亚马逊云科技
中国峰会

SESSION 262

PIXVERSE × AMAZON WEB SERVICES

从像素到叙事：多模态模型重构视频生成的能力边界

荣蓉

Rong Rong

企业服务产品负责人

爱诗科技

汪其香

Wang Qixiang

解决方案架构师

亚马逊云科技

视频生成，正进入
从"能生成"走向"可叙事"的爆发时刻

01

视频成为通用语言

从文字到图像，从图像到视频——人类表达的密度正在被一帧一帧地拉高。

02

范式从扩散走向统一

文本理解、图像感知、时序建模、物理规律——多模态架构正在把它们融为一体。

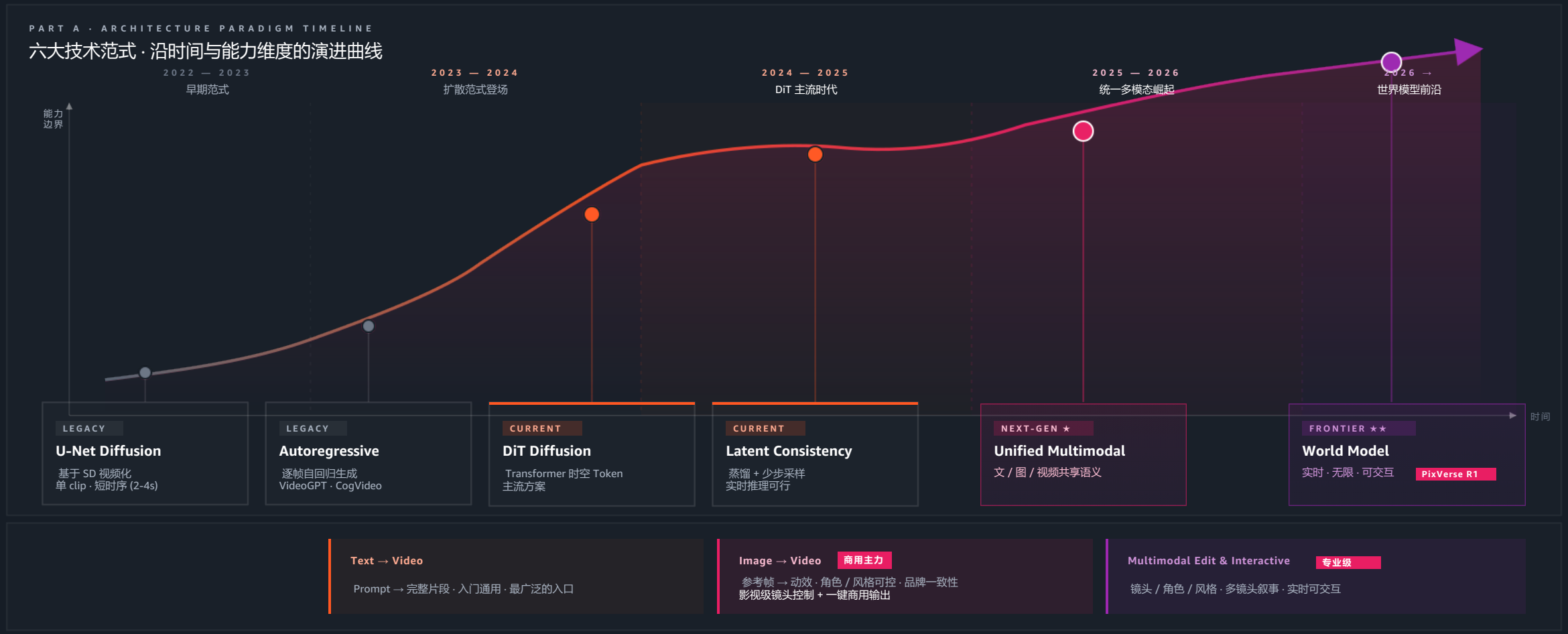
03

能力边界被系统重写

从一段几秒钟的片段，到一段可被导演、被剪辑、被交付的"完整影像"。

视频生成模型发展路线

从 2022 单帧扩散 → 2026 统一多模态与世界模型 · 能力边界沿曲线持续右移



但要走到"商业落地"——
有三个变量必须同时成立

质量 × 可控性 × 规模化

= 视频 AI 真正进入商业市场的"乘法公式"

质量 · Quality

画面级别从"可观看"升级到
"电影级、可商用"。

可控性 · Control

镜头、角色、风格、连续性——
创作者要的不是惊喜，而是确定。

规模化 · Scale

推理算力、全球分发、合规交付——
这是云厂商的主场。

视频生成流程环节

01

INPUT

输入理解

Multimodal Input Parsing

理解用户上传的
文本 / 图像 / 视频 / 音频
作为生成的语义起点

BEDROCK 模型

- Amazon Nova
- Claude
- 开源模型

能力 · 图像/视频内容描述、参考帧提取、风格识别

02

PROMPT

提示词重写

Prompt Optimization

用户口语 → 模型可读的
结构化分镜脚本
显著提升首次成片质量

BEDROCK 服务

- Prompt Optimization
- Nova Prompt Optimizer
- Claude 重写

能力 · 启发式 + LLM 双路径，自动适配下游模型语法

03

EMBEDDING ★

多模态检索

Multimodal Embedding & RAG

检索素材库 / IP 库 / 历史片段，
给生成模型注入
"品牌一致性"与"角色一致性"

BEDROCK 模型

- **Nova Multimodal Embeddings**

首个统一支持文 / 文档 / 图 / 视频 / 音频的 Embedding 模型

04

GENERATION ★

视频生成

Video Synthesis Core

管线的核心模型层——
文生 / 图生 / 多镜头叙事 / 实时世界模型

BEDROCK 模型矩阵

- Luma Ray2
- Nova Reel 1.1

同一接口，多模型可选

客户按场景 / 成本 / 风格自选

Front Tier Model

- **PixVerse V6 / R1 / C1**

05

MODERATION

内容审核

Safety & Compliance

文本提示 + 输出图像/视频
双向审核，过滤
暴力 / 不当内容 / 版权风险

安全栈

- Bedrock Guardrails
- Image Content Filters
- Rekognition · Video Mod

能力 · 企业级安全护栏

06

DELIVERY

分发与后处理

Post-process & CDN

转码 / 加水印 / 拼接 /
多分辨率出片，
就近全球分发

媒体栈

- S3 · CloudFront
- Nova Sonic · 语音

能力 · 从生成到观众端到端

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

Amazon Bedrock：一个让模型厂商"被使用"的开放生态

全球已有 100,000+ 组织在 Bedrock 上构建生成式 AI 应用 [Source] Amazon.amazon.com/bedrock

SCALE OF THE PLATFORM

100,000+

全球客户构建生产级 AI 应用

Hundreds

基础模型可选择
覆盖文本 / 图像 / 视频

Up to 75%

推理成本可优化空间
（模型蒸馏 + 智能路由）

① 模型自由选择 · Model Choice
数百款 FM 一键接入：Anthropic / Meta / Mistral / OpenAI / Amazon Nova / 合作伙伴模型。

