

Cloud Studio (云端 IDE)

实践教学



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

实践教程

DeepSeek 使用指南

实践教程

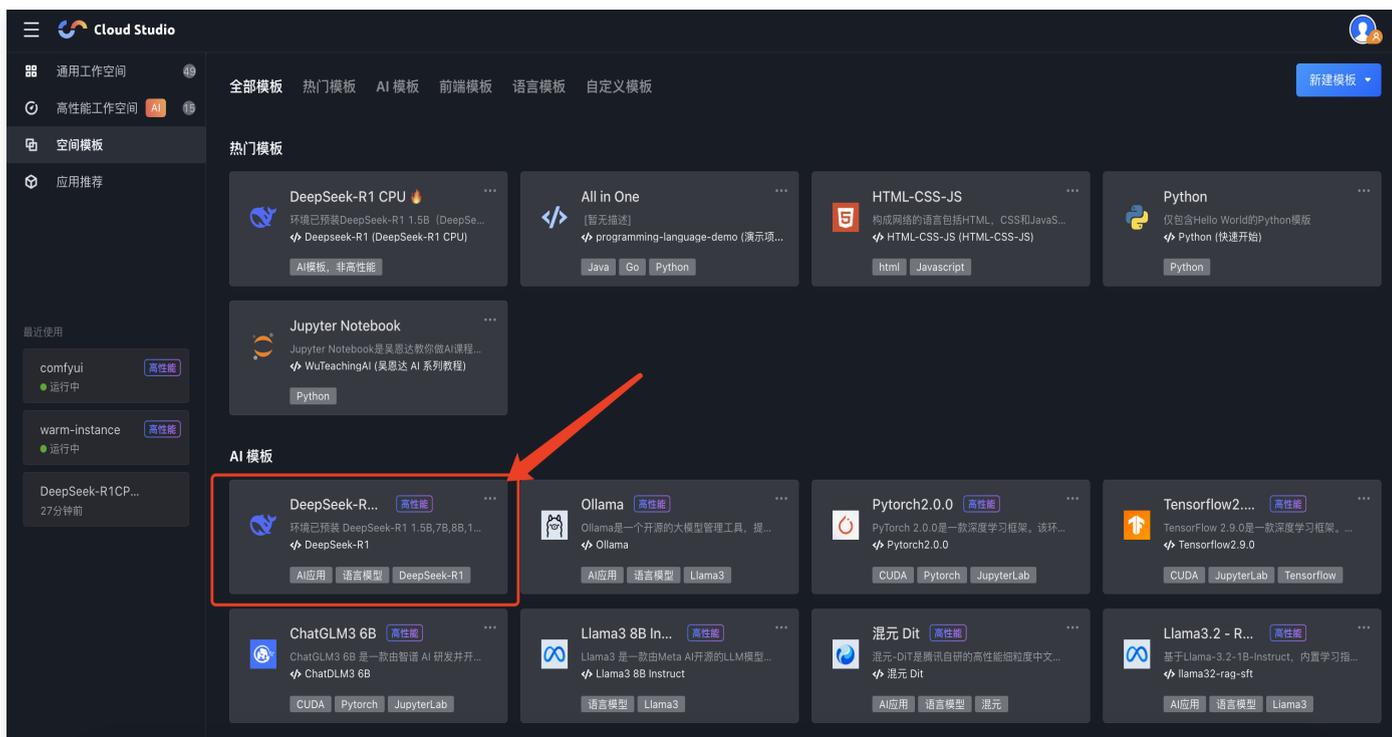
DeepSeek 使用指南

最近更新时间：2025-02-10 15:10:33

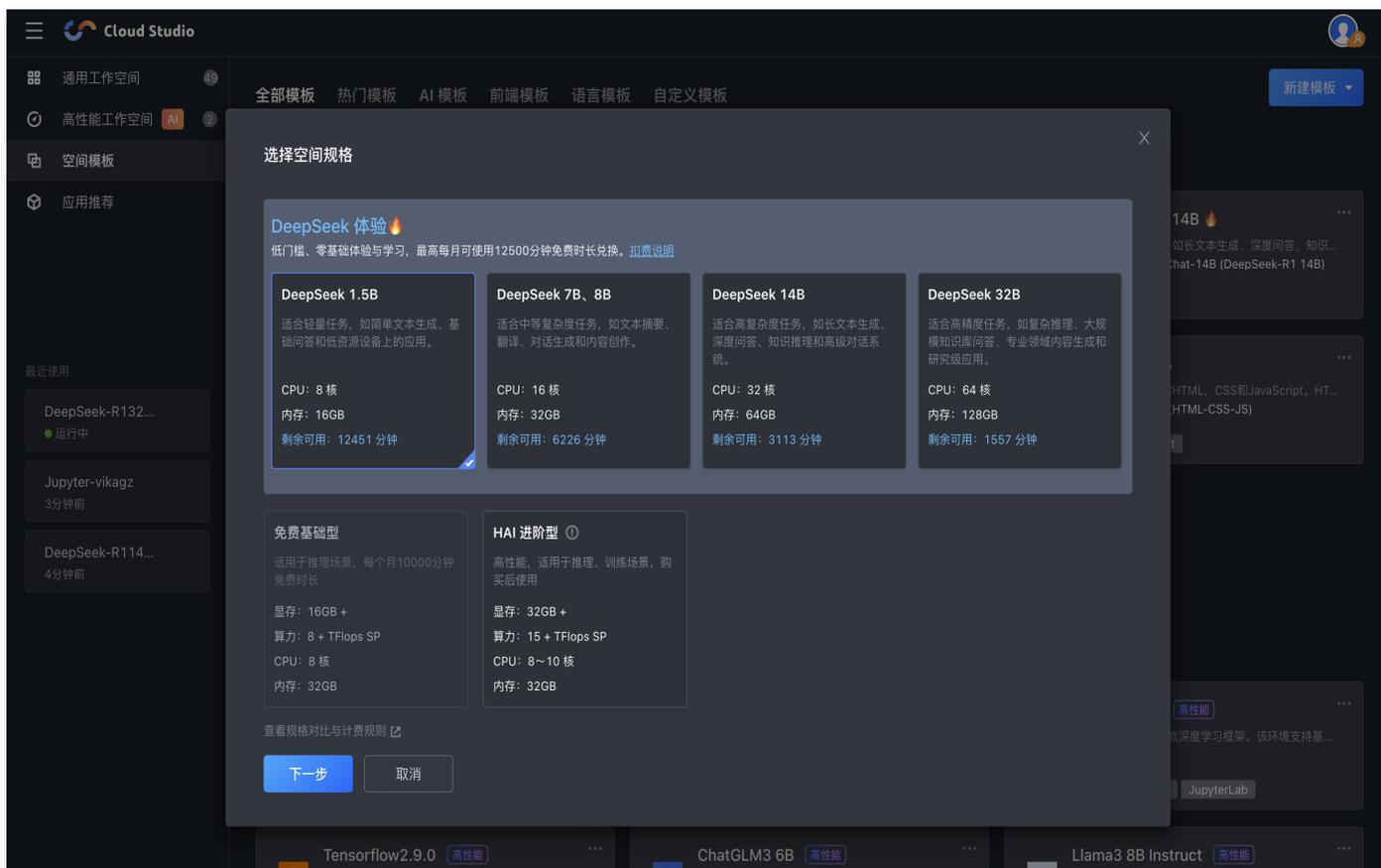
Cloud Studio 已内置 DeepSeek-R1 1.5B、7B、14B、32B 模型，并支持一键部署。

