

Deepseek:

账号注册与创建API

进入到Deepseek官网: <https://www.deepseek.com/>

点击左上方的API开放平台 (英文界面为API Platform)



1. 在此登陆或注册一个账号



Code login Password login

Only login via phone number, WeChat, email is supported in your region.

+86 Phone number

Code

Send code

I confirm that I have read, consent and agree to DeepSeek's [Terms of Use](#) and [Privacy Policy](#).
New phone numbers will be automatically registered.

Log in

OR

Log in with WeChat

2. 在此进入到用户界面，点击左边侧栏的API keys

The screenshot shows the DeepSeek Platform dashboard. On the left, there is a sidebar with various navigation options: Usage, API keys (which is highlighted with a red arrow), Top up, Billing, Docs, FAQ, Pricing, Integration, Contact us, Profile, Chat for free, Service Status, Terms of Use, Privacy Policy, and two Chinese license numbers. The main content area is titled "Usage". It displays financial information: Topped-up balance (¥5.95 CNY), Granted balance (¥0.00 CNY), and Monthly expenses (¥0.04 CNY). Below this is a "Monthly usage" section with a bar chart for "deepseek-chat" showing API requests (9) and Tokens (392). The chart has a Y-axis from 0 to 5 and an X-axis from 3-1 to 3-31. The bar for March 1st is blue, and the bar for March 21st is yellow.

3. 点击新建一个API Key

API keys

Your API keys are listed below. The API key is only visible and can be copied once at creation. Save it securely. Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, DeepSeek may also automatically disable any API key that we've found has leaked publicly. Usage of API keys created before April 25, 2024, was not tracked.

| Name | Key | Created | Last used | Actions |
|------------|-------------------|------------|------------|---------|
| demo1(mac) | sk-732ad*****9a99 | 2025-03-10 | 2025-03-10 | |

Create new API key



4. 为创建的key命名

Create new API key



Name

Name of the API key

Cancel

Create API key

5. 注意！！！ 这里一定要将API key复制并保存好，一旦点击Done关闭了窗口就无法再看到这个API key了，丢失了只能重新创建！！！

Create new API key

X

Please save this API key somewhere safe and accessible. For security reasons, you won't be able to view it again through your account. If you lose this API key, you'll need to generate a new one.

sk-b204c34ef4534ad4882a92b3f1f6834f

Done

Copy

账户充值与查看Token用量

- 接下来点击左边侧栏中的Top up向账户中充值，可以点击蓝色箭头所指的Custom自定义充值金额（比如一块钱）

The screenshot shows the DeepSeek Platform interface. On the left, there is a sidebar with various links: Usage, API keys, Top up (highlighted with a red arrow), Billing, Docs, FAQ, Pricing, Integration, Contact us, Profile, Chat for free, Service Status, Terms of Use, and Privacy Policy. The main content area is titled 'Top up'. It has two tabs: 'Online recharge' (selected) and 'Bank transfer'. Below the tabs is a 'Amount' section with a 'Custom' button highlighted by a blue arrow. A note below says 'Off-Peak Discounts: DeepSeek-V3 with 50% off and DeepSeek-R1 with 75% off at off-peak hours (16:30-00:30 UTC daily). Optimize your workflow while enjoying these exclusive savings. [View Details]'. There is a 'Payment method' section with 'ALIPAY' selected (indicated by a checkmark). Below that is a 'WeChat Pay' option. At the bottom are 'Next step' and 'Note' sections. The 'Note' section contains two points: 1. The recharge amount is solely for API service usage. Conversations on the web version and app are free and do not require any recharge. 2. After recharging, you can visit the billing page to request an invoice.

- 充值完成后回到用户界面，可以看到账户中的余额

deepseek Platform

Usage

All dates and times are UTC-based, and data may be delayed up to 5 minutes.

Topped-up balance **¥5.95 CNY**

Granted balance [Expiry date](#) **¥0.00 CNY**

Monthly expenses **¥0.04 CNY**

[Top up](#) [Balance alert enabled \(Settings\)](#)

Monthly usage

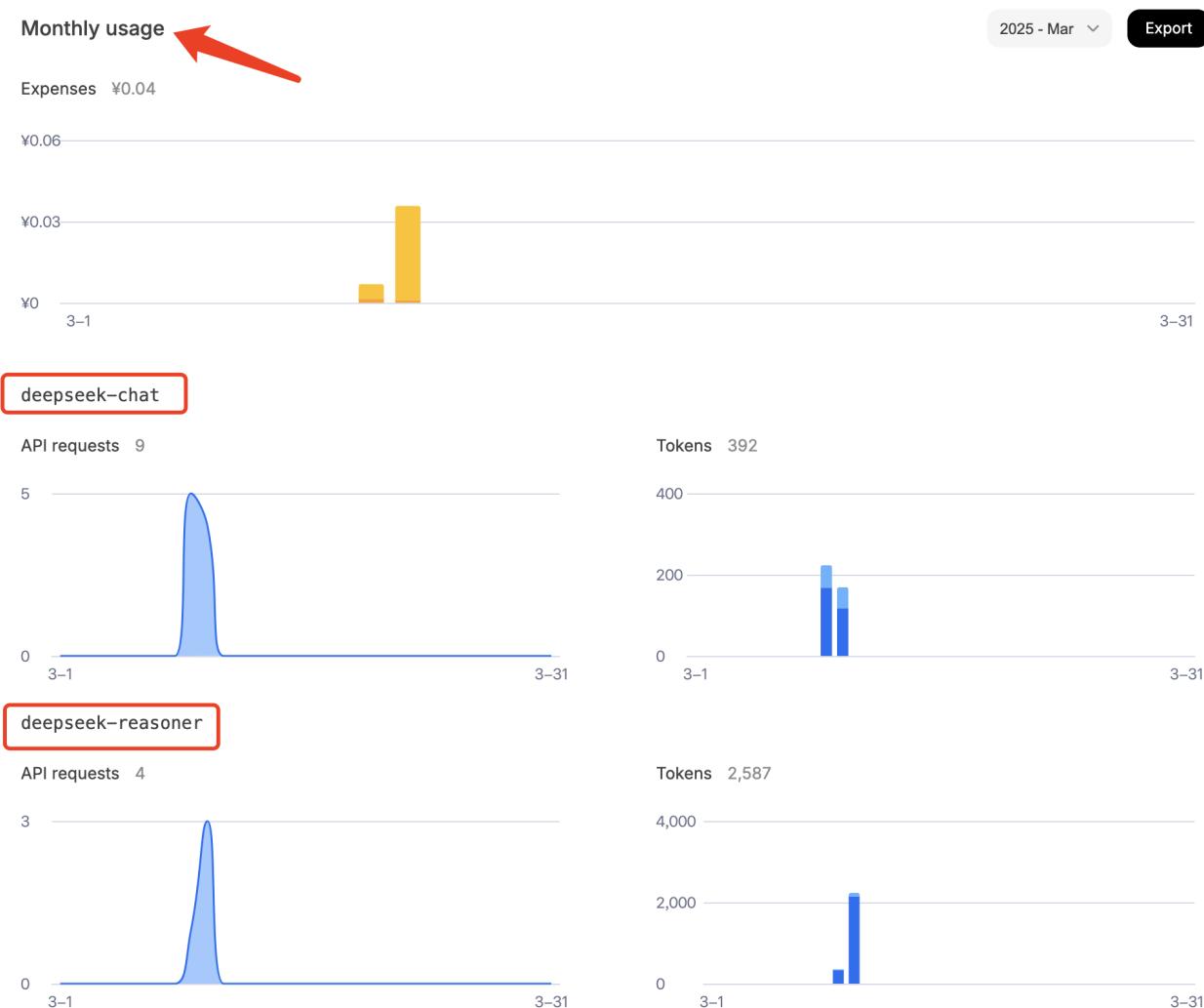
2025 - Mar [Export](#)