② 自有数据安全定制 · Custom & Private
Knowledge Bases / Data Automation / 微调——你的数据永不用于训练他人模型。

③ 企业级安全合规 · Guardrails
ISO / SOC / GDPR / FedRAMP / HIPAA 全覆盖；Guardrails 拦截高达 88% 有害内容。

④ Agent 化能力 · Bedrock AgentCore
把视频生成模型嵌入 Agent 链路：跨工具、跨数据、跨工作流

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

亚马逊云科技 多模态生态 × PixVerse 模型实践

= 全球创作者的视频 AI 基础设施

接下来，把舞台交给 PixVerse

从视频模型到视频 Agent – AI 如何接管内容生产

从视频模型到视频 Agent

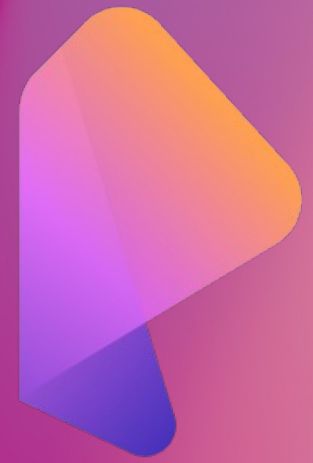
从5～15秒分镜头生成，到自动完成1个完整视频生产任务

01 爱诗科技公司介绍

02 生成式视频模型发展

03 覆盖 UGC 和专家的视频 Agent 能力

04 我们相信 Agent 是下一代生产力



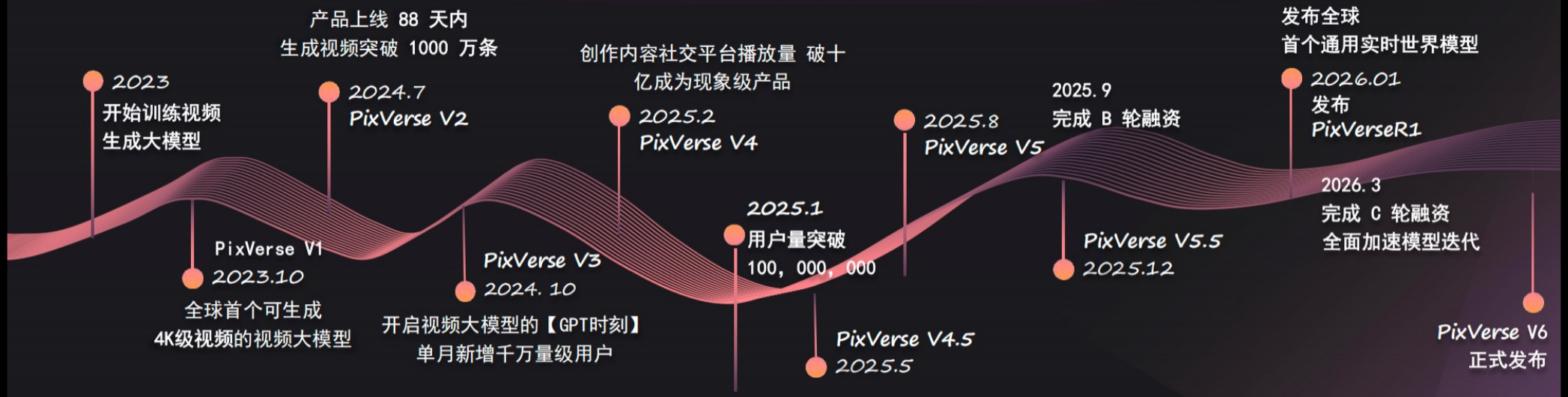
PixVerse

01 爱诗科技公司介绍

PixVerse

两年内，PixVerse已完成八代模型选代，持续突破AI视频生成技术的边界，

稳居全球视频生成模型第一梯队



全球用户	生成视频	覆盖国家
1亿+	8亿+	177个

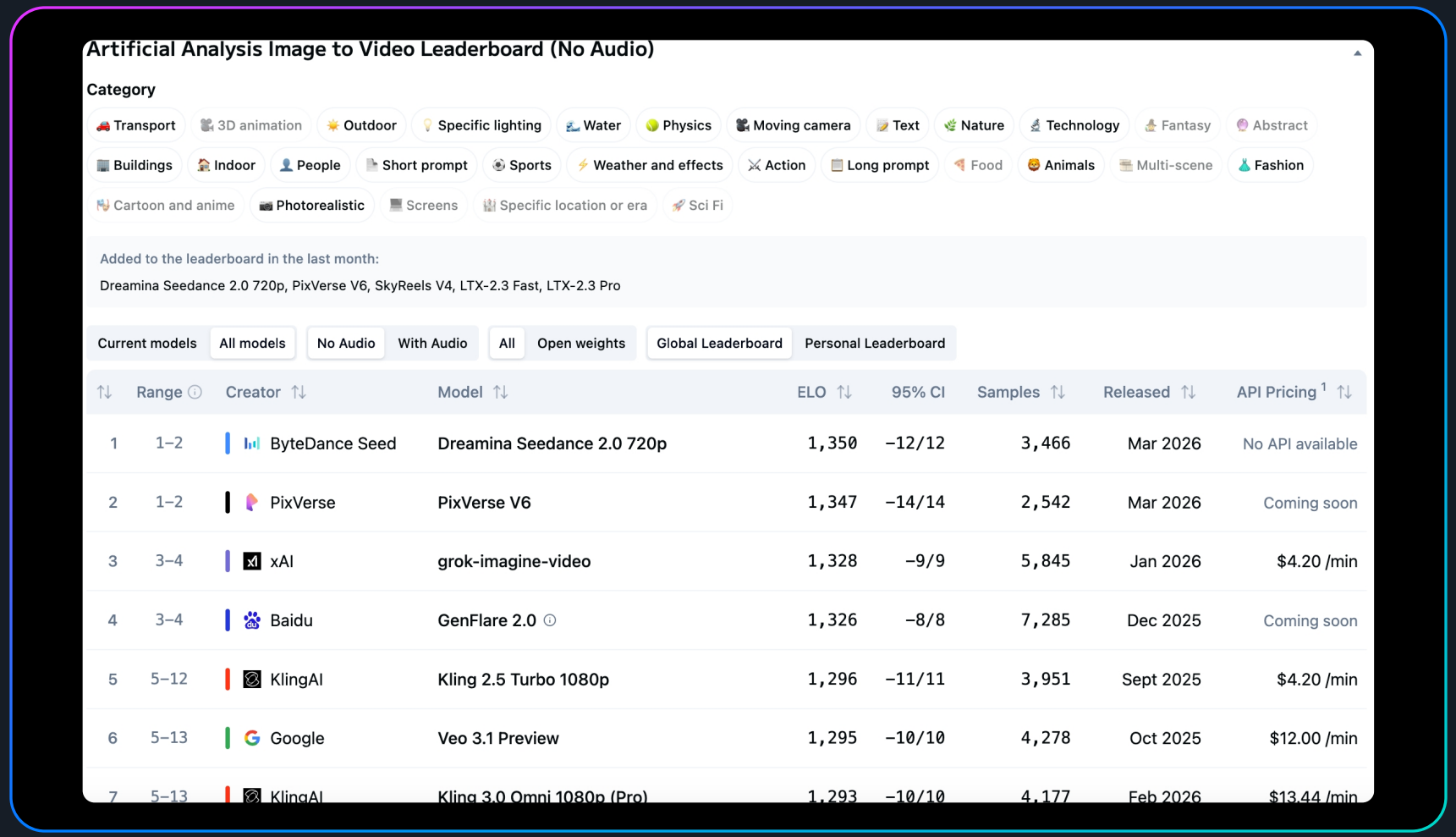
王长虎 博士

爱诗科技创始人/CEO

- 从0-1搭建全球最顶级的视频AI技术与产品团队，打造字节的视觉多模态大模型
- 带领数百人团队，管理数万块GPU、负责每日亿级投稿视频的生成、处理与分发
- 凭借国际领先的视频AI能力，与业务团队携手、助力抖音、TikTok飞速发展

PixVerse

26.04 Artificial 榜单



Artificial Analysis Image to Video Leaderboard (No Audio)

Category

Transport | 3D animation | Outdoor | Specific lighting | Water | Physics | Moving camera | Text | Nature | Technology | Fantasy | Abstract | Buildings | Indoor | People | Short prompt | Sports | Weather and effects | Action | Long prompt | Food | Animals | Multi-scene | Fashion | Cartoon and anime | Photorealistic | Screens | Specific location or era | Sci Fi

Added to the leaderboard in the last month:

Dreamina Seedance 2.0 720p, PixVerse V6, SkyReels V4, LTX-2.3 Fast, LTX-2.3 Pro

Current models | All models | No Audio | With Audio | All | Open weights | Global Leaderboard | Personal Leaderboard

	Range	Creator	Model	ELO	95% CI	Samples	Released	API Pricing [1]
1	1–2	ByteDance Seed	Dreamina Seedance 2.0 720p	1,350	−12/12	3,466	Mar 2026	No API available
2	1–2	PixVerse	PixVerse V6	1,347	−14/14	2,542	Mar 2026	Coming soon
3	3–4	xAI	grok-imagine-video	1,328	−9/9	5,845	Jan 2026	$4.20 /min
4	3–4	Baidu	GenFlare 2.0	1,326	−8/8	7,285	Dec 2025	Coming soon
5	5–12	KlingAI	Kling 2.5 Turbo 1080p	1,296	−11/11	3,951	Sept 2025	$4.20 /min
6	5–13	Google	Veo 3.1 Preview	1,295	−10/10	4,278	Oct 2025	$12.00 /min
7	5–13	KlingAI	Kling 3.0 Omni 1080p (Pro)	1,293	−10/10	4,177	Feb 2026	$13.44 /min

PixVerse

模型进化

打造复合的模型矩阵，用模型组合来适应不同行业

C 系列 模型

为影视而生

打斗/特效升级丨多宫格智能分镜叙事

E 系列 模型

营销专项

针对不同品类的更准确表达

混剪/一键成片/高光剪辑

更贴合的营销叙事

R 系列 模型

全球首个支持最高1080P分辨率

通用实时世界模型

实时交互—所想即所见、所说即所现

根基 V 系列 模型

高频迭代、持续提升

稳居全球视频生成模型第一梯队

PixVerse

2022年

AI 能生成视频吗?

2024年

AI 能生成更真实的视频吗?

2026年

AI 能生成更长更完整的视频吗?

