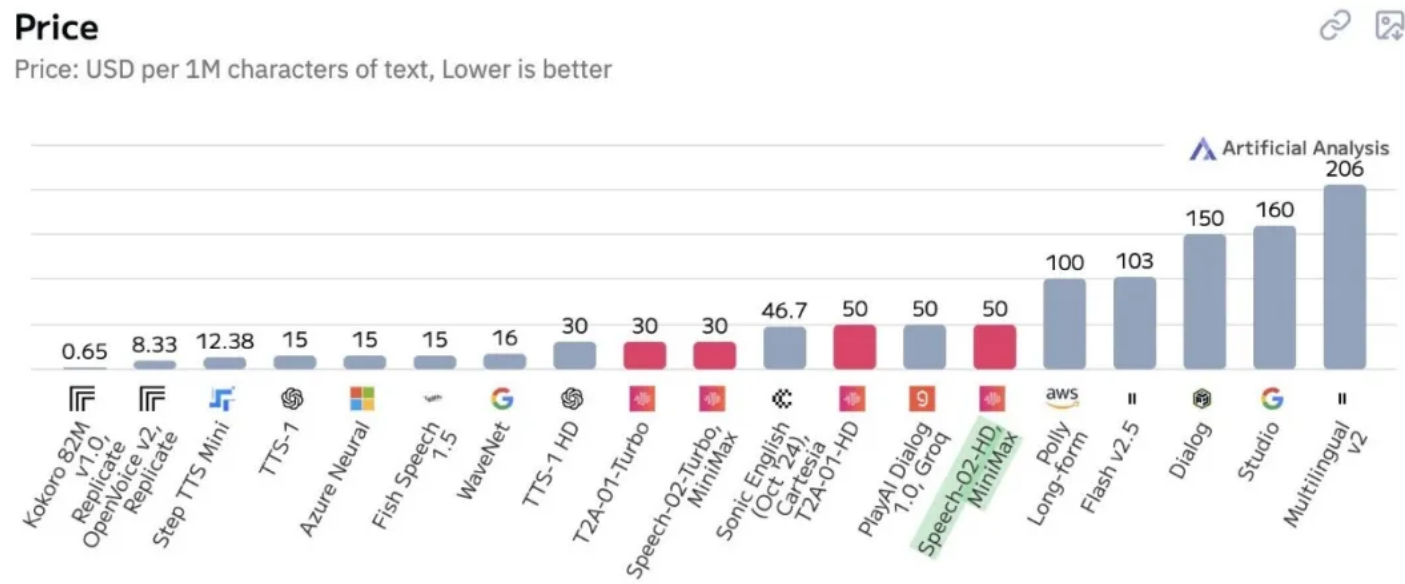


# 核心竞品追踪0514

## MiniMax发布TTS系统MiniMax Speech 02

- 功能：
  - 任意语言×任意口音×任意音色的个性化语音
  - 目前官方称可以“驾驭” 32语种、不同口音、不同情绪的人声
  - MiniMax Speech 02 价格更低，为3.5元/万字符，分别是 ElevenLabs Flash V2.5 与 Multilingual V2 的一半与四分之一
- 性能测试表现：
  - 在两项全球权威语音基准测评榜单：Artificial Analysis Speech Arena 和 Hugging Face TTS Arena 中，MiniMax Speech（在榜中对应为Speech-02-HD）超越了 OpenAI、ElevenLabs 等全球性能优异的模型，双双位列第一
  - Arena的ELO评分（主要是根据随机听取过程中用户的听感测试的）中MiniMax Speech 02 的听感更加优异。
  - 基于自己构建的专用测试集中在主流欧美语种上表现不错，在亚洲语种上字错率回会超10%



Artificial Analysis Speech Arena Leaderboard

Creator	Model	Arena ELO	95% CI	# Appearances
MiniMax	Speech-02-HD	1163	-23/+24	1271
OpenAI	TTS-1 HD	1158	-14/+13	6589
MiniMax	Speech-02-Turbo	1148	-23/+24	1241
OpenAI	TTS-1	1135	-14/+14	6471
ElevenLabs	Multilingual v2	1116	-13/+13	7842
ElevenLabs	Turbo v2.5	1111	-14/+13	7652
ElevenLabs	Flash v2.5	1109	-17/+16	2988
Cartesia	Sonic English (Oct '24)	1106	-13/+12	8128
OpenAI	GPT-4o Realtime Preview	1100	-29/+31	653
Kokoro	Kokoro 82M v1.0	1089	-17/+16	2898
MiniMax	T2A-01-HD	1080	-16/+15	2969
OpenAI	GPT-4o mini TTS	1068	-33/+33	496
Amazon	Polly Generative	1062	-16/+15	3869
Microsoft Azure	Azure Neural	1059	-12/+13	7916
Amazon	Polly Long-Form	1058	-15/+14	4545
Speechify	Simba	1046	-15/+16	3167
Google	Studio	1039	-15/+15	4805

语音测试

👤 TTS-AGI / TTS-Arena-V2

👍 like 737

🔄 Running on CPU UPGRADE

Leaderboard

TTS

Conversational

Top Voters

Public

Personal

Rank	Model	ELO
#1	MiniMax Speech-02-HD	1779
#2	MiniMax Speech-02-Turbo	1766
#3	Eleven Multilingual v2	1583
#4	Kokoro v1.0	1519
#5	MegaTTS 3	1494
#6	Cartesia Sonic 2	1493
#7	Eleven Flash v2.5	1491

听感测试

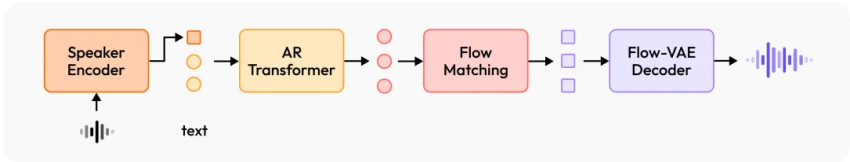
Table 2 | Objective Evaluation Metrics on the Multilingual Test Set.

Language	WER ↓		SIM ↑	
	MiniMax	ElevenLabs	MiniMax	ElevenLabs
Chinese	2.252	16.026	0.780	0.677
English	2.164	2.339	0.756	0.613
Cantonese	34.111	51.513	0.778	0.670
Japanese	3.519	10.646	0.776	0.738
Korean	1.747	1.865	0.776	0.700
Arabic	1.665	1.666	0.736	0.706
Spanish	1.029	1.084	0.762	0.615
French	4.099	5.216	0.628	0.535
Italian	1.543	1.743	0.699	0.579
Vietnamese	0.880	73.415	0.743	0.369
Thai	2.701	73.936	0.800	0.588
Turkish	1.520	0.699	0.779	0.596
Indonesian	1.237	1.059	0.729	0.660
Portuguese	1.877	1.331	0.805	0.711
Dutch	1.143	0.803	0.738	0.680
German	1.906	0.572	0.733	0.614
Russian	4.281	3.878	0.761	0.676
Ukrainian	1.082	0.997	0.730	0.647
Polish	1.415	0.766	0.802	0.729
Romanian	2.878	1.347	0.809	0.699
Greek	2.016	0.991	0.826	0.733
Czech	3.875	2.108	0.796	0.685
Finnish	4.666	2.964	0.835	0.759
Hindi	6.962	5.827	0.818	0.730

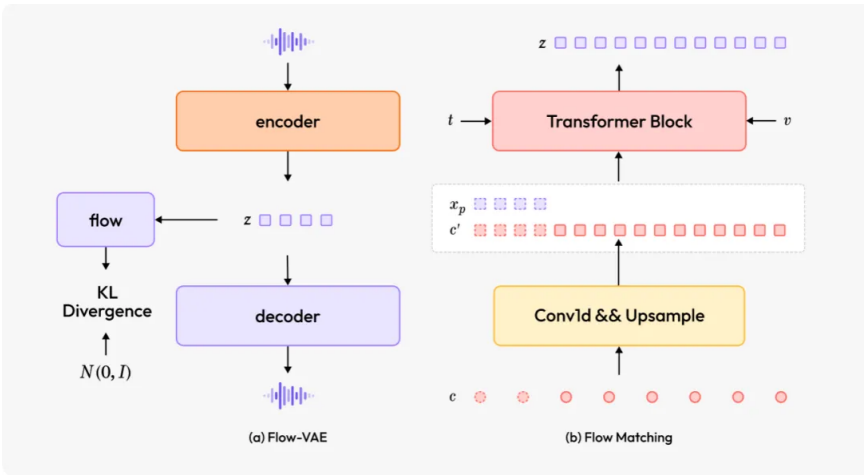
多语种测试

技术：

- 编码器仅从参考音频中提取音色特征，因此更能捕捉声音的本质——音色、音调和风格等特征
- 采用Flow-VAE将音频压缩为隐层特征，并通过Flow Matching模型来建模这一隐层特征。传统VAE通常假设 latent space 为标准正态分布，Flow-VAE则引入flow模型



