

面向审计行业 DeepSeek 大模型操作指南

版本 1.0 | 适用对象：审计从业人员

南京审计大学

计算机学院大模型团队提供

2025 年 2 月 8 日

目录

1 DeepSeek 基本概况	3
2. DeepSeek 主要版本	4
3. DeepSeek 审计能力	5
4. DeepSeek 部署方法	6
4.1 官方渠道	6
4.1.1 网页版使用	6
4.1.2 手机版使用	8
4.2 第三方渠道	8
4.2.1 硅基流动&华为云	8
4.2.2 纳米 AI 搜索	9
4.2.3 阿里云	10
4.2.4 百度智能云	11
4.2.5 火山引擎	11
4.2.6 其他平台	12
4.3 本地部署	12
4.3.1 下载 ollama	13
4.3.2 合适版本安装	13
4.3.3 输入安装代码	15
4.3.4 测试部署模型	17
4.3.5 部署非量化模型	18
5. DeepSeek 审计助手	20
5.1 基础操作场景	20
5.2 审计工作辅助	21
5.3 审计学习考试	22
5.4 其他提示	23

1. DeepSeek 基本概况

DeepSeek 是由杭州深度求索人工智能基础技术研究有限公司（简称“深度求索”）开发的一系列人工智能模型。该模型拥有数以亿计甚至更多的参数，通过在海量文本数据上进行预训练，学习到丰富的语言结构和语义信息；并支持智能对话、准确翻译、创意写作、高效编程、智能解题和文件解读等多种功能。其“深度思考”和“联网搜索”功能能够更全面地理解用户问题并提供准确答案。

杭州深度求索人工智能基础技术研究有限公司公司成立于 2023 年 7 月 17 日，专注于开发先进的大语言模型（LLM）和相关技术。自成立以来，公司在 AI 领域取得了显著成果，主要使用数据蒸馏技术，得到更为精炼、有用的数据。2024 年 1 月 5 日，发布 DeepSeek LLM（深度求索的第一个大模型），目前，DeepSeek-R1、V3、Coder 等系列模型已上线国家超算互联网平台。英伟达称，DeepSeek-R1 是最先进的大语言模型，亚马逊和微软也接入 DeepSeek-R1 模型。DeepSeek 大模型在多个基准测试中表现优异，尤其是在代码和数学任务上，超越了其他开源模型，甚至与领先的闭源模型（如 GPT-4 和 Claude-3.5-Sonnet）不相上下。

DeepSeek 被业界认为“以高性价比著称的 AI 模型服务商”，原因是这家公司的出现极大地降低了大模型训练和应用的成本，如该公司开发的 DeepSeek-V3 训练成本仅 557.6 万美元，而 OpenAI 训练 GPT-4 所花费的成本高达 7800 万美元甚至是 1 亿美元，双方的成本相差至少 10 倍。DeepSeek-V3 在数学、代码能力和中文知识问答方面还超过了 GPT-4，可以说是性价比超高。此外，DeepSeek 团队只有 139 名研发人员，而开发 GPT 的 OpenAI 团队则有 1200 名研究人员。

在审计领域，DeepSeek 大模型能够帮助审计人员高效处理各类多源异构的审计数据、识别风险、提升审计质量；通过自动化的数据处理、智能化的风险识别和定制化的报告生成等功能，帮助审计人员降低人工成本、提高审计质量和效率。

2. DeepSeek 主要版本

目前，DeepSeek 的核心版本主要有 DeepSeek-V3、DeepSeek-R1、Janus Pro，表 1 中列出了这 3 个核心版本的特点和适用场景。

表 1 DeepSeek 核心版本与适用场景

模型版本	发行时间	模型大小	核心能力	适用场景示例
DeepSeek-V3	2024-12-26	671B	通用自然语言处理（NLP），支持长文本理解、多语言交互	合同条款解析、政策法规匹配、审计报告生成
DeepSeek-R1	2025-1-20	671B	复杂逻辑推理，强化数学与代码生成能力	财务数据分析、异常检测、风险建模
DeepSeek Janus Pro	2025-1-28	7B	多模态（文本+图像+表格）理解与生成	票据识别、图表数据关联分析、可视化报告

这三个版本原始模型权重已经在 [hugging face](#) 上开源，用户可以免费下载。国内使用可以通过其镜像（<https://hf-mirror.com/>）获取。DeepSeek-V3 和 DeepSeek-R1 的模型参数量较大，达到了 671B，直接部署这两个模型需要 1.3~2 TB（FP16）的显存支持（如 128 卡 H100 的集群）。为方便一般用户本地使用，DeepSeek 团队使用 Qwen2.5 和 Llama3.3，以 DeepSeek-R1 为教师模型，蒸馏了 6 款小模型，包含 1.5B~70B 在内共有 6 个尺寸，如表 2 所示。

表 2 DeepSeek-R1 蒸馏的 6 个尺寸的模型

蒸馏的模型	基座模型	下载地址
DeepSeek-R1-Distill-Qwen-1.5B	Qwen2.5-Math-1.5B	HuggingFace
DeepSeek-R1-Distill-Qwen-7B	Qwen2.5-Math-7B	HuggingFace
DeepSeek-R1-Distill-Llama-8B	Llama-3.1-8B	HuggingFace
DeepSeek-R1-Distill-Qwen-14B	Qwen2.5-14B	HuggingFace
DeepSeek-R1-Distill-Qwen-32B	Qwen2.5-32B	HuggingFace
DeepSeek-R1-Distill-Llama-70B	Llama-3.3-70B-Instruct	HuggingFace

即使经过了蒸馏，7B 模型也需要 20~25G 的显存，即使是 24G 的 4090 显卡，部署也存在一定的风险。为此，在个人使用时，很多会将这类模型进行进一步量

化，以缩减模型大小，ollama 官方拉取的 DeepSeek 模型即是通过 4bit 量化后的模型。

这里需要注意：无论是模型蒸馏还是量化，都会或多或少降低模型的能力。

3. DeepSeek 审计能力

（一）数据采集与预处理

DeepSeek 支持多种数据源的接入，包括财务系统、ERP 系统和数据库等，确保数据获取的全面性。

通过数据清洗、缺失值填补、异常值检测和格式转换等操作，DeepSeek 能够自动清洗、转换和整合数据，确保数据质量，并将不同来源的数据统一格式化，为后续分析提供高质量的数据基础。

（二）数据分析与挖掘

DeepSeek 提供多种数据分析工具，如趋势分析、比率分析和异常检测等，帮助审计人员快速识别潜在的风险区域。DeepSeek 还可以进行时序分析，揭示财务数据中的潜在问题。

DeepSeek 利用机器学习算法识别潜在风险和异常交易，通过结合历史数据训练风险识别模型，实时监控异常交易、非正常模式和潜在的舞弊行为。

（三）支持自定义分析模型

用户可以根据具体审计需求自定义分析模型，针对特定场景（如税务审计、资产管理审计等）设定独特的分析参数。

（四）风险识别与评估

DeepSeek 基于预设规则和机器学习模型识别潜在风险领域，通过预设的审计规则和数据驱动的机器学习模型，自动识别潜在风险区域，帮助审计人员发现财务漏洞、操作风险或法律风险。

