

一，下载 / 安装 ollama 工具 : <https://ollama.com>

[Blog](#)[Discord](#)[GitHub](#)[Models](#)

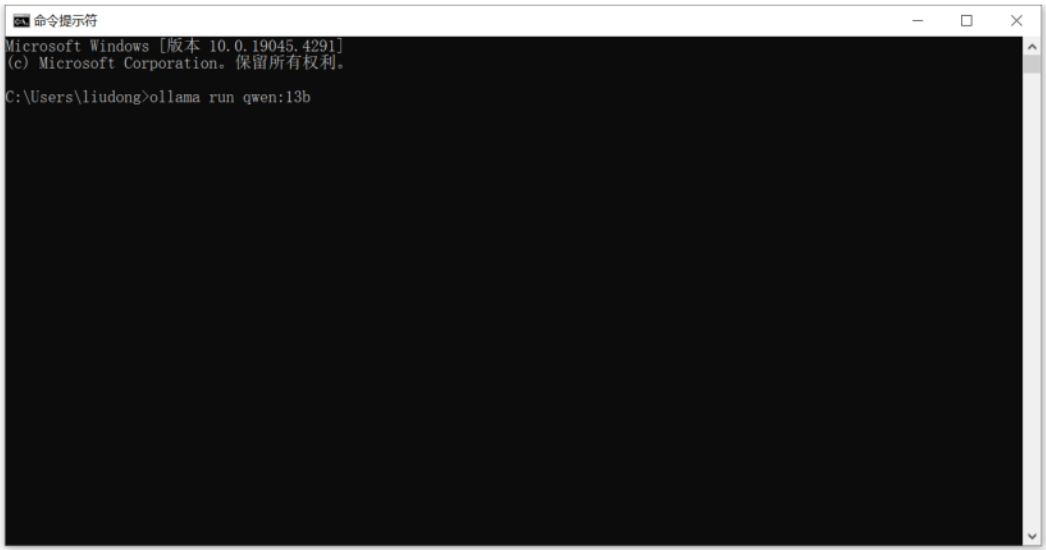
Get up and running with large  
language models.

Run [Llama 3](#), [Phi 3](#), [Mistral](#), [Gemma](#), and  
other models. Customize and create your  
own.

Download ↓

Available for macOS, Linux,  
and Windows (preview)

二，打开电脑终端，使用 ollama 下载通义千问模型和 embedding 模型



- 1. 输入 ollama run qwen:14b - 安装大模型
- 2. 输入 ollama pull nomic-embed-text - 安装嵌入模型

Model	Parameters	Size	Download
Llama 3	8B	4.7GB	ollama run llama3
Llama 3	70B	40GB	ollama run llama3:70b
Phi-3	3,8B	2.3GB	ollama run phi3
Mistral	7B	4.1GB	ollama run mistral
Neural Chat	7B	4.1GB	ollama run neural-chat
Starling	7B	4.1GB	ollama run starling-1m
Code Llama	7B	3.8GB	ollama run codellama
Llama 2 Uncensored	7B	3.8GB	ollama run llama2-uncensored
LLaVA	7B	4.5GB	ollama run llava
Gemma	2B	1.4GB	ollama run gemma:2b
Gemma	7B	4.8GB	ollama run gemma:7b
Solar	10.7B	6.1GB	ollama run solar

