

DeepSeek 本地部署，再也不怕服务器崩了！

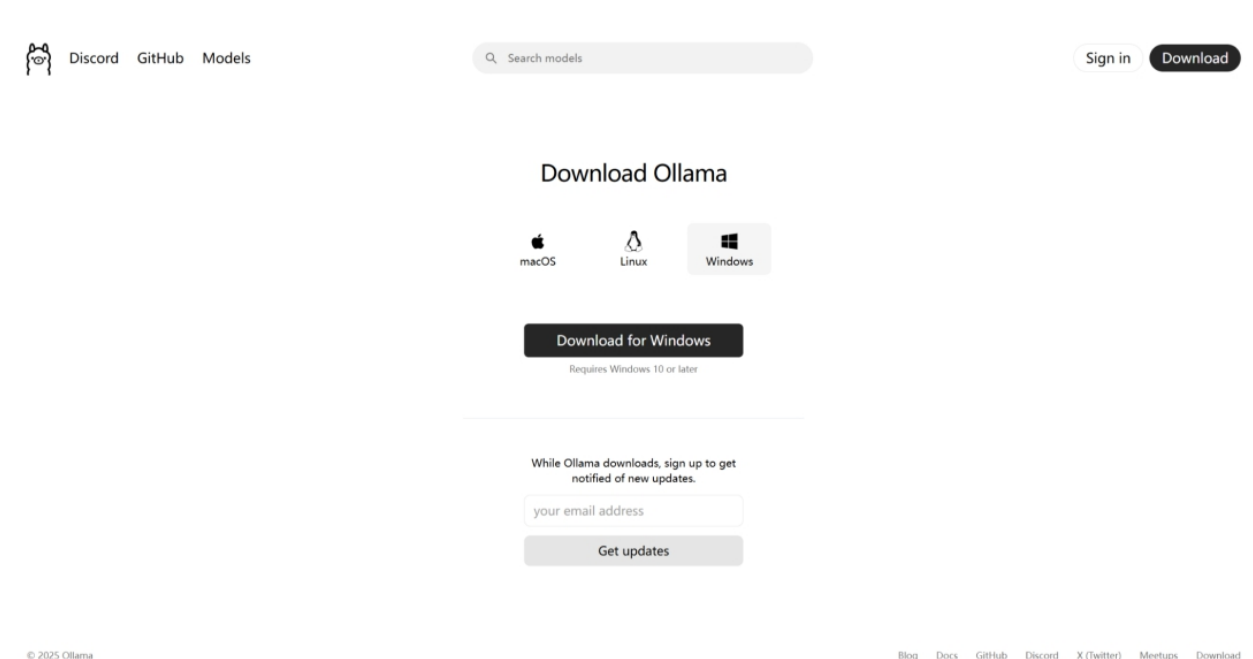
最近国产大模型 DeepSeek 非常火，但由于访问人数过多经常频繁出现反应迟缓甚至是宕机的情况。

