Vimo AI 开放 API

修订：2026-11-13。本版本支持多素材、首尾帧、用户分组价格；失败说明优先保留上游可公开的原文，仅过滤隐私和内部诊断信息。

1. 地址与鉴权

基地址：https://vimo.asia/api/v1。模型 ID、能力、价格以实时模型目录为准。

除模型目录外，所有接口携带以下任一请求头：

```
Authorization: Bearer sk-YOUR_API_KEY
X-Api-Key: sk-YOUR_API_KEY
```

登录站点 /api-console 创建、复制或吊销密钥。Webhook 签名密钥在创建时保存。密钥只放在服务端。本站接口不是 OpenAI SDK 兼容协议。

2. 端点

方法与路径	说明
GET /api/v1/models	公开模型目录，无需 Key
POST /api/v1/videos/generations	创建视频、图片、文本任务
GET /api/v1/videos/generations/{id}	查询当前账户任务
GET /api/v1/account	余额、月净消耗及限频配置
GET /api/tasks/{id}/content	视频与图片下载，不在 /api/v1 下

业务成功响应直接返回资源对象。模型目录返回 {object:"list",data:[...]}。错误保留 {error:{code,message}}，并增加原因、来源和重试建议字段，见第 12 节。下载返回媒体字节，错误可能是纯文本。

3. 模型目录

```
curl https://vimo.asia/api/v1/models
```

id、label 使用后台“定价”设置的模型展示名称。创建任务请从实时目录取 id；任务查询与回调的 model 使用对应展示名称。供应商字段及内部模型标识不对外返回。修改展示名称可能影响调用方，请提前通知接入方更新配置。

不携带 Key 时返回默认报价；携带有效 API Key 时返回该 Key 所属用户的分组报价。无效或已吊销的 Key 返回 401。同一用户的全部密钥与网站内生成共用用户分组。组价未配置的字段回退默认价；明确设置为 0 的价格表示免费。请携带自己的 Key 查询目录后估算费用，不要缓存或公开其他账户报价。

字段	含义
id / label / description	档位 ID、名称、说明；请求 model 使用 id
kind	video / image / text
billing.mode	per_second / per_request / per_image / per_token
billing.credits	基础单价；per_token 时不能据此估算费用
resolution_credits	对应分辨率的积分单价，已包含倍率
token_pricing	按 token 档位的 input_per_million / output_per_million，单位积分；其他为 null
seconds	min / max / allowed / default；allowed 非空时精确选档
aspect_ratios / resolutions / sizes	支持的画幅、分辨率、像素尺寸数组
modes	text / image / reference；视频 image 表示首尾帧
limits	max_images / max_audios / max_videos；本 API 每类最多 100 个，实际不得超过模型上限
tags	展示标签

素材数量上限不代表所有素材组合均被上游接受。素材格式、大小和时长还受对应模型限制。

4. 创建任务字段

请求 POST /api/v1/videos/generations，Content-Type 为 application/json。

JSON 请求体上限 1 MiB，超限返回 400。素材使用 HTTPS URL，不提交文件内容或 base64。

字段	类型与要求
model	string，必填，1-64 字符
prompt	string，必填，去首尾空白后 1-15000 字符
seconds	正整数，视频必填，符合模型时长范围
aspect_ratio	string，视频与 size 至少传一项；图片可选
resolution	string，可选，原样使用目录值；省略取默认
size	string，如 1280x720；按目录尺寸匹配画幅及分辨率
mode	text / image / reference；可选，按素材与模型能力推断
images	HTTPS URL 数组，多参考图或图生图

字段	类型与要求
audios	HTTPS URL 数组，参考音频
videos	HTTPS URL 数组，参考视频
first_frame	HTTPS URL，视频首帧
last_frame	HTTPS URL，视频尾帧；必须同时提供 first_frame
input_reference	旧版单素材 HTTPS URL，保留兼容
client_request_id	string，1-64 字符，账户级订单幂等编号
callback_url	HTTPS URL，任务回调地址

