HOW TO CREATE A PERSONAL AI SERVER USING NVIDiA NANO – Installation & setup instructions

Nvidia Orin Nano setup - Ubuntu Jetson OS Install, Ollama AI Server install, Open Web UI Install

Context: What is this document purpose?

These document will provide a complete end to end and consilidated instructions and guidance- provide steps, examples, options, screens & details. The document is created on a Windows laptop and Windows laptop is used for flusing the Ubuntu Jetson OS image flushed to the miniSD card; similar or echivalent instructions will apply when an Ubuntu laptop is used to Flush the Ubuntu Jetson OS image on the miniSD card (or USB), or Mac OS laptop is used to Flush the image on miniSD card (or USB).

The process is complex and requires a lot of time and skills, consolidating various pieces of information and completing actions of firrenent nature. This document intent is to easily avoi d known blockers and challanges.

Related information: *On some particular sections this document will utilize installation steps & documentation from **Nvidia AI LAB tutorial** and might be considered an **extention** of these original documents & instructions:*

Nvidia official documentation:

- https://www.jetson-ai-lab.com/initial_setup_jon.html
- https://www.jetson-ai-lab.com/tutorial_ollama.html
- https://www.jetson-ai-lab.com/tutorial_openwebui.html

YouTube videos:

- **NVIDIA Jetson Orin Nano Super COMPLETE Setup Guide & Tutorial**: https://www.youtube.com/watch?v=-PjMC0gyH9s
- **3 Minute Fix for Chromium and other Snaps not launching**: https://www.youtube.com/watch?v=x6bccF3xtRE&t=79s
- **Use These! Jetson Docker Containers Tutorial**: https://www.youtube.com/watch?v=HlH3QkS1F5Y

Objective: What are we trying to achieve?

This document is a detailed step by step documentation & installation instructions on how to setup a **Home Personal AI Server** by using a low cost dedicated **Nvidia Jetson Orin Nano server.** To create a **Personal AI Server** we will use also **Ollama AI server**, **Open Web UI server**, **Jetson Containers** deployed & confugured on **Jetson Orin Nano.**

What is the Nvidia Jetson Orin Nano? https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/product-development/

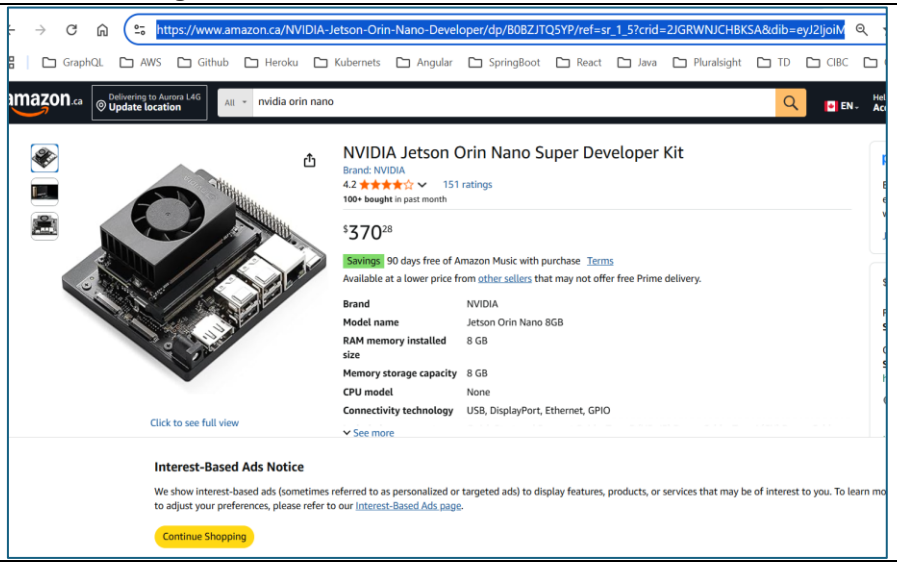
Required Hardware & Software: What do we need?