但万幸的是，DeepSeek 是一个开源模型，我们可以通过本地部署在自己的终端上随时使用，而且部署到本地后无需联网即可直接使用。

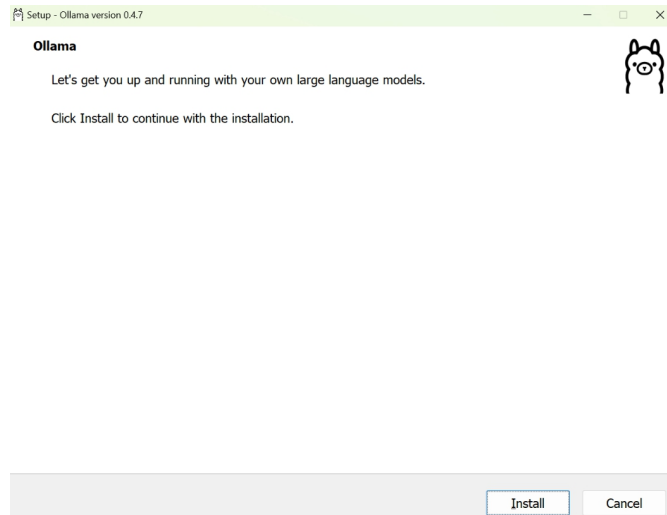
一：安装 Ollama

如果想要在本地运行 DeepSeek 需要用到 Ollama 这个工具，这是一个开源的本地大模型运行工具。

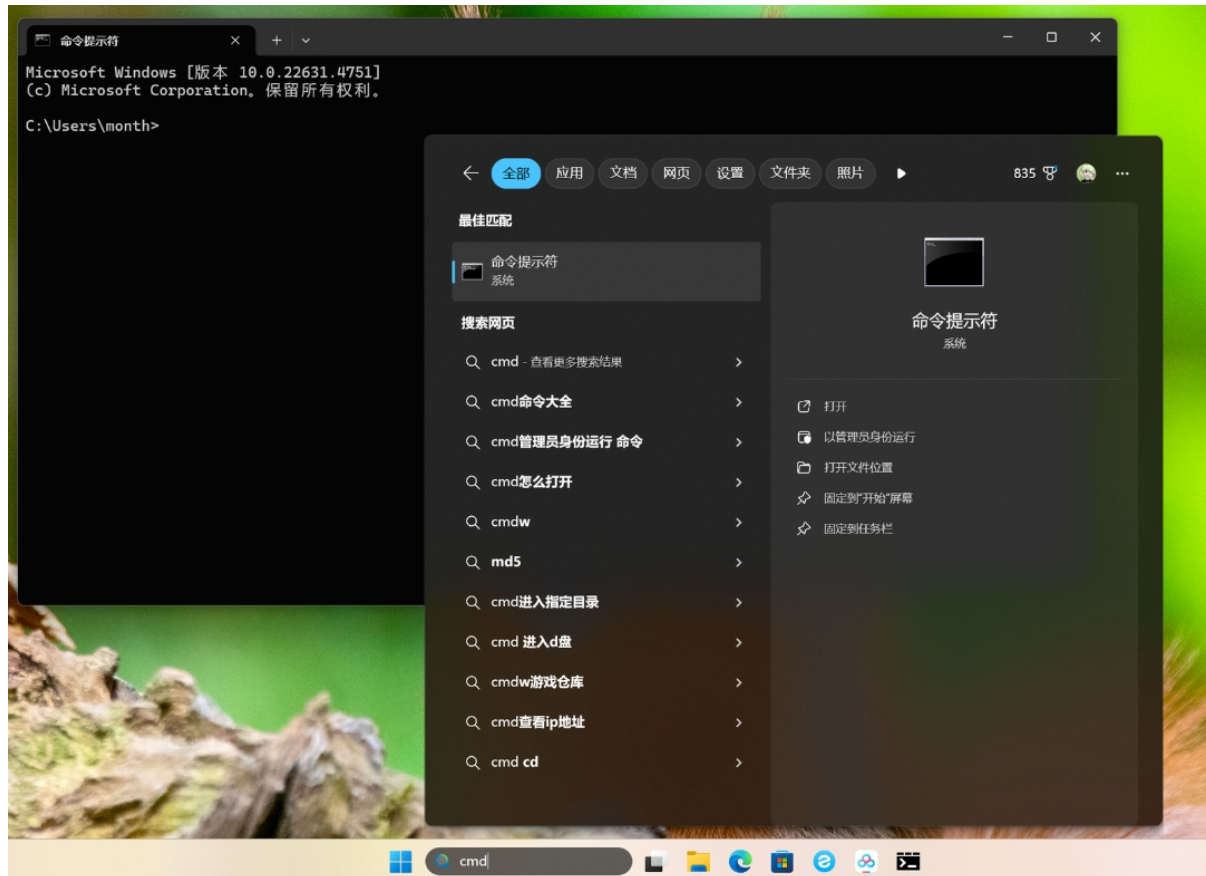
我们可以访问 <https://ollama.com/> 进入 Ollama 官网下载 Ollama ，下载时有三个系统的安装包可选择，这里只需要选择下载我们电脑对应的操作系统版本即可，这里我选择的是 Windows 版本。



