“???”?????????

???? OpenClaw+ ?????????? llama.cpp ??????????????????

?????? GTR9Pro ??? OpenClaw??????????????????????????????

> “ ?????????OpenClaw+?????????

---

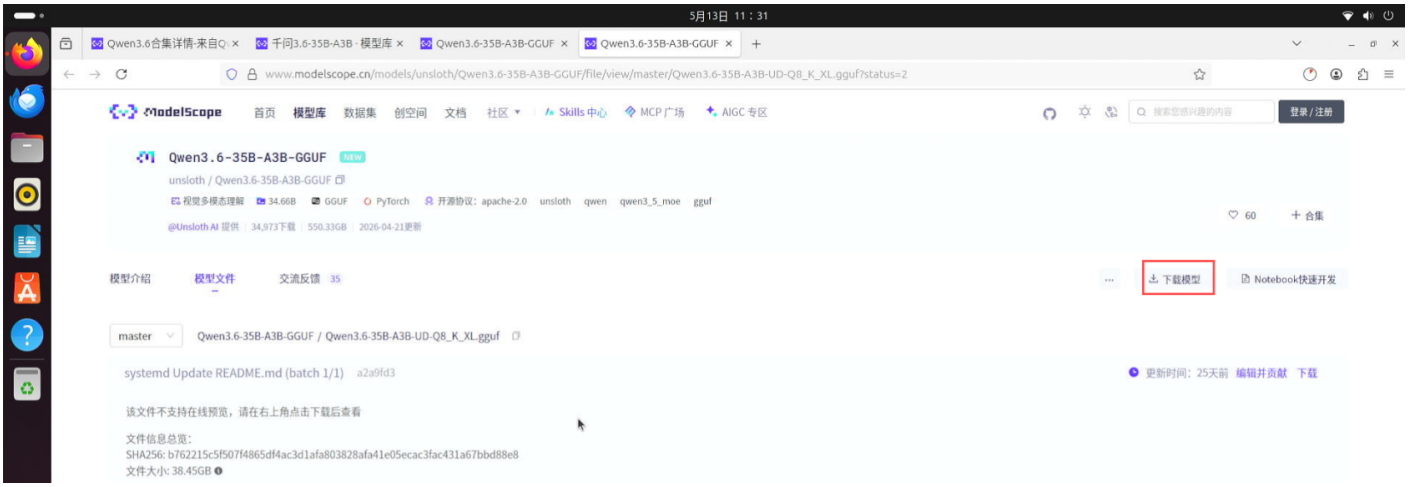
1. ????

llama.cpp ?? GGUF ????????????????????

0. Hugging Face??**????**?
1. ModelScope??????

???????????? `Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf` ????????? 8 ????????????????????????????????????

> “ ????????????????



????????????????????????

```
sudo mv /home/□□□/□□/Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf /opt/models/
```

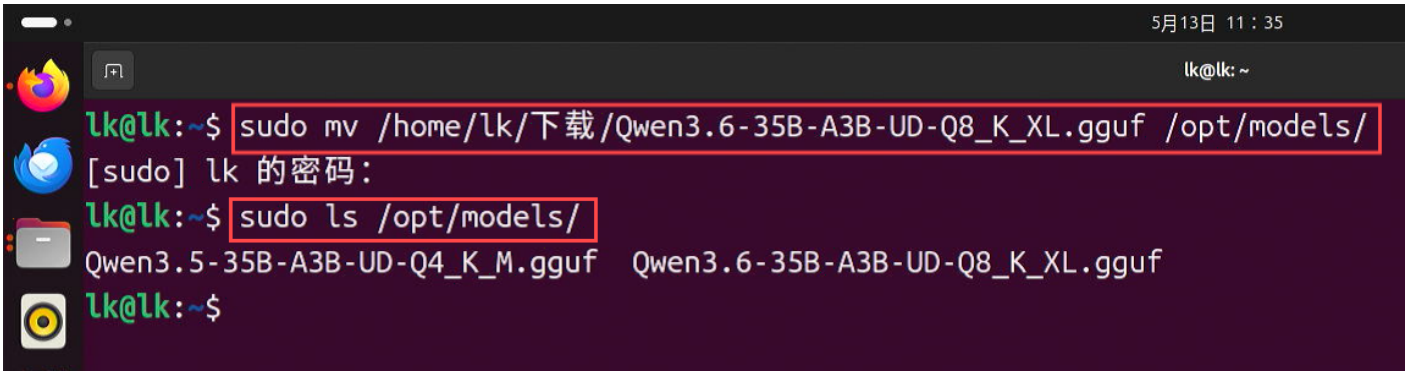
>

> ??????????????????????????

