

腾讯混元大模型

动态与公告



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。

文档目录

动态与公告

产品动态

旧版本模型下线迁移公告

动态与公告

产品动态

最近更新时间：2026-04-27 14:35:11

⚠ 注意：

为进一步提升大模型服务体验，腾讯混元大模型相关功能将逐步迁移至 [TokenHub](#)。迁移后，原平台将不再新增模型能力，并停止支持新购模型服务。用户已购买的模型服务可继续使用，暂不受影响。如需开通新的模型服务或使用更多模型能力，请前往 [TokenHub](#)。

2025年12月

动态名称	动态描述	发布时间	相关文档
Tencent HY Vision 1.5 Instruct 上线	基于文本 TurboS 基座生产的图生文快思考模型，相比上一版本在图像基础识别、图像分析推理等维度都有明显的效果提升。	2025-12-17	查看详情

2025年11月

动态名称	动态描述	发布时间	相关文档
Tencent HY 2.0 Think 上线	发布特性： 1. 模型底座从 TurboS 升级为混元2.0，模型能力全面提升。 2. 显著增强模型的复杂指令遵循、多轮及长文理解、代码、Agent、推理能力。	2025-11-09	查看详情
Tencent HY 2.0 Instruct 上线	发布特性： 1. 模型底座从 TurboS 升级为混元2.0，模型能力全面提升。 2. 显著增强模型的指令遵循、多轮及长文理解、文学创作、知识准确性、代码及推理能力。	2025-11-11	查看详情

2025年10月

动态名称	动态描述	发布时间	相关文档
hunyuan-translation 模型版本更新	支持语种齐全，33种语种互译和5种民族语言互译。同尺寸模型中效果最优，WMT25比赛30种语言获得第一，开源测试集 Flores200效果领先。	2025-10-14	查看详情

支持语种：中文，英语，法语，葡萄牙语，西班牙语，日语，土耳其语，俄语，阿拉伯语，韩语，泰语，意大利语，德语，越南语，马来语，印尼语，菲律宾语，印地语，中文繁体，波兰语，捷克语，荷兰语，高棉语，缅甸语，波斯语，古吉拉特语，乌尔都语，泰卢固语，马拉地语，希伯来语，孟加拉语，泰米尔语，乌克兰语，藏语，哈萨克语，蒙古语，维吾尔语，粤语。

2025年09月

动态名称	动态描述	发布时间	相关文档
hunyuan-large-role-latest 上线	1. 适用场景：AI 数字分身、AI 角色扮演、AI 情感陪聊等。 2. 特性说明：显著提升了角色一致性与对话深度。通过在大规模高质量角色对话数据上的强化训练，模型能深度理解并稳定维持角色设定，有效减少 OOC（脱离角色）问题。不仅在多轮互动中保持上下文连贯，更大幅提升了聊天的趣味性和沉浸感，使每次对话都生动而富有深度。	2025-09-24	查看详情
hunyuan-t1-vision-20250916	混元最新版 t1-vision 视觉深度思考模型，相比上一版模型在通用图文问答、视觉定位、OCR、图表、拍题解题、看图创作等任务上全面提升，显著优化了英文和小语种能力。	2025-09-16	查看详情

2025年08月

动态名称	动态描述	发布时间	相关文档
hunyuan-t1-latest 模型版本更新	一、发布特性： 大幅提升主模型慢思考模型的高难数学、复杂推理、高难代码、指令遵循、文本创作质量等能力。 二、效果指标： 通用能力上，相比线上版本，数学难题提升5pp，逻辑推理提升1.8pp，科学提升3pp，代码竞赛提升4pp，文创写作质量提升3pp，知识问答提升4.8pp。	2025-08-22	查看详情

2025年07月

动态名称	动态描述	发布时间	相关文档
hunyuan-turbos-	适用于视频理解场景。 发布特性：基于混元 turbos-vision 的视频理解模型，支持视频描述、视频内容问答等基本的视频理解能	2025-07-28	查看详情

vision-video 上线	力。		
hunyuan-turbos-20250716	<ol style="list-style-type: none"> 1. 训练策略：升级 PostTrain 训练策略，大幅提升模型的文理能力。 2. 通用优化：提升文创的内容质量和丰富度，提升文科通用的理解能力、专业知识能力和指令遵循能力，提升理科的推理能力，解题能力。 3. 产品特性：文创-优化讲稿体裁。 	2025-07-16	查看详情
hunyuan-t1-20250711	能力和特征：大幅提升高难度数学、逻辑和代码能力，优化模型输出稳定性，提升模型长文能力。	2025-07-11	查看详情

