

腾讯混元大模型

产品简介



腾讯云

【 版权声明 】

©2013–2024 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

产品简介

产品概述

产品优势

应用场景

产品简介

产品概述

最近更新时间：2024-11-08 17:43:22

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。

混元生文

混元文生文目前已上线10个版本，以下为模型特性的简要介绍，可作为业务接入时版本选择的参考依据。

模型类型	模型名称	版本更新时间	能力和特征	输入输出	相关文档
Turbo	hunyuan-turbo-latest	2024-10-17	Hunyuan-turbo 模型动态更新版本，是混元模型系列效果最好的版本，与C端（腾讯元宝）保持一致。	最大输入28k 最大输出4k	调用示例 立即体验
	hunyuan-turbo	2024-09-18	Hunyuan-turbo 模型默认版本，采用全新的混合专家模型（MoE）结构，相比 hunyuan-pro 推理效率更快，效果表现更强。	最大输入28k 最大输出4k	调用示例 立即体验
Large	hunyuan-large	2024-11-05	Hunyuan-large 模型总参数量约 389B，激活参数量约 52B，是当前业界参数规模最大、效果最好的 Transformer 架构的开源 MoE 模型	最大输入28k 最大输出4k	调用示例 立即体验 腾讯混元-large
	hunyuan-large-	2024-11-08	擅长处理长文任务如文档摘要和文档问答等，同时也具	最大输入128k	调用示例 立即体验

	longcontext		备处理通用文本生成任务的能力。在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。	最大输出6k	腾讯混元-large
Pro	hunyuan-pro	2024-09-05	万亿级参数规模 MOE-32K 长文模型。在各种 benchmark 上达到绝对领先的水平，复杂指令和推理，具备复杂数学能力，支持 functioncall，在多语言翻译、金融法律医疗等领域应用重点优化。	最大输入28k 最大输出4k	调用示例 立即体验
Standard	hunyuan-standard-256K	2024-10-28	采用更优的路由策略，同时缓解了负载均衡和专家趋同的问题。长文方面，大海捞针指标达到99.9%。MOE-256K 在长度和效果上进一步突破，极大的扩展了可输入长度。	最大输入250k 最大输出6k	调用示例 立即体验
	hunyuan-standard	2024-10-28	采用更优的路由策略，同时缓解了负载均衡和专家趋同的问题。MOE-32K 性价比相对更高，在平衡效果、价格的同时，可对实现对长文本输入的处理。	最大输入30k 最大输出2k	调用示例 立即体验
Lite	hunyuan-lite	2024-10-30	升级为 MOE 结构，上下文窗口为256k，在 NLP，代码，数学，行业	最大输入250k 最大输出6k	调用示例 立即体验

			等多项评测集上领先众多开源模型。		
角色扮演	hunyuan-role	2024-07-04	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。	最大输入28k 最大输出4k	调用示例 立即体验
Function Call	hunyuan-functioncall	2024-09-04	混元最新 MOE 架构 FunctionCall 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。	最大输入28k 最大输出4k	调用示例 立即体验
代码	hunyuan-code	2024-07-04	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	最大输入4k 最大输出4k	调用示例 立即体验

混元多模态模型

混元多模态模型重磅发布，支持图片生文，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。以下为模型特性的简要介绍，可作为业务接入时版本选择的参考依据。

模型类型	模型名称	版本更新时间	能力和特征	输入输出	相关文档
------	------	--------	-------	------	------

混元图生文	hunyuan-vision	2024-11-04	<p>混元最新多模态模型，支持图片+文本输入生成文本内容。</p> <p>图片基础识别：对图片中主体、元素、场景等进行识别</p> <p>图片内容创作：对图片进行概述、创作广告文案、朋友圈、诗词等</p> <p>图片多轮对话：输出单张图片进行多轮交互问答</p> <p>图片分析推理：对图片中逻辑关系、数学题、代码、图表进行统计分析</p> <p>图片知识问答：对图片包含的知识点进行问答，例如历史事件、电影海报</p> <p>图片 OCR：对自然生活场景、非自然场景的图片识别文字</p>	<p>最大输入 6k</p> <p>最大输出 2k</p>	<p>调用示例 立即体验</p>
-------	----------------	------------	---	-------------------------------	--

