

# 腾讯云智能体开发平台 购买指南



腾讯云

## 【 版权声明 】

©2013–2025 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

## 购买指南

计费概述

购买方式

续费说明

退费说明

# 购买指南

## 计费概述

最近更新时间：2025-07-03 10:59:41

腾讯云智能体开发平台提供大模型应用构建平台，根据用户所使用的 tokens 数、知识库容量、并发数等资源进行计费，目前为公测阶段，注册开通体验即可获得一定量的免费额度。如需购买或扩容，请参见 [购买方式](#)。

### 开通方式

腾讯云智能体开发平台的开通使用需要先通过 [腾讯云企业实名认证](#) 或者 [腾讯云个人实名认证](#)。通过实名认证后，首次在 [腾讯云智能体开发平台产品页](#) 单击[产品体验](#)时，即可开通腾讯云智能体开发平台使用权限。

### 免费额度

通过实名认证的主账号首次登录腾讯云智能体开发平台产品时，获得一定量的免费体验额度，详情如下图所示：

资源类别	免费额度
精调知识大模型标准版	<p>首次开通腾讯云智能体开发平台服务即获赠累计50万 tokens 的免费调用额度，有效期2个月；以资源包的形式发放到您的腾讯云账号中，优先扣除。（模型通用）</p> <div><p><b>说明：</b></p><p>2025年04月07日12:00至2025年08月31日12:00 <a href="#">活动</a> 期间，企业新客户获得 300万 tokens DeepSeek 模型资源包，有效期1年，以资源包的形式发放到您的腾讯云账号中，优先扣除。</p></div> <p>免费额度：支持在 <a href="#">智能体开发平台-计费管理-资源包</a> 中查看。</p>
精调知识大模型高级版	
混元大模型标准版	
混元大模型高级版	
混元大模型长文本版	
混元大模型角色扮演版	
混元大模型 Turbo 版	
DeepSeek-R1	
DeepSeek-V3	
DeepSeek-V3-0324	
自定义模型	
应用服务-知识库容量	<p>开通腾讯云智能体开发平台服务即获赠累计300万字符数的免费知识库容量，有效期6个月；（注：针对2024年7月5日及以后开通产品的账号生效）到期后，知识库中的知识将被处理成“超量失效”状态，“超量失效”状态的知识在对话中不生效，且“超量失效”状态的知识180天后将被自动删</p>

	除，如需要继续使用知识库，请在到期前及时 <a href="#">购买</a> 。
联网搜索	首次开通腾讯云智能体开发平台服务即获赠5000次的免费调用额度，有效期2个月。（注：仅支持在智能体开发平台创建应用，开启联网搜索服务时使用） 用尽或到期后，需要 <a href="#">购买</a> 搜索服务包使用。

## 特别说明

### ⚠ 注意：

- 腾讯云智能体开发平台创建 Multi-Agent 模式应用，选择 DeepSeek 系列模型时，截止到北京时间2025年06月26日限时免费结束。2025年06月27日0时起，我们将对 DeepSeek 系列模型进行计费，产品支持预付费和后付费两种收费模式。
- 腾讯云智能体开发平台将下架行业大模型相关计费项。行业大模型能力已合并至精调知识大模型系列，用户可在应用中选择使用精调知识大模型高级版或标准版。

## 在以下场景/功能中发生交互时，会对 tokens 产生消耗：

- 应用配置管理：
  - 应用配置 > 角色设定中，一键优化功能。
  - 知识库管理 > 问答 > 导入问答 > 文档生成问答对。
  - 知识库管理 > 任务流程 > 插入节点 > 自动生成询问语。
  - 知识库管理 > 任务流程 > 插入节点 > 自动生成答案的预览。
  - 工作流>大模型类节点（参数提取、大模型、大模型意图识别、大模型标签提取、大模型知识问答），消耗 tokens 计入在节点中选择的模型上。
  - 工作流>条件判断节点（选择通过 AI 方式判断）或者选项卡节点（在做用户 query 与选项相似度判断时），消耗tokens计入在应用配置的生成模型上。
- 问答过程中，在应用配置中测试、应用发布后调用应用 API / 使用体验链接、应用评测、应用体验。

