



#7

安装和使用定制私有 AI 工作站



课程目标

- 了解定制私有 AI 工作站的基础知识
- 安装和配置工作站
- 使用主要 AI 工具

课程内容

定制私有 AI 工作站
简介

工作站安装和配置

AI 工具概览

系统维护和优化

定制私有 AI 工作站简介

定制私有 AI 工作站简介

- AI 工作站解决方案概述
 - 本地端运行 AI 工具
 - 运用自己的 GPU 做运算
- 主要特点和优势
 - 无需网络
 - 保障用户资料
 - 无需订阅费用

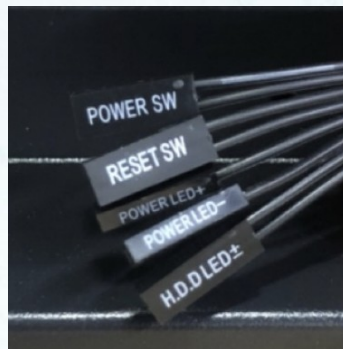
工作站安装和配置

工作站安装和配置

- 逐步安装指南
- 安装作业系统

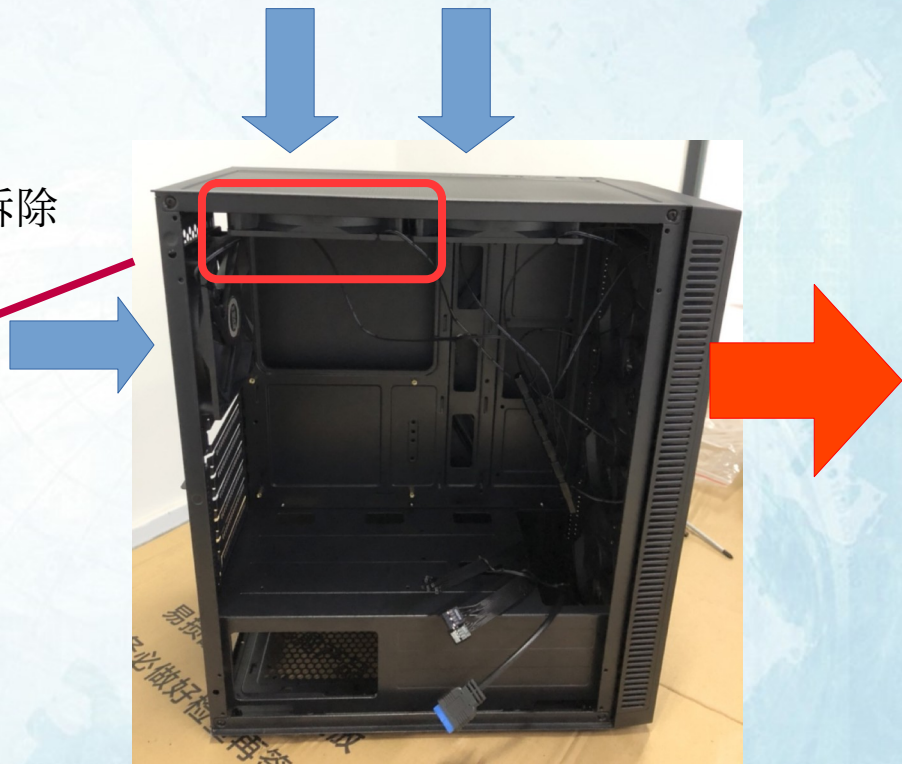
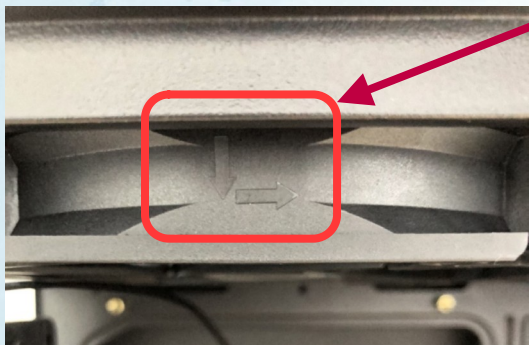
工作站安装和配置

- 机箱
- 机箱外置面板控制线



工作站安装和配置

- 机箱风扇
- 从上方、后方进气，从前方出气
- 为了方便操作，会先把进气的三把风扇拆除



工作站安装和配置

- 主机板
 - 华南 X99-F8D PLUS
- 主机后挡板
- 导热膏
- 胶钉



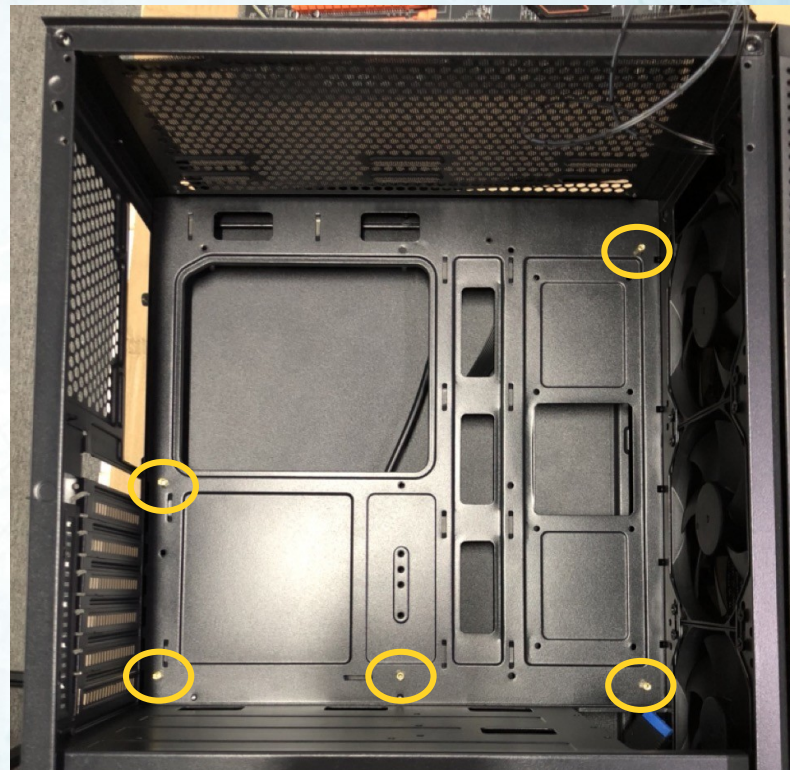
工作站安装和配置

- 在机箱装上主机板的挡板
- 注意后置 I/O 面板上的插口位置
- 由里面往外推，装上主机板的后挡板



工作站安装和配置

- 根据主机板，改变机箱的铜柱位置
- 对不到铜柱的位置，在主机板上放上铰钉



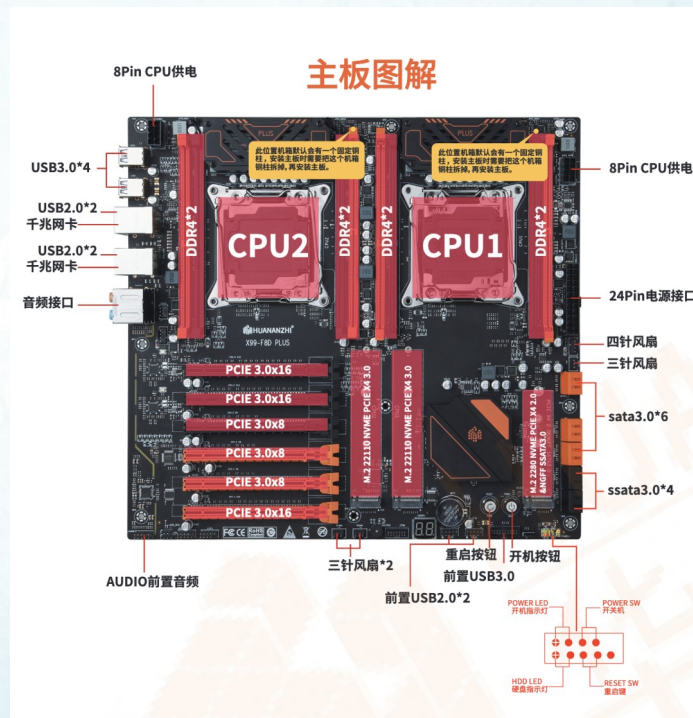
工作站安装和配置

- 安置主机板到机箱里
- 在能对上铜柱的位置锁上螺丝



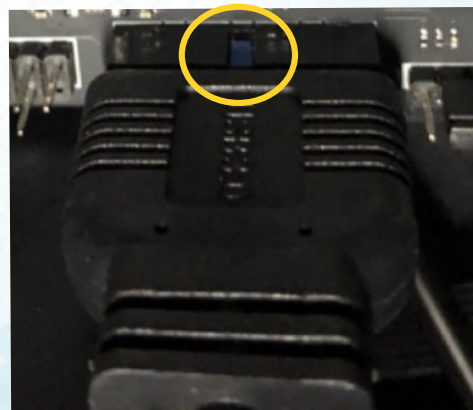
工作站安装和配置

- 查看主板说明书



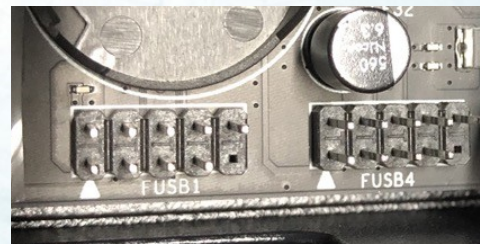
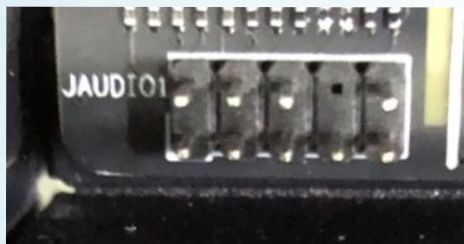
工作站安装和配置

- 连接机箱外置面板控制线 - USB3.0



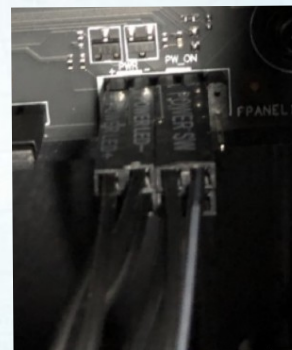
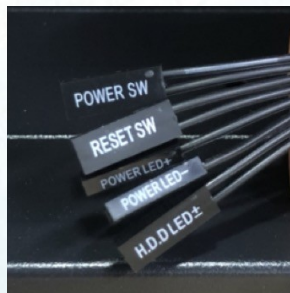
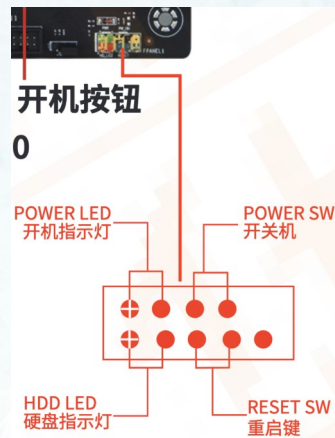
工作站安装和配置

- 连接机箱外置面板控制线 - USB2.0 和 AUDIO



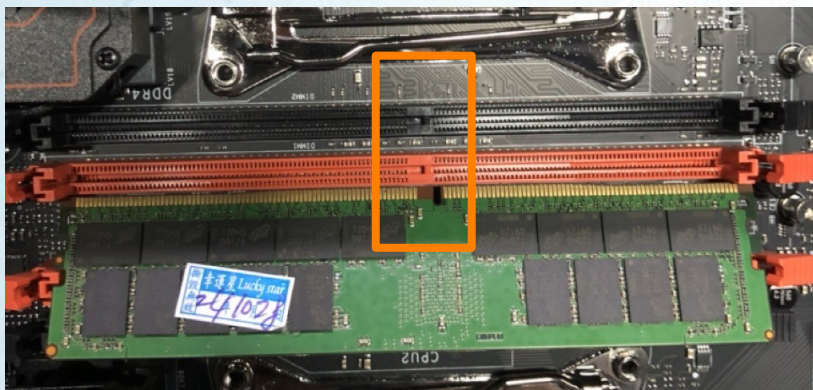
工作站安装和配置

- 连接机箱信号线和控制线
- 查看主板说明书



工作站安装和配置

- 插上 RAM
- 中间有凹槽
- 对应主机板的插槽



工作站安装和配置

- 安装 CPU 散热风扇 - 气流方向
- 例子：塔式散热器

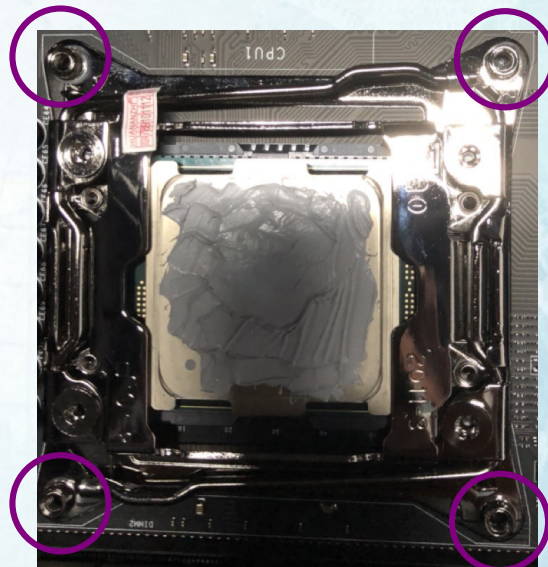


热管



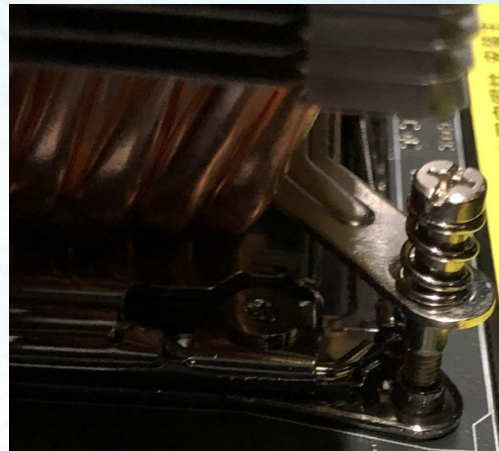
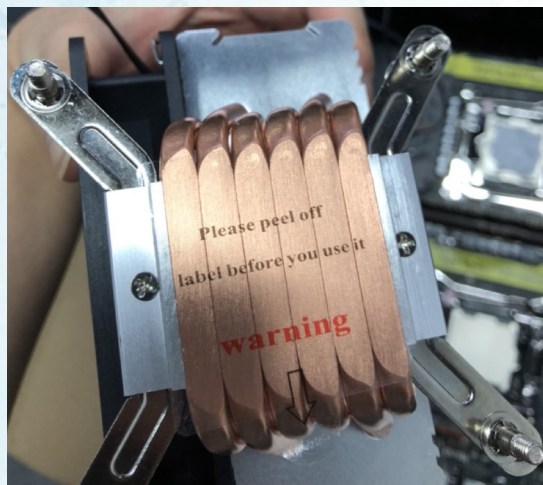
工作站安装和配置

