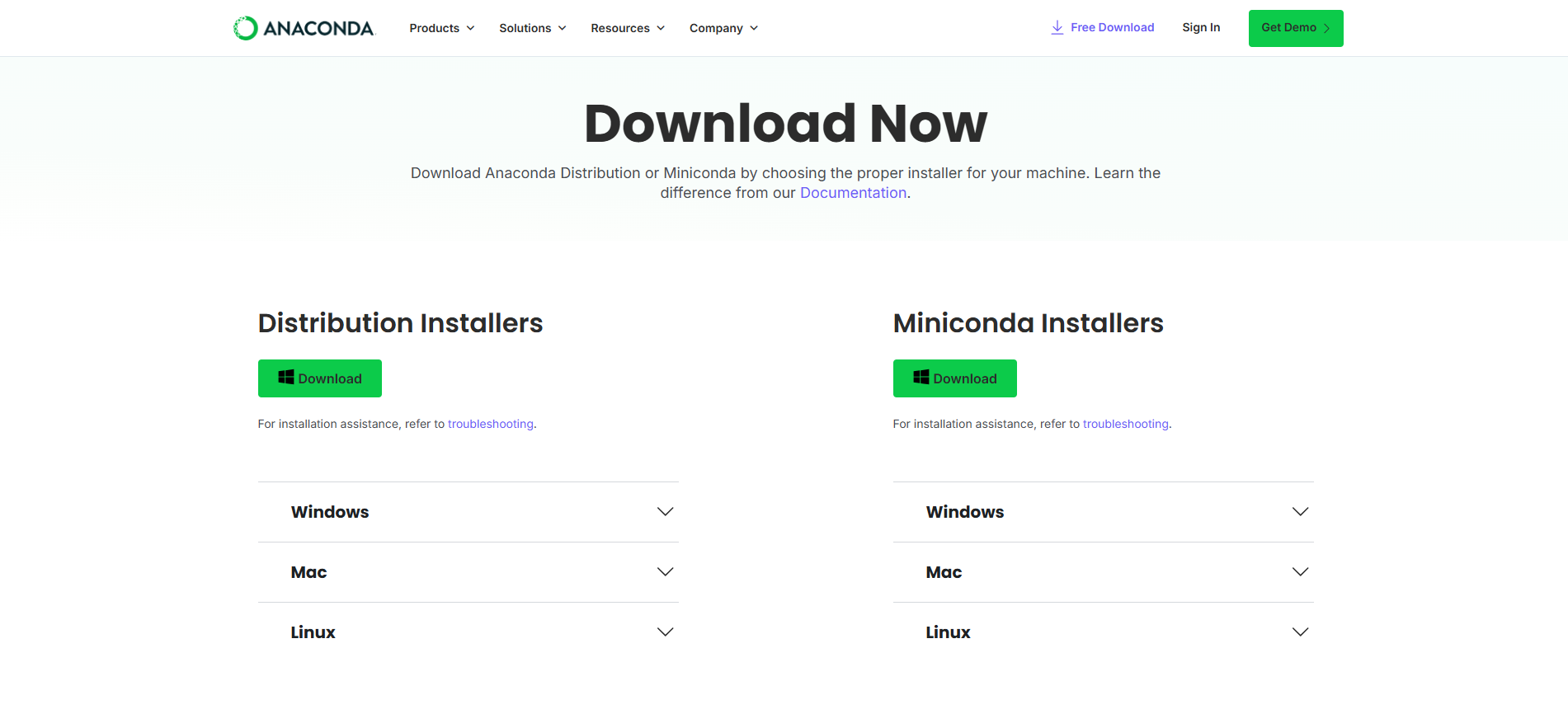