每个 URL 最长 2048 字符，不允许 URL 用户名密码。链接须供上游读取，且有效期覆盖排队和生成过程。不要提供内网链接。

规则：

- images、audios、videos 可混合使用；逐类检查模型数量上限。
- 视频 image 模式只接受一至两张图，按首帧、尾帧排列；不接受音视频。
- first_frame / last_frame 不能与素材数组混用，必须使用 image 模式。
- input_reference 不能与素材数组或首尾帧同时传入；按 URL 路径后缀区分音视频，其余按图片。
- 图片接受多张 images，仅限声明支持参考图的档位；不接受音频、视频及首尾帧字段。
- 文本目前只接收 prompt，不支持参考素材或非 text 模式。
- 同时传 size、画幅、分辨率时必须一致。图片不再静默覆盖分辨率；K 档位按实际像素尺寸匹配。
- 未声明的字段返回 422，不再静默忽略。模型专属字段尚未建立契约，不能任意透传。

5. 视频混合参考示例

MODEL_ID、时长、分辨率、画幅及素材数量须按实时目录替换。URL 为示例，不可直接用作真实素材。

```
curl -X POST https://vimo.asia/api/v1/videos/generations \
  -H "Authorization: Bearer sk-YOUR_API_KEY" \
  -H "Content-Type: application/json" \
  -d '{
    "model": "MODEL_ID",
    "prompt": "保持产品外观，参考视频运镜",
    "seconds": 5,
    "aspect_ratio": "16:9",
    "resolution": "720P",
    "mode": "reference",
    "images": ["https://example.com/a.jpg", "https://example.com/b.jpg"],
    "videos": ["https://example.com/motion.mp4"],
    "audios": ["https://example.com/sound.wav"],
    "client_request_id": "your-unique-order-id"
  }'
```

6. 首尾帧、图片与文本示例

首尾帧视频请求体：

```
{
  "model": "VIDEO_MODEL_ID",
  "prompt": "从起始画面自然过渡到结束画面",
  "seconds": 5,
  "aspect_ratio": "16:9",
  "mode": "image",
  "first_frame": "https://example.com/first.jpg",
  "last_frame": "https://example.com/last.jpg"
}
```

图生图请求体：

```
{
  "model": "IMAGE_MODEL_ID",
  "prompt": "保持产品与标签，改为自然侧光背景",
  "aspect_ratio": "1:1",
  "images": ["https://example.com/product.jpg"]
}
```

文本请求体：

```
{"model":"TEXT_MODEL_ID","prompt":"为产品写三条简短广告语"}
```

纯文生图片或视频不传参考素材。视频仍需 seconds 和画幅。

7. 查询与结果

创建成功 HTTP 201，返回当前任务对象，不保证状态固定为 queued。间隔 10-15 秒查询：

```
curl https://vimo.asia/api/v1/videos/generations/TASK_ID \
  -H "Authorization: Bearer sk-YOUR_API_KEY"
```

字段	含义
id / model	任务 ID / 模型展示名称
object / kind	video / image / text
mode	实际 text / image / reference 模式
status	queued / in_progress / completed / failed / cancelled
seconds / aspect_ratio / resolution	本次有效参数
video_url	完成的视频或图片下载地址，其他为 null
output_text	文本正文，其他为 null
usage	credits 为费用，refunded 为已退积分

字段	含义
error	失败时保留 code="generation_failed"，新增 reason_code、具体 message、source、upstream_status、param、retryable、retry_action；否则 null
created_at / updated_at	Unix 整数秒

completed、failed、cancelled 为终态。没有公开取消端点；客户端等待超时不会取消服务器任务或自动退款。无权访问与不存在统一返回 404。

8. 下载

原样使用 video_url，并携带同账户有效 Key。校验地址属于本站后再发送密钥。

```
curl --fail "https://vimo.asia/api/tasks/TASK_ID/content" \
  -H "Authorization: Bearer sk-YOUR_API_KEY" -o result.bin
```

按 Content-Type 确定文件格式；支持 Range，分段响应可能为 206。不要把错误正文保存成视频。代理地址稳定不代表文件永久保存，重要结果及时下载。

9. 幂等与重试

相同账户、client_request_id、有效参数返回原任务，HTTP 201，带 X-Idempotent-Replay: true，不再次扣费。改变模型、提示词、素材顺序、模式、画幅、分辨率、视频时长或 callback_url 返回 409。文本只比较有效的模型、提示词与 callback_url。

新订单使用新编号；网络超时重试保留原编号和请求。失败任务不会自动重新生成；要再次生成需新编号。

10. 账户、计费与限频

GET /api/v1/account 需要 Key，返回 object、pricing_group、balance、consumed_this_month、rate_limit。pricing_group 是当前账户的定价组；只有管理员可调整。月净消耗按服务端时区及账本扣费、退款发生时间汇总。