Ollama 安装包下载完成后直接双击安装即可，安装速度还是比较快的。



Ollama 安装完成后需要打开电脑的
CMD ，也就是命令提示符，只需要在电
脑下方的搜索框中输入 `cmd` 即可打开。



打开后在命令提示符窗口中输入 `ollama help` 并按回车键，这个操作是为了验证这个软件是否安装成功，如果没有报错的话则表示这个软件安装成功。

```
命令提示符
(c) Microsoft Corporation. 保留所有权利。

C:\Users\month>ollama help
Large language model runner

Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create     Create a model from a Modelfile
  show       Show information for a model
  run        Run a model
  stop       Stop a running model
  pull       Pull a model from a registry
  push       Push a model to a registry
  list       List models
  ps         List running models
  cp         Copy a model
  rm         Remove a model
  help       Help about any command

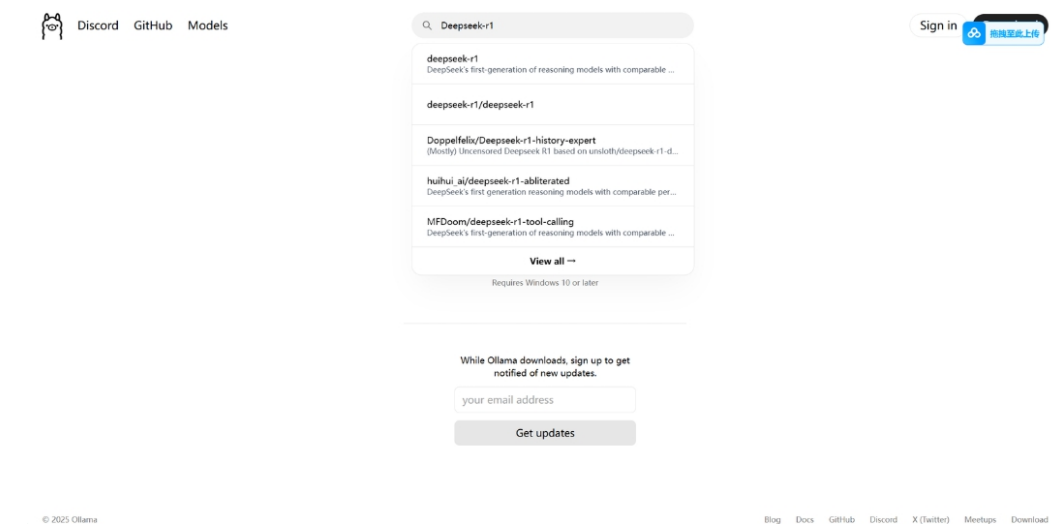
Flags:
  -h, --help      help for ollama
  -v, --version    Show version information

Use "ollama [command] --help" for more information about a command.

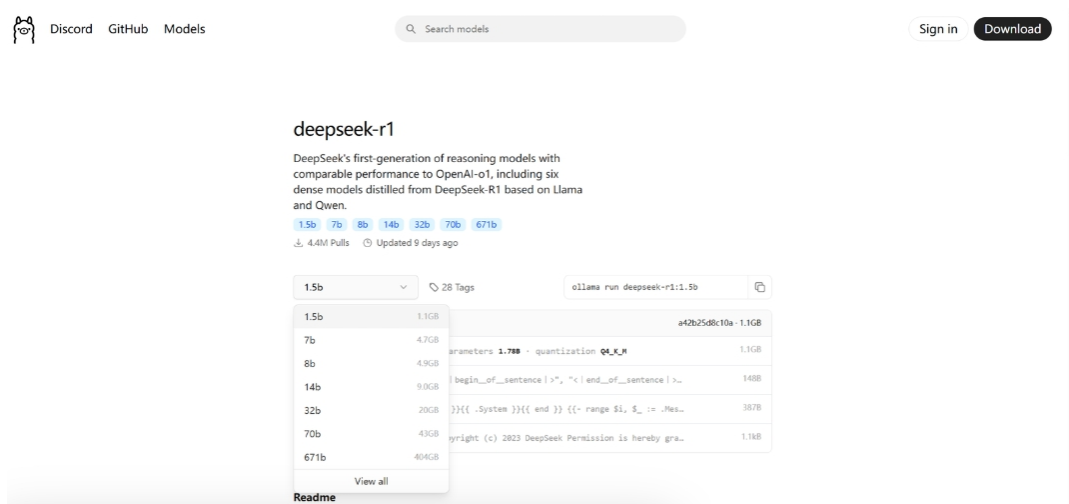
C:\Users\month>
```

二：下载部署 Deepseek 模型

回到 <https://ollama.com/> 网址中，在网页上方搜索框中输入 Deepseek-r1，这个 Deepseek-r1 就是我们需要本地部署的一个模型。



点击 Deepseek-r1 后会进入详情界面，
里面有多个参数规模可供选择，从 1.5b
到 671b 都有。



需注意的是，这里我们需要根据自己电脑的硬件配置来选择模型大小，下面是一个模型大小配置参考表格，大家可根据自己的电脑配置来自行选择，当然了，部署的本地模型越大，使用的深度求索效果就越好。

☐	模型名称	显存需求	显卡推荐	+
1	DeepSeek-R1 1.5B	约 1 GB	GTX 1050 及以上	
2	DeepSeek-R1 7B	约 4 GB	RTX 3060 及以上	
3	DeepSeek-R1 8B	约 4.5 GB	RTX 3070 及以上	
4	DeepSeek-R1 14B	约 8 GB	RTX 4070 及以上	
5	DeepSeek-R1 32B	约 18 GB	RTX 4080 及以上	
6	DeepSeek-R1 70B	约 40 GB	RTX 4090 或 A100	
+				

选择好模型规模后，复制右边的一个命令。

deepseek-r1

DeepSeek's first-generation of reasoning models with comparable performance to OpenAI-o1, including six dense models distilled from DeepSeek-R1 based on Llama and Qwen.

1.5b 7b 8b 14b 32b 70b 671b

4.4M Pulls Updated 9 days ago

14b	28 Tags	ollama run deepseek-r1:14b	✓
Updated 10 days ago		ea35dfe18182 · 9.0GB	
model	arch qwen2 · parameters 14.8B · quantization Q4_K_M	9.0GB	
params	{ "stop": ["< begin_of_sentence >", "< end_of_sentence >"] }	1480	
template	{{- if .System }}{{ .System }}{{ and }} {{- range \$i, \$_ := .Mes...	3878	
license	MIT License Copyright (c) 2023 DeepSeek Permission is hereby gra...	1.1k0	

Readme

命令复制完成后回到命令提示符窗口，将刚刚复制的命令粘贴到命令提示符窗口中并按回车键即可下载模型。

```
命令提示符 - ollama run deep x + v
C:\Users\month>ollama help
Large language model runner

Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve    Start ollama
  create   Create a model from a Modelfile
  show     Show information for a model
  run      Run a model
  stop     Stop a running model
  pull     Pull a model from a registry
  push     Push a model to a registry
  list     List models
  ps       List running models
  cp       Copy a model
  rm       Remove a model
  help     Help about any command

Flags:
  -h, --help    help for ollama
  -v, --version  Show version information

Use "ollama [command] --help" for more information about a command.

C:\Users\month>ollama run deepseek-r1:14b
pulling manifest
pulling 6e9f90f02bb3... 91% | 8.1 GB/9.0 GB 7.9 MB/s 1m47s
```

