

MiniMax H3 接入API文档

生产环境 · 文档版本 1.9 · 更新日期 2026-08-28

通过 Stargate 调用 MiniMax-H3。示例使用 768P / 5 秒 / 16:9；其他输出规格需先确认已开通。

1. 地址与鉴权

API Base URL: <https://console.tokenforu.com/v1>

所有接口均需携带 API Key；发送 JSON 请求体时添加 Content-Type: application/json。

```
HTTP
Authorization: Bearer <STARGATE_API_KEY>
```

API Key 由业务后端保管，仅发送到 Stargate 服务。查询、下载、取消和删除建议使用创建任务时的同一 Key；无权访问的任务返回 404。

接口列表

以下路径相对于 API Base URL，**不要重复拼接** /v1。

方法	路径	用途
GET	/models	查询当前 Key 可见的模型
POST	/videos	创建视频任务
GET	/videos/{id}	查询任务
GET	/videos?limit=20	查询最近任务
GET	/videos/{id}/content	下载视频
POST	/videos/{id}/cancel	请求取消任务
DELETE	/videos/{id}	删除任务或资源

2. 创建视频任务

视频生成采用异步方式：**创建任务** → **保存 id** → **查询状态** → **下载视频**。

使用前确认 API Key 已开通 MiniMax-H3。GET /models 可查询当前 Key 可见的模型 ID。

请求字段

字段	类型	要求	说明
model	string	必填	固定为 MiniMax-H3
prompt	string	必填	所有生成方式均需非空的视频描述
seconds	string / number	建议填写	整数秒，示例为 "5" 或 5
size	string	建议填写	分辨率档位，示例为 "768P"
input_reference	string	可选	单张首帧图片 URL，不接受数组
callback_url	string	可选	公网可访问的 HTTP(S) 回调地址，建议 HTTPS
metadata	object	可选	以下嵌套字段的容器；不可传 null
metadata.ratio	string	建议填写	数字画幅比例，示例为 "16:9"
metadata.content	array	使用素材时填写	图片、视频、音频元素，格式见下文

画幅使用 metadata.ratio，不支持 adaptive。时长和分辨率使用顶层 seconds、size，不要在 metadata 中重复设置。

接口接收 JSON，不接受本地文件的 multipart/form-data 上传。描述写入顶层 prompt，素材数组写入 metadata.content，不要改用顶层 content。

文生视频请求

将以下内容保存为 video-request.example.json：

```
JSON
{
  "model": "MiniMax-H3",
  "prompt": "电影感航拍，一艘帆船穿过清晨薄雾，柔和日光，镜头缓慢推进",
  "seconds": "5",
  "size": "768P",
  "metadata": {
    "ratio": "16:9"
  }
}
```

素材字段

metadata.content 中每个元素包含 type、role 和对应 URL：

用途	type	role	URL 字段
首帧	image_url	first_frame	image_url.url
尾帧	image_url	last_frame	image_url.url
参考图	image_url	reference_image	image_url.url
参考视频	video_url	reference_video	video_url.url
参考音频	audio_url	reference_audio	audio_url.url

- 单张首帧可直接用 input_reference；首尾帧各最多一张，可通过数组同时提交。首尾帧模式的输出画幅由输入图片决定。
- **首尾帧不能与任何 reference_* 素材混用。** 使用参考图、参考视频或参考音频时，不填写 input_reference。

- 仅使用 input_reference 时可省略 metadata.content；使用数组提交首帧时，不要再通过 input_reference 重复提交。

多参考图请求

保存为 video-request.multiref.example.json，将图片地址替换为上游可访问的真实图片 URL：

```
JSON
{
  "model": "MiniMax-H3",
  "prompt": "结合两张参考图生成一段风格一致、镜头连贯的视频",
  "seconds": "5",
  "size": "768P",
  "metadata": {
    "ratio": "16:9",
    "content": [
      {
        "type": "image_url", "role": "reference_image",
        "image_url": {"url": "https://example.com/1.jpg"}
      },
      {
        "type": "image_url", "role": "reference_image",
        "image_url": {"url": "https://example.com/2.jpg"}
      }
    ]
  }
}
```

参考视频与音频

在上面的 metadata.content 数组中追加所需元素；以下不是完整请求。

参考视频：

```
JSON
{
  "type": "video_url",
  "video_url": {"url": "https://example.com/reference.mp4"},
  "role": "reference_video"
}
```