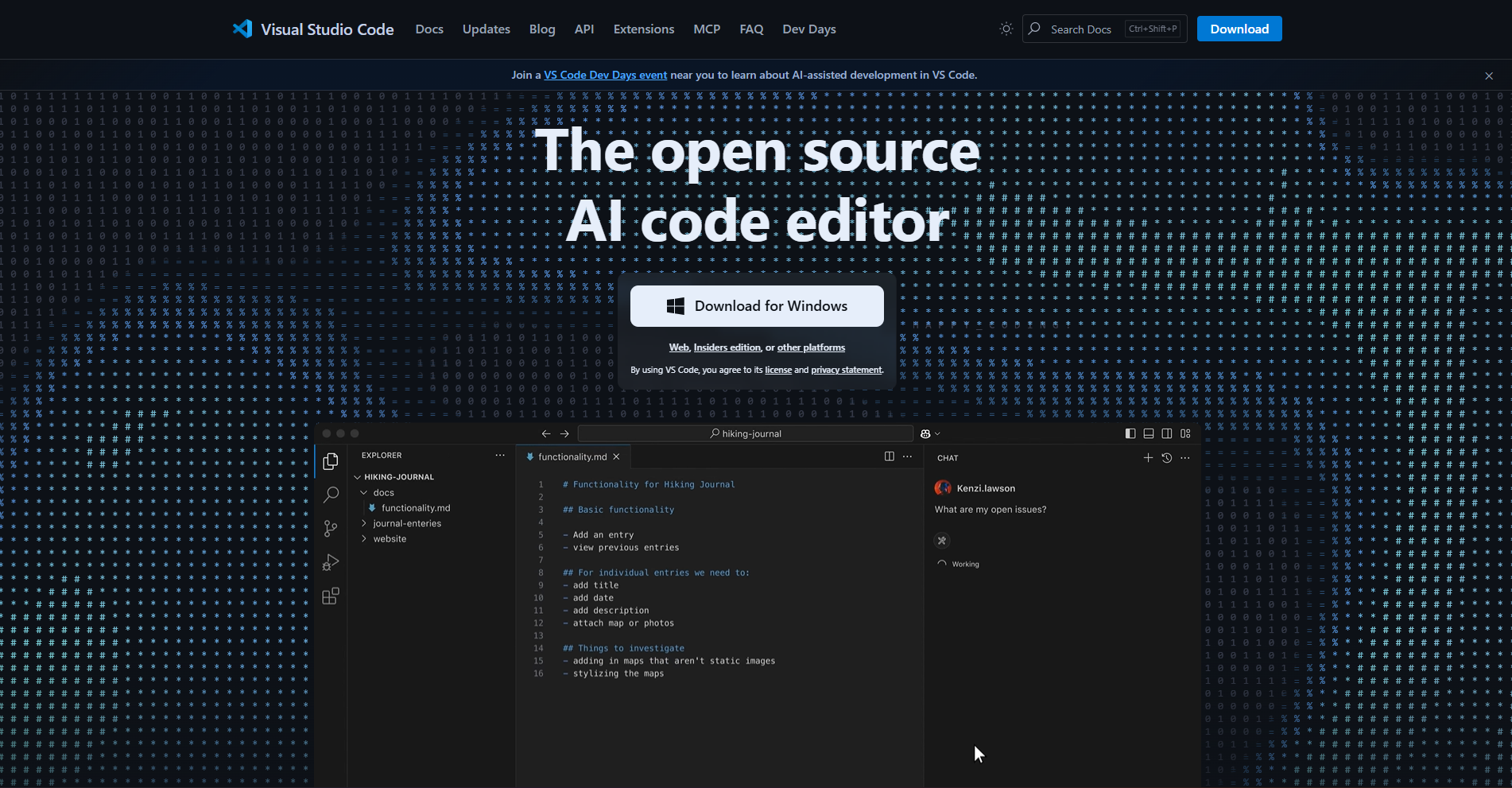
Hướng dẫn cài đặt