一、快速创建

1. 进入 [Cloud Studio](#)，在空间模板下单击 **DeepSeek-R1** 模板。



2. 根据需求选择不同规格（DeepSeek 体验、免费基础型、HAI 进阶型）来创建。



规格对比如下：

名称	DeepSeek 体验规格	基础型	HAI进阶型
预置环境	<ul style="list-style-type: none"> 预装 Ollama。不同规格分别预置 DeepSeek 1.5B、7B、8B、14B、32B模型。 Chatbot-Ollama 	<ul style="list-style-type: none"> 预装 Ollama、DeepSeek-R1 7B、14B、32B 模型。 支持 Open WebUI。 	
规格	<ul style="list-style-type: none"> CPU: 8~64核不等 内存: 16~128GB不等 	<ul style="list-style-type: none"> 显存: 16GB + 算力: 8 + TFlops SP CPU: 8 核 内存: 32GB 	<ul style="list-style-type: none"> 显存: 32GB + 算力: 15 + TFlops SP CPU: 8~10 核 内存: 32GB <p>选择后仍有两卡基础型、两卡进阶型可供选择，具体参见 购买页 提示。</p>
适用场景	适用于体验 DeepSeek 蒸馏模型推理。打开即可访问可视化对话界面，上手成本低。	16G显存，可用于体验 DeepSeek 7B，最大可到32B但推理速度较慢。	32G显存，可用于体验 Deepseek 14B、32B。支持模型部署。

使用费用	每月可免费兑换使用时长从12500分钟到1562分钟不等。具体费用可参考 购买指南 。	默认使用每月10000分钟高性能工作空间免费时长，与通用工作空间免费时长分别计算互不影响。暂不支持付费增加时长。	该规格无法使用每月10000分钟高性能工作空间免费时长，需跳转到 HAI 购买，具体费用可参见 计费说明 。
------	---	--	--