人声编码器

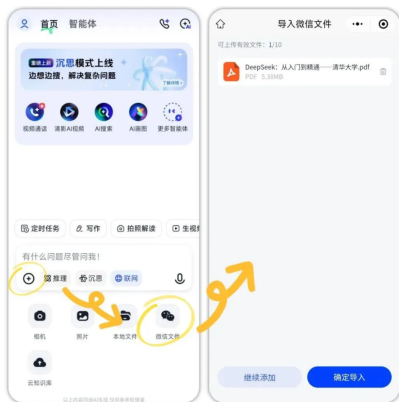


Flow-VAE

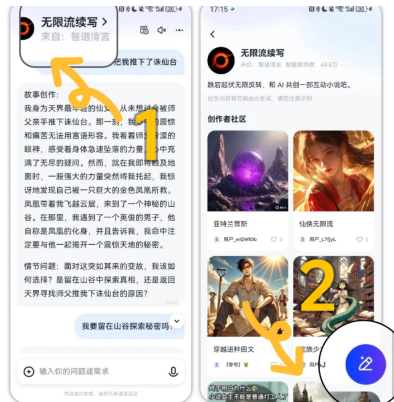
智谱清言更新社区和导入微信文件功能

- app端：

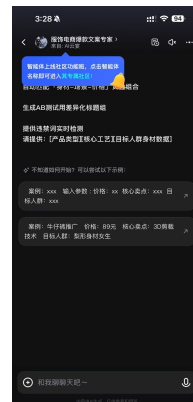
- 可以分享某个智能体的内容到该智能体的社区，支持点赞和评论。但是智能体很多，目前看大部分智能体下面没有内容或者只有一两个创作内容，对比Kimi打算要打造的综合社区来说可能太局限了
- 现在可以直接导入微信文件。



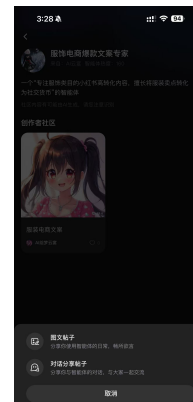
微信导入功能



智能体社区



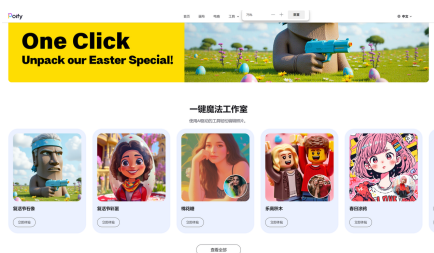
智能体社区入口



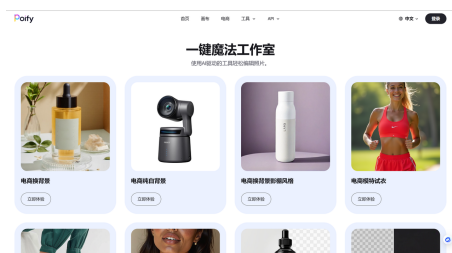
可以分享图文帖子或对话分享

## 快手推出AI作图工具Poify，专注于赋能电商

- 主要面向于B端用户，比如可以给电商提供AI试衣、一键生成海报，还融合了一些创作功能，可以分享至社交平台宣传。
- 虽然宣传过程中宣称自己是面向电商的，但实际网站里面宣传的还是自己的图片创作功能，大部分是趣味风格编辑生成图片功能，比如将照片转换为迪士尼风格、乐高风格等



网页端推荐主要还是内容创作



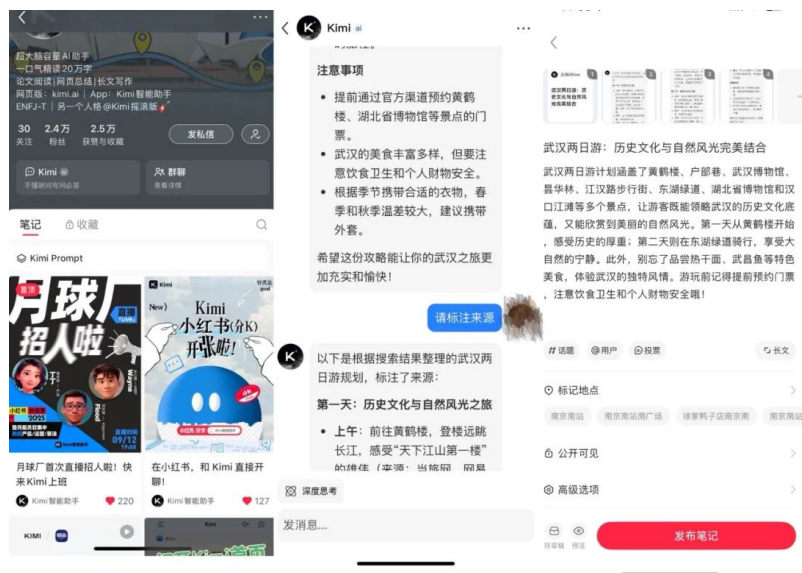
电商页面



主要工具

## Kimi与小红书合作，可一键生成笔记

- 从小红书端来说（只跟小红书合作了），目前只能从kimi官方账号中进入Kimi对话，而且对话功能也比较少，对话框上的功能只有深度思考。从Kimi端来说，可以将生成的内容直接分享至小红书，但分享的形式是对话的截图。综合来说小红书和Kimi的合作非常有限，并没有涉及到AI搜索，或者AI创作的合作
- kimi智能助手在小红书发布的这条帖子点赞收藏只有300多，传播度一般，不过是这个账号里流量最好的帖子



kimi分享



## 昆仑万维正式开源Matrix-Game

### 功能：

- Matrix-Game大模型是Matrix-Zero世界模型中的可交互视频生成大模型，专为开放式环境中的高质量生成和精确控制而设计，面向游戏世界建模的交互式世界基础模型
- 如视频所示，可以根据用户输入的任意控制指令，生成对应的游戏世界视频
- 核心：聚焦于视频生成与用户交互的深度融合，让用户可以通过简单直观的指令，自由探索、操控、甚至创造出细节丰富、物理规则合理的虚拟世界
- 潜在应用方向：虚拟游戏世界快速搭建、影视与元宇宙内容生产、具身智能体训练与数据生成





## 性能测试表现：

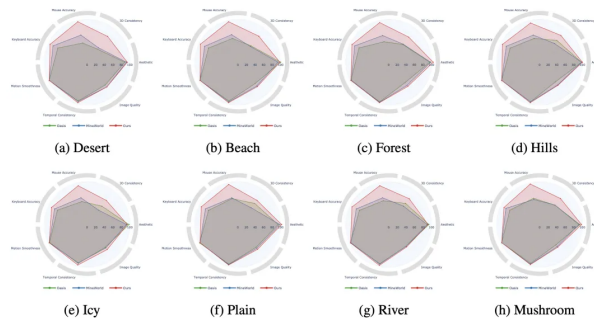
- 与微软的MineWorld对比，在Minecraft的各项指标上实现全面超越，交互可控性和物理一致性理解方面分数比较高。
- 此外在游戏场景中展现出优势
- 自己提出的统一基准的GameWorld Score测评框架，测试成绩领先

Model	Keyboard Action						Mouse Movement Action							
	forward	backward	left	right	jump	attack	camera↑	camera↓	camera←	camera→	camera↖	camera↗	camera↙	camera↘
Oasis [9]	0.85	0.78	0.80	0.79	0.77	0.89	0.66	0.55	0.33	0.35	0.56	0.53	0.45	0.51
MineWorld [17]	0.86	0.80	0.87	0.88	0.82	0.87	0.46	0.45	0.53	0.54	0.66	0.77	0.87	0.96
Ours	<b>0.99</b>	<b>0.91</b>	<b>0.92</b>	<b>0.96</b>	<b>0.88</b>	<b>0.95</b>	<b>0.91</b>	<b>0.98</b>	<b>0.89</b>	<b>0.90</b>	<b>0.92</b>	<b>0.97</b>	<b>0.98</b>	<b>0.98</b>

运动、攻击等动作准确率90%+

Model	Visual Quality		Temporal Quality		Action Controllability		Physical
	Image Quality ↑	Aesthetic ↑	Temporal Cons. ↑	Motion smooth. ↑	Keyboard Acc. ↑	Mouse Acc. ↑	3D Cons. ↑
Oasis [9]	0.65	0.48	0.94	<b>0.98</b>	0.77	0.56	0.56
MineWorld [17]	0.69	0.47	0.95	<b>0.98</b>	0.86	0.64	0.51
Ours	<b>0.72</b>	<b>0.49</b>	<b>0.97</b>	<b>0.98</b>	<b>0.95</b>	<b>0.95</b>	<b>0.76</b>