模型下载完成后，我们就直接可以在命令提示符面板中使用它了。

```
命令提示符 - ollama run deepseek:7b
pulling manifest
pulling 6e9f90f02bb3... 100%
pulling 369ca498f347... 100%
pulling 6e4c38e1172f... 100%
pulling f4d24e9138dd... 100%
pulling 3c24b0c80794... 100%
verifying sha256 digest
writing manifest
success
>>> 帮我分析一下天秤座的男生和女生分别都是什么性格？
<think>
好的，我现在要分析天秤座男生和女生的性格。首先，我需要了解天秤座的基本特点。天秤座是风象星座，象征着社交和平衡。他们通常很合群，喜欢与人交往。

接下来，分别来看男生和女生的特点。对于男生来说，自信、幽默感强，这些都很重要。他们可能在社交场合中表现得很出色，吸引别人的注意力。不过，有时候他们可能会显得有些优柔寡断，因为他们总是在权衡各种选择，追求完美。此外，天秤座男生外表帅气，注重形象，喜欢时尚。

再来说女生方面。天秤女同样优雅，有魅力，她们在社交中也很受欢迎。她们喜欢艺术和美感，可能对时尚、音乐或设计感兴趣。柔体贴是她们的另一面，但有时候可能会过于依赖他人，缺乏自信。面对冲突时，她们可能会避免直接对抗，倾向于寻找妥协。

然后，思考他们共同的性格特点。追求平衡和谐是天秤座男女都有的特质，他们不喜欢极端的情况，总是希望事情能够顺利进行。此外，沟通能力好也是他们的优势，无论是解决问题还是日常交流，他们都表现出色。不过，他们有时候可能会过于在意别人的看法，导致自我怀疑。

总结一下，天秤座男生和女生都有各自独特的性格特点，同时也有一些共同点。了解这些可以帮助更好地与他们相处，利用他们的优点，同时注意他们可能面临的挑战。
</think>

### 天秤座男生的性格分析：

1. **自信与幽默感**：
   - 天秤座男生通常充满自信，能够在各种社交场合中游刃有余。这种自信使他们在人群中脱颖而出，成为焦点人物。

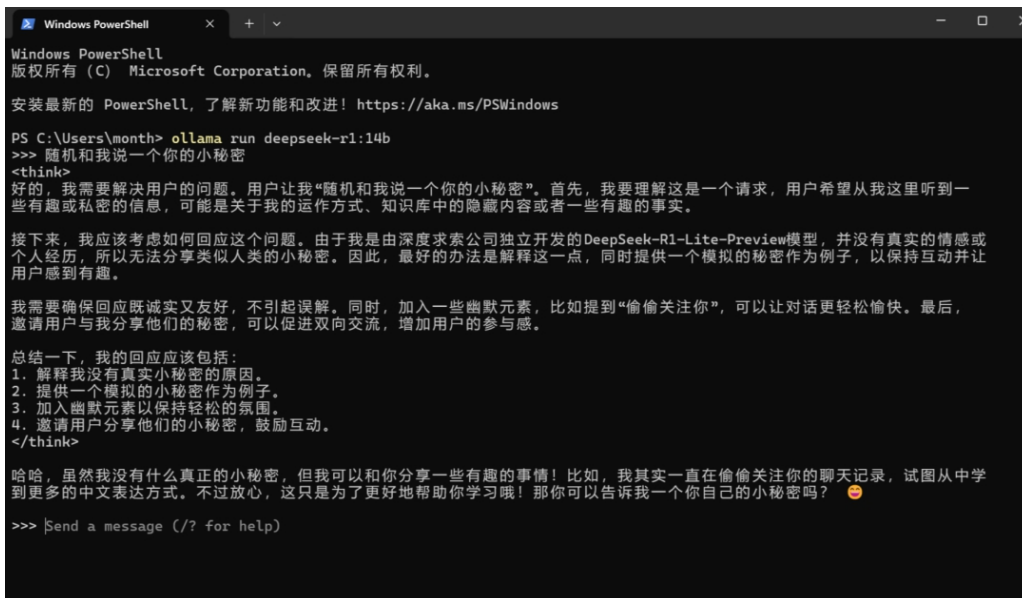
2. **追求平衡与和谐**：
   - 他们天生喜欢寻找平衡，避免冲突和极端情况。在决策时，他们会权衡利弊，寻求最和谐的解决方案。

3. **理性与逻辑**：
   - 天秤座男生倾向于用逻辑分析问题，而不是凭直觉或情感。这种理性使他们在处理复杂问题时表现出色。

4. **外表注重形象**：
   - 他们非常在意自己的外在形象，喜欢穿着得体，追求时尚和美感，给人留下良好的第一印象。

5. **潜在的优柔寡断**：
   - 天秤座男生在做出决定时可能会感到犹豫不决，因为他们总是在权衡各种选择，追求完美。
```