Docs, FAQ, Pricing, Integration, Contact us, Profile, Chat for free ↗

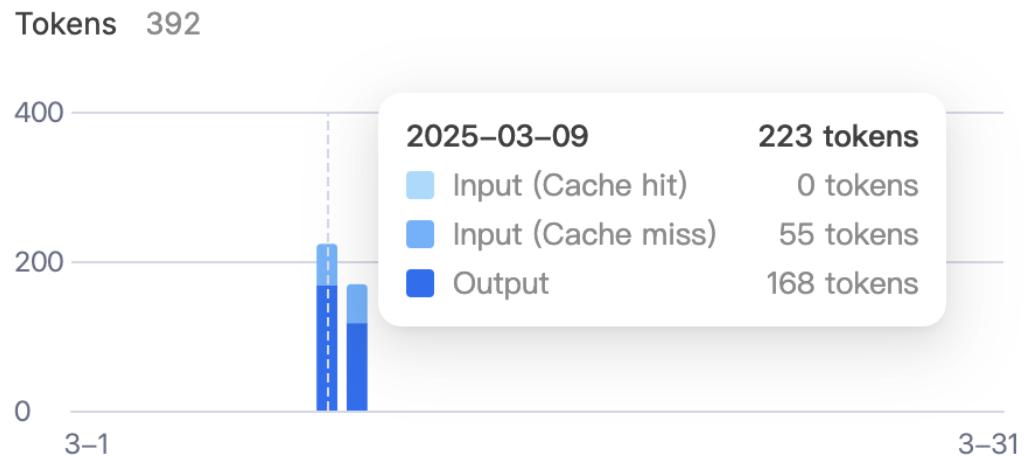
Service Status · Terms of Use · Privacy Policy ·

This screenshot shows the 'Usage' section of the deepseek Platform. It displays a summary of financials: Topped-up balance (¥5.95 CNY), Granted balance (¥0.00 CNY), and Monthly expenses (¥0.04 CNY). Below this, there's a chart titled 'Monthly usage' for March 2025. The chart has two sections: 'deepseek-chat' (API requests: 9, Tokens: 392) and 'deepseek-reasoner' (API requests: 4, Tokens: 2,587). The x-axis represents the month from 3-1 to 3-31. A red arrow points to the 'Monthly usage' title.

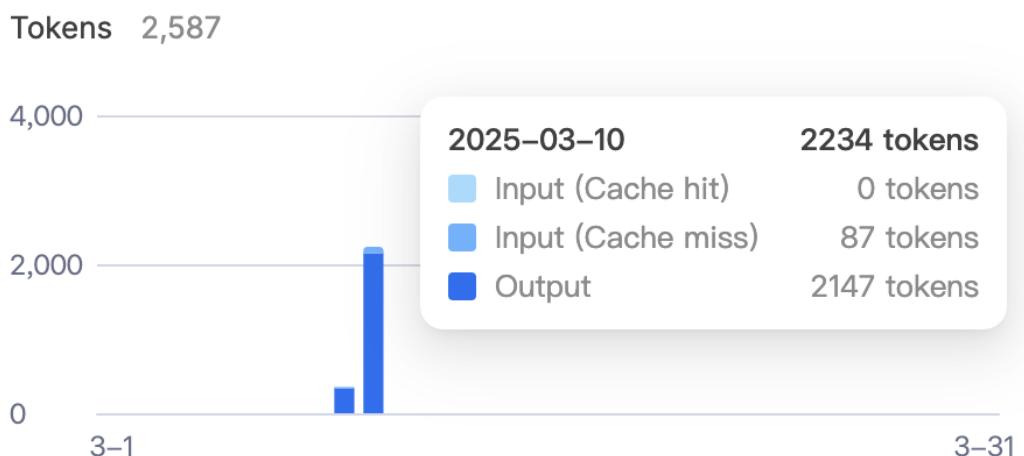
3. 往下看，可以看到API调用的用量，上方左图是deepseek-chat模型的调用次数，右图是token的用量，将光标移动到柱形图上可以显示出具体的token吞吐