视频按秒或按次、图片按张，按对应分辨率计价；目录分辨率价格已包含倍率，不重复相乘。按 token 档位按实际输入、输出用量结算，单价见 token_pricing。视频和图片的常规按秒、按次、按张档位建单预扣，失败退回；文本事后扣费，可能因并发结算出现负余额。费用以 usage 为准。

限频维度	限制
单 Key	每秒 10 次、每分钟 60 次
同账户全部 Key	每分钟 120 次
同 Key 查询同任务	至少间隔 5 秒

429 按 Retry-After 秒数等待。当前不限制运行任务数，接入方应控制并发；不代表无限容量。模型目录与下载不计入上述业务限频。

11. Webhook

callback_url 使用 HTTPS。视频成功发送 video.completed；统一失败链路发送 video.failed，图片和文本也可能收到失败事件，但不发送成功事件。图片和文本以查询结果为准。

```
X-Vimo-Event: video.completed | video.failed
X-Vimo-Signature: t=<Unix秒>,v1=<HMAC十六进制>
```

事件包含 id、object=event、type、created、data。data 包含任务 id、object=video、status、model、video_url、usage；失败额外含 error，与任务查询使用同一套安全原因结构，见第 12 节。data.object 是历史固定字段，真实类型查询 kind。

用创建该任务的 Key 对应 webhook_secret 对 t + "." + 原始请求体计算 HMAC-SHA256。不要对解析后重新序列化的 JSON 验签。

```
import { createHmac, timingSafeEqual } from "node:crypto";
function verify(rawBody, header, secret) {
  const m = /^t=(\d+),v1=([0-9a-f]{64})$/.exec(header || "");
  if (!m || !Buffer.isBuffer(rawBody)) return false;
  if (Math.abs(Date.now()/1000 - Number(m[1])) > 300) return false;
  const expected = createHmac("sha256", secret)
    .update(m[1] + ".").update(rawBody).digest();
  return timingSafeEqual(expected, Buffer.from(m[2], "hex"));
}
```

接收后去重、保存并尽快返回 2xx。每次超时 10 秒，首次失败后最多重试 3 次，间隔 1/4/16 秒；4xx 除 408/429 外停止重试。无持久化投递队列或补发接口，需保留低频查询兜底。吊销旧 Key 后历史任务仍使用旧 secret，轮换时保留过渡验签能力。

12. 错误处理

12.1 错误来源与兼容

error.message 优先返回上游可安全公开的失败原文，保留原来的语言、尺寸、数量、参数要求和不确定性，不替换为固定归类文案。例如上游返回“号池无号”时仍显示“号池无号”；英文原因保留英文。
reason_code 用于分类处理，不代表 message 必须是固定句子。未命中分类的新原因，只要可以安全分离，也会返回。

上游明确报密钥无效、账户额度不足、响应解析配置错误时，分别显示“站长密钥失效”“站长余额不足”“站长通道配置错误”。未配置密钥单独显示“站长未配置密钥”。不返回真实密钥、账户、剩余额度、供应商名称或地址。

请求拒绝保留原 HTTP 状态及原 code；已失败任务的查询仍为 HTTP 200、status="failed"、error.code="generation_failed"。上游的 403、429 等写在 error.upstream_status，不冒充我们接口本次返回的 HTTP 状态。客户端请读取响应正文，不要只看 HTTP 200 判断生成成功。

```
{
  "error": {
    "code": "generation_failed",
    "reason_code": "site_balance_insufficient",
    "message": "站长余额不足，请等待站长充值",
    "source": "upstream",
    "upstream_status": 403,
    "param": null,
    "retryable": false,
    "retry_action": "wait_for_site"
  }
}
```