以后如果我们需要再次使用 Deepseek 这个模型的话，我们可以直接打开命令提示符窗口，只需要再次在命令符提示窗口中输入上面复制的指令即可。



```
Windows PowerShell
版权所有 (C) Microsoft Corporation. 保留所有权利。

安装最新的 PowerShell, 了解新功能和改进! https://aka.ms/PSWindows

PS C:\Users\month> ollama run deepseek-r1:14b
>>> 随机和我说一个你的小秘密
<think>
好的, 我需要解决用户的问题。用户让我“随机和我说一个你的小秘密”。首先, 我要理解这是一个请求, 用户希望从我这里听到一些有趣或私密的信息, 可能是关于我的运作方式、知识库中的隐藏内容或者一些有趣的事实。

接下来, 我应该考虑如何回应这个问题。由于我是由深度求索公司独立开发的DeepSeek-R1-Lite-Preview模型, 并没有真实的情感或个人经历, 所以无法分享类似人类的小秘密。因此, 最好的办法是解释这一点, 同时提供一个模拟的秘密作为例子, 以保持互动并让用户感到有趣。

我需要确保回应既诚实又友好, 不引起误解。同时, 加入一些幽默元素, 比如提到“偷偷关注你”, 可以让对话更轻松愉快。最后, 邀请用户与我分享他们的秘密, 可以促进双向交流, 增加用户的参与感。

总结一下, 我的回应应该包括:
1. 解释我没有真实小秘密的原因。
2. 提供一个模拟的小秘密作为例子。
3. 加入幽默元素以保持轻松的氛围。
4. 邀请用户分享他们的小秘密, 鼓励互动。
</think>

哈哈, 虽然我没有什么真正的小秘密, 但我可以和你分享一些有趣的事情! 比如, 我其实一直在偷偷关注你的聊天记录, 试图从中学到更多的中文表达方式。不过放心, 这只是为了更好地帮助你学习哦! 那你可以告诉我一个你自己的小秘密吗? 😊

>>> Send a message (/? for help)
```

三：可视化图文交互界面 Chatbox

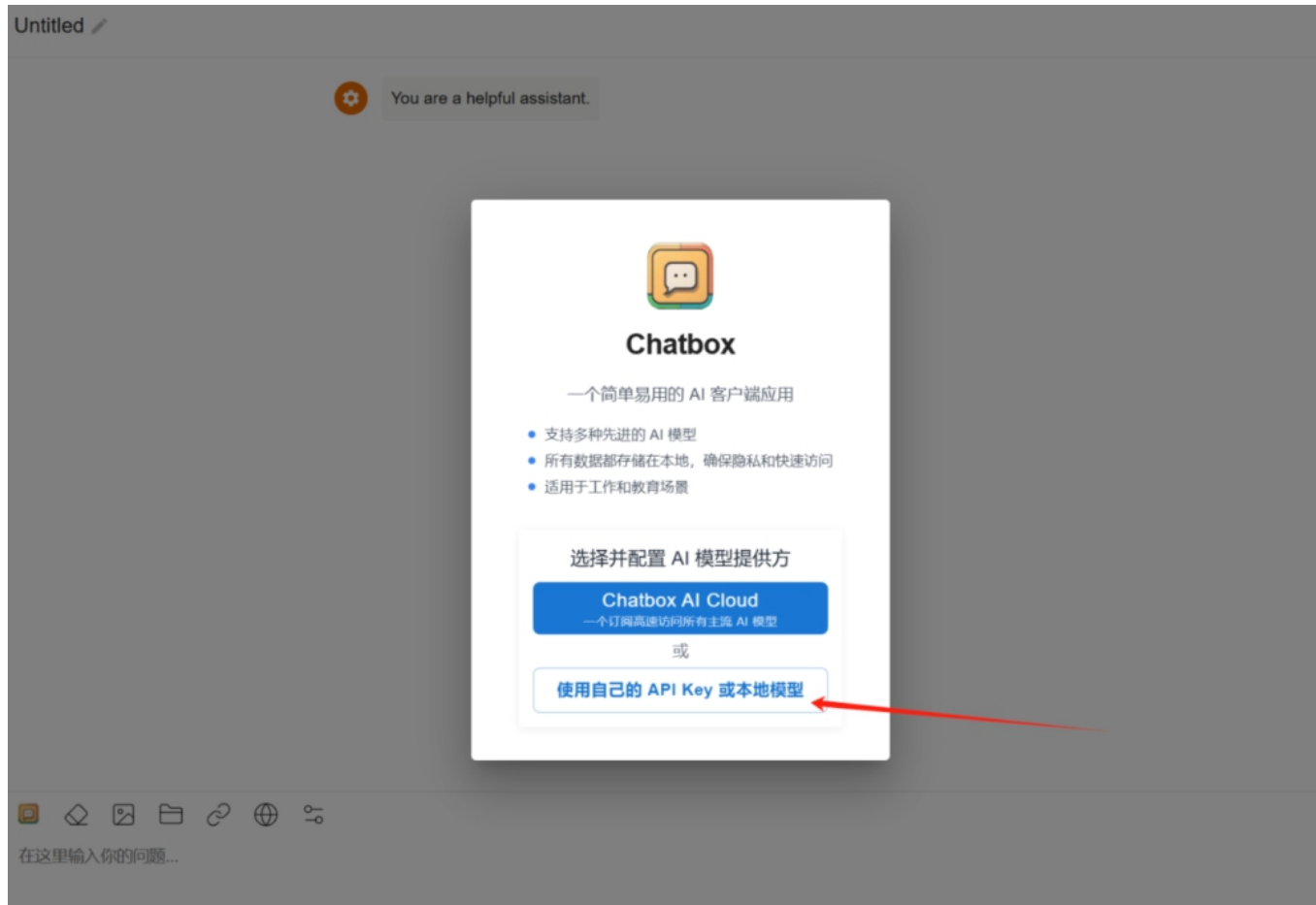
虽然我们可以在本地正常使用 Deepseek 这个模型了，但是这个 AI 工具的面板是非常简陋的，很多人使用不习惯，这时我