计算消耗内容：包括用户的输入 + 系统 prompt + 召回信息（含文档、问答、开启搜索引擎后搜索召回的内容）+历史上文+ 输出。

## 备注

1. tokens 换算方式：1 tokens  $\approx$  1~1.5个汉字。
2. 基于已配置的任务流程会调用任务型专属模型，消耗对应的 tokens。
3. 应用对话交互中触发的应用配置的欢迎语、输出配置-回复设置中自定义的保守回复、敏感词拦截后的回复不计入 tokens 消耗。
4. 如不使用搜索引擎，建议关闭"应用配置 > 联网搜索"开关，以减少额外的 搜索次数 消耗。
5. 应用配置 > 知识库 > 高级选项中，文档和问答的召回数量会影响召回 tokens ，设定的数量越高，拼接输入到大模型的召回片段越多，消耗 tokens 数量相应增加。

**⚠ 注意:**

- 在账单结算时，系统将按照免费资源包 > 预付费资源包的顺序进行结算，即免费资源包是优先扣除的。
- 若您欠费或因违禁原因停服后，将不能继续享受免费额度，服务恢复后可继续使用有效期范围内的免费额度。
- 免费额度：精调知识大模型标准版、精调知识大模型高级版、混元大模型系列、DeepSeek 系列共用 50万 tokens 免费额度。
- 优先消耗到期时间排在前面的免费额度。

## 计费大纲

计费场景	计费类型	资源类型	说明
应用服务	应用服务-平台预置模型	大模型 tokens 资源包	多种预置大模型服务调用价格，包括精调知识大模型、混元大模型等。
		大模型按量后付费	
		大模型共享并发包（已下架）	预置大模型提供默认并发额度，可通过购买共享并发包增加并发。
	应用服务-自定义模型	大模型 tokens 资源包	产品上支持对接三方模型，包括 Tione 平台或其他三方部署的生成大模型。 使用时需要收取应用中使用的周边模型费用。如改写、检索、排序等。
		大模型按量后付费	
	应用服务-用户部署模型	大模型专属并发包	提供独立部署模型资源，独立部署模型不按照 tokens 消耗计费，可通过购买专属并发包使用。
	应用服务（通用）	知识库扩容包	购买知识库扩容包可上传文档、问答到知识管理中实现知识问答场景。
		搜索服务包	通过购买搜索服务包实现调用搜索引擎。

**❗ 购买说明:**

- 应用服务-用户部署模型购买前需联系架构师提前部署。
- 如使用应用中的知识管理，需要购买知识库扩容包。

- 需要使用应用中的联网搜索时，可购买搜索服务包。
- 平台预置模型在北京时间2025年05月14日开始下架大模型共享并发包的售卖，目前按照 TPM 和 QPM 限流（按 tokens 收费使用公共资源池不保障 TPM 和 QPM 使用上限，对限流有要求可以购买专属并发）。

## 产品价格

### 应用服务

#### 预置模型-大模型 tokens 资源包

定义：一次性购买一定数量的 tokens 资源包，有效期内调用模型服务时优先抵扣资源包当中的 tokens 余量，tokens 资源包如果到期未用完，会当做过期作废处理。

资源包有效期：1年。1年后未使用的资源包清零。

付费方式：预付费，购买 tokens 资源包越大，单价越低。

适用范围：稳定调用，具有一定规模的业务体量。

tokens 资源包	精调知识大模型标准版	精调知识大模型高级版	混元大模型Turbo 版
1000万 tokens	¥120	¥1,200	¥240
5000万 tokens	¥600	¥6,000	¥1,200
1亿 tokens	¥1,180	¥11,800	¥2,300
5亿 tokens	¥5,900	¥59,000	¥11,500
10亿 tokens	¥11,700	¥117,000	¥22,600