此为chat模型的token调用

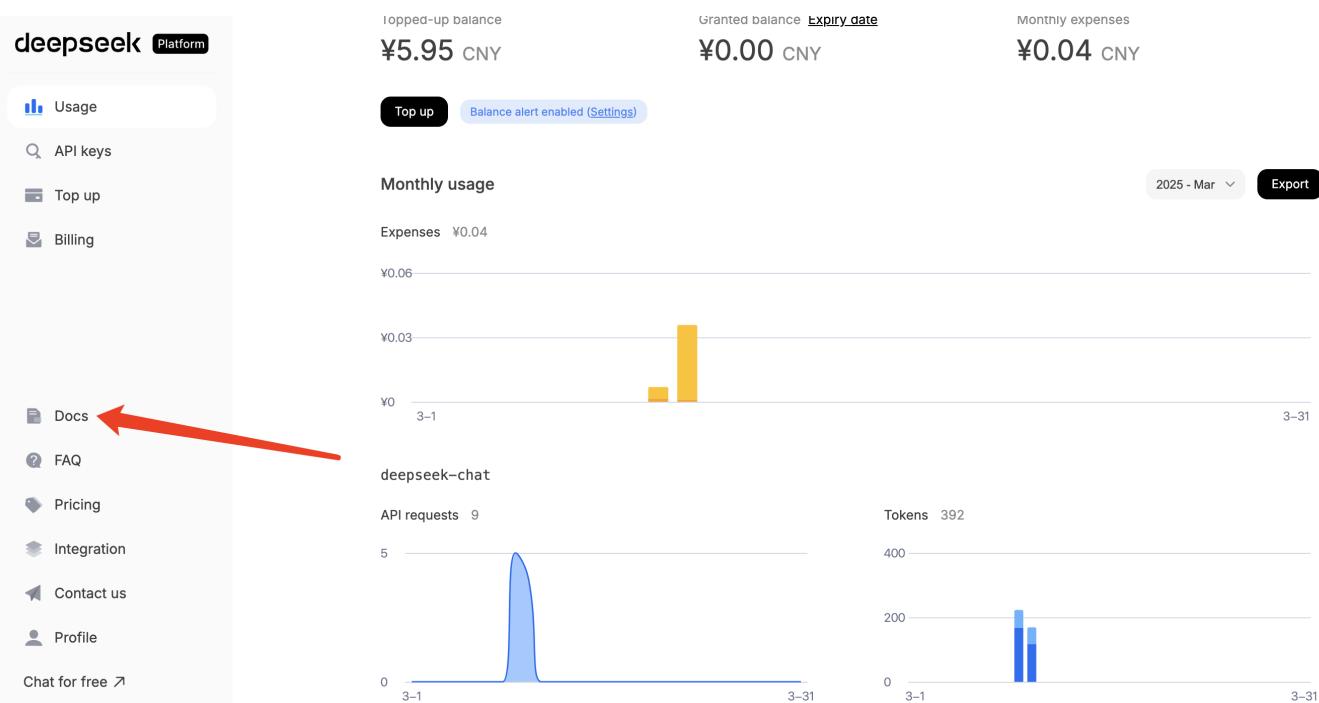


此为R1 (reasoner) 模型的token调用，可以明显看出R1模型的token输出非常多，这是因为思考过程也是算在token输出里的



使用指南

1. 点击左边侧栏的Docs



2. 切换成中文

The screenshot shows the DeepSeek API Docs page. The left sidebar includes sections like Quick Start, Your First API Call, Models & Pricing, The Temperature Parameter, Token & Token Usage, Rate Limit, Error Codes, News, and a list of releases. The main content area is titled "Your First API Call". It explains the API's compatibility with OpenAI and provides a table for configuration parameters:

| PARAM | VALUE |
|------------|---|
| base_url * | https://api.deepseek.com |
| api_key | apply for an API key |

Below the table are several notes and a section titled "Invoke The Chat API". A red arrow points to the "English" dropdown at the top right of the page.

3. 在此可以看到几个关键信息

DeepSeek API 文档

首次调用 API

DeepSeek API 使用与 OpenAI 兼容的 API 格式，通过修改配置，您可以使用 OpenAI SDK 来访问 DeepSeek API，或使用与 OpenAI API 兼容的软件。

| PARAM | VALUE |
|------------|---------------------------------------|
| base_url * | <code>https://api.deepseek.com</code> |
| api_key | apply for an API key |

* 出于与 OpenAI 兼容考虑，您也可以将 `base_url` 设置为 `https://api.deepseek.com/v1` 来使用，但注意，此处 v1 与模型版本无关。

* `deepseek-chat` 模型已全面升级为 DeepSeek-V3，接口不变。通过指定 `model='deepseek-chat'` 即可调用 DeepSeek-V3。

* `deepseek-reasoner` 是 DeepSeek 最新推出的推理模型 DeepSeek-R1。通过指定 `model='deepseek-reasoner'`，即可调用 DeepSeek-R1。

调用对话 API

在创建 API key 之后，你可以使用以下样例脚本的来访问 DeepSeek API。样例为非流式输出，您可以将 stream 设置为 true 来使用流式输出。

4. 该页面往下，选择python。

1. 红色方框中的<>内需要输入刚刚查看到的API keys。
2. 蓝色方框中为模型选择，"deepseek-chat"即是deepseek的普通版本，"deepseek-reasoner"即是Deepseek-R1
3. 绿色方框中的内容为您给AI设定的角色，例如“你是一个有用的助手”。然后设定测试提问词，例如“你好”

在创建 API key 之后，你可以使用以下样例脚本的来访问 DeepSeek API。样例为非流式输出，您可以将 stream 设置为 true 来使用流式输出。

curl **python** nodejs

```
# Please install OpenAI SDK first: `pip3 install openai`

from openai import OpenAI

client = OpenAI(api_key="<DeepSeek API Key>", base_url="https://api.deepseek.com")

response = client.chat.completions.create(
    model="deepseek-chat",
    messages=[
        {"role": "system", "content": "You are a helpful assistant"},
        {"role": "user", "content": "Hello"},
    ],
    stream=False
)

print(response.choices[0].message.content)
```

5. 调用价格和Token用量计算

Token 用量计算

token 是模型用来表示自然语言文本的基本单位，也是我们的计费单元，可以直观的理解为“字”或“词”；通常 1 个中文词语、1 个英文单词、1 个数字或 1 个符号计为 1 个 token。

一般情况下模型中 token 和字数的换算比例大致如下：

- 1 个英文字符 ≈ 0.3 个 token。
- 1 个中文字符 ≈ 0.6 个 token。

但因为不同模型的分词不同，所以换算比例也存在差异，每一次实际处理 token 数量以模型返回为准，您可以从返回结果的 `usage` 中查看。

离线计算 Tokens 用量

您可以通过如下压缩包中的代码来运行 tokenizer，以离线计算一段文本的 Token 用量。

[deepseek_v3_tokenizer.zip](#)

模型 & 价格

下表所列模型价格以“百万 tokens”为单位。Token 是模型用来表示自然语言文本的最小单位，可以是一个词、一个数字或一个标点符号等。我们将根据模型输入和输出的总 token 数进行计量计费。

模型 & 价格细节

| 模型(1) | deepseek-chat | deepseek-reasoner |
|---|---------------------------------|-------------------|
| 上下文长度 | 64K | 64K |
| 最大思维链长度(2) | - | 32K |
| 最大输出长度(3) | 8K | 8K |
| 标准时段价格 (北京时间 08:30-00:30) | 百万tokens输入（缓存命中） ⁽⁴⁾ | 0.5元 |
| | 百万tokens输入（缓存未命中） | 2元 |
| | 百万tokens输出 ⁽⁵⁾ | 8元 |
| 优惠时段价格 ⁽⁶⁾ (北京时间 00:30-08:30) | 百万tokens输入（缓存命中） | 0.25元 (5折) |
| | 百万tokens输入（缓存未命中） | 1元 (5折) |
| | 百万tokens输出 | 4元 (5折) |

