

Configuration Prerequisites for Large Models (Must Read)

Configuration Prerequisites for Large Models (Must Read)

1. English Version

OpenRouter Operation Steps

Important prerequisite for large model-related examples: a Wi-Fi connection is required.

For detailed instructions on how to connect, please refer to Section 6 of Chapter 1, "Connecting to Wi-Fi." This tutorial provides a brief overview.

1. First, download the XGO Bluetooth remote control app.
2. Open the app, click on the network configuration information in the upper right corner, enter the Wi-Fi name and password, and click to generate a QR code.
3. Use the on-screen buttons on the RiderPI to select the wireless network function, and point the QR code generated by XGO at the RiderPI's camera.
4. After a short wait, the wireless network function will automatically exit, indicating a successful Wi-Fi connection.

This tutorial concerns the configuration of AI large models. To run the large model-related examples in the boot-up example correctly, you need to follow the configuration steps below.

1. English Version

The online large model in the English version boot-up example uses the OpenRouter platform middleware.

OpenRouter Operation Steps

1. First, log in to the following official website:

<https://openrouter.ai/>

2. Then log in. You can use various login methods; this article recommends using your 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

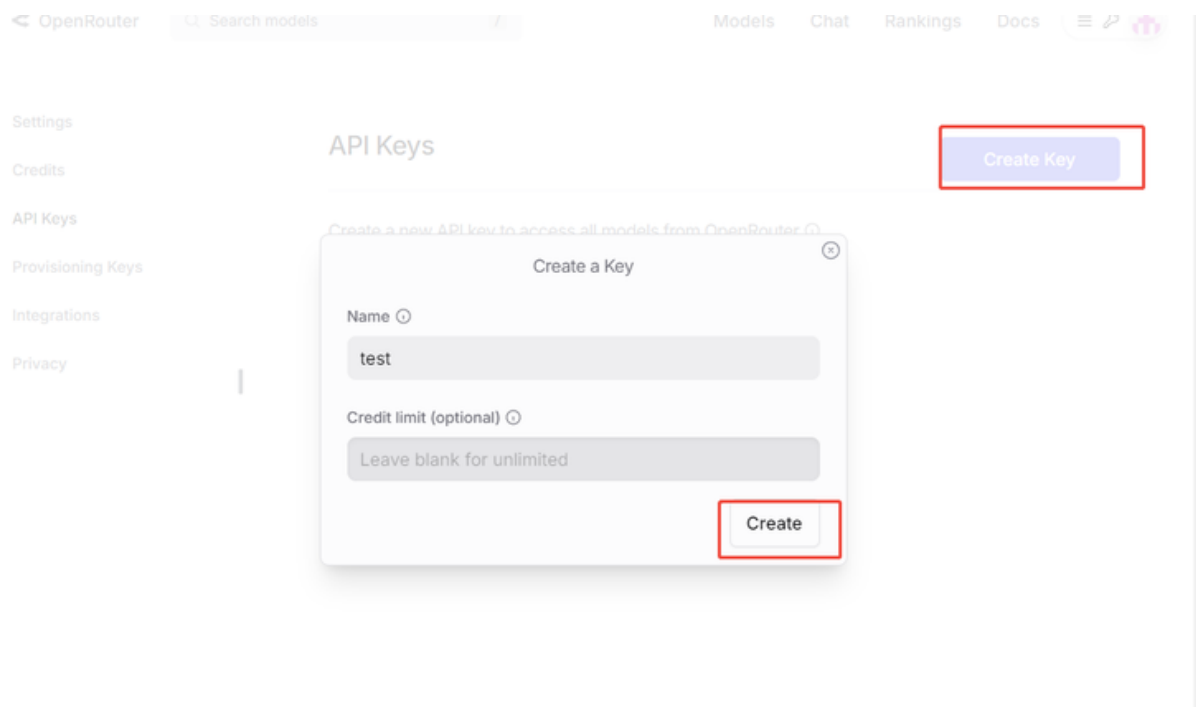
50+

Active Providers

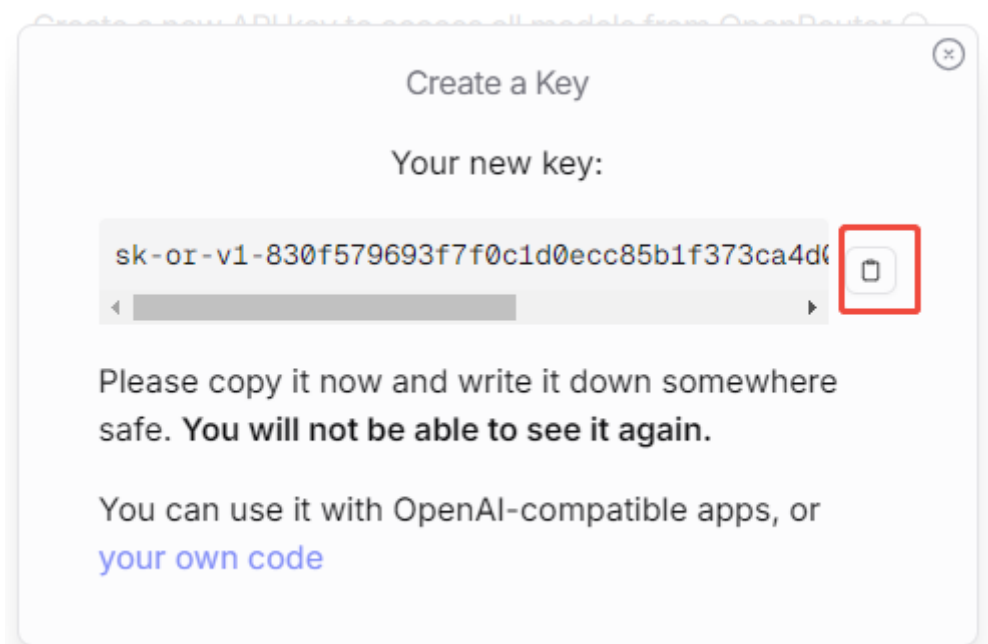
300+

Models

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



4. Then copy the created key and save it to your computer. **You can only copy this key the first time you create it.** Once you leave this interface, you cannot view it again.



5. Then we find a free model to test whether the registered account works normally.