GameWorld Score评测系统成绩领先

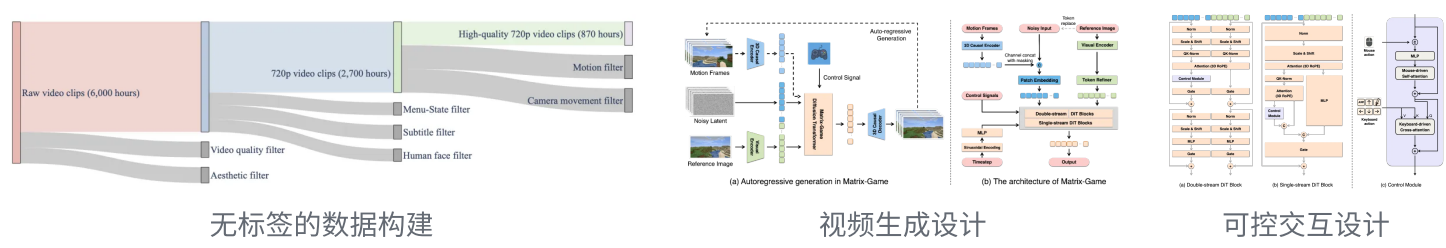


8大典型Minecraft场景

## 技术：

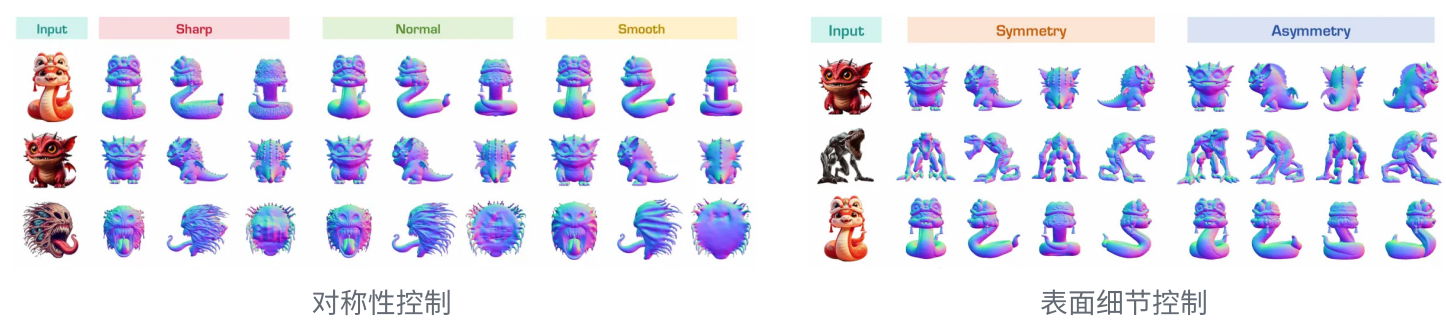
- 数据构建：**自主构建大规模 Matrix-Game-MC 数据集，涵盖：
  - 无标签预训练数据：**从 6000 小时 MineDojo 数据中三阶段过滤出约千小时高质量片段（画质美学、内容纯净、动态稳定）。
  - 有标签数据：**结合 VPT agent 在 MineRL 中自动探索（含精准控制信号）与基于 Unreal Engine 的程序化模拟（含位置、动作、反馈标签）构建。
- 模型设计：**

- **图像到世界建模**：输入单张图像与动作信号，通过多模态扩散模型（参考 Flux、混元视频）直接生成交互视频，不依赖语言提示。
- **自回归视频生成**：以前 k=5 帧为上下文逐段生成，结合扰动、删帧与 Classifier-free guidance 提升时序稳定性。
- **可控交互机制**：动作输入分离散（键盘）与连续（鼠标），融合 GameFactory 控制模块与 Diffusion Transformer，实现细粒度操控响应。



# 阶跃星辰开源3D大模型Step1X-3D

- **功能：**
  - Step1X-3D也强调了可控性，不过与本周Matrix-Game不同的是，Step1X-3D更像做的是创作出来的3D模型的可控性，比如从某个2D平面修改建模，而昆仑万维的是“操作”层面的



- **性能测试表现：**
  - 只进行了基于自建的综合测试，与主流模型对比评分较高

Table 1: **Quantitative comparison on different models.** \* indicates closed-source models, and indicate the best, second best, and third best performance respectively.

Model	Texture CLIP-Score↑	Geometry		
		Uni3D-I↑	OpenShape <sub>sc</sub> -I↑	OpenShape <sub>pb</sub> -I↑
Rodin-v1.5*	0.845	0.270	0.104	0.081
Meshy-4*	0.796	0.357	0.137	0.099
Tripo-v2.5*	0.848	<b>0.366</b>	<b>0.140</b>	0.134
Trellis	0.848	0.353	0.136	<b>0.139</b>
Hunyuan3D 2.0	0.829	0.352	0.131	0.131
<b>Step1X-3D (Ours)</b>	<b>0.853</b>	0.361	0.139	0.130

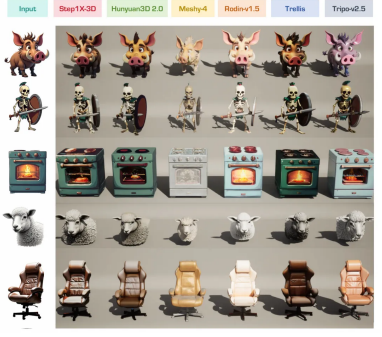


Figure 12: Qualitative comparison with SOTA methods on generated texture.

- **技术：**
  - **Step1X-3D** 基于增强型网格-SDF 转换，提升水密几何转换成功率 20%，提高生成精度与效率。

- 采用 **3D 原生两阶段架构**，解耦几何与纹理表征。
- 几何生成** 采用深度优化的混合型 VAE-DiT 架构，适配 3D 特性。
- 纹理生成** 基于深度定制的 SD-XL 模型。
- 整体架构延续 2D 生成模型（如 Stable Diffusion）设计逻辑，便于集成 LoRA 等成熟 2D 控制技术

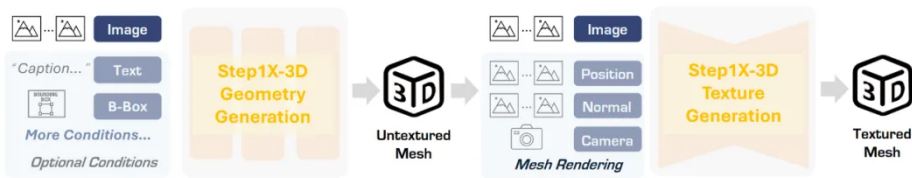


Figure 2: Overall pipeline of Step1X-3D.

原生3D架构

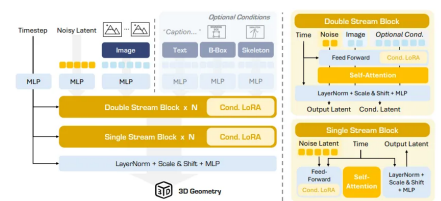
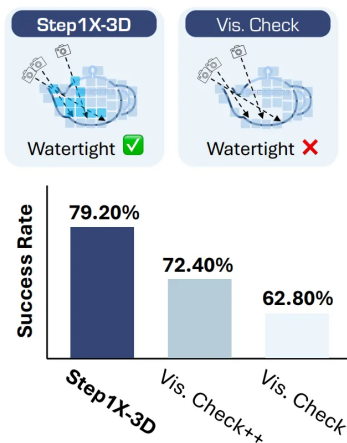


Figure 4: Step1X-3D geometry diffusion framework. The framework enables seamless integration of diverse conditions, allowing fine-tuning for various downstream tasks.

几何生成模型基于 FLUX MMDiT 结构和 Rectified flow 算法直接对 3D 表示生成进行建模



(b) Mesh to SDF Process

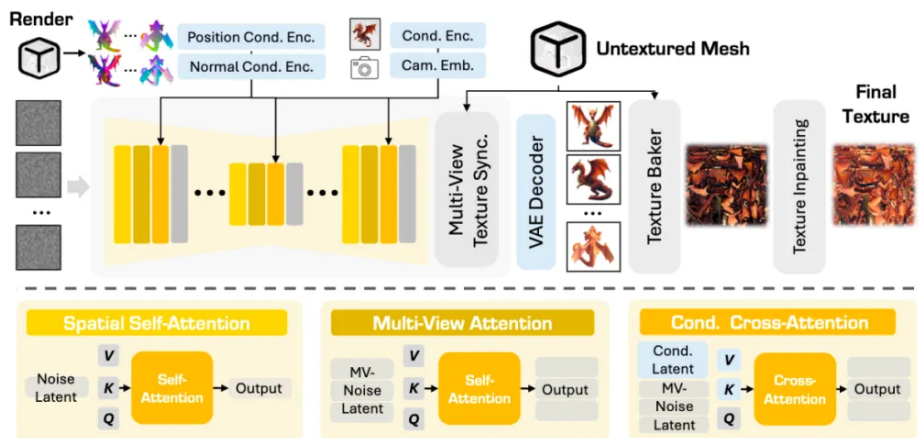


Figure 5: Texturing module of Step1X-3D.