- 安装 CPU 散热风扇 - 涂导热膏
- 目标为广范围分布，填充与热管相接处的孔隙



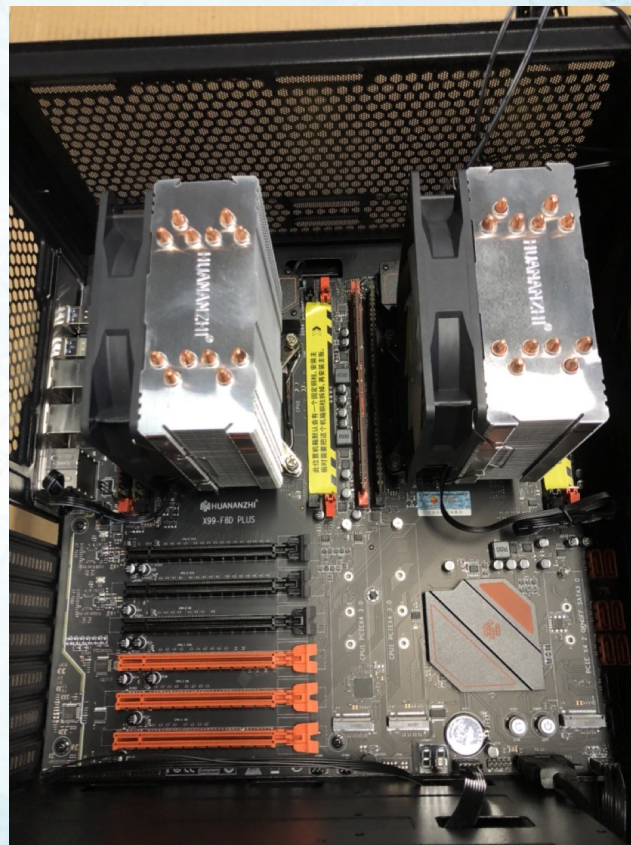
工作站安装和配置

- 撕去热管的防尘胶纸
- 放上塔式散热器
- 对好螺丝的位置，装到主机板上



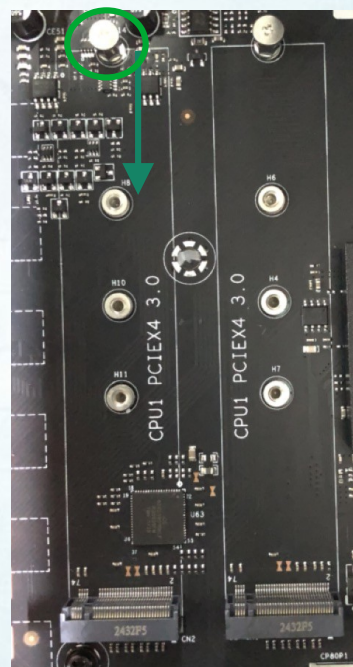
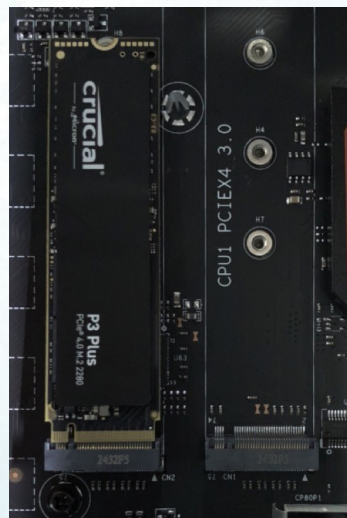
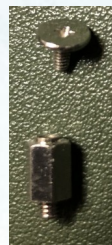
工作站安装和配置

- 连接 CPU 风扇的电线
- 电线有凹槽，对应主板上的特定插槽
- 重复步骤 安装另外一个 CPU 风扇



工作站安装和配置

- 安装硬碟（例子： M.2 硬碟）
- 移除 螺丝 和 螺丝底座
- 把 螺丝底座 拧紧旋入到长度相符的位置



工作站安装和配置

- 将 M.2 硬碟以 30 度角插到 M.2 接口
- 将 螺丝 放在 M.2 硬碟后缘缺口上并拧紧到 螺丝底座



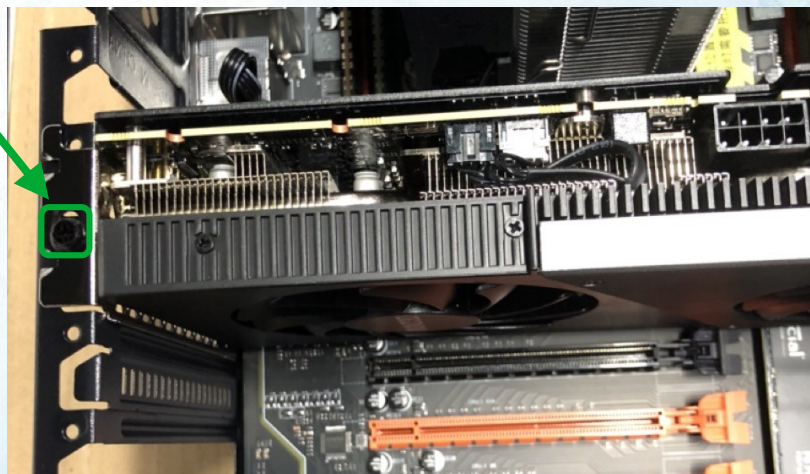
工作站安装和配置

- 安装显卡
- 确认插槽位置
- 拆除不需要的机箱挡板



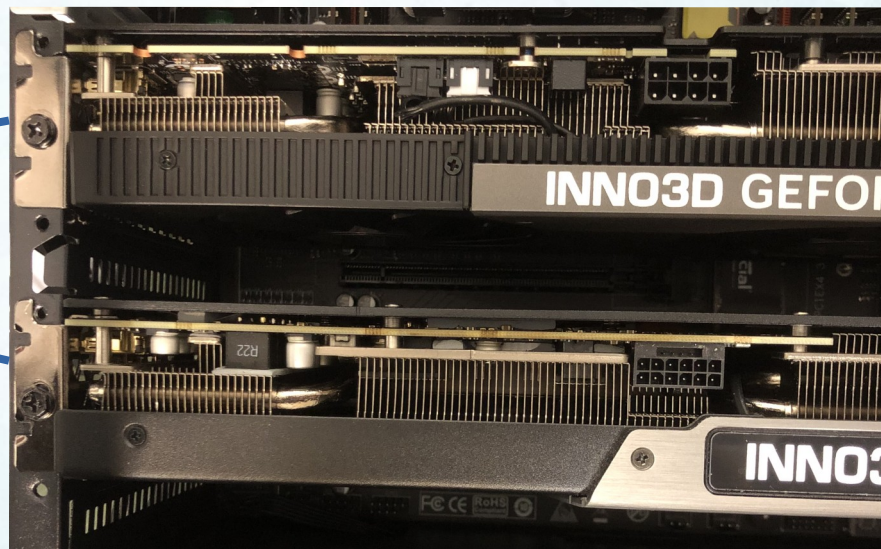
工作站安装和配置

- 安装显卡（例子： NVIDIA RTX 4070 Ti Super 和 NVIDIA RTX 4060 Ti ）
- 卡上有凹槽，对应插槽的方向
- 插上后用螺丝固定在机箱上



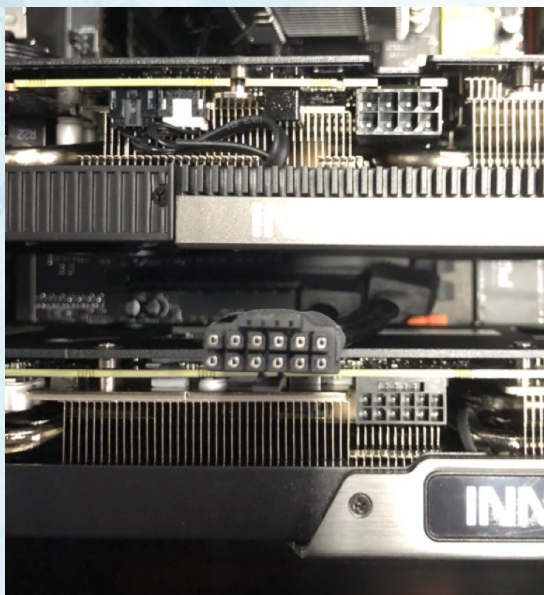
工作站安装和配置

- 如此重复安装另外一张显卡



工作站安装和配置

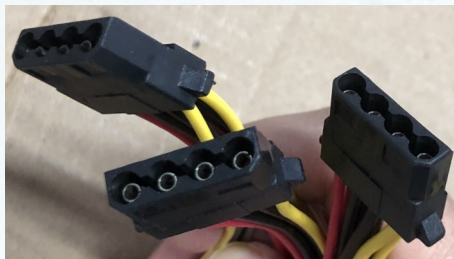
- 某些显卡会配备特殊的电线
- 为 RTX 4070 Ti Super 接上配备的电线



工作站安装和配置

- 供电器
- 电线类型

机箱风扇电线



24Pin 主板电线



8Pin CPU 电线



8Pin GPU 电线



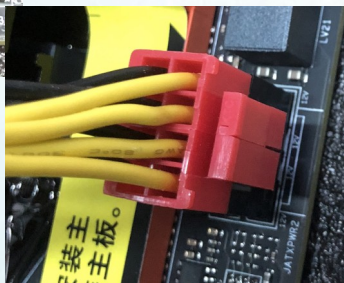
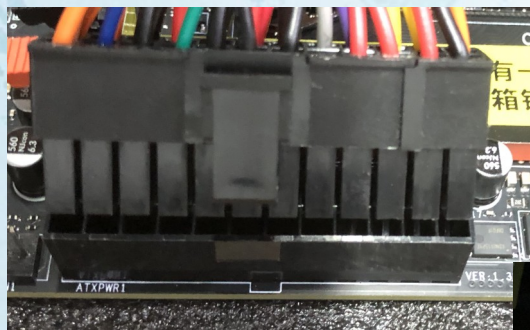
工作站安装和配置

- 放置供电器
 - 供电器的风扇朝下
 - 从另一边的侧方放进机箱
- 用螺丝固定



工作站安装和配置

- 连接 24Pin 主板的供电
- 连接两个 8Pin CPU 供电



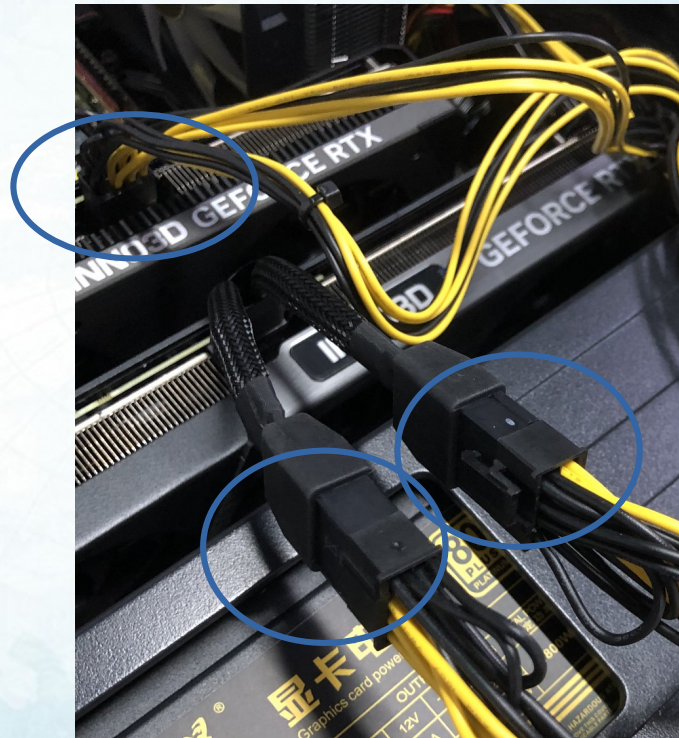
工作站安装和配置

- 记得还要把刚拆下的风扇装回去
- 连接风扇供电



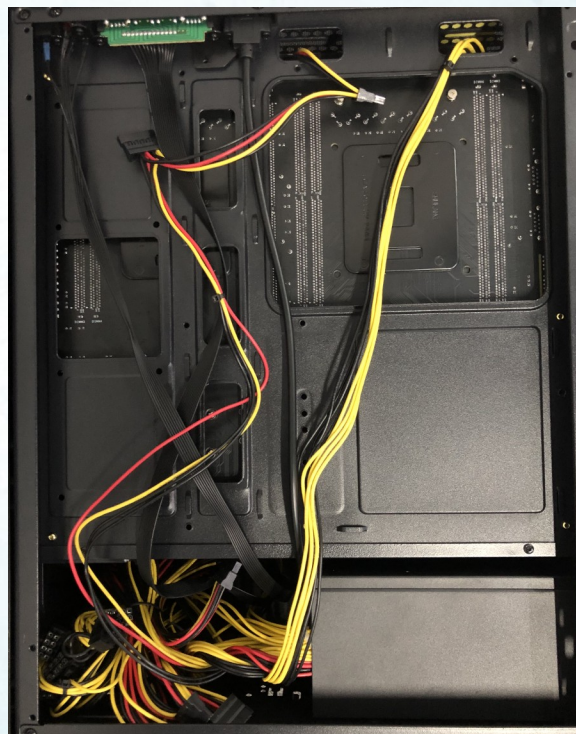
工作站安装和配置

- 连接显卡供电
- 4060 Ti 需要一条 8Pin
- 4070 Ti Super 需要两条 8Pin



工作站安装和配置

- 整理电线，防止卡到风扇 或 造成短路



工作站安装和配置

- 封起机箱
- 接上网络数据线、键盘、滑鼠、显示器（HDMI 接在显卡上）
- 接上电源
- 下一步，安装作业系统（OS）

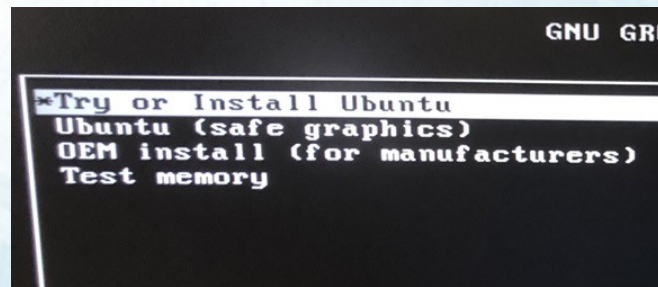
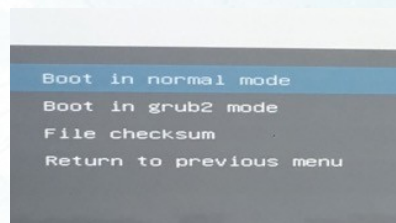


工作站安装和配置

- 安装 OS , 这次选用 Ubuntu 22.04.5 Desktop
- 需要制作 可启动 U 盘
 - 使用 Rufus 或者 Ventoy
- 下载并放置 ubuntu-22.04.5-desktop-amd64.iso 文件在 U 盘里

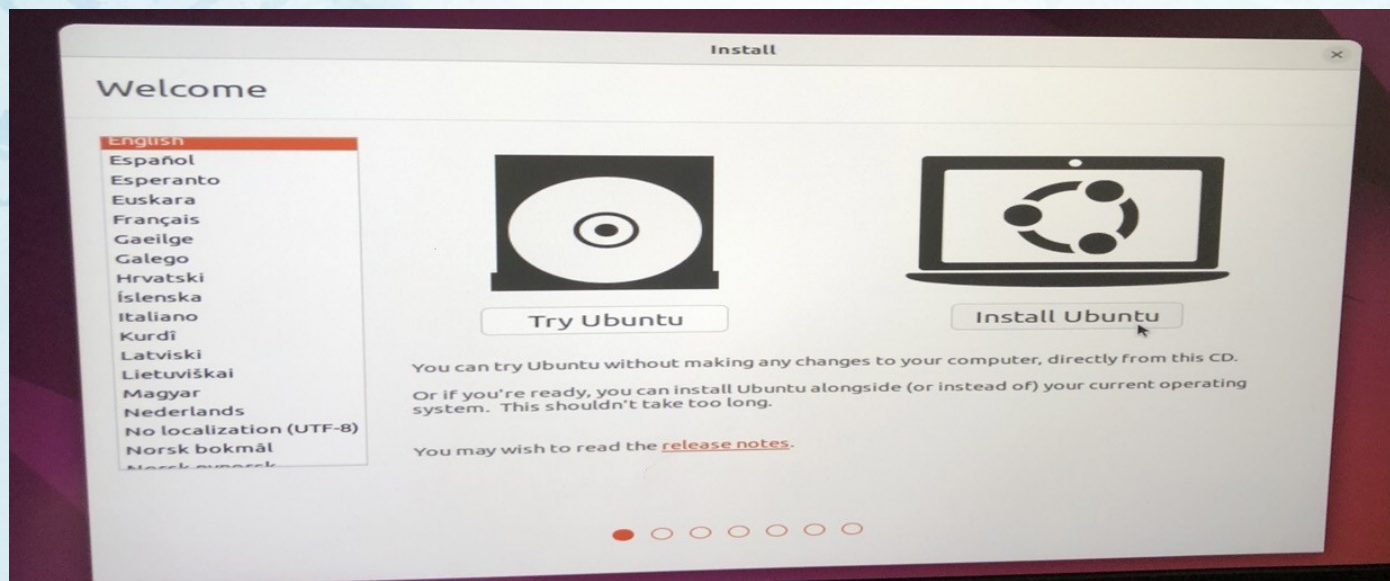
工作站安装和配置

- 插上 可启动 U 盘，开启电脑
- 进入 Ventoy 画面，选择 ubuntu-22.04.5-desktop-amd64.iso
- Boot in normal mode