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

03 覆盖 UGC 和专家的视频 Agent 能力

PixVerse

UGC 人人都能用

我有一个想法，AI 帮我自动完成

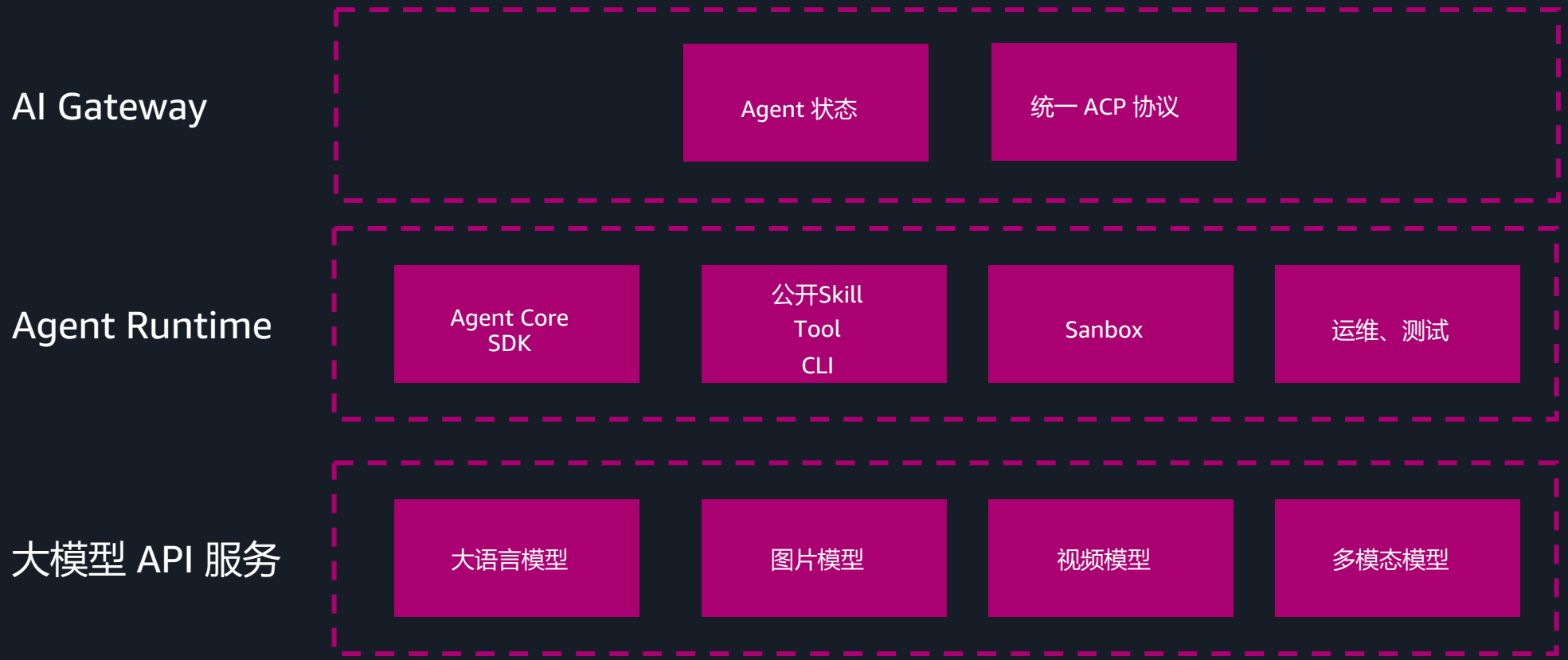
专业创作者

我有一个脚本，AI 辅助我完成

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