混元生图

混元生图提供文生图 API 技术服务，基于混元文生图大模型，可结合输入的文本描述智能创作出与输入相关的图像内容，提供更高质量的图像创作、更强大的中文语义理解能力以及更广泛的应用场景。

模型类型	模型名称	版本更新时间	能力和特征	输入输出	相关文档
混元生图	混元生图	2024-09-11	<p>输入 prompt 文本，生成图像。</p> <p>prompt 扩写：支持 prompt 自动扩写优化。</p> <p>绘画风格：支持18种预置风格选项 + prompt 自定义风格。</p>	<p>最大输入1024字符文本</p> <p>最大输出4张图片</p>	<p>调用示例 立即体验</p>

			分辨率：支持8种分辨率。 生成数量：支持1-4张。 随机种子：支持随机种子生成或自定义固定种子生成。		
	混元生图 (多轮对话)	2024-08-28	输入 prompt 文本，生成图像，支持通过多轮对话的方式不断调整图像内容。	最大输入100轮对话文本 输出图片	调用示例
	文生图轻量版	2024-07-15	输入 prompt 文本，生成图像，支持更轻量的同步请求方式和更高的性能。	最大输入256字符文本 输出图片	调用示例

腾讯元器平台

腾讯元器平台是一款基于混元大模型开发的一站式智能体创作与分发平台，用户无需编程经验即可轻松创建属于自己的智能体。在这个平台上，用户可以通过描述来设定智能体的角色和流程，利用丰富的插件和知识库资源让智能体更加智能化。平台还提供基于 AI 大模型的智能对话功能，以及智能体商店，汇聚了众多富有创意和趣味性的智能体玩法，为用户带来全新的智能体验。

产品优势

最近更新时间：2024-09-20 14:53:51

腾讯混元大模型是由腾讯全链路自研的通用大语言模型，拥有万亿参数规模，具有强大的中文理解与创作能力、逻辑推理能力，以及可靠的任务执行能力。

腾讯混元通过采用混合专家模型(MoE)结构，推动了性能提升和推理成本下降。在中文表现尤其是在文本生成、数理逻辑和多轮对话上性能表现卓越，整体处于业界领先水平。支持AI 搜索联网插件，通过整合腾讯优质的内容生态（如微信公众号、视频号等）和全网搜索，使混元具备强大的时新、深度内容获取和AI问答能力。

同时，混元还采用了各种技术手段来提高模型的性能和效果，例如使用掩码策略、使用不同的优化算法、进行数据增强等。这些技术手段可帮助模型更好地处理文本数据，提高模型的泛化能力和生成效果。

腾讯混元大模型目前覆盖**四大核心能力**：

多轮对话

具备上下文理解和长文记忆能力，流畅完成各专业领域的多轮问答。

知识增强

有效解决事实性、时效性问题，快速检索相关内容，提升内容生成效果。

逻辑推理

准确理解用户意图，擅长数学推导，基于输入数据或信息进行推理、分析和规划。

内容创作

支持文学创作、文本概要、角色扮演、文生图和图生文等多模态创作能力，生成内容流畅、规范、中立、客观。

应用场景

最近更新时间：2024-03-29 15:46:41

文档场景

可提供文档创作、文本润色、文本校阅、表格公式及图表生成等能力，提高创作效率，提升创作体验。

会议场景

可提供会中问答、会议总结、会议待办项整理等能力，简化会议操作并提高会议效率。

广告场景

可提供智能化的广告素材创作，提升营销内容创作工作效率。

营销场景

构建智能导购，帮助商家提升服务质量和效率。