DeepSeek 会对识别出的风险进行评估和排序，根据风险的严重程度、发生概率和影响范围，自动评估并排序，帮助审计人员优先处理最关键的风险点。

（五）审计证据收集与管理

通过 DeepSeek 的自动化分析，系统能够生成详细的审计底稿，包括审计过程、分析方法、数据来源及审计结果等内容，确保审计工作的透明性和可追溯性。

DeepSeek 支持审计证据的电子化存储和管理，审计证据以电子形式存储，支持文档管理、版本控制和权限管理，方便审计人员快速查阅和追溯。

（六）可视化与报告生成

DeepSeek 提供丰富的可视化图表，包括图表、热力图和流程图等，帮助审计人员直观展示分析结果，提升报告的可读性和说服力。

系统能够自动生成标准化的审计报告，包含详细的数据分析结果、风险评估和审计结论等内容，显著减少报告编写时间。

4. DeepSeek 部署方法

使用 DeepSeek 主要有三种渠道：官方渠道、第三方渠道、本地部署。这三种渠道各自特点如表 3 所示。

表 3 DeepSeek 三种使用渠道对比

渠道	优点	缺点
官方渠道	功能齐全、操作简单（联网搜索/跨设备同步）	高峰期易崩溃，取决于流量，看运气
第三方渠道	规避官方崩溃风险，国产 GPU 加速或白嫖算力	功能受限（如对话记录不保存），需实名认证/复杂配置
本地部署	隐私性强、永久离线，定制化模型选择	依赖硬件性能（需高配电脑），技术门槛较高，大部分部署的是蒸馏版本

4.1 官方渠道

DeepSeek 官方分为网页版和移动版，网页版用户点击“开始对话”并注册后即可使用；移动版需通过手机下载注册后使用，两者功能相同。

4.1.1 网页版使用

访问链接：<https://chat.DeepSeek.com/>，可以在任何设备和浏览器打开。之前从未登录过的用户需要进行登陆，使用手机号、微信或者邮箱登陆即可。如图 1

所示，输入自己的手机号，点击发送验证码，然后通过接收到的验证码登录即可。



The image shows the DeepSeek login interface. At the top is the DeepSeek logo. Below it are two tabs: '验证码登录' (Login with verification code) and '密码登录' (Login with password). A message states: '您所在地区仅支持 手机号 / 微信 / 邮箱 登录' (Your region only supports login with phone number / WeChat / email). There is a text input field for a phone number with a '+86' country code and a '请输入手机号' (Please enter phone number) placeholder. Below this is another input field for a verification code with a '#' placeholder and a '请输入验证码' (Please enter verification code) placeholder, followed by a '发送验证码' (Send verification code) button. A checkbox is present with the text: '我已阅读并同意 用户协议 与 隐私政策, 未注册的手机号将自动注册' (I have read and agree to the User Agreement and Privacy Policy, unregistered phone numbers will be automatically registered). A large blue '登录' (Login) button is below the checkbox. At the bottom, there is a separator '或' (or) and a button with a WeChat icon and the text '使用微信扫码登录' (Use WeChat QR code to login).

图 1 DeepSeek 注册页面

登录成功后，进入图 2 所示界面，然后点击“开始对话”就可以使用。



图 2 DeepSeek 官方网站主界面

不过需要注意，那就是如何选择 V3 还是 R1 模型，可以参考下图。此外还

可根据需要，选择是否勾选“联网搜索”。

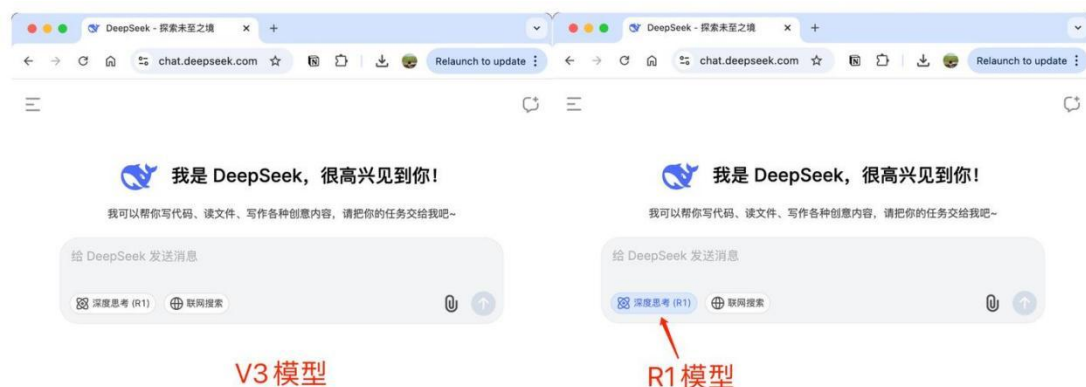


图 3 不同版本的 DeepSeek 选择

4.1.2 手机版使用

手机版的使用和电脑版基本一样，根据需要是否激活 R1 即可。唯一不同的是需要在手机安装对应的 App。安装方法如下：

方法 1：进入 DeepSeek 官网后，将鼠标移动至“获取手机 APP”处，扫描弹出的 APP 下载二维码（<https://download.DeepSeek.com/app/>），然后选选择对应的下载渠道即可。

方法 2：在手机自带的应用商城中，搜索 DeepSeek，点击下载安装即可。

4.2 第三方渠道

随着 DeepSeek 模型迅速走红，官方平台面临访问量激增的压力，经常遇到服务拥堵的情况。不过，国内主流云计算平台已全面接入 DeepSeek 模型，为用户提供了稳定可靠的替代方案。这些非官方渠道提供了三类模型选择：DeepSeek-V3 模型、完整版 DeepSeek-R1 模型（671B 参数）、轻量级 DeepSeek-R1 模型（参数规模从 1.5B 到 70B 不等）。其中，完整版 R1 模型保留了全部 671B 参数，能发挥出最佳性能，但对计算资源要求较高，通常需要支付一定费用。轻量级模型则通过知识蒸馏技术，在保持核心功能的同时大幅降低了参数规模，可在普通算力环境下流畅运行，为用户提供了更灵活的选择。

4.2.1 硅基流动&华为云

硅基流动与华为云团队联合首发并上线了基于华为云昇腾云服务的 DeepSeek R1/V3, 推理服务目前支持 V3 和 R1 大模型, 以及多款 R1 蒸馏小模型。

在硅基流动的一站式大模型云服务平台 SiliconCloud 上 (网址为: <https://siliconflow.cn/zh-cn/>), 用户注册后可以在网页右侧选择 DeepSeek-R1 等模型进行体验使用, 如图 4 所示。

