



咚文AI

DONGWEN AI HUB PLATFORM

<https://www.dwen.cn>

完全开放 · 免费使用

咚文AI

面向开发者的大模型调用聚合平台

一个地址，一个 Key，直连 15 家大模型厂商

完全开放、免费使用，为个人与企业开发者省去重复对接的麻烦

对话 · 多模态

语音转写

文本向量化

文生图

文生视频

图生视频

本地模型



这份材料讲什么

从「为什么做」到「怎么用」，再落到「已经能用的功能」。

01

背景与定位

多厂商接入的重复劳动，以及咚文AI 的答案

02

平台架构与能力全景

三层结构 · 请求全流程 · 六类大模型能力

03

厂商与模型矩阵

15 家厂商 · 六类能力 · 模型广场

04

开发者接入

四步接入 · 多语言示例 · 错误码与 FAQ

05

控制台与演进路线

8 个管理页面 · 免费策略 · 后续规划

多厂商接入，是每个开发者的重复劳动

12+

API 格式各不相同

OpenAI 之外，Anthropic 用 x-api-key、Google 走 URL 参数，签名与报文都要重新实现一遍

N

密钥四处散落

每个厂商一个控制台、一套配额策略，轮换与吊销全靠人工记忆，安全审计无从谈起

\$\$

计费规则不统一

按 Token、按次、按并发各不相同，成本核算与限额管控难以落地到业务侧

?

故障难定位

调用失败时，到底是网络、密钥、限流还是模型名写错，没有统一日志可查



咚文AI 把这些重复劳动收敛成一次：统一协议、统一密钥、统一日志、统一观测。

两种用法都免费，按你的需要挑一种

模式 A

公共额度 · 开箱即用

适合个人开发者、学习与快速验证

- **零配置**
注册后直接拿到 ymk- 接入 Key，无需自备厂商账号
- **一个地址**
<https://api.dwen.cn/v1>，所有模型统一入口
- **免费额度**
平台承担上游成本，按 Key 统计用量，防滥用限速
- **覆盖全能力**
对话 / 多模态 / 转写 / 向量化 / 图片 / 视频
- **永久免费**
不充值、不绑卡、不设付费墙

适合：写 Demo、做课程作业、验证想法、小工具

模式 B

自建渠道 · 完全自主

适合企业团队、私有模型与合规场景

- **BYOK**
填入自己的厂商 API Key，费用与配额自己掌控
- **多账号并行**
同一厂商可配多个渠道，按项目隔离
- **预算与限流**
厂商侧的额度、账单、白名单策略完全自主
- **本地模型**
接 Ollama / vLLM / LM Studio，数据不出内网
- **密钥安全**
AES-GCM 加密落库，前端永不回显明文