2025年06月

动态名称	动态描述	发布时间	相关文档
hunyuan-a13b 上线	适用场景：绝大部分场景，同时兼顾效果及推理性能。 模型能力和特征：混元第一个混合推理模型，hunyuan-standard-256K 的升级版，总参数 80B，激活 13B，默认是慢思考模式，支持通过参数或者指令进行快慢思考模式切换，慢快思考切换方式为 query 前加/ no_think；整体能力相对上一代全面提升，特别数学、科学、长文理解和 Agent 能力提升显著。 模型结构：混元 MoE 结构。	2025-06-25	查看详情
hunyuan-t1-vision-20250619 上线	特性：混元最新版 t1-vision 多模态理解深度思考模型，支持多模态原生长思维链，相比上一代默认版本模型全面提升。	2025-06-19	查看详情
hunyuan-turbos-vision-20250619 上线	特性：混元最新版 turbos-vision 视觉语言旗舰大模型，在图文理解相关的任务上，包括基于图片的实体识别、知识问答、文案创作、拍照解题等方面相比上一代默认版本模型全面提升。	2025-06-19	查看详情
hunyuan-translation-lite 模型版本更新	<ol style="list-style-type: none"> 1. 模型能力和特征：混元翻译专项模型，基于混元 2B-Dense 模型进行翻译能力专项优化，通过迭代高质量多语言 SFT 数据，强化多语种翻译能力。支持简体中文、繁体中文、粤语、印尼语、英语、日语、法语、葡萄牙语、西班牙语、土耳其语、俄语、阿拉伯语、韩语、意大利语、德语、越南语、马来语、印尼语等 18 种语言互译。 2. 模型结构：混元 Dense 结构。 	2025-06-06	查看详情

hunyuan-turbos-20250604 上线	能力和特征：预训练底座升级，写作、阅读理解能力提升，较大幅度提升代码和理科能力，复杂指令遵循等持续提升。	2025-06-04	查看详情
----------------------------	--	------------	----------------------

2025年05月

动态名称	动态描述	发布时间	相关文档
hunyuan-t1-20250529 上线	能力和特征：优化文本创作、作文写作，优化代码前端、数学、逻辑推理等理科能力，提升指令遵循能力。	2025-05-29	查看详情
hunyuan-t1-latest 模型版本更新	更新至 hunyuan-T1 模型的动态最新版本。	2025-05-28	查看详情
hunyuan-large-vision 上线	此模型适用于图文理解场景，是基于混元 Large 训练的视觉语言大模型，支持任意分辨率多张图片+文本输入，生成文本内容，聚焦图文理解相关任务，在多语言图文理解能力上有显著提升。	2025-05-26	查看详情
hunyuan-turbos-vision-v1.1更新	<ol style="list-style-type: none"> 此模型适用于图文理解场景，是基于混元最新 turbos 的新一代视觉语言旗舰大模型，聚焦图文理解相关任务，包括基于图片的实体识别、知识问答、文案创作、拍照解题等方面相比前一代模型全面提升。 发布特性：模型上下文窗口从8k 升级至32k，提升通用图片理解效果。 	2025-05-23	查看详情
hunyuan-t1-20250521	【最新版本】 <ol style="list-style-type: none"> 数学、代码、科学、知识问答能力显著增强。 优化文科数据，提升文创、文本理解能力。 	2025-05-21	查看详情
hunyuan-turbos-latest 模型版本更新	更新至 hunyuan-TurboS 模型的动态最新版本。	2025-05-20	查看详情
hunyuan-turbos-20250515	一、发布特性： <ol style="list-style-type: none"> 长短思维链自适应融合。 全面提升数学、代码、逻辑和科学等理科效果，文科效果维持。 二、效果指标： 理科：数学 ood 测试集提升7%；逻辑难题提升5%；代码能力及体验提升5%；科学提升8%。	2025-05-20	查看详情

hunyuan-t1-vision 正式版发布	混元多模态理解深度思考模型，支持多模态原生长思维链，擅长处理各种图片推理场景，在理科难题上相比快思考模型全面提升。	2025-05-16	查看详情
----------------------------	---	------------	----------------------