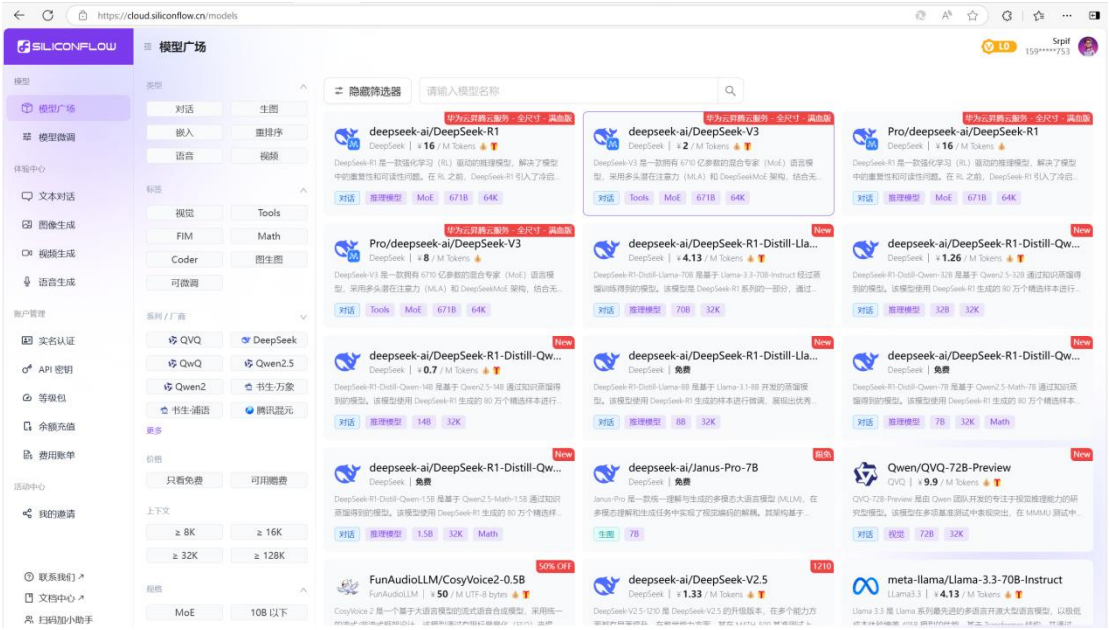


图 4 硅基流动模型广场主界面

4.2.2 纳米 AI 搜索

360 宣布在其旗下纳米 AI 搜索中开通 “DeepSeek 高速专线”, 用户可在手机应用商店中下载并安装 “纳米 AI 搜索”, 注册登录后点击底部 “大模型”, 进入如图 5 所示左边画面。随后选择 “DeepSeek-R1-满血版高速专线”, 即可进入图 5 所示右边画面。接着就可以在下方的输入框中输入你需要的问题了。

大模型

搜索模型

热门

DeepSeek

智脑

豆包



DeepSeek-R1-满血版高速专线

在数学、代码、自然语言推理等任务上性能比肩 o...



DeepSeek-R1-360高速专线

在数学、代码、自然语言推理等任务上性能比肩 o...



AI助手（混合模型）

自动调度最强大模型回答问题，满足多场景交互



智脑（360gpt-Pro）

擅长搜索、总结、思维导图，生成速度快，意图识...



文心一言（ERNIE-4.0-Turbo-8K）

适合内容创作、对比判断类知识问答领域，具备超...



多模型协作

文心一言（ERNIE-4.0-Turbo-8K）、智脑（360g...



多模型辩论

豆包（Doubao-Pro-32k）、文心一言（ERNIE-4...



AI搜索



AI视频



AI工具



大模型

< DeepSeek-R1-满血版高速专线



DeepSeek-R1-满血版高速专线

本模型为DeepSeek-R1-671B全尺寸版本，由华为 910B GPU服务器提供推理加速，更擅长代码编程、数学计算和逻辑推理。老周决定这两天大酬宾，给大家免费使用，快来薅羊毛！

试着问问我：

如何用5升和6升的水壶从池塘取3升水？



以唐山的粗钢产量，能产多少根金箍棒？



怎么0基础做款酷炫的多人竞技游戏？



联网搜索



对话历史



新话题



输入任何问题

以上内容均由AI大模型生成，仅供参考

图 5 360 纳米 AI 搜索 app

4.2.3 阿里云

阿里云 PAI Model Gallery 支持用户通过云平台一键部署 DeepSeek-V3、DeepSeek-R1 模型及其蒸馏版本。用户可以根据业务需求选择部署不同参数量的模型。具体请参考链接：

<https://help.aliyun.com/zh/pai/user-guide/one-click-deployment-DeepSeek-v3-model>

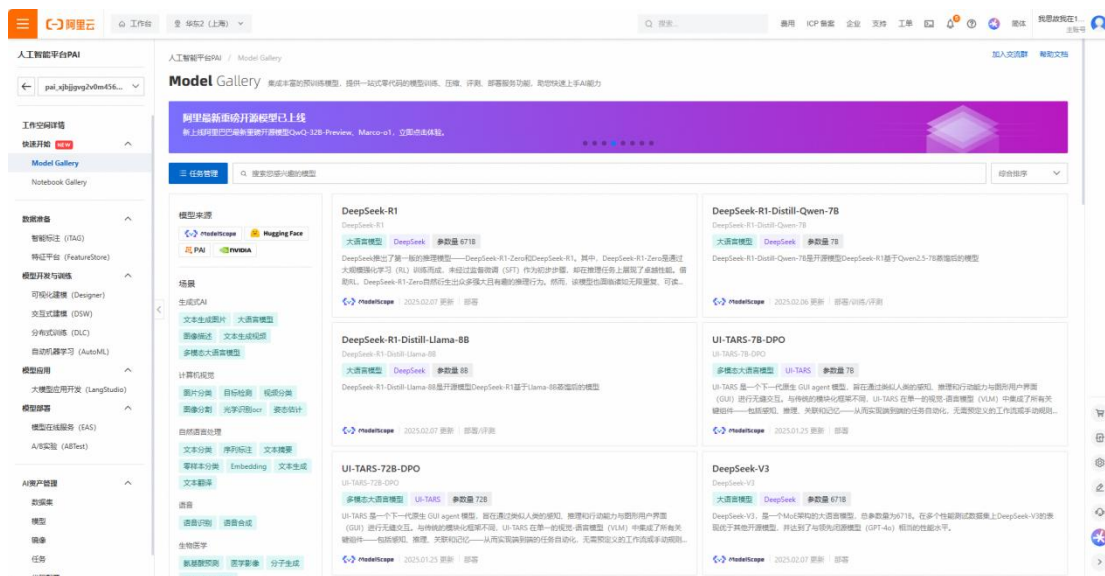


图 6 阿里云模型界面

4.2.4 百度智能云

百度智能云提供两种使用 DeepSeek 模型的途径：一是在模型广场调用 DeepSeek 的 V3 和 R1 模型的 API；二是在体验中心与这两款模型直接对话。其使用网址为：https://cloud.baidu.com/product-s/qianfan_modelbuilder。

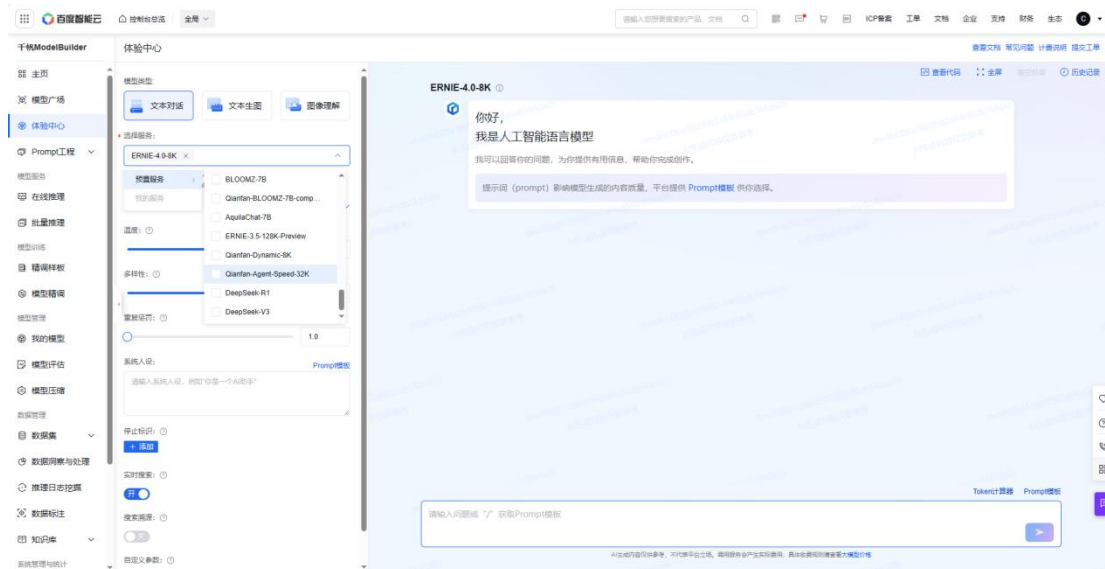


图 7 百度智能云会话界面

4.2.5 火山引擎

火山引擎支持 V3/R1 等不同规模的 DeepSeek 开源模型，用户可以通过两种方式使用这些模型。一是在火山引擎机器学习平台 veMLP 中部署，适用于自己

进行模型定制、部署、推理的企业；二是在火山方舟中调用模型，适用于期望通过 API 快速集成预训练模型的企业，目前已经支持 4 个模型版本。其访问的网络地址为：<https://www.volcengine.com/product/ark>