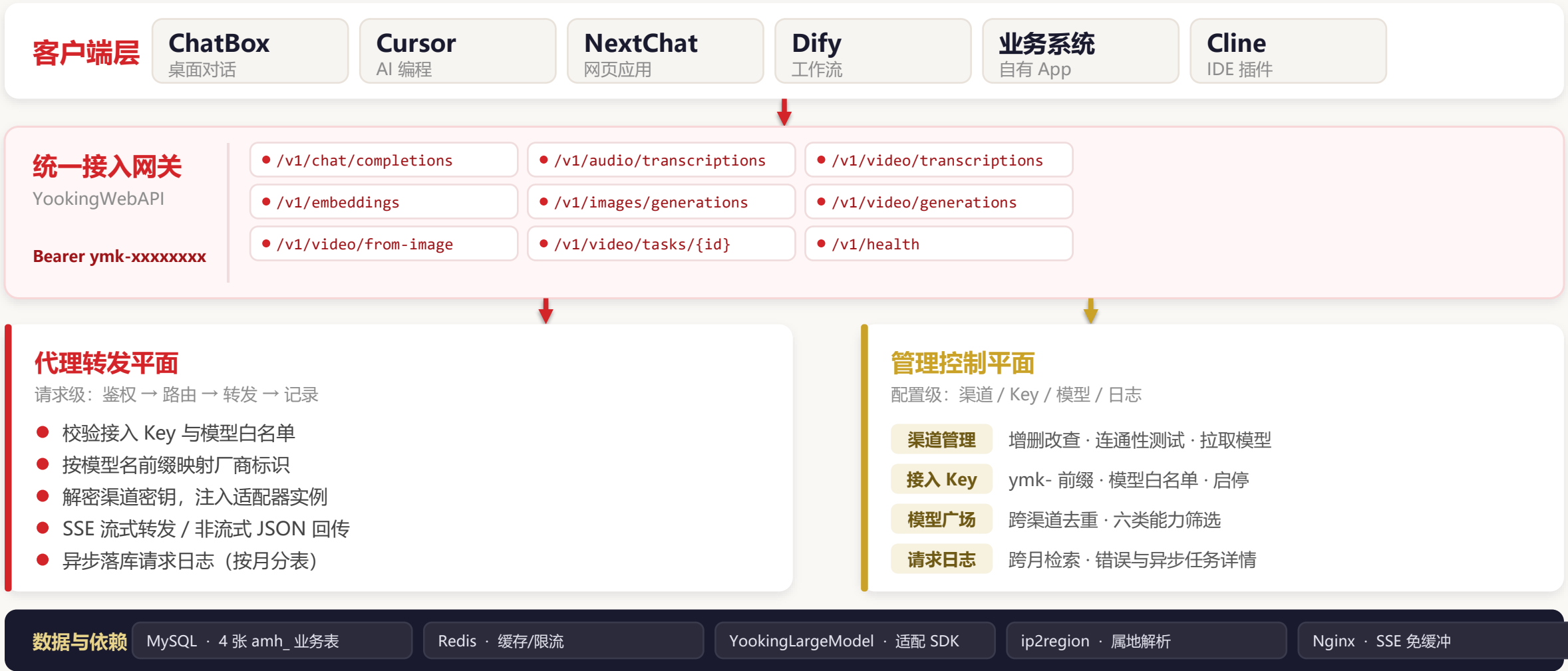
适合：生产系统、企业内网、数据合规、成本核算

或

两者共用同一套 OpenAI 兼容接口与同一份请求日志 —— 先用模式 A 跑通，需要时切到模式 B，业务代码零改动。

切换方式：在「上游渠道」里配置自己的 Key 即可，平台 Key 与渠道配置互不影响。

三层结构：客户端 / 统一网关 / 双平面



一次调用的六个环节，全程透明转发



请求 Request

```
POST /v1/chat/completions HTTP/1.1
Host: www.dwen.cn
Authorization: Bearer ymk-d23bcf9f...
Content-Type: application/json
{ "model": "deepseek-v4-flash",
  "messages": [ { "role": "user", "content":
                  "帮我总结这份周报" } ],
  "stream": true }
```

响应 Response (SSE 流式)

```
HTTP/1.1 200 OK
Content-Type: text/event-stream
data: {"choices":[{"delta":{"content":"本"}}]}
data: {"choices":[{"delta":{"content":"周"}}]}
data: {"choices":[{"delta":{"content":"已完成"}}]}
data: [DONE]
1,247 tokens · 320ms · ttft 86ms
```