2025年04月

动态名称	动态描述	发布时间	相关文档
hunyuan-functioncall 模型版本更新	hunyuan-functioncall 模型版本更新，增强了复杂指令和多轮理解能力，增强了开源 BenchMark 的效果，相比上个版本在长输入意图、结构化输出等场景下效果提升明显。	2025-04-22	查看详情
hunyuan-turbos-latest 模型版本更新	作为 hunyuan-TurboS 模型的动态最新版本。	2025-04-16	查看详情
hunyuan-turbos-20250416 重磅推出	hunyuan-TurboS 混元旗舰大模型最新版本，具备更强的思考能力，更优的体验效果。预训练底座升级，增强底座的指令理解及遵循能力；对齐阶段增强数学、代码、逻辑、科学等理科能力；提升文创写作质量、文本理解、翻译准确率、知识问答等文科能力；增强各领域 Agent 能力，重点加强多轮对话理解能力等。	2025-04-16	查看详情
hunyuan-turbos-vision 重磅推出	此模型适用于图文理解场景，是基于混元最新 turbos 的新一代视觉语言旗舰大模型，聚焦图文理解相关任务，包括基于图片的实体识别、知识问答、文案创作、拍照解题等方面，相比前一代模型全面提升。	2025-04-07	查看详情
hunyuan-t1-latest 模型版本更新	腾讯发布首个深度推理模型 Hunyuan-T1 版本更新。本次更新：提升项目级代码生成能力；提升文本生成写作质量；提升文本理解 topic 的多轮、tob 指令遵循和字词理解能力；优化繁简混杂和中英混杂输出问题。	2025-04-03	查看详情
hunyuan-t1-20250403 重磅推出	全面提升模型文本生成能力、文本逻辑严谨性与专业规范性，强化项目级代码生成能力，提升文本逻辑与专业性，优化多轮 ToB 场景理解。	2025-04-03	查看详情

2025年03月

动态名称	动态描述	发布时间	相关文档
hunyuan-turbos-longtext-	混元全新一代 MOE 模型长文版本，基于混元 Turbos 底座进一步增训，擅长处理长文任务如文档摘要和文档问答等，同时也具备处理通用文本生成任务的能力。在	2025-03-25	查看详情

128k-20250325 重磅推出	长文本的分析和生成上表现优异，能有效应对复杂和详尽的长文内容处理需求。		
hunyuan-t1-latest 重磅推出	腾讯正式发布首个深度推理模型 Hunyuan-T1，该模型基于3月初发布的业界首个超大规模 Hybrid-Transformer-Mamba MoE 大模型 TurboS 快思考基座，扩展了推理能力，并进一步对齐人类偏好。	2025-03-21	查看详情
hunyuan-t1-20250321 重磅推出	全面构建模型文理科能力，长文本信息捕捉能力强。支持推理解答各种难度的数学/逻辑推理/科学/代码等科学问题。腾讯首个推理模型。	2025-03-21	查看详情
hunyuan-turbos-latest 模型版本更新	作为 hunyuan-TurboS 模型的动态最新版本，与 C 端（腾讯元宝）保持一致。	2025-03-13	查看详情
hunyuan-turbos-20250313 模型版本更新	hunyuan-TurboS 混元旗舰大模型最新版本，具备更强的思考能力，更优的体验效果。预训练底座训练 token 数升级；数学/逻辑/代码等思考能力提升；中英文通用体验效果提升，包括文本创作、文本理解、知识问答、闲聊等。	2025-03-13	查看详情

2025年02月

动态名称	动态描述	发布时间	相关文档
hunyuan-turbos-20250226 重磅推出	hunyuan-TurboS pv2.1.2 固定版本。	2025-02-26	查看详情
hunyuan-large 模型版本更新	hunyuan-large 模型版本更新，各项能力上相比前一代模型全面提升。	2025-02-10	查看详情
hunyuan-standard-256K 模型版本更新	持续优化通用基础效果，长文多文档问答、长文翻译等高质量精标数据，在长文检索和长中文问答任务上优于市场同规模竞品模型。	2025-02-10	查看详情
hunyuan-standard 模型版本更新	持续优化通用基础效果，训练数据中融合更多领域数据，在指令遵循、逻辑、理解等基础能力上也得到进一步提升。	2025-02-10	查看详情

