



# **aiDAPTIV+ Pro Suite 2.0**

## **Model Download and Preparation**

**Version 1.0**

**Phison Electronics Corporation**

Tel: +886-37-586-896 Fax: +886-37-587-868

E-mail: sales@phison.com / support@phison.com

Phison may make changes to specifications and product description at any time without notice. PHISON and the Phison logo are trademarks of Phison Electronics Corporation, registered in the United States and other countries. Products and specifications discussed herein are for reference purposes only. Copies of documents which include information of part number or ordering number, or other materials may be obtained by emailing us at sales@phison.com or support@phison.com.

©2024 Phison Electronics Corp. All Rights Reserved.

## REVISION HISTORY

Revision	Draft Date	History	Pro Suite Version	Author
0.1	2025/10/28	Draft version 0.1	NXUN_2.0.7	Jarred Walton

# 1. INTRODUCTION

Phison aiDAPTIV+ provides the hardware and software necessary to fine-tune train many of the most popular AI models. However, licensing and distribution agreements preclude companies from shipping systems with most Large Language Models (LLMs) pre-installed. This document will cover the necessary steps to create an account and gain access to Meta's Llama 3 models for use with aiDAPTIV+.

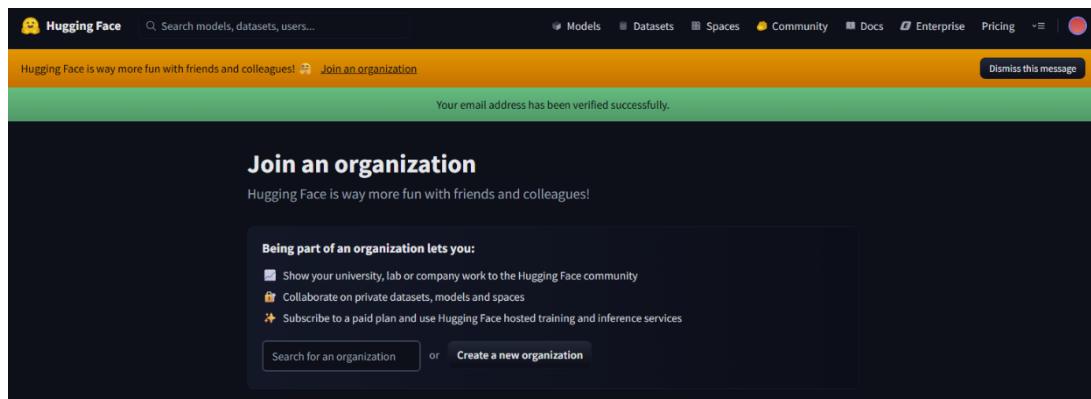
# 2. ACCOUNT CREATION

HuggingFace.co is one of the most popular websites for AI models, including Meta Llama, DeepSeek, Qwen, and many more. Most publicly available LLMs have restricted access to their models and require an account at HuggingFace, along with signing an agreement on how the model(s) will be used. This is intended to prevent end users from unwittingly downloading modified models that don't represent the original LLMs. Create an account by opening [HuggingFace.co](https://huggingface.co) in the browser of your choice to get started.

After giving your email and choosing a password, followed by choosing a unique username, you'll receive a confirmation email from Hugging Face. Follow the link provided to confirm your account.

## 2.1. Organizations

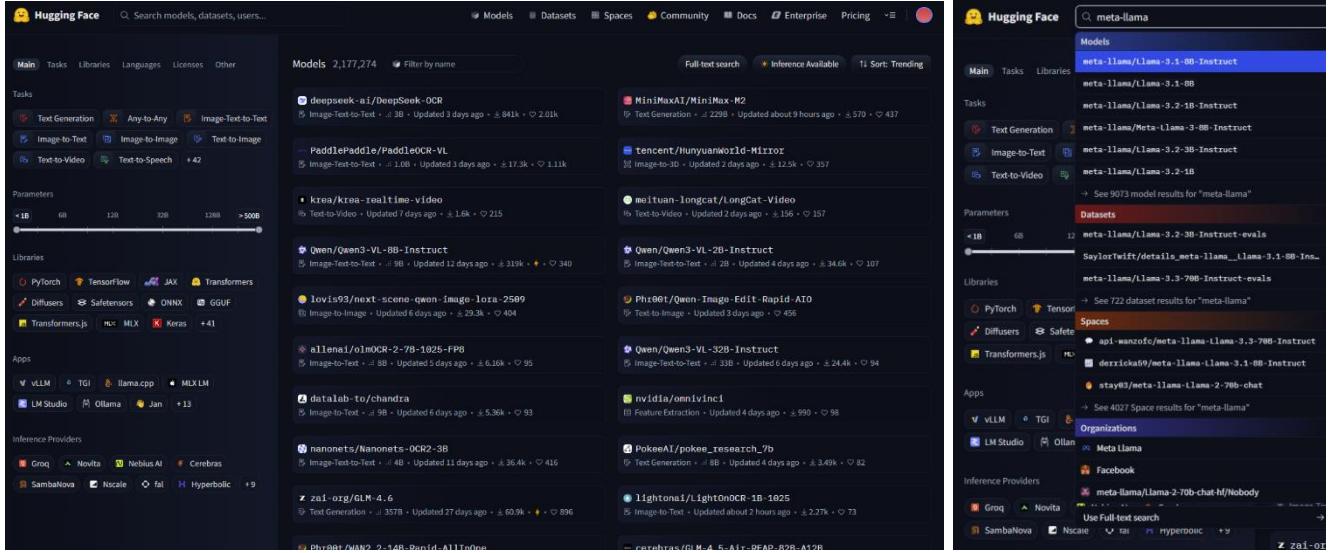
An optional step is to create or join an organization, which allows people from the same group — a business, university, lab, etc. — to more readily collaborate and work together. You can also provide information for accounts with X, GitHub, and LinkedIn, and provide information on your AI and ML interests. Once you've completed these steps (or skipped them), you'll land on your Hugging Face profile page.