1. `deepseek-chat` 模型对应 **DeepSeek-V3**; `deepseek-reasoner` 模型对应 **DeepSeek-R1**。
2. 思维链为 `deepseek-reasoner` 模型在给出正式回答之前的思考过程，其原理详见[推理模型](#)。
3. 如未指定 `max_tokens`，默认最大输出长度为 4K。请调整 `max_tokens` 以支持更长的输出。
4. 关于上下文缓存的细节，请参考[DeepSeek 硬盘缓存](#)。
5. `deepseek-reasoner` 的输出 token 数包含了思维链和最终答案的所有 token，其计价相同。
6. DeepSeek API 现实行错峰优惠定价，每日优惠时段为**北京时间 00:30-08:30**，其余时间按照标准价格计费。请求的计价时间为该请求完成的时间。
6. Temperature的设置，可以根据需求自行调整。

Temperature 设置

`temperature` 参数默认为 1.0。

- 我们建议您根据如下表格，按使用场景设置 `temperature`。

| 场景 | 温度 |
|------------|-----|
| 代码生成/数学解题 | 0.0 |
| 数据抽取/分析 | 1.0 |
| 通用对话 | 1.3 |
| 翻译 | 1.3 |
| 创意类写作/诗歌创作 | 1.5 |

Volces

账号注册与充值与创建API

进入火山方舟官网：<https://www.volcengine.com/product/ark>

1. 点击控制台



2. 在此登陆或注册一个账号（第一次注册的账号需要实名认证）



3. 在此充值（与Deepseek一样，可以只充一块钱）

火山引擎 总览

搜索产品或文档

企业 工具 费用 支持 备案 文档

2

快捷导航

最近访问 火山方舟

自定义快捷入口 + 添加入口

费用信息

可用余额 **¥11.00**

月账单概览

费用中心 >

充值 收支明细

代金券 (张) 0

可开发票 **¥0.00**

应付金额 欠费金额

2095手机用户#LNrxXx
ID: 2105048867

账号中心 >

主账号 2 已实名 未绑定邮箱

待支付订单 > 0 待续缴项 > 0 待处理工单 > 0

访问控制

用户 > 0 用户组 > 0 角色 > 2

新增用户 新增用户组 新增角色

尚未开通多云安全防护

75%的安全事件是云配置错误引发的，云上资产可能存在以下风险：

服务器异常 原因：安全组配置错误和密码设置不合理等

网站异常 原因：Web服务器配置错误、数据库配置错误等

数据丢失 原因：数据备份策略不当、存储配置错误或文件系统不稳定等

一键开通多云安全平台

限时免费

4. 充值完成后先将光标移到左上角火山引擎logo左侧的三条杠处以打开侧栏，然后点击“火山方舟”

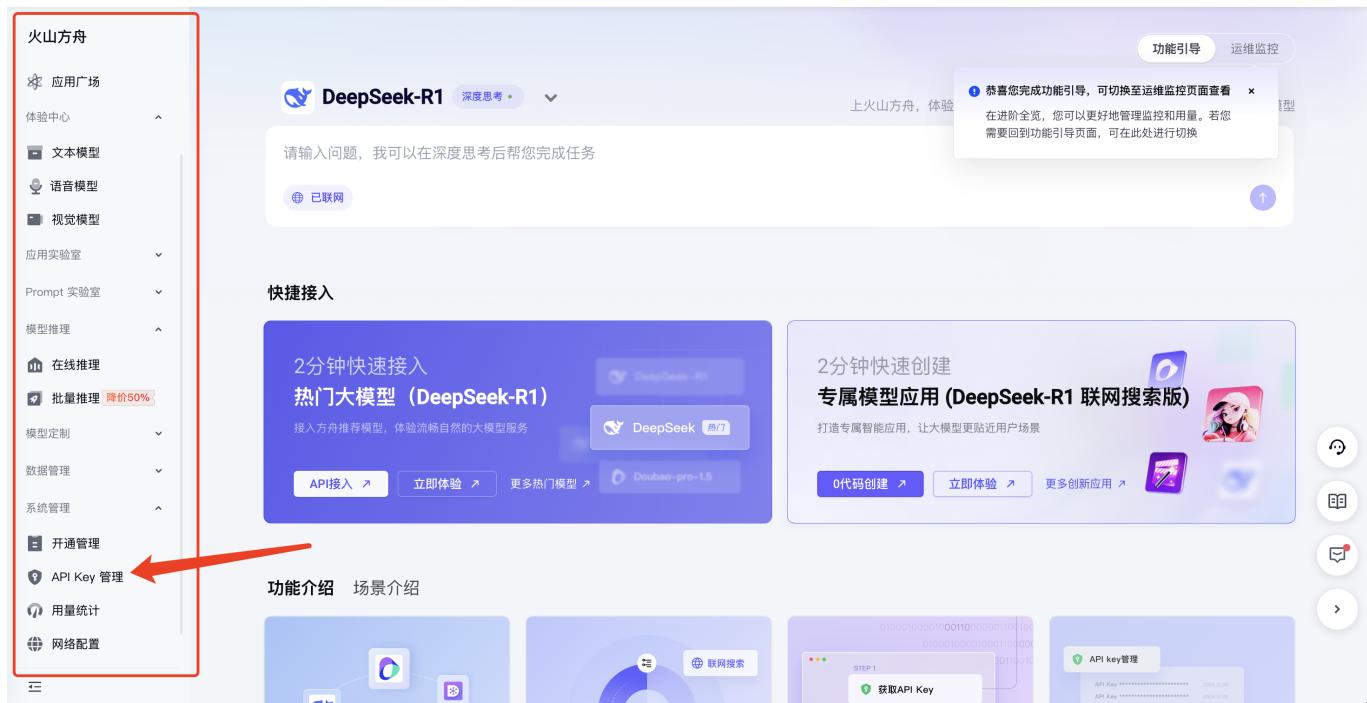
The screenshot shows the main interface of the Huoshan Engine Control Console. At the top, there is a navigation bar with links for '总览' (Overview), '企业' (Enterprise), '工具' (Tools), '费用' (Cost), '支持' (Support), '备案' (Record Filing), '文档' (Documentation), and a red-highlighted '控制台门户升级说明' (Console Portal Upgrade Instructions). Below the navigation bar is a search bar labeled '搜索产品或文档' (Search products or documents) and another search bar labeled '最近访问' (Recently visited) with the result '火山方舟' (Volcano Ark).

The left sidebar lists '热门产品' (Hot Products) including '火山方舟' (Volcano Ark), '扣子专业版' (Kappo Professional Edition), '云服务器' (Cloud Server), '对象存储' (Object Storage), and '域名服务' (Domain Name Service). A blue arrow points from the '火山方舟' link in the sidebar to its corresponding category in the main content area.