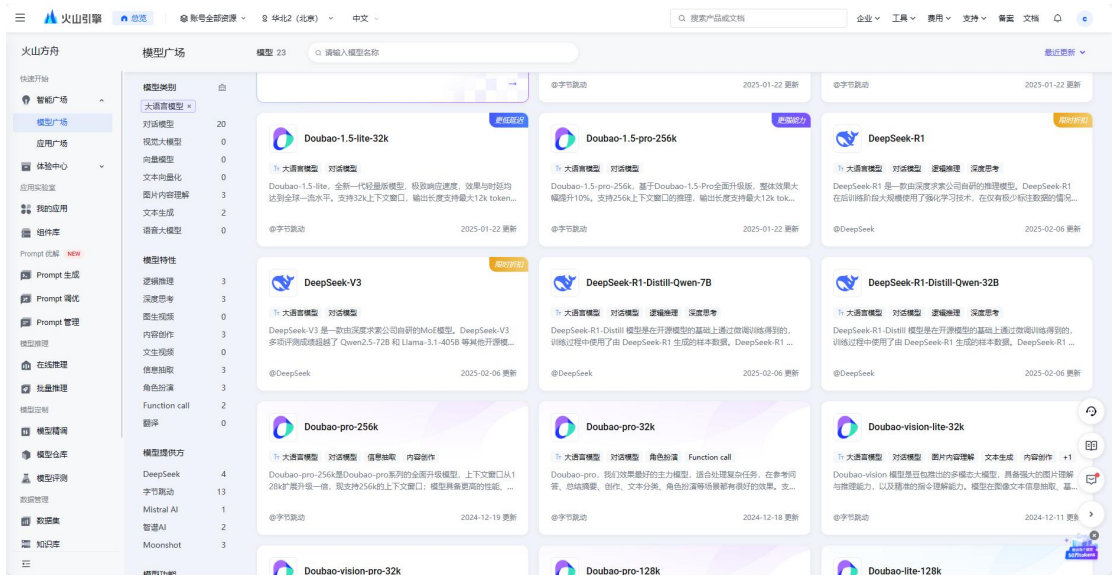


图 8 火山引擎模型界面

4.2.6 其他平台

- (1) 无问芯穹（目前提供蒸馏版 R1）：<https://cloud.infini-ai.com/genstudio>
- (2) 国家超算互联网平台（目前提供蒸馏版 R1）：<https://chat.scnet.cn/>
- (3) 腾讯云 TI-ONE：<https://console.cloud.tencent.com/tione/v2/aimarket>
- (4) 秘塔 AI 搜索：<https://metaso.cn/>

4.3 本地部署

这是最保险的一种方法，即在本地电脑上安装 DeepSeek。只要电脑正常运行，用户就可以随时使用 DeepSeek。

但这种方法有一个缺点，那就是一般来说个人电脑的性能有限，只能装“蒸馏版”的 DeepSeek，这个版本的 DeepSeek 需要占用的电脑资源要比满血版少的多，同样性能也差不少。

这种方式适合的人群为：1 需要保证数据安全，不能联网；2 对于性能的忍受度较高；3 只需要普通的 AI 功能。

如果你符合上述描述，那你可以选择本地部署，方法其实非常简单。

4.3.1 下载 ollama

Ollama 是一个集成主流 AI 大模型的免费平台，用户可以通过访问 <https://ollama.com/> 下载并安装。

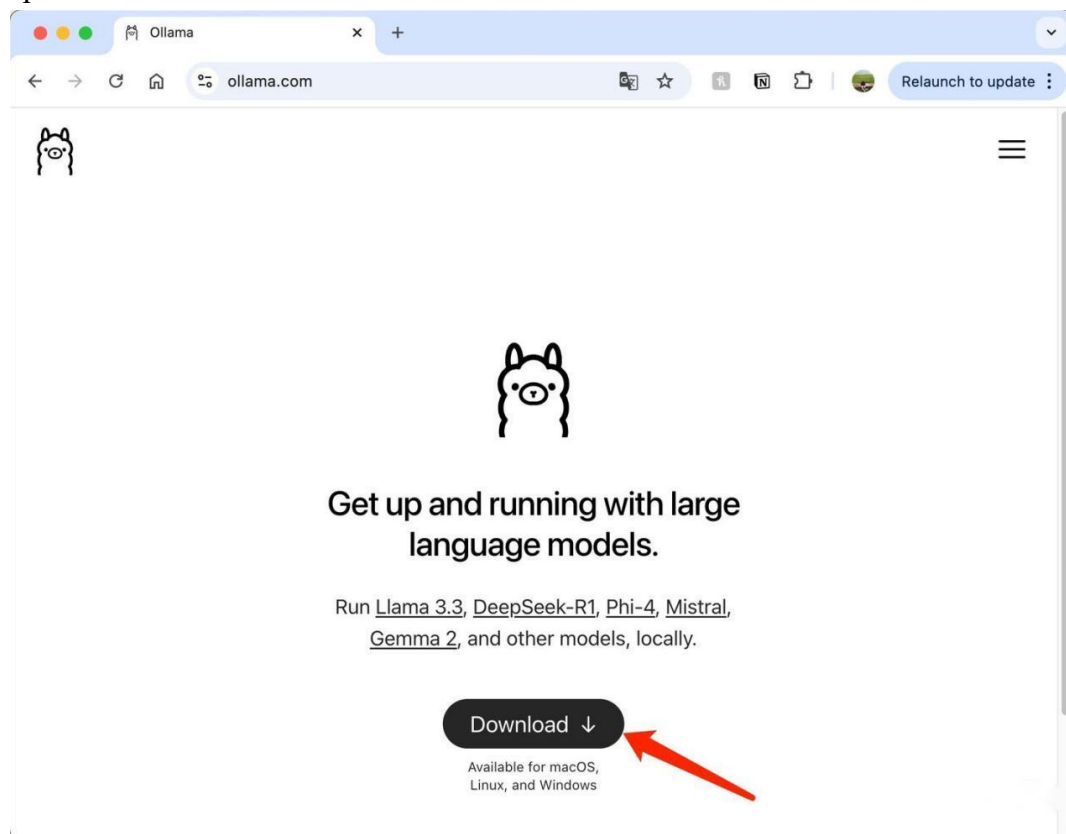


图 9 ollama 官网

4.3.2 合适版本安装

安装 Ollama 后，用户可以选择下载 DeepSeek 的不同版本，其中 R1 和 V3 是较为推荐的版本。

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

↓ 5.8M Pulls ↗ 28 Tags ⌚ Updated 10 days ago

deepseek-coder-v2

An open-source Mixture-of-Experts code language model that achieves performance comparable to GPT4-Turbo in code-specific tasks.

16b 236b

↓ 520.5K Pulls ↗ 64 Tags ⌚ Updated 4 months ago

deepseek-coder

DeepSeek Coder is a capable coding model trained on two trillion code and natural language tokens.