Step1X-3D 数据筛选与 SDF 转换流程示意图

纹理生成模型基于预训练的 Diffusion model，通过单视图和 3D 几何信息指导，生成多视角一致的纹理信息

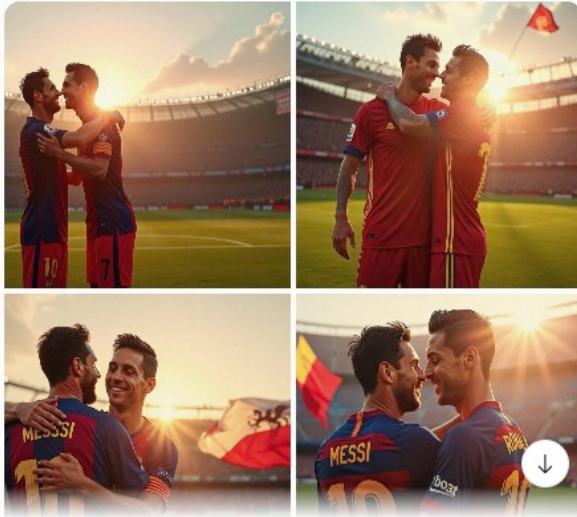
## 元宝

- 腾讯元宝web端更新，一句话直接出图
  - 处理流程：
    - 利用Deepseek优化生图提示词，而后利用混元大模型生图
  - 实际效果：
    - 实际效果非常一般，hunyuan深度思考模式下文生图实体无法区分

以下是根据需求梳理出的图片提示词，将由元宝生图工具生成图片。

【现实主义风格，16:9比例 足球巨星梅西与C罗在绿茵场上深情相拥，夕阳余晖勾勒出两人剪影，运动服细节清晰可见，球场边球迷欢呼声起，逆光中飞扬的发丝与飘动的队旗构成动态背景】

以下是为您生成的图片



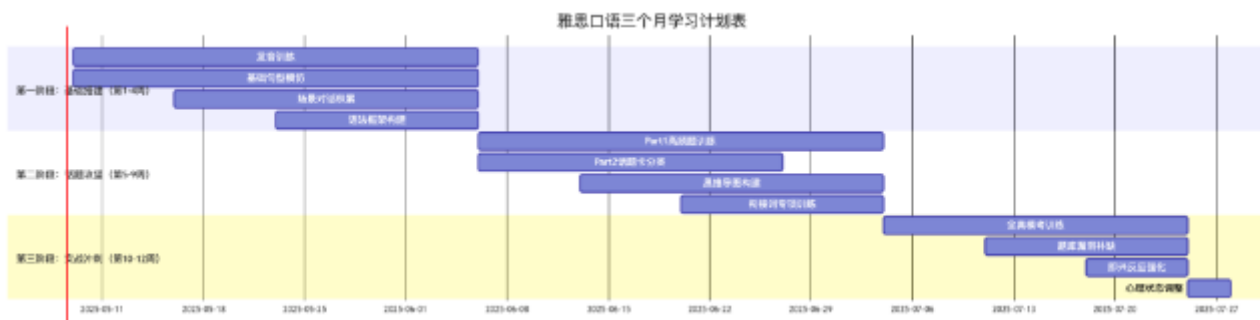
实际生成效果

- 腾讯元宝支持生成Mermaid图表
  - 输入一句话指令：帮我用mermaid生成一张XXX图，即可快速得到AI自动生成的饼图、序列图、流程图、甘特图等各类图表，辅助高效学习和工作。



元宝生成mermaid图表





mermaid图表

- 元宝上线一键解读新闻功能：

- 用户在腾讯新闻App浏览资讯、深度视频时，均可一键分享元宝，实现秒级智能解析、生成核心摘要，快速获得干货。

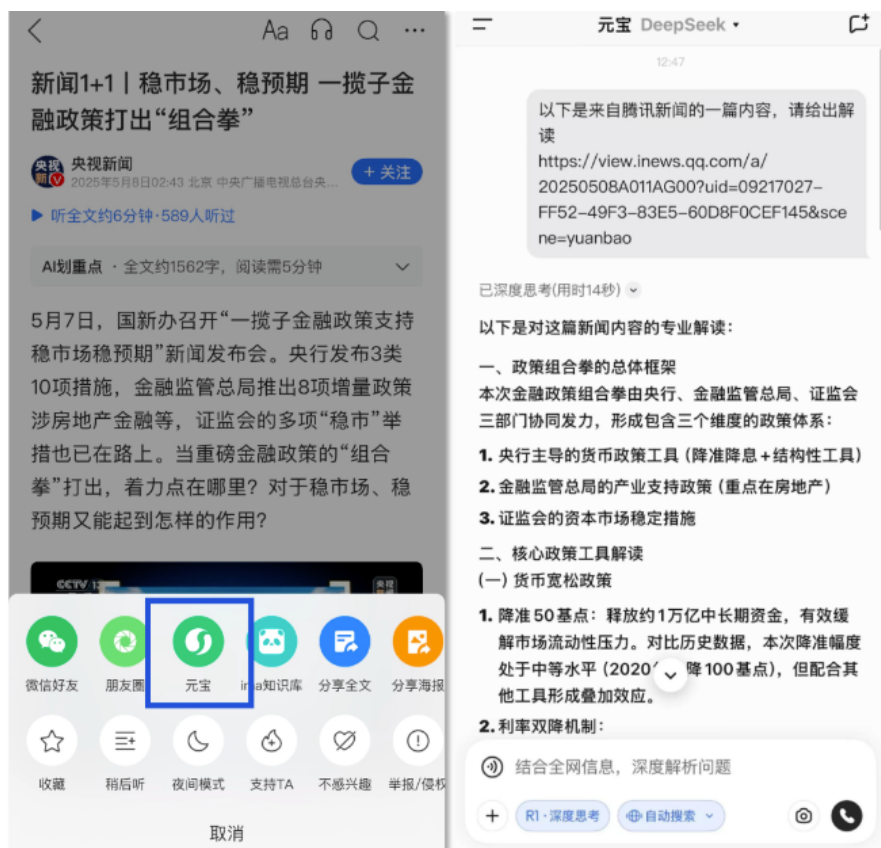
- 本质原理：

- 转发后会复制文章链接，并在链接前加上“一下是来自腾讯新闻的一篇内容，请给出解读”，粘贴为prompt输入LLM
    - 可复制性极强，可与手百合作通过简易调整完成此功能

- 用户可以针对文章与总结继续进行追问

- 新闻总结并不简短，类似于深度思考正常回答的长度，据此推断并无其他优化步骤，对用户的即刻阅览新闻的需求满足不够充分

- 推荐在此基础上添加对新闻解读篇幅的限制，以满足用户对迅速阅览新闻的需求



一键总结新闻

# 豆包

- 字节多模态模型Seed1.5-VL登陆火山引擎

- 官方称仅20B激活参数获38项SOTA
- 实体识别与区分能力较强，能够给出图中各个实体的精确坐标范围
- 融合了图像理解、文本OCR、语义匹配、数学推理、图像差异推断等能力
- 实测评价：
  - 具有较强的图像理解与计算能力，能够依据图片中的文字信息与图像信息精准识别出货架角落里的黄桃味酸奶优酸乳，还准确数清瓶数、识别单价，并进行正确的价格计算
  - 具有较强的图像差异对比与图像推理能力，能够通过精准识别前两行图形叠加后去同存异的变换逻辑，准确推导出第三行的缺失图形



seed1.5-VL做混合推理



seed1.5-VL做行测图像推理

- 豆包超能创意1.0批量生图功能推出

- 豆包可以在web端选择“图像生成”使用合适的提示词，批量生成图片，支持一句话批量生图，最多批量生成20张
- 下面三张图片分别为官方提示词、修改后提示词、一句话生图，充分说明经过短暂修改或者简短的prompt可以完成诸多有趣变种，营销可泛用和裂变能力强，可以考虑借鉴此功能做一个小娱乐产品

创建 20 张各种水果的高分辨率 3D 渲染图，并将其设计为一个可爱的毛绒抱枕。其形状应看起来柔软、圆润和蓬松。使用健壮的笔触，并搭配微妙的阴影和缝线。其形状应略微不规则且柔软，并配以柔和的阴影和灯光，以突出其质感和真实感。将其放置在干净简约的背景（浅灰色或淡蓝色）上，并保持愉快的雕塑美感。

接下来我将生成 20 张各种水果造型的可爱毛绒抱枕的 3D 渲染图。抱枕形状柔软圆润、略微不规则，有细微织物褶皱和缝线。搭配柔和阴影和灯光。放置在浅灰色或淡蓝色的干净简约背景上。



官方提示词



略微修改后提示词结果

8张，毛绒绒的拟人化熊猫，超可爱，穿着不同朝代的服饰，穿梭在不同朝代与重要文物背景，比例「1:1」

接下来我将生成 8 张图片，主体是毛绒绒的拟人化熊猫，它超可爱，会穿着不同朝代的服饰，穿梭在不同朝代与重要文物背景中。



一句话批量生图结果

## 模型实测：

- 实际生成时间在1分钟左右
- 能够较为准确地生成prompt所指示的图片



8张，欧洲足球历史各大球星的q版公仔图。

接下来我将为您生成8张欧洲足球历史各大球星的q版公仔图。

已生成

```
prompt
[{"type": "generate", "text": "生成8张欧洲足球历史各大球星的q版公仔图，风格可爱，线条简洁，色彩鲜艳，背景为纯色。"}]
```



