

大模型服务平台 TokenHub

产品简介



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。

产品简介

最近更新时间：2026-04-17 23:32:41

什么是大模型服务平台 TokenHub?

大模型服务平台 TokenHub 是一站式 AI 大模型服务平台，面向企业和开发者提供统一的模型接入能力。平台聚合了腾讯混元、DeepSeek、MiniMax、Kimi、智谱 GLM、通义千问等多家行业领先的大模型，覆盖文本生成、图片生成、视频生成、3D 生成等多种 AI 能力场景。

无论您是希望快速集成 AI 能力的应用开发者，还是需要稳定可靠的模型推理服务的企业用户，大模型服务平台 TokenHub 都为您提供了一站式解决方案：**统一的 API 接口、灵活的计费模式、完善的监控体系以及便捷的在线体验**，帮助您以较低的接入成本获得优质的大模型服务。

核心功能

功能	说明
模型广场	汇聚多家厂商的主力大模型，提供多维度筛选（类型、服务商、体验状态）、模型对比和详细信息查看，帮助您快速选择合适的模型使用。平台提供了部分模型的免费体验包，供用户免费试用平台部分模型。
体验中心	在线体验大模型能力，支持文本对话（深度思考、流式输出）、图片生成（多风格、多尺寸）、视频生成以及 3D 生成，便于您在接入 API 之前充分评估模型效果。
AI 创作	视频特效中心展示了海量创意视频特效模板，支持一键复刻爆款同款和 API 灵活调用，方便根据个人需求选择、体验和调用特效。
在线推理	创建和管理模型的推理服务实例，获取专属 API Endpoint。支持按需选择计费方式——免费体验、按 Token 计费，满足从测试验证到生产部署的全阶段需求。
模型监控	展示模型、服务性能相关指标，如：TTFT (Time To First Token)、TPOT (Time Per Output Token)、RPM (Request Per Minute)。
用量统计	展示模型、服务性能计费用量指标，如：输入 Token、输出 Token、TPM (Tokens Per Minute)、插件调用次数。
API Key 管理	集中管理 API 访问密钥，支持精细化的权限控制（全部模型及服务或限定范围），便捷地启停切换和调用统计，保障接口调用的安全性与可控性。
Token Plan	是面向龙虾和编程场景设计的专属订阅套餐，覆盖腾讯混元、MiniMax、GLM、Kimi 等国产主流模型，兼容热门龙虾工具和主流编程工具，内设更多套餐档位可供选择。
Coding Plan	面向开发者的 AI 编程工具套餐，提供 Lite 和 Pro 两档选择，支持 OpenClaw、Cursor 等常见 AI 工具集成，提升开发效率。

平台支持的模型

- 平台支持的模型清单，请参见 [模型列表](#)。
- 模型计费方式，请参见 [计费方式](#)。

快速开始使用

您可以通过 [快速入门](#) 文档了解如何领取体验包、创建服务并调用 API。