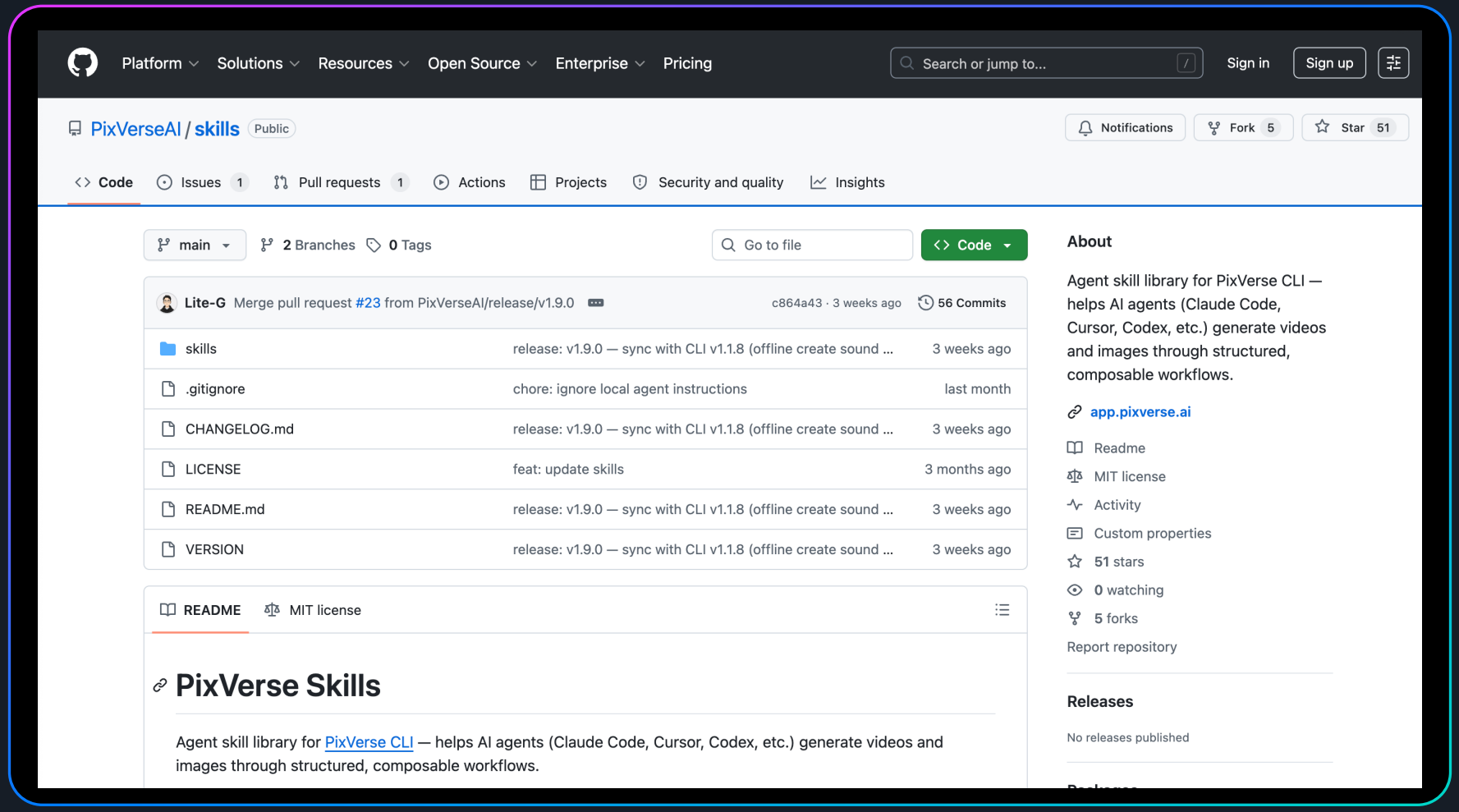
爱诗科技视频 Agent 技术架构



*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

爱诗科技公开 skills



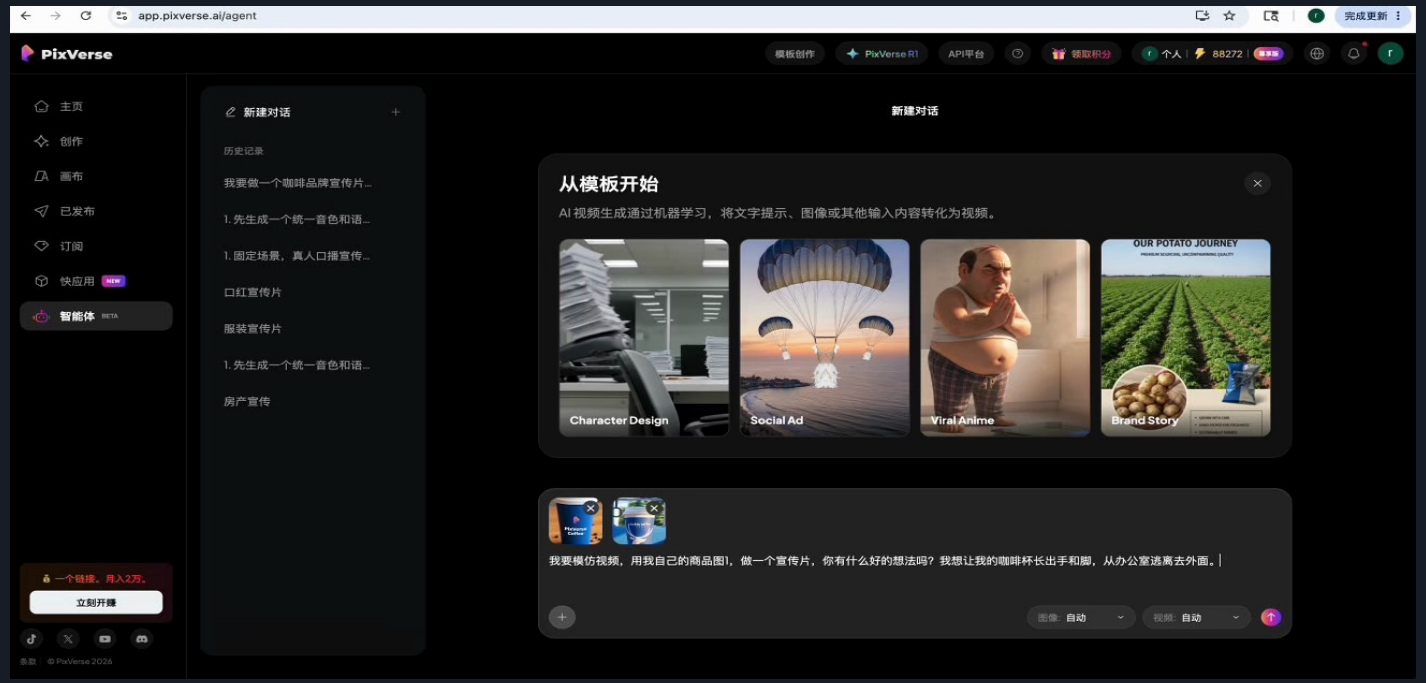
*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