一个接口，六类大模型能力开箱可用



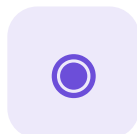
对话

Chat

01

文本生成 · 流式输出

Function Calling · JSON Mode



多模态理解

Vision

02

图片 / 音频 / 视频理解

统一走 messages content



语音转写

ASR

03

音视频转文字

发言人区分 · 时间戳分段



文本向量化

Embedding

04

RAG 检索 · 语义搜索

分类聚类 · 8 家厂商



文生图

Text→Image

05

DALL-E 3 / CogView-4

万相 / FLUX / Kolors



视频生成

Text/Image→Video

06

文生视频 · 图生视频

异步任务 + 状态轮询

全部能力统一遵循 OpenAI 兼容协议 —— 现有 SDK、框架、客户端几乎零改造即可接入。

15 家厂商 × 六类能力，一表看清实现边界

下面这张表来自 SDK 的适配器清单与上游协议差异分析，是全部能力的实现边界。

厂商	对话	视觉理解	语音转写	向量化	文生图	视频生成
OpenAI（美国）	✓	✓	—	✓	DALL-E 3	—
Anthropic（美国）	✓	✓	—	—	—	—
Google（美国）	✓	✓	—	—	—	—
DeepSeek 深度求索	✓	—	—	✓	—	—
通义千问（阿里云）	✓	✓	✓	✓	万相	万相 T2V/I2V
智谱 AI（GLM）	✓	✓	—	✓	CogView-4	—
Moonshot（Kimi）	✓	—	—	—	—	—
硅基流动	✓	✓	—	✓	FLUX/Kolors	—
豆包（火山引擎）	✓	✓	—	✓	豆包图像	Seedance T2V/I2V
零一万物（Yi）	✓	✓	—	✓	—	—
MiniMax（海螺）	✓	✓	—	—	—	—
小米 MiMo	✓	—	—	—	—	—
Ollama（本地部署）	✓	—	—	✓	—	—
自定义 OpenAI 兼容	✓	按需	—	按需	—	—
咚文AI 统一出口	✓	✓	✓	✓	✓	✓

「自定义 OpenAI 兼容」指任何遵守 /chat/completions 规范的服务，改 Base URL 即可接入。

流式输出 + 工具调用，完整对齐 OpenAI 协议

客户端

ChatBox / SDK

POST /v1/chat/completions

Bearer ymk-... model messages

咚文AI 网关

鉴权 · 路由 · 转发

校验 Key → 前缀映射 → 取渠道

AES-GCM 解密上游密钥

上游厂商

15 家适配器

建立 SSE 长连接

厂商流式返回增量 token

流式输出 (SSE) 实时打字效果

咚 文 AI 已 为 你 完 成 聚 合

```
data: {"choices":[{"delta":{"content":"咚"}}]}
```

```
data: {"choices":[{"delta":{"content":"文"}}]}
```

```
data: [DONE]    首字 86ms · 总耗时 320ms
```

开箱可用的对话特性

- **流式 / 非流式**
stream=true 走 SSE, false 返 JSON
- **JSON Mode**
结构化输出, 便于业务解析
- **多轮上下文**
messages 数组原样转发, 无状态
- **Function Calling**
工具定义与调用结果完整透传
- **思维链与缓存**
推理型模型的 reasoning / cached token
- **模型前缀路由**
改 model 名即切换厂商, 无需改代码

图片 / 音频 / 视频，统一走 messages content



多模态请求体

```
{ "model": "qwen-vl-max",  
  "messages": [ { "role": "user",  
    "content": [  
      { "type": "text", "text": "这张图讲了什么? " },  
      { "type": "image_url",  
        "image_url": { "url": "https://.../01.jpg" } } ] } ] }
```

支持的输入形态

- 图片** 公网 URL / Base64 · png · jpg · webp · gif
openai / qwen / glm / doubao ...
- 音频** 公网 URL · mp3 · wav · m4a · flac
qwen-audio 系列
- 视频** 公网 URL · mp4 · mov · mkv · avi
qwen-vl 系列

音视频转文字，自带发言人区分与时间戳

输入

- 音频** /v1/audio/transcriptions
mp3 / wav / m4a / flac / ogg
- 视频** /v1/video/transcriptions
mp4 / mkv / avi / mov (自动抽音轨)

处理链路



异步任务：提交 → 轮询 → 下载结果；结果含时间戳分段与发言人标识

转写结果（含时间戳与发言人）

diarization_enabled = true

- S1** 00:00 - 00:07 大家好，今天我们来评审咚文AI 聚合平台的接入方案。
- S2** 00:07 - 00:16 我关注的是渠道侧：十五家厂商的密钥怎么统一托管和轮换？
- S1** 00:16 - 00:25 平台按用户维度加密存储，AES-GCM，前端永不回显。
- S3** 00:25 - 00:34 那限额和成本核算呢？我们希望按项目做隔离。
- S2** 00:34 - 00:42 可以用接入 Key + 模型白名单来切分，日志里都带 Key 维度。