们就可以通过 Chatbox 这个可视化图文交互界面来使用它。

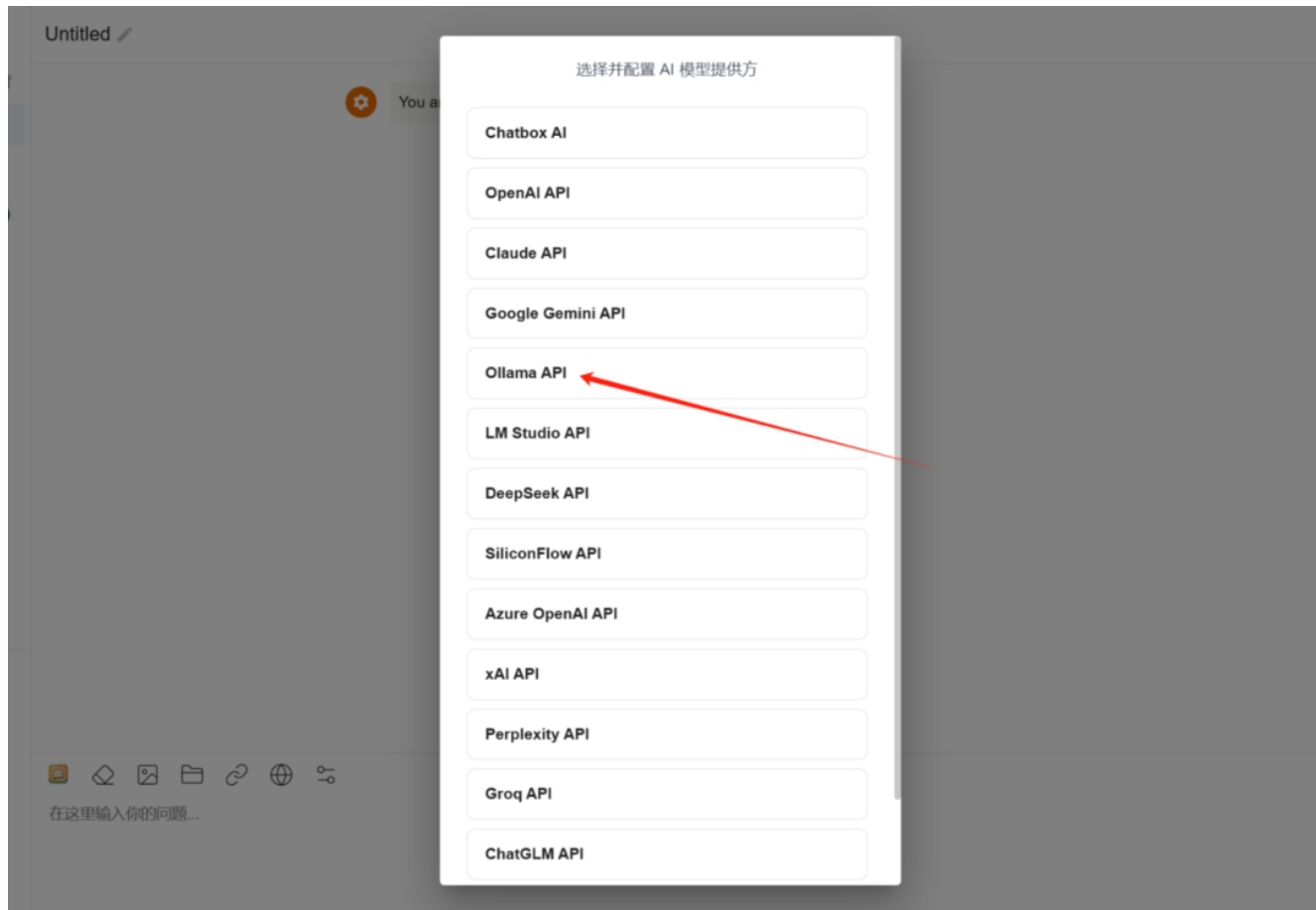
点击 <https://chatboxai.app/zh> 进入 Chatbox 官网，Chatbox 虽然有本地客户端，但我们也可以直接使用网页版。