参考音频：

```
JSON
{
  "type": "audio_url",
  "audio_url": {"url": "https://example.com/reference.mp3"},
  "role": "reference_audio"
}
```

素材要求

以下为上述输入方式的 [H3 素材限制](#)：

素材	数量与大小	格式与时长
参考图	最多 9 张，每张 ≤30 MB	JPEG/JPG、PNG、WEBP、HEIC、HEIF
参考视频	最多 3 个，每个 ≤50 MB	MP4/MOV；每个 2~15 秒，总时长 ≤15 秒
参考音频	最多 3 个，每个 ≤15 MB	WAV/MP3；每个 2~15 秒，总时长 ≤15 秒

图片及视频宽、高均为 256 ~ 5760 像素，宽高比为 0.4 ~ 2.5；首尾帧也遵守图片文件限制。视频需使用 H.264/H.265，帧率 23.976 ~ 60 fps，内含音轨使用 AAC/MP3。素材 URL 必须在生成期间持续可访问，不附加 Stargate API Key。

提交请求

以下命令使用 Bash 语法，API Key 从环境变量读取：

```
BASH
export STARGATE_BASE_URL='https://console.tokenforu.com/v1'
read -r -s -p 'Stargate API Key: ' STARGATE_API_KEY
export STARGATE_API_KEY

curl --fail-with-body -sS -X POST "$STARGATE_BASE_URL/videos" \
  -H "Authorization: Bearer $STARGATE_API_KEY" \
  -H "Content-Type: application/json" \
  --data-binary @video-request.example.json
```

多参考图请求使用 @video-request.multiref.example.json；其余接口调用不变。

受理后返回 HTTP 202 Accepted：

```
JSON
{
  "id": "task_example",
  "object": "video",
  "model": "MiniMax-H3",
  "status": "queued",
  "progress": 0,
  "created_at": 1787792400,
  "seconds": "5",
  "size": "768P"
}
```

保存返回的真实 id，替换后文的 task_example。202 只表示请求已受理；创建响应也可能直接返回终态，按实际 status 处理。

3. 查询任务

```
BASH
export TASK_ID='task_example'

curl --fail-with-body -sS "$STARGATE_BASE_URL/videos/$TASK_ID" \
  -H "Authorization: Bearer $STARGATE_API_KEY"
```

status	含义	处理
queued	排队中	继续查询
in_progress	生成中	继续查询
unknown	状态暂不确定	保留 ID，稍后查询
completed	已完成	下载视频
failed	失败	读取 error.code、error.message
cancelled	已取消	结束查询
expired	任务已过期	结束查询

完成响应示例：

```
JSON
{
  "id": "task_example",
  "object": "video",
  "model": "MiniMax-H3",
  "status": "completed",
  "progress": 100,
  "created_at": 1787792400,
  "completed_at": 1787792460,
  "result_url": "https://console.tokenforu.com/v1/videos/task_example/content"
}
```

- 以 status 判断生成结果；HTTP 200 或 progress=100 不代表生成成功。
- created_at、completed_at 为 Unix 秒时间戳；结束时间也可用于失败、取消等终态。
- seconds、size 可在创建响应回显，后续查询不保证返回；自行保存原始请求参数。
- 建议每 5 秒查询一次。本地等待超时不会取消任务，保留原 ID 可继续查询。

最近任务： GET /videos?limit=20 返回 object="list" 和数组 data，按最新任务优先排序。limit 默认 20，接受 1~500，实际最多返回 100 条，不提供分页。列表包含当前 API Key 访问范围内的视频任务，客户端按 model="MiniMax-H3" 筛选；跟踪进度使用单任务查询。

4. 下载视频

任务为 completed 后调用：

```
BASH
curl --fail-with-body -sS "$STARGATE_BASE_URL/videos/$TASK_ID/content" \
  -H "Authorization: Bearer $STARGATE_API_KEY" \
  -o output.mp4
```

响应为视频二进制；文件格式以实际 Content-Type 为准。result_url 也需要鉴权，若为以 / 开头的路径，按 <https://console.tokenforu.com> 解析。

- 支持传入 Range、If-Range；返回 206 表示分段内容，不能当作完整文件。
- 下载失败时不要将输出文件当作有效视频；完成后及时转存，下载地址不代表永久保存。