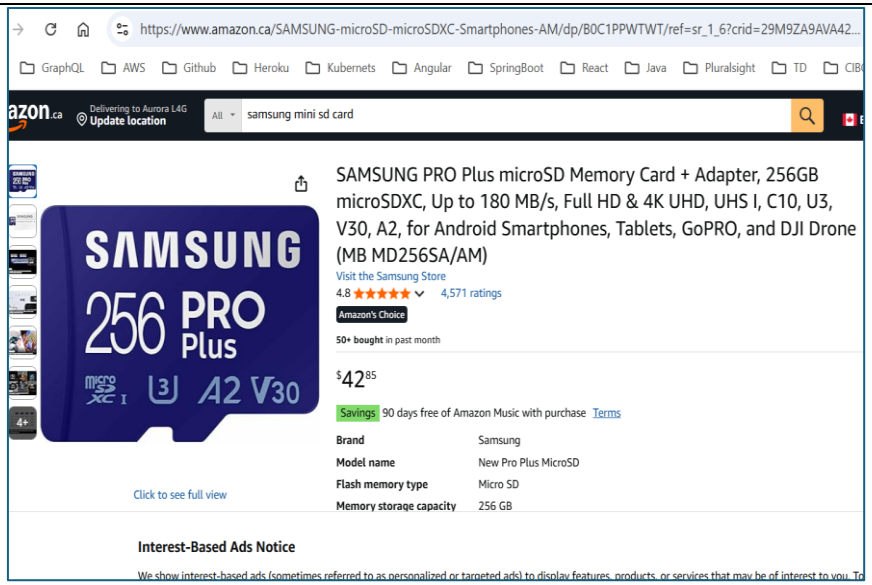
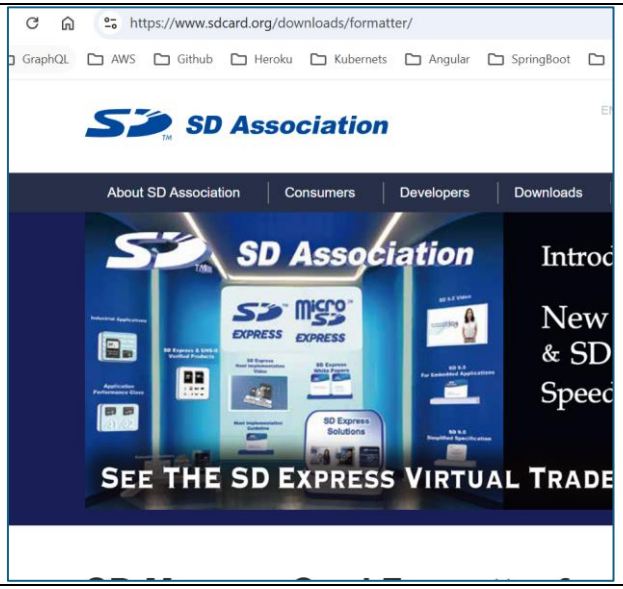
1. Nvidia Jetson Orin Nano server - Super 8 Gb type
2. Nvidia Ubuntu Jetson bootable OS Image install, Ollama AI Server, Open Web UI
3. SD Card, SD formatting App, Balena Etcher App to flash install on miniSD (USB), ZIP App
4. Keyboard, mouse, Monitor, Ethernet Internet connection, Ethernet cable, DP Video connection cable or DP to HMDI adapter, Orin Nano power adapter
5. Laptop, Internet connection, access to download software, access to releated Web documentation

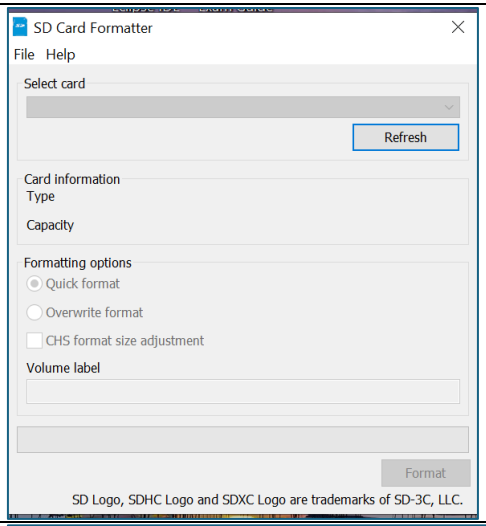
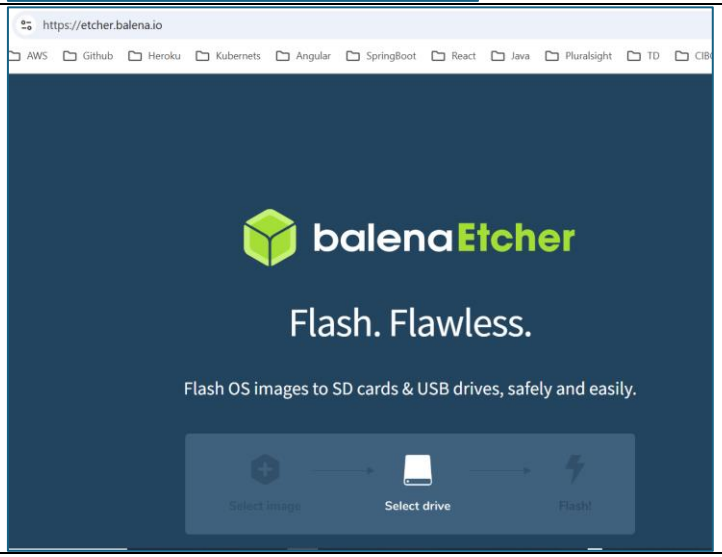
High level installation steps: High level what are we doing?