其他模型安装方式

三，启动 ollama 服务，后续供 anythingLLM 调用

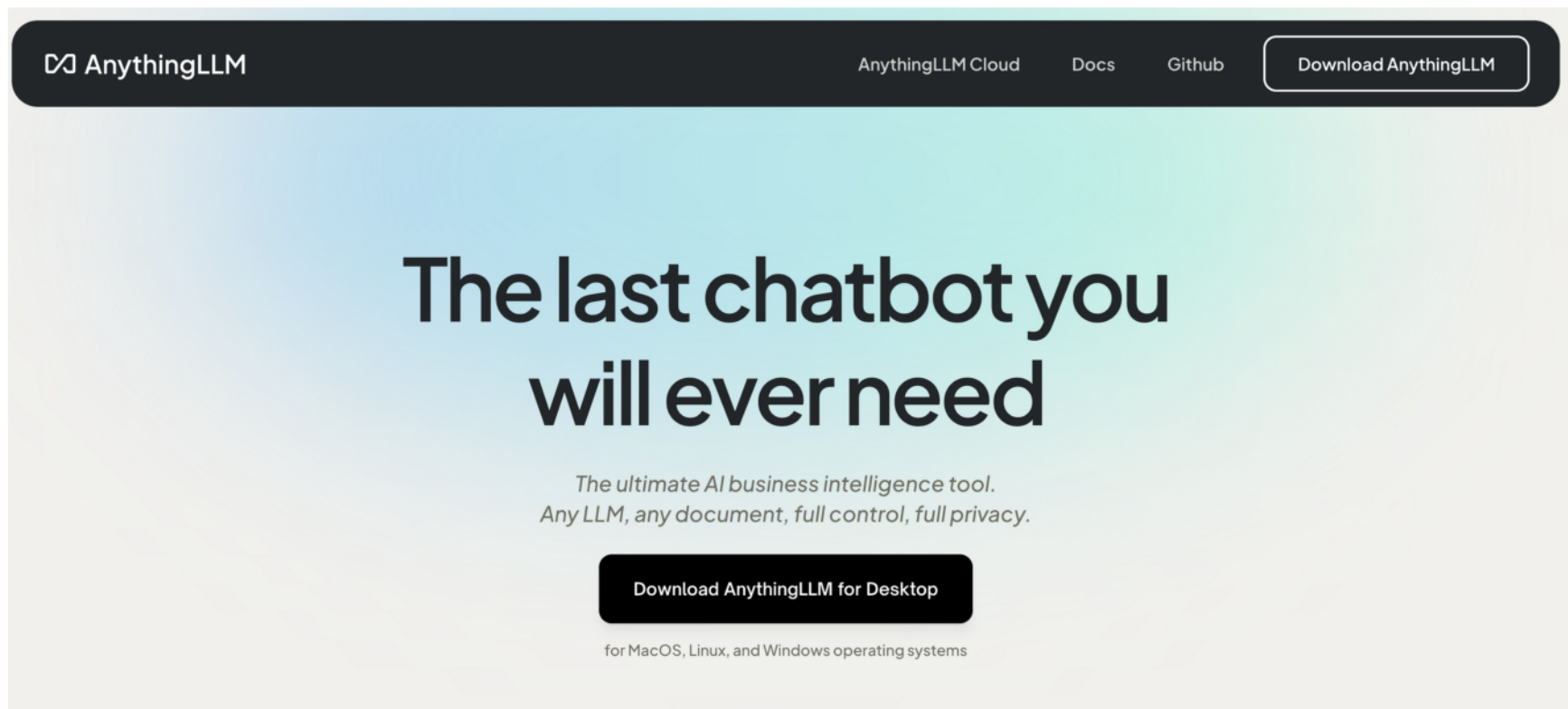
在终端中输入：ollama serve

```
(base) liudong@192 ~ % ollama pull nomic-embed-text
pulling manifest
pulling 970aa74c0a90... 100% 274 MB
pulling c71d239df917... 100% 11 KB
pulling ce4a164fc046... 100% 17 B
pulling 31df23ea7daa... 100% 420 B
verifying sha256 digest
writing manifest
removing any unused layers
success
(base) liudong@192 ~ % ollama serve
```

