

Configure AI large model

Configure AI large model

1. Overseas Version of OpenRouter
OpenRouter Operation Steps
2. Overseas Version of Tongyi Qianwen
3. Local DIFY Configuration
 - 3.1 Enter the API key.

For large-scale model examples, an important prerequisite is connecting to a Wi-Fi network with internet access. For details on how to connect, see "Raspberry Pi Basic Settings" under "Raspberry Pi System Configuration" - 04. Network Settings for specific instructions.

This tutorial is about configuring large AI models. To run large model-related examples properly, follow the steps below.

1. Overseas Version of OpenRouter

Some online large models in the overseas version startup examples use the middleware of the OpenRouter platform, a third-party platform. This large model can be used 50 times per day for free. If you wish to continue using it, refer to the usage rules for expansion.

The following examples require this large model: scene description, dual-model intelligent agent application, intelligent action choreography, and free conversation.

OpenRouter does not have free models for video analysis or text-to-image creation, so models from the Tongyi Qianwen platform are required.

OpenRouter Operation Steps

1. Log in to the official website:

<https://openrouter.ai/>

2. Log in. You can use a variety of methods, but this article recommends using a GitHub account.

The Unified Interface For LLMs

Better **prices**, better **uptime**, no subscription.

Start a message...



Featured Models

[View Trending](#)

Gemini 2.5 Pro Preview

New

by google

83.1B

Tokens/wk

13.9s

Latency

+7.42%

Weekly growth

GPT-4.1

New

by openai

39.9B

Tokens/wk

584ms

Latency

+13.06%

Weekly growth

Claude 3.7 Sonnet

by anthropic

324.5B

Tokens/wk

1.5s

Latency

-3.37%

Weekly growth

7.9T

Monthly Tokens

1.9M

Global Users

50+

Active Providers

300+

Models

3. Go to the API-KEY page and create one.

Credits

Keys

Activity

Settings

Sign out

The Unified Interface For LLMs

Better **prices**, better **uptime**, no subscription.

Start a message...



Featured Models

[View Trending](#)