Two side-by-side screenshots of the Hugging Face interface. The left screenshot shows the 'New Organization' creation form with fields for 'Organization Username', 'Organization type', 'Homepage', 'Logo', 'Organization Full name', 'GitHub username', 'Twitter username', 'LinkedIn profile', and 'AI &amp; ML interests'. The right screenshot shows the resulting organization profile page with sections for 'Models' (0, None yet), 'Datasets' (0, None yet), 'Edit profile', 'Settings', 'AI &amp; ML interests' (None yet), and 'Organizations' (None yet). A large red circle obscures the profile picture.

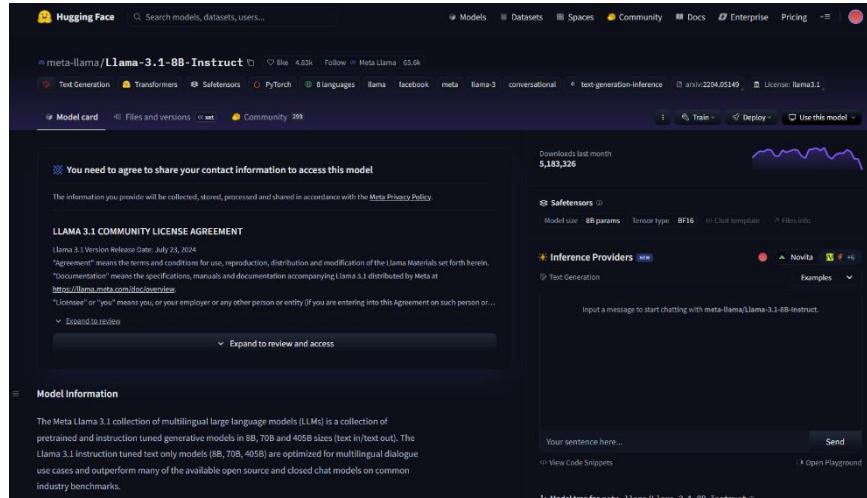
### 3. ACCESSING MODELS

To get started with aiDAPTIV+, the most useful area will be the Models section of HuggingFace.co, linked along the top of the webpages. There are thousands of models that have been uploaded by companies and users. For the purposes of this guide, we'll get Meta's Llama 3 models for use with aiDAPTIV+.



#### 3.1. Meta Llama 3.1 8B

Click in the search bar and type “meta-llama” and a list of matches will populate as you type. Not all models are in the required format to work with aiDAPTIV+, and for Llama 3.1 8B specifically we want the Instruct version: Llama-3.1-8B-Instruct.



Before you can download Llama 3, you need to agree to the community license agreement. Click the “Expand to review and access” button, then fill in the form with the requested information. Once you submit a request for access, you’ll see a notification. It typically takes a few minutes for the request to be granted, at which point you’ll receive an email notification of approval. Clicking the “your settings” link will show the status of your request.

By agreeing you accept to share your contact information (email and username) with the repository authors.

First Name	<input type="text"/>
Last Name	<input type="text"/>
Date of birth	<input type="text"/> mm/dd/yyyy
Country	<input type="text"/>
Affiliation	<input type="text"/>
Job title	<input type="text"/>

Your country and region (based on approximate internet address) will be shared with the model owner.

By clicking Submit below I accept the terms of the license and acknowledge that the information I provide will be collected, stored, processed and shared in accordance with the Meta Privacy Policy

**Model Information**

The Meta Llama 3.1 collection of multilingual large language models (LLMs) is a collection of pretrained and instruction tuned generative models in 1B, 70B and 405B sizes (text in/text out). The Llama 3.1 instruction tuned text only models (8B, 70B, 405B) are optimized for multilingual dialogue.