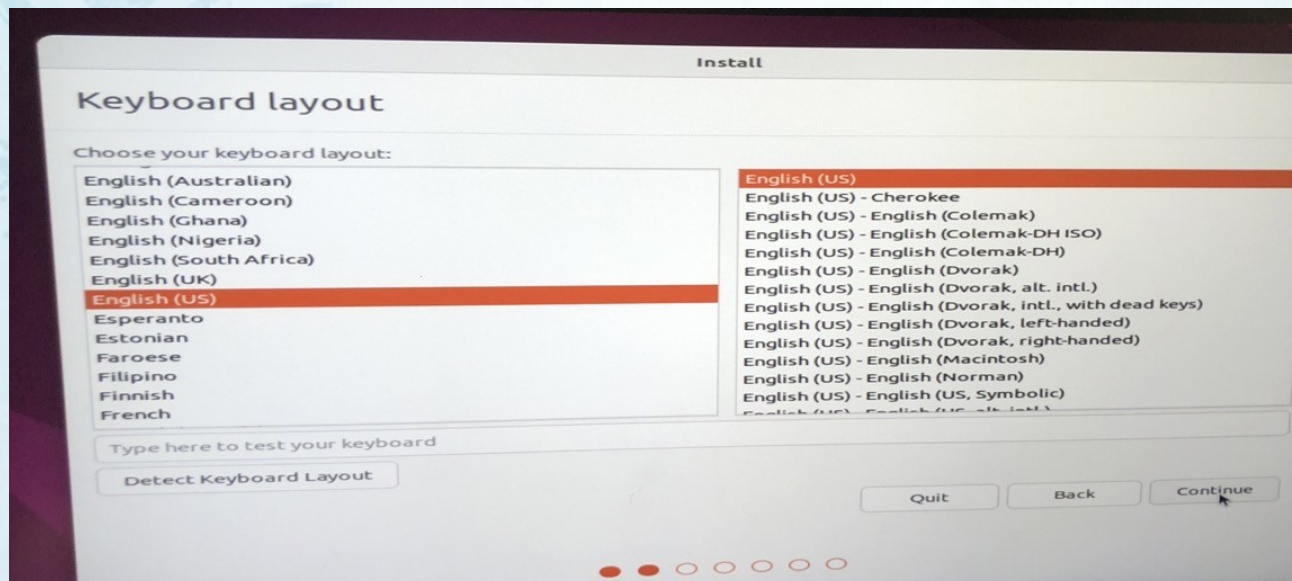
工作站安装和配置

- 开始 Ubuntu 安装程序，选择 Install Ubuntu



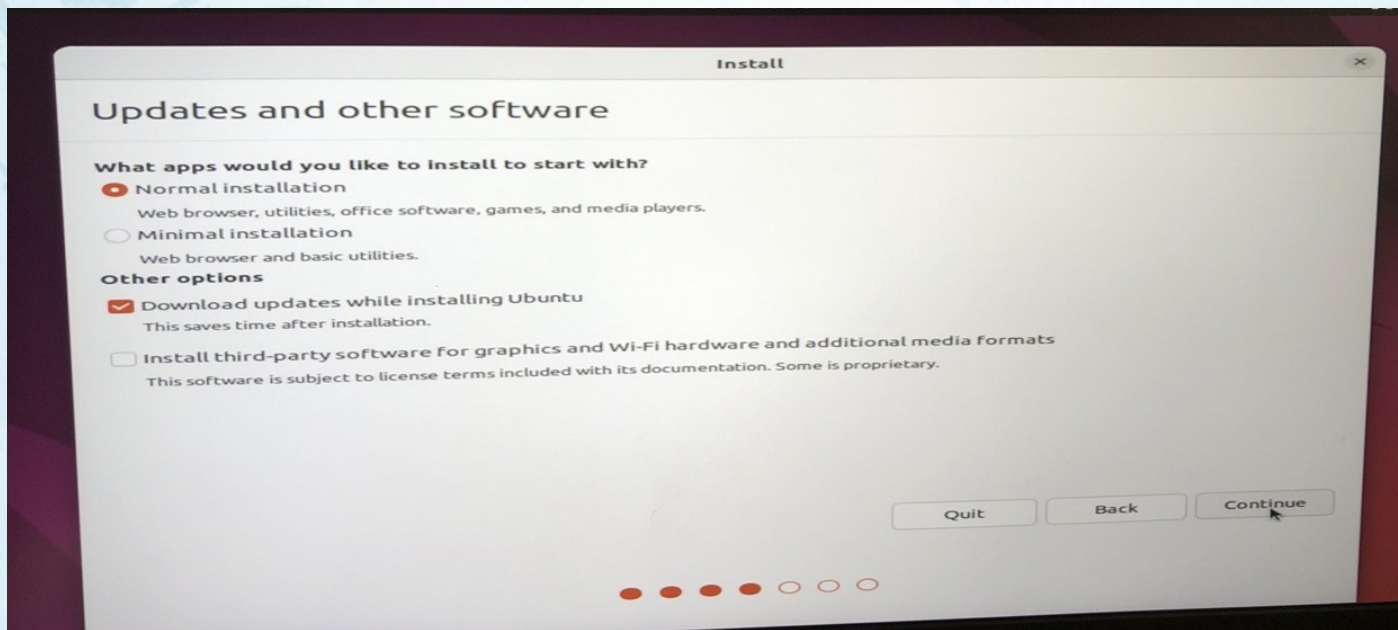
工作站安装和配置

- 选择键盘设定，通常都预设就可以继续。



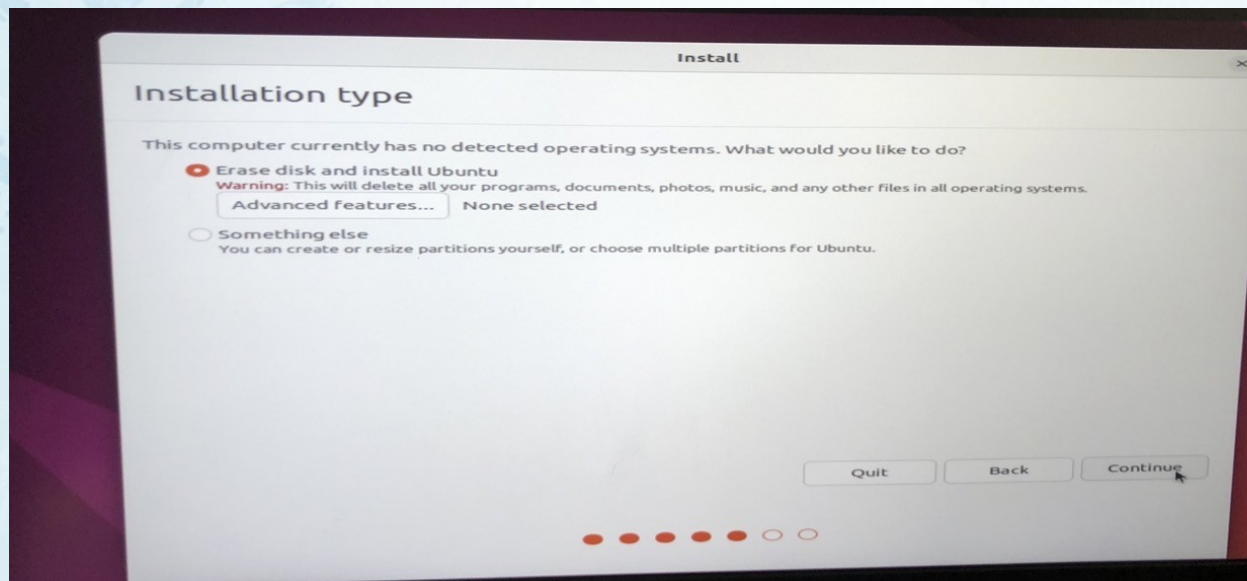
工作站安装和配置

- 选择 Normal installation , 自动安装浏览器, 影像播放器之类的基本工具



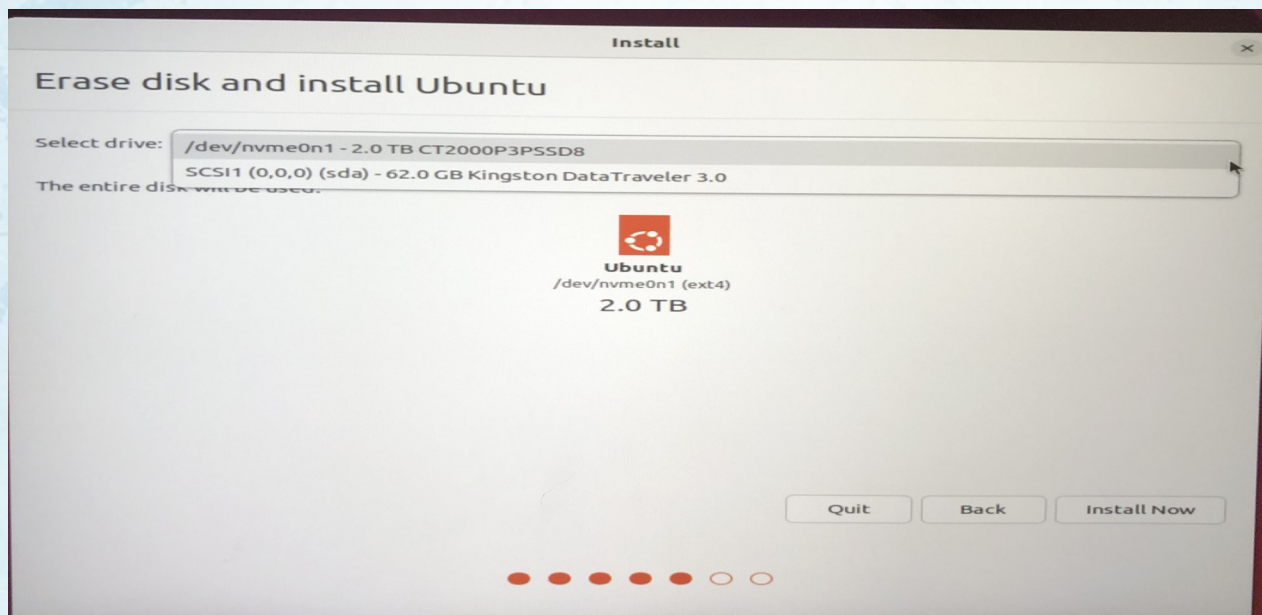
工作站安装和配置

- 选择 Erase disk and install Ubuntu



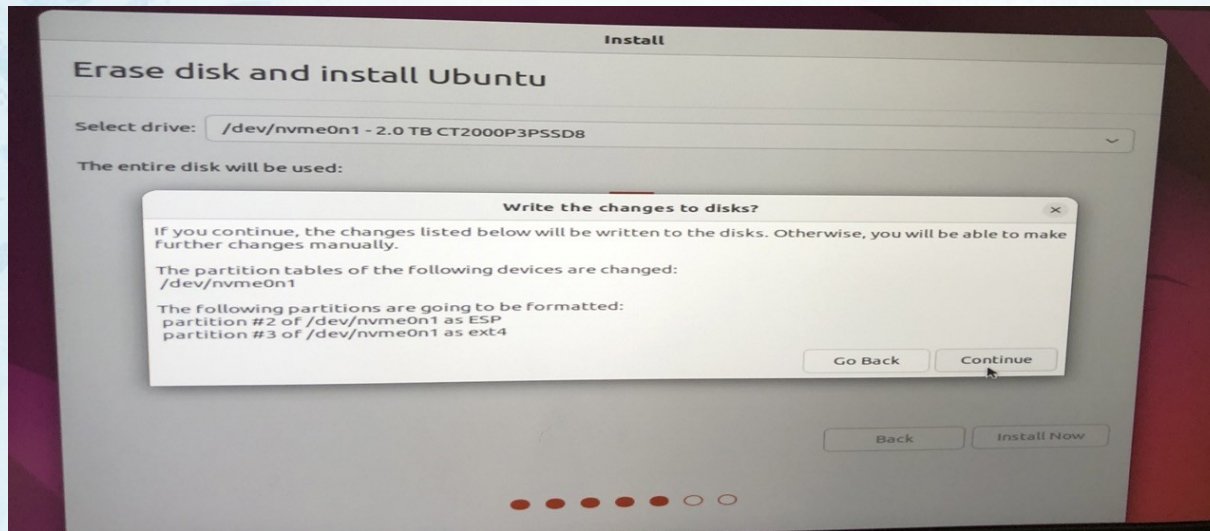
工作站安装和配置

- 选取唯一的硬碟，然后 Install Now



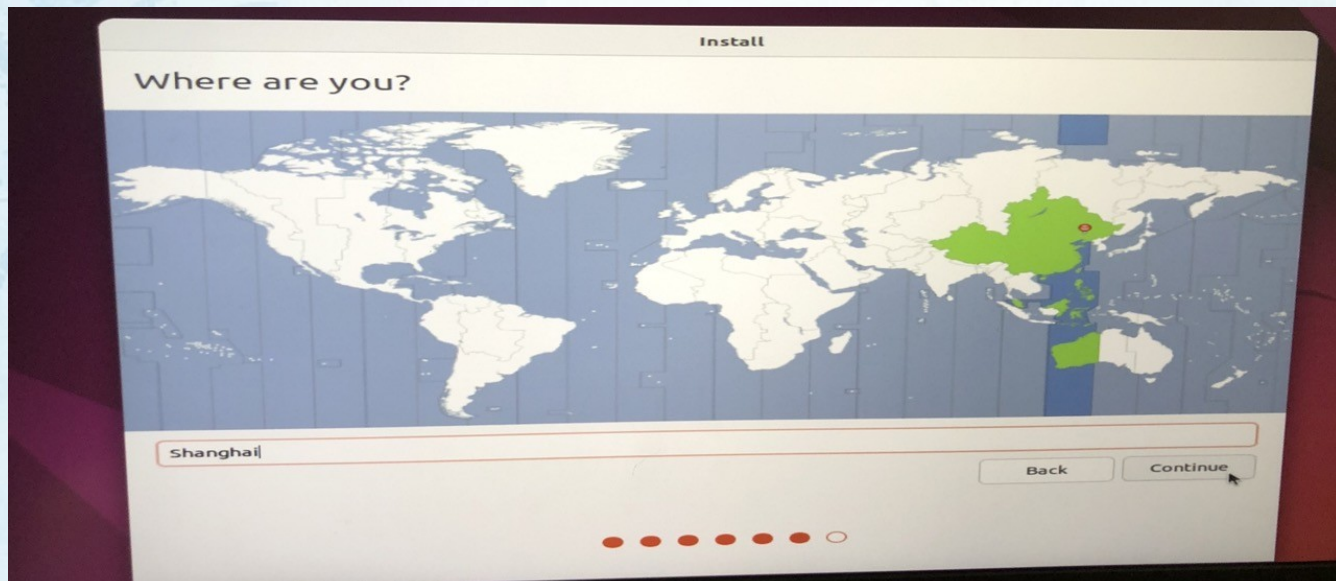
工作站安装和配置

- 点击 Continue 继续安装



工作站安装和配置

- 选择地区和时区



工作站安装和配置

- 设置账户，然后开始安装。

Install

Who are you?

Your name: ✓

Your computer's name: ✓
The name it uses when it talks to other computers.