文本转向量，为 RAG 与语义检索打底

语义相近 → 向量相近



典型用途与支持范围

RAG 知识库

把企业文档切片向量化，检索时召回最相关片段

语义搜索

摆脱关键词匹配，按意思找内容

文本分类聚类

零样本归类工单、评论、舆情

去重与推荐

用余弦相似度判重、找相似内容

八家厂商开箱可用

OpenAI

DeepSeek

通义千问

智谱 GLM

硅基流动

豆包

零一万物

Ollama

5 家厂商，同步与异步协议都抹平

POST /v1/images/generations, 请求体与 OpenAI 图像接口完全一致。

○ OpenAI DALL·E 3

标准 OpenAI 图像协议，同步返回图片 URL

○ 智谱 CogView-4 / 4-Plus

OpenAI 兼容协议，同步模式

○ 通义万相 wan2.7-image / Pro

DashScope 原生异步，必须带 X-DashScope-Async 头

○ 硅基流动 FLUX.1 / Kolors

OpenAI 兼容协议，聚合多家开源文生图模型

○ 豆包图像生成

火山方舟异步任务：提交 → 轮询 → 取图，指数退避 3s→15s

○ 统一参数

size / n / quality / negative_prompt / seed / watermark 自动映射

模型广场 · 13 家预置厂商



OpenAI
openai



Anthropic
anthropic



Google
google



DeepSeek
deepseek



通义千问
qwen



智谱 AI
glm



Moonshot
moonshot



硅基流动
siliconflow



豆包
doubao



零一万物
yi



MiniMax
minimax



小米 MiMo
mimo



Ollama
ollama

模型列表由厂商接口实时拉取，不维护静态注册表。

文生视频 / 图生视频, 异步任务 + 状态轮询

两种入口

T2V

文生视频

POST /v1/video/generations

prompt → video · wan2.7-t2v · seedance-2-0

I2V

图生视频

POST /v1/video/from-image

image_url + prompt → video · wan2.7-i2v · seedance-1-0-lite-i2v

异步任务时序

- 1 ① 提交任务
厂商返回 task_id, 日志记为 Pending
- 2 ② 轮询状态
GET /v1/video/tasks/{taskId}?model=...
- 3 ③ 生成完成
task_status = Succeeded, 返回 video_url
- 4 ④ 回写日志
更新 media_url / 状态码 / 耗时

任务状态流转

Pending 待处理

Running 处理中

Succeeded 已成功

Failed 已失败

Expired 已过期

Unknown 未知

分辨率

480p / 720p / 1080P, 适配各厂商大小写

时长与比例

1-60 秒; 16:9 / 9:16 / 1:1 / 4:3 / 3:4

提示词改写

万相支持 prompt_extend, 豆包支持原生音频

8 个页面覆盖接入全生命周期

Vue 3 + Element Plus 前端，通过 BizRouter 插件与后端通信，插件 DLL 热部署。

- 仪表盘

渠道状态、今日请求量 / 成功率 / 平均延迟、最近请求
- 上游渠道管理

增删改查、启停、连通性测试、从厂商拉取模型
- 接入 Key 管理

生成 ymk- 凭证、模型白名单、启停与吊销
- 可用模型

跨渠道去重，按厂商 / 六类能力 / 名称筛选
- 请求日志

跨月检索，错误详情与异步任务详情可展开
- 渠道商速查

13 家厂商的 Key 申请入口、Base URL、认证方式
- 开发者接入指南

5 步接入、9 个端点一览、多语言示例、错误码与 FAQ
- API 测试工具

8 个测试卡片，在浏览器里直接验证每个端点

1,247

今日请求量

99.2%

成功率

320ms

平均延迟

最近请求

实时刷新

21:30:12	gpt-4o	对话	1,250	320ms	✓
21:28:45	deepseek-v4-flash	对话	2,100	180ms	✓
21:25:03	fun-asr	音频转写	—	2.1s	✓
21:22:18	wan2.7-t2v	视频生成	—	8.5s	✓
21:20:17	qwen-vl-max	对话	1,100	280ms	✓
21:18:56	text-embedding-v3	向量化	5	95ms	✓
21:15:30	deepseek-v4-flash	对话	500	160ms	✗

统一托管各厂商 API Key

增删改查、启停、连通性测试、从厂商拉取模型；密钥 AES-GCM 加密落库。

名称	类型	能力	状态	延迟	模型数	操作
我的百炼	通义千问 (qwen)	视觉 · ASR · 文生图 · 视频	正常	220ms	38	编辑 拉取模型 禁用 删除
DeepSeek 官方	DeepSeek (deepseek)	—	正常	180ms	4	编辑 拉取模型 禁用 删除
OpenAI	OpenAI (openai)	视觉 · 文生图	正常	680ms	21	编辑 拉取模型 禁用 删除
Claude	Anthropic (anthropic)	视觉	正常	920ms	9	编辑 拉取模型 禁用 删除
豆包方舟	豆包 (doubao)	视觉 · 文生图 · 视频	正常	410ms	17	编辑 拉取模型 禁用 删除
智谱 GLM	智谱 AI (glm)	视觉 · 文生图	正常	350ms	14	编辑 拉取模型 禁用 删除
本地 Ollama	Ollama (ollama)	—	禁用	—	6	编辑 拉取模型 启用 删除