输入：ollama serve

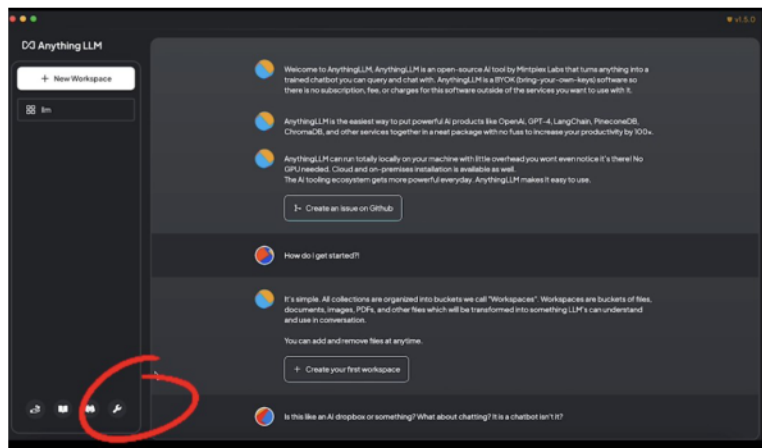
## 四，安装 anythingLLM

下载地址：<https://useanything.com/>

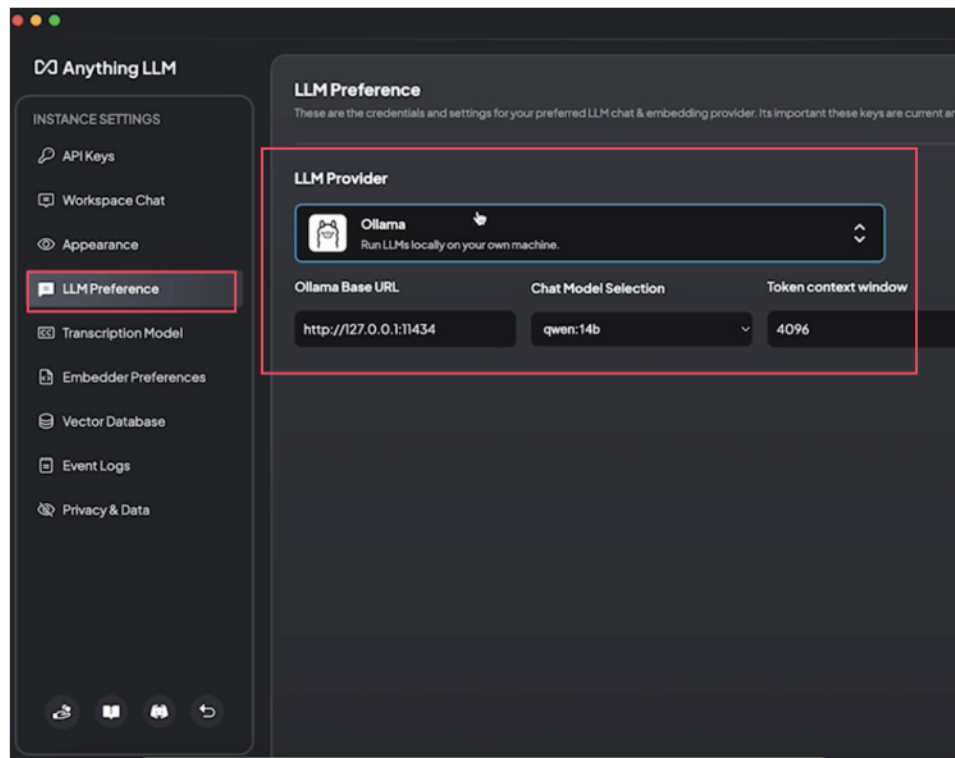