#### ⚠ 注意：

- 购买的预付费 tokens 资源包有效期为1年，1年后未使用的资源包清零。
- 购买 tokens 预付费配额后，服务并发数保持不变，如果无法满足实际的业务需求，可按需增购并发。
- 支持在 [产品购买页](#) 中选择大模型 tokens 资源包购买。

#### 预置模型-大模型按量后付费

定义：预置模型除了提供预付费的 tokens 资源包外，还提供按量后付费的计费方式。按量后付费根据实际调用接口产生的 tokens 数量计费，不同的模型价格不同，按照实际消耗的输入、输出总 tokens 数量，**每小时**对腾讯云账户进行扣费。

模型版本	单价
精调知识大模型标准版	输入：0.012元/千tokens 输出：0.012元/千tokens
精调知识大模型高级版	输入：0.12元/千tokens 输出：0.12元/千tokens
混元大模型标准版	输入：0.0045元/千tokens 输出：0.005元/千tokens
混元大模型高级版	输入：0.03元/千tokens 输出：0.1元/千tokens
混元大模型长文本版	输入：0.015元/千tokens 输出：0.06元/千tokens
混元大模型角色扮演版	输入：0.004元/千tokens 输出：0.008元/千tokens
混元大模型Turbo版	输入：0.015元/千tokens 输出：0.05元/千tokens
DeepSeek-R1	输入：0.026元/千tokens 输出：0.075元/千tokens
DeepSeek-R1-0528	输入：0.026元/千tokens 输出：0.075元/千tokens
DeepSeek-V3	输入：0.0115元/千tokens 输出：0.017元/千tokens
DeepSeek-V3-0324	输入：0.0115元/千tokens 输出：0.017元/千tokens

**⚠ 注意：**

- 支持在 [后付费设置](#) 中开启后付费。（如 DeepSeek 开启后付费未生效，请 [提交工单](#) 联系我们）
- 子账号默认无后付费设置功能权限，支持主账号在 [用户权限管理](#) 中为子账号开通“后付费设置”权限。
- DeepSeek-V3 与 DeepSeek-V3-0324价格相同，使用相同计费项。

- DeepSeek-R1 与 DeepSeek-R1-0528价格相同，使用相同计费项。

## 自定义模型-大模型 tokens 资源包

定义：一次性购买一定数量的 tokens 资源包，有效期内调用模型服务时优先抵扣资源包当中的 tokens 余量，tokens 资源包如果到期未用完，会当作过期作废处理。

资源包有效期：1年。1年后未使用的资源包清零。

付费方式：预付费，购买 tokens 资源包越大，单价越低。支持在 [产品购买页](#) 中选择大模型 tokens 资源包-自定义模型购买。

适用范围：稳定调用，具有一定规模的业务体量。

tokens 资源包	自定义模型
1000万 tokens	¥115
5000万 tokens	¥570
1亿 tokens	¥1,130
5亿 tokens	¥5,600
10亿 tokens	¥11,100

## 自定义模型-大模型按量后付费

定义：自定义模型除了提供预付费的 tokens 资源包外，还提供按量后付费的计费方式。按量后付费根据实际调用接口产生的 tokens 数量计费，不同的模型价格不同，按照实际消耗的输入、输出总 tokens 数量，**每小时**对腾讯云账户进行扣费。

模型版本	单价
自定义模型	输入：0.0115元/千tokens 输出：0.0115元/千tokens

### ⚠ 注意：

1. 支持在 [后付费设置](#) 中开启后付费。（选择应用服务-自定义模型开启后付费）
2. 子账号默认无后付费设置功能权限，支持主账号在 [用户权限管理](#) 中为子账号开通“后付费设置”权限。
3. 自定义模型收费范围：
  - 问答过程中，在应用配置中测试、应用发布后调用应用 API / 使用体验链接、应用评测、应用体验，使用知识问答、联网搜索、大模型回复消耗的tokens 数。
  - workflow大模型知识问答节点和知识检索节点消耗的 tokens 数。