共 7 个渠道 ● 正常 6 ○ 禁用 1

渠道密钥以 AES-GCM 加密落库，编辑时前端不回显明文；「拉取模型」调用 SDK 的 ListModelsAsync 同步厂商模型。

接入 Key 管理

为你的应用分配 ymk- 前缀的接入凭证，可限制可调用的模型范围。

名称	Key	状态	模型限制	创建时间	操作
我的应用 Key	ymk-d23bcf9f****f6e9	启用	全部	2026-06-21 13:46	复制 禁用 删除
生产环境	ymk-7a41c0e8****21bd	启用	3 个模型	2026-06-28 09:12	复制 禁用 删除
测试用 Key	ymk-0f9e34ab****7c05	启用	全部	2026-07-02 15:38	复制 禁用 删除
移动端 App	ymk-b5d1e70c****8a44	禁用	全部	2026-07-05 11:20	复制 启用 删除
内部能力引擎	ymk-3c8fa216****d90e	启用	全部	2026-07-09 16:45	复制 禁用 删除

使用说明：把接入 Key 填入 AI 客户端（ChatBox、Cursor、NextChat、Dify...） ， API 地址填 <https://api.dwen.cn/v1> 即可。

Key 创建成功

请立即复制保存，关闭后无法再次查看完整 Key

ymk-d21bcf9fb2123c3657afea6b083f6e9

复制 Key

完整 Key 只在创建成功时展示一次，此后仅保留脱敏前缀，从机制上避免明文外泄。

请求日志

每一次调用的模型、端点、Token、延迟、状态与客户端属地全部留痕；按月分表存储。

时间范围

2026-07-01 ~ 2026-07-09

模型

全部模型

端点

全部端点

状态

全部状态

查询

时间	模型	端点	Token	延迟	状态	客户端 IP	属地	媒体信息
2026-07-09 21:30:12	gpt-4o	对话	1,250	320ms	✓	203.0.113.24	上海 上海	—
2026-07-09 21:28:45	deepseek-v4-flash	对话	2,100	180ms	✓	203.0.113.24	上海 上海	—
2026-07-09 21:25:03	fun-asr	音频转写	—	2.1s	✓	120.55.18.7	浙江 杭州	2m 14s
2026-07-09 21:22:18	wan2.7-t2v	视频生成	—	8.5s	✓	120.55.18.7	浙江 杭州	Succeeded
2026-07-09 21:20:17	qwen-vl-max	对话	1,100	280ms	✓	39.156.66.10	北京 北京	—
2026-07-09 21:18:56	text-embedding-v3	向量化	5	95ms	✓	39.156.66.10	北京 北京	—
2026-07-09 21:15:30	deepseek-v4-flash	对话	500	160ms	✗	112.65.12.88	广东 深圳	—

展开行 · 错误与异步任务详情

upstream_error: 502 Bad Gateway — api.deepseek.com 连接超时（重试 2 次后放弃）

可用模型（模型广场）

模型列表由各厂商接口实时拉取，跨渠道自动去重，支持按厂商 / 六类能力筛选。

厂商

全部厂商

类型

全部类型

搜索模型名称

查询

模型名称	厂商	类型	上下文	所属渠道	渠道状态
gpt-4o	OpenAI	对话	128K	OpenAI	正常
deepseek-v4-flash	DeepSeek	对话	64K	DeepSeek 官方	正常
qwen3.7-max	通义千问	对话	1M	我的百炼	正常
qwen-v1-max	通义千问	对话	32K	我的百炼	正常
fun-asr	通义千问	ASR	—	我的百炼	正常
text-embedding-v3	通义千问	Embedding	8K	我的百炼	正常
wan2.7-image	通义千问	文生图	—	我的百炼	正常
wan2.7-t2v	通义千问	文生视频	—	我的百炼	正常