??????????????

```
sudo ls /opt/models/
```

??????????Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf?????????



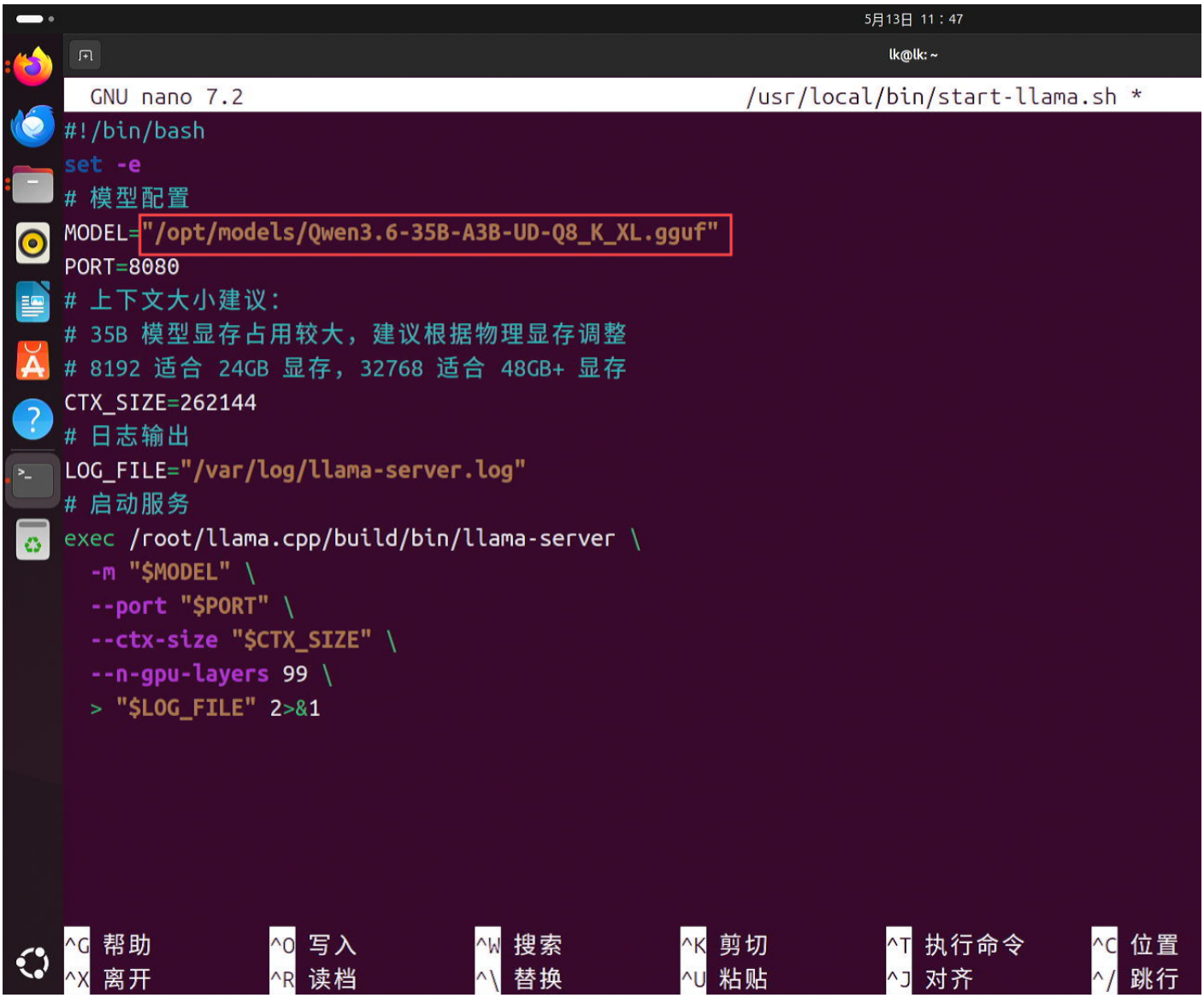
2. ?? llama ????