Pick a username: ✓

Choose a password: ⓘ Short password

Confirm your password: ✓

☐ Log in automatically

☒ Require my password to log in

☐ Use Active Directory

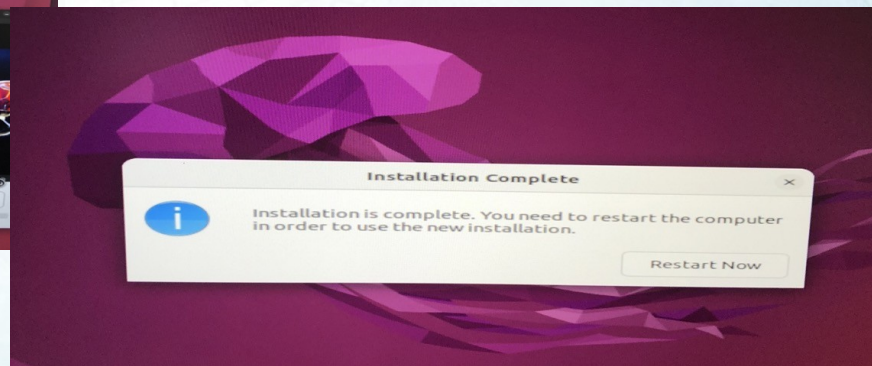
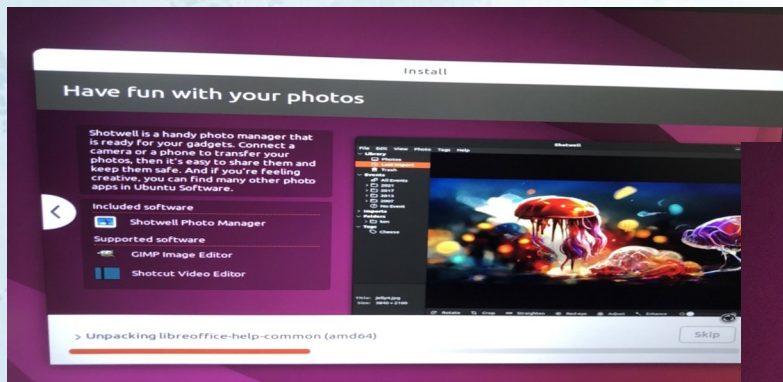
You'll enter domain and other details in the next step.

Back Continue

•••••••

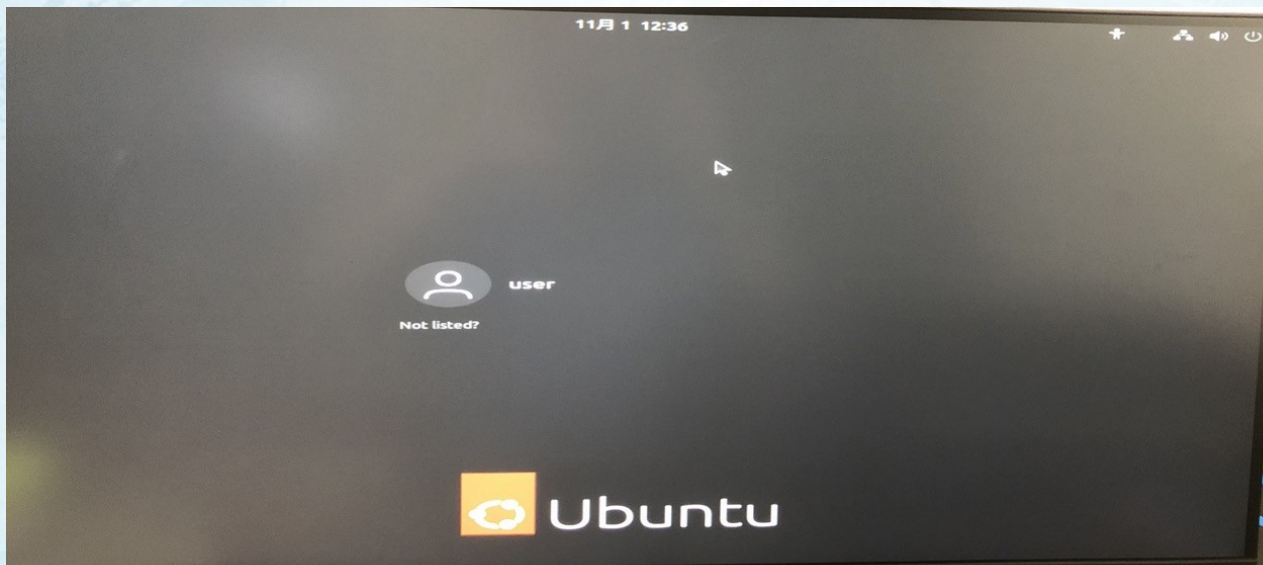
工作站安装和配置

- 等待安装完毕，然后重启电脑
- 记得要拔掉可启动 U 盘



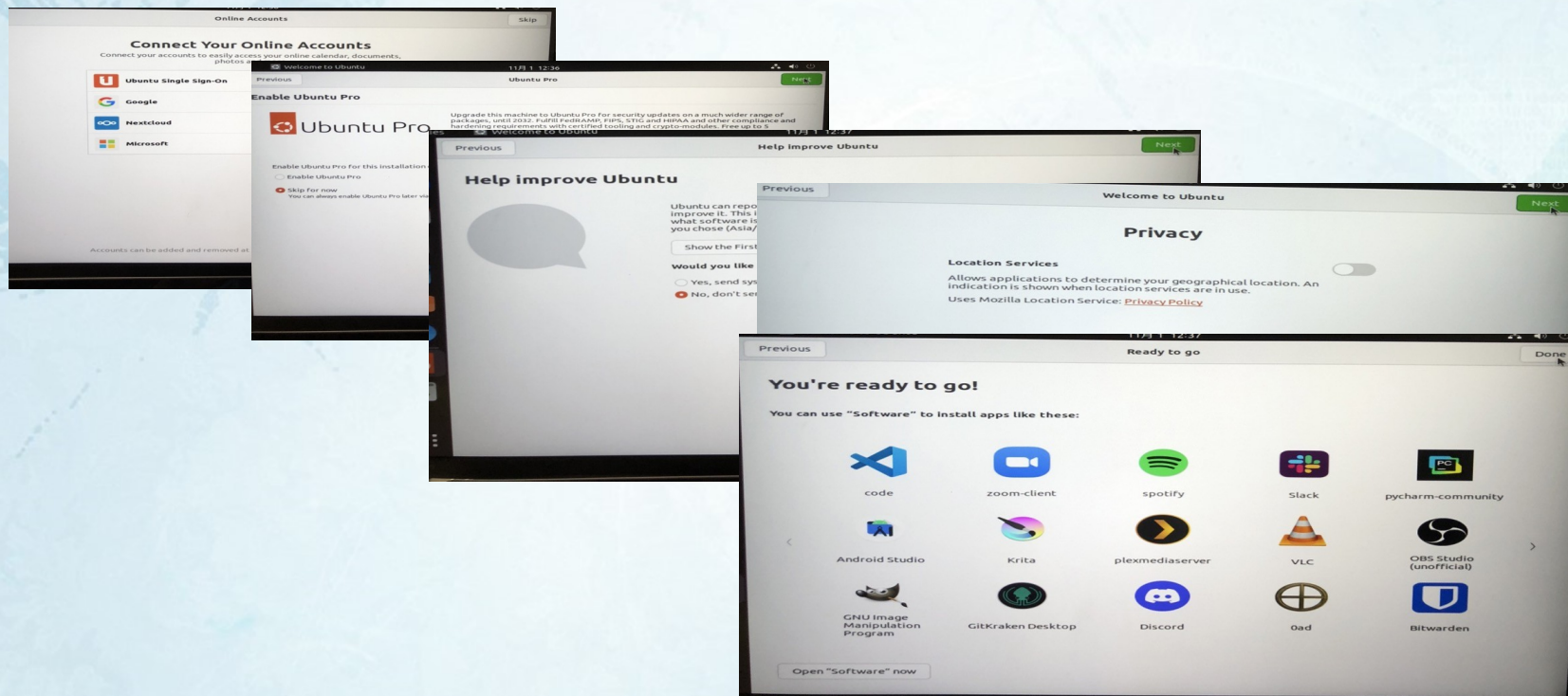
工作站安装和配置

- 电脑重启后用刚才设置的账户登入
- 可以发现画面显示的比例有点怪。



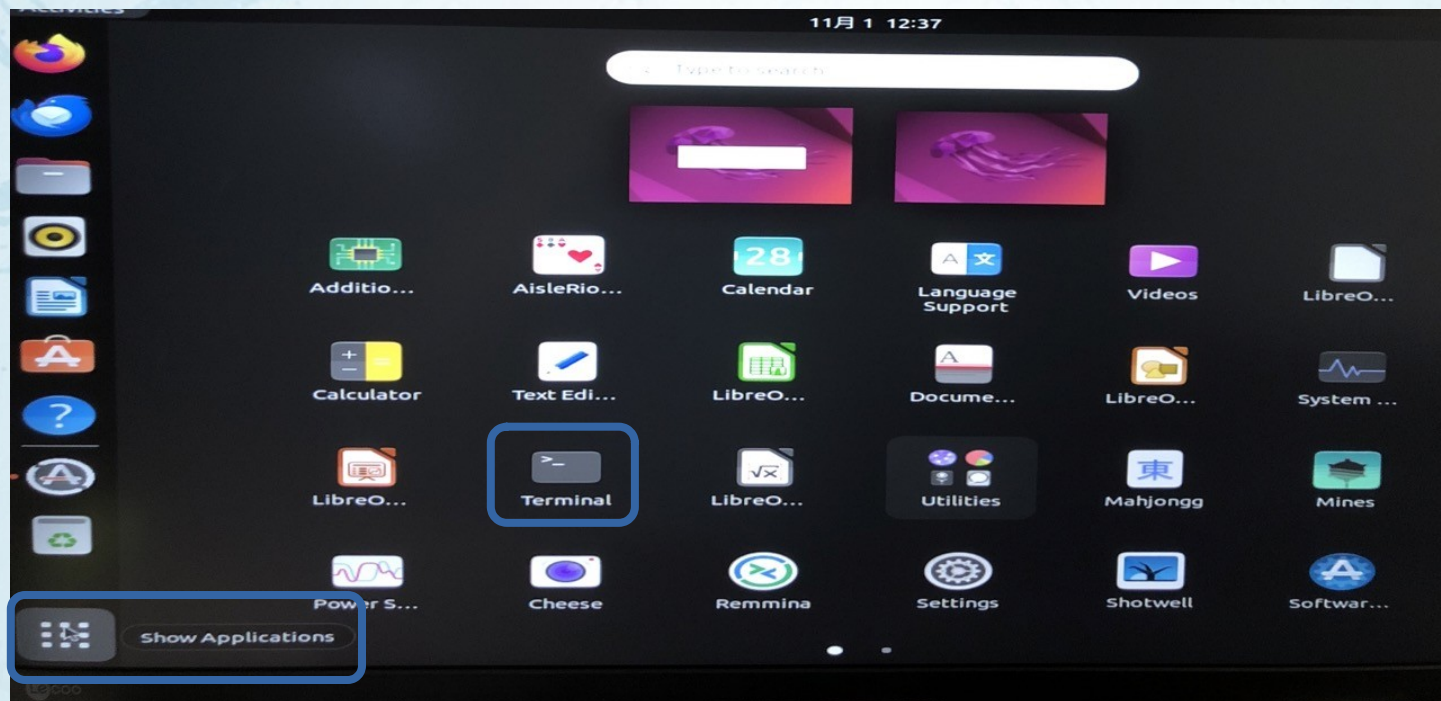
工作站安装和配置

- 跳过一堆弹窗



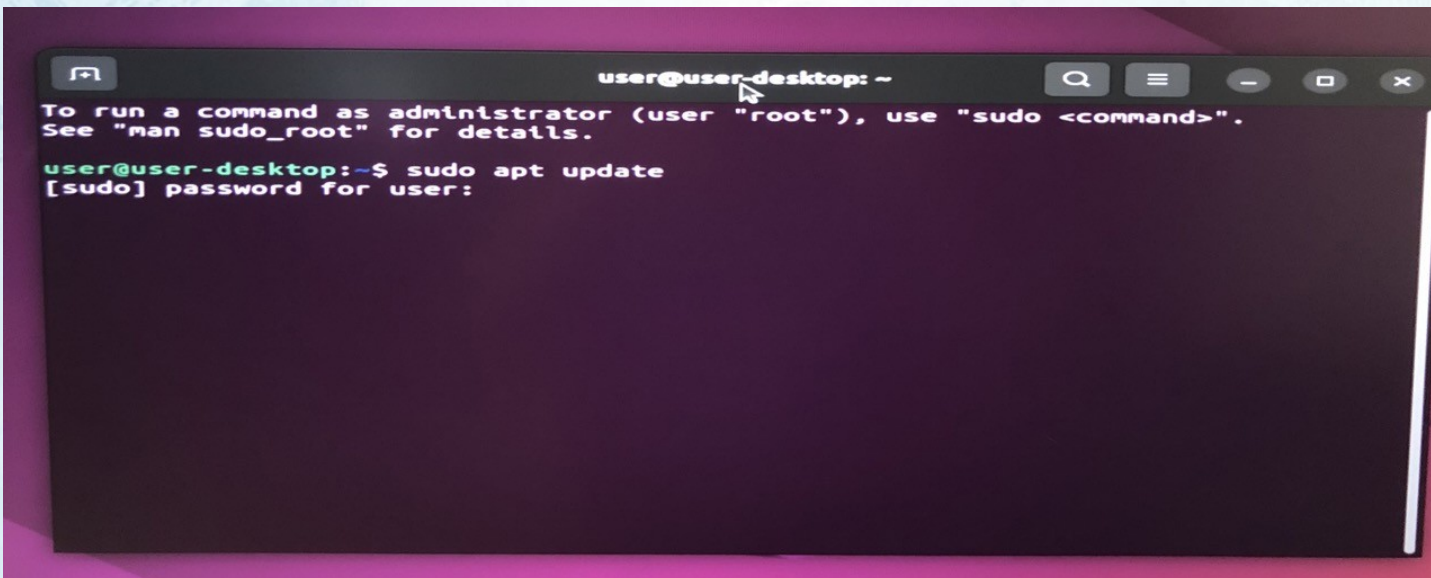
工作站安装和配置

- 开启 Terminal 做基本的更新，让画面显示正常



工作站安装和配置

- 运行指令 `sudo apt update`
- 凡使用 `sudo` 指令都需要输入当前用户的密码



```
user@user-desktop: ~  
To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo_root" for details.  
user@user-desktop:~$ sudo apt update  
[sudo] password for user:
```