共 128 个模型 对话 96 ASR 4 Embedding 28 文生图 5 文生视频 3 图生视频 2

开发者接入指南（控制台内置）

OpenAI 兼容协议 · 即插即用



端点	方法	说明	示例模型
POST /v1/chat/completions	POST	对话完成（含多模态理解）	deepseek-v4-flash
POST /v1/audio/transcriptions	POST	音频转写（ASR）+ 发言人区分	fun-asr
POST /v1/video/transcriptions	POST	视频转写（自动抽取音频轨）	fun-asr
POST /v1/embeddings	POST	文本向量化（Embedding）	text-embedding-v3
POST /v1/images/generations	POST	文生图（5 家厂商）	dall-e-3 / cogview-4
POST /v1/video/generations	POST	文生视频（异步任务）	wan2.7-t2v

curl 示例（复制即可运行）

```
curl https://api.dwen.cn/v1/chat/completions \
-H "Authorization: Bearer ymk-xxxxxxx" \
-H "Content-Type: application/json" \
-d '{"model": "deepseek-v4-flash",
  "messages": [{"role": "user", "content": "你好"}],
  "stream": false}'
```

接入要点

- Base URLhttps://api.dwen.cn/v1
- 认证方式Authorization: Bearer ymk-xxxxxxx
- 请求格式JSON (Content-Type: application/json)
- 流式响应SSE, 需带 Accept: text/event-stream
- 迁移成本OpenAI SDK 仅改 base_url, 业务代码零改动

四步接入，五分钟跑通第一个请求

01

注册并领取 Key

在控制台创建接入 Key,
拿到 ymk- 前缀凭证

```
访问 www.dwen.cn  
创建 Key 并复制
```

预计耗时
约 1 分钟

02

替换 Base URL

把 OpenAI SDK 的
base_url 指向平台

```
base_url="https://  
api.dwen.cn/v1"
```

预计耗时
约 1 分钟

03

挑选模型名

在「可用模型」里选一个
模型名填进 model 字段

```
model="deepseek-v4-flash"  
model="qwen-v1-max"
```

预计耗时
约 1 分钟

04

发起调用

原有 SDK 代码不变,
流式与非流式都支持

```
client.chat.completions  
.create(...)
```

预计耗时
约 1 分钟

Python: 从 OpenAI SDK 迁到咚文AI, 只改两行

```
from openai import OpenAI  
client = OpenAI(  
    base_url="https://api.dwen.cn/v1", # ← 改这里  
    api_key="ymk-xxxxxxxxxxxxxxxx", # ← 改这里  
)
```

```
resp = client.chat.completions.create(  
    model="deepseek-v4-flash",  
    messages=[{"role": "user",  
                "content": "你好, 咚文AI"}],  
    stream=True)
```

只改两行：从 OpenAI SDK 迁到咚文AI

平台完整兼容 OpenAI Chat Completions 的请求与响应格式，包括流式 SSE、Function Calling、JSON Mode 与多模态 content 数组。

① base_url

指向 <https://api.dwen.cn/v1>，其余保持不变

② api_key

使用控制台创建的 ymk- 接入 Key

流式输出

stream=True 即返回 text/event-stream，需带 Accept 头

模型切换

改 model 字段即可换厂商，前缀自动映射到对应适配器

```
Python · 对话（流式）

from openai import OpenAI

client = OpenAI(
    base_url="https://api.dwen.cn/v1",
    api_key="ymk-d23bcf9fb2123c30...",
)

stream = client.chat.completions.create(
    model="deepseek-v4-flash",
    messages=[{"role": "user",
                "content": "介绍一下咚文AI"}],
    stream=True)

for chunk in stream:
    delta = chunk.choices[0].delta.content
    if delta:
        print(delta, end="", flush=True)
```