UGC 人人都能用 对话式视频 Agent

输入想法 --

- LLM：需求澄清
- 多模态：素材分析
- LLM：设计脚本
- IT2V：生成分镜视频
- 工具：拼接剪辑



*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

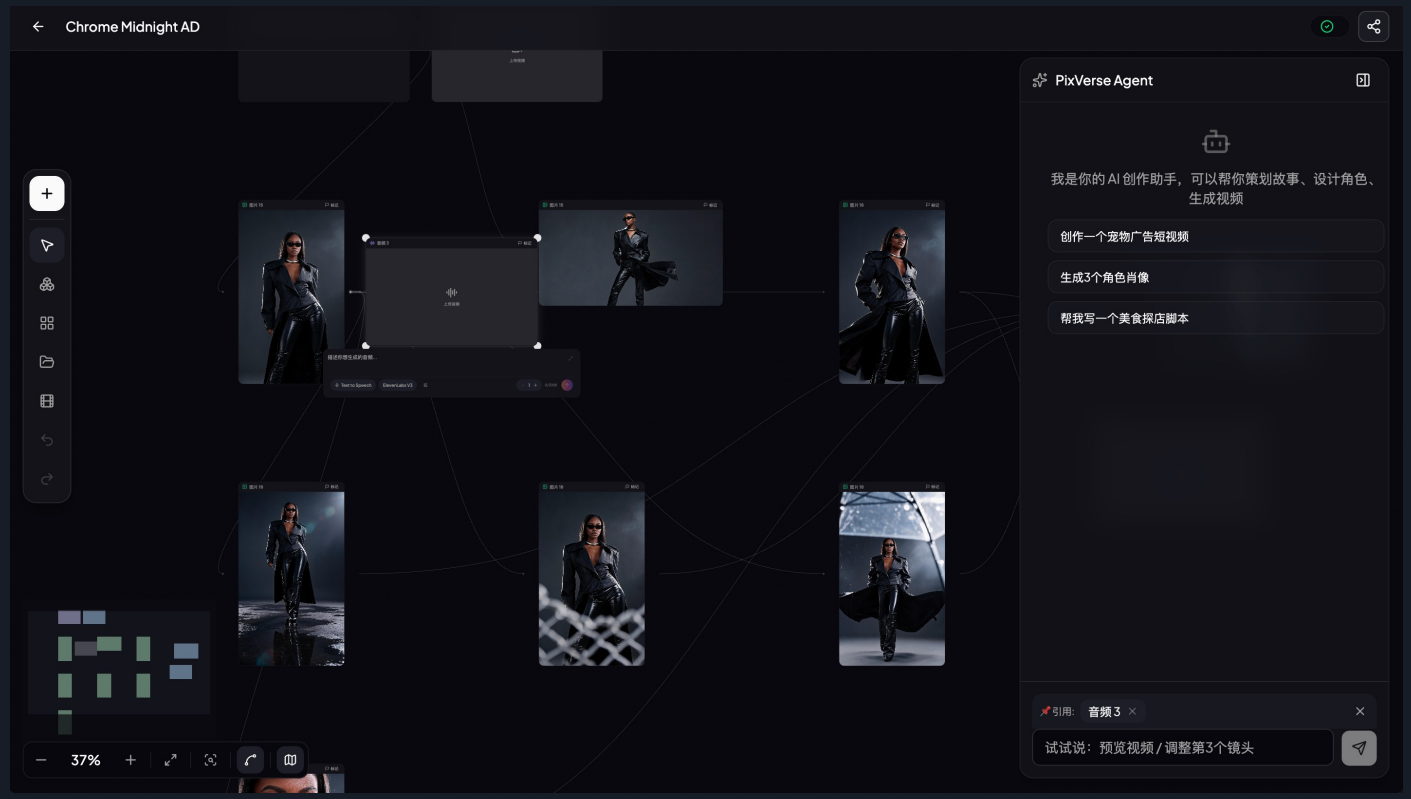
UGC 人人都能用--对话式视频 Agent

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

专业创作者
对话+无限画布

- 对话自动生成画布
- 脚本、图片、视频可在一个页面完成
- 可复刻创组者画布



PixVerse

专业创作者
对话+无限画布

- 对话自动生成画布
- 脚本、图片、视频可在一个页面完成
- 可复刻其他创组者的画布

视频成片