?????????????? llama ????????

```
sudo nano /usr/local/bin/start-llama.sh
```

?? `MODEL` ??????????????????????????????

???????? `Ctrl+X` - `Y` - `□□` ?????????



3. ?????????

?????? llama ?????????????

```
reboot
```

??????? `127.0.0.1:8080` ????????????? `Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf`
????????????????????????????????????



4. OpenClaw ?????

??????????? OpenClaw ???????????????

```
openclaw config
```

?? `Local` - `Model`

???? `vLLM`?

vLLM base URL ??? `http://127.0.0.1:8080/v1`?

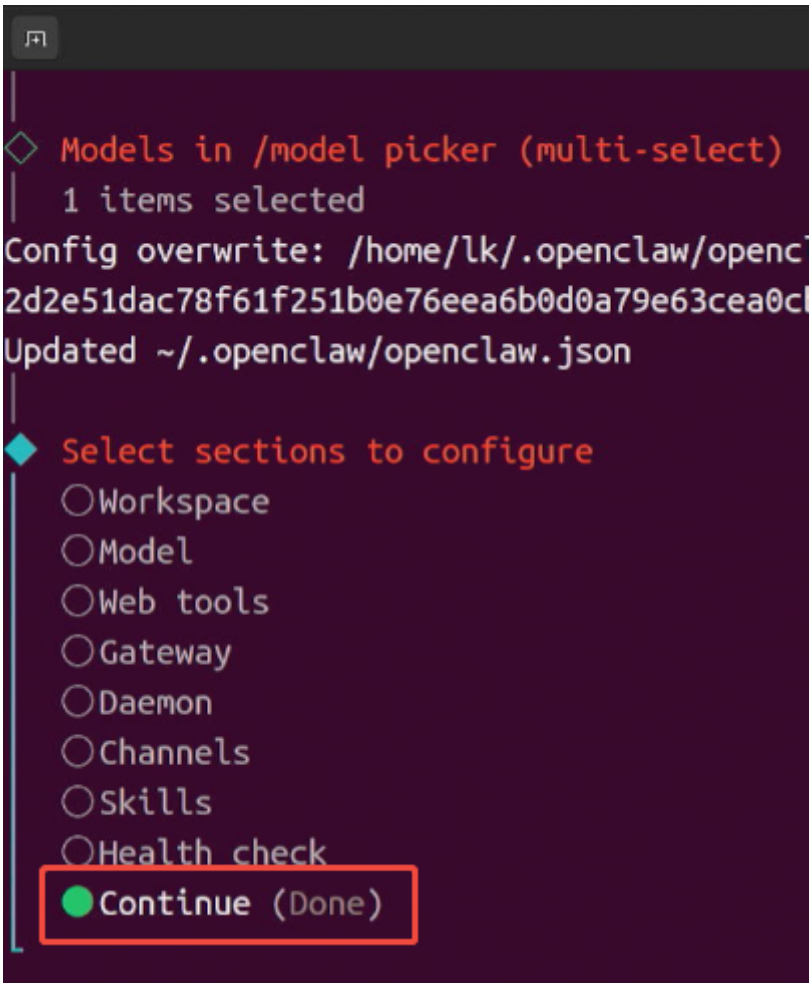
vLLM API Key ?? `sk-local`????????

vLLM model ?? `Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf`

??????????????


Where will the Gateway run?
Local (this machine)
Select sections to configure
Model
Model/auth provider
vLLM
vLLM base URL
http://127.0.0.1:8080/v1
vLLM API key
sk-local
vLLM model
Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf
Model configured
Default model set to vllm/Qwen3.6-35B-A3B-UD-Q8_K_XL.gguf

????????????? `Continue` ??????????