九个代理端点 · 全部兼容 OpenAI 协议

认证统一为 Authorization: Bearer ymk-xxxxxxx; 流式响应统一为 SSE，无需认证。

端点	方法	能力	说明
/v1/chat/completions	POST	对话 / 多模态	文本生成、图片/音频/视频理解，流式 SSE + 非流式 JSON
/v1/audio/transcriptions	POST	语音转写	mp3 / wav / m4a / flac / ogg，支持发言人区分与时间戳分段
/v1/video/transcriptions	POST	语音转写	mp4 / mkv / avi / mov，服务端自动抽取音频轨后转写
/v1/embeddings	POST	向量化	单条或批量输入，8 家厂商，可指定输出维度
/v1/images/generations	POST	文生图	5 家厂商，同步与异步协议由适配器统一抹平
/v1/video/generations	POST	文生视频	异步任务：提交后返回 task_id
/v1/video/from-image	POST	图生视频	图片 + 提示词 → 视频，异步任务模式
/v1/video/tasks/{taskId}	GET	任务查询	轮询视频生成任务状态，返回 task_status 与 video_url
/v1/health	GET	健康检查	返回支持的对话 / 向量化 / 文生图 / 视频厂商清单，无需认证

视频生成采用「提交—轮询」异步模式；其余端点均为同步或流式实时返回。

常见失败原因与对应排查路径

每次调用都会写入请求日志，包含状态码、错误信息、延迟与客户端属地。

状态码	错误类型	含义	排查建议
400	invalid_request	缺少必要参数（model / messages / input）	检查请求体 JSON 格式与必填字段
401	invalid_api_key	接入 Key 无效、已禁用或不存在	在「接入 Key」页确认状态为启用
404	model_not_found	模型名在当前渠道中不存在	在「可用模型」页确认模型名拼写，或先拉取模型
429	rate_limit_exceeded	上游厂商限流或账户额度不足	检查厂商账户余额与配额，或切换渠道
500	internal_error	服务端内部错误或上游返回异常	在「请求日志」展开查看 error_msg
502	upstream_error	上游渠道不可达	检查渠道 Base URL 与网络连通性
504	timeout	请求超时（默认 60 秒，Ollama 300 秒）	换用更快的模型，或在渠道高级选项调大超时

所有错误响应均为 OpenAI 兼容的错误 JSON，可直接被现有 SDK 解析。

面向个人与企业开发者，完全开放、永久免费

¥0

使用费用

不充值、不绑卡、不设付费墙

15 家

厂商通道

一个地址直连，无需逐个注册

9 个

标准端点

六类大模型能力统一出口

2 种

使用模式

公共额度或自建渠道，随时切换

两种模式，按需选择

模式 A · 公共额度

注册即得 Key，平台承担上游成本，按 Key 限速

模式 B · 自建渠道

填入自己的厂商 Key，费用与配额完全自主

三点承诺

● 不改能力边界

文档写明支持的模型与端点，线上完全一致

● 不藏日志

每一次调用都能在请求日志里查到状态与耗时

● 不做静默降级

容量不足时明确报错，不会偷偷换成小模型

免费不等于无边界：公共额度有速率与额度上限，重负载场景建议切换到自建渠道模式。

已经能用的，和正在做的



路线图以「开发者用得上」为唯一取舍标准 —— 每一项能力都必须先在这套系统里自用跑通，才对外开源免费。

咚文AI 解决的是一件事：把「接模型」变成「填地址」

1 个

接入地址

api.dwen.cn/v1, 兼容 OpenAI

1 个

接入凭证

ymk- 前缀, 可限制可用模型

0 行

业务改造

现有 SDK 与框架零改动

∞

调用方式

流式、非流式、异步任务全部支持

- 对个人开发者

省下注册与对接十几家厂商的时间, 把精力放回业务本身

- 对企业团队

统一密钥托管、统一日志、统一配额, 满足审计与合规要求

- 对本地部署

Ollama / vLLM 直接接入, 数据不出内网

- 下一步

访问 www.dwen.cn 创建第一个接入 Key, 五分钟跑通第一个请求

官网: <https://www.dwen.cn> • 接入地址: <https://api.dwen.cn/v1>



咚文AI

DONGWEN AI HUB PLATFORM

一个地址，接入所有大模型

完全开放 · 免费使用 · 为开发者省去重复对接

官方网站

<https://www.dwen.cn>

Ai Hub → 控制台

接入地址

<https://api.dwen.cn/v1>

Authorization: Bearer ymk-.....

支持能力

对话 · 多模态 · 转写

向量化 · 文生图 · 视频生成



扫码访问官网

咚文AI · 大模型调用聚合平台 · 面向个人与企业开发者免费开放

让每一次模型调用，都只是一行 base_url 的距离