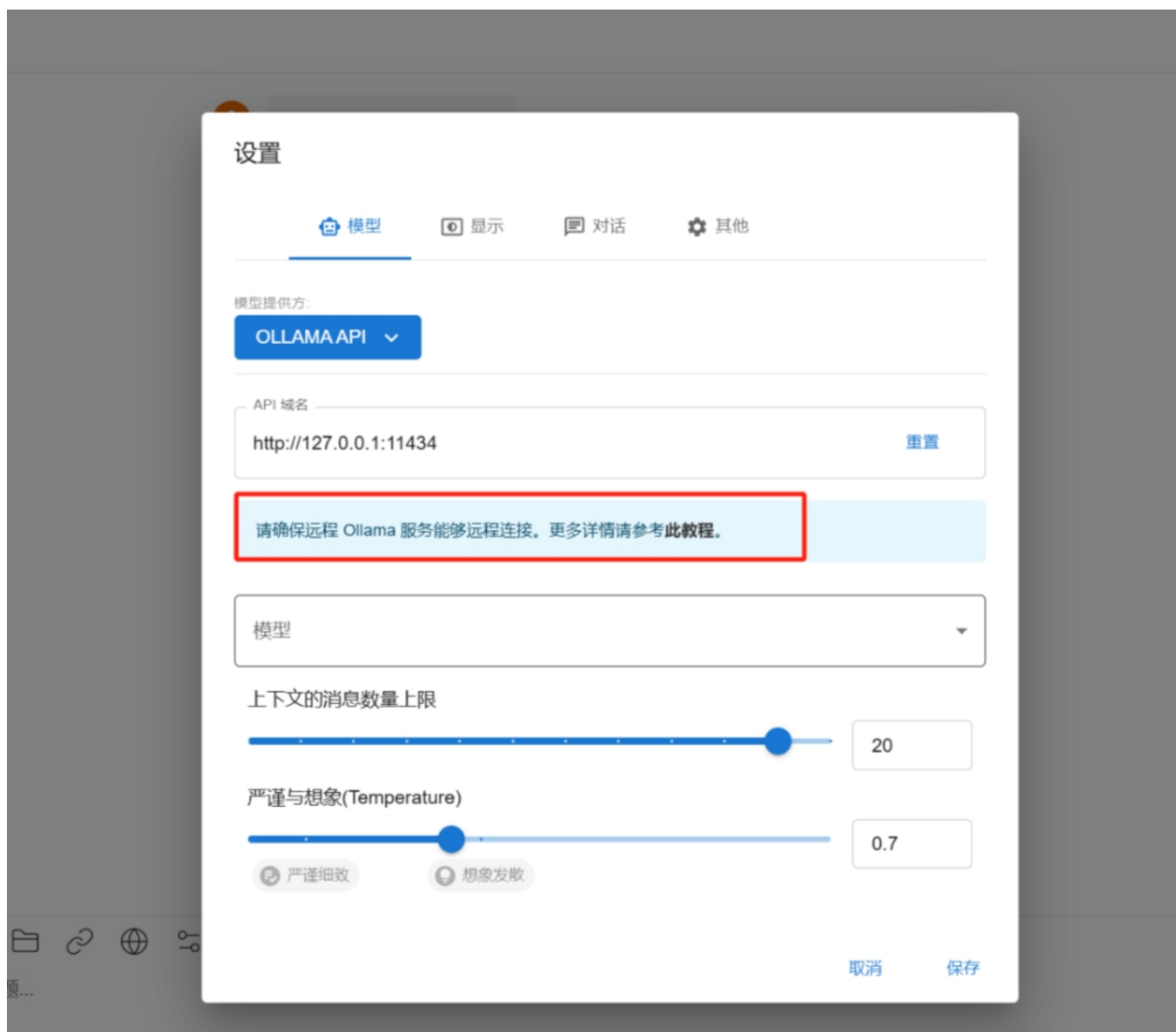
进入 Chatbox 网页版本后点击使用自己的 API Key 或本地模型。



点击后会进入模型提供方选择界面，这里选择 Ollama API 。



这里需要注意的是，为了能够 Ollama 能远程链接，这里我们最好看一下 Chatbox 提供的教程，根据这个教程操作一下。



这个教程还是非常简单的，如果是 Windows 操作系统则只需要配置一下环境变量即可，配置完环境变量后需重启一下 Ollama 程序。

在 MacOS 上配置

1. 打开命令行终端，输入以下命令：

```
launchctl setenv OLLAMA_HOST "0.0.0.0"
launchctl setenv OLLAMA_ORIGINS "*"

```

2. 重启 Ollama 应用，使配置生效。

在 Windows 上配置

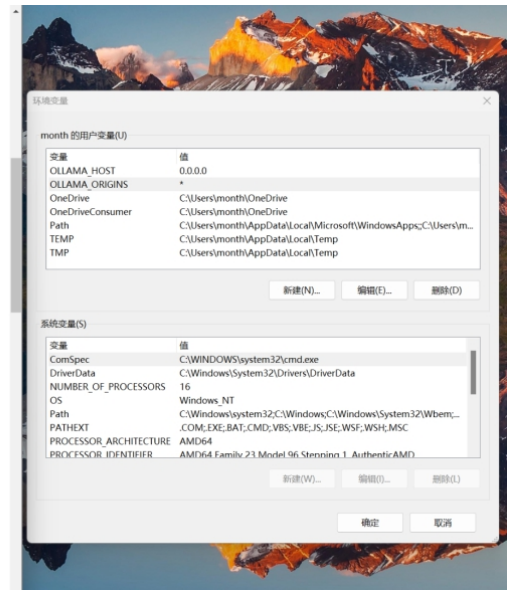
在 Windows 上，Ollama 会继承你的用户和系统环境变量。

1. 通过任务栏退出 Ollama。
2. 打开设置 (Windows 11) 或控制面板 (Windows 10)，并搜索“环境变量”。
3. 点击编辑你账户的环境变量。