Gemini 2.5 Pro Preview

New

by google

83.1B

Tokens/wk

13.9s

Latency

+7.42%

Weekly growth

GPT-4.1

New

by openai

39.9B

Tokens/wk

584ms

Latency

+13.06%

Weekly growth

Claude 3.7 Sonnet

by anthropic

324.5B

Tokens/wk

1.5s

Latency

-3.37%

Weekly growth

7.9T

Monthly Tokens

1.9M

Global Users

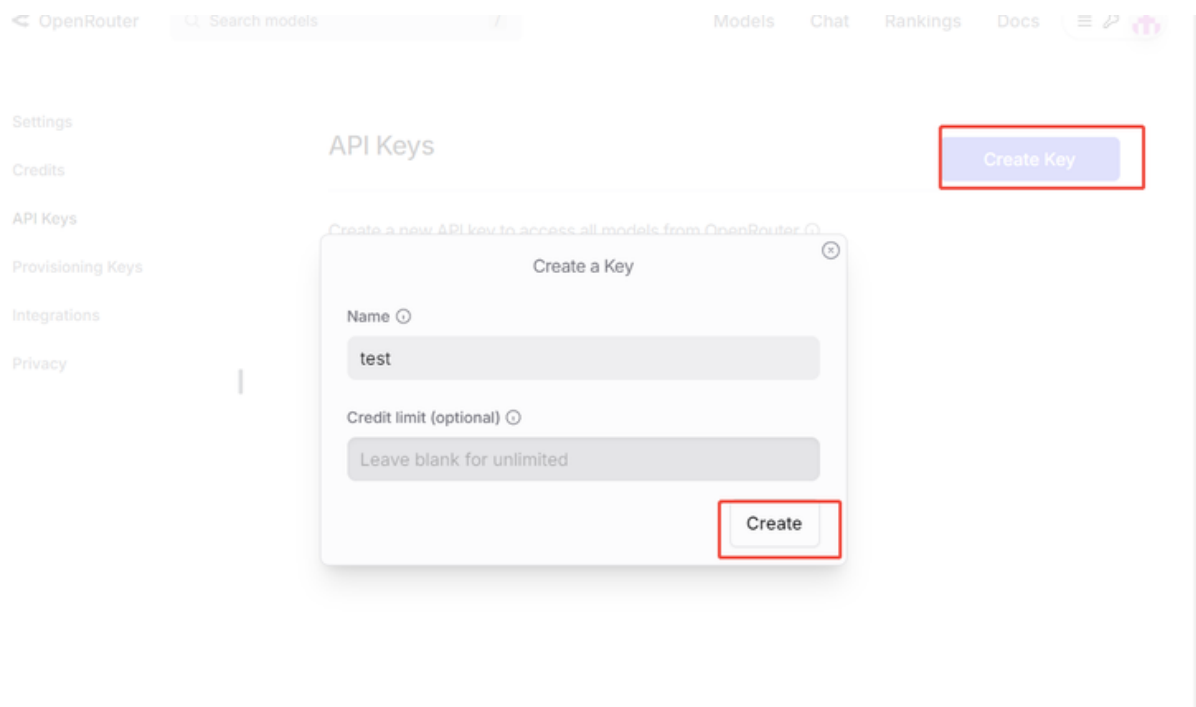
50+

Active Providers

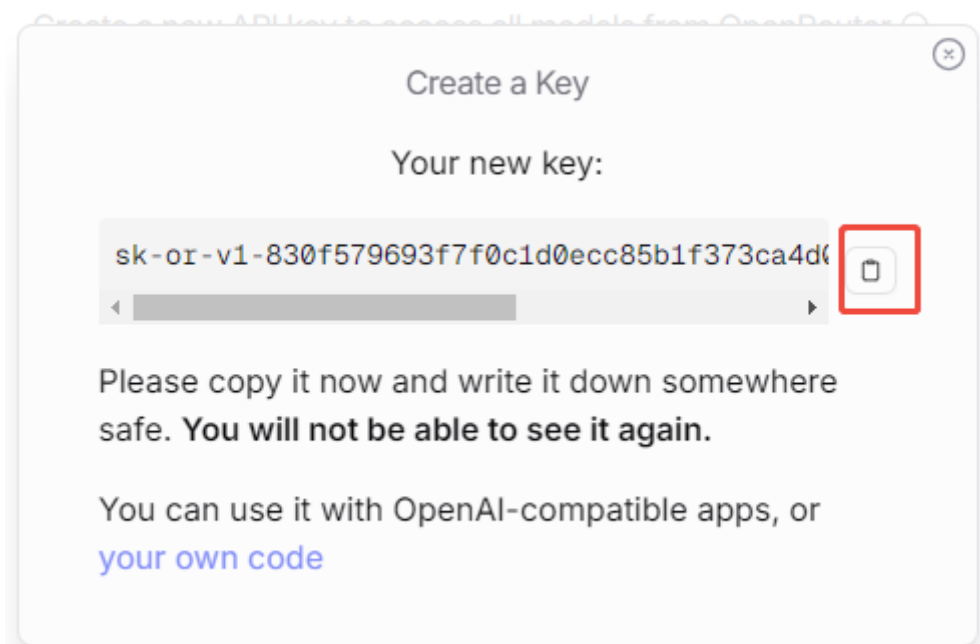
300+

Models

<https://openrouter.ai/settings/keys>



4. Copy and save the created key to your computer. **You can only copy this key the first time you create it;** you won't be able to view it again after leaving this page.



5. Next, we'll use a free test model to see if the registered account is functioning properly.

OpenRouter

Search models

/

ModelsChatRankingsDocs

The Unified Interface For LLMs

Better prices, better uptime, no subscription.

Start a message...>

Featured Models

Gemini 2.5 Pro Preview

by google

83.1B

13.9s

+7.42%

Tokens/wk

Latency

Weekly growth

GPT-4.1

by openai

39.9B

584ms

+13.06%

Tokens/wk

Latency

Weekly growth

Claude 3.7 Sonnet

by anthropic

324.5B

1.5s

-3.37%

Tokens/wk

Latency

Weekly growth

View Trending

7.9T

Monthly Tokens

1.9M

Global Users

50+

Active Providers

300+

Models

https://openrouter.ai/models

OpenRouter

Search models

/

ModelsChatRankingsDocs

Input Modalities

Text

Image

File

Context length

4K

64K

1M

Prompt pricing

FREE

\$0.5

\$10+

Series

GPT

Claude

Gemini

Models

59 models

Reset Filters

free

Sort

DeepSeek: DeepSeek V3 0324 (free)

98.8B tokens

DeepSeek V3, a 685B-parameter, mixture-of-experts model, is the latest iteration of the flagship chat model family from the DeepSeek team. It succeeds the DeepSeek V3 model and performs ...

by deepseek | 164K context | \$0/M input tokens | \$0/M output tokens

DeepSeek: R1 (free)

44.2B tokens

DeepSeek R1 is here: Performance on par with OpenAI o1, but open-sourced and with fully open reasoning tokens. It's 671B parameters in size, with 37B active in an inference pass. Fully open...