工作站安装和配置

- 运行 `sudo apt upgrade`
- 输入 Y 继续安装

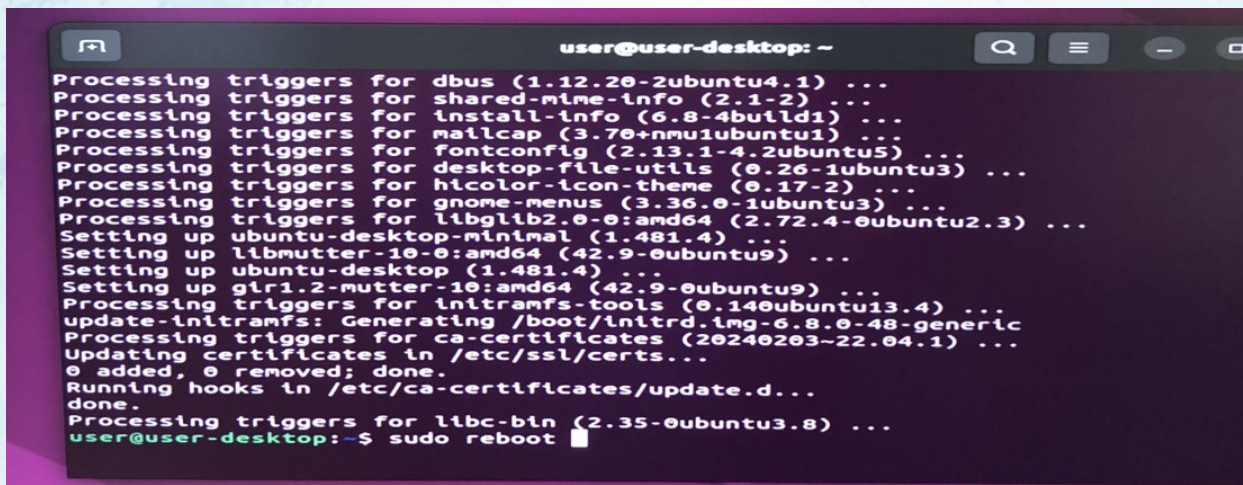
```
See 'man sudo_root' for details.
user@user-desktop:~$ sudo apt update
[sudo] password for user:
Sorry, try again.
[sudo] password for user:
Hit:1 http://cn.archive.ubuntu.com/ubuntu jammy InRelease
Hit:2 http://cn.archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:3 http://cn.archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:4 http://security.ubuntu.com/ubuntu jammy-security InRelease
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
98 packages can be upgraded. Run 'apt list --upgradable' to see them.
user@user-desktop:~$ sudo apt upgrade
```

```
xserver-xorg-legacy xwayland xxd
98 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
77 standard LTS security updates.
Need to get 0 B/510 MB of archives.
After this operation, 51.1 MB of additional disk space will be used.
Do you want to continue? [Y/n] Y
```

```
user@user-desktop: ~
Preparing to unpack .../96-libreoffice-ogltrans_1%3a7.3.7-0ubuntu0.22.04.7_all.d
eb ...
Unpacking libreoffice-ogltrans (1:7.3.7-0ubuntu0.22.04.7) over (1:7.3.7-0ubuntu
0.22.04.6) ...
Preparing to unpack .../97-libreoffice-pdfimport_1%3a7.3.7-0ubuntu0.22.04.7_all
.deb ...
Unpacking libreoffice-pdfimport (1:7.3.7-0ubuntu0.22.04.7) over (1:7.3.7-0ubuntu
0.22.04.6) ...
Setting up python3-pkg-resources (59.6.0-1.2ubuntu0.22.04.2) ...
Setting up libxpat1:amd64 (2.4.7-1ubuntu0.4) ...
Setting up libgssr1-common (1:1.14.47-1ubuntu0.1) ...
Setting up bubblewrap (0.6.1-1ubuntu0.1) ...
Setting up libbapparmori:amd64 (3.0.4-2ubuntu2.4) ...
Setting up xdg-desktop-portal (1.14.4-1ubuntu2-22.04.2) ...
Setting up libgjs0g:amd64 (1.72.4-0ubuntu0.22.04.4) ...
Setting up openvpn (2.5.11-0ubuntu0.22.04.1) ...
Setting up ubuntu-drivers-common (1:0.9.0-2-0.22.04.0) ...
Setting up libfontembed1:amd64 (1.20.15-0ubuntu1.4) ...
Setting up linux-firmware (20220329.git681281e4-0ubuntu3.35) ...
update-initramfs: Generating /boot/initrd.img-6.8.0-48-generic
update-initramfs: Generating /boot/initrd.img-6.8.0-48-generic
Progress: [ 55%] [=====]
```

工作站安装和配置

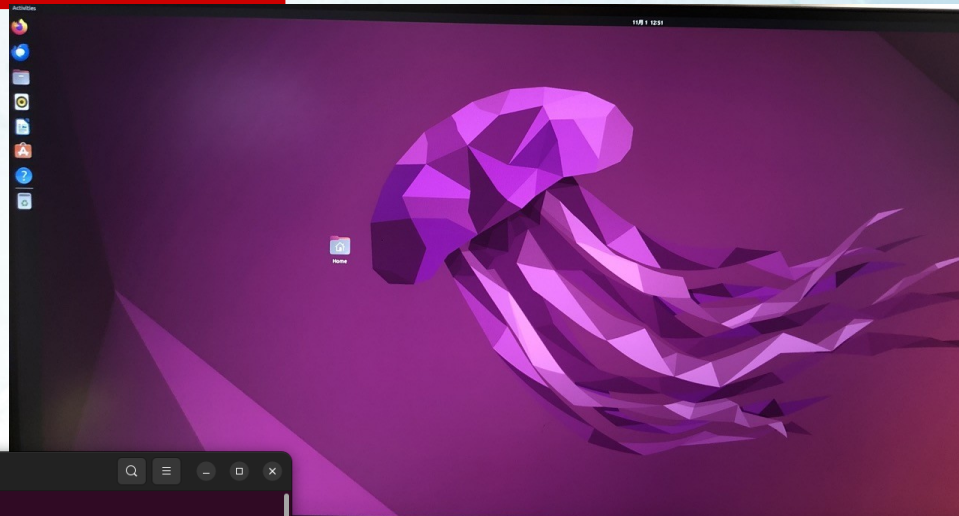
- 安装完成后需要重启电脑，可以运行指令 `sudo reboot`



```
user@user-desktop: ~  
Processing triggers for dbus (1.12.20-2ubuntu4.1) ...  
Processing triggers for shared-mime-info (2.1-2) ...  
Processing triggers for install-info (6.8-4build1) ...  
Processing triggers for mailcap (3.70+nmuiubuntu1) ...  
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...  
Processing triggers for desktop-file-utils (0.26-1ubuntu3) ...  
Processing triggers for hicolor-icon-theme (0.17-2) ...  
Processing triggers for gnome-menus (3.36.0-1ubuntu3) ...  
Processing triggers for libgl1b2.0-0:amd64 (2.72.4-0ubuntu2.3) ...  
Setting up ubuntu-desktop-minimal (1.481.4) ...  
Setting up libmutter-10-0:amd64 (42.9-0ubuntu9) ...  
Setting up ubuntu-desktop (1.481.4) ...  
Setting up gir1.2-mutter-10:amd64 (42.9-0ubuntu9) ...  
Processing triggers for initramfs-tools (0.140ubuntu13.4) ...  
update-initramfs: Generating /boot/initrd.img-6.8.0-48-generic  
Processing triggers for ca-certificates (20240203~22.04.1) ...  
Updating certificates in /etc/ssl/certs...  
0 added, 0 removed; done.  
Running hooks in /etc/ca-certificates/update.d...  
done.  
Processing triggers for libc-bin (2.35-0ubuntu3.8) ...  
user@user-desktop:~$ sudo reboot
```


工作站安装和配置

- 电脑重启后，画面比例变正常
- 运行指令 `sudo lshw -C display`
- 可以见到 driver 不是 nvidia
- 下一步要安装 nvidia driver



```
user@user-desktop: ~  
user@user-desktop: $ sudo lshw -C display  
[sudo] password for user:  
*-display  
  description: VGA compatible controller  
  product: NVIDIA Corporation  
  vendor: NVIDIA Corporation  
  physical id: 0  
  bus info: pci@0000:05:00.0  
  logical name: /dev/fb0  
  version: a1  
  width: 64 bits  
  clock: 33MHz  
  capabilities: pm msi pciexpress vga_controller bus_master cap_list rom fb  
  configuration: depth=32 driver=nouveau latency=0 resolution=3440,1440  
  resources: irq:66 memory:00000000-00000000 memory:b0000000-bfffffff memory:c0000000-c1ffffff ioport:7000(size=128) memory:c0000-dffff  
*-display UNCLAIMED  
  description: VGA compatible controller  
  product: NVIDIA Corporation  
  vendor: NVIDIA Corporation  
  physical id: 0  
  bus info: pci@0000:04:00.0  
  version: a1  
  width: 64 bits  
  clock: 33MHz  
  capabilities: pm msi pciexpress vga_controller cap_list  
  configuration: latency=0  
  resources: memory:fa000000-faffffff memory:e0000000-efffffff memory:f0000000-f1ffffff ioport:f000(size=128) memory:fb000000-fb07ffff
```


工作站安装和配置

- 运行指令 `sudo apt purge nvidia`
 - 清除 nvidia 相关的 package
- 运行指令 `sudo ubuntu-drivers autoinstall`
- 完成后，重启电脑

```
user@user-desktop:~$ sudo apt purge nvidia
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Package 'nvidia' is not installed, so not removed
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
user@user-desktop:~$
```

```
user@user-desktop:~$ sudo ubuntu-drivers autoinstall

Found initrd image: /boot/initrd.img-6.8.0-40-generic
Found memtest86+ image: /boot/memtest86+.elf
Found memtest86+ image: /boot/memtest86+.bin
Warning: os-prober will not be executed to detect other
Systems on them will not be added to the GRUB boot config
Check GRUB_DISABLE_OS_PROBER documentation entry.
done
user@user-desktop:~$ sudo reboot
```

工作站安装和配置

- 重新开机后，运行指令 `sudo lshw -C display`
- 可见 driver 更新了，现在是用 nvidia driver
- 运行指令 `nvidia-smi` 可以看见显卡详细资料

```

user@user-desktop: ~
user@user-desktop:~$ sudo lshw -C display
[sudo] password for user:
*-display
   description: VGA compatible controller
   product: NVIDIA Corporation
   vendor: NVIDIA Corporation
   physical id: 0
   bus info: pci@0000:05:00:0
   version: a1
   width: 64 bits
   clock: 33MHz
   capabilities: vga_controller bus_master cap_list rom
   configuration: driver=nvidia latency=0
   resources: irq:30-c0ffffff memory:b0000000-bffffff memory:c0000000-c1ffffff
*-display
   description: VGA compatible controller
   product: NVIDIA Corporation
   vendor: NVIDIA Corporation
   physical id: 0
   bus info: pci@0000:04:00:0

```

```

user@user-desktop: ~
user@user-desktop:~$ nvidia-smi
Thu Nov  7 11:52:07 2024

+-----+
| NVIDIA-SMI 550.120                  Driver Version: 550.120          CUDA Version: 12.4          |
+-----+-----+
| GPU   Name                               Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf              Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
|=====+=====+
|  0  NVIDIA GeForce RTX 4070 ...        Off | 00000000:05:00:0 On |         1%    Default |
|  0%   35C   P8              7W / 285W | 248MiB / 16376MiB |             MIG M. |
+-----+-----+
|  1  NVIDIA GeForce RTX 4060 Ti         Off | 00000000:04:00:0 Off |         0%    Default |
|  0%   34C   P8              4W / 165W |   9MiB / 16380MiB |             MIG M. |
+-----+-----+

+-----+
| Processes:                               GPU Memory                               |
|  GPU   GI    CI        PID   Type   Process name                        Usage                               |
|=====+=====+
|    0   N/A   N/A     2252      G   /usr/lib/xorg/Xorg                   127MiB                             |
|    0   N/A   N/A     2403      G   /usr/bin/gnome-shell                 113MiB                             |
|    1   N/A   N/A     2252      G   /usr/lib/xorg/Xorg                    4MiB                               |
+-----+

```

工作站安装和配置

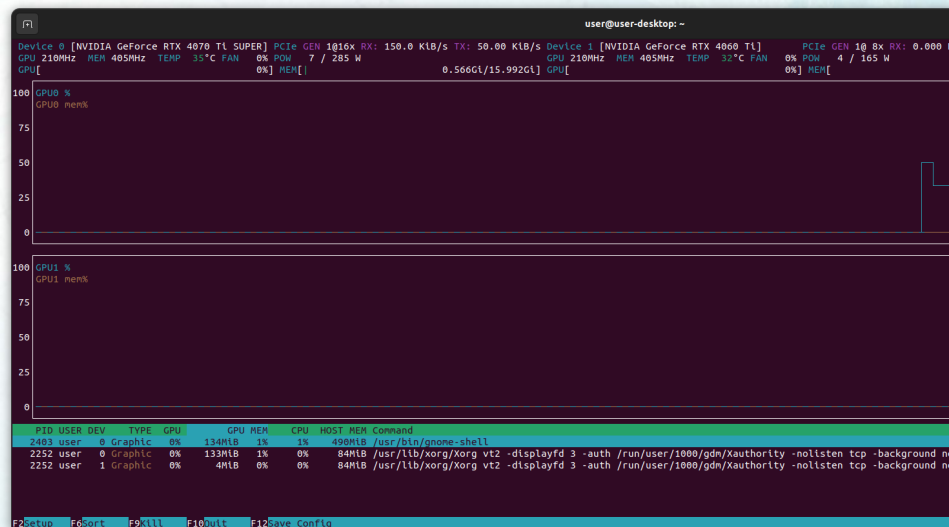
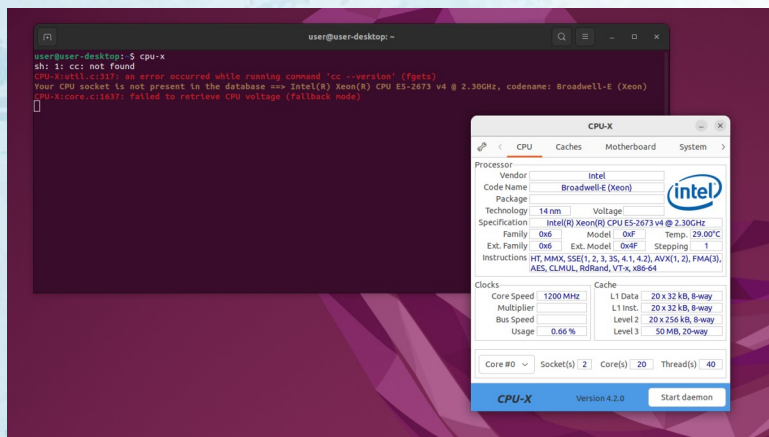
- 安装监察工具
- 运行指令 `sudo apt install cpu-x`
- 运行指令 `sudo apt install nvt`

```
user@user-desktop: ~  
user@user-desktop:~$ sudo apt install cpu-x  
Reading package lists... Done  
Building dependency tree... Done  
Reading state information... Done  
The following additional packages will be installed:  
  libc6  libc6-i386  libc6-udeb  libc6-udeb-i386  
Suggested packages:  
  libc6-dev  libc6-doc  libc6-i386-dev  
The following NEW packages will be installed:  
  cpu-x  libc6  libc6-i386  libc6-udeb  libc6-udeb-i386  
0 upgraded, 5 newly installed, 0 to remove and 0 not upgraded.  
Need to get 1,766 kB of archives.  
After this operation, 3,259 kB of additional disk space will be used.  
Do you want to continue? [Y/n] Y
```

```
user@user-desktop: ~  
user@user-desktop:~$ sudo apt install nvt  
Reading package lists... Done  
Building dependency tree... Done  
Reading state information... Done  
The following NEW packages will be installed:  
  nvt  
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.  
Need to get 43.9 kB of archives.  
After this operation, 106 kB of additional disk space will be used.  
Get:1 http://cn.archive.ubuntu.com/ubuntu jammy/multiverse amd64 nvt amd64 1.2.2-1 [43.9 kB]  
Fetched 43.9 kB in 0s (1,400 kB/s)  
Selecting previously unselected package nvt.  
(Reading database ... 203693 files and directories currently installed.)  
Preparing to unpack .../nvt_1.2.2-1_amd64.deb ...  
Unpacking nvt (1.2.2-1) ...  
Setting up nvt (1.2.2-1) ...  
Processing triggers for man-db (2.10.2-1) ...
```