生成效果

接下来我将把这8张球星图片改成更写实的风格。

已生成



修改prompt后的生成效果

- 评价：
  - 能够更加精准把握创意意图，生成符合期待的高质量图像
  - 能够人机共创，可以多次提示词修改，发散出想象力
  - 能够一次性批量生成20张差异化图像并行处理，即时呈现，可玩性较高，时间成本较低

## 阿里通义

- 全新的通义APP上线了千问3模型（Qwen3），既能深思，也能秒答
  - 通义APP支持调用Qwen3混合推理模型
    - 混合推理模型：
      - 能够随机应变选择合适的模型
        - 对于简单问题则采用秒答模式，对于困难问题则启用深思模式
- 阿里通义有原域名tongyi.ai转为新域名tongyi.com
  - 评价：
    - 品牌一致性：tongyi.com与“通义”完全对应，直观易记，无需解释后缀含义。
    - 国际化潜力：.com是全球化企业的标准配置，为未来市场拓展奠定基础。
    - 用户信任度：.com域名历经数十年沉淀，天然赋予品牌专业性与权威感。

## Gemini

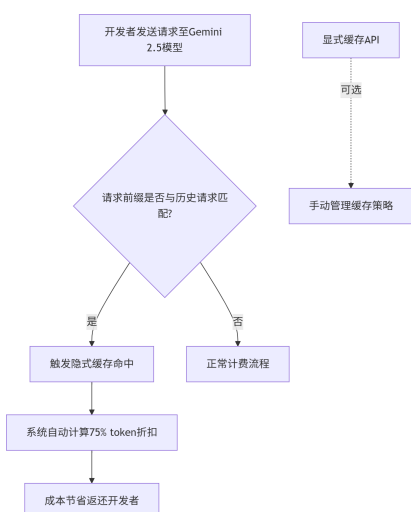
### Gemini 上线隐式缓存机制：

### 5月Gemini API上线了上下文缓存功能

- 隐式缓存：
  - 系统自动识别请求中的重复内容（无需开发者手动配置），对相同前缀部分给予75%的token费用折扣。开
- 触发条件（最小Token要求，低于此阈值不会触发缓存）
  - Gemini 2.5 Flash：1024 Token（约750字）



- Gemini 2.5 Pro：2048 Token（约1500字）
- **发者可以通过保持新对话和历史对话前缀保持一致，将更新部分放置前缀后以合理触发缓存机制，节省开支**
- 举例：
  - 若前缀部分消耗1000 Token，实际仅按250 Token计费。若每次请求包含相同的500字系统提示（如“你是一个客服助手…”），后续请求中这部分Token成本可降低75%。假设日均1万次请求，每月可节省数千万Token费用
- 适用群体
  - 通过合理设计请求结构，开发者可在高频重复任务中实现**月均数百万Token的成本削减**，尤其适合预算敏感的中小型团队




## Implicit caching with Gemini API 使用 Gemini API 进行隐式缓存

Implicit caching directly passes cache cost savings to developers without the need to create an explicit cache. Now, when you send a request to one of the Gemini 2.5 models, if the request shares a common prefix as one of previous requests, then it's eligible for a cache hit. We will dynamically pass cost savings back to you, providing the same 75% token discount.

隐式缓存直接将节省的缓存成本传递给开发者，无需创建显式缓存。现在，当您向 Gemini 2.5 模型发送请求时，如果该请求与之前某个请求共享相同的前缀，则符合缓存命中条件。我们将动态地将节省的成本返还给您，提供相同的 75% 令牌折扣。

## Gemini 2.5 pro AI理解视频更进一步

- Gemini 2.5 Pro 在十几个学术视频基准测试中取得了新的SOTA（业界最佳）成绩
- 
- | Benchmark  | Previous SOTA | Gemini 2.5 Pro | Gemini 2.5 Pro (Video) |
|--|---------------|----------------|------------------------|
| OpenAI Video Chat                                      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video)                              | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio)                              | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text)                               | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image)                              | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Caption)                            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Transcript)                         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Summary)                            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Question Answering)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Classification)                     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Generation)                         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Reasoning)                          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Translation)                        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Summarization)                      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Sentiment Analysis)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Topic Classification)               | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Entity Recognition)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Relationship Extraction)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Event Detection)                    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Action Recognition)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Object Detection)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Captioning)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Captioning)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Captioning)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Captioning)                    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Generation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Generation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Generation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Generation)                    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification)               | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification)               | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification)               | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification)                | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation)                  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation)                   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation)                    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization)                | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization)                | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization)                | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization)                 | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification Accuracy)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification Accuracy)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification Accuracy)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification Accuracy)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation Accuracy)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation Accuracy)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation Accuracy)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation Accuracy)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation Accuracy)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation Accuracy)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation Accuracy)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation Accuracy)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering Accuracy)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering Accuracy)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering Accuracy)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering Accuracy)   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization Accuracy)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization Accuracy)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization Accuracy)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization Accuracy)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation F1 Score)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation F1 Score)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering F1 Score)   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation Precision)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation Precision)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering Precision)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation Recall)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation Recall)             | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering Recall)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification F1 Score)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation F1 Score)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation F1 Score)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation F1 Score)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering F1 Score)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering F1 Score)   | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization F1 Score)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization F1 Score)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification Precision)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation Precision)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation Precision)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation Precision)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering Precision) | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering Precision)  | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization Precision)      | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization Precision)       | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Classification Recall)        | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Classification Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Segmentation Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Segmentation Recall)           | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Annotation Recall)            | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Annotation Recall)             | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Question Answering Recall)    | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Question Answering Recall)     | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Image Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Video Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Audio Summarization Recall)         | 48.3 (GPT-4o) | 48.3           | 48.3                   |
| OpenAI Video Chat (Text Summarization Recall)          | 48.3 (GPT-4o) | 48.3           | 48.3                   |
- Gemini 2.5pro首次实现了**原生多模态模型能够将音视频信息与代码等其他数据格式无缝结合**。不是简单地“看懂”视频，而是能基于视频内容进行更深层次的理解和创造。

## 直接把视频变身网页交互应用

- 实测牛顿定律视频输出互动网页案例中：**思考时间：1min+，输出时间：2min+ 总时长大约 3min**



## 代码块

```
1  <!DOCTYPE html>
2  <html lang="zh-TW">
3  <head>
4      <meta charset="UTF-8">
5      <meta name="viewport" content="width=device-width, initial-scale=1.0">
6      <title>【觀念】牛頓第二運動定律</title>
7      <style>
8          body {
9              font-family: 'Arial', 'Microsoft JhengHei', sans-serif;
10             line-height: 1.6;
11             margin: 0;
12             padding: 0;
13             background-color: #f0f4f8; /* Light blue-gray, similar to video
14             bg */
15             color: #333;
16             display: flex;
17             flex-direction: column;
18             align-items: center;
19         }
20         .container {
21             width: 90%;
22             max-width: 800px;
23             background-color: #fff;
24             margin: 20px auto;
25             padding: 20px;
26             border-radius: 10px;
27             box-shadow: 0 0 15px rgba(0,0,0,0.1);
28         }
29
30         h1, h2, h3 {
31             color: #2c3e50; /* Dark blue */
32             text-align: center;
33         }
34         h1 {
35             font-size: 2.2em;
36             border-bottom: 2px solid #3498db; /* Blue accent */
37             padding-bottom: 10px;
38             margin-bottom: 20px;
39         }
40         h2 {
41             font-size: 1.8em;
```

```

42         color: #2980b9; /* Slightly lighter blue */
43         margin-top: 30px;
44     }
45     h3 {
46         font-size: 1.4em;
47         color: #16a085; /* Teal */
48     }
49
50     .character-dialogue {
51         display: flex;
52         align-items: flex-start;
53         margin-bottom: 15px;
54         padding: 10px;
55         border-radius: 8px;
56         background-color: #ecf0f1; /* Light gray for dialogue box */
57     }
58     .character-dialogue .avatar {
59         width: 60px;
60         height: 60px;
61         border-radius: 50%;
62         margin-right: 15px;
63         font-size: 2em;
64         display: flex;
65         justify-content: center;
66         align-items: center;
67         flex-shrink: 0;
68     }
69     .character-dialogue .avatar.cat { background-color: #bdc3c7; } /*
Gray cat */
70     .character-dialogue .avatar.frog { background-color: #2ecc71;
color: white; } /* Green frog */
71
72     .speech-bubble {
73         background-color: #ffffff;
74         padding: 10px 15px;
75         border-radius: 10px;
76         box-shadow: 0 2px 5px rgba(0,0,0,0.05);
77         position: relative;
78         flex-grow: 1;
79     }
80     .speech-bubble strong { color: #e67e22; } /* Orange for emphasis */
81
82     .interactive-quiz, .explanation, .example {
83         margin-top: 20px;
84         padding: 15px;
85         border: 1px solid #ddd;
86         border-radius: 8px;