OpenRouter

Search models

/

Models

Chat

Rankings

Docs

The Unified Interface For LLMs

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

Start a message...

→

Featured Models

[View Trending](#)

Gemini 2.5 Pro Preview

by google

83.1B Tokens/wk

13.9s Latency

+7.42% Weekly growth

GPT-4.1

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/models

OpenRouter

Search models

/

Models

Chat

Rankings

Docs

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 ...

正在等待 openrouter.ai 的响应...

6. If there is a normal response, it means it works normally.

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?

1577958961@qq.com

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 created API keys can directly use free models. If you want to use non-free models, you need to purchase the quantity of non-free models with your credit card. This is not explained in the tutorial; you can research it yourself.
- Remotely access the RiderPI system via VNC and open the terminal. For instructions on how to access the system via VNC, please refer to the tutorial in Chapter 1.
- Terminal input:

```
#cm4 run
nano ~/RaspberryPi-CM4-main/demos/API_KEY.py

#cm5 run
nano RaspberryPi-CM5/demos/API_KEY.py
```

```
####国内key相关的 Domestic key related
#星火大模型 Starfire Large Model
XINGHOU_APPID = 'xxxxxxx' #填写星火大模型的APPID
XINGHOU_APISecret = 'xxxxxxx' #填写星火大模型的APISecret
XINGHOU_KEY = "xxxxxxx" #填写星火大模型的APIKEY
```

```
#通义千问 Tongyi Qianwen
TONYI_key='xxxxxxxxx' #填写通义千问的APIKEY
```

```
#####国外key相关的 Foreign key related
#注册的地址: https://openrouter.ai/ Registered address: https://openrouter.ai/
openAI_KEY = 'xxxxxxx' #填写openrouter的APIKEY
```

Then, enter the OpenRouter API-KEY as prompted, save with Ctrl+S, and exit with Ctrl+X. This will allow you to run the large model example in the English version.