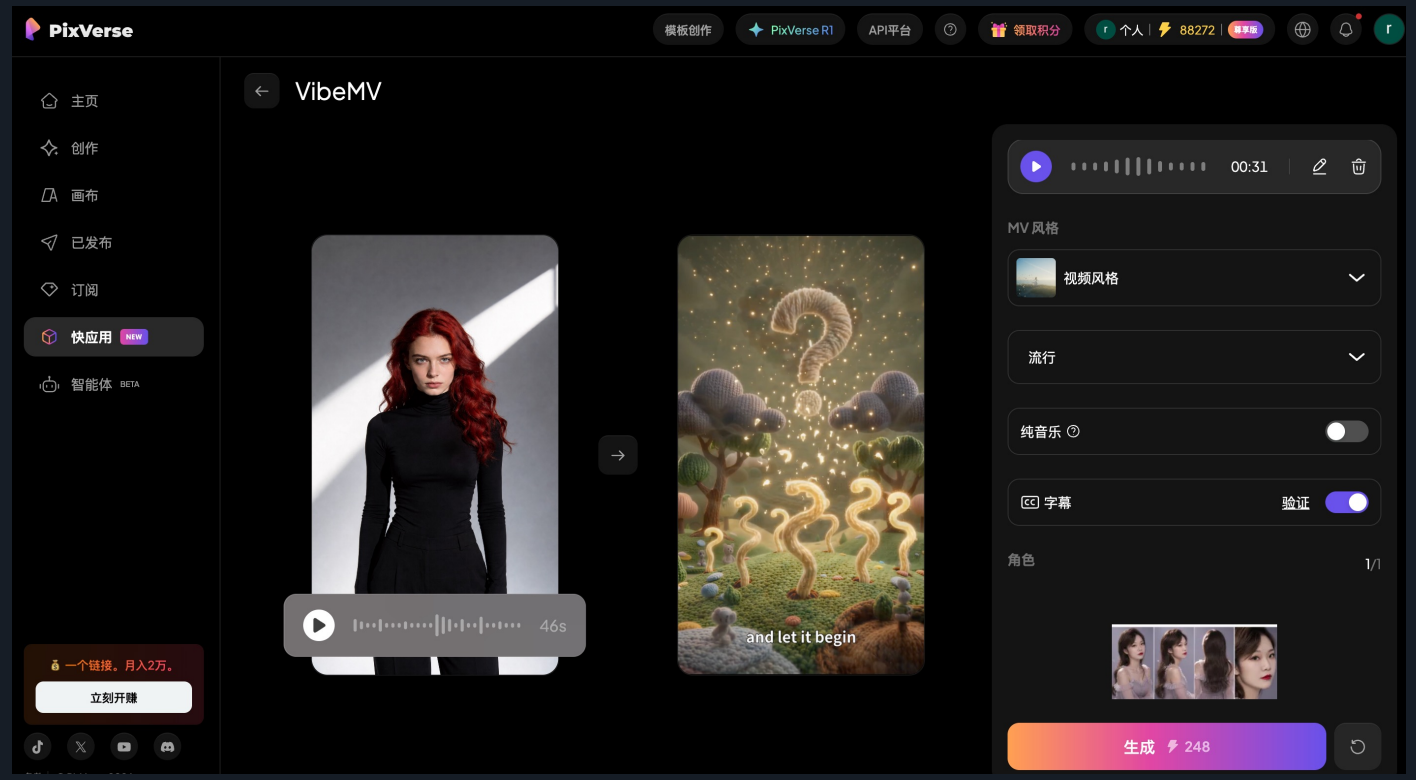
*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

垂直行业 Agent

输入音频 MP3 自动做音乐 MV 视频 --

- LLM：歌词分析
- LLM：分镜脚本
- IT2V：生成分镜视频



*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

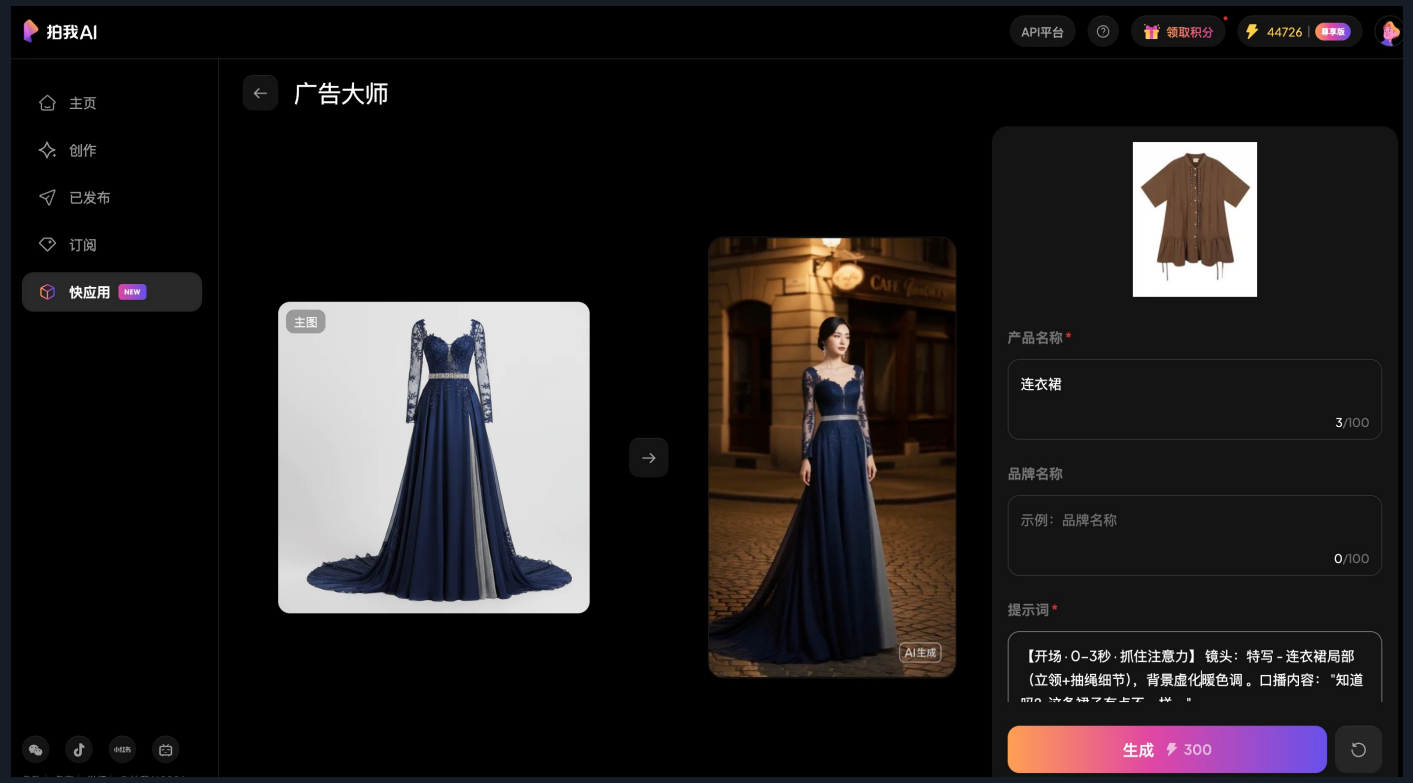
行业 Agent--音乐 mv

PixVerse

行业 Agent--营销视频

- 输入：商品信息
- Agent工作：
  - LLM：意图、商品信息理解
  - 多模态：素材分析
  - LLM：设计脚本、写营销文案
  - IT2V：生成分镜视频
  - 工具：拼接剪辑
- 输出：完整的营销视频