2025年01月

动态名称	动态描述	发布时间	相关文档
hunyuan-turbo 模型版本更新	本版本优化：数据指令 scaling，大幅提升模型通用泛化能力；大幅提升数学、代码、逻辑推理能力；优化文本理解字词理解相关能力；优化文本创作内容生成质量。	2025-01-10	查看详情
hunyuan-turbo-latest 模型版本更新	通用体验优化，包括 NLP 理解、文本创作、闲聊、知识问答、翻译、领域等；提升拟人性，优化模型情商；提升意图模糊时模型主动澄清能力；提升字词解析类问题的处理能力；提升创作的质量和可互动性；提升多轮体验。	2025-01-10	查看详情
hunyuan-turbo-20241223 上线	turbo pv2.0.6 固定版本 本版本优化：数据指令 scaling，大幅提升模型通用泛化能力；大幅提升数学、代码、逻辑推理能力；优化文本理解字词理解相关能力；优化文本创作内容生成质量。	2025-01-10	查看详情

2024年12月

动态名称	动态描述	发布时间	相关文档
hunyuan-standard-vision 重磅推出	混元最新多模态模型，支持多语种作答，中英文能力均衡。	2024-12-31	查看详情
hunyuan-turbo 模型版本更新	相比之前版本大幅提升模型通用泛化能力；大幅提升数学、代码、逻辑推理能力；优化文本理解字词理解相关能力；优化文本创作内容生成质量。	2024-12-23	查看详情
hunyuan-turbo-latest 模型版本更新	大幅提升模型通用泛化能力；大幅提升数学、代码、逻辑推理能力；优化文本理解字词理解相关能力；优化文本创作内容生成质量。	2024-12-23	查看详情
hunyuan-turbo-20241120 上线	hunyuan-turbo 2024 年 11 月 20 日固定版本，介于 hunyuan-turbo 和 hunyuan-turbo-latest 之间的一个版本。	2024-12-23	查看详情
hunyuan-large-longcontext 模型版本更新	hunyuan-large-longcontext 模型版本更新，持续提升多文档、多轮、摘要总结复杂指令效果。	2024-12-18	查看详情

hunyuan-lite-vision 重磅推出	混元最新7B 多模态模型，支持中英文场景的多模态对话、图像物体识别、文档表格理解、多模态数学等。	2024-12-12	查看详情
--------------------------	--	------------	----------------------

2024年11月

动态名称	动态描述	发布时间	相关文档
hunyuan-turbo-vision 重磅推出	混元新一代视觉语言旗舰大模型，采用全新的混合专家模型（MoE）结构，在图文理解相关的基础识别、内容创作、知识问答、分析推理等能力上相比前一代模型全面提升。	2024-11-26	查看详情
hunyuan-translation-lite 重磅推出	混元翻译模型支持自然语言对话式翻译；支持中文和英语、日语、法语、葡萄牙语、西班牙语、土耳其语、俄语、阿拉伯语、韩语、意大利语、德语、越南语、马来语、印尼语15种语言互译。	2024-11-25	查看详情
hunyuan-turbo-latest 模型版本更新	hunyuan-turbo-latest 模型版本更新，优化解决特殊字符导致重复问题，全面提升各项能力，提升 markdown 输出的稳定性，降低不合理拒答率，持续提升文本理解、创作、知识问答效果。	2024-11-20	查看详情
hunyuan-large 模型版本更新	hunyuan-large 模型版本更新，修复人设和长文数据效果。	2024-11-20	查看详情
hunyuan-functioncall 模型版本更新	hunyuan-functioncall 模型版本更新，融合最新 SFT 训练数据，相比上个版本在长输入意图、结构化输出等场景下效果提升明显。	2024-11-15	查看详情
hunyuan-code 模型版本更新	hunyuan-code 模型版本更新，通过半年高质量 SFT 数据迭代训练，其效果在各方向综合代码任务的高质量人工评测上均有提升。	2024-11-12	查看详情
hunyuan-large-longcontext 重磅推出	此模型采用混元 MoE 结构，拥有 A52B 的参数规模，基于开源的 hunyuan-large 模型进一步增训，特别优化了对长文任务的处理能力。在长文本解析和生成方面展现显著的效果，与元宝 C 端文档阅读场景使用的是同款模型。	2024-11-08	查看详情
hunyuan-large 重磅推出	腾讯开发的开源业界参数规模最大、效果最好的 transformer 结构的 MoE 模型。在 CMMLU、MMLU、CEval 等多学科综合评测集、中英文 NLP 任务、代码和数学等9大维度全面领先，处于行业领先水平。	2024-11-05	查看详情

hunyuan-vision 模型版本更新	提升多图分析推理、多图对比理解、复杂指令遵循能力，优化输出格式 bad case。	2024-11-04	查看详情
-----------------------	---	------------	----------------------