??????????? OpenClaw Gateway ?????????

```
openclaw gateway restart
```

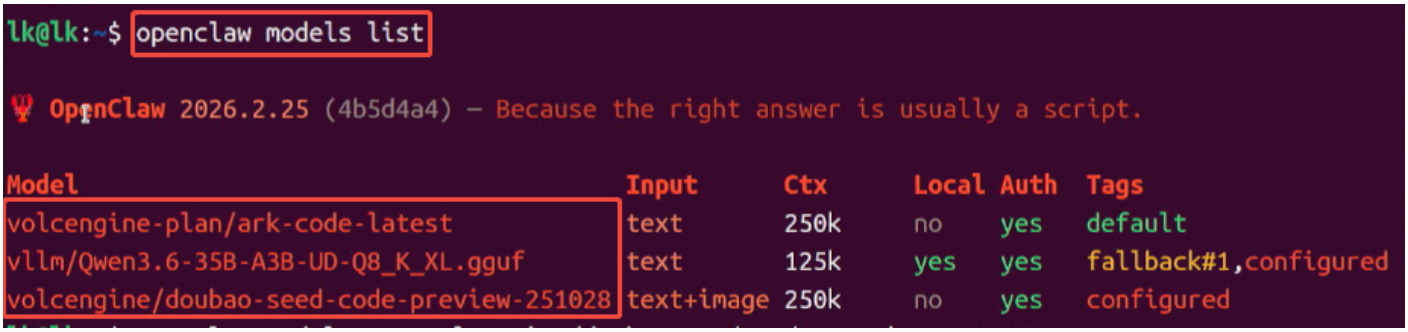
??????????? OpenClaw?



?????

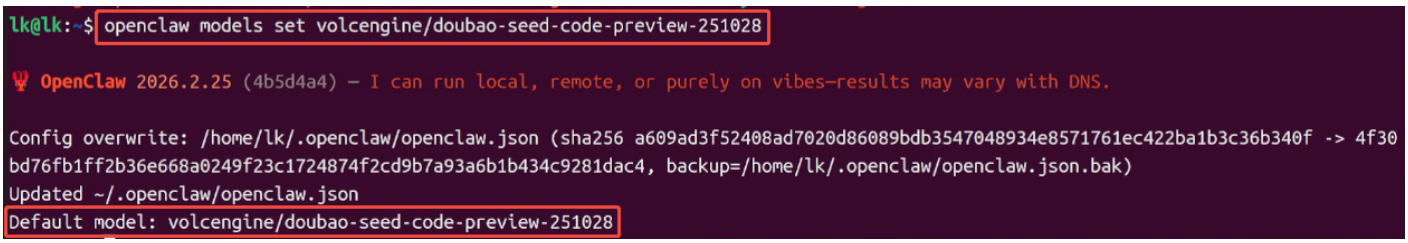
??

```
openclaw models list
```

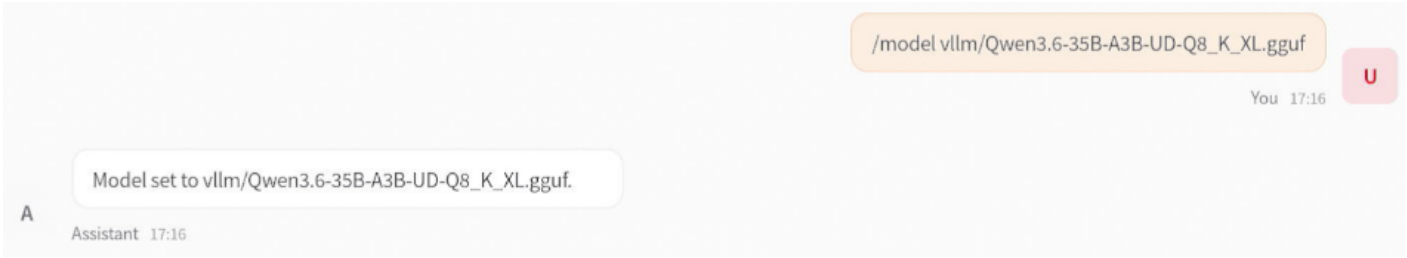


???????????????? Gateway????

```
openclaw models set □□□□
```



????????????? OpenClaw ???????`/model □□□□` ??????????



??????