The main content area is organized into several sections:

- 计算 (Compute):** Includes '弹性计算' (Elastic Computing), 'GPU云服务器' (GPU Cloud Server), '弹性裸金属服务器' (Elastic Bare Metal Server), '弹性伸缩' (Elastic Scaling), and '操作系统' (Operating System) sections.
- 网络 (Network):** Includes '云上网络' (Cloud Network), '公网IP' (Public IP), '私有网络' (Private Network), 'NAT网关' (NAT Gateway), and '应用型负载均衡' (Application-type Load Balancing).
- 存储 (Storage):** Includes '基础存储服务' (Basic Storage Services), '对象存储' (Object Storage), '弹性块存储' (Elastic Block Storage), '文件存储 vePFS' (File Storage vePFS), '文件存储 NAS' (File Storage NAS), '日志服务' (Log Service), '大数据文件存储' (Big Data File Storage), and '数据迁移与工具' (Data Migration and Tools).
- 数据库 (Database):** Includes '关系型数据库' (Relational Database) sections for '云数据库 MySQL 版' (MySQL Edition), '云数据库 veDB MySQL 版' (veDB MySQL Edition), '云数据库 PostgreSQL 版' (PostgreSQL Edition), and '云数据库 SQL Server 版' (SQL Server Edition); and a 'NoSQL数据库' (NoSQL Database) section for '缓存数据库 Redis 版' (Redis Edition).
- 混合云网络 (Hybrid Cloud Network):** Includes '专线连接' (Point-to-Point Connection) and 'VPN连接' (VPN Connection).
- 容器与中间件 (Container and Middleware):** Includes '跨地域网络' (Cross-Region Network), '云企业网' (Cloud Enterprise Network), and '容器与中间件' (Container and Middleware) sections.
- 安全 (Security):** Includes '云安全' (Cloud Security) sections for 'DDoS高防' (DDoS High Defense), 'DDoS基础防护' (DDoS Basic Protection), 'Web应用防火墙' (Web Application Firewall), and '云安全中心' (Cloud Security Center).
- 其他 (Others):** Includes '控制台门户升级说明' (Console Portal Upgrade Instructions), '账号中心' (Account Center), '待处理工单' (Pending Work Orders) with count '0', '角色' (Role) with count '2', '新增角色' (Add Role), and '存在以下风险' (Exists the following risks) with a warning icon.

5. 左边侧栏往下滑，点击API Key管理



6. 在此界面创建API Key并点击蓝色箭头所指的“眼睛”以查看API Key，将API Key复制并保存。与Deepseek不同的是，火山方舟可重新查看API Key。

| 名称 | API Key | 创建时间 | 权限 | 操作 |
|------------------------|---------------|------------------|-----|--------------------|
| api-key-20250309104726 | *****@#\$%^&* | 2025-03-09 10:47 | all | 删除 |

服务开通与使用指南

1. 点击左侧栏中的“开通管理”，然后选择开通图中框中的两个模型：Doubao-1.5-lite-32k与Doubao-1.5-pro-32k

每个模型赠送50万tokens 推理额度 适用于「豆包」等语言大模型，快来体验吧

开通管理

火山方舟协作奖励计划二期全面升级

享个人权益 每日单模型 50万 免费tokens | 升级企业权益，每日单模型 500万 免费tokens | 活动时间：即日起至2025年5月31日

立即参与

Tr 大语言模型 **语音大模型** **视觉大模型** 切换百万tokens单位展示

| 模型名/提供方 | 状态 | 免费推理额度 | 推理(输入)定价 | 操作 |
|--|-----|--------------------------|----------------------|---|
| Doubao-1.5-vision-pro-32k 字节跳动 | 未开通 | 剩500,000/共500,000 tokens | 0.0030 元/千tokens | 0.0 开通服务 接入文档 |
| Doubao-1.5-lite-32k 字节跳动 | 已开通 | 剩500,000/共500,000 tokens | 0.0001 元/千tokens 折后价 | 0.0 关闭服务 接入文档 |
| Doubao-1.5-pro-32k 字节跳动 | 已开通 | 剩499,911/共500,000 tokens | 0.0004 元/千tokens 折后价 | 0.0 关闭服务 接入文档 |
| DeepSeek-R1-Distill-Owen-32B DeepSeek | 未开通 | 剩500,000/共500,000 tokens | 0.0015 元/千tokens | 0.0 开通服务 接入文档 |
| DeepSeek-R1-Distill-Owen-7B DeepSeek | 未开通 | 剩500,000/共500,000 tokens | 0.0006 元/千tokens | 0.0 开通服务 接入文档 |
| DeepSeek-R1 | | | | |

1 / 2 / 3 / 10条/页

2. 点击蓝色箭头所指的“接入文档”，然后点击左侧栏中的“开始使用”然后点击“产品简介”。在此可以看到各种模型的介绍(更详细的模型介绍可以点击左侧栏中的“模型能力”然后点击“模型列表”)

火山引擎 | 文档中心

火山方舟大模型服务平台

搜索目录或文档标题

开始使用

产品简介 (当前)

快速入门-调用模型服务

快速入门-速览平台功能

产品公告

模型能力

产品计费

API 参考

介绍

对话(Chat) API

对话(Chat)-文本 API

对话(Chat)-视觉 API

视频生成 API

向量化 API

最近更新时间：2025.03.12 14:44:14 首次发布时间：2023.06.27 23:47:27

在本产品文档中搜索

产品简介

为您提供大模型服务的开发平台，提供功能丰富、安全以及具备价格竞争力的模型调用服务，同时提供模型数据、精调、推理、评测等端到端功能，全方位保障您的 AI 应用开发落地。

方舟定位

选择并开通模型

使用模型前请访问[开通管理页面](#) 开通对应模型服务。

| doubao-1.5-pro | doubao-1.5-lite | doubao-1.5-vision |
|--|--|--|
| 最新一代专业版大模型，单价不提升的同时，模型能力有大幅提升，在知识、代码、推理、中文等相关的多项测评中获得高分，达到行业SOTA水平。 <ul style="list-style-type: none">最高支持 256k 最大上下文长度可配置 12k 最大输出长度高质量、低成本、极具性价比 | 最新一代轻量版大模型，模型版本效果比肩doubao 1.0 pro系列，超低延迟。每个token输出时间仅为10ms，将给汽车座舱、角色扮演等场景应用，带来全新的速度体验。 <ul style="list-style-type: none">支持 32k 最大上下文长度 / 50 张图片可配置 12k 最大输出长度更快、更便宜 | 最新一代视觉理解模型，升级多模态融合、动态分辨率、多模态对齐等技术，增强模型在视觉推理、文字识别、细粒度信息理解、指令遵循的能力，模型回复变得更加精简、友好。 <ul style="list-style-type: none">支持 32k 上下文长度 / 50 张图片可配置 12k 最大输出长度低成本、强能力 |