1.3b 6.7b 33b

↓ 455.3K Pulls ↗ 102 Tags ⌚ Updated 13 months ago

deepseek-v3

A strong Mixture-of-Experts (MoE) language model with 671B total parameters with 37B activated for each token.

671b

↓ 146.9K Pulls ↗ 5 Tags ⌚ Updated 2 weeks ago

图 10 ollama 中的 DeepSeek 模型

这两个的不同点为 R1 提供了从 1.5B 到 671B 不同大小的模型，而 V3 只有 671B，而 671B 需要的电脑性能单个电脑几乎不可能满足，所以建议大家可以直接安装并且部署 R1 模型。

DeepSeek R1 的链接：<https://ollama.com/library/DeepSeek-r1:7b>

可以看到 R1 有 7 个版本，其中最小的是 1.5b，它需要的内存大小为 1.1GB，

这个要求几乎所有的电脑都可以满足，可以作为尝试。

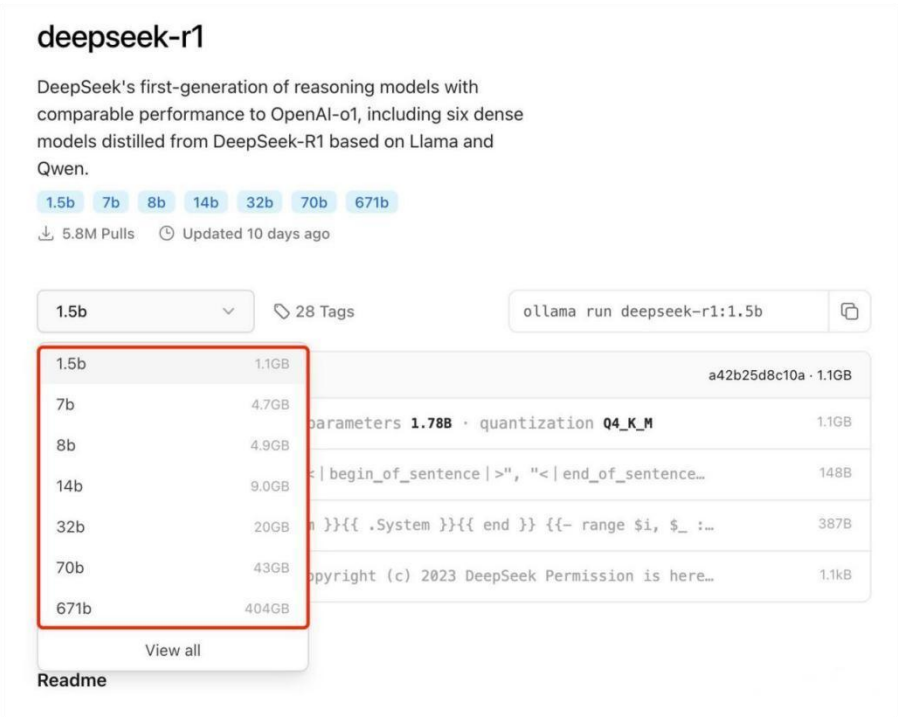


图 11 ollama 中的 DeepSeek-r1 模型

4.3.3 输入安装代码

这一部分需要一些基础的编程知识。Windows 用户可以通过搜索打开命令行，而 Mac 用户则需要打开 Terminal，界面如下。

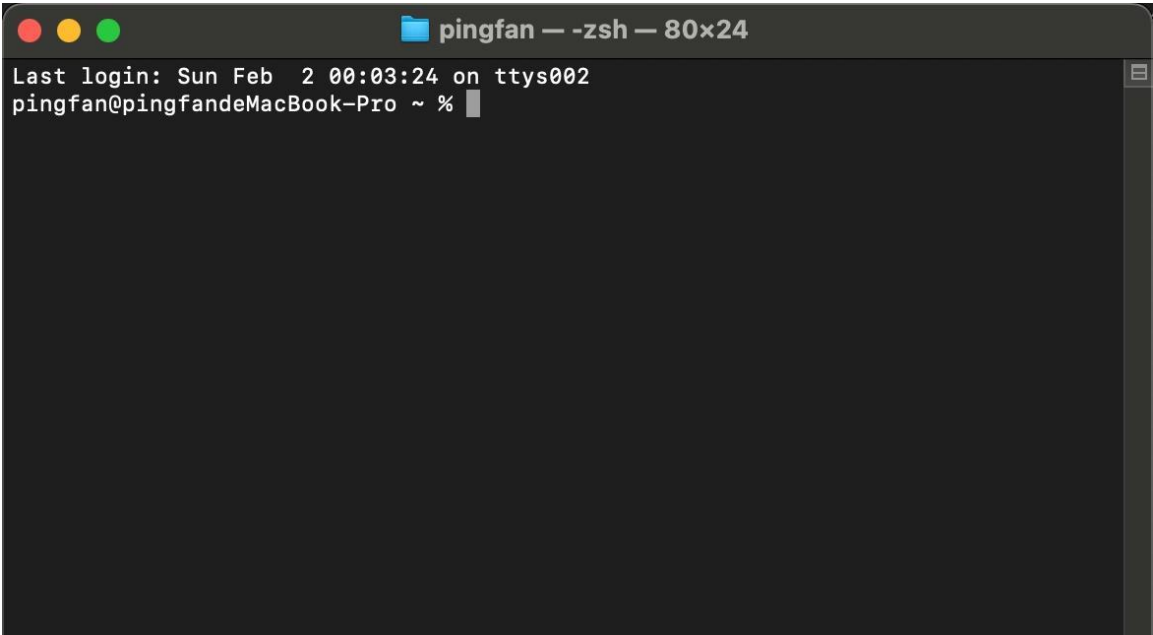


图 12 命令行界面

然后再返回 Ollama 页面，选择 1.5b 这个模型，相对应的代码会自动更新，可以点击复制按钮一键完成复制操作。

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

↓ 5.8M Pulls · ⌚ Updated 10 days ago

7b	28 Tags	ollama run deepseek-r1:7b	
Updated 12 days ago		0a8c26691023 · 4.7GB	
model	arch qwen2 · parameters 7.62B · quantization Q4_K_M	4.7GB	
params	{ "stop": ["< begin_of_sentence >", "< end_of_sentence_"	148B	
template	{{- if .System }}{{ .System }}{{ end }} {{- range \$i, \$_ :-	387B	
license	MIT License Copyright (c) 2023 DeepSeek Permission is here...	1.1kB	

图 13 复制模型下载及运行指令

复制并粘贴代码成如下方式，然后回车键即可。

```
pingfan — -zsh — 80x24
Last login: Sun Feb  2 00:03:24 on ttys002
pingfan@pingfandeMacBook-Pro ~ % ollama run deepseek-r1:1.5b
```