by deepseek | 164K context | \$0/M input tokens | \$0/M output tokens

Google: Gemini 2.0 Flash Experimental (free)

31.2B tokens

Gemini Flash 2.0 offers a significantly faster time to first token (TTFT) compared to Gemini Flash 1.5, while maintaining quality on par with larger models like Gemini Pro 1.5. It introduces notable ...

by google

6. If the response is normal, it means it's functioning properly.

OpenRouter Search models /

Models Chat Rankings Docs

DeepSeek: DeepSeek V3 0324 (free)

deepseek/deepseek-chat-v3-0324:free

Created Mar 24, 2025 | 163,840 context | \$0/M input tokens | \$0/M output tokens

DeepSeek V3, a 685B-parameter, mixture-of-experts model, is the latest iteration of the flagship chat model family from the DeepSeek team. It succeeds the [DeepSeek V3](#) model and performs really well on a variety of tasks.

Free Model weights

Overview Providers Versions Apps Activity Uptime API

Providers for DeepSeek V3 0324 (free)

OpenRouter routes requests to the best providers that are able to handle your prompt size and parameters, with fallbacks to maximize uptime.

Sort by

Chutes	Context	Max Output	Input	Output	Latency	Throughput	Uptime
fp8	164K	164K	\$0	\$0	1.60s	40.37t/s	III

正在连接...

New Room To: DeepSeek V3 0324 (free) + Add model

hello,who are you?

hello,who are you?

DeepSeek V3 0324 (free) | Targon

Hello! I'm DeepSeek Chat, an AI assistant created by DeepSeek. My purpose is to help you with a wide range of tasks—whether it's answering questions, brainstorming ideas, assisting with coding, or just having a friendly chat. 😊

How can I help you today?

Web Search New Room

Start a message...

- The API key you create directly supports free models. If you wish to use non-free models, you will need to purchase the required amount with your credit card. This tutorial will not be explained in detail, so please research it on your own.

2. Overseas Version of Tongyi Qianwen

Some of the online large models in the overseas version startup examples use Tongyi Qianwen's large visual model. This platform is a third-party platform, and the large model is a powerful visual model provided by Alibaba. After registration and authentication, you can use some of the large model for free, or pay for additional model usage according to the website.

The following examples require this large model: dual-model intelligent agent-object tracking, video description, and text-based image creation.

1. Open the website and register.

Register account

The screenshot shows the Alibaba Cloud registration page. On the left, under 'Account Benefits', there are sections for 'Free Trial' (Get free hands-on experience with 50+ products. Now up to 12 Months for Elastic Compute Service!) and 'Premium Support Services' (1-on-1 pre-sale consultation, 24/7 after-sales technical support with 6 free tickets per quarter). On the right, the 'Sign up to Alibaba Cloud' form is displayed. It asks to 'Please select your account type *'. Two options are shown: 'Business Account' (For purchasing services required by businesses. Enjoy premium support services and exclusive offers.) and 'Individual Account' (For purchasing services required by individuals or for personal use.). The 'Individual Account' option is selected and highlighted with a red box. Below the options is a 'Next' button. There are also links for 'Sign up with Google' and 'Sign up with Github'. At the bottom, it says 'Already a member? Sign In'.

This screenshot shows the next step of the registration process. The 'Sign up to Alibaba Cloud' form now has input fields for 'Email Address *' (Enter your email), 'Password *' (Enter your password), and 'Confirm Password *' (Confirm your password). There are 'Sign Up (Step 1 of 2)' and 'Go Back' buttons. The 'Already a member? Sign In' link is still present at the bottom.

2. After registration, enter your contact information based on your country and verify your identity. Without verification, you will not be able to receive the free large model.

Alibaba Cloud

AI SearchContact SalesEnglishCartConsole

Why UsPricingProductsSolutionsMarketplaceDevelopersPartnersDocumentationServicesModel StudioComplete Sign Up

Successfully registered!

Keep update information.

Complete your account information to use Alibaba Cloud cloud computing services. You can complete the information later.
[Return to Previous Page \[+\]](#)

Security Verification

For account security reasons, please bind your current mobile phone number and use it only in security scenarios such as one time password (OTP) and password retrieval scenarios. [\[+\]](#)

Mobile Phone Number *

+64

Enter a mobile phone number

Verification Code *

Enter a verification code

Obtain Verification Code

Failed to obtain a verification code? [\[+\]](#)

Country/Region

New Zealand

☐ Alibaba Cloud may use your information to contact you about updates and special offers. You can unsubscribe at any time.