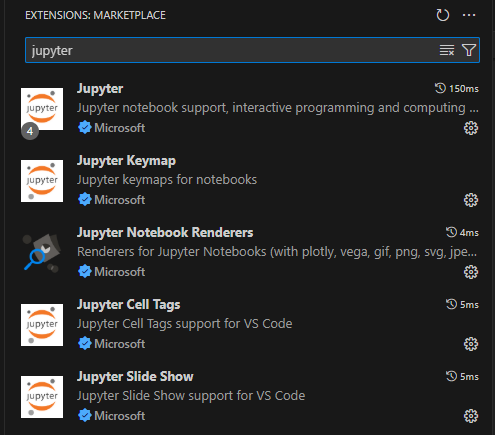
1. Tải Anaconda - [https://www.anaconda.com/download/](https://www.anaconda.com/download/success)

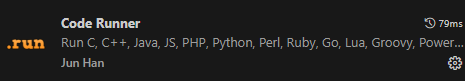
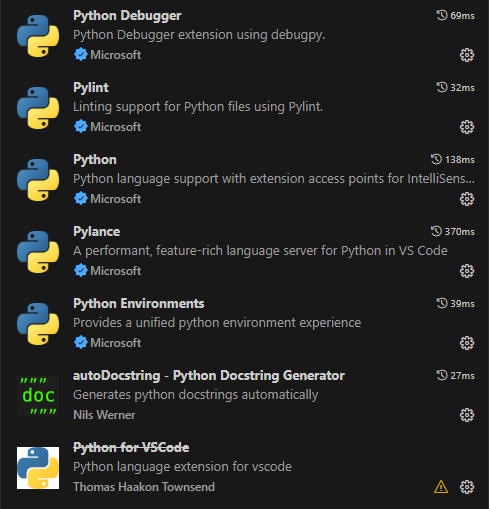
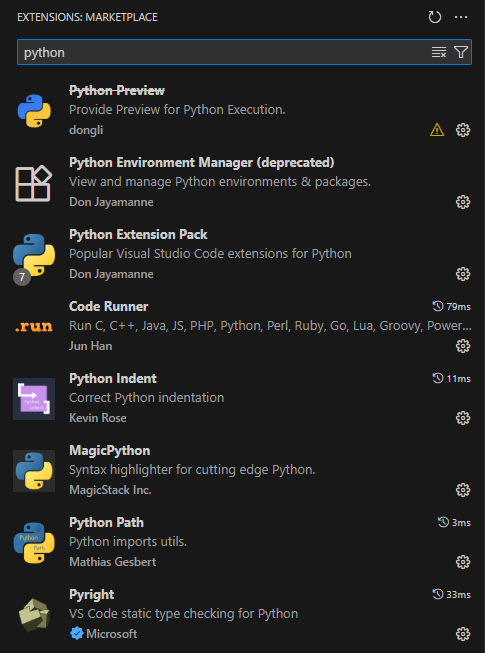


1. Tải Visual Studio Code - <https://code.visualstudio.com/>



Trong VS code cần cài các extension sau để thuận tiện và dễ sử dụng.