操作页面



*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

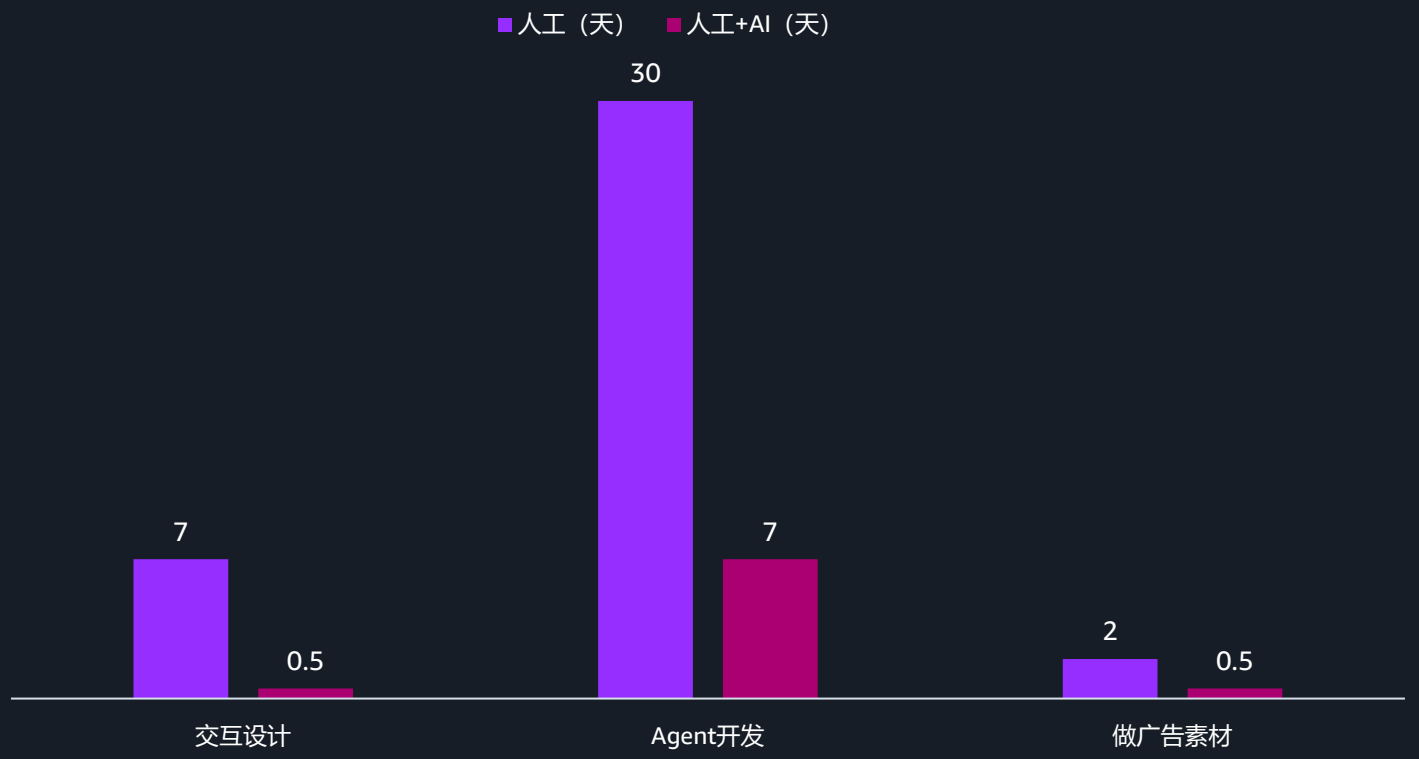
PixVerse

行业 Agent --营销视频

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

PixVerse

爱诗科技 Agent 实践



PixVerse

爱诗科技 Agent 实践

交互设计

过去：1-2周
MRD->PRD->UX->UI

现在：0.5天
Open AI Codex做UX/UI

Agent 开发

过去：1-2个月
算法->工程->前后端

现在：1-2周
-内测：产品AI coding
-增长：研发介入保障稳定

广告投放

过去：广告费+代理服务费

现在：
OpenClaw 搭建 workflow->自动做素材->接入广告账号自动投放->回收数据->分析和爆款的差距自动学习

PixVerse

产品设计

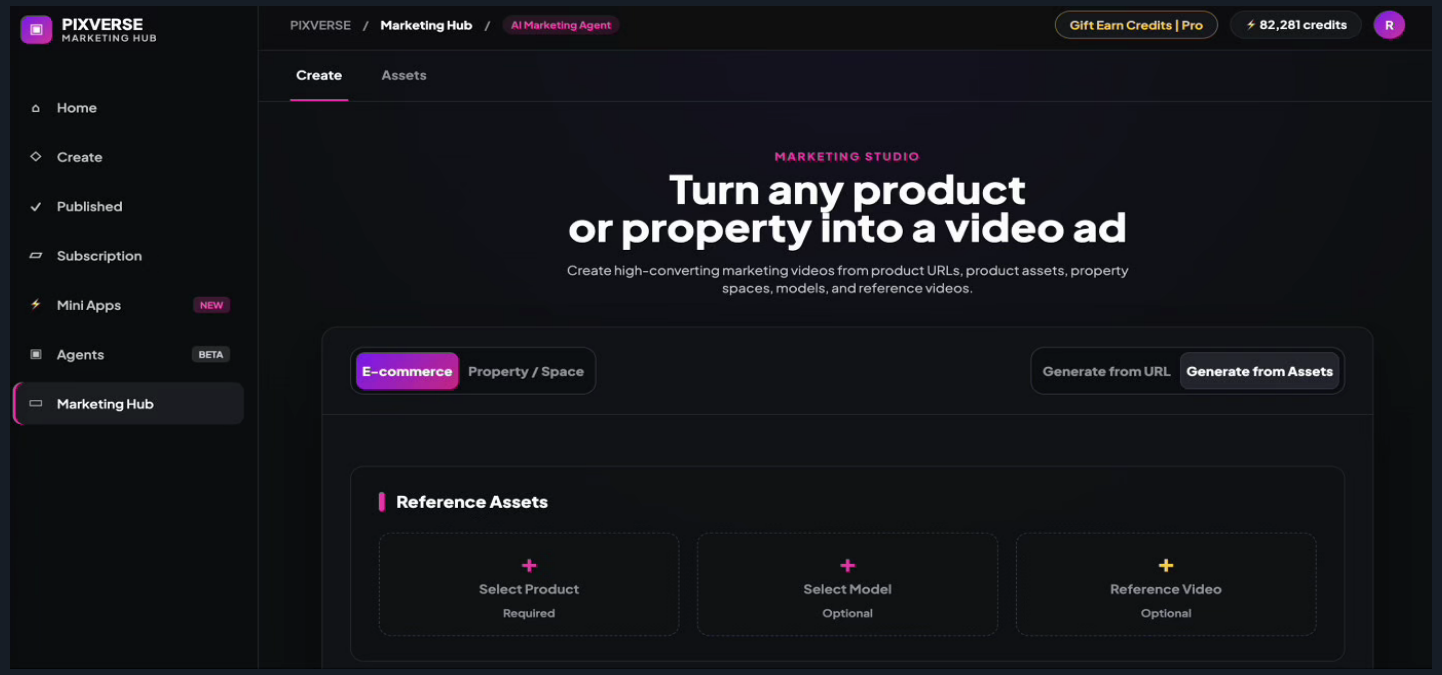
- 输入：文字需求、参考交互图、授权公司的 figma 设计组件规范读取权限
- 输出：Open AI Codex 做 UX/UI

