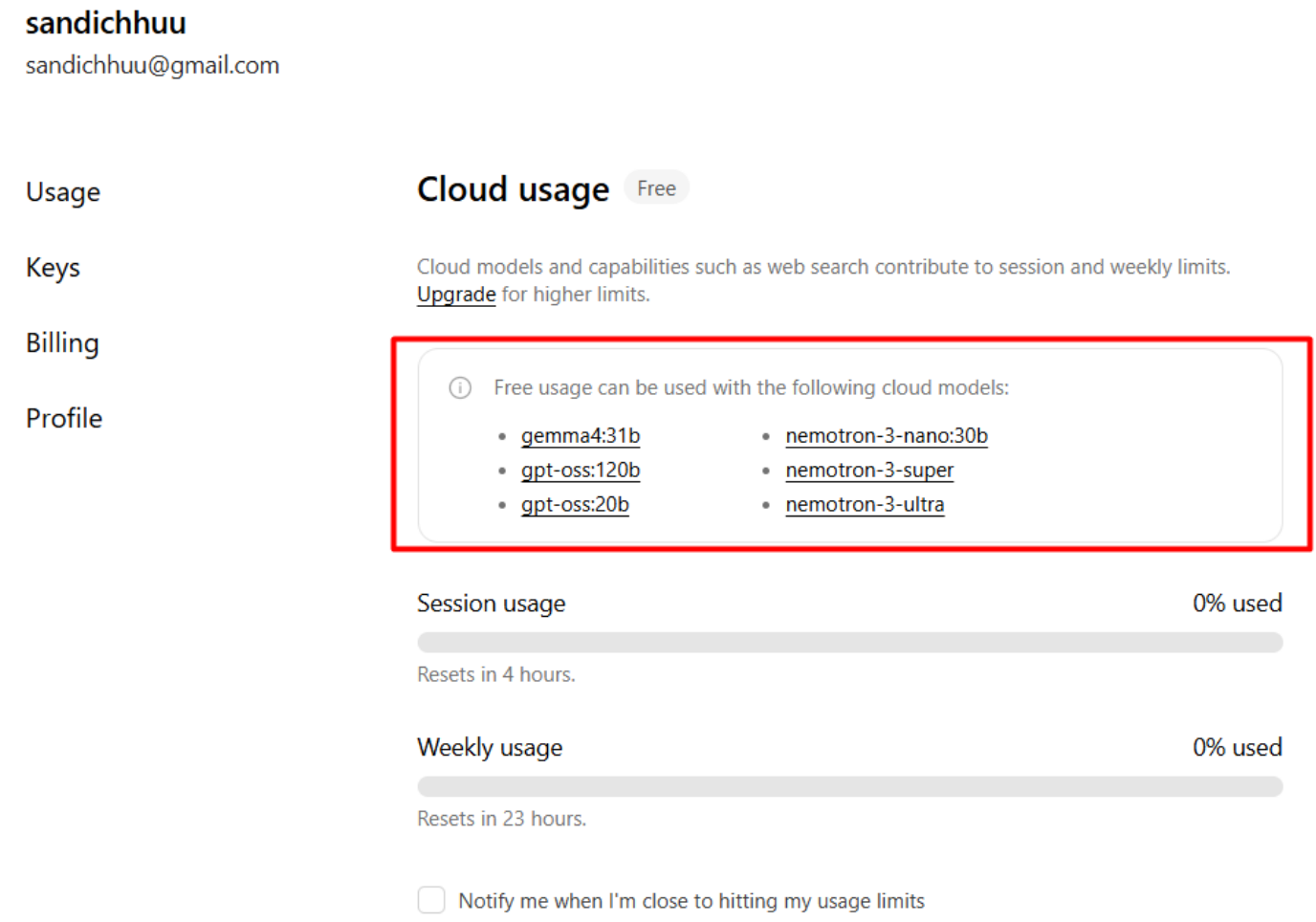
Giảm chi phí với Ollama free token hàng tuần.