工作站安装和配置

- 运行指令 `cpu-x` 会弹出一个弹窗
- 运行指令 `nvt` 会直接在 Terminal 里显示资讯，按 Q 键退出



AI 工具概览

AI 工具概览

- Ollama 是一款可以本地端运行大型语言模型服务
- Pinokio 是一款整合工具，整合了几乎所有开源的 AI 工具
 - 简单开始使用 AI 工具



AI 工具概览

- 安装 Ollama
- 运行指令 `sudo apt install curl`
 - 安装 curl 指令包
- 运行指令 `curl -fsSL https://ollama.com/install.sh | sh`
 - 安装 Ollama

```
user@user-desktop: ~  
user@user-desktop:~$ sudo apt install curl
```

```
user@user-desktop: ~  
user@user-desktop:~$ curl -fsSL https://ollama.com/install.sh | sh  
>>> Installing ollama to /usr/local  
>>> Downloading Linux amd64 bundle  
##### 14.3%
```

```
user@user-desktop: ~  
user@user-desktop:~$ curl -fsSL https://ollama.com/install.sh | sh  
>>> Installing ollama to /usr/local  
>>> Downloading Linux amd64 bundle  
##### 100.0%  
>>> Creating ollama user...  
>>> Adding ollama user to render group...  
>>> Adding ollama user to video group...  
>>> Adding current user to ollama group...  
>>> Creating ollama systemd service...  
>>> Enabling and starting ollama service...  
Created symlink /etc/systemd/system/default.target.wants/ollama.service → /etc/systemd/system/ollama.service.  
>>> NVIDIA GPU installed.  
user@user-desktop:~$
```

AI 工具概览

- 运行指令 `ollama list`
 - 去查看已经获取的模型
- 运行指令 `ollama pull llama3.2`
 - 去获取模型储存到本地
- 可以前往官网查看模型列表
 - <https://ollama.com/library>

```
user@user-desktop: ~  
user@user-desktop:~$ ollama list  
NAME      ID      SIZE      MODIFIED  
user@user-desktop:~$ ollama pull llama3.2  
pulling manifest  
pulling dde5aa3fc5ff... 100% [REDACTED] 2.0 GB  
pulling 966de95ca8a6... 100% [REDACTED] 1.4 KB  
pulling fcc5a6bec9da... 100% [REDACTED] 7.7 KB  
pulling a70ff7e570d9... 100% [REDACTED] 6.0 KB  
pulling 56bb8bd477a5... 100% [REDACTED] 96 B  
pulling 34bb5ab01051... 100% [REDACTED] 561 B  
verifying sha256 digest  
writing manifest  
success  
user@user-desktop:~$ ollama pull qwen2.5  
pulling manifest  
pulling 2bada8a74506... 100% [REDACTED] 4.7 GB  
pulling 66b9ea09bd5b... 100% [REDACTED] 68 B  
pulling eb4402837c78... 100% [REDACTED] 1.5 KB  
pulling 832dd9e00a68... 100% [REDACTED] 11 KB  
pulling 2f15b3218f05... 100% [REDACTED] 487 B  
verifying sha256 digest  
writing manifest  
success  
user@user-desktop:~$ ollama list  
NAME      ID      SIZE      MODIFIED  
qwen2.5:latest 845dbda0ea48 4.7 GB 9 seconds ago  
llama3.2:latest a80c4f17acd5 2.0 GB About a minute ago  
user@user-desktop:~$
```

AI 工具概览

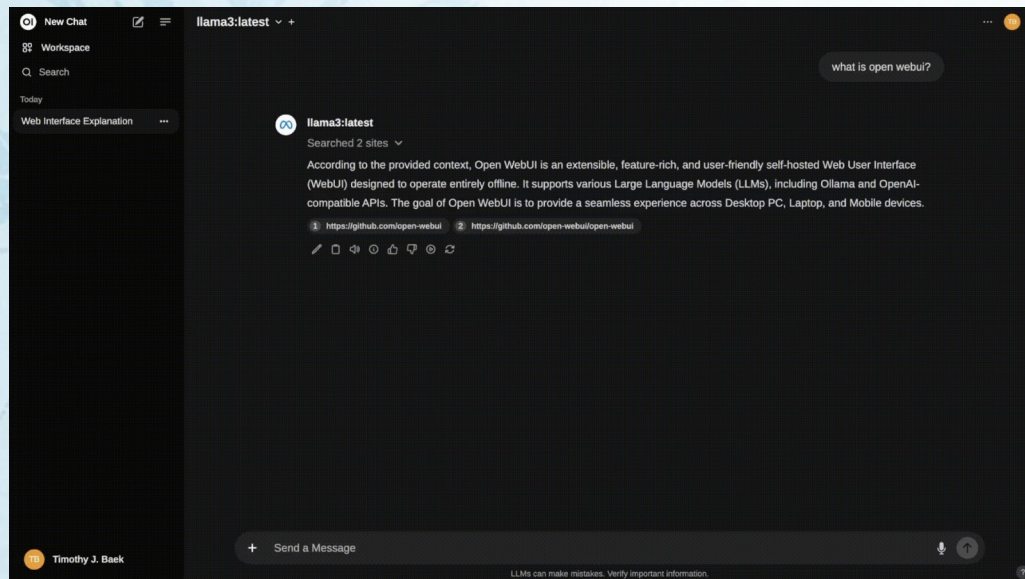
- 可以运行指令去提问题
- 结果如下图:

```
top:~$ curl http://localhost:11434/api/generate -d '{ "model": "llama3.2", "prompt": "Why is the sky blue?" }'
```

```
user@krystal-desktop: ~
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.418965354Z", "response": " the", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.429153995Z", "response": " Earth", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.436971153Z", "response": " 's", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.444789838Z", "response": " atmosphere", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.452624715Z", "response": " makes", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.461396502Z", "response": " the", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.469822369Z", "response": " sky", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.477674357Z", "response": " appear", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.485547259Z", "response": " blue", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.493424675Z", "response": " to", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.502476949Z", "response": " our", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.510326101Z", "response": " eyes", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.518112331Z", "response": " ", "done": false}
{"model": "llama3.2", "created_at": "2024-11-12T03:41:37.526012932Z", "response": "", "done": true, "done_reason":
":stop", "context": [128006, 9125, 128007, 271, 38766, 1303, 33025, 2696, 25, 6790, 220, 2366, 18, 271, 128009, 128006, 882
, 128007, 271, 10445, 374, 279, 13180, 6437, 30, 128009, 128006, 78191, 128007, 271, 791, 13180, 8111, 6437, 1606, 315, 264, 2
5885, 2663, 13558, 64069, 72916, 11, 7086, 1306, 279, 8013, 83323, 10425, 13558, 64069, 11, 889, 1176, 7633, 433, 304, 279, 33
89, 220, 777, 339, 9478, 382, 8586, 596, 1148, 8741, 1473, 16, 13, 3277, 40120, 29933, 9420, 596, 16975, 11, 433, 35006, 13987,
35715, 315, 45612, 1778, 439, 47503, 320, 45, 17, 8, 323, 24463, 320, 46, 17, 4390, 17, 13, 4314, 35715, 45577, 279, 3177, 304, 6
82, 18445, 11, 719, 814, 45577, 24210, 320, 12481, 8, 93959, 810, 1109, 5129, 320, 1171, 8, 93959, 627, 18, 13, 1115, 374, 1606,
279, 9333, 35715, 527, 810, 7524, 520, 72916, 279, 24210, 93959, 4245, 311, 872, 9333, 1404, 323, 7191, 7479, 3158, 8844, 311,
872, 3148, 627, 19, 13, 1666, 264, 1121, 11, 279, 6437, 3177, 374, 38067, 304, 682, 18445, 323, 25501, 1057, 6548, 505, 682, 559
6, 315, 279, 13180, 11, 3339, 433, 5101, 6437, 627, 20, 13, 578, 5129, 93959, 320, 1171, 11, 19087, 11, 323, 14071, 8, 3136, 311,
5944, 304, 264, 7833, 1584, 323, 5662, 1057, 6548, 6089, 505, 279, 5216, 315, 279, 7160, 11, 902, 374, 3249, 584, 1518, 1124, 43
9, 1694, 12401, 311, 279, 35174, 382, 4516, 11, 279, 10824, 315, 13558, 64069, 72916, 323, 279, 9420, 596, 16975, 3727, 279, 13
180, 5101, 6437, 311, 1057, 6548, 13], "total_duration": 4144921098, "load_duration": 2014614011, "prompt_eval_count
": 31, "prompt_eval_duration": 176000000, "eval_count": 222, "eval_duration": 1951000000}
user@krystal-desktop: $
```


AI 工具概览

- Open WebUI 是一个为大型语言模型设计的友好型 Web 用户界面



How to Install 🚀

Installation via Python pip 🐍

Open WebUI can be installed using pip, the Python package installer. Before proceeding, ensure you're using Python 3.11 to avoid compatibility issues.

1. Install Open WebUI: Open your terminal and run the following command to install Open WebUI:

```
pip install open-webui
```

2. Running Open WebUI: After installation, you can start Open WebUI by executing:

```
open-webui serve
```

This will start the Open WebUI server, which you can access at <http://localhost:8080>

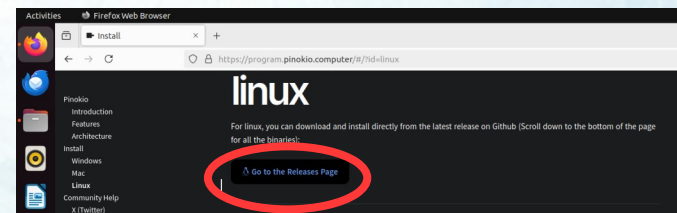
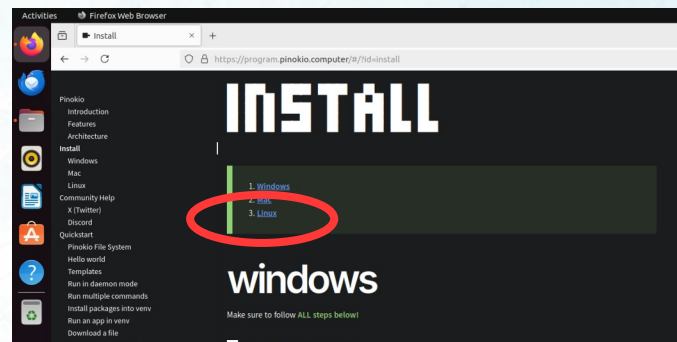
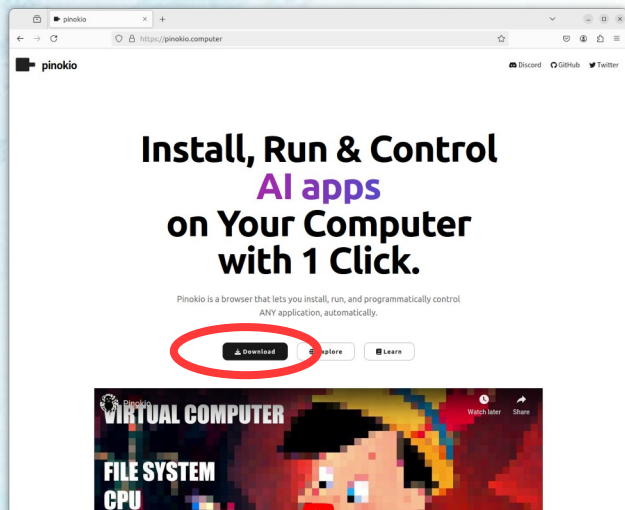
AI 工具概览

- Pinokio 是一款整合工具，整合了几乎所有市面上开源的 AI 工具
 - 简单开始使用 AI 工具
 - 一键安装



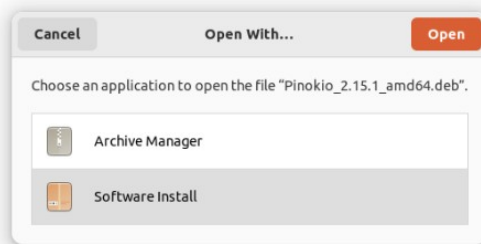
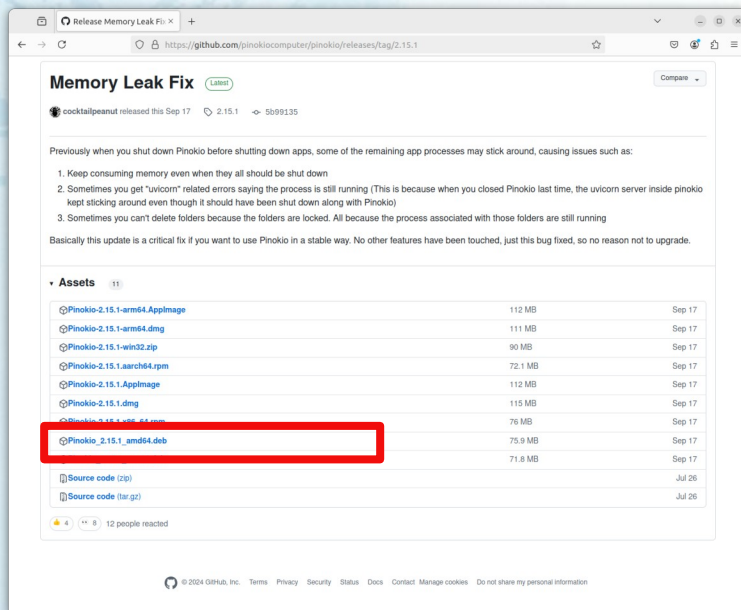
AI 工具概览

- 前往 Pinokio 官网 <https://pinokio.computer>
- 点击 Download
- 选择 Linux 并前往 GitHub 的页面



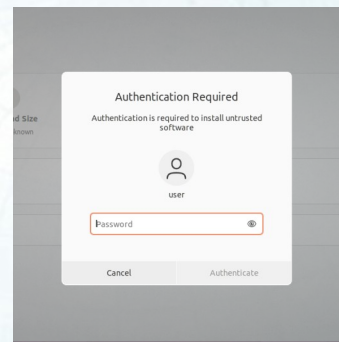
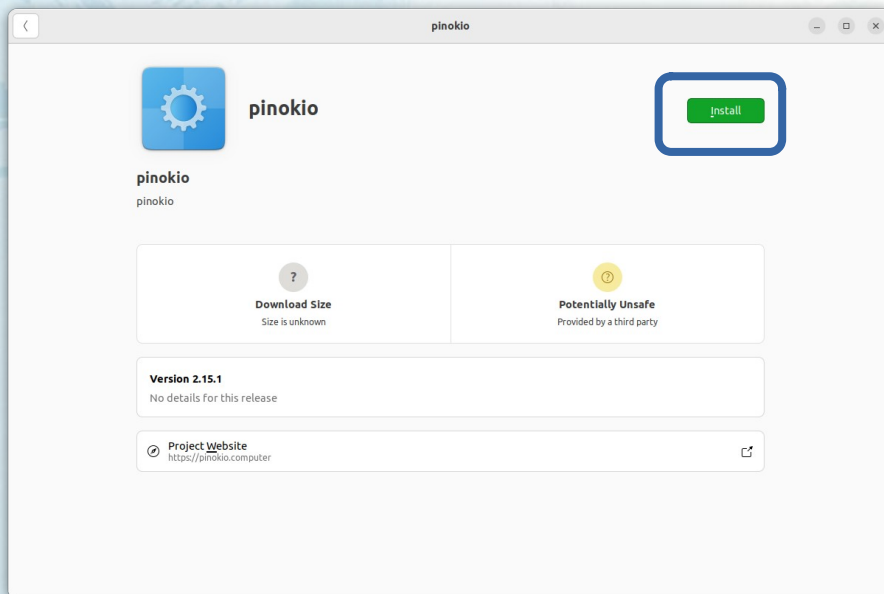
AI 工具概览