根据电脑型号，选择合适的版本

## 五, anything LLM 相关参数配置



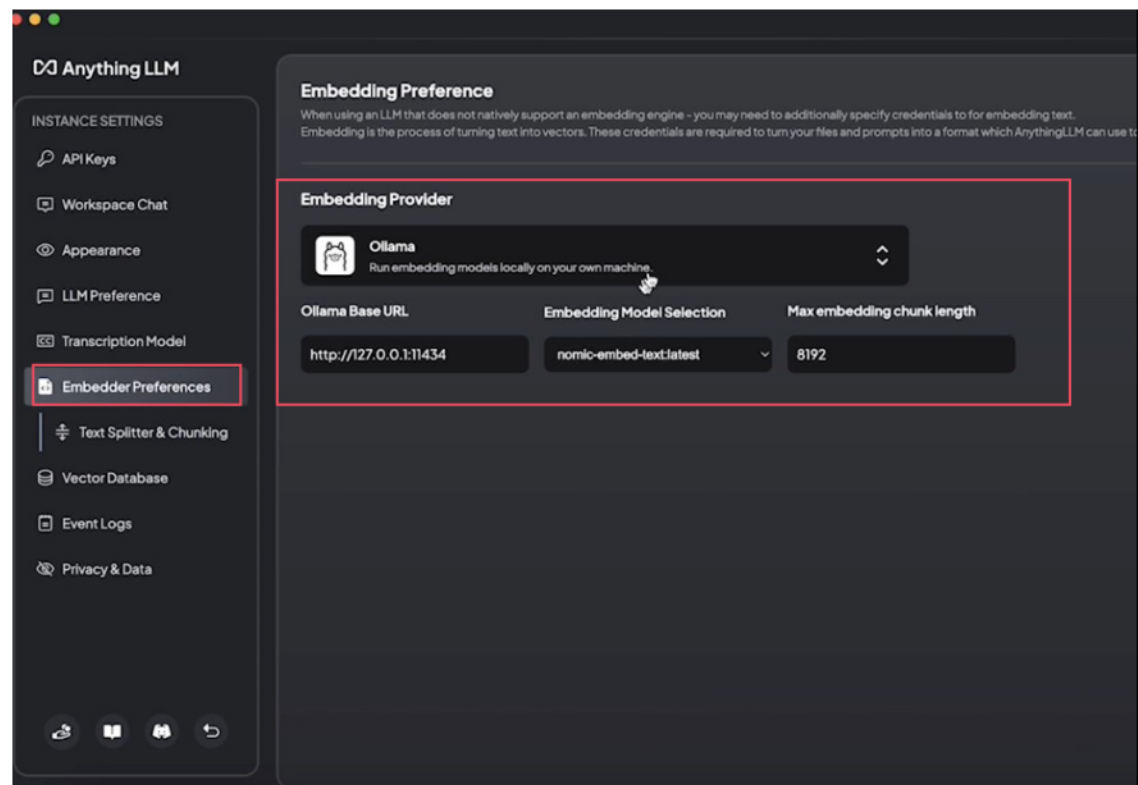
点击这个地方, 进入设置



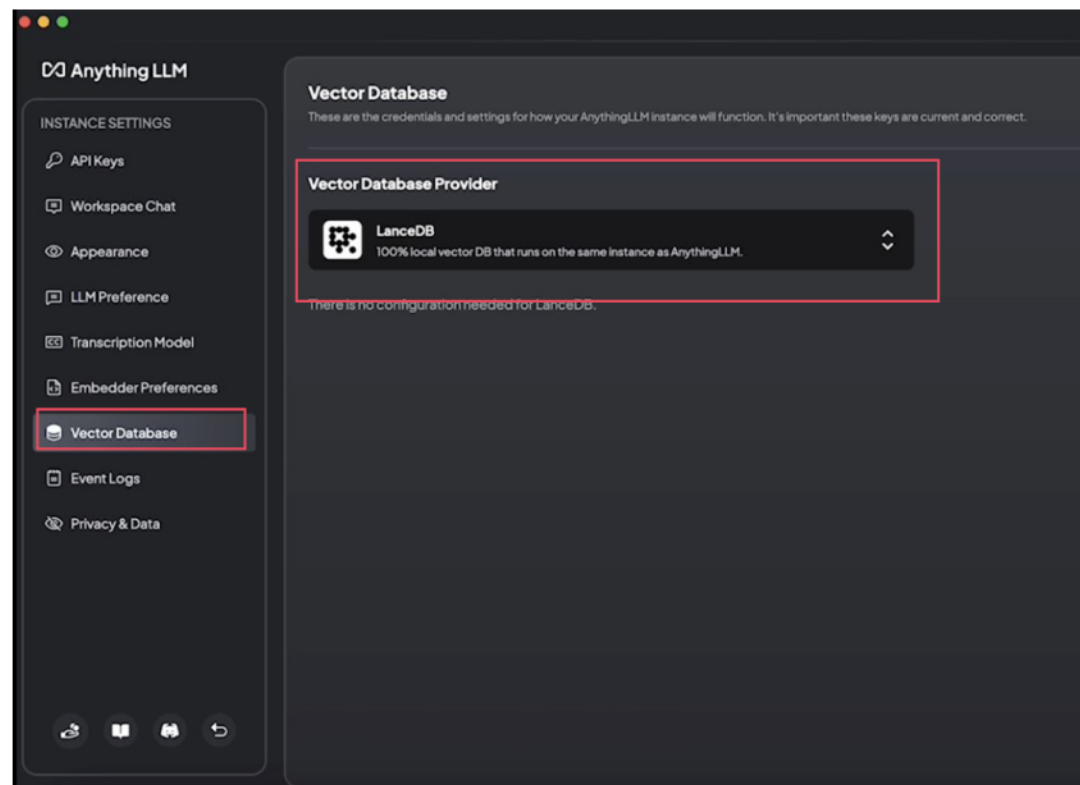
按照截图中的配置即可