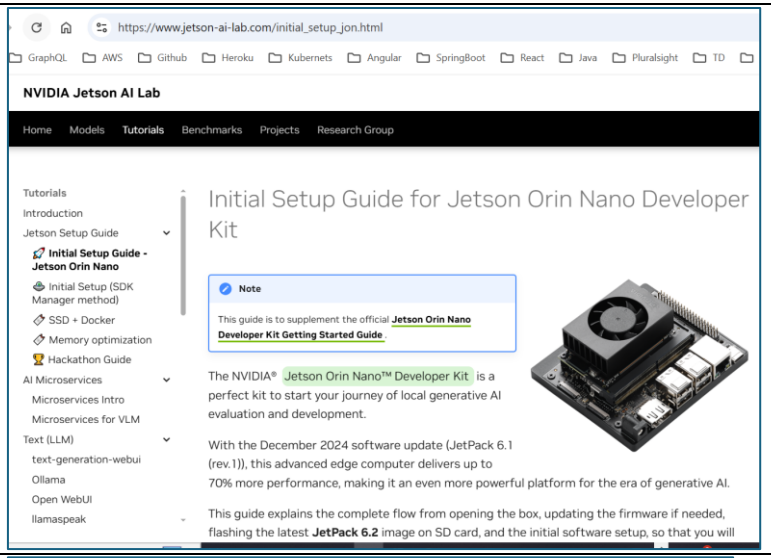
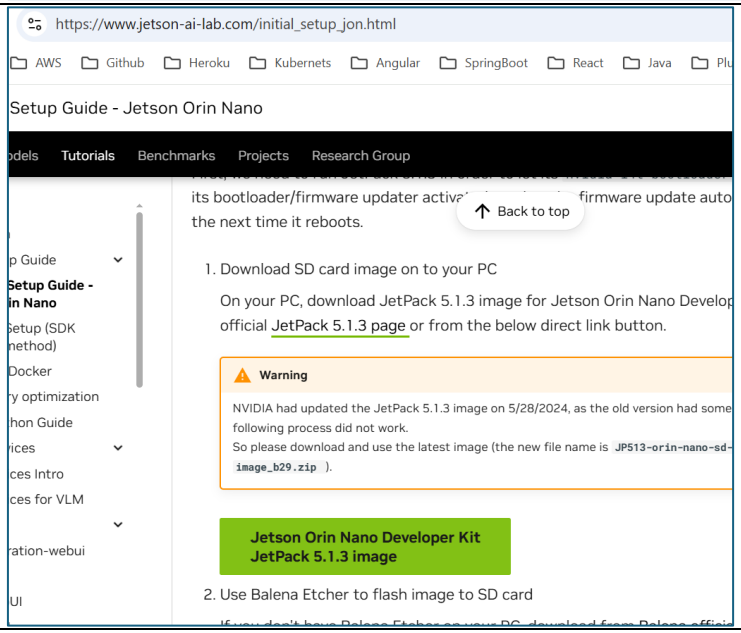
1. Create a Ubuntu Jetson OS image on mini SD card and install on the Jetson Nano device
2. Install Jetson Docker containers, install Ollama using Docker, load LLM model, start Ollama & Client Prompt AI, install Open Web UI using Docker
3. Use Open Web UI AI prompt, pointing to Ollama AI server hosted on local Nano device; have private AI server at very affordable proce (USD $ 250)

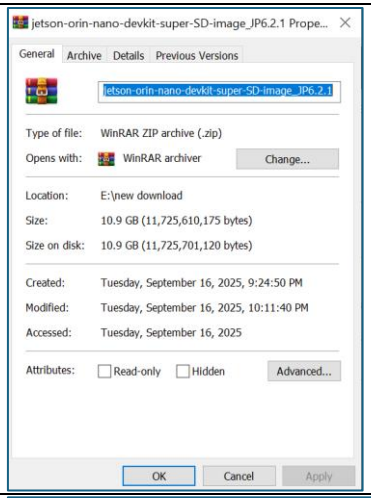
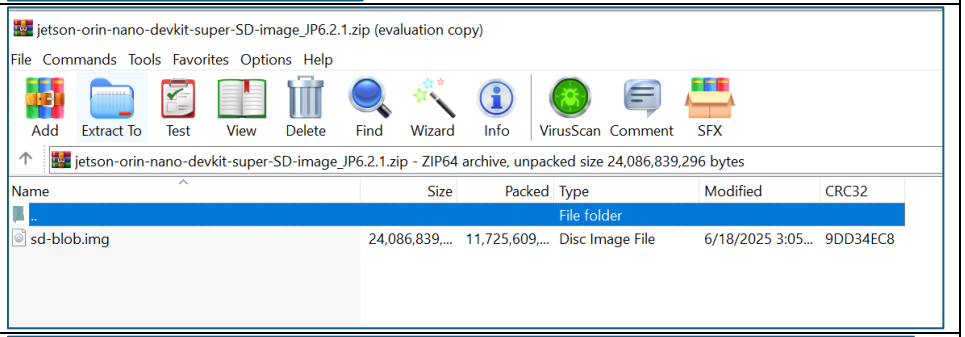
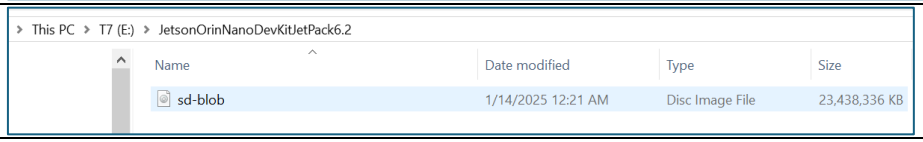
Detailed level installation steps: What are the detailed steps?