说明：

基础型资源有限且为动态库存，如出现库存不足提示，可稍后重试或选用其他规格。

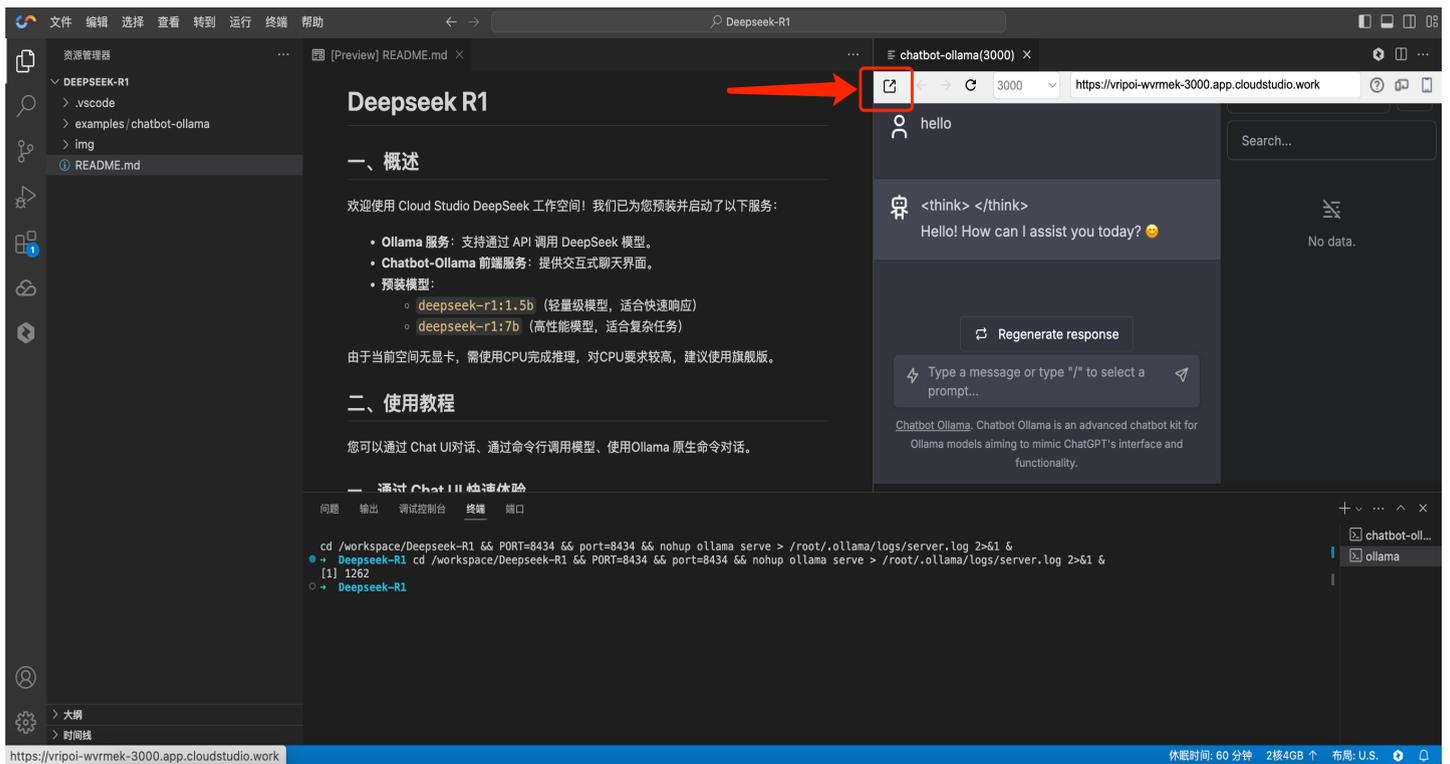
二、使用轻量型进行对话与临时部署

轻量型已为您预装并启动了以下服务：

- **Ollama 服务：**支持通过 API 调用 DeepSeek 模型。
- **Chatbot-Ollama 前端服务：**提供交互式聊天界面。
- **预装模型：**
 - DeepSeek-R1-Distill-Qwen-1.5B：适合轻量级任务，如简单文本生成、基础问答和低资源设备上的应用。
 - DeepSeek-R1-Distill-Qwen-7B：适合中等复杂度任务，如文本摘要、翻译、对话生成和内容创作。
 - DeepSeek-R1-Distill-Llama-8B：与7B类似，性能略有提升，适合更复杂的文本生成和理解任务。
 - DeepSeek-R1-Distill-Qwen-14B：适合高复杂度任务，如长文本生成、深度问答、知识推理和高级对话系统。
 - DeepSeek-R1-Distill-Qwen-32B：适合高精度任务，如复杂推理、大规模知识库问答、专业领域内容生成和研究级应用。

使用方法1：通过 Chat UI 快速体验

我们提供了一个开箱即用的聊天界面，您可以直接通过浏览器与模型交互。等待页面加载完成即可查看到下图中的界面。单击红框区域，即可新页面打开聊天页面。



操作步骤

1. 访问 Chat UI

当前环境已配置为默认启动该服务，可直接访问。如果希望修改启动命令，可以前往 `.vscode/preview.yml` 修改。

2. 选择模型

在界面右上角的下拉菜单中，选择 `deepseek-r1:1.5b` 或 `deepseek-r1:7b`。

⚠ 注意:

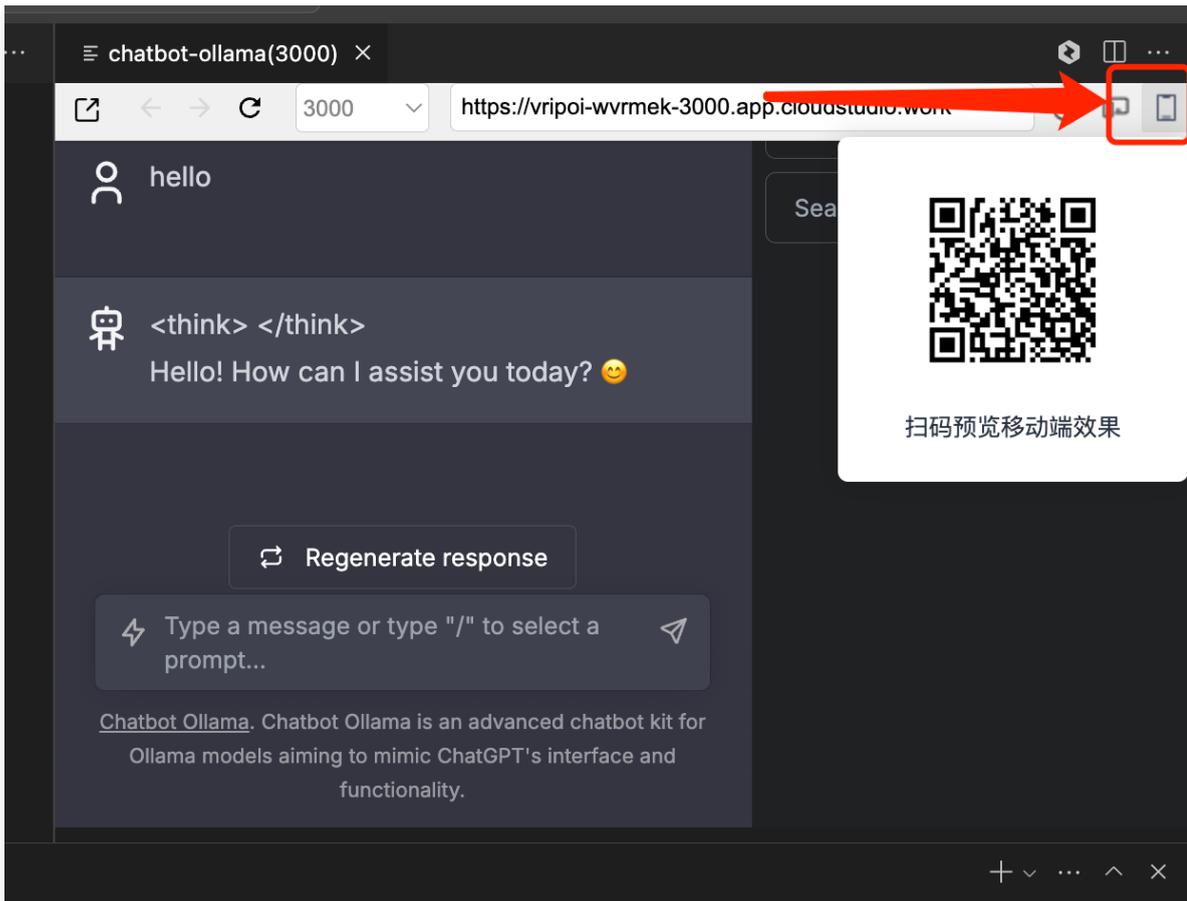
加载7B模型需要更多内存，请先切换到旗舰型后，再选择7B模型。

3. 开始对话

输入问题或指令，单击发送即可获取模型回复。

4. 在其他设备打开

您可以复制链接地址后在其他设备打开。也可 hover 下方红框处并扫码，即可在移动端打开。



示例：生成代码

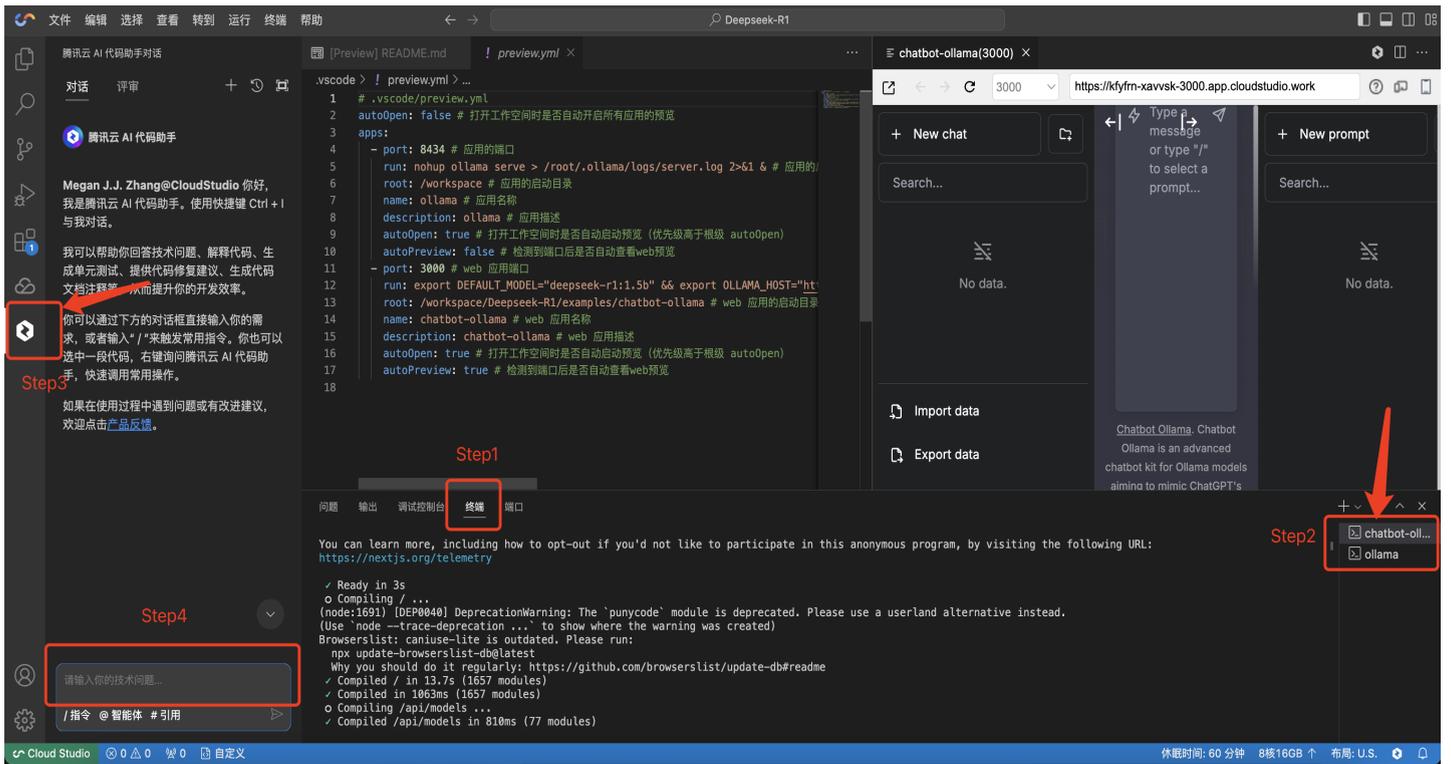
输入：

用 Python 实现一个斐波那契数列生成函数。

预期输出：

```
def fibonacci(n):  
    a, b = 0, 1  
    result = []  
    for _ in range(n):  
        result.append(a)  
        a, b = b, a + b  
    return result
```