- 选择下载 Pinokio_2.15.1_amd64.deb
- 点击下载好的档案，然后选择 Software Install



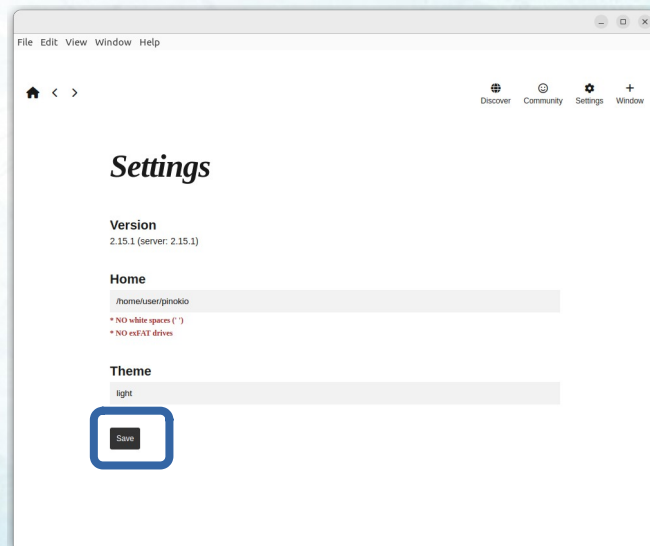
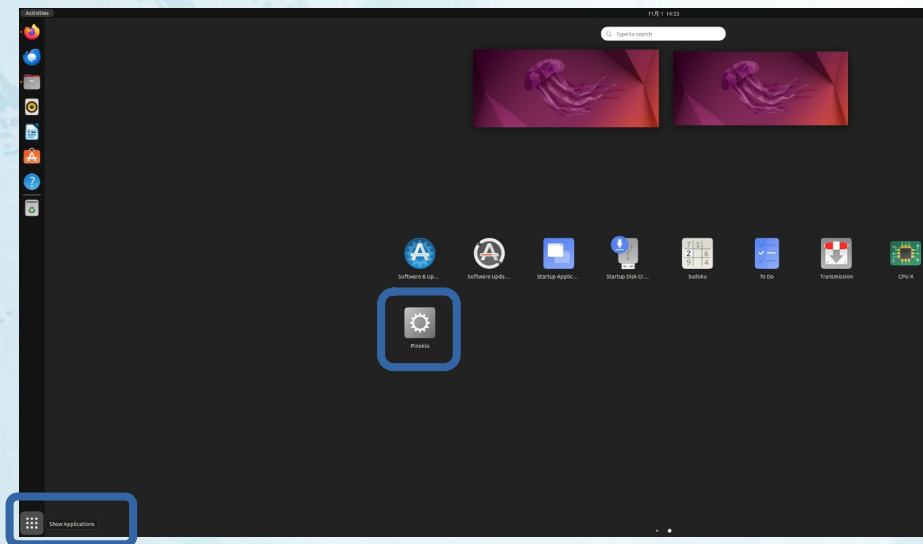
AI 工具概览

- 在 Pinokio 的安装画面，点击 Install
- 输入当前用户的密码



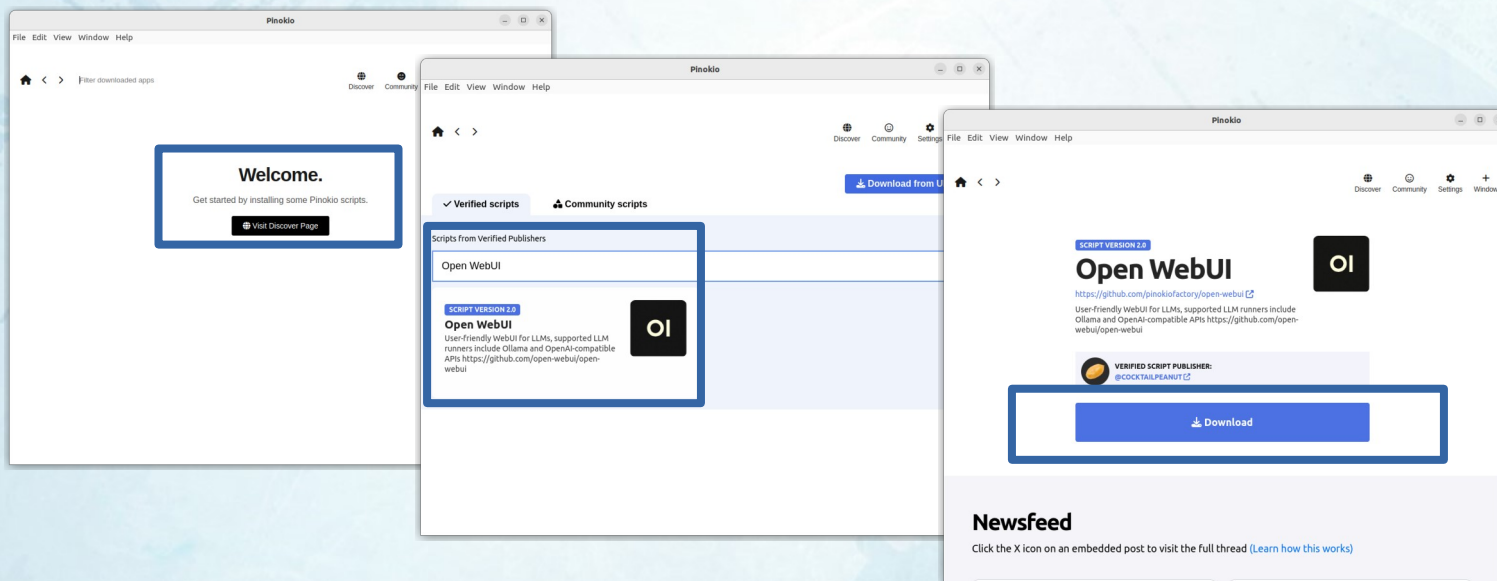
AI 工具概览

- 点击左下角的 Show Applications，可以看见 Pinokio 已经安装了
- 第一次开启 Pinokio 需要确认设定，直接使用预设的就可以



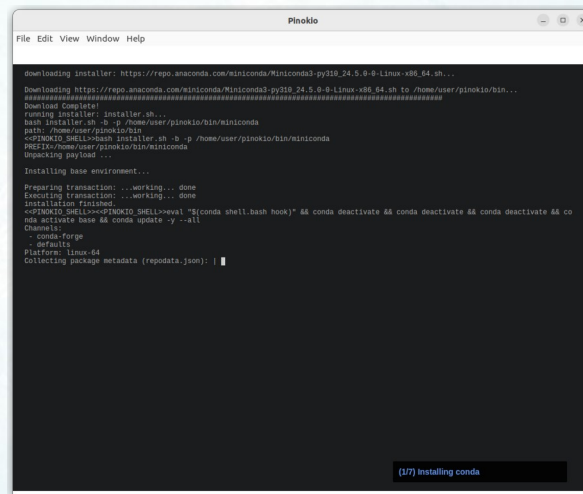
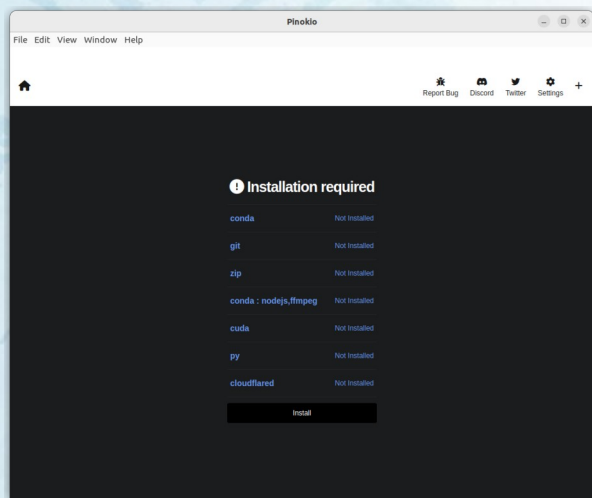
AI 工具概览

- 点击 **Discover** 前往搜索 Open WebUI
- 点击 **Download** 开始下载



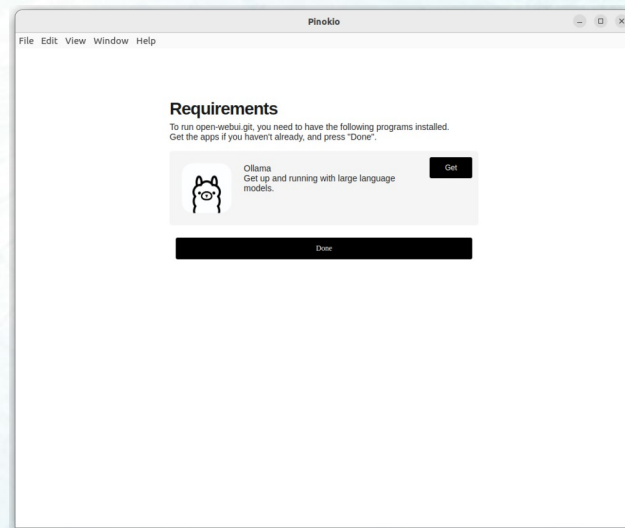
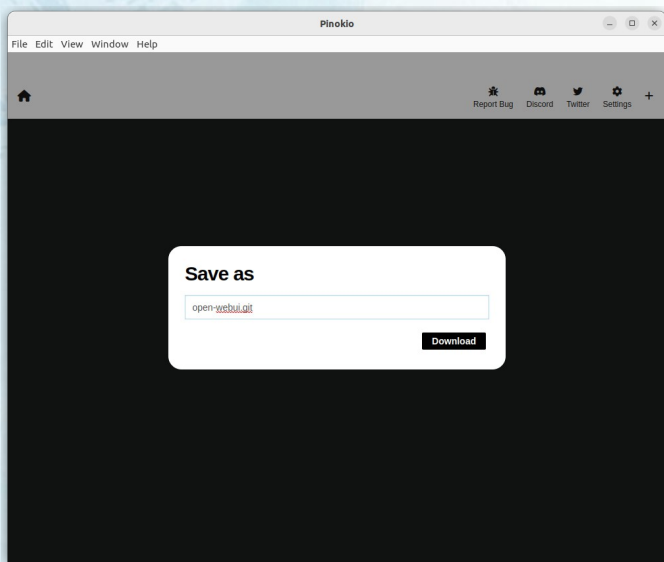
AI 工具概览

- 有些必须的前置工具，直接按 **Download** 开始下载



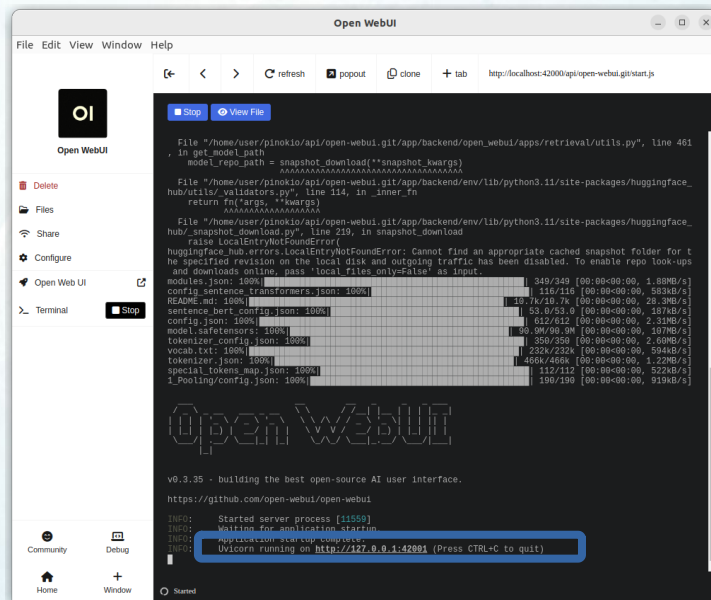
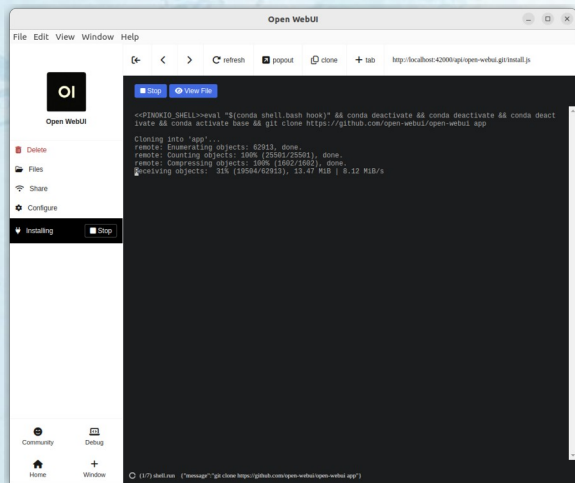
AI 工具概览

- 前置工具安装好后，弹出弹窗请求从 **github** 下载安装档案，点选 **Download**
- 下一步会提醒需要先装 **Ollama**，点选 **Done** 继续下载



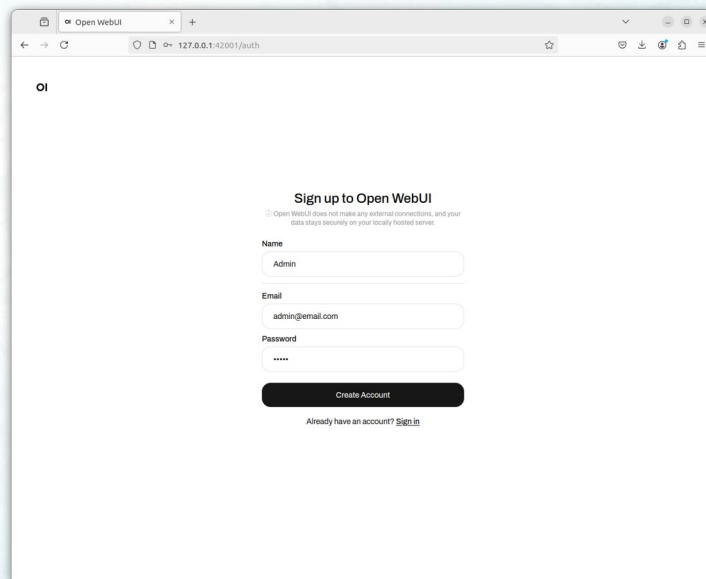
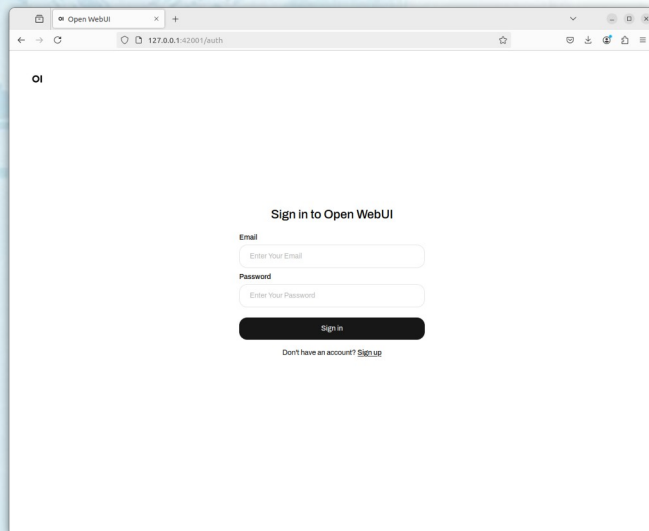
AI 工具概览

- 等到安装完成，可以见到 Open WebUI 的 ASCII 图案
- 点击里面的链接前往 Open WebUI 网页



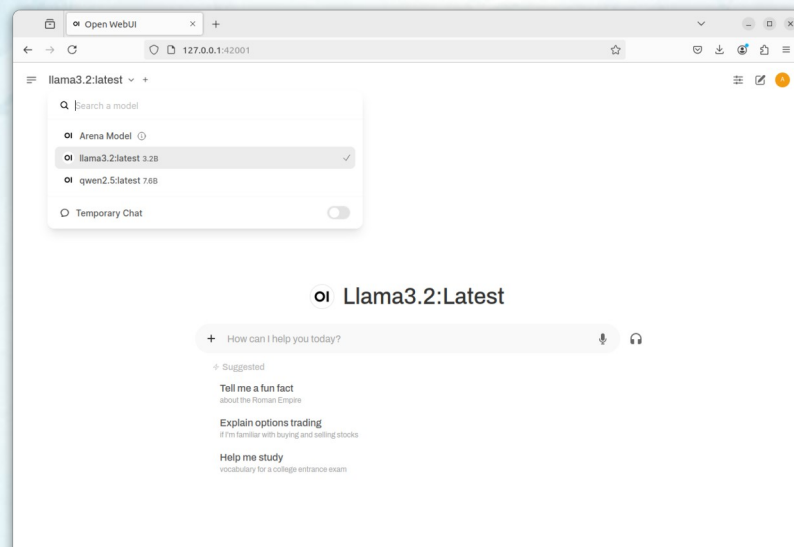
AI 工具概览

- 初次设定需要创建一个管理者账户，点击下方的 Sign up
- 随意填写 email 和 password



AI 工具概览

- 登入后可以在左上角选择已有的模型，在下方对话框
- 还可以按右方的 + 号去获取其他模型



AI 工具概览

- Open WebUI 的用户管理
- 新增用户需要管理者的确认

01

Sign up to Open WebUI

Open WebUI does not make any external connections, and your data stays securely on your locally hosted server.