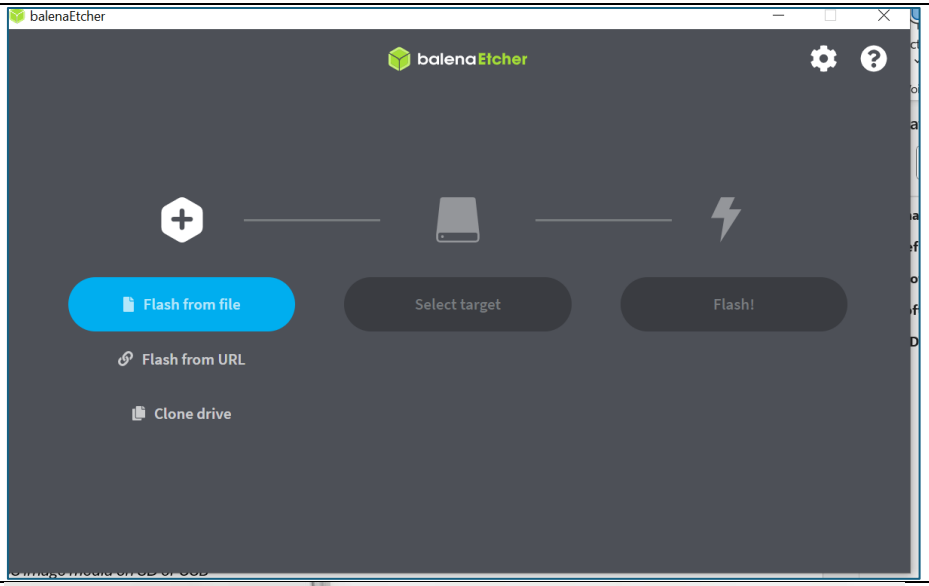
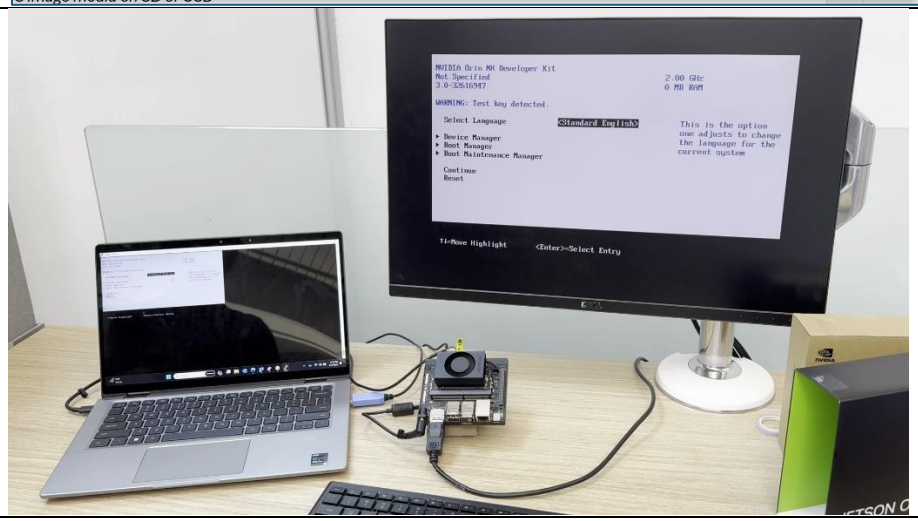
	Step	Details or Image
1	*Buy & get the **Nvidia Jetson Orin Nano server – Super 8 Gb RAM** – http://amazon.ca* *Observation: get Super versiuon with 8 Gb RAM*	

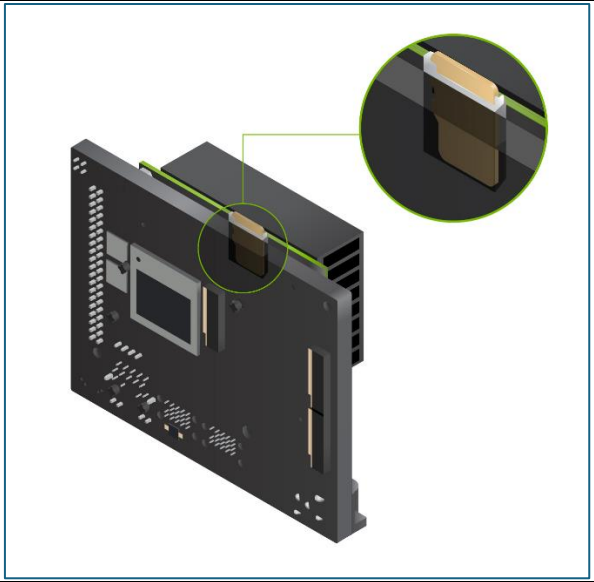
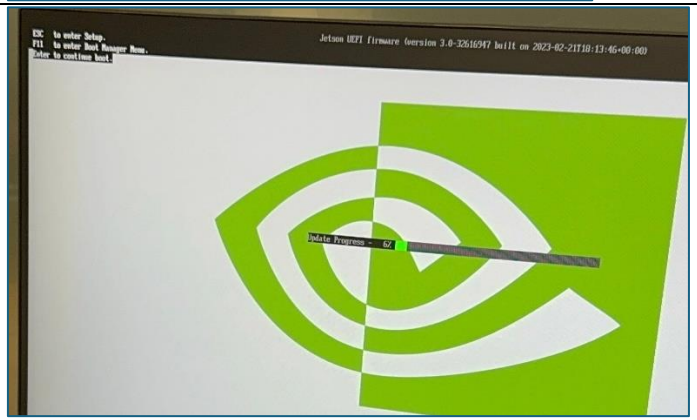
2	*Buy & get the **mini XD SD card** http://amazon.ca* • *Samsung mini SD Card – min 64 Gb or 256 GB (highly recommended),* • *Bigger the XD card size the better - even 512 GB, 1 TB* *Observation: You might need an adapter for miniSD card as well; tha can be inserted in your laptop USB or into a laptop card port (or laptop mini car dport if available)*	
3	*Download & Install the **SD Card Formatter Application** from https://www.sdcard.org/downloads/formatter/*	

