Got DeepSeek R1 running locally - Full setup guide and my personal review (Free OpenAI o1 alternative that runs locally??)

> Edit: I double-checked the model card on Ollama(https://ollama.com/library/deepseek-r1), and it does mention DeepSeek R1 Distill Qwen 7B in the metadata. So this is actually a distilled model. But honestly, that still impresses me!

Just discovered DeepSeek R1 and I'm pretty hyped about it. For those who don't know, it's a new **open-source AI model that matches OpenAI o1 and Claude 3.5 Sonnet** in math, coding, and reasoning tasks.

You can check out Reddit to see what others are saying about DeepSeek R1 vs OpenAI o1 and Claude 3.5 Sonnet. For me it's really good - good enough to be compared with those top models.

And the best part? **You can run it locally on your machine, with total privacy and 100% FREE!!**

I've got it running locally and have been playing with it for a while. Here's my setup - super easy to follow:

*(Just a note: While I'm using a Mac, **this guide works exactly the same for Windows and Linux users***! 👌)*

1) Install Ollama

Quick intro to Ollama: It's a tool for running AI models locally on your machine. Grab it here: https://ollama.com/download

2) Next, you'll need to pull and run the DeepSeek R1 model locally.

Ollama offers different model sizes - basically, bigger models = smarter AI, but need better GPU. Here's the lineup:

> 1.5B version (smallest):
> ollama run deepseek-r1:1.5b
>
> 8B version:
> ollama run deepseek-r1:8b
>
> 14B version:
> ollama run deepseek-r1:14b
>
> 32B version:
> ollama run deepseek-r1:32b
>
> 70B version (biggest/smartest):
> ollama run deepseek-r1:70b

Maybe start with a smaller model first to test the waters. Just open your terminal and run:

```
ollama run deepseek-r1:8b
```

Once it's pulled, the model will run locally on your machine. Simple as that!

Note: The bigger versions (like 32B and 70B) need some serious GPU power. Start small and work your way up based on your hardware!

3) Set up Chatbox - a powerful client for AI models

Quick intro to Chatbox: a free, clean, and powerful desktop interface that works with most models. I started it as a side project for 2 years. It’s privacy-focused (all data stays local) and super easy to set up—no Docker or complicated steps. Download here: https://chatboxai.app

In Chatbox, go to settings and switch the model provider to Ollama. Since you're running models locally, you can ignore the built-in cloud AI options - **no license key or payment is needed!**

Then set up the Ollama API host - the default setting is http://127.0.0.1:11434 , which should work right out of the box. That's it! Just pick the model and hit save. Now you're all set and ready to chat with your locally running Deepseek R1! 🚀

Hope this helps! Let me know if you run into any issues.

Here are a few tests I ran on my local DeepSeek R1 setup (loving Chatbox's **artifact preview** feature btw!) 👇

Explain TCP:

Honestly, this looks pretty good, especially considering it's just an 8B model!

Make a Pac-Man game:

It looks great, but I couldn't actually play it. I feel like there might be a few small bugs that could be fixed with some tweaking. (Just to clarify, this wasn't done on the local model — my mac doesn't have enough space for the largest deepseek R1 70b model, so I used the cloud model instead.)

Honestly, I've seen a lot of overhyped posts about models here lately, so I was a bit skeptical going into this. But after testing DeepSeek R1 myself, I think it's actually really solid. It's not some magic replacement for OpenAI or Claude, but it's **surprisingly capable** for something that runs locally. The fact that it's free and works offline is a huge plus.

What do you guys think? Curious to hear your honest thoughts.