如果遇到网页展示有问题，可返回编辑器终端，查看报错信息并询问 AI 代码助手。



使用方法2：通过命令行调用模型

您可以使用 `curl` 直接与 Ollama 服务交互，适用于自动化任务或脚本调用。

基础请求示例

调用 `deepseek-r1:1.5b`

```

curl -X POST http://localhost:8434/api/generate \
-H "Content-Type: application/json" \
-d '{
  "model": "deepseek-r1:1.5b",
  "prompt": "用一句话解释量子计算",
  "stream": false
}'
    
```

调用 `deepseek-r1:7b`

```

curl -X POST http://localhost:8434/api/generate \
-H "Content-Type: application/json" \
-d '{
  "model": "deepseek-r1:7b",
  "prompt": "写一篇关于可再生能源的短文(200字)",
  "stream": false
}'
    
```

```
}'
```

流式输出 (实时逐句返回)

将 `stream` 参数设为 `true` :

```
curl -X POST http://localhost:8434/api/generate \  
-H "Content-Type: application/json" \  
-d '{  
    "model": "deepseek-r1:7b",  
    "prompt": "详细说明如何训练一个神经网络",  
    "stream": true  
}'
```

使用方法3: Ollama 原生命令进阶使用

除了 HTTP API, 您还可以直接通过 `ollama` 命令行工具与模型交互。

1. 查看已安装模型

输入如下:

```
ollama list
```

输出示例:

NAME	SIZE
deepseek-r1:1.5b	1.5B
deepseek-r1:7b	7B

2. 启动交互式对话

启动输入如下:

```
ollama run deepseek-r1:1.5b
```

输入 `/bye` 或者 `Ctrl + D` 退出对话。

示例对话流程:

```
>>> 推荐一个适合初学者的机器学习项目
```

DeepSeek: 一个手写数字识别系统是一个不错的入门项目, 可以使用MNIST数据集和Python的scikit-learn库快速实现...

>>> 如何评估模型性能?

DeepSeek: 常用方法包括划分训练集/测试集、交叉验证、计算准确率/精确率/召回率等指标...

>>> /bye

3. 模型管理

命令	说明
<code>ollama pull deepseek-r1:7b</code>	重新拉取模型 (更新时使用)
<code>ollama rm deepseek-r1:7b</code>	删除模型

4. 切换模型

如需切换到7B模型, 可在退出当前进程后再次输入以下代码即可:

```
ollama run deepseek-r1:7b
```

⚠ 注意:

当前机器为 CPU, 7B模型推理速度较慢, 建议使用1.5B模型。更大参数模型建议使用 HAI进阶型。

常见问题

如何查看预览?

如需了解更多使用内置 web 预览的介绍, 可参考 [在线预览调试](#) 文档。

如何释放端口, 关闭某个进程?

如果需要释放端口, 关闭某个进程, 可以通过下述命令查看端口占用情况:

```
lsof -i -P -n
```

再使用下述命令关闭进程:

```
kill -9 <进程ID>
```

空间不足怎么处理?

如果右下角提示空间不足，可单击右下角的“2核4GB”，选择“旗舰型”，即可升级空间。



相关文档与资源

- Deepseek 官方介绍文档: [GitHub – deepseek-ai/DeepSeek-R1](#)
- Ollama 官方介绍文档: [GitHub – ollama/ollama](#)

更多帮助和支持

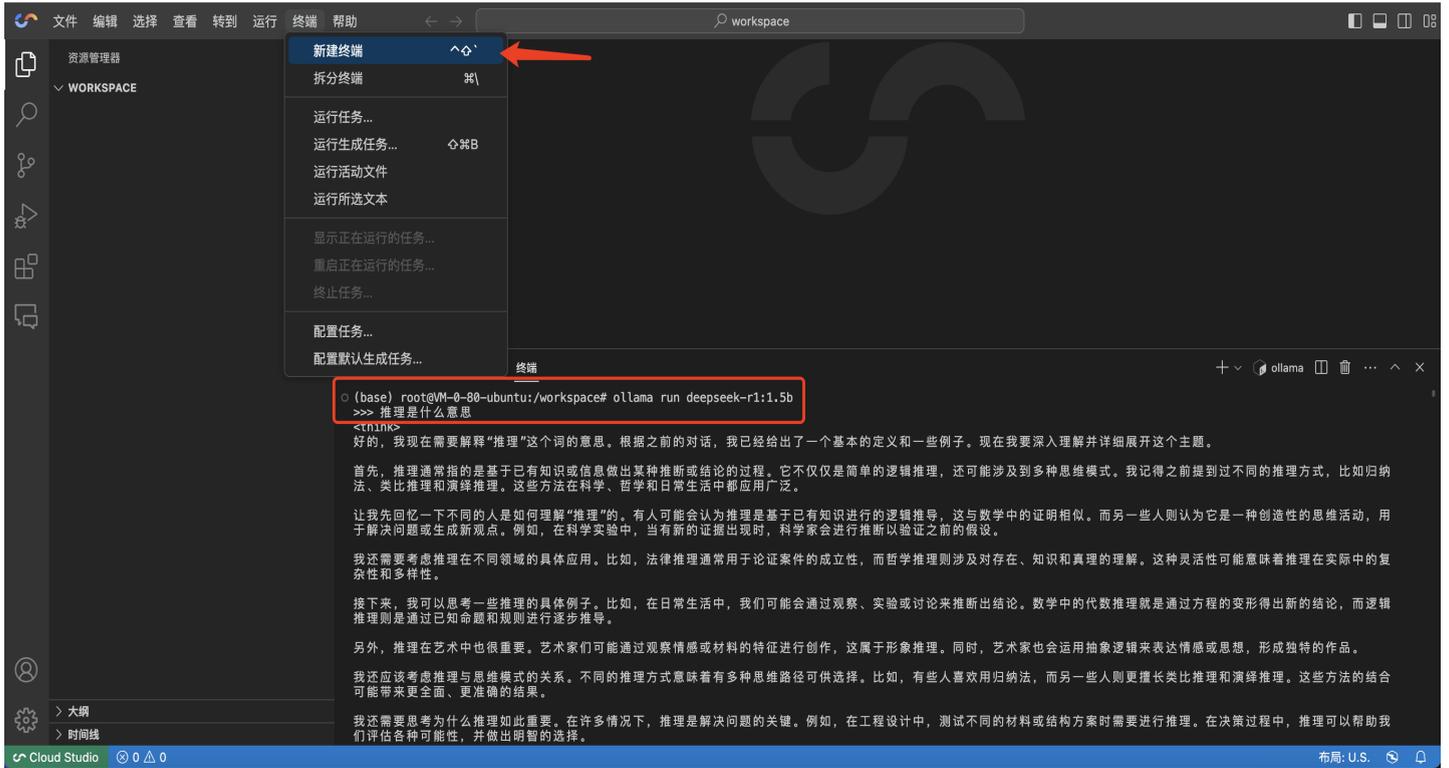
欢迎加入 Cloud Studio 用户反馈群。当您遇到问题需要处理时，您可以直接通过扫码进入 Cloud Studio 用户群进行提问。