Ollama

Đầu tiên hãy tạo tài khoản Ollama ở đây: Ollama

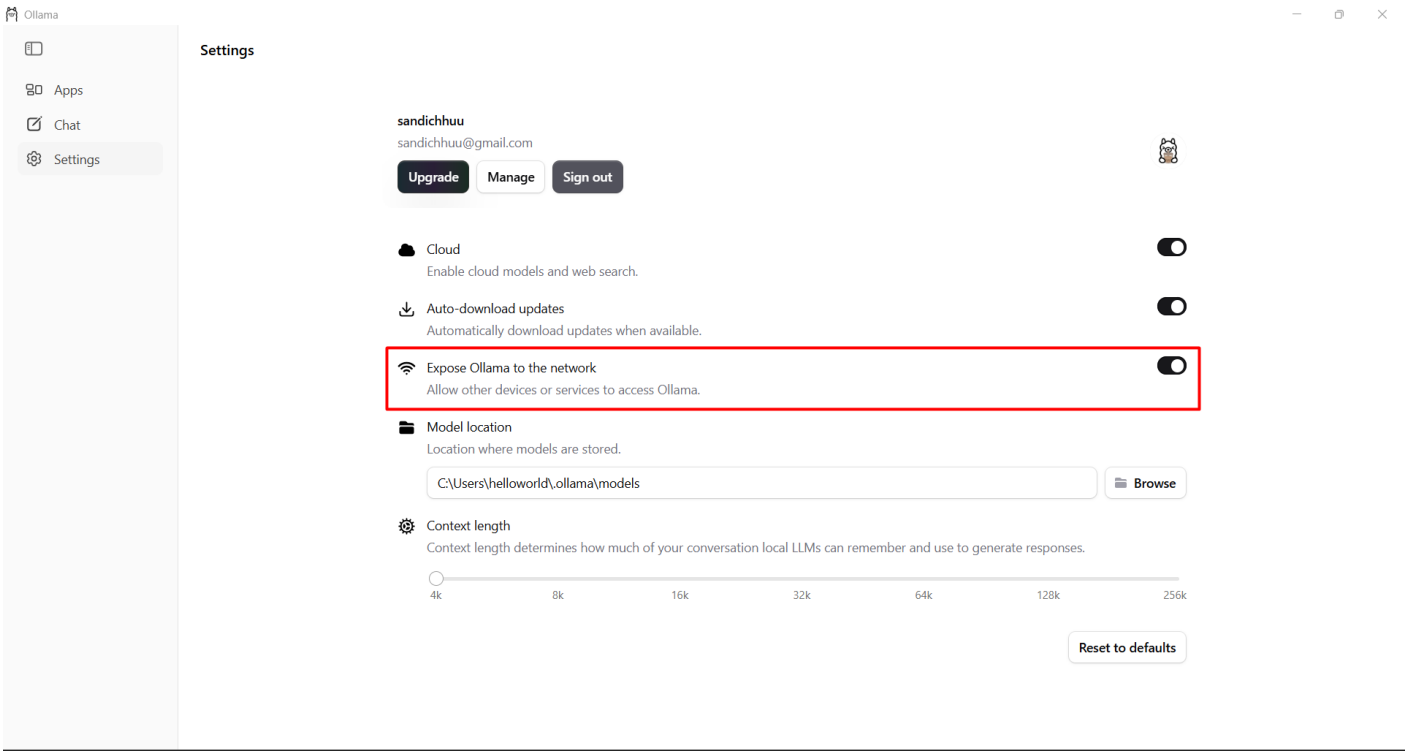
Tiếp theo hãy vào đây xem mô hình nào đang free: Ollama



Ví dụ ở đây họ đang cho free 6 mô hình này.
Trong đó chúng ta sẽ chừa tụi gpt-oss ra, mấy mô hình này hơi dởm khi code.
Khuyến nghị dùng mô hình Gemma4:31b.