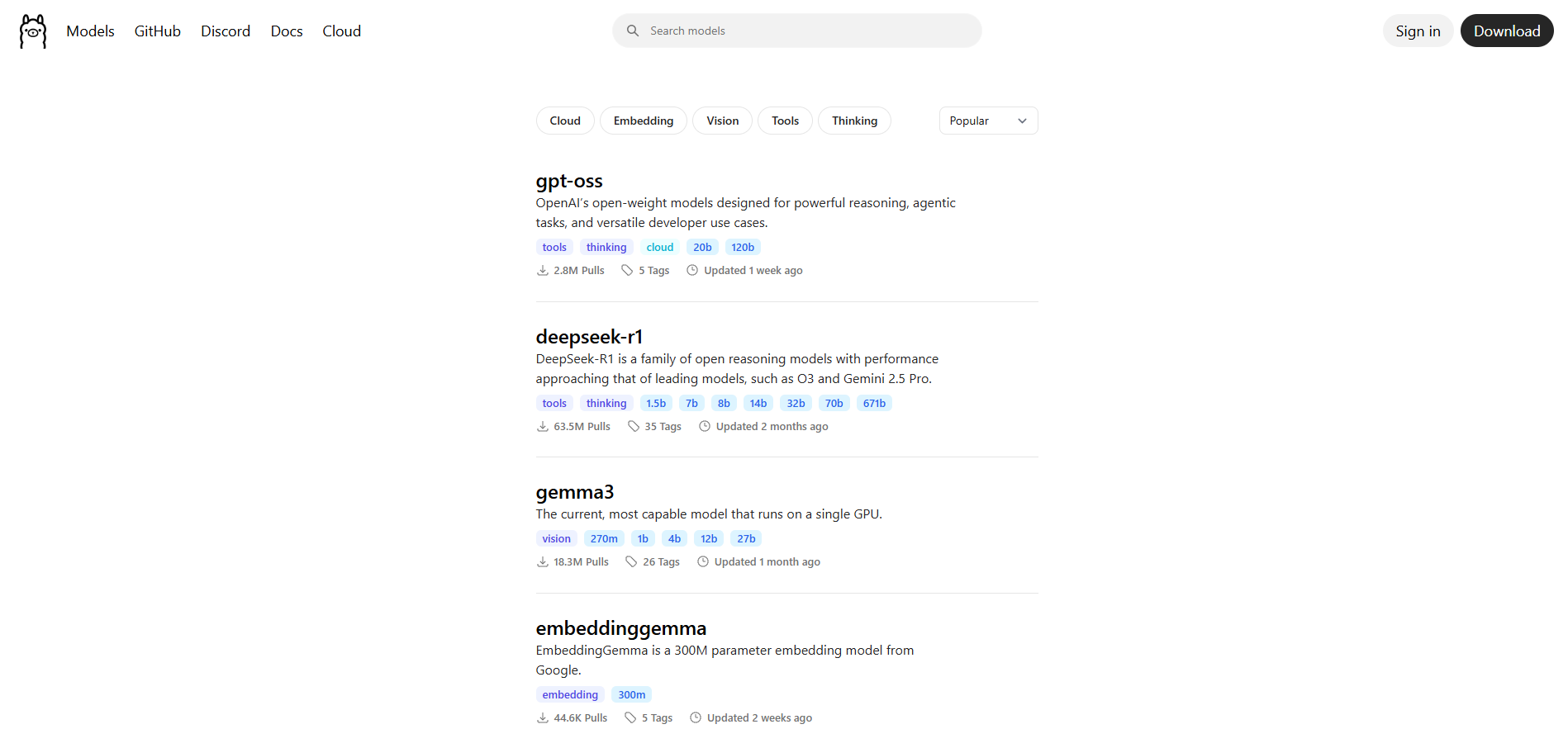
1. Bật anaconda prompt

* Nhập: conda create -n mykernel python=3.9.21
* Tiếp theo nhâp: activate mykernel
* Nhập như hình để kiểm tra version sao cho đúng 3.9.21
* Sau đó cài đặt một loạt các packages.

pip install python-dotenv langchain langchain-ollama langchain-community langchain-core langgraph requests chromadb sentence-transformers

1. Cài đặt Ollama - [https://ollama.com/download/](https://ollama.com/download/windows)

* Tùy thuộc vào laptop hay PC mạnh hay yếu để có thể chọn được model phù hợp.
* Model nhỏ (0 – 3GB) 🡪 Kết quả thường lệch lạc, không có chiều sâu, câu trả lời đơn giản và rời rạc.
* Model vừa (3-15GB) 🡪 Kết quả có thể chấp nhận được, tuy nhiên cũng có nhiều hạn chế. Thời gian xử lý ổn phù hợp.
* Model lớn (15 – 45GB) 🡪 Kết quả tốt, tuy nhiên hạn chế với máy yếu, đặc biệt là laptop (VRAM Card yêu cầu lớn >=20GB. Chip phải nhiều core và thread, RAM >= 64GB).
* Model rất lớn (>45GB) 🡪 Mạnh như Chat GPT, DeepSeek, Gemini…
* Vào đường dẫn sau để tìm model: <https://ollama.com/search>



* Sau đó tùy vào model bạn chọn sẽ có lệnh: ollama run llama3:8b – Chạy trong kernel anaconda 🡪 mykernel. Để cài đặt.