图 14 执行模型下载及运行指令

安装速度取决于网速和所在地区。1.5B 的模型通常可以在 5 分钟内完成安装，安装界面如下，箭头所指的位置是输入问题的地方。

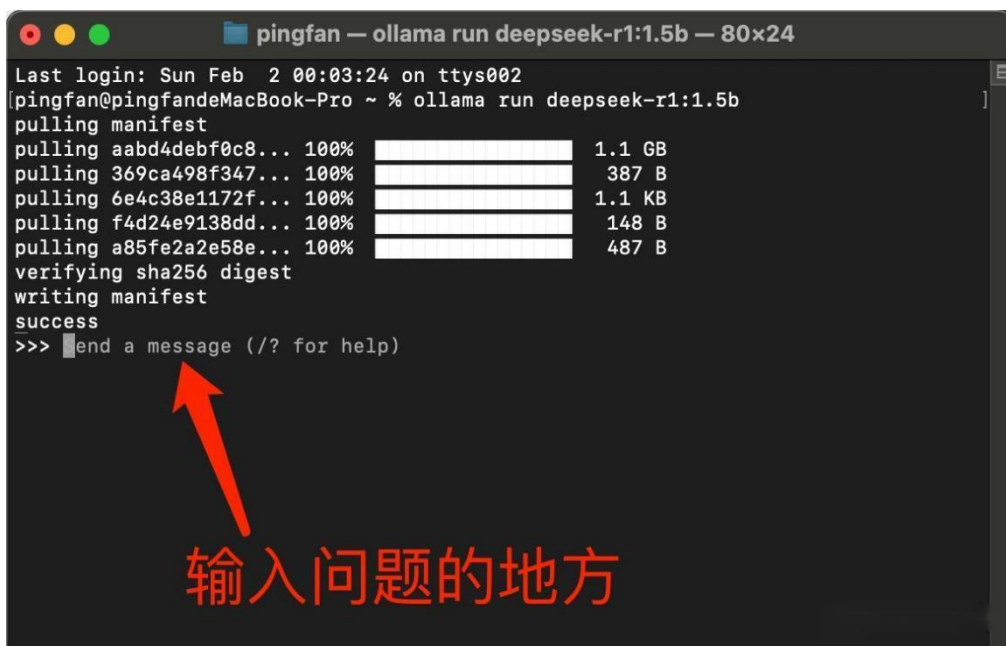


图 15 安装成功界面

4.3.4 测试部署模型

使用方法非常简单，用户只需输入问题并等待输出结果即可。



图 16 运行结果界面

1.5b 的效果受限于模型大小，性能不会很好，可以根据自己电脑内存大小尝试 14b 或者 32b 的模型。

4.3.5 部署非量化模型

另外，如果想部署未量化版本的 DeepSeek 或者原始版本的 DeepSeek，可以进入网站 “<https://hf-mirror.com/>”，选择对应版本的模型，按照其指南依次进行部署。下面以 32B 未量化版本为例，说明该过程。

进入模型所对应的页面：

<https://hf-mirror.com/DeepSeek-ai/DeepSeek-R1-Distill-Qwen-32B>，点击 “Use this model”，如下图所示。