Next

Payment Information

3. Go to Large Models, select Qwen-VL-Max, and receive your free model. Withdraw some large model credits

Alibaba Cloud

AI SearchContact SalesEnglishCartConsole

Why UsPricingProductsSolutionsMarketplaceDevelopersPartnersDocumentationServicesModel StudioFree Trial

TiDB Cloud Available for Singapore region now

Creating the cutting-edge TiDB cloud-native architecture on Alibaba Cloud

Learn More

01 ApsaraDB for MongoDB 8.0 Released

02 Qwen3 Models

03 Simple Application Server Special Offer

04 Cloud Drive for Enterprises

Free Tier

Promo Center

Alibaba Cloud Model Studio

TiDB

SAS - the New VPS ChoiceOnly \$9.9/yearSAS 2vCPUx 1GBBuy Now

Alibaba Cloud

AI SearchContact SalesEnglishCartConsole

Why UsPricingProductsSolutionsMarketplaceDevelopersPartnersDocumentationServicesModel StudioComplete Sign Up

Alibaba Cloud > Products > Alibaba Cloud Model Studio

Model Studio: Supercharge Your AI Journey Effortlessly With Industry-Leading GenAI Models

You can seamlessly integrate our advanced AI models—such as Qwen-Max (MoE), Qwen-VL (visual understanding), and the *Wan* series (video generation)—via Model Studio's purpose-built APIs

Activate Now

Contact Sales

Qwen2.5 and Qwen2.5-VL Open-Source Free Trial Models will be commercialized from 2025-06-04 00:00:00 (UTC+08). Click here for more details.