[模型列表](#) | [模型定价详情](#)

方舟定位

选择并开通模型

快速入门

实践教程

产品优势

文档反馈

3. 向下滑可以看到一个与Deepseek类似的调用代码。

快速入门

[快速入门-调用模型服务](#) | [直接体验](#)

帮助您发起第一次的大模型服务调用。

Python SDK | [Curl](#) | [OpenAI Python SDK](#)

```

Python
import os
from volcenginesdkarkruntime import Ark
# 获取并配置方舟API Key，并从环境变量中读取。https://www.volcengine.com/docs/82379/1399008#_1-%E7%9A%84%E5%8F%96%E9%80%91
client = Ark(api_key=os.environ.get("<ARK_API_KEY>"))
client = Ark(
    api_key=os.environ.get("<ARK_API_KEY>"),
    base_url="https://ark-cn-beijing.bytedance.net/api/v3",
)
completion = client.chat.completions.create(
    # 您的方舟推理接入点。
    # 替换为模型 ID，获取模型 ID: https://www.volcengine.com/docs/82379/1330310
    model="<Model>",
    messages=[
        {"role": "user", "content": "你好"}
    ]
)
print(completion.choices[0].message.content)

```

您也可将 `<Model>` 配置为推理接入点ID，来获取调用限流、购买保障包、监控调用明细等附加能力。详细请参见 [获取推理接入点 ID](#)。

4. 点击左侧栏中的“产品计费”，然后点击“模型服务计费”，可以看到火山方舟的收费标准

模型服务计费

最近更新时间: 2025.03.14 20:53:31 首次发布时间: 2023.06.27 23:50:07

[我的收藏](#) | [有用](#) | [无用](#)

不同模型支持的能力及单价各不相同，存在多种计费方式和多种计费逻辑。不同业务类型，适合的服务以及计费方式不同，本文会从计费项目、计费方式、计费逻辑等维度帮助您了解模型服务产生的费用，帮助您规划费用支出。

Token 概念

模型服务用量的一个重要概念是 token。在调用模型推理服务时，会将输入内容进行分词（tokenize），转化为模型可以理解的 token，经过模型处理后，同样输出 token，并转化为您需要的文本或者其他内容载体。而模型处理（包括输入、输出）的 token 数量会被作为模型推理服务用量的一个重要计量单位。如语言大模型推理服务费用的一种较为常见的计算方式如下：

```

费用 = token 使用数量 * token 单价

```

- 由于不同模型采用的分词策略不同，同一段文本可能会被转化为不同数量的 token。
- 您可以使用[Token计算器](#)或者[分词 API](#)接口，来估算内容对应的 token 数量。

产品价格

说明

下列价格仅供参考，请以[定价详情页](#)以及实际购买页价格为准。

在线推理

AutoDL

账号注册与充值

AutoDL官网地址: <https://www.autodl.com/home>



1. 账号创建，在此登陆或创建一个账号（第一次注册的账号需要实名认证）



2. 租赁服务器

AutoDL AI 算力云

弹性、好用、省钱

立即注册 了解详情

注册礼包 GPU选型 开具发票 新手入门

炼丹会员及租用价格

AutoDL坚持为您提供服务稳定、价格公道的GPU租用服务。更为学生提供免费升级会员通道，享极具性价比的会员价格。[如何升级会员？](#)

[炼丹会员](#) [普通用户](#)

选择一个有空闲的服务器

内蒙B区 / 001机 可租用至：2025-10-02

RTX 4090 / 24 GB 空闲/总量 0 / 8

每GPU分配 硬盘 其它

CPU: 16 核, Xeon(R) Platinum 8352V 系统盘: 30 GB GPU驱动: 560.35.03 ¥2.08/时 ¥2.10/时 9.5折

内存: 120 GB 数据盘: 50 GB, 可扩容 130 GB CUDA版本: ≤ 12.6 已租完

内蒙B区 / 022机 可租用至：2025-10-02

RTX 4090 / 24 GB 空闲/总量 6 / 8

每GPU分配 硬盘 其它

CPU: 16 核, Xeon(R) Gold 6430 系统盘: 30 GB GPU驱动: 560.35.03 ¥2.08/时 ¥2.10/时 9.5折

内存: 120 GB 数据盘: 50 GB, 可扩容 1500 GB CUDA版本: ≤ 12.6 6卡可租

内蒙B区 / 191机 可租用至：2025-10-01

RTX 4090 / 24 GB 空闲/总量 0 / 8

每GPU分配 硬盘 其它

CPU: 16 核, Xeon(R) Gold 6430 系统盘: 30 GB GPU驱动: 560.35.03 ¥2.08/时 ¥2.10/时 9.5折

先在镜像选择处选择基础镜像，然后根据图中所示选择框架。选择好框架后点击创建并开机

算力市场 / 创建实例 ① 严禁使用WebUI等算法生成违禁图片、严禁挖矿，一经发现立即封号！

计费方式: 按量计费 包日 包周 包月 计费规则

创建完主机后仍然可以转换计费方式。如选择按量计费，价格发生变动以实例开机时的价格为准。

| 选择主机: | 主机ID | 算力型号/显存 | 空闲GPU | 每GPU分配 | CPU型号 | 硬盘 | 最高CUDA | 价格(单卡) |
|---------------------------------------|------------------|---------|-----------------------|------------------------|-------------------------|------|--------------------|--------|
| <input checked="" type="radio"/> 121机 | RTX 4090 24GB | 1 / 8 | CPU: 16核 内存: 120GB | Xeon(R) Platinum 8352V | 数据盘: 50GB 可扩容: 923GB | 12.6 | ¥2.08/时 ¥2.10/时 | |

GPU数量: 框架名称 框架版本 Python版本 CUDA版本

| | | | |
|--------------|----------|---------------------|--------|
| PyTorch | > 2.5.1 | > 3.12(ubuntu22.04) | > 12.4 |
| TensorFlow | > 2.3.0 | > | |
| Miniconda | > 2.1.2 | > | |
| JAX | > 2.1.0 | > | |
| PaddlePaddle | > 2.0.0 | > | |
| TensorRT | > 1.11.0 | > | |
| Gromacs | > 1.10.0 | > | |
| Jittor | > 1.9.0 | > | |
| 镜像: | 1.8.1 | > | |