以上示例成立的前提是上游明确说明余额不足。只有 403、没有具体说明时，保留其可公开的状态说明或返回通用 403 说明，不猜测余额或密钥。

字段	含义
code	兼容原客户端的错误码；任务失败仍为 generation_failed
reason_code	稳定的具体原因标识，见下面清单
message	优先为过滤隐私后的上游原文；涉及站长密钥、账户或配置时返回站长原因；无安全原文才使用归类说明
source	request=接入用户请求或权限；upstream=模型服务生成链路报错；site=站长服务内部异常
upstream_status	已记录的上游 HTTP 状态；未记录则 null；不是我们的 HTTP 响应状态
param	可确认的公开参数路径，例如 seconds、images[1]；不能确认则 null
retryable	true=条件允许时可重试；false=需改参数或等待修复；null=无法安全判断。不是自动重发指令
retry_action	推荐动作，见第 12.4 节
request_id	请求被拒绝时的排查编号，同时通过 X-Request-ID 响应头返回；任务查询及 Webhook 的 error 不保证包含此字段

source 表示错误的报告链路，不等于责任归属。例如上游报“站长余额不足”时 source=upstream，负责处理的是站长。

12.2 我们接口返回的 HTTP 状态

HTTP / code	含义与处理
400 BAD_REQUEST / VALIDATION_FAILED	JSON、字段类型、请求体或业务校验不正确；按 message 修正
401 INVALID_API_KEY / UNAUTHORIZED	接入用户没有通过鉴权；检查本网站签发的 API Key。无效、吊销和账户不可用不进一步区分，避免账户探测
402 INSUFFICIENT_CREDITS	接入用户在本网站的积分不足；与站长上游余额不足分别表示
403 FORBIDDEN	当前接入用户没有执行操作的权限；不包含其他用户的资源详情
404 NOT_FOUND	任务不存在或不属于当前账户；两者不进一步区分
409 CONFLICT	同请求编号被用于不同参数，或当前状态发生冲突
422 UNPROCESSABLE	参数、素材组合、时长、分辨率、比例或模型不符合要求
429 RATE_LIMITED	本网站限频；按 Retry-After 秒数等待
500 INTERNAL	站长服务内部异常；保留原请求编号、参数并先核对是否已建单
502 PROVIDER_ERROR	模型服务链路异常；查看 reason_code、message、upstream_status，不要当成下游 Key 失效

下载接口成功时返回文件或分段字节；错误可能是纯文本及非 2xx 状态，不要把报错正文保存成视频。生成已成功与结果下载失败是两个不同阶段，不会因本次下载失败自动改写任务状态。

12.3 上游 HTTP 状态可能对应的原因

下面是可能原因，不是根据数字直接下结论。服务商可能用同一个状态表示不同业务问题，应优先使用具体业务说明。未记录上游状态时 upstream_status=null，不从任意三位数字、任务号或业务错误码猜 HTTP 状态。

上游状态	常见含义及本接口处理
400	请求或业务参数被拒绝；有具体原因则反馈参数、素材问题
401	站长通道鉴权失败；明确密钥失效才标 site_key_invalid，未说明时标 site_authentication_failed
402	对方要求付款或额度校验未通过；明确余额不足才标 site_balance_insufficient
403	可能是模型权限、IP 白名单、区域限制、账号池暂无可用账号、访问策略、素材访问权限等。明确哪项就反馈哪项；只给 Forbidden 则不猜
404	可能是模型、任务、资源或接口目标不存在；只给 404 不直接认定通道配置错误
405	请求方式不受支持，提示站长通道配置问题

上游状态	常见含义及本接口处理
408	对方请求等待超时；先核对原任务，不假定生成已失败
409	任务、幂等或状态冲突；先核对原任务
413	请求或素材过大；调整素材或请求体
415	媒体类型或格式不受支持
422	语法可解析但业务内容不符合要求；按实际提示修正
429	可能是限频、并发或排队限制；明确原因分别反馈。没有可靠等待秒数时不编造 `Retry-After`
500 / 502 / 503	模型服务内部错误、网关异常、繁忙或维护；优先反馈实际业务原因
504	对方网关等待超时，不等于已确认生成超时；先查原任务

HTTP 语义参考：https://www.rfc-editor.org/rfc/rfc9110.html 。429 参考：https://www.rfc-editor.org/rfc/rfc6585.html 。实际业务解释以响应中的可公开原因和已记录事实为准。