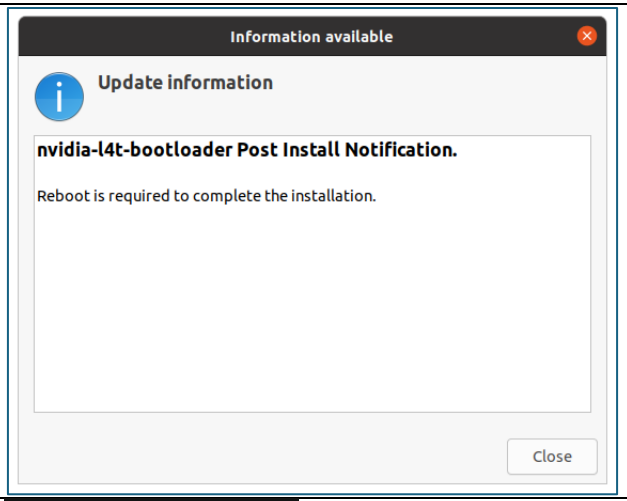
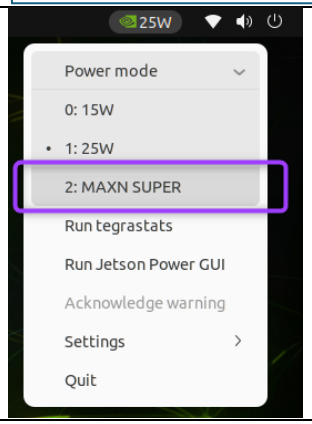
4	*Insert the mini SD card into the Windows laptop and start the **SD Card Formatter Application,** then format the mini XD SD Card* *Observation: this will take few seconds*	
5	*Download & Install the **Balena Etcher Application** from:* *https://etcher.balena.io/* *to Flash Ubuntu Jetson OS Image into the mini XD SD Card – this will be bootable OS installer for Unbuntu Jetson OS*	

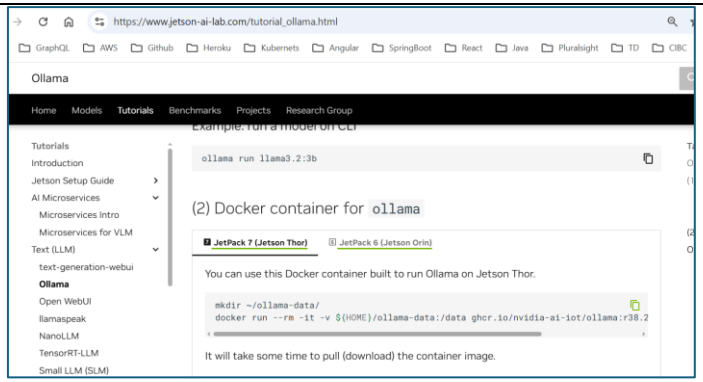
6	*Download **Ubuntu Jetson OS Image Archive installer** from* *login to Nvidia DEV zone with user ID and passoword from the laptop* *https://www.jetson-ai-lab.com/* *go to NVIDIA AI LAB Tutorial: https://www.jetson-ai-lab.com/initial_setup_jon.html* ***Observation**: current Nvidia Orin Nano install guide will contain different installation notes with various options and download locations.* *AI LAB tutorial link Jetson download is the only Orin Nano Jetson bootable installer image, that actually will allow you to setup Ubunto OS on the Nvidia Orin Nano. The other download locations * images are not bootable, cannot really be used successfully. Unfortunately.*	
7	*Press **Download** button in the **Jetson Orin Dev Kit Jetpack 5.1** image* *Page location NVIDIA AI LAB Tutorial: https://www.jetson-ai-lab.com/initial_setup_jon.html* *Save the JAR to local laptop drive* *Observation: 11Gb JAR archive download can take up to 10 min*	