2024年10月

动态名称	动态描述	发布时间	相关文档
hunyuan-lite 模型版本更新	模型版本正常迭代，融合最新 SFT 训练数据，相比较上一版本在数学、专业领域效果提升明显。	2024-10-30	查看详情
hunyuan-standard 模型版本更新	更新 SFT 训练数据，优化通用基础效果，相比上个版本在数学、多轮对话、知识问答领域上具有更好的效果。	2024-10-28	查看详情
hunyuan-standard-256K 模型版本更新	补充更新最新长文数据，在长文检索和长中文问答任务上优于市场同规模竞品模型。	2024-10-28	查看详情
hunyuan-translation 重磅推出	支持中文和英语、日语、法语、葡萄牙语、西班牙语、土耳其语、俄语、阿拉伯语、韩语、意大利语、德语、越南语、马来语、印尼语15种语言互译，基于多场景翻译评测集自动化评估 COMET 评分，在十余种常用语种中外互译能力上整体优于市场同规模模型。	2024-10-25	查看详情
hunyuan-vision 重磅推出	支持多图输入能力，优化通用单图基础模型效果。	2024-10-23	查看详情
hunyuan-turbo-latest 重磅推出	作为 hunyuan-turbo 模型的动态最新版本，与 C 端（腾讯元宝）保持一致。	2024-10-22	查看详情

2024年09月

动态名称	动态描述	发布时间	相关文档
hunyuan-standard-256K 模型版本更新	优化通用基础效果，补充长文多文档问答、长文翻译等高质量精标数据，在长文检索和长中文问答任务上优于市场同规模竞品模型。	2024-09-26	查看详情

hunyuan-standard 模型版本更新	在通用效果提升的基础上，训练数据中融合了医疗、金融等更多领域数据，在指令遵循、逻辑、理解等基础能力上也得到进一步提升。	2024-09-26	查看详情
hunyuan-lite 模型版本更新	重点优化了模型的数值计算和逻辑推理能力，同时模型的文本创作和多轮对话能力也得到了显著提升。	2024-09-18	查看详情
混元生图模型版本更新	模型升级，优化通用生图效果。	2024-09-11	查看详情
重磅推出混元全新 MOE 模型 hunyuan-turbo	混元新一代大语言模型，采用全新的混合专家模型（MoE）结构，性能和推理效率提升1倍以上。在中文创作能力、逻辑推理能力、数学计算和多轮对话等通用场景上，效果提升显著，综合能力优于国内同规模大模型。	2024-09-05	查看详情
hunyuan-function 模型版本更新	混元 function call 模型版本升级，支持并行工具调用、支持多轮、支持 Workflow 形式的 System Prompt、支持长输入意图、支持结构化输出。	2024-09-04	查看详情

2024年08月

动态名称	动态描述	发布时间	相关文档
hunyuan-standard-256K 模型版本更新	在通用效果提升的基础上，训练数据中融合了医疗、金融领域的长文数据、长文翻译数据和长文多文档问答等高质量精标数据，使其效果在检索任务和中文问答任务上显著优于市场同规模竞品模型。	2024-08-28	查看详情
hunyuan-standard 模型版本更新	在通用效果提升的基础上，训练数据中融合了医疗、金融等更多领域数据，在指令遵循、逻辑、理解等基础能力上也得到进一步提升。	2024-08-28	查看详情
混元生图（多轮对话）上线	混元生图（多轮对话）API 上线，支持通过多轮对话的方式不断调整图像内容。	2024-08-28	查看详情
混元生图模型版本更新	模型升级，优化通用生图效果。	2024-08-23	查看详情
hunyuan-vision 模型版本更新	优化图生文指令理解能力，优化知识问答、OCR 识别、翻译、初中数学、多轮对话等场景能力。	2024-08-21	查看详情
混元生文新增搜索结果返回	支持打开搜索插件开关并命中搜索时，返回搜索结果链接，支持在正文中打开引用角标开关。	2024-08-02	API 更新历史