???????????????

```
sudo apt install mesa-utils
```

???????????

```
glxinfo | grep -i "video memory\|vram"
```

??????????

```
lk@lk:~$ glxinfo | grep -i "video memory\|vram"
    Video memory: 98304MB
    Dedicated video memory: 98304 MB
    Currently available dedicated video memory: 55475 MB
```

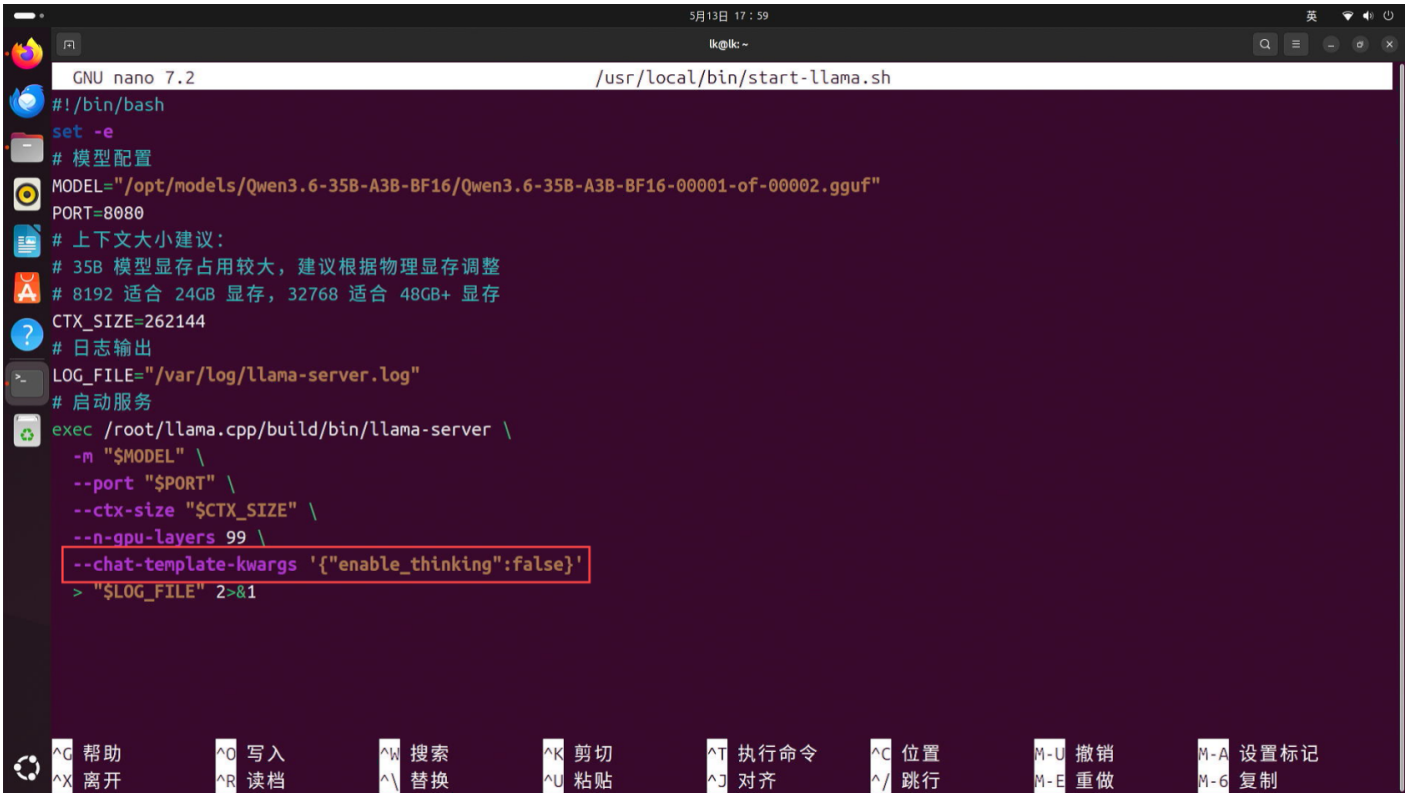
> ❝ Video Memory??????
>
> Currently available dedicated video memory???????

????????????? 54G??? GTR9Pro ????????????????????????????

??????

Qwen3.6-35B-A3B ???????????

???????????????????????? "--n-gpu-layers 99 " ??????? `--chat-template-kwargs '{"enable_thinking":false}'` ???



???????? `Ctrl+X` - `Y` - `□□` ?????????

?????????



Revision #6
Created 2026-05-13 07:18:26 CST by xiaoshao
Updated 2026-05-13 10:11:41 CST by xiaoshao