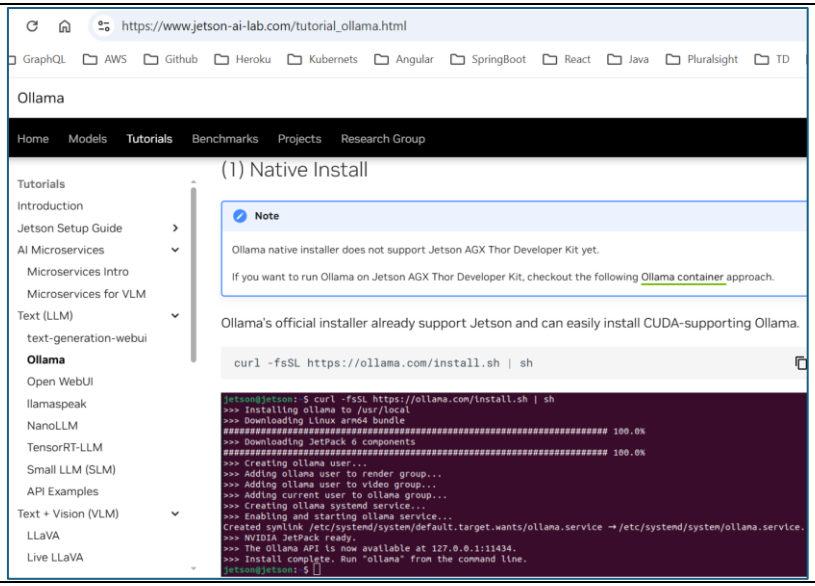
8	*Make sure you download this JAR:* ***jetson-orin-nano-devkit-super-SD-image_JP6.2.1.jar*** ***Observation**: current Nvidia Orin Nano install guide will contain different installation notes with various options and download Ubunto Jetson software locations.* *The installation is not clear and other download links will provided non-bootable OS installer downloads; NVIDIA AI LAB tutorial link Jetson download is the only Orin Nano Jetson bootable installer image, that actually will allow you to setup Ubunto OS on the Nvidia Orin Nano. The other download images are not bootable, cannot really be used successfully. Unfortunately.* *Observation: JAR size will be approximately 11 Gb*	
9	*Download Archive Application to unzip the JAR if necessary: WinRar, WinZIP, 7zip* *Unzip the Archive Image:* ***jetson-orin-nano-devkit-super-SD-image_JP6.2.1.jar*** *with WinRAR or other archive Application*	
10	*Ubuntu Jetson OS image should look like this:* ***sd-blob.img*** *Observation: sd-blob.dmg image size will be approximately 24 Gb*	

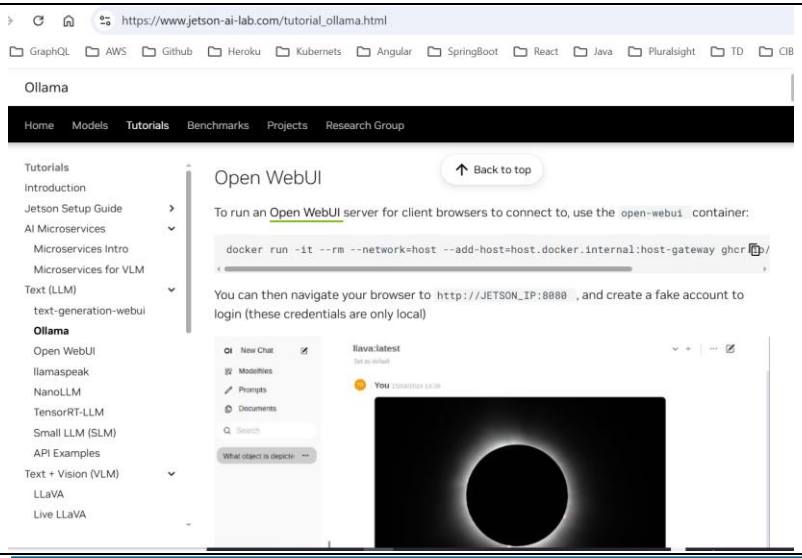
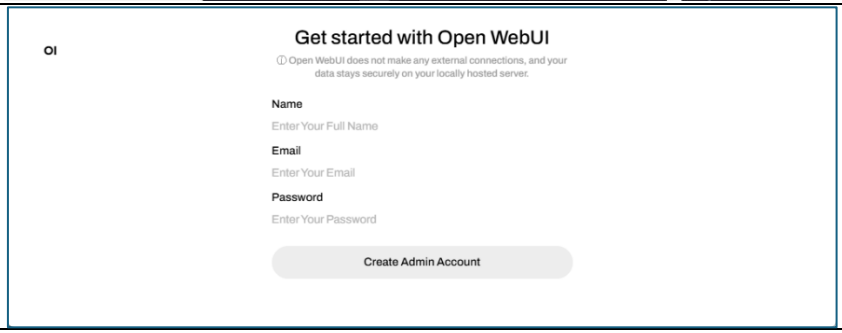
11	**FLASH the Ubuntu Jetson OS bootable image to the miniSD card** *Start the **Balena Etcher Application** and lookup the Ubuntu Jetson image **sd-blob.img** from your file system location; then press "Flash from File" **sd-blob.img** image file, then" Select Target " ad miniSDS card drive, and press* *Observations:* • *This process will flash the Ubuntu OS Jetson to the miniSD card* • *you can flash the image to USB stick as well* • *the process is very time consuming and can take up to 10 minutes, depending on the laptop* • *at the end the flash image miniSD Card size will be approximatelly 24 Gb*	
12	**Check Jetson Orin Nano boot capabilities** • Connect keyboard, mouse, Ethernet cable to Nano device • Connect Ethernet cable from Nano to the Internet Router or switch ***Power ON** Jetson Orin nano – plug the power cable, and when the boot program sequence is shown please check **the boot Jetson UEFI Firmware version is higher that 36.0*** *Observation:* • *Nano device doens't have an ON/OFF switch* • *UEFI boot application upgrade Instructions will be provided in future version of this document* • *if Nano doesn't have the UEFI Formware version higher than 36.0, then OS mini SD card image will **not boot; you are stuck and have to update the Firmware instead***	
12	**Power OFF** *Jetson Orin Nano device; plug out the power cable* *Observation: Nano device doens't have an ON/OFF switch*	