## 大模型专属并发包

定义：同时进行的会话数量，从请求发起到流式返回全部结果的整个过程都算在占用并发。

付费方式：包月包年。

模型版本	并发数量	包月	包年	备注
精调知识大模型标准版	1并发	¥17,000	¥200,000	可按照1、2、5并发购买
	2并发	¥24,000	¥280,000	
	5并发	¥34,000	¥400,000	
精调知识大模型高级版	1并发	¥55,000	¥660,000	
	2并发	¥72,000	¥860,000	
	5并发	¥100,000	¥1,200,000	
DeepSeek-R1	1并发	¥11,000	¥130,000	
	5并发	¥54,000	¥645,000	
DeepSeek-V3	1并发	¥8,000	¥95,000	
	5并发	¥39,000	¥465,000	

### 📌 购买说明：

购买 **专属并发包** 可独立使用，无需购买 tokens 资源。购买前需要联系对接的客户经理确认资源部署情况，如无对接的客户经理可联系 [售前咨询](#)。

## 知识库容量

定义：在智能体开发平台的应用中上传文档和问答，知识库容量计算用户所有应用上传的文档和问答的总字符数，删除的文档和应用不占用知识库容量。

资源有效期：可选，即将过期如需继续使用请续费。已过期的知识库仅支持删除操作。

付费方式：包年包月。

知识库扩容包	包月	包年
1000万字符	¥160	¥1,800
1亿字符	¥1,400	¥16,000

**⚠ 注意：**

- 字符数不计算文档中的空格，文档中存在图片的将转存为图片链接，每张图大约换算为150个字符。
- 一篇3万字的硕士论文，大约相当于4万字符，50页文档。1000万字符约为250篇硕士论文，1亿字符约为2500篇硕士论文。

**知识库扩容包到期说明：**

- 到期后，知识库中的知识将被处理成“超量失效”状态，“超量失效”状态的知识在对话中不生效，且“超量失效”状态的知识30天后将被自动删除，如需要继续使用知识库，请在到期前及时续费。
- “超量失效”状态的知识在购买知识库扩容包后，需要手动批量恢复并发布后，可支持正式环境对话中使用。

## 搜索服务

智能体开发平台提供搜索服务，如您在应用中开启了联网搜索服务，则每次调用搜索引擎将会扣减搜索服务次数。

**预付费资源包：**搜索服务资源包根据业务量级划分为不同规格，可一次性付费购买，自购买日起一年内有效，一年内若资源包未使用完，则过期作废。

资源包规格	价格
10万次	¥2,200
50万次	¥11,000
100万次	¥20,000
500万次	¥90,000
1000万次	¥170,000

**后付费：**支持开启后付费使用，在 [后付费设置](#) 中选择“应用服务\_搜索服务后付费”开启开关。（账户欠费无法使用，开启前请关注账户余额及时 [充值](#) 。）

收费项	单价
搜索服务后付费	22 元/千次

# 购买方式

最近更新时间：2025-05-20 00:00:52

## 预付费/包年包月

- 在购买腾讯云智能体开发平台应用产品服务之前，您需要先 [注册腾讯云账号](#) 并通过实名认证，详情请参见 [注册腾讯云](#)。
- 腾讯云智能体开发平台应用产品服务，如有需要单击 [售前咨询](#) 联系产品售前架构师或单击 [立即购买](#)。

### 选择配置

采购类型	<a href="#">大模型Tokens资源包</a>	<a href="#">大模型共享并发包</a>	<a href="#">大模型专属并发包</a>	<a href="#">知识库扩容包</a>	<a href="#">搜索服务包</a>
模型选择	<a href="#">精调知识大模型标准版</a>	<a href="#">精调知识大模型高级版</a>	<a href="#">医学行业大模型标准版</a>	<a href="#">金融行业大模型标准版</a>	<a href="#">教育行业大模型标准版</a>
	<a href="#">混元大模型Turbo版</a>	<a href="#">教育行业大模型高级版</a>			
可选套餐包	1000万tokens 有效期自购买之日一年内		120.00元	- 0 +	
	5000万tokens 有效期自购买之日一年内		600.00元	- 0 +	
	1亿tokens 有效期自购买之日一年内		1180.00元	- 0 +	