5. 取消与删除

取消： POST /videos/{id}/cancel，无需请求体。HTTP 200 返回当前任务对象；以 status=cancelled 确认取消完成，若仍在进行中则继续查询。已进入终态的任务返回原状态。

删除： DELETE /videos/{id}，行为取决于任务状态：

状态	删除行为
queued	刷新状态后仍排队时尝试取消，结果以实际任务状态为准
completed、failed、expired	处理资源删除，并移除可访问的任务记录
in_progress、cancelled	返回 409 task_delete_not_supported

删除成功返回：

```
JSON
{
  "id": "task_example",
  "object": "video.deleted",
  "deleted": true
}
```

排队任务删除后仍可能以 cancelled 状态查询到。

6. 终态回调

创建时设置 callback_url 后，任务结束会发送 JSON POST 通知。完成通知示例：

```
JSON
{
  "id": "task_example",
  "object": "video",
  "model": "MiniMax-H3",
  "status": "completed",
  "progress": 100,
  "result_url": "https://console.tokenforu.com/v1/videos/task_example/content",
  "result_urls": [
    "https://console.tokenforu.com/v1/videos/task_example/content"
  ]
}
```

失败、取消、过期通知使用相应状态，有错误时附带 error.code、error.message。

接收端应：

1. 确认 id 属于自己已保存的任务，可靠入队后在 10 秒内返回 2xx。
2. 使用自己的 API Key 查询该任务，以查询结果驱动下载和交付，不直接信任通知中的状态或任意 URL。
3. 按任务 ID 幂等处理，避免重复交付；保留轮询作为回调未到达时的兜底。

回调可能重复或延迟，失败后的重试次数由服务配置决定。

7. 错误与重试

错误响应示例：

```
JSON
{
  "error": {
    "type": "invalid_request_error",
    "code": "model_not_found",
    "message": "The model is not available for this token."
  }
}
```

HTTP	常见错误	处理
400	invalid_json、model_missing、invalid_metadata、invalid_limit	修正请求参数
400	video_request_invalid	参数未通过校验，如 ratio=adaptive

HTTP	常见错误	处理
400	video_task_billing_unconfigured	核对模型和参数，仍报错时联系平台
401	鉴权错误	检查 Key 是否有效
403	权限、认证或 insufficient_quota	检查账号、授权和额度
404	model_not_found、task_not_found	核对模型 / 任务 ID 及访问权限
409	video_not_ready、task_missing_upstream_id	保留任务 ID，稍后查询
429	限流	查询请求退避后重试
5xx	服务或内容获取错误	查询、下载可有限重试；保留错误信息

部分响应没有 `error.code`，应同时处理 HTTP 状态码和 `error.message`。

创建请求不要自动重试：当前接口不提供可依赖的 Idempotency-Key 去重保证。已有 ID 时继续查询原任务；创建超时或断连且没有 ID 时，先确认原请求是否受理，再决定是否重提。创建响应含 `metadata.recovery_status` 等待确认信息时，同样保留原 ID 查询。

查询遇到网络超时、响应中断、HTTP 408、429 或 5xx 可退避后重试，不能通过重新创建来恢复任务。

价格与额度按开通约定执行；视频任务对象不返回结算金额，取消或删除不代表免除已产生的费用。

8. Python 示例

分发包附带 `video_client.py`、文生视频请求 `video-request.example.json` 和多参考图请求 `video-request.multiref.example.json`，使用 Python 3.10+ 标准库，无需安装依赖。

设置 `STARGATE_API_KEY` 后运行，随附请求文件已填写上述 MiniMax-H3 参数：

```
BASH
python video_client.py --request video-request.example.json --state task-001.json --output video-001.mp4
```

多参考图调用将 `--request` 的文件名改为 `video-request.multiref.example.json`。

恢复已有任务，不重复创建：

```
BASH
python video_client.py --task-id task_example --state task-001.json --output video-001.mp4
```

Windows PowerShell 7.1+ 可先隐藏输入 Key，再执行上述 Python 命令：

```
POWERSHELL
$env:STARGATE_API_KEY = Read-Host 'Stargate API Key' -MaskInput
```

每个新任务使用独立的状态文件与输出路径。默认查询间隔 5 秒、轮询等待预算 30 分钟，可通过 `--interval`、`--max-wait` 调整；创建和下载另有请求超时。无 ID 且提交结果不明确时，先确认原请求，不要删除状态文件后直接重提。