Introducing Qwen3

Why Us

Features

Scenarios

Models and Pricing

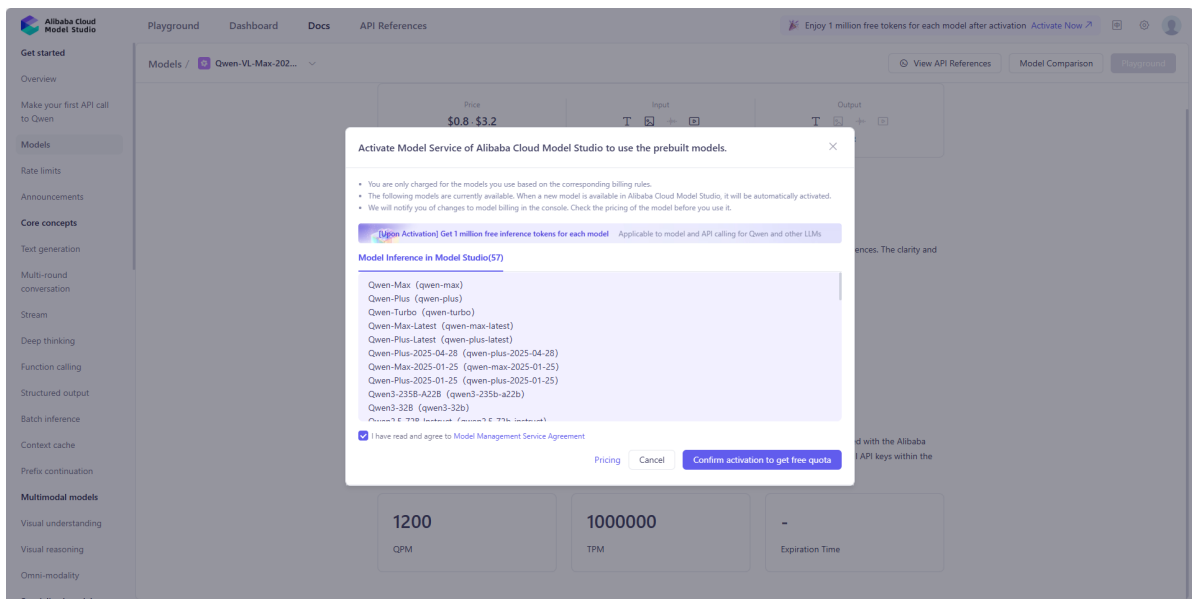
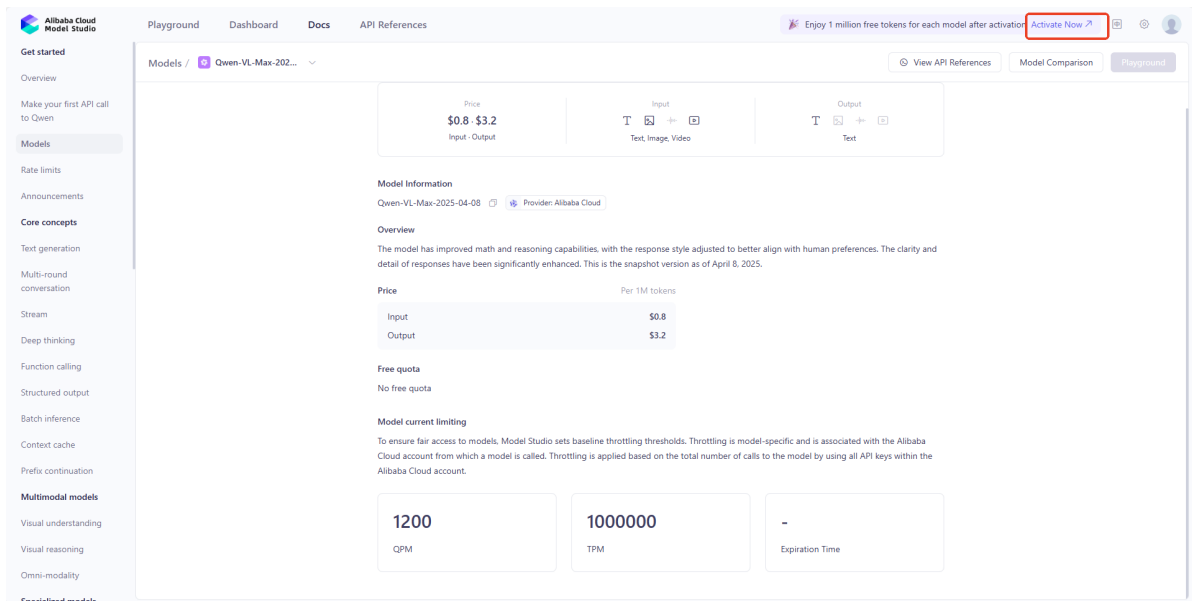
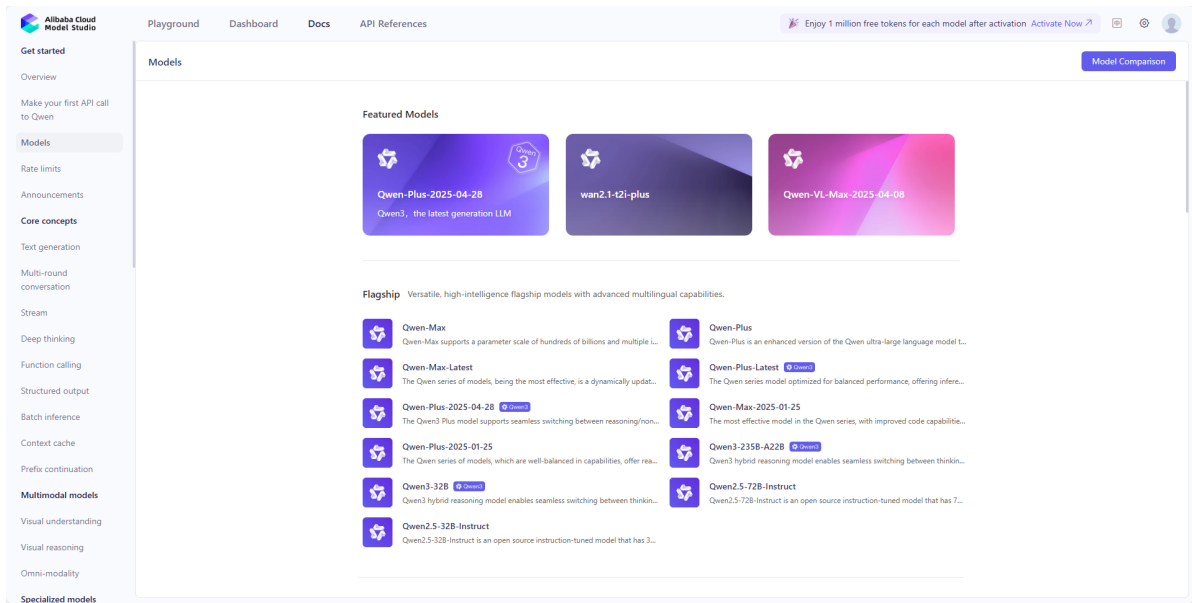
Customer

Documentation

Qwen3: Major Models Unveiled!

All Qwen3 models seamlessly integrate "thinking" and "non-thinking" modes, allowing you to switch modes during conversation. [Try It Now in Model Studio!](#)

Chat now with Alibaba Cloud Customer Service to assist you in finding the right products and services to meet your needs.