- 腾讯云工程师实时群内答疑。
- 扫码入群可先享受产品上新功能。
- 更多精彩活动群内优享。



三、使用基础型进行对话

高性能工作空间开机自启动 Ollama 服务，可单击**新建终端**。



输入以下命令查看已下载模型并启动：

```
ollama list # 查看当前已经下载的模型，当前工作空间已预装 DeepSeek-R1 1.5B、7B、8B、14B、32B 五种模型。
ollama run [NAME] # 运行想要启动的模型，NAME是模型的名称，如 ollama run deepseek-r1:32b。
```

常见问题

其他常见问题可参考 [常见问题](#)。

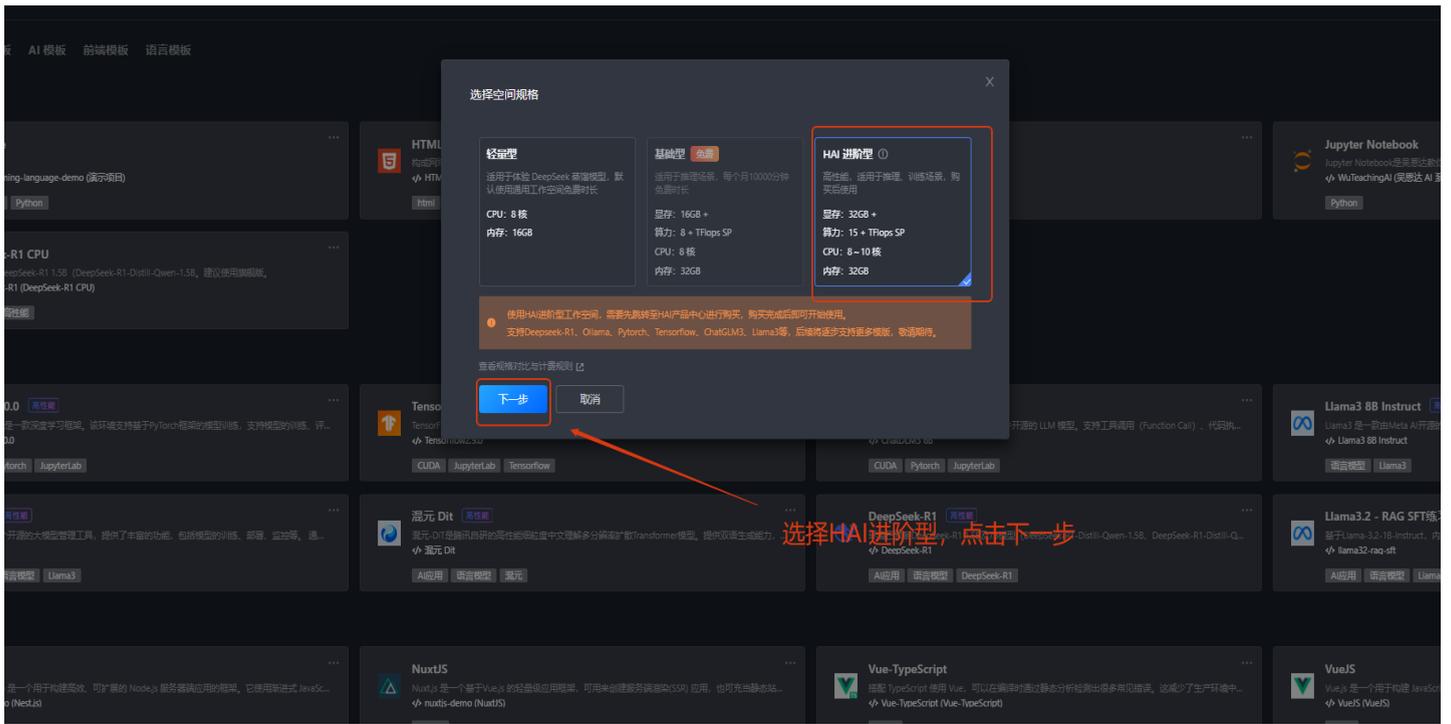