2024年07月

动态名称	动态描述	发布时间	相关文档
混元生图模型版本更新	新增 seed 参数，支持随机或固定种子生成图片。 新增一次生成多图功能，最多支持4张图片。 新增自定义水印图片功能。	2024-07-26	查看详情
混元生图模型版本更新	模型升级，优化通用生图效果。 prompt 输入长度上限从100扩展至1024个字符。	2024-07-18	查看详情
文生图轻量版上线	文生图轻量版 API 上线。	2024-07-15	查看详情
重磅推出混元多模态模型 hunyuan-vision	支持在对话中输入图片生成文本：包括图片基础识别、图片内容创作、图片多轮对话、图片知识问答、图片分析推理、图片 OCR 等能力。	2024-07-04	查看详情
新增角色扮演专属模型 hunyuan-role	混元最新版角色扮演模型，混元官方精调训练推出的角色扮演模型，基于混元模型结合角色扮演场景数据集进行增训，在角色扮演场景具有更好的基础效果。	2024-07-04	查看详情
新增代码生成专属模型 hunyuan-code	混元最新代码生成模型，经过 200B 高质量代码数据增训基座模型，迭代半年高质量 SFT 数据训练，上下文长窗口长度增大到 8K，五大语言代码生成自动评测指标上位居前列；五大语言10项考量各方面综合代码任务人工高质量评测上，性能处于第一梯队。	2024-07-04	查看详情
新增 functioncall 专属模型	混元最新 MOE 架构 FUNCCALL-SFT 模型，经过高质量的 FunctionCall 数据训练，上下文窗口达32K，在多个维度的评测指标上处于领先。	2024-07-04	查看详情

2024年05月

动态名称	动态描述	发布时间	相关文档
hunyuan-lite 升级	<ul style="list-style-type: none"> 升级为多专家 MOE 架构，模型效果极大提升。 上下文输入输出扩展为256K。 	2024-05-22	查看详情
混元生图模型体验中心上线	腾讯混元大模型控制台上线模型体验中心，支持在线体验混元生图。	2024-05-13	模型体验中心

混元生图上线	混元生图 API 上线，支持根据输入的 Prompt 文本生成图像。	2024-05-13	查看详情
新增 hunyuan-standard-256K	<ul style="list-style-type: none"> 新增超长上下文的 hunyuan-standard-256K，输入突破到250K，输出6K hunyuan-standard-256K 独立定价，详见 购买指南。 	2024-05-11	查看详情

2024年04月

动态名称	动态描述	发布时间	相关文档
混元大模型各模型版本品牌升级	原“混元大模型高级版”更名为“hunyuan-pro”；原“混元大模型标准版”更名为“hunyuan-standard”。	2024-04-25	/
hunyuan-standard 模型升级	hunyuan-standard 模型升级为32K 长文大规模多专家模型，效果、性能输出限制极大提升。其中，输入上限扩展为30Ktokens，输出上限2Ktokens	2024-04-25	/
新增 hunyuan-lite	新增更具性价比的 hunyuan-lite 版本。	2024-04-25	查看详情
混元大模型高级版模型升级	<ul style="list-style-type: none"> 底座升级，预训练数据量在原有基础上进一步丰富和扩展。 上下文限制极大提升，从16K 扩展到支持32K上下文长度（输入24K，输出8K）。 优化格式化输出效果。 优化文本生成效果。 	2024-04-22	/
新增流式调用开关	混元大模型高级版、混元大模型标准版中，新增流式调用开关。支持选择大模型输出方式为流式或者非流式一次性输出。	2024-04-11	流式请求成功示例 非流式请求成功示例
新增流式输出审核开关	混元大模型高级版、混元大模型标准版中，新增流式输出审核开关。 <ul style="list-style-type: none"> 开启流式输出审核时，混元安全将分别对每个流式片段独立审核，提升首字出话效率； 	2024-04-11	详见 StreamModeration

- 未充值时，默认为审核完所有输出内容后再返回。

[n 参数说明](#)

2024年01月

动态名称	动态描述	发布时间	相关文档
腾讯混元-Embedding 上线	新增 Embedding 接口：可以将文本转化为高质量的向量数据。	2024-01-05	查看详情

2023年11月

动态名称	动态描述	发布时间	相关文档
腾讯混元大模型标准版上线	新增模型：腾讯混元大模型标准版，由腾讯研发的大语言模型，具备强大的中文创作能力，复杂语境下的逻辑推理能力，以及可靠的任务执行能力。	2023-11-23	查看详情

2023年09月

动态名称	动态描述	发布时间	相关文档
腾讯混元大模型上线	腾讯混元大模型（Tencent Hunyuan）是由腾讯研发的大语言模型，具备强大的中文创作能力，复杂语境下的逻辑推理能力，以及可靠的任务执行能力。	2023-09-06	产品简介