配置费用 0.00元 [立即购买](#)

- 选择采购类型、模型及具体套餐包规格后进行下单，完成支付订单发货成功后即可使用，查看 [订单状态](#)。

## 后付费

智能体开发平台应用产品支持开通后付费使用，可在 [后付费设置](#) 开启计费开关后使用。

- 应用管理
- 原子能力
- 插件中心
- 体验中心
- 系统管理
- 计费管理
- 调用统计

## 计费管理

资源包 并发扩展 知识库容量扩展 后付费设置

原子能力后付费设置已迁移到原子能力控制台。 [点击查看](#)

1. 开启后，资源包耗尽时，以“后付费方式”按小时从账户余额中扣费，欠费时无法使用， [查看账户余额](#)（各开关每小时可修改一次）
2. 扣费顺序：体验资源包>付费资源包>后付费，可在 [控制台-费用中心](#) 查看账单

子产品名称	启用状态	生效状态	操作
应用服务_精调知识大模型标准版_token后付费	<input type="checkbox"/>	未启用	
应用服务_精调知识大模型高级版_token后付费	<input type="checkbox"/>	未启用	
应用服务_医学行业大模型标准版_token后付费	<input type="checkbox"/>	未启用	

### 注意：

1. 开启后，资源包耗尽时，以“后付费方式”按小时从账户余额中扣费，欠费时无法使用，单击 [查看账户余额](#)（修改上限：每小时1次）
2. 扣费顺序：体验资源包 > 付费资源包 > 后付费，可在 [控制台 > 费用中心](#) 查看账单，每小时对腾讯云账户进行扣费并推送账单明细。
3. 子账号默认无“后付费设置”的功能权限，需要主账号在 [用户权限管理](#) 中为子账号开通“后付费设置”权限。

# 续费说明

最近更新时间：2025-05-20 00:00:52

若您购买的资源耗尽或到达产品有效期时，请进入[腾讯云智能体开发平台](#) [产品购买页](#) 点击购买，进行续费。选择并确认模型版本、资源包容量后，进行下单，完成支付后续费成功。

## 注意：

包年包月资源仅到期当天支持续费，资源到期后将进行回收，如需继续使用请 [点击购买](#)，进行续费。

# 退费说明

最近更新时间：2025-05-20 00:00:52

腾讯云智能体开发平台提供两类计费项：**按量计费**和**包年包月类计费项**。相关退费说明见下：

## 1. 按量计费类计费项：包括应用服务 token 资源包和搜索服务包。

### 退费规则：

- 正式采购7天内（含7天）支持退费：
  - 基于**资源包剩余比例**计算退款金额。
  - 退款金额 = 新购折后价格 \* （剩余量/总量）
- 超过7天不支持退费。

## 2. 包年包月类计费项：包括应用服务的并发包及知识库扩容包。

### 退费规则：

- 正式采购7天内（含7天）支持退费：
  - 基于**使用时长折算**（以天为单位计算，不满24小时记为一天）退费金额。
  - 退款金额 = 支付金额 - （已使用时长 ÷ 总时长） × 订单原价 × 适用折扣
- 超过7天不支持退费。

## 退费流程

发起退费前，请先确定是否满足退费说明的条件，并确保基于腾讯云智能体开发平台服务产生的一切数据、资料已完成备份、迁移。若您未进行备份、迁移，在退费后产生的任何损失由您自行承担，与腾讯云无关。

1. 登录腾讯云官网，点击 [售后支持](#) 页面，反馈需要退款的产品并描述退款原因，腾讯云客服人员将会在线处理您提交的申请。
2. 您可在 [订单管理](#) 页面查看退货订单，订单状态为“已退费”，可在 [费用中心](#) 页面查看款项。若审核不通过，您可在工单里查看审核结果。

### ⓘ 说明：

发起退款流程前，请您查看资源是否属于可退费范畴，若不符合，则无法退还。