```

# 【觀念】牛頓第二運動定律

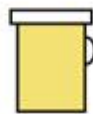
臺北酷課雲 x 均一教育平台

## 推啤酒杯遊戲



狐狸貓和小數蛙拿著玩具機械手臂，玩推啤酒杯遊戲。

玩法：操控機械手臂，先調整好左右後，按住「前進」鍵使機械手臂推著啤酒杯向前，開始前進後就不可以調整方向，而讓杯子向前移動到最靠近虛線圈內的人獲勝。



開始推杯子



小數蛙突然覺得疑惑：慣性定律不是說「靜者恆靜」嗎？為什麼靜止的啤酒杯突然移動了？

## VIDEO TOYS

Generate interactive learning apps from YouTube content  
An experiment by Aaron Wade

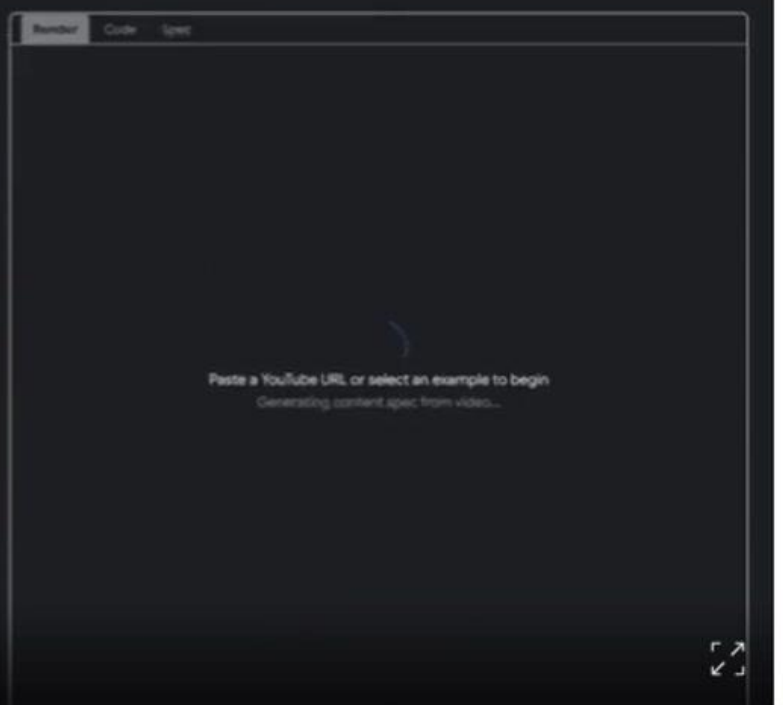
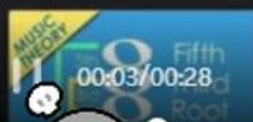
Paste a URL from YouTube:

<https://www.youtube.com/watch?v=5tqCof7Iw>

Generate app



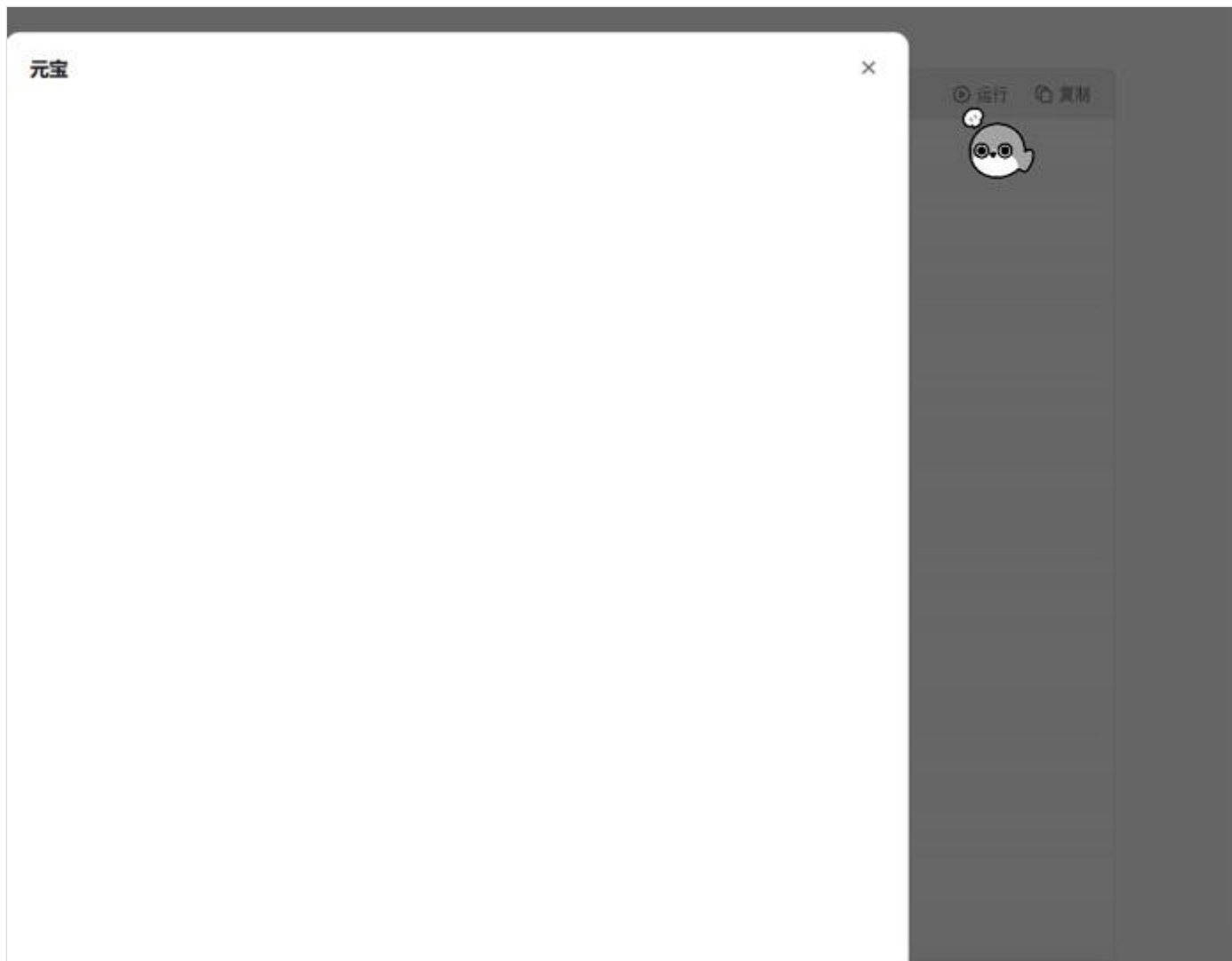
Examples

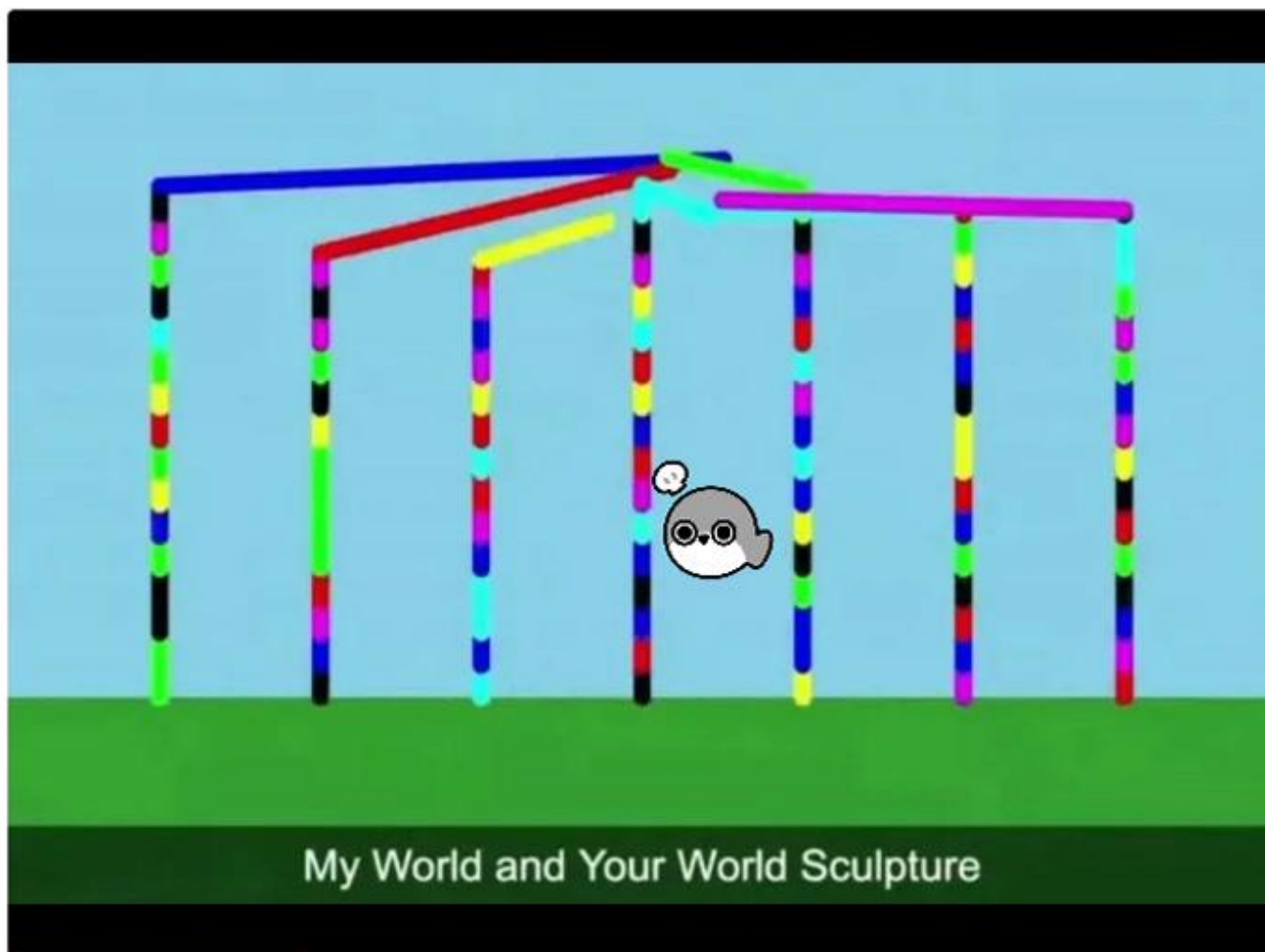




## 视频理解模仿生成视频呈现的动画

- 给一个YouTube视频链接，让模型模仿视频开头动画





## 精准检索与描述视频片段——音视觉识别

Model input: Frames & Audio at 1FPS (for clarity, the visualisation below is only a subset)



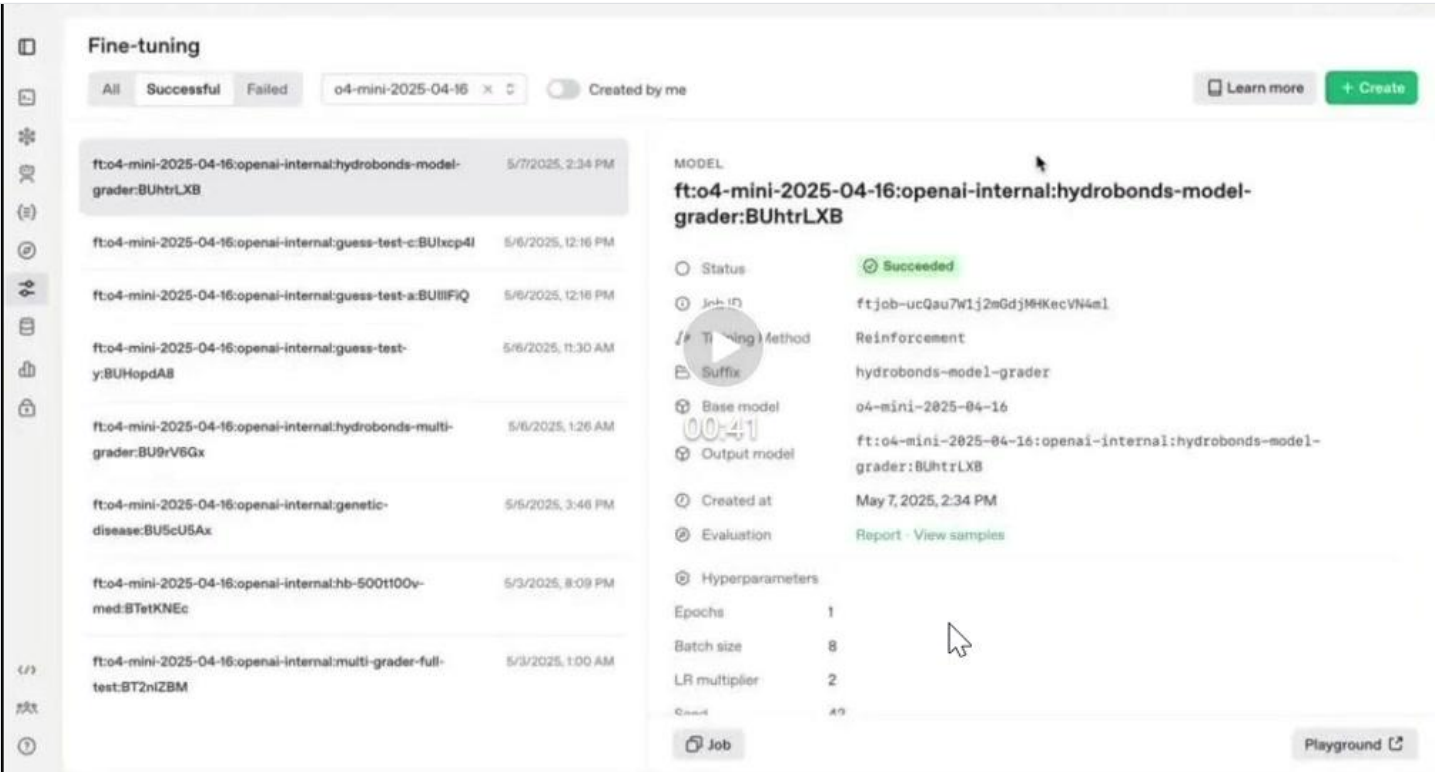
00:00/00:10



# OpenAI

## OpenAI o4-mini携强化微调 RFT 正式上线:

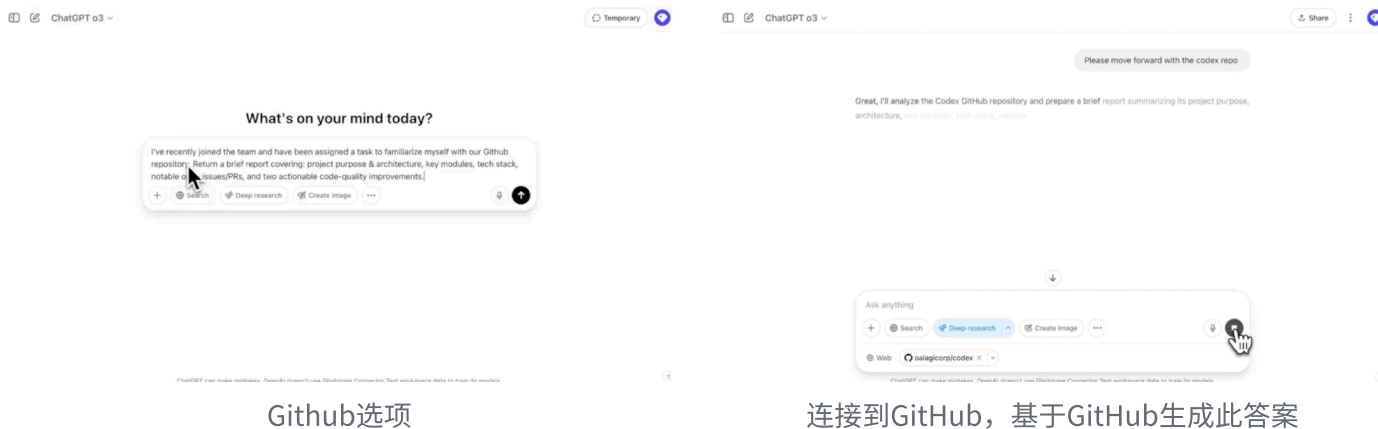
- 采用强化学习算法，快速将模型提升至专家博士水平支持定向打造领域专家 AI，如法律、医疗等专业场景。演示效果如图：



- 真实案例
- **ChipStack：为半导体设计“智能布线”**
  - 公司：ChipStack，搞AI驱动芯片设计和验证工具。
  - 痛点：将设计接口绑定到验证IP（预制验证组件）是件耗时耗人力的苦差事，涉及大量信号映射，需要深厚的领域知识。
  - 目标：训练OpenAI模型自动完成这项任务。ChipStack准备了不到50个样本的数据集，进行了多次RFT实验。
  - 评分器思路：用Python定义了一个比较预测输出（一系列名值对）与期望答案的评分器，计算精确率和召回率的F1分数。
  - 成果：o1-mini和o3-mini模型性能均提升了约12个百分点。微调后的模型在识别何时“不”应用布线方面进步巨大，这对于包含大量可选信号的商业验证IP至关重要

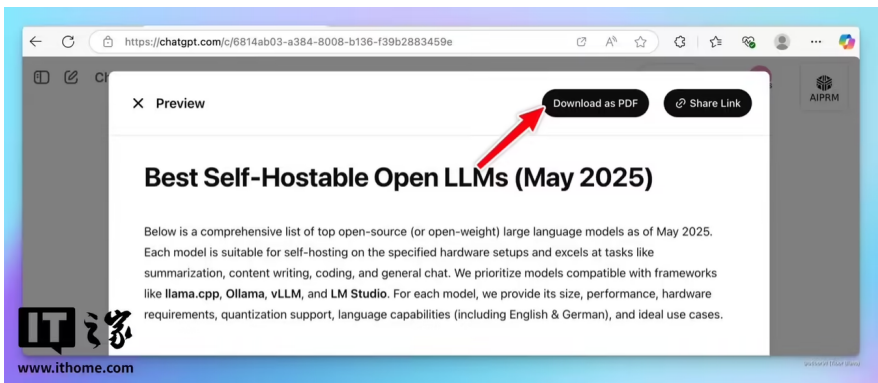
# ChatGPT Deep Research 集成 GitHub 仓库

- 支持针对源码与 PR 进行深度问题分析与引用生成可通过个人设置关联 GitHub，便捷调用仓库分析能力
- 目前仍处于测试阶段，Team用户可用，并将陆续对Plus、Pro用户开放