旧版本模型下线迁移公告

最近更新时间：2026-05-20 16:01:21

随着腾讯云大模型的升级与迭代发布，我们将逐步对旧版本模型进行下线规划调整。

根据调整规划，混元旧版本模型（详见下方模型列表）将于北京时间 2026 年 6 月 22 日 00:00 起正式下线，届时不再提供以下模型的接入服务。

同时原 [腾讯混元大模型平台](#) 将于 9 月 30 日 全面停服下线，此次下线旧版本模型建议迁移至 TokenHub 使用最新版本模型。详见 [TokenHub 迁移指南](#)。

下线模型列表

下线 model 参数	建议迁移 model 参数
<code>hunyuan-2.0-instruct-20251111</code>	<code>hy3-preview</code>
<code>hunyuan-2.0-thinking-20251109</code>	<code>hy3-preview</code>
<code>hunyuan-2.0-thinking-dev-20251012</code>	<code>hy3-preview</code>
<code>hunyuan-t1-latest</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250321</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250403</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250521</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250529</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250711</code>	<code>hy3-preview</code>
<code>hunyuan-t1-20250822</code>	<code>hy3-preview</code>
<code>hunyuan-t1-dev-20250822</code>	<code>hy3-preview</code>
<code>hunyuan-turbos-latest</code>	<code>hy3-preview</code>
<code>hunyuan-turbos-20250226</code>	<code>hy3-preview</code>
<code>hunyuan-turbos-customized</code>	<code>hy3-preview</code>
<code>hunyuan-turbos-20250313</code>	<code>hy3-preview</code>

hunyuan-turbos-20250416	hy3-preview
hunyuan-turbos-20250515	hy3-preview
hunyuan-turbos-20250604	hy3-preview
hunyuan-turbos-20250716	hy3-preview
hunyuan-turbos-20250926	hy3-preview
hunyuan-turbos-longtext-128k-20250325	hy3-preview
hunyuan-turbo-latest	hy3-preview
hunyuan-turbo	hy3-preview
hunyuan-customized	hy3-preview
hunyuan-turbo-customized	hy3-preview
hunyuan-turbo-20241120	hy3-preview
hunyuan-turbo-20241223	hy3-preview
hunyuan-standard	hy3-preview
hunyuan-standard-32K	hy3-preview
hunyuan-standard-256K	hy3-preview
hunyuan-code	hy3-preview
hunyuan-funcall	hy3-preview
hunyuan-large	hy3-preview
hunyuan-large-longcontext	hy3-preview
deepsearch	hy3-preview
hunyuan	hy3-preview
hunyuan-pro	hy3-preview
hunyuan-lite	hy3-preview

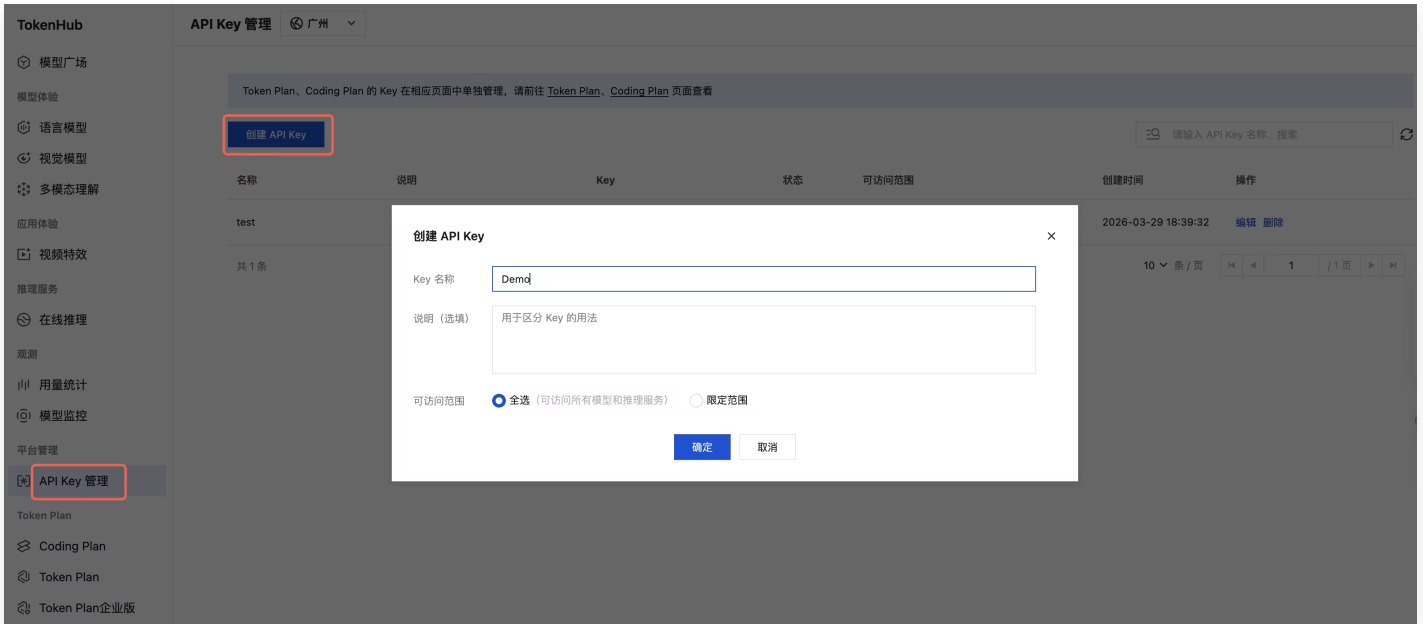
hunyuan-large-role	hunyuan-role-latest
hunyuan-large-role-latest	hunyuan-role-latest
hunyuan-large-role-plus	hunyuan-role-latest
hunyuan-role	hunyuan-role-latest
hunyuan-role-plus	hunyuan-role-latest
hunyuan-large-role-20250822	hunyuan-role-latest
hunyuan-turbos-role-20251114	hunyuan-role-latest
hunyuan-translation-preview	hy-mt2-pro