Name

NewUser

Email

newuser@email.com

Password

Create Account

Already have an account? [Sign in](#)

Account Activation Pending

Contact Admin for WebUI Access

Your account status is currently pending activation. To access the WebUI, please reach out to the administrator. Admins can manage user statuses from the Admin Panel.

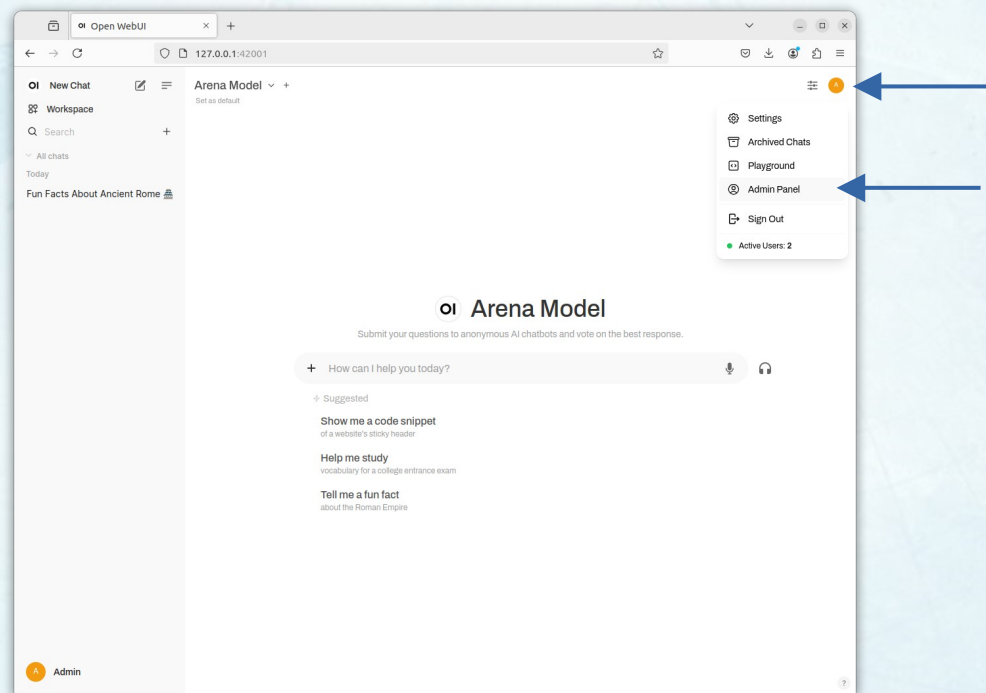
Admin: Admin (admin@email.com)

Check Again

[Sign Out](#)

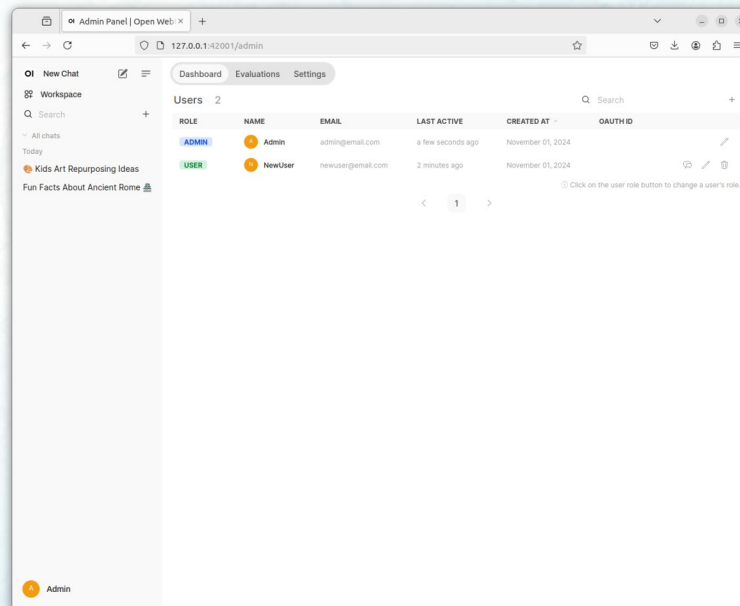
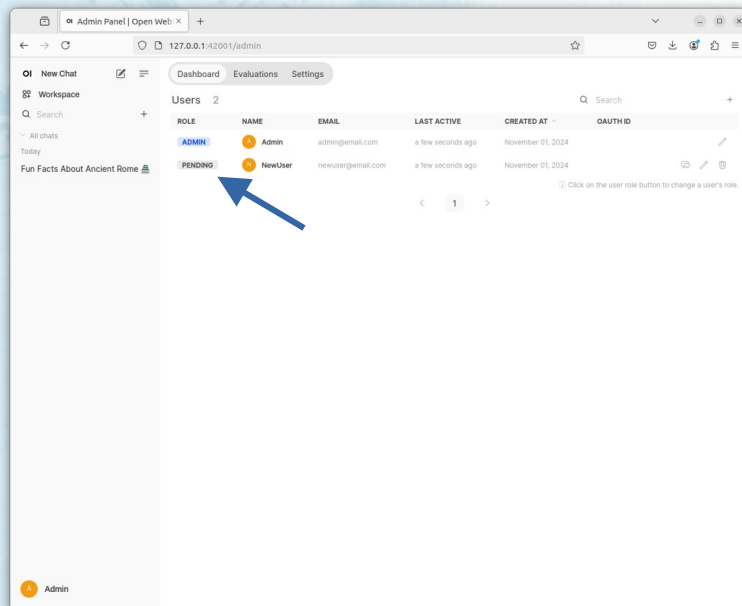
AI 工具概览

- 需要管理者登入系统，点击右上角头像，在下拉列表里按 **Admin Panel**



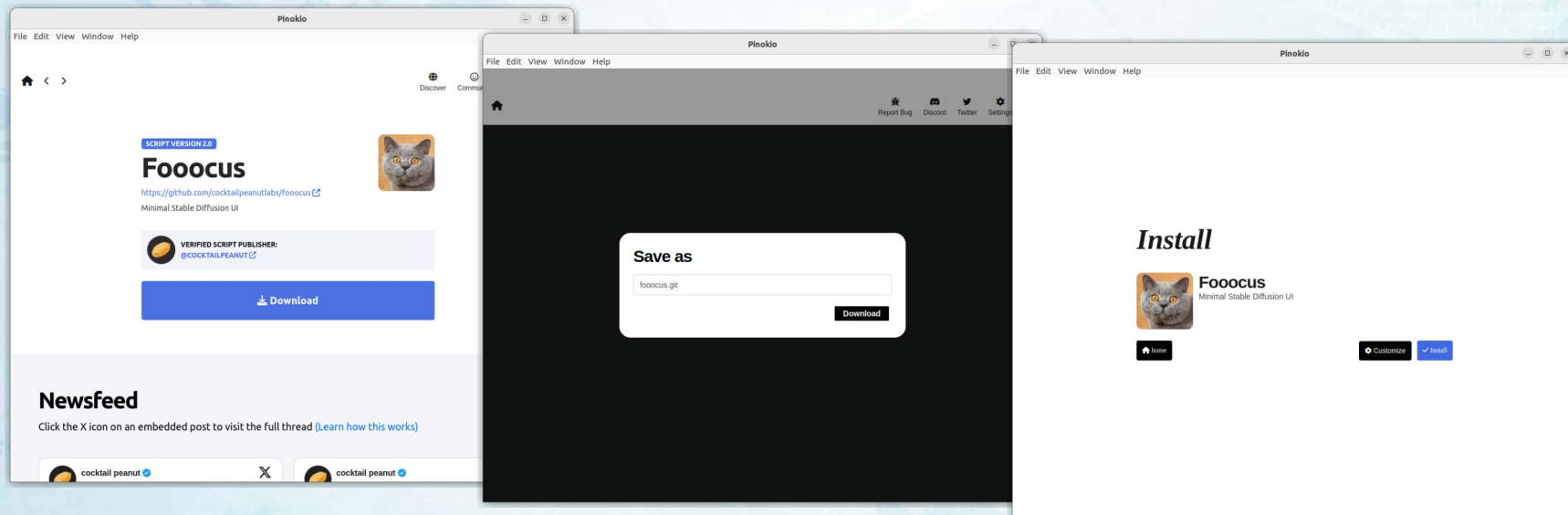
AI 工具概览

- 可以见到 NewUser 的角色，现在是 PENDING (待处理)
- 直接点击 PENDING，转变角色成 USER



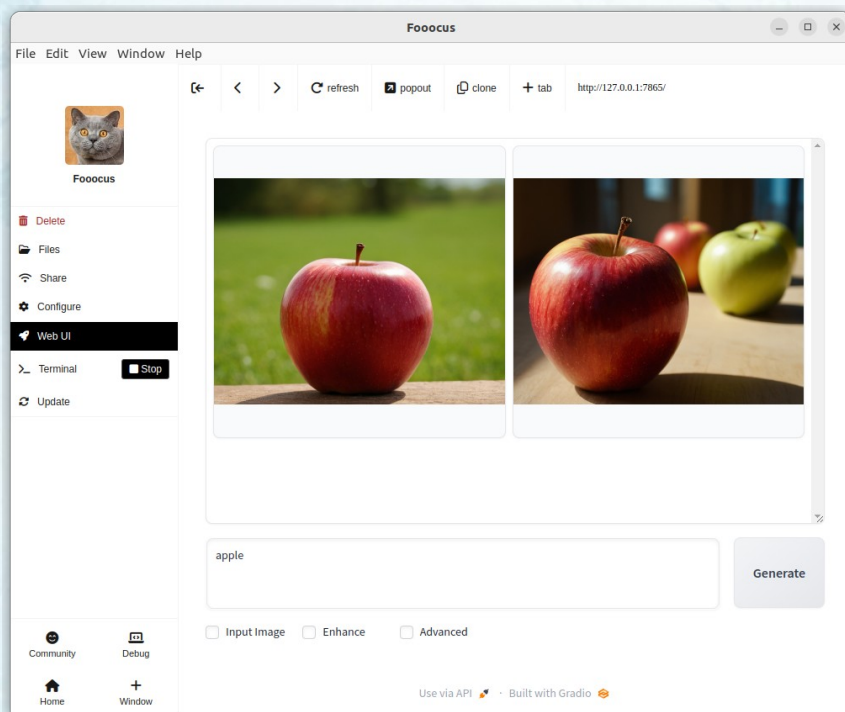
AI 工具概览

- 尝试安装更多不同的工具 foocus
 - Stable Diffusion 图像生成工具



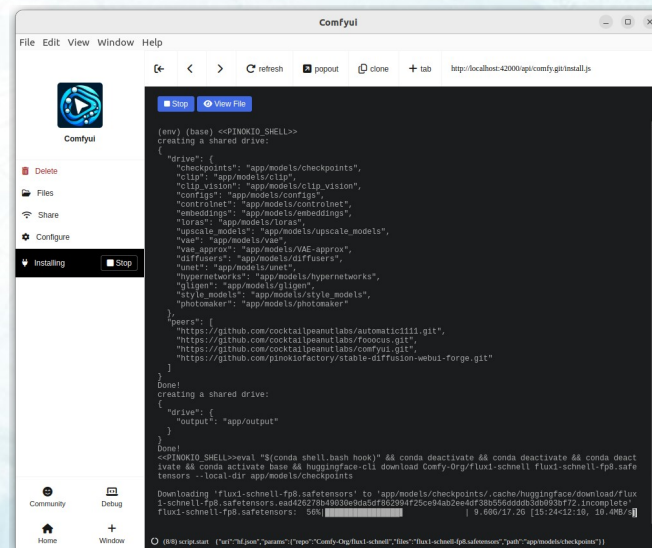
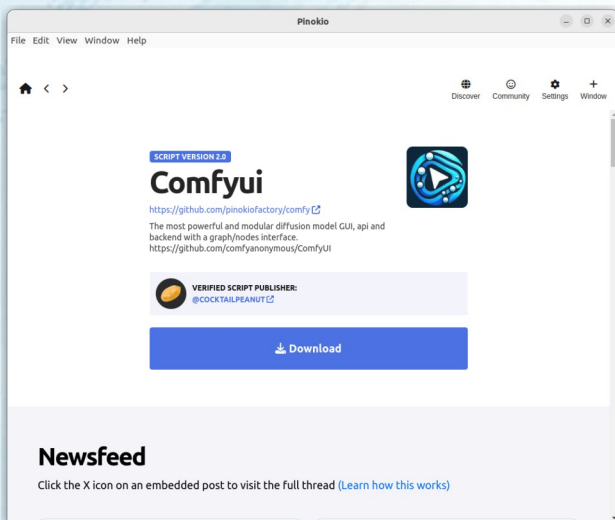
AI 工具概览

- 直接在下方输入文字，按右方按钮开始生成



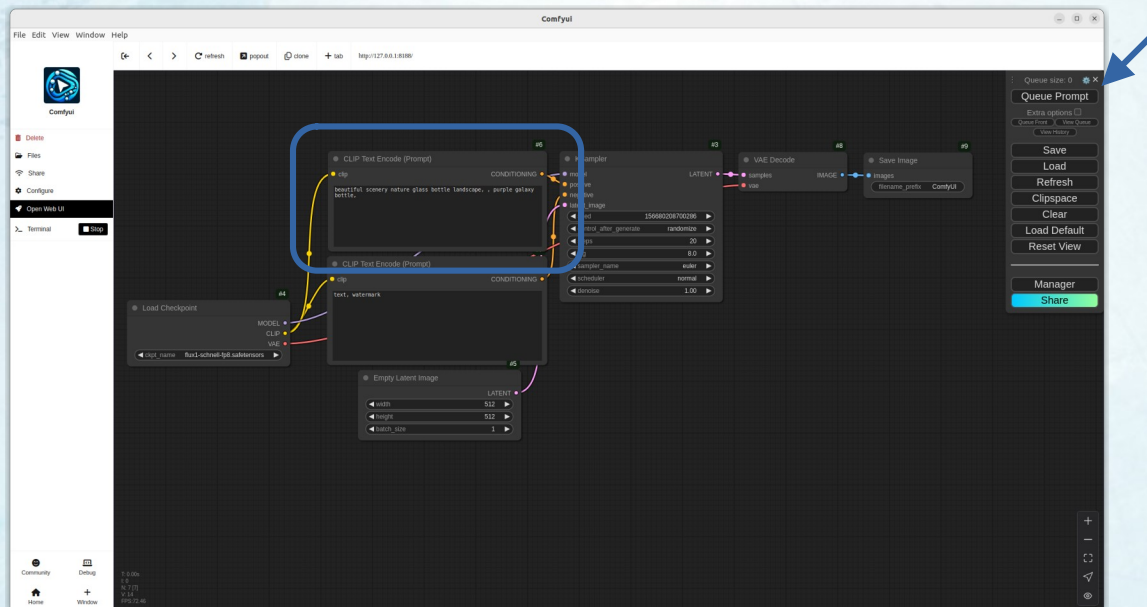
AI 工具概览

- 尝试安装更多不同的工具 ComfyUI
 - FLUX.1 图像生成工具



AI 工具概览

- 安装完成后会见到这个画面
- 在蓝色区域改变文字描述，然后按右上角的 Queue Prompt 开始生成图片



系统维护和优化

系统维护和优化

- 保持系统更新以确保拥有最新的安全补丁
 - `sudo apt update`
 - `sudo apt upgrade`
- 用 `nvidia-smi` 指令查看 GPU 用量
 - 考虑升级配置（更换更高阶的 GPU 卡）

结语

结语

- 硬件组装
- 安装作业系统
- 安装 AI 工具
- 维护工作站

谢谢各位