配置大模型

## 六，anything LLM 相关参数配置

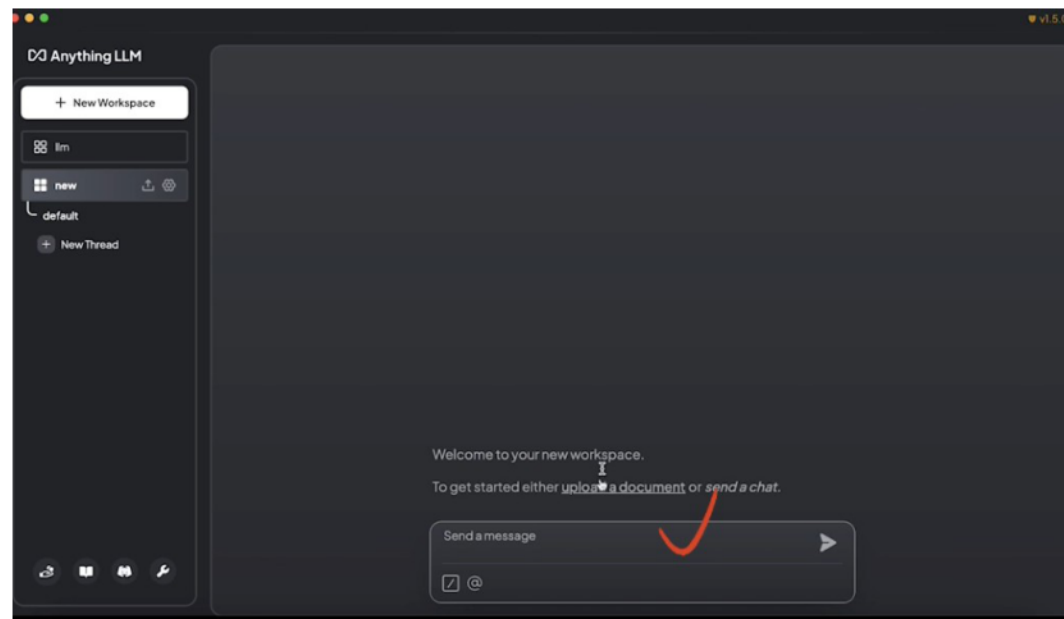
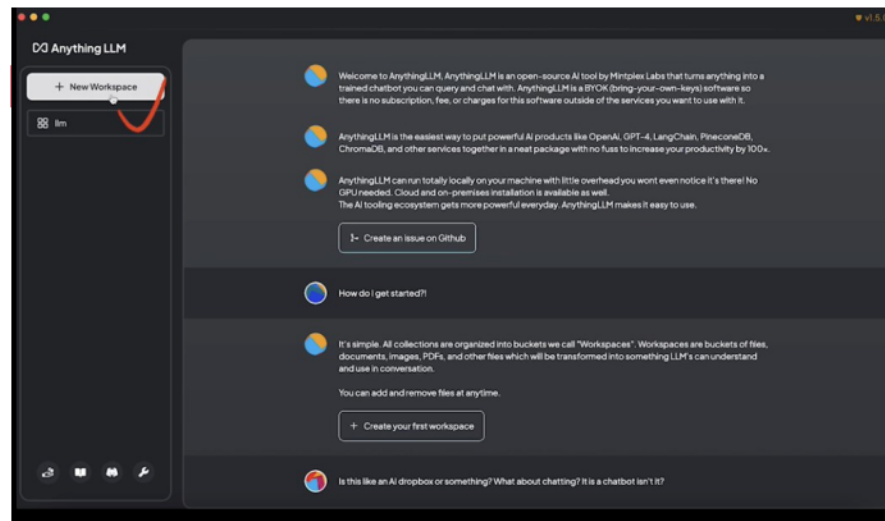


配置嵌入模型



配置向量数据库

## 七，知识库测试



上传文档

