DeepSeek

Usage instructions:

1. Launch the product via 1-click. **Please wait until** the instance passes **all** status checks and is running. You can connect using your Amazon private key and '**ubuntu**' login via your SSH client.

   To update software, use: **sudo apt update && sudo apt upgrade -y**

2. Set up your IP address. Replace the server_name with your **instance Public IPv4 address.**

sudo nano /etc/nginx/sites-available/open-webui

```
  GNU nano 7.2
server {
    listen 80;
    server_name 54.236.6.1;

    location / {
        proxy_pass http://localhost:8080;
        proxy_set_header Host $host;
        proxy_set_header X-Real-IP $remote_addr;
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
        proxy_set_header X-Forwarded-Proto $scheme;
    }
}
```

- **Save & Exit**

3. Check syntax and restart nginx.

   sudo nginx -t

   sudo systemctl restart nginx

4. Install NVIDIA GPU and run the Deepseek model. ***Be patient. It may take a few minutes….***

curl -fsSL https://ollama.com/install.sh | sh

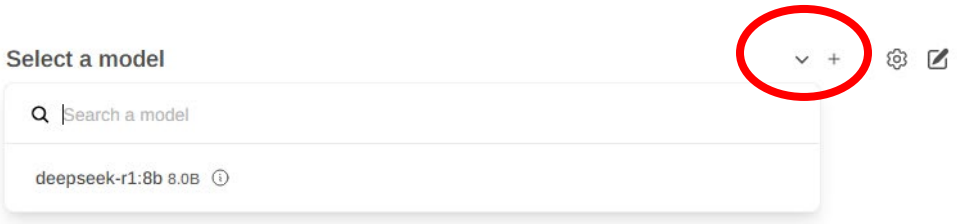
ollama run deepseek-r1:8b

5. When complete, access the Web Interface. Open your web browser and navigate to **http://your_instance_Public_IPv4_address** to interact with DeepSeek via Open WebUI.

Ex: http://23.32.3.2

Note: Open WebUI Community platform is **NOT** required to run Open WebUI. The account you use here does not sync with your self-hosted Open WebUI instance, and vice versa. This will help notify you of updates.

Select the model inside the GUI



AWS Data

- Data Encryption Configuration: This solution does not encrypt data within the running instance.
- User Credentials are stored: /root/.ssh/authorized_keys & /home/ubuntu/.ssh/authorized_keys
- Monitor the health:
  - Navigate to your Amazon EC2 console and verify that you're in the correct region.
  - Choose Instance and select your launched instance.
  - Select the server to display your metadata page and choose the Status checks tab at the bottom of the page to review if your status checks passed or failed.

Extra Information: **(Optional)**

Allocate Elastic IP

To ensure that your instance **keeps its IP during restarts** that might happen, configure an Elastic IP. From the EC2 console:

1. Select ELASTIC IPs.
2. Click on the ALLOCATE ELASTIC IP ADDRESS.
3. Select the default (Amazon pool of IPv4 addresses) and click on ALLOCATE.
4. From the ACTIONS pull down, select ASSOCIATE ELASTIC IP ADDRESS.
5. In the box that comes up, note down the Elastic IP Address, which will be needed when you configure your DNS.
6. In the search box under INSTANCE, click and find your INSTANCE ID and then click ASSOCIATE.
7. Your instance now has an elastic IP associated with it.
8. For additional help: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/elastic-ip-addresses-eip.html