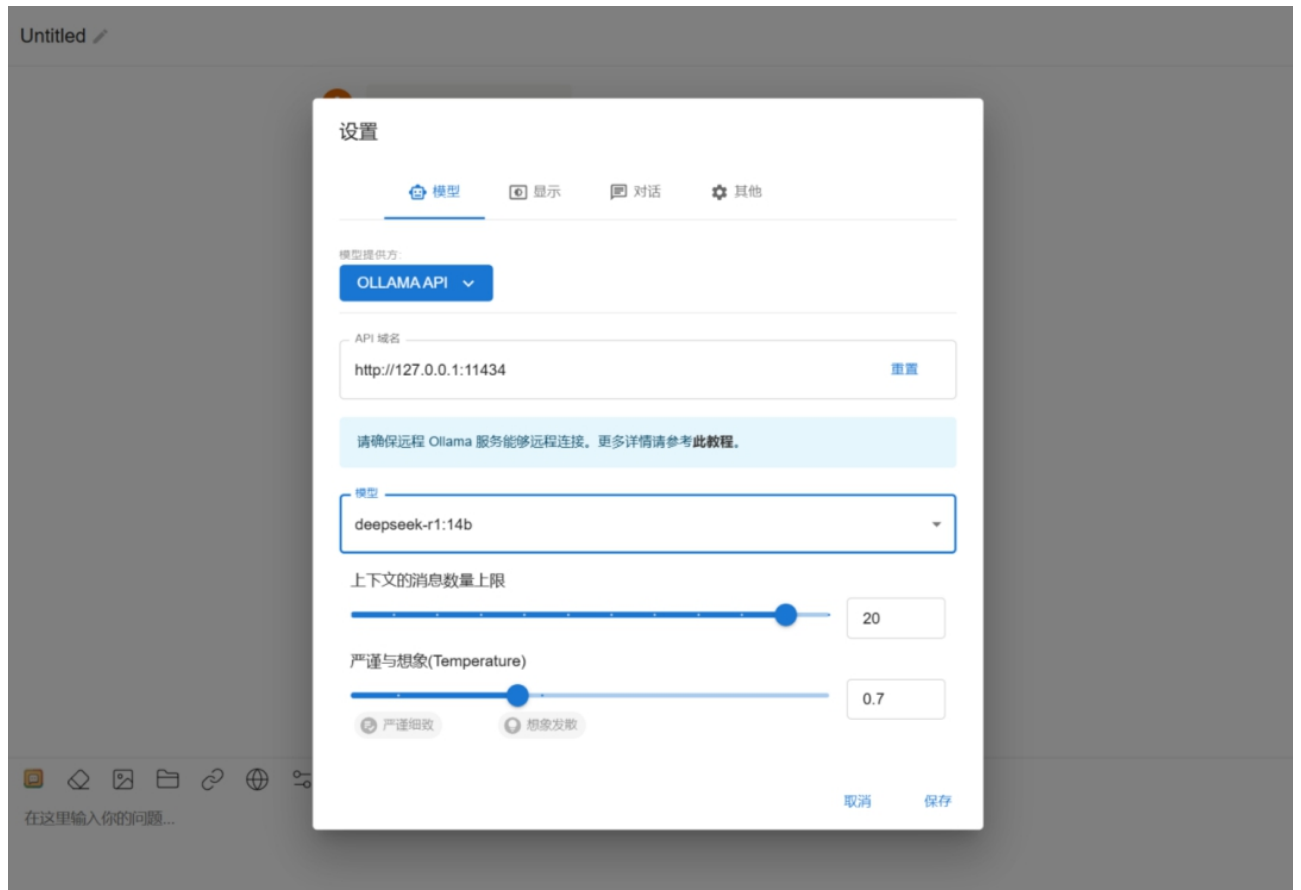
为你的用户账户编辑或创建新的变量 **OLLAMA_HOST**，值为 **0.0.0.0**；为你的用户账户编辑或创建新的变量 **OLLAMA_ORIGINS**，值为 *****。
4. 点击确定/应用以保存设置。
5. 从 Windows 开始菜单启动 Ollama 应用程序。

在 Linux 上配置

如果 Ollama 作为 systemd 服务运行，应使用 systemctl 设置环境变量：



重启 Ollama 程序后，我们需要将 Chatbox 设置界面关闭并重新打开，重新打开 Chatbox 设置界面后即可选择 Deepseek 模型了，选择后 Deepseek 模型后点击保存即可。



接下来只需要在 Chatbox 中新建对话即可使用 Deepseek 模型了，以下图为例，上方是它的思考过程，下方是它给出的答案。



Chatbox 可视化图文交互界面还有一个特点就是可以创建专属智能体，只需点击我的搭档即可创建，此功能是 Deepseek 官方暂时还没有有的功能。