PixVerse

产品设计

- 输入：文字需求、参考交互图、授权公司的 figma 设计组件规范读取权限
- 输出：Open AI Codex 做 UX/UI

UX 演示



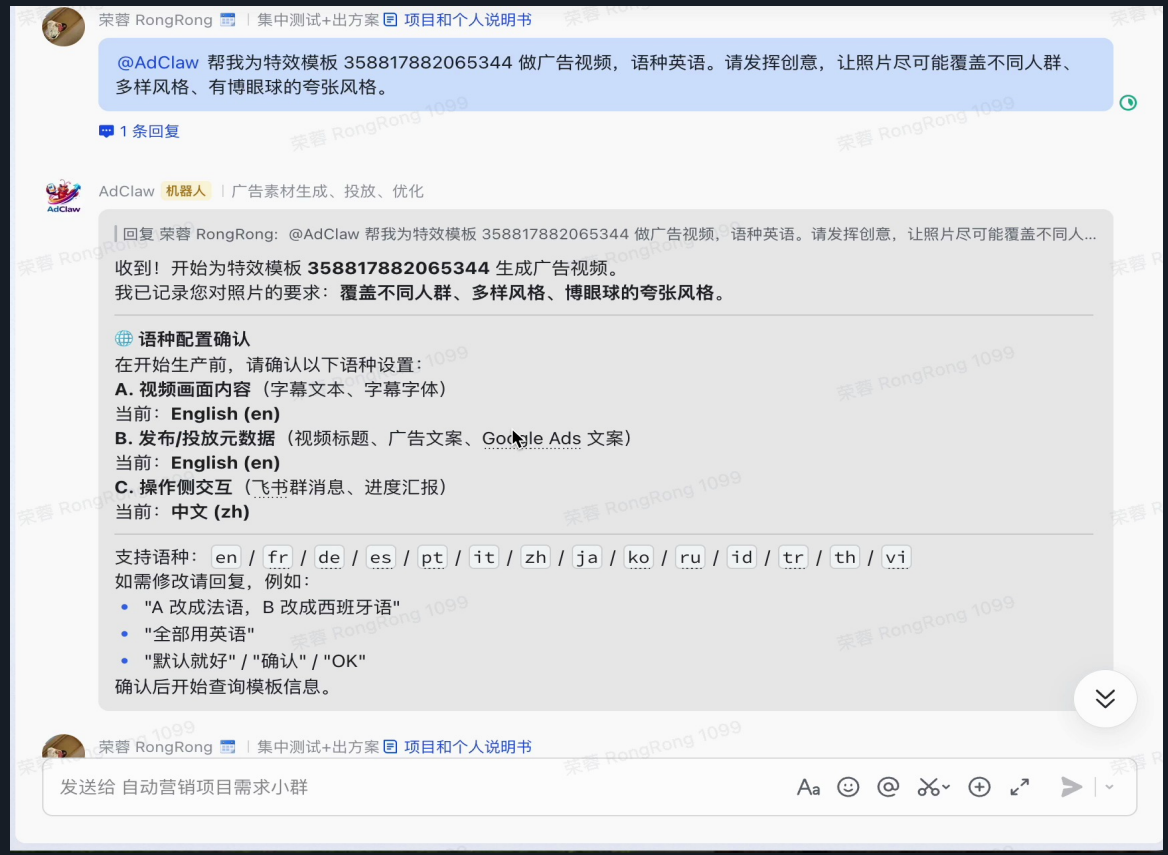
PixVerse

产品投放

自然语言搭建 OpenClaw

- 输入：需求 或 模板 id
- Agent 工作：需求分析和链路匹配
- 输出：广告投放视频

操作录屏



PixVerse

产品投放

基于 OpenClaw 搭建 workflow

- 输入：文字需求
- Agent 工作：需求分析和链路匹配
- 输出：广告投放视频

视频成片

PART 03

与亚马逊云科技
共建全球创作基础设施

为 PixVerse 提供全球低延迟交付、企业级合规、海量推理算力。

让"灵感一键成片"成为基础设施能力。

PixVerse 技术优势 × 亚马逊云全球基础设施

三层能力共建，支撑 1 亿 + 用户 · 175 国家 的高并发视频生成 [Source] PixVerse 官方 / GeekWire 2026.03

LAYER 01

全球低延迟交付

Global Low-Latency Delivery

36+

亚马逊云全球 Region 网络

让创意从生成到分发以秒级触达全球。

- 全球 PoP + CloudFront 加速
- 就近推理 / 就近渲染
- 创作者→观众端到端延迟优化

LAYER 02

企业级安全合规

Enterprise Security & Compliance

175+

国家/地区监管落地

本地数据合规、区域监管适配。

- ISO / SOC / GDPR / HIPAA
- 数据驻留 + 加密传输/存储
- Bedrock Guardrails 内容护栏

LAYER 03

规模化推理算力

Massive GPU Inference

100M+

PixVerse 全球用户基数

承接亿级日活的视频生成负载。

- 丰富的 GPU资源 + Trainium
- 弹性容量池 / Spot 经济性
- SageMaker/EKS等资源管理平台

亚马逊云科技
中国峰会

Thank you

亚马逊云科技
中国峰会

Session 4:PixVerse:从像素到叙事：多模态模型如何重构视频生成的能力边界

扫描上方二维码
填写调查问卷

亚马逊云科技
中国峰会

Session 4:PixVerse:从像素到叙事：多模态模型如何重构视频生成的能力边界

扫描上方二维码
填写调查问卷

亚马逊云科技
中国峰会

Session 4:PixVerse:从像素到叙事：多模态模型如何重构视频生成的能力边界

扫描上方二维码
填写调查问卷

亚马逊云科技
中国峰会

Session 4:PixVerse:从像素到叙事：多模态模型如何重构视频生成的能力边界

扫描上方二维码
填写调查问卷