迁移方式

第一步：登录 [TokenHub](#) 控制台，前往 [在线推理](#) 页面，选择您需要的模型，开启免费体验以及启用后付费，开启后将优先消耗免费体验额度，免费体验额度使用完之后按 Token 用量后付费，模型后付费价格请参见 [模型价格](#)。

服务名称	服务 ID	类型	状态	免费额度用量	用量监控	计费方式	操作
deepseek-v4-flash	deepseek-v4-flash	默认	运行中	用0 / 共1000000 tokens	山	免费体验	启用后付费 删除
glm-5-turbo	glm-5-turbo	默认	已停止	-	山	-	启用后付费 删除
glm-5v-turbo	glm-5v-turbo	默认	运行中	用0 / 共500000 tokens	山	免费体验	启用后付费 删除
hy3-preview	hy3-preview	默认	运行中	用33 / 共1000000 tokens	山	免费体验 按 Token 计费	停用后付费 删除
kimi-k2.6	kimi-k2.6	默认	运行中	用0 / 共500000 tokens	山	免费体验 按 Token 计费	停用后付费 删除
kimi-k2.5	kimi-k2.5	默认	运行中	用33 / 共500000 tokens	山	免费体验 按 Token 计费	停用后付费 删除

第二步：在 TokenHub [API Key 管理](#) 页面中创建新的 API Key。操作详情请参见 [API Key 管理](#)。



第三步: 参考 [模型详情页](#) 中的调用示例（以 Hy3 preview 为例），更换调用的 API Key。

- 若通过 OpenAI SDK 访问，则 base_url 更换为：`https://tokenhub.tencentmaas.com/v1`
- 若通过 URL 直接访问，则 URL 为：`https://tokenhub.tencentmaas.com/v1/chat/completions`

调用方式 | 通过 API 快速集成模型能力

首次成功调用模型后，平台将自动为您创建与模型同名的[在线推理服务](#)。

- 1 获取 API Key**
调用推理服务需要使用 API Key 进行身份鉴权，请先前往 API Key 管理页面创建 API Key。
[前往 API Key 管理](#)
- 2 替换 API Key 并调用服务**
Chat Completions API Responses API

```
curl -X POST 'https://tokenhub.tencentmaas.com/v1/chat/completions' \
-H 'Authorization: Bearer YOUR_API_KEY' \
-H 'Content-Type: application/json' \
-d '{
  "model": "hy3-preview",
  "messages": [
    {"role": "system", "content": "You are a helpful assistant."},
    {"role": "user", "content": "你好"}
  ],
  "stream": false
}'
```

更多 API 调用示例请参见 [API 使用说明](#)。

联系我们

如果您在迁移过程中遇到任何问题或需要帮助，您可以咨询 [在线客服](#) 或 [提交工单](#) 来与我们联系。