4. Register your own API Key

The top screenshot shows the 'API References' page in the Alibaba Cloud Model Studio console. It includes a sidebar with navigation options like 'Preparations', 'Chat', 'Image generation', 'Video generation', 'Text Embedding', and 'OpenAI compatibility'. The main content area is titled 'Obtain an API key' and provides instructions on how to obtain an API key. It includes a section 'Obtain API key' with steps: 1. Go to 'API Key Management' and click 'Create My API Key'. 2. In the 'Actions' column, click 'View'. Then, you can view the API KEY.

The bottom screenshot shows the 'API Key Management' page. It has a table with columns: ID, API Key, Workspace, Description, Creation Time, and Actions. The table is currently empty, showing 'No data'. A 'Create My API Key (0/10)' button is highlighted in the top right corner.

Then remotely access the system of the car using VNC and open a terminal. For instructions on how to access the system using VNC, please refer to the tutorial in Chapter 1.

7. Enter in the terminal

```
shell
nano /home/pi/project_demo/09.AI_Big_Model/API_KEY.py
```

```
20
21
22 #通义千问 Tongyi Qianwen
23 TONYI_key='sk-b[REDACTED]:6' #填写通义千问的APIKEY
24
25
26
27 #####国外key相关的 Foreign key related
28 #注册的地址 : https://openrouter.ai/ Registered address : https://openrouter.ai/
29 openAI_KEY = 'sk-or-v1-[REDACTED]408d0f' #填写openrouter
30 的APIKEY
```

Python 2 Tab Width: 8 Ln 6, Col 1 INS

Then, following the prompts, enter the OpenRouter API-KEY and the API-KEY for TONYI_KEY. Press Ctrl+S to save and Ctrl+X to exit. You can now run the Big Model demo in the English version. The Spark Big Model key used by international customers is already included in the relevant examples and does not need to be entered here.

3. Local DIFY Configuration

The image's Dify function is disabled by default. Execute the above command to enable it.

```
cd ~/dify-1.6.0/docker
./docker-compose-linux-aarch64 up -d
```

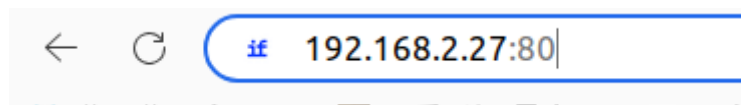
This will enable it.

If you don't want Dify to be running all the time (not recommended for 2GB of RAM), execute the following command to disable it.

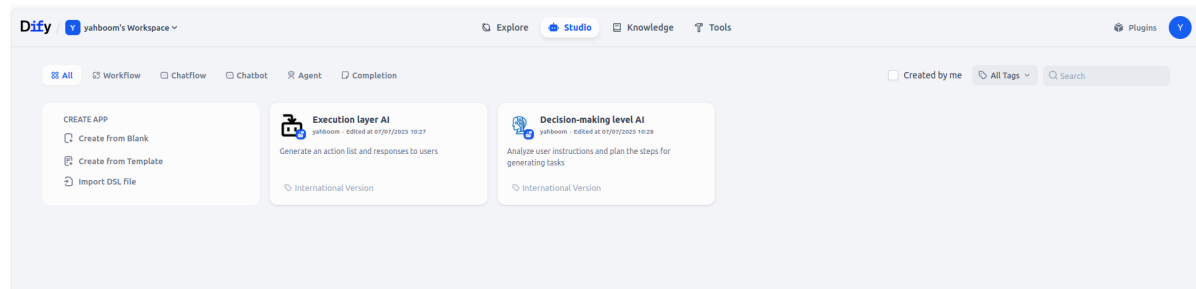
```
cd ~/dify-1.6.0/docker
./docker-compose-linux-aarch64 down
```

Enter the Dify configuration page

- **Observe the IP address of the OLED screen**
- Open a browser on any computer in the Raspberry Pi system on the same network segment as the car and enter the IP address + :80 in the address bar. For example:



The page after entering dify is as follows:

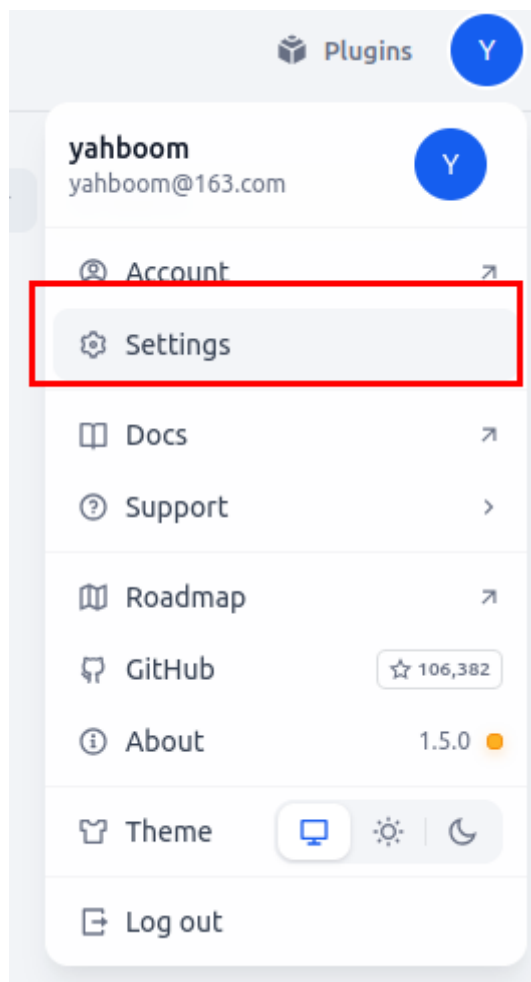


If you are accessing the website from an unfamiliar device, you will need to log in with your account:

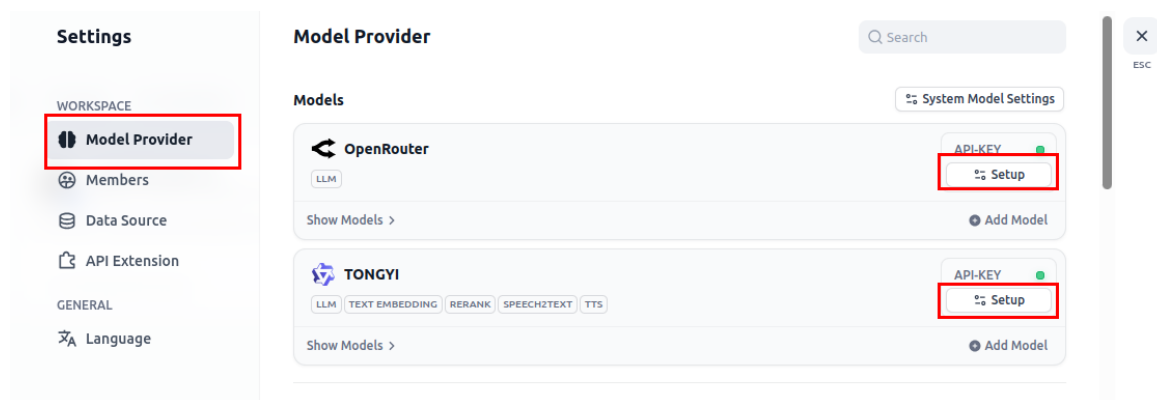
- Username: yahboom@163.com
- Password: yahboom123

3.1 Enter the API key.

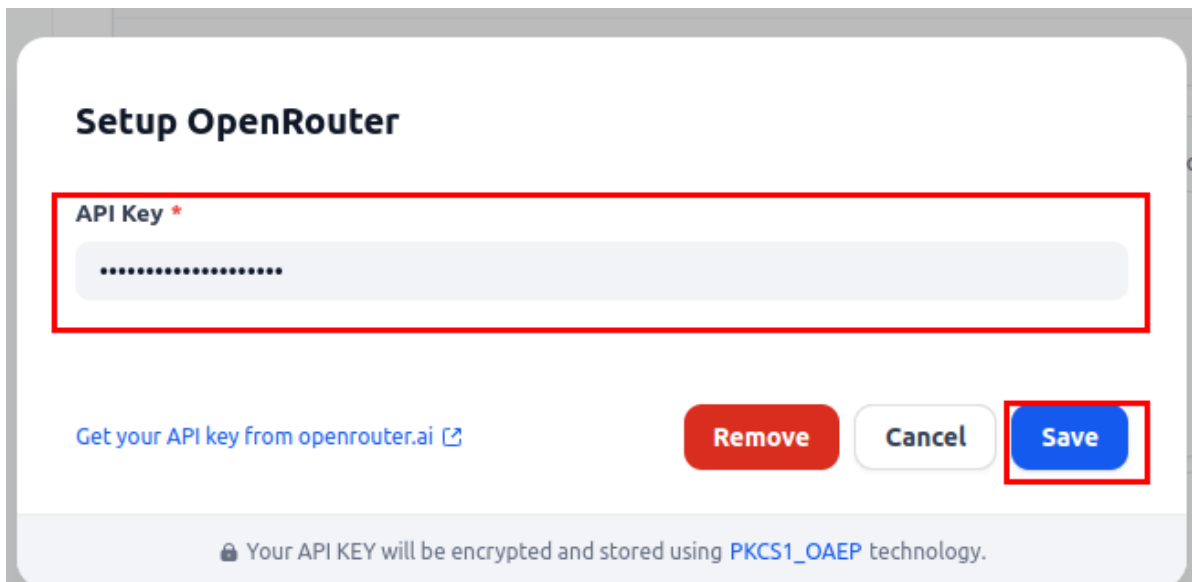
Click your profile picture in the upper right corner, then click Settings.



