

# 大模型服务平台 TokenHub

## 模型列表



腾讯云

## 【 版权声明 】

©2013–2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

## 【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

## 【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

## 【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

# 模型列表

最近更新时间：2026-05-07 16:31:42

## 语言模型

模型名称	model (调用参数)	能力支持	上下文窗口 (Token)	最大输入 (Token)	最大输出 (Token)
Hy3 preview	hy3-preview	<ul style="list-style-type: none"><li>深度思考 (交错式思考)</li><li>结构化输出</li><li>Function Calling</li><li>Cache 缓存</li></ul>	256k	192k	128k
HY 2.0 Think	hunyuan-2.0-thinking-20251109	<ul style="list-style-type: none"><li>深度思考</li><li>Function Calling</li></ul>	192k	128k	64k
HY 2.0 Instruct	hunyuan-2.0-instruct-20251111	<ul style="list-style-type: none"><li>Function Calling</li></ul>	144k	128k	16k
Hunyuan-role	hunyuan-role-latest	角色扮演模型 适用 AI 数字分身、AI 角色扮演、AI 情感陪聊等场景	32k	28k	4k
DeepSeek-V4-Flash	deepseek-v4-flash	<ul style="list-style-type: none"><li>深度思考</li><li>结构化输出</li><li>Function Calling</li><li>Cache 缓存</li></ul>	1M	1M	384k
DeepSeek-V4-Pro	deepseek-v4-pro	<ul style="list-style-type: none"><li>深度思考</li><li>结构化输出</li><li>Function Calling</li><li>Cache 缓存</li></ul>	1M	1M	384k

Deepseek-v3.2	deepseek-v3.2	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> </ul>	128k	96k	32k
Deepseek-v3.1	deepseek-v3.1-terminus	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> </ul>	128k	96k	32k
Deepseek-r1-0528	deepseek-r1-0528	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> </ul>	128k	96k	16k
Deepseek-v3-0324	deepseek-v3-0324	<ul style="list-style-type: none"> <li>● Function Calling</li> </ul>	128k	128k	16k
GLM-5.1	glm-5.1	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k
GLM-5V-Turbo	glm-5v-turbo	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k
GLM-5-Turbo	glm-5-turbo	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k
GLM-5	glm-5	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k

Kimi-K2.6	kimi-k2.6	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	256k	256k	256k
Kimi-K2.5	kimi-k2.5	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● 结构化输出</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	256k	256k	256k
MiniMax-M2.7	minimax-m2.7	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k
MiniMax-M2.5	minimax-m2.5	<ul style="list-style-type: none"> <li>● 深度思考</li> <li>● Function Calling</li> <li>● Cache 缓存</li> </ul>	200k	200k	128k

## 视觉模型

### 图像生成

模型名称	model (调用参数)	模型介绍	任务类型	默认并发数
HY-Image-V3.0	hy-image-v3.0	基于混元大模型，能够去思考图像的布局、构图、笔触，利用世界知识去推理常识性的画面。同时可以解析千字级别的复杂语义，生成长文本文字、复杂漫画、表情包，还能生成生动有趣的科普插画。	<ul style="list-style-type: none"> <li>● 文生图</li> <li>● 图生图</li> </ul>	1
HY-Image-Lite	hy-image-lite	采用超高压压缩编解码器，实现图像生成快速响应与高品质输出。支持电商商品图美化、设计工具素材生成、游戏场景迭代等场景。	文生图	1

### 视频生成

模型名称	model (调用参数)	模型介绍	任务类型	默认并发数
HY-Video-1.5	hy-video-1.5	支持文本、图像多模态输入生成高清视频，可实现场景切换与多角色交互，简化制作流程、降低成本，应用于企业广告营销与个人创意落地场景。	<ul style="list-style-type: none"> <li>文生视频</li> <li>图生视频</li> </ul>	5
YT-Video-2.0	yt-video-2.0	支持生成动态连贯性高、画面过渡自然的视频，适用于对质量要求较高的广告创意、影视片段和产品展示等场景。	图生视频	5
YT-Video-HumanActor	yt-video-humanactor	单张参考照片即可驱动生成动态人像视频，精准还原表情、姿态，支持写实、二次元等多风格切换。	图生视频	5
YT-Video-FX	yt-video-fx	通过上传图片 and 选择特效模板，生成一段特效视频，将静态图像转化为充满活力、动感、有趣的视频画面。	图生视频	5

### 3D 生成

模型名称	model (调用参数)	模型介绍	任务类型	默认并发数
HY-3D-3.0	hy-3d-3.0	采用混元生3D 3.0模型，可生成更高精度以及更高质量的3D 模型，支持文生3D、图生3D、多视图生3D、单几何生成（白模），草图生3D、智能拓扑生3D功能。	<ul style="list-style-type: none"> <li>文生3D</li> <li>图生3D</li> </ul>	3
HY-3D-3.1	hy-3d-3.1	采用混元生3D 3.1模型，可生成更高精度以及更高质量的3D 模型，支持文生3D、图生3D、八视图生3D、单几何生成（白模）功能。	<ul style="list-style-type: none"> <li>文生3D</li> <li>图生3D</li> </ul>	3
HY-3D-Express	hy-3d-express	采用混元生3D 极速版模型，可将生成模型时间缩短至1分30秒内，可在较短时间内生成3D 模型文件。	<ul style="list-style-type: none"> <li>文生3D</li> <li>图生3D</li> </ul>	1

### 多模态理解模型

模型名称	model (调用参数)	模型介绍	上下文窗口 (Token)	最大输入 (Token)	最大输出 (Token)
YT-VITA	youtu-vita	VITA 是一款多模态理解模型，支持对视频和图片内容进行分析，可用于视频结构解析、图像目标检测等场景。	128k	100k	15k

## 能力说明

### 深度思考

模型在生成最终回答前，先进行内部思维链 (Chain-of-Thought) 推理，通过逐步分析和拆解问题，提升复杂任务 (如数学、逻辑推理、代码生成等) 的回答准确性。

### 联网搜索

模型支持在推理过程中访问互联网，检索并整合实时信息，从而为时效性问题 (如新闻、天气、最新数据等) 提供更准确的回答。

### 结构化输出

模型支持按照指定的格式 (如 JSON Schema) 输出结构化数据，便于下游程序直接解析和使用，适用于信息抽取、数据填充、API 响应构建等场景。

## Function Calling

模型支持函数调用能力，可在推理过程中根据用户意图自动识别并触发预定义的外部工具或 API，实现查询数据库、调用第三方服务等扩展操作。

### Cache 缓存

模型 Cache 缓存能力可复用历史请求中的上下文计算结果，减少重复计算开销，从而提升响应速度并降低调用成本。