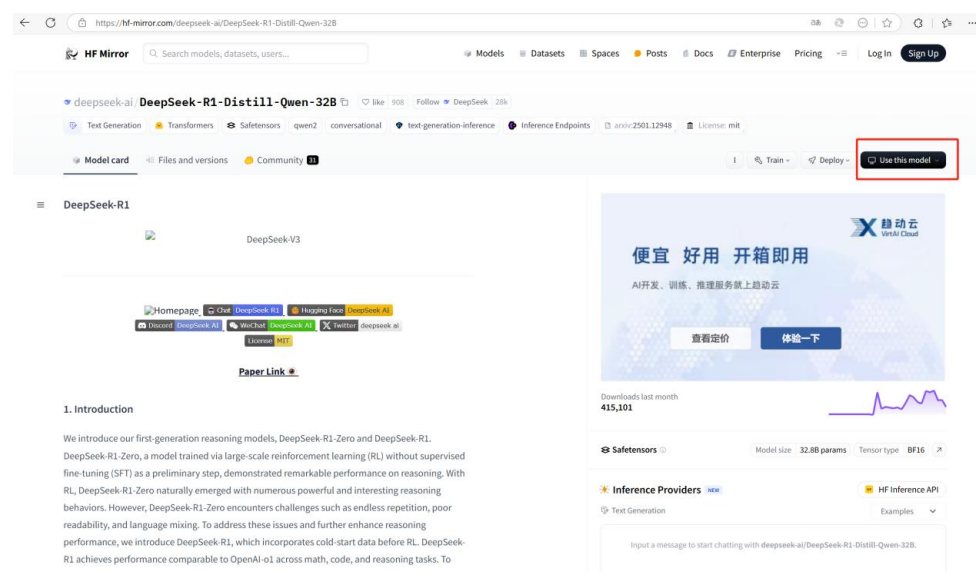


图 17 hugging face 国内镜像 DeepSeek-R1-32B 模型页面

在上步点击后，将弹出不同的使用方式。如下图。

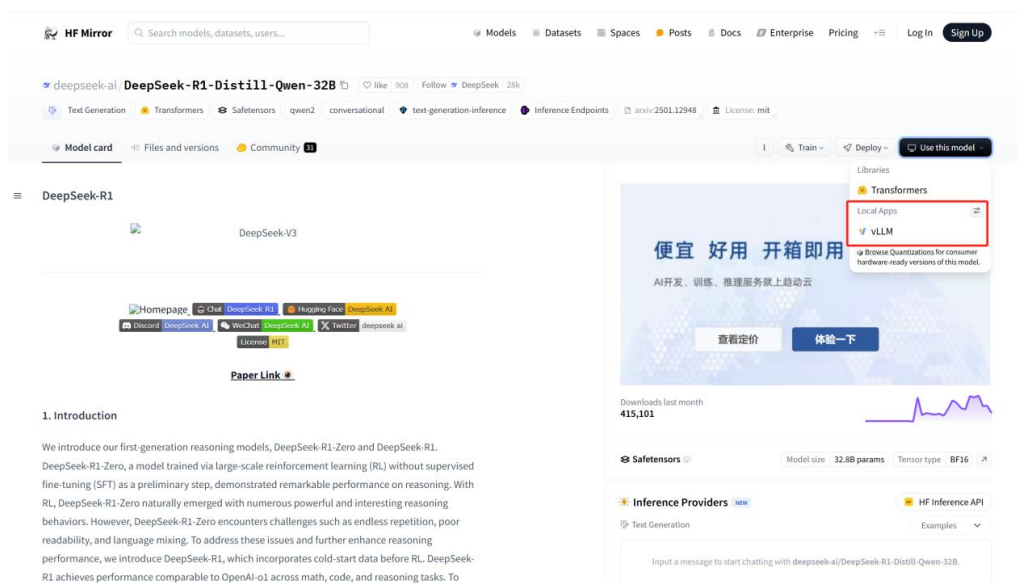


图 18 使用方式界面

选择“vLLM”方式，点击后将显示该方式的部署和测试步骤。此处提供两种部署方式，一种是基于 pip 的安装方式，如下图所示。

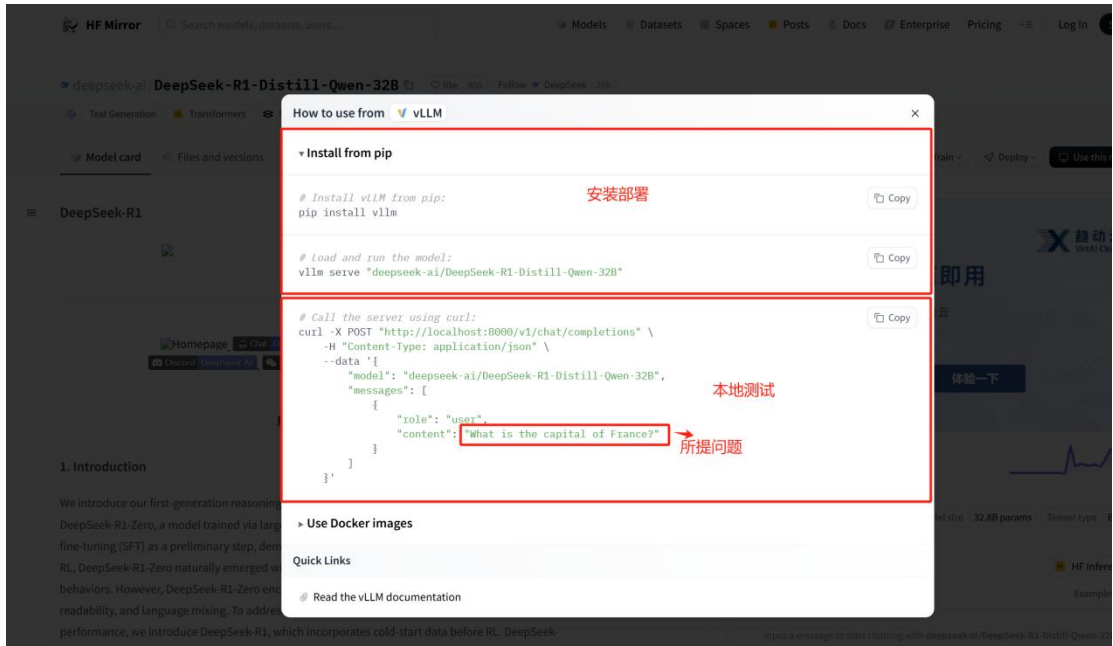


图 19 基于 pip 的安装部署及本地测试说明

另一种是基于 docker 的安装方式，如下图所示。

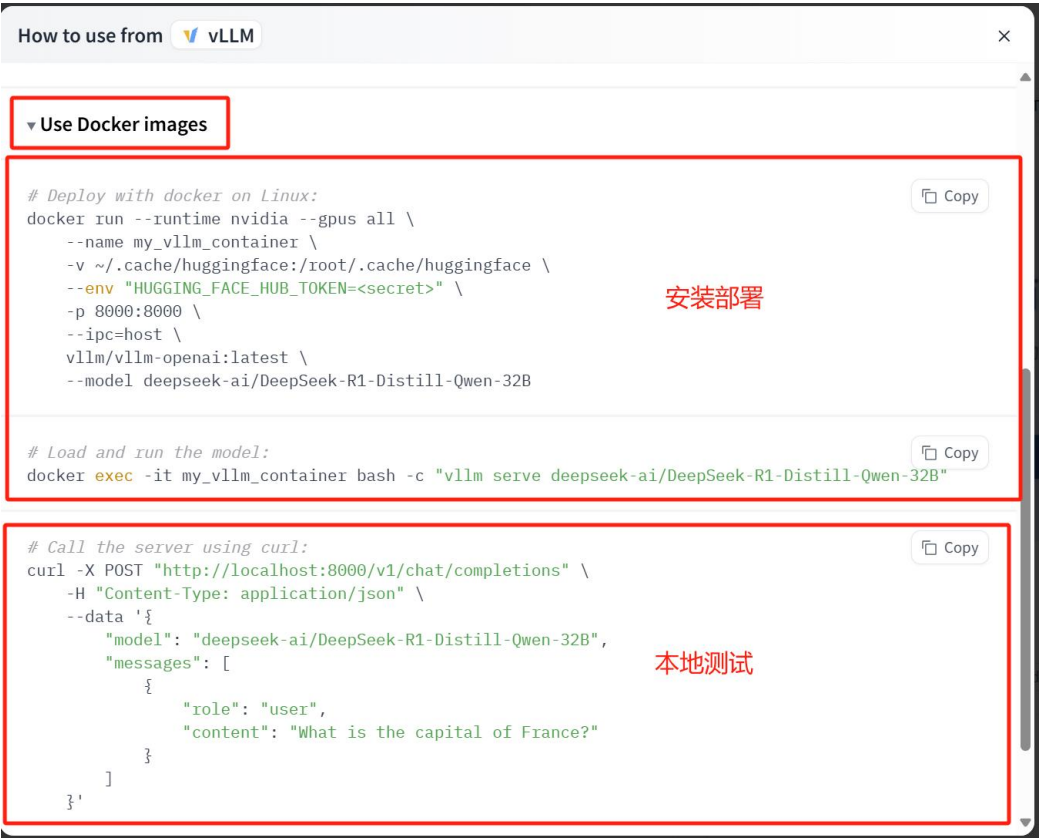


图 20 基于 docker 的安装部署及本地测试说明

5. DeepSeek 审计助手

5.1 基础操作场景

访问官网：在浏览器输入 www.DeepSeek.com，用户可以直接选择“开始对话”或下载手机 App。使用前需注册账号，可选择邮箱或手机注册（建议使用常用邮箱），验证身份时需查看收件箱中的验证邮件；也可选择手机注册，绑定手机号并输入验证码。

由于用户访问量过大，服务器负载过高，与 DeepSeek 对话时经常出现“服务器繁忙，请稍后再试”的提示。除了官网之外，还有硅基流动、英伟达、秘塔 A 搜索等平台部署了 DeepSeek 供用户使用。

5.1.1 快速理解概念

操作方式：用户可以直接向 DeepSeek 提问（中英文均可），提问时需注意明确需求、提供背景、指定格式、控制长度和及时纠正等，这些通常被称为“提示词”。与大模型的交互需要专业的提示词和逻辑性交流，将其视为工作伙伴和助手，以便更好地理解人类语言。

（1）明确需求

- ✕ 错误示例：「帮我写份审计底稿」
- ✔ 正确示例：「我需要一份关于财政审计中“挪用公款”的审计底稿，并附上相关法律法规的具体条款，时间范围为 2024-2025 年。」

（2）提供背景

- ✕ 错误示例：「分析这个数据」
- ✔ 正确示例：「这是一家企业过去三个月的销售数据，请分析与过去三年同期的销量差异（附审计数据）」

（3）指定格式

- ✕ 错误示例：「给出审计整改方案」
- ✔ 正确示例：「请用表格形式逐条列出上述审计整改方案，包含审计问题对象和整改预期成果」

(4) 控制长度

• ✕ 错误示例：「详细说明」

• ✔ 正确示例：「请用 100 字以内说明区块链审计技术的应用，确保完全不懂技术的人也能理解」

(5) 及时纠正

• ✕ 错误示例：「审计证据」

当回答不满意时，可以：

• ✔ 正确示例：「审计证据的充分性和适当性有什么区别？我不理解，请再举个例子。追问」

5.1.2 审计过程模拟

操作方式：通过场景示例进行演练（参考）

复制

你是一家服装公司的审计新人，现在需要：

1. 识别收入确认风险 → DeepSeek会提示关键风险点
2. 选择审计程序 → 提供选择题并解释每个选项的适用场景
3. 分析发现异常 → 自动生成可视化分析图表

5.1.3 审计案例库

操作方式：输入“#审计案例#+关键词”

示例输入：

"审计案例：存货舞弊" → 获取瑞幸咖啡等真实案例解析

"审计案例函证失败" → 展示典型错误及应对方案

案例学习后输入"生成思维导图"可自动梳理要点

5.2 审计工作辅助

5.2.1 审计工作清单

操作方式：通过描述任务自动生成定制化清单

示例输入：

"我要审计一家电商公司的销售费用情况，请生成审计新手注意事项清单"

示例参考：

markdown复制

```
1. **重点核查项**:  
- 刷单返现的会计处理是否合规  
- 网红推广费与销售量的匹配性  
- 运费计提的完整性  
  
2. **新手易错提示**:  
 注意区分平台佣金与广告费性质  
 警惕大额"其他费用"明细  
  
3. **推荐审计程序**:  
【程序1】 同比分析各平台ROI（自动生成分析模板）  
【程序2】 对头部KOL执行穿行测试（附操作指引）
```

5.2.2 审计文档解析

操作方式：上传审计文档

示例输入：

- 合同/凭证 → 自动提取关键条款（如退货权期限）
- 财务报告 → 异常波动自动标注（如应收账款周转率骤降）
- 会议纪要 → 识别潜在风险信号（如"放宽信用政策"表述）