Cài đặt & Cấu hình

Cài ollama bằng lệnh này trong PowerShell: `irm https://ollama.com/install.ps1 | iex`.
Sau đó vào Settings mở cái phần khoanh đỏ lên để DSH kết nối được.

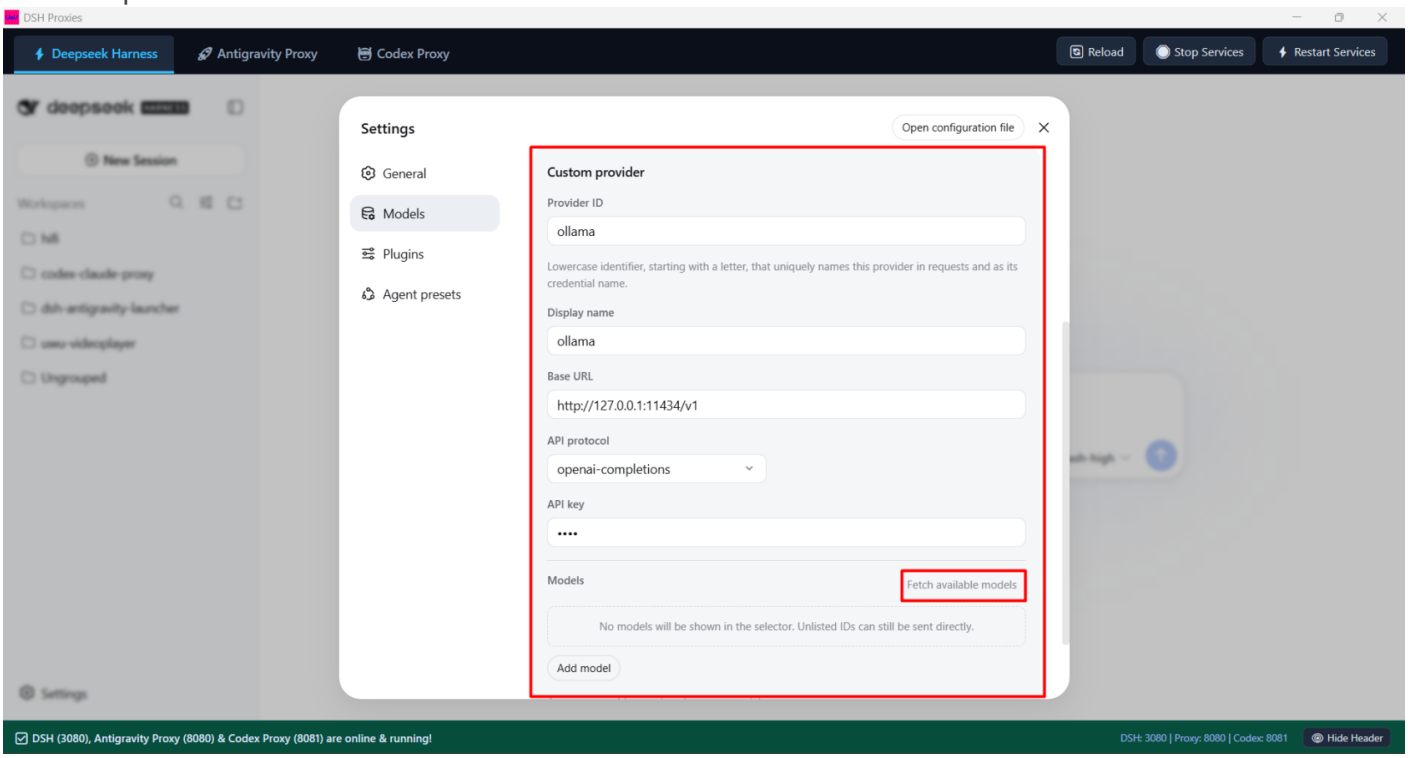


Tiếp theo chạy lần lượt các lệnh sau trong PowerShell để dùng các mô hình free:

```
ollama pull gemma4:31b-cloud
ollama pull gpt-oss:120b-cloud
ollama pull gpt-oss:20b-cloud
ollama pull nemotron-3-nano:30b-cloud
ollama pull nemotron-3-super:cloud
ollama pull nemotron-3-ultra:cloud
```