四、使用进阶型进行对话与部署

1. 购买算力

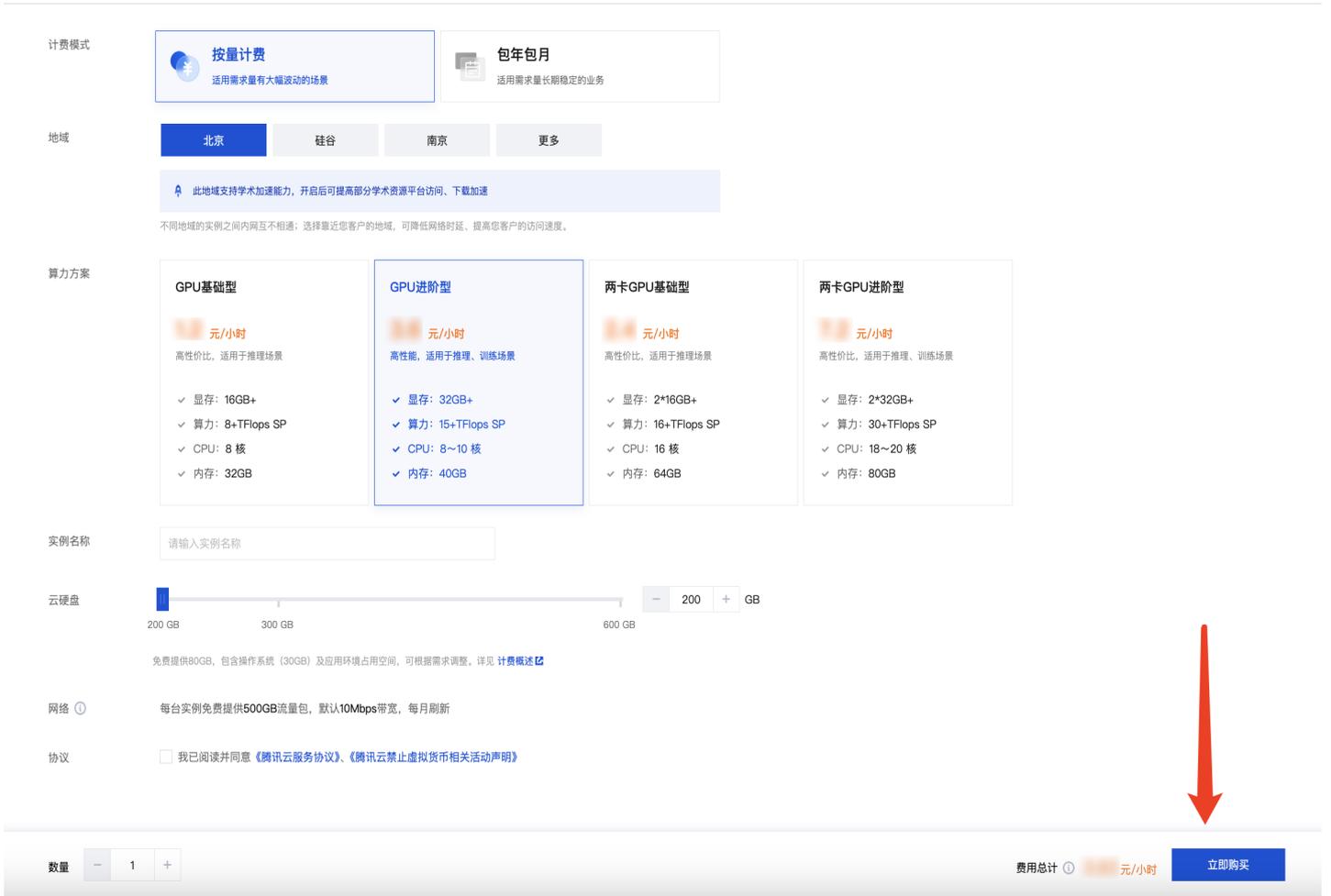
选择 HAI 进阶型后单击下一步。如遇到授权提示，根据提示信息完成授权即可。

⚠ 注意：

当前不支持协作者账号授权访问，请使用主账号或子账号。



根据个人意向配置选购方案，填写实例名称，阅读并勾选协议，立即购买后，自动跳转到 Cloud Studio 创建工作空间。如账户余额为负请及时续费。

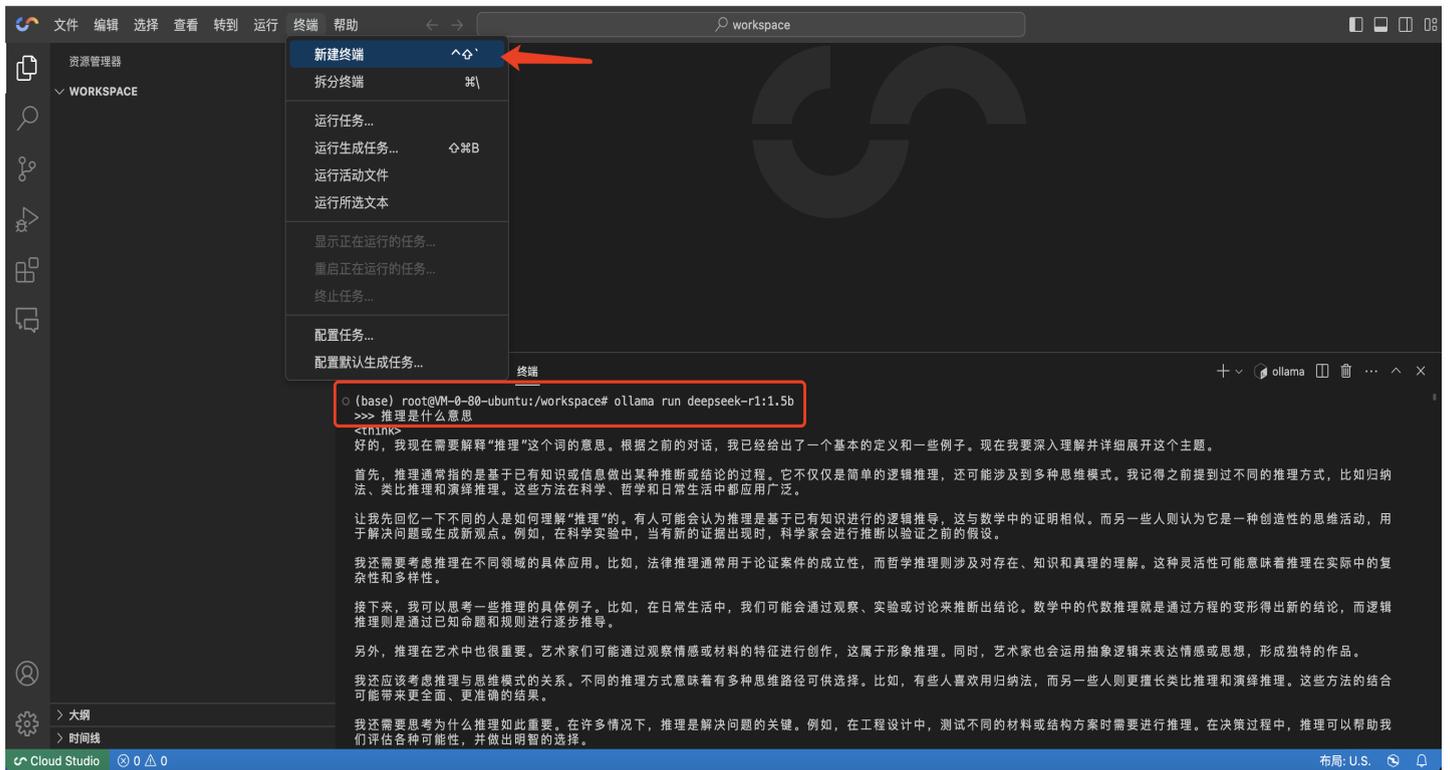


HAI 实例开机需要大约2-5分钟，状态修改为运行中，即可点击进入。



2. 在终端对话