PyTorch / 2.5.1 / 3.12(ubuntu22.04) / 12.4

创建完成后仍然可以更换其他镜像

日常费用: ¥0.00/日 配置费用: ¥2.08/时 费用明细 取消 创建并开机

账户余额: ¥9.80

将光标移动到个人下拉界面，点击容器实例即可看到刚才租赁的服务器

算力市场 / 创建实例 ① 严禁使用WebUI等算法生成违禁图片、严禁挖矿，一经发现立即封号！

计费方式: 按量计费 包日 包周 包月 计费规则

创建完主机后仍然可以转换计费方式。如选择按量计费，价格发生变动以实例开机时的价格为准。

| 选择主机: | 主机ID | 算力型号/显存 | 空闲GPU | 每GPU分配 | CPU型号 | 硬盘 | 最高CUDA | 价格(单卡) |
|---------------------------------------|------------------|---------|-----------------------|------------------------|-------------------------|------|--------------------|--------|
| <input checked="" type="radio"/> 121机 | RTX 4090 24GB | 1 / 8 | CPU: 16核 内存: 120GB | Xeon(R) Platinum 8352V | 数据盘: 50GB 可扩容: 923GB | 12.6 | ¥2.08/时 ¥2.10/时 | |

GPU数量: 1

数据盘: 免费50GB 需要扩容

| | | | | | |
|---------------|-------|-------|------|------------|-----|
| 实例规格: | GPU型号 | CPU | 内存 | 系统盘 | 数据盘 |
| RTX 4090 * 1卡 | 16核心 | 120GB | 30GB | 免费50GB SSD | |

基础镜像 hot 我的镜像 没有我要的环境?

基础镜像包含常用基本软件，如：深度学习框架、Miniconda等。如需其他软件可创建后安装

PyTorch / 2.5.1 / 3.12(ubuntu22.04) / 12.4

创建完成后仍然可以更换其他镜像

炼丹师2095 [个人认证] ID: b55e95cb-ab35-4883-be54-80dc49528e40

炼丹会员 可用余额: ¥9.80 充值

冻结余额: ¥0.00

代金券: ¥0.00 容器实例: 1

退出登录

3. 服务器配置

点击JupyterLab

容器实例 实例连续关机15天会释放实例，实例释放会导致数据清空且不可恢复，释放前实例在数据在。

租用新实例 批量续费

订阅GPU通知 设置密钥登录 小程序管理实例 搜索实例名称/ID

| 实例ID / 名称 | 状态 | 规格详情 | 本地磁盘 | 健康状态 | 付费方式 | 释放时间/停机时间 | SSH登录 | 快捷工具 | 操作 |
|------------------------------------|-----|------------------------|------------------------|------|------------------|--------------------|-----------------------------|--|-------|
| 内蒙B区 / 241机 6302498b56-74ba2fed | 运行中 | vGPU-32GB * 1卡 查看详情 | 系统盘 0.62% 数据盘 0.00% | 正常 | 按量计费 余额不足24小时 | 关机15天后释放 设置定时关机 | 登录指令 ssh***** 密码***** | JupyterLab AutoPanel 实例监控 自定义服务 | 关机 更多 |

点开auodl-tmp文件夹，将给出的requirements.txt上传

文件 编辑 查看 运行 内核 标签页 设置 帮助

启动页

笔记本

Python 3 (ipykernel)

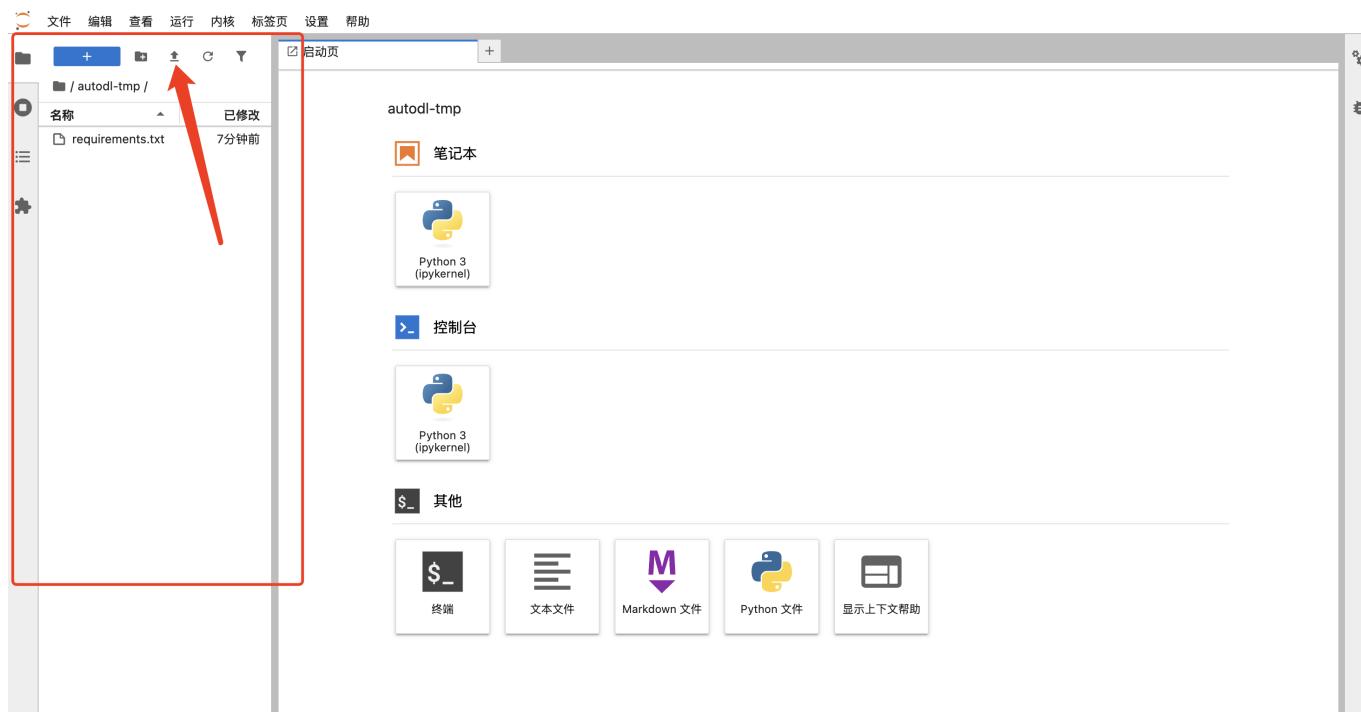
控制台

Python 3 (ipykernel)