12.4 原因标识和处理动作

reason_code	原因
invalid_parameter	接入用户参数不正确，具体要求见 message / param
invalid_api_key / authentication_required	接入用户 Key 或身份验证问题
insufficient_credits	接入用户账户积分不足
access_denied / not_found / request_conflict / rate_limited	接入权限、资源、请求冲突或本网站限频
reference_download_failed / reference_access_denied	参考素材无法下载或访问被拒绝
reference_expired	上游明确表示素材链接过期
reference_import_unavailable	素材导入失败；对方只说“过期或不可用”时保留这种不确定性，不直接认定过期
reference_count_exceeded	素材数量超限
reference_format_invalid	格式、编码或文件内容不符合要求
reference_constraints_invalid	素材大小、尺寸或时长不符合要求
frame_invalid / reference_subject_unsupported	首尾帧或素材内容类型不受支持
reference_content_rejected / output_content_rejected / content_rejected	参考图片、生成结果或输入内容未通过审核
copyright_similarity_rejected	生成内容未通过版权或 IP 相似性检查；不代表平台自行认定侵权
context_length_exceeded / model_parameter_invalid	提示词上下文过长或模型参数不受支持
site_key_invalid / site_key_missing	站长密钥失效，或明确未配置密钥；两者分别标识
site_authentication_failed	站长通道鉴权失败，但未明确密钥具体问题
site_balance_insufficient / site_payment_required	站长余额不足，或付款 / 额度校验未通过
site_channel_config_error	站长通道配置或响应解析错误
site_account_pool_unavailable	对方明确反馈账号池暂无可用账号，需要站长处理通道资源
site_access_restricted / site_model_permission_denied / site_permission_denied	站长通道白名单、区域限制、模型权限或其他已明确的权限不足

reason_code	原因
site_network_error	站长通道连接、域名解析或证书问题；隐藏实际地址和网络细节
site_task_interrupted	站长服务处理过程中断，先核对原任务；请求中断不会擅自解释成生成超时
service_queue_full / service_concurrency_limited / service_rate_limited	模型服务队列已满、并发达到上限或请求受限
model_unavailable / service_unavailable	模型维护、停用或服务异常
generation_timeout	上游明确说明生成任务超时
service_timeout	请求或通道响应超时；不据此新增失败状态
service_access_denied / service_not_found / service_conflict / service_request_rejected	对方拒绝访问、目标不存在、冲突或拒绝请求，未提供更具体的可公开原因
upstream_business_error	其他可安全分离的上游失败说明，见 message 原文
site_internal_error	站长服务内部异常，原始诊断留在后台
generation_failed	生成失败，暂无具体原因
generation_failed_or_cancelled	对方只反馈“视频生成失败或被取消”。不能据此认定是接入用户主动取消，也没有更细原因

retry_action	处理
fix_request	修正参数或素材；需要新建任务时使用新请求编号
fix_credentials	检查接入用户自己的本网站 API Key
top_up	补足接入用户在本网站的积分
wait_for_site	等待站长修复密钥、余额、权限或配置
retry_later	等待后有限重试，未确定是否已建单时保留原编号
retry_same_request	保留 client_request_id 及完全相同参数重试，先核对已有任务
query_task	先查原任务，避免重复生成
new_task	原任务已明确失败；决定再次生成时用新编号。这会创建新任务并按规则计费，客户端不得无上限自动重发
none	没有可安全推荐的重试动作，先核对已有状态与原因

12.5 隐私边界与历史记录

公开问题结论，不公开供应商名称、真实域名 / IP / 接口路径、内部模型标识、上游任务号、真实密钥、Authorization、Cookie、签名参数、账户资料、精确余额、采购价、服务器路径、堆栈、SQL 或完整原始响应。公开模型显示名、已公示参数限制、本网站任务号和请求排查编号保留。

上游原文“Video generation failed or was cancelled (request id: ...)”返回“Video generation failed or was cancelled”；只移除上游请求追踪编号，不另行翻译、扩写或推断原因。不能将其中的“或被取消”理解为已确认用户取消。原文中的地址等隐私用隐藏标记替换；涉及凭据、账户或内部配置而无法安全保留的内容，继续使用对应的站长原因说明。

没有可返回的具体原因时，显示“生成失败，暂无具体原因”。旧任务查询会根据已保存的诊断返回失败说明，不重发历史 Webhook，不修改成功失败、积分或退款。无法恢复未保存的上游原因。

客户端应展示 error.message，并结合 reason_code 和 retry_action 处理；不要自行将所有错误再改成“联系管理员”。Webhook 接收端同样读取 data.error。网络异常先查已有任务，避免重复付费。不提供任务列表、批量产物、任意上游参数透传、文本流式响应或充值 API。