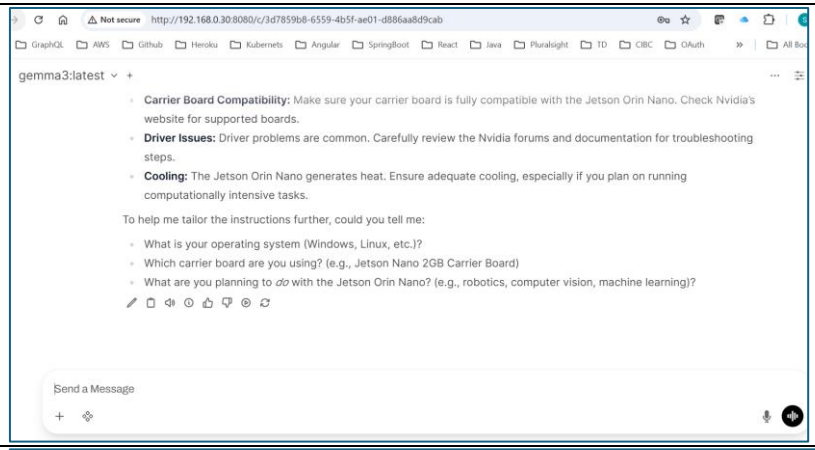
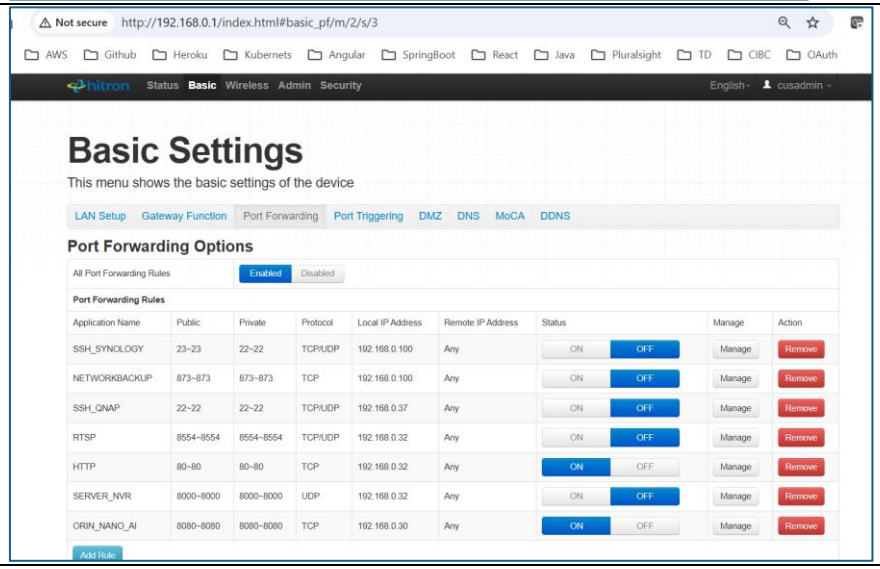
13	**Add miniSD Card to Orin Nano** • *Plug in the miniSD Card into Nvidia Orin Nano server, as shown in the details* *Observation: you can use & flash the OS image into a bootable USB drive, and can boot the Jetson OS from bootable USB drive; in that case plud the USB stick into the Nanor device USB port*	
14	**START UP - Power ON the Nvidia Jetson Orin Nano device** *Observation: Nano device doens't have an ON/OFF switch* • *Plug in the power adapter in the Nano server* • *You should see a Nvidia UEFI Boot screen similar to the image on the right*	