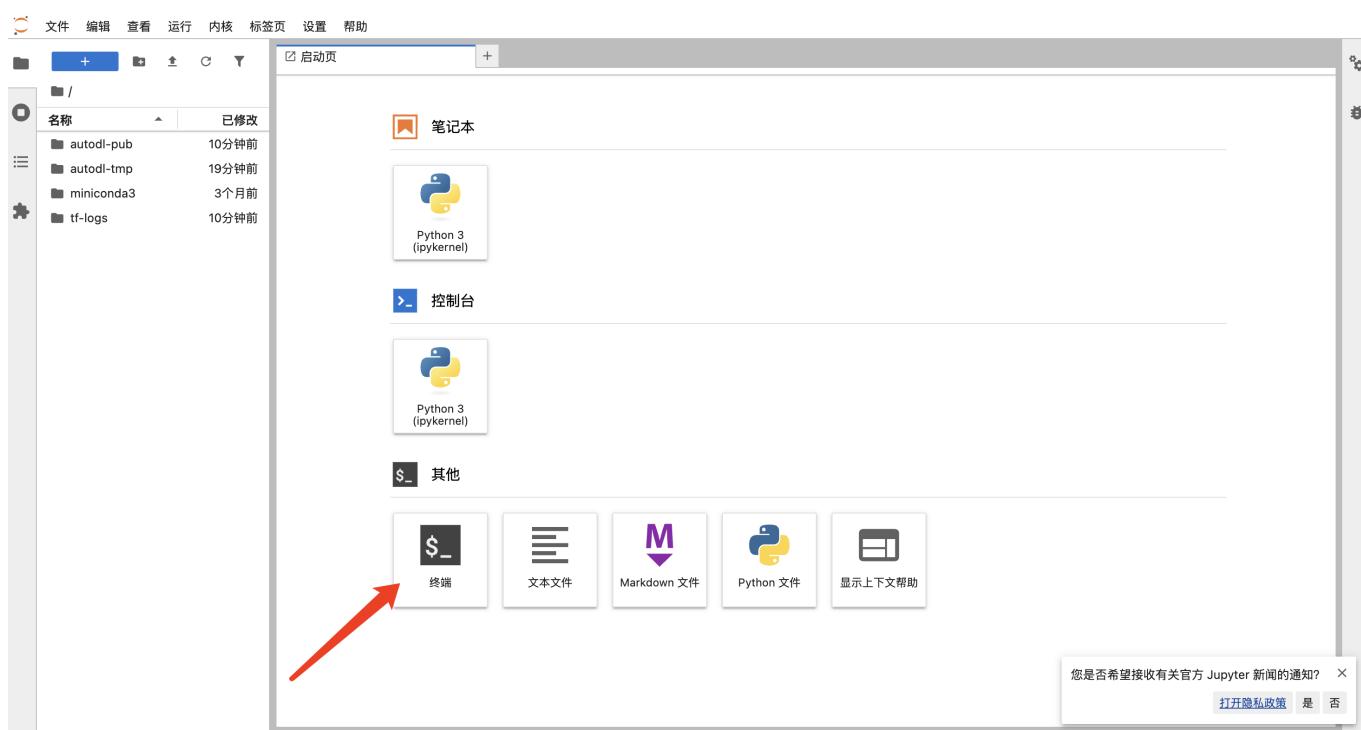
其他

终端 文本文件 Markdown 文件 Python 文件 显示上下文帮助

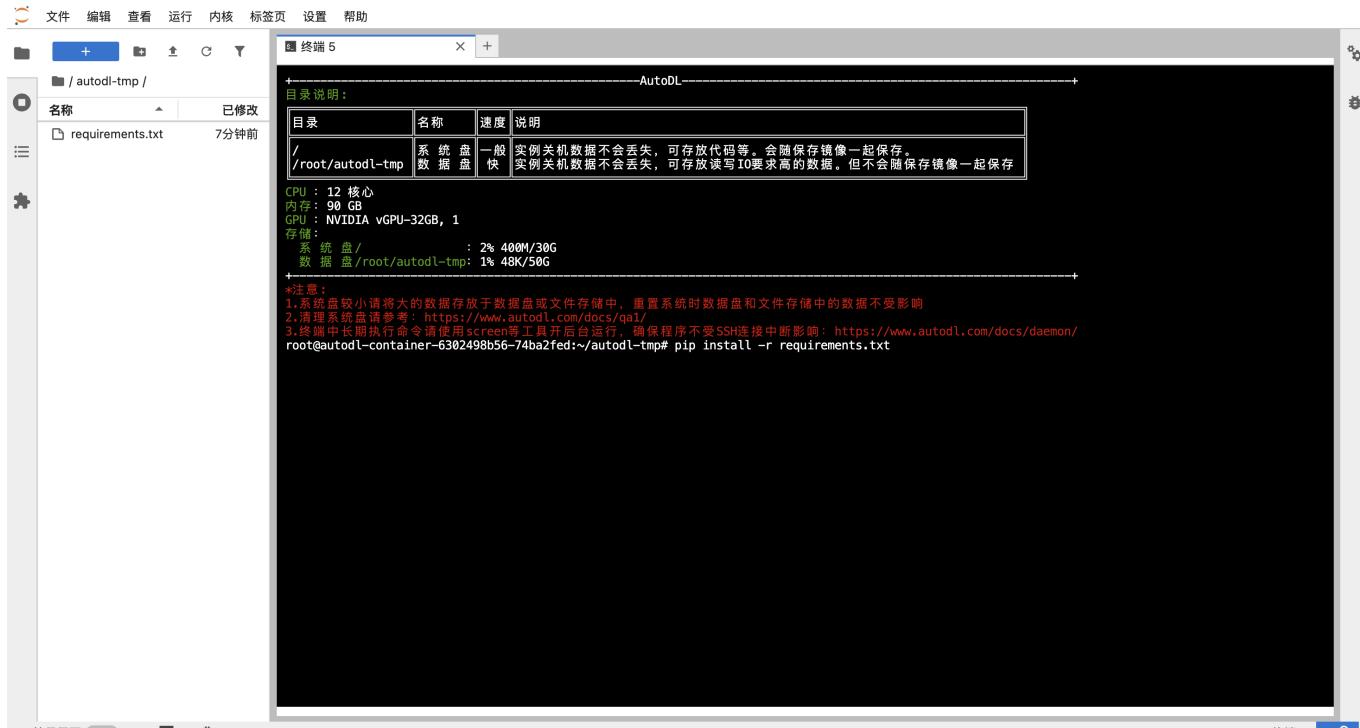
您可以将文件拖拽上来，也可以自行选择文件目录以上传



点击终端



在终端中输入 pip install -r requirements.txt 并按下回车键



至此，部署成功