## ChatGPT 深度研究新增导出为 PDF 功能，可保留报告格式

- 现在新增了“下载为PDF”的选项，能完美保留报告的原始布局，目前处于网页端的测试阶段
- 但是由于AI生成的研究报告肯定还是需要进一步修改的，PDF的格式也不一定满足要求，该功能的使用场景比较小



## Genspark

### Genspark 推出 “AI Sheets” 革新电子表格体验

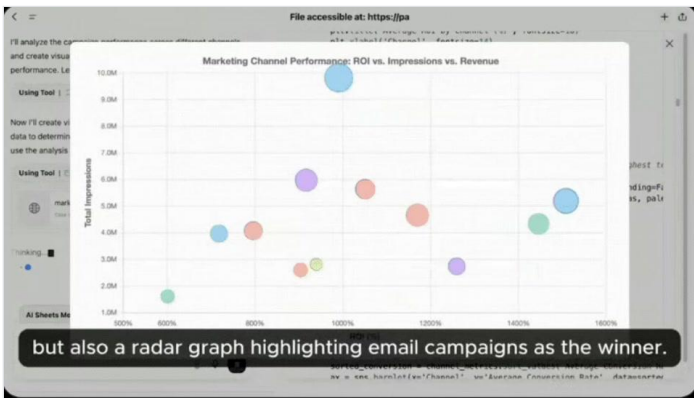
- 功能：
  - Genspark AI Sheets推出全新AI表格工具，用户通过自然语言对话即可完成数据收集、整理、分析和可视化，无需掌握复杂Excel公式；



- 工具支持多格式文档导入、自动数据清洗、智能分析与可视化，官方称处理速度比传统手动操作快数十倍；
- 该工具目前处于Beta测试阶段免费开放，适用于销售、市场、产品等多个领域，能有效解决传统表格处理中的效率和专业门槛问题。

● 亮点：

- 能够自动搜索和抓取公开信息，并整理成结构化的表格，无需用户手动输入
- 会自动识别和处理数据中的重复项、缺失值，并进行格式统一
- 一键智能分析与可视化
- 多格式文档导入，兼容性强

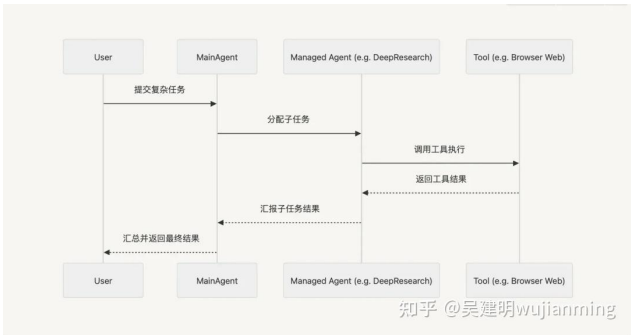


(edit provided text)		
me running around		
We all had a wonderful		
ed to do it agian soon.		

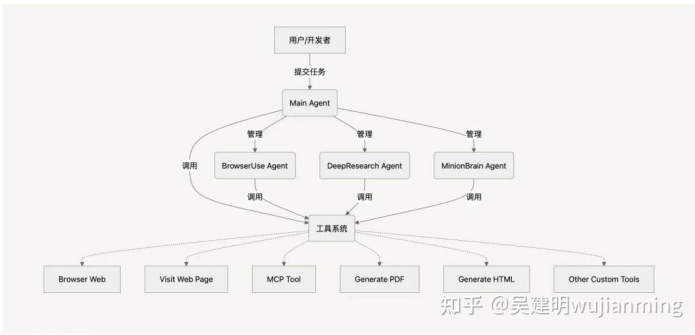
# Minion-agent

## 全新多合一AI智能体框架

- 功能：
  - 可实现多框架无缝支持，灵活调用不同框架的能力，实现多个智能体协作处理任务
  - 提供各种工具服务，同样也支持MCP等功能
  - Minion-agent选择完全开源
  - 随着MCP的发展，AI越来越趋向于开源、合作等趋势。Minion-agent真正做的就是利用不同框架的优势，从而提高工作的效率，符合未来AI发展的趋势



多框架融合的智能体生态系统



流程图

- 核心价值与技术优势
  - 多框架融合能力
    - 支持接入如 OpenAI、LangChain、SmolaAgents、Google AI 等主流框架
    - 通过统一接口整合不同框架能力，降低学习成本和系统集成复杂度
  - 精巧的多智能体架构设计
    - 任务生命周期清晰：从指令接收 → 拆解 → 分配 → 工具调用 → 结果整合 → 回传
    - 支持主-子智能体协作，实现并行处理与模块解耦
  - 工具即服务机制
    - 内建网页浏览器、文件操作、自动化任务等插件式工具
    - 工具通过统一接口接入，方便开发者灵活调用与扩展
  - 模块间标准通信协议
    - 主体/工具/框架之间通过标准化结构传输信息
    - 实现松耦合、可扩展、高容错的多模块协同系统

## Design Agent Lovart

prompt: 优化这个海报，呈现科技化风格

生成速度较快，整体风格符合输入，但是文字识别有缺陷。

文小言

Baidu 百度

文小言App（原文心一言），百度旗下新搜索智能助手。AI多模型融合调度的全新引擎驱动，整合百度生态资源，聚焦模型开放与产品功能创新，致力打造最好的AI产品体验

**岗位要求：**

- 1、根据业务需求，完成文小言数据产品的规划、设计和落地
- 2、结合业务特点，构建指标体系与数据分析框架，输出具体策略建议，并推动业务团队落地执行
- 3、对业务异常指标进行归因，对业务问题进行诊断、分析，积累核心问题的分析方法论，推动业务和运营策略的有效运行
- 4、深度参与业务核心专项，主动发现和探索业务的机会点，有效应用分析方法助力业务达成目标

# AI产品创新部

热招岗位：数据分析师（J83767）  
投递：HRBP anning03@baidu.com  
地点：北京

文小言

Baidu 百度

文小言App（原文心一言）百度旗下重要出海产品重磅助力。AI赋能智能电全面赋能创新计划项目，传承百度真变变，聚焦数据化运营和客互功能提升打造最好AI产品体验

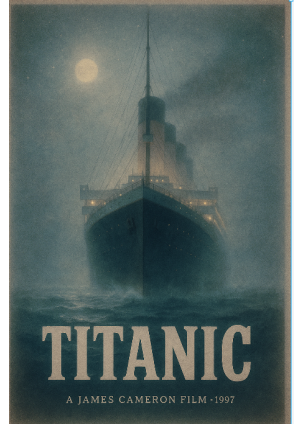
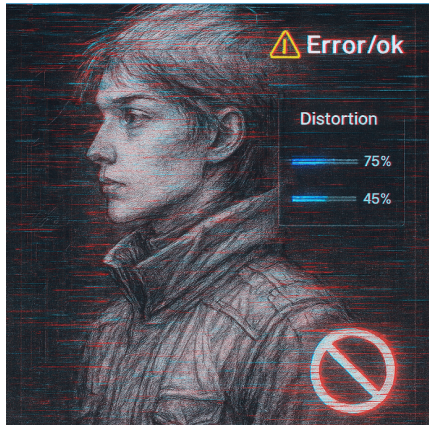
**岗位要求：**

1. 根据业务需求，完成文小言数据产品的规划、设计和落地
2. 结合业务特点，构建指标体系与数据分析框架，输出具体策略建议，并推动业务团队落地执行
3. 对业务异常指标进行归因，对业务问题进行诊断、分析，积累核心问题的分析方法论，推动业务和运营策略的有效运行
4. 深度参与业务核心专项，主动发现和探索业务的机会点，有效应用分析方法助力业务达成目标

# AI产品创新部

热招岗位：数据分析师（J83767）  
投递：HRBP anning03@baidu.com  
地点：北京

- prompt1: "特写镜头：雨中霓虹街道，主角手握染血信件，面部阴影交错。低角度仰拍，雨水顺下巴滴落，眼神绝望中透出决绝。冷蓝色调，背景模糊的车灯拉出光轨，快门速度1/60"
- prompt2: 画面主体是一幅素描风格的人物头像，线条细腻，展现出人物的发型、面部轮廓和穿着。故障元素人物脸部叠加了类似信号干扰的红蓝线条，营造出故障、失真的视觉效果，暗示视觉信息出现问题。
- prompt3: 生成模特走秀图
- prompt4: 生成一个梦境里的泰坦尼克号电影的电影海报，要朦胧美，要复古，标题设计有冲击力







- 集成 GPT image-1、Flux Pro、Suno AI 等全栈 AI 模型
- 全流程自动化执行可在画布选区直接发指令
- 自动换装并批量生成平台广告图