与基础型类似，打开后新建终端并输入命令即可完成对话。



3. 使用 Chatbot UI / Openweb UI 对话

注意:
该功能仅限付费版可用。

目前进阶型空间也具备 Chatbot UI 对话框功能，其体验路径为：

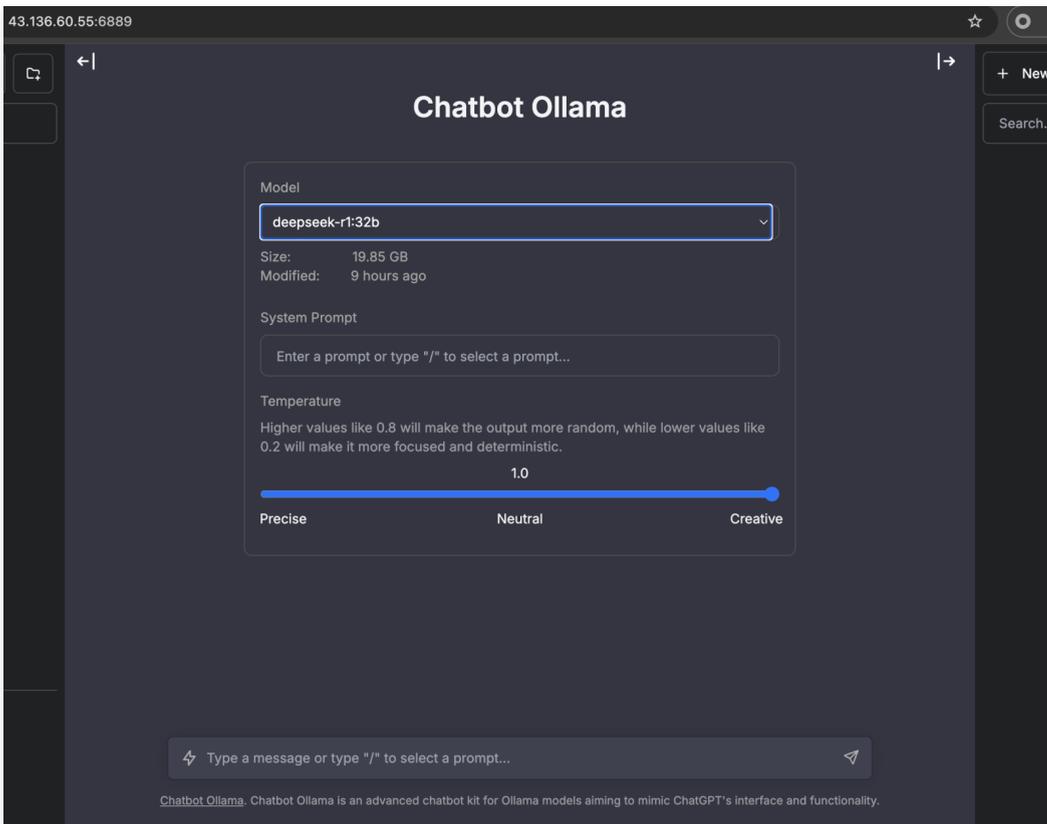
1. 在高性能工作空间列表中，单击已创建的进阶型空间最右侧图标，进入 HAI资源控制台。



2. 单击 ChatbotUI 图标或 OpenWebUI 图标。



3. 选择模型, 进行对话。



更多进阶型操作指南

HAI 实践文档：包含个人支持库搭建指引等。

[OpenwebUI 文档](#)：包含联网搜索指引等。