Click Model Provider, then click Setup for the corresponding model provider.



Enter the API key you applied for in [2.3 Registering an OpenRouter Platform Account], then click Save.



Setup OpenRouter

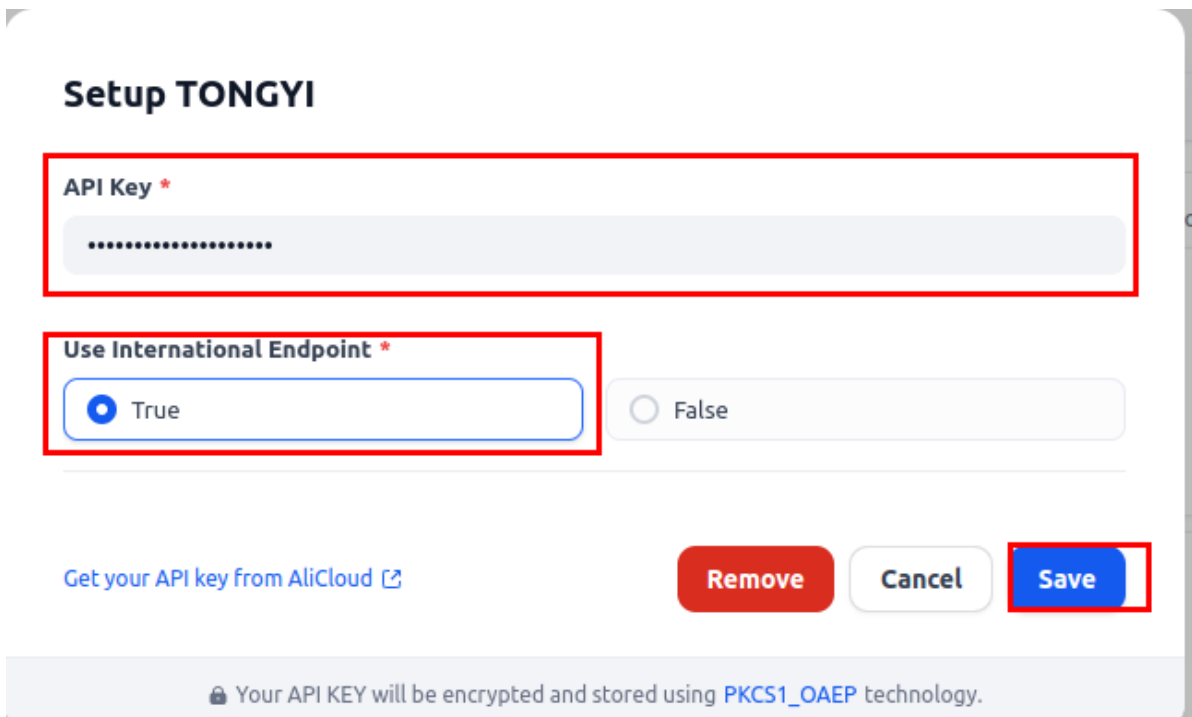
API Key *

.....

[Get your API key from openrouter.ai](#) [Remove](#) [Cancel](#) [Save](#)

🔒 Your API KEY will be encrypted and stored using [PKCS1_OAEP](#) technology.

Click Setup for TONGYI, then enter the API-KEY you applied for in [2.4 Registering an Alibaba Bailian Model International Platform Account], then select Use International. Endpoint, then click Save.



Setup TONGYI

API Key *

.....

Use International Endpoint *

☒ True ☐ False

[Get your API key from AliCloud](#) [Remove](#) [Cancel](#) [Save](#)

🔒 Your API KEY will be encrypted and stored using [PKCS1_OAEP](#) technology.

At this point, set `DIFY_SWITCH=True` in `API_KEY.py` to run the configured agent in dify.

```
API_KEY.py
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30 # 国外 dify 的开关
31 DIFY_SWITCH = True
32
33
34
35
36
37
```