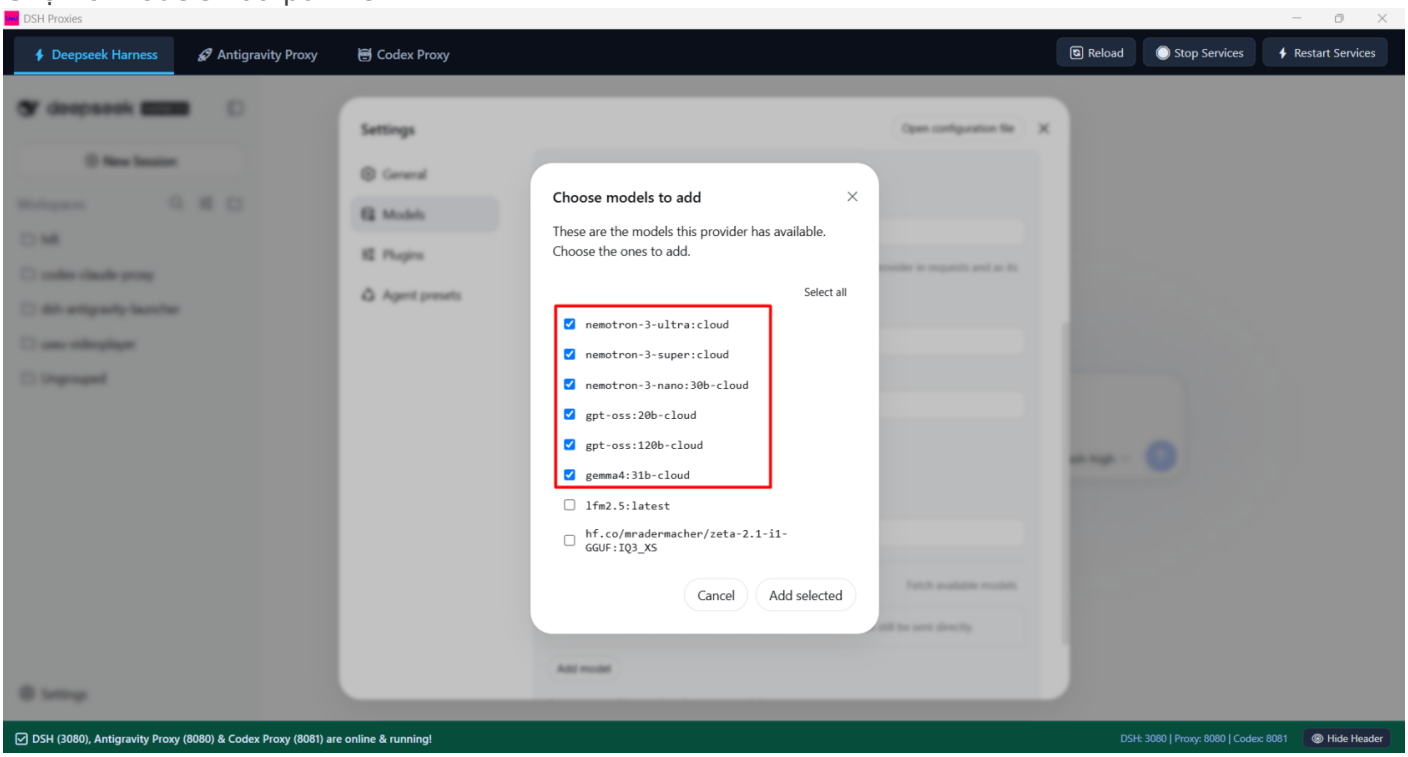
Deepseek Harness

Vào Deepseek Harness để cấu hình thêm mô hình như ảnh:

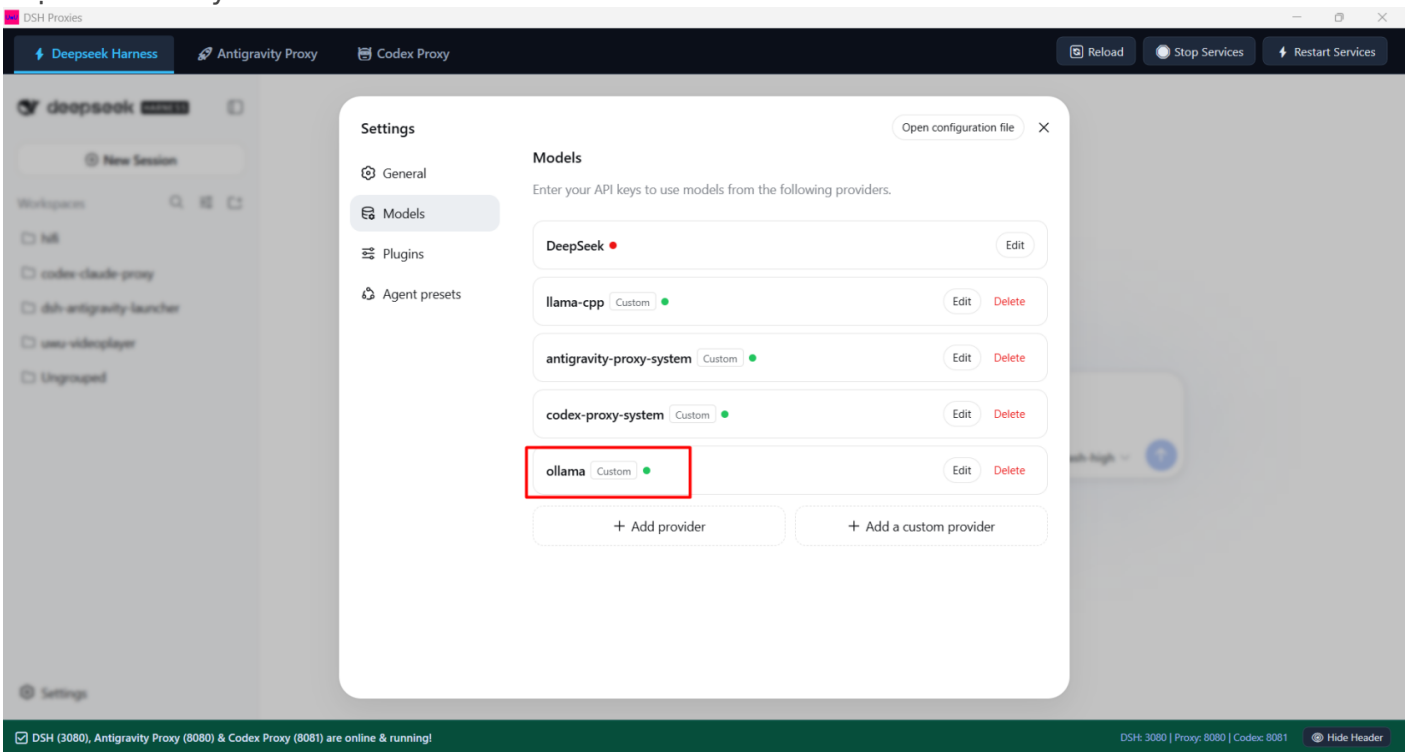


Sau đó chọn nút `Fetch Available Models`:

Chọn 6 models vừa pull về



Hiện ra như này là ok



Revision #1
Created 2026-08-30 01:16:59 UTC by sandichhuu
Updated 2026-08-30 01:46:13 UTC by sandichhuu