5.2.3 审计访谈

操作方式：输入#模拟访谈#+角色开启演练

示例参考：

#模拟访谈# 应付账款会计复制

```
#模拟访谈# 应付账款会计  
DeepSeek回复（扮演被审计方）：  
"这些应付款项确实超过了账期，主要是因为..."  
  
你追问："请问逾期付款的违约金如何处理？"  
DeepSeek自动评估：  
✅ 优秀提问！延伸建议：  
1. 核对合同违约金条款  
2. 检查营业外支出明细  
3. 关注是否构成资金占用
```

5.3 审计学习考试

5.3.1 智能错题本

操作方式：上传错题截图/拍照→自动解析

特色功能：

- 自动归类知识点（如"审计抽样风险"）
- 推送相关真题（按难度分级）
- 生成易混淆点对比表（如"信赖不足风险 vs 过度信赖风险"）

5.3.2 CPA 审计训练

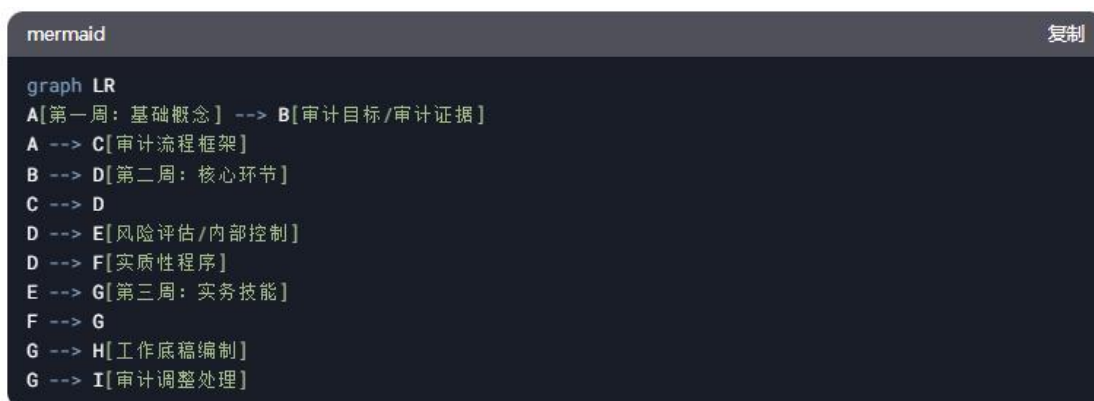
操作方式：输入#CPA 特训#+章节编号

示例参考：

"CPA 特训第 5 章" → 获取近 5 年考点热力图

"生成记忆口诀" → 获得如："认定三兄弟，存在完整性，权利与义务"

5.3.3 学习路径辅导



学习建议（参考）：

- 早上：用 5 分钟问 1 个基础概念（如"什么是分析程序"）
- 午间：解析 1 个真实审计文档（上传公司年报试分析）
- 晚上：完成 1 个情景模拟（如存货监盘突发情况处理）

5.4 其他提示

隐私保护：

- 上传文件自动加密（24 小时后删除）
- 支持匿名学习模式

效率技巧：

- 对复杂问题说"请用表格对比说明"

- 在答案后追加"生成知识卡片"获取复习素材

新手常见问题库：

输入#常见误区#查看：

- 审计=查账？
- 细节测试越多越好？
- 如何判断审计证据是否足够？

附录：常见问题解答（FAQ）

一、模型选择与部署

Q1：如何选择适合审计任务的 DeepSeek 版本？

- **DeepSeek-R1:**
 - 适用场景：复杂逻辑推理（如风险建模、异常交易检测）。
 - 推荐版本：
 - **R1-1.5B:** 本地部署（显存 $\geq 1.1\text{GB}$ ），适合基础问答和文档解析。
 - **R1-32B:** 团队使用（显存 $\geq 32\text{GB}$ ），支持多维度数据分析。
- **DeepSeek-V3:**
 - 适用场景：长文本处理（合同条款解析、法规匹配）。
 - 推荐版本：仅 671B 完整版（需云端部署）。
- **Janus Pro:**
 - 适用场景：多模态任务（票据识别、图表关联分析）。

Q2：本地部署需要哪些硬件配置？

- 显存需求对照表：

模型版本	显存需求（4bit 量化）	显存需求（非量化）
R1-1.5B	1.1GB	3GB
R1-7B	8GB	20GB
R1-32B	24GB	64GB
- 最低配置建议：
 - CPU：Intel i7 或 AMD Ryzen 7 及以上。
 - 内存：16GB（1.5B 模型） / 32GB（7B 及以上模型）。
 - 操作系统：Windows 10 / macOS Monterey / Ubuntu 20.04。

Q3：第三方平台如何收费？（参考价格如下，收费价格以各平台官网发布为准）

- 华为云：
 - 按需计费：R1-671B 模型 输入为¥4 元/M tokens，输出为¥16 元/M tokens。
- 火山引擎：
 - R1-671B 模型 输入为¥2 元/M tokens，输出为¥8 元/M tokens。
 - 企业套餐：按年订阅可享 85 折优惠。
- 百度云帆：

- 一键部署：R1-671B 模型 输入为 ¥2 元/M tokens，输出为 ¥8 元/M tokens。目前可享受限时免费服务。
-

二、性能与优化

Q4：量化模型性能下降明显吗？如何权衡？

- 性能对比：
 - 速度提升：4bit 量化模型推理速度提高 40%-60%。
 - 精度损失：通用任务损失 5%-8%，数学/代码任务损失 10%-15%。
 - 推荐策略：
 - 对精度要求高：优先选择非量化版本（如 R1-32B）。
 - 资源有限：使用量化版本（如 R1-7B-4bit）。
-

Q5：如何提升本地模型的响应速度？

- 优化方案：
 1. 关闭后台程序：释放显存占用（如游戏、视频软件）。
 2. 启用 GPU 加速：安装 CUDA 工具包并配置 `--gpu-layers=20` 参数。
 3. 限制输出长度：在提问时添加“请用 200 字内回答”。
-

三、错误与故障排除

Q6：部署时报错“显存不足”如何解决？

- 步骤排查：
 1. 检查显存：使用 `nvidia-smi`（NVIDIA）或 `radeontop`（AMD）查看占用情况。
 2. 降低模型规格：换用更小版本（如从 32B 切换至 7B）。
 3. 启用量化：在 Ollama 中添加 `--quantize q4_0` 参数。
-

Q7：访问官网频繁提示“服务器繁忙”怎么办？

- 应急方案：
 1. 切换第三方渠道：如火山引擎、秘塔 AI 搜索。
 2. 非高峰时段操作：建议工作日上午 8:00-10:00 访问。
 3. 本地部署备用：安装 R1-1.5B 轻量版临时使用。
-

四、数据安全与隐私

Q8：上传审计文件是否安全？

- 存在一定数据泄露风险，建议：
 - 1. 敏感数据优先选择本地部署版本。
 - 2. 必须使用在线服务，尽量选择知名度高的在线服务，并对数据做脱敏处理。
-

五、其他高频问题

Q9：如何联系 DeepSeek 技术支持？

- 官方渠道：
 - 邮箱：support@deepseek.com（24 小时内响应）。
 - 在线客服：官网右下角“帮助中心”入口。
-

Q10：模型更新后原有功能失效怎么办？

- 应对措施：
 1. 检查版本日志：在官网“更新公告”查看兼容性说明。
 2. 回滚旧版本：用旧版本模型更换。
 3. 提交反馈：通过客服通道描述具体问题。