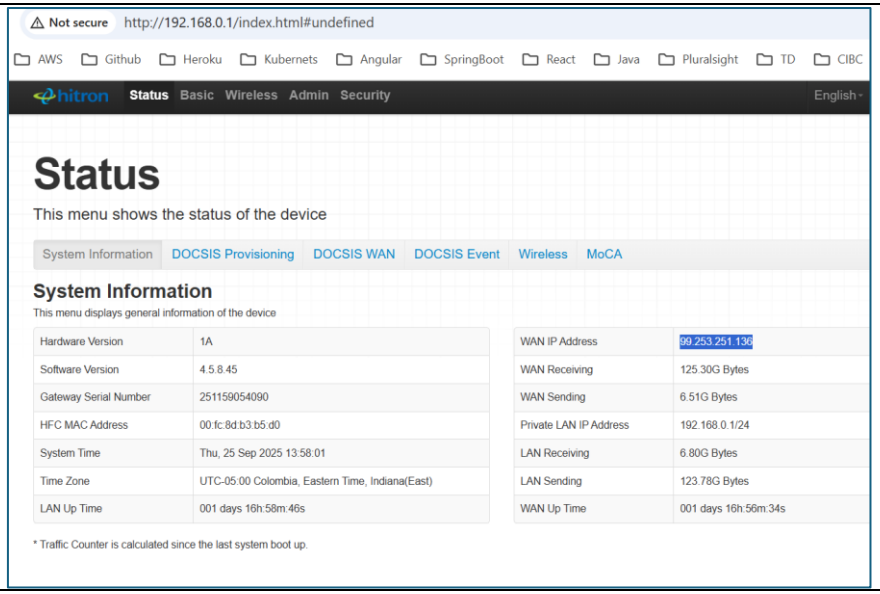
15	**START Jetson OS first time on Nano** • *Ubuntu Jetson OS should boot automatically from miniSD card (or USB)* • *Ubuntu Nvidia Jetson OS Desktop installation sequence will appear on the Nano Monitor* *You will be prompted to with configuration screens; time zone, language, user root admin ID, setup Nano admin password etc* *Observation: software update followed by a reboot might be necessary and required* *For software update you can open **Terminal** application and update Ubunto OS and other applications like – python, docker, java etc*	
	SETUP Jetson Orin Power Mode *Observation: this will setup Orin Nano power mode – 15 W, 25 W, or MAXN SUPER*	
16	**Rollback the snap Application version** *Open command prompt **Terminal** application on Nano: enter the following command to rollback and lock the **snap** program version* ***Observation**; web browsers available on Orin Nano (Chromium , Firefix) cannot be launched from Jetson Nano desktop due to a security related defect for **snap** version; you have to downgrade the snap app version with specific command*	Details steps and commands are provided in this YouTube video **3 Minute Fix for Chromium and other Snaps not launching:** https://www.youtube.com/watch?v=x6bccF3xtRE&t=79s
17	**Install Jetson Docker Containers** jetson-containers run $(autotag text-generation-webui)	Details steps and commands are provided in this YouTube video **Use These! Jetson Docker Containers Tutorial:** https://www.youtube.com/watch?v=HlH3QkS1F5Y

18	**Start Chromium Web browser on Nano** *Login to AI LAB DE Zone with user ID & password* *Follow the Tutorial https://www.jetson-ai-lab.com/* *Observation: for this step you have to login to Nvidia DEV zone with user ID and password directly from your Orin Nano Desktop using Chromium*	
19	**INSTALL OLLAMA AI SERVER** • *Open **Terminal** application on Nano* • *Install Ollama AI server from Docker Image using following command* mkdir ~/ollama-data/ docker run --rm -it -v ${HOME}/ollama-data:/data ghcr.io/nvidia-ai-iot/ollama:r38.2.arm64-sbsa-cu130-24.04 *or* *Copy command from tutorial* *https://www.jetson-ai-lab.com/tutorial_ollama.html*	
20	*After Ollama server installation was completed Please use the command **Terminal** Application in the same command window* Starting ollama server	

	For Native Install Ollama AI server on Nano - you should follow the AI LAB Tutorial details: https://www.jetson-ai-lab.com/tutorial_ollama.html ***Observation:*** *Native Ollama installation is much more complex setup and for first Ollama installation we recommend Docker container Ollama installation above; user has the option and can decide to install native Ollama server*	
21	***LOAD OLLAMA SERVER*** *with gemma3 LLM MODEL* *Load LLM model in Ollama*	
22	**USE OLLAMA CLIENT AS AI PROMPT** • *In the same* ***Terminal*** *command you can use directly the* ***Ollama Client AI Prompt***	

23	**INSTALL OPEN WEB UI SERVER** *Open another second command **Terminal** window to install Open Web UI docker installation* *Install Open WebUI AI server from Docker Image using the following command:* docker run -it --rm --network=host --add-host=host.docker.internal:host-gateway ghcr.io/open-webui/open-webui:main *or* *Copy command from AI LAB tutorial:* *https://www.jetson-ai-lab.com/tutorial_ollama.html*	
25	*After install completes, please check the Open Web UI install using Chromium Web Browser;* Default port is 8080 *http://192.168.0.30:8080/* *On first page load of Open Web UI page, Create Admin account, you have to Register with username and password*	

26	*Start using Web UI AI features from Nano local web browser by accessing the Open Web UI App endpoint* *From Orin Nano on local host: http://localhost:8080* *From Local home intranet: http://JETSON_NANO-LOCAL-IP:8080*	
	Setup home router to allow Nano server Open Web UI access from anywhere Internet *Internet access endpoint http://HOME-ROUTER-IP:8080* ***Observation**; to access the Ollama WEB UI from the internet, outside your local network you have to configure & create a PORT FOWARDING rule on the Internet & Router. Configuration might be different for specific Internet Provider and specific routers, or specifc home setup:* *As an example we present Canadian Rogers high Speed internet provider utilizing Hytron router*	

	RULE FORWARD CONFIG IN HOME ROUTER • *Configure Home or business network router with Forwarding rule* • *Get your static router IP to access from home local network* • *Use browser or mobile phone – from anywhere on Internet:* *Connect to your home router Web admin page:* *http://192.168.0.1/login.html* *Login to your router using admin account with* • *cusadmin* • *password*	
	Start using Web UI AI features & Ollama Server AI running on your personal network on Nano, from anywhere on the internet *Internet access IP endpoint http://HOME-ROUTER-IP:8080* *For example: http://99.253.251.136:8080*	
	Secure the internet access using VPN tunneling *For personal use and for much more secure access, a VPN secure solution and connection is recommended* • *Download VNC Server* • *Download VNC Client* • *Generate secure encryption keys* • *Setup VNC connection* • *Access Nano server Desktop remotely*	