Hugging Face Search models, datasets, users...

meta-llama/Llama-3.1-8B-Instruct 4.8K Follow Meta Llama 65.6k

Text Generation Transformers Safetensors PyTorch Languages Llama Facebook meta llama-3 conversational test-generation-inference univc204.05149 License: Llama-1

Model card Files and versions Community 298

You need to agree to share your contact information to access this model

The information you provide will be collected, stored, processed and shared in accordance with the [Meta Privacy Policy](#).

LLAMA 3.1 COMMUNITY LICENSE AGREEMENT

Llama 3.1 Version Release Date: July 23, 2024

"Agreement" means the terms and conditions for use, reproduction, distribution and modification of the Llama Materials set forth herein.

"Documentation" means the specifications, manuals and documentation accompanying Llama 3.1 distributed by Meta at <https://llama.meta.com/docOverview>.

"Owner" or "you" means you, or your employer or any other person or entity (if you are entering into this Agreement on such person's behalf).

Your request to access this repository has been submitted and is awaiting a review from the repository authors. You can check the status of all your access requests in [your settings](#).

If approved, you'll get access to all 17 repositories in the [Meta's Llama 3.1 models & evals Getting Started Collection](#), including this one.

Downloads last month 5,183,326

Inference Providers Text Generation Examples

Model size: 8B params Tensor type: BF16 Chat template: File info

Input a message to start chatting with meta-llama/Llama-3.1-8B-Instruct.

Your sentence here... Send View Code Snippets Open Playground

Hugging Face Search models, datasets, users...

Gated Repos Status

View the gated repositories that you have requested access to.

Profile Account Authentication Organizations Billing

Filter by repo name

Repo Name	Type	Date	Request Status
Meta's Llama 3.1 models & evals	Gating Group Collection	Oct 27	<span style="background-color: yellow;">Pending</span>

Once your request has been accepted, you'll have access to all the related repositories — Llama 3.x 1B, 3B, 8B, 70B, and 405B.

[Access granted] Your request to access model meta-llama/Llama-3.3-70B-Instruct has been accepted

huggingface <website@huggingface.co> to me 11:10 AM (6 minutes ago)

Hi

This is to let you know your request to access model "meta-llama/Llama-3.3-70B-Instruct" on [huggingface.co](#) has been accepted by the repo authors.

You can now access the repo [here](#), or view all your gated repo access requests [in your settings](#).

Cheers,

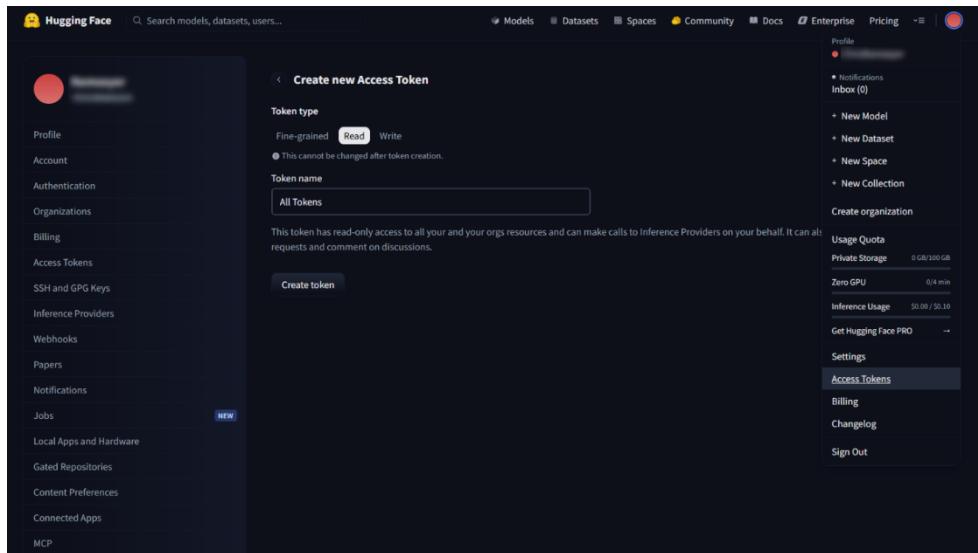
The Hugging Face team

Hugging Face: The AI community building the future.

Reply Forward

### 3.2. Creating an Access Token for use with Pro Suite

Once an account has been granted access to the appropriate LLM, Llama 3 in this example, Pro Suite needs an access token in order to download the files from Hugging Face. To create a token, click on your user profile icon (in the top-right corner of the Hugging Face site), and the fourth option from the bottom of the list should be Access Tokens — click that.



If this is a new account with no existing tokens, you need to create one. The simplest option is to create a read-only token with access to all the repositories you have on Hugging Face. Set the token type to “Read” and give it an appropriate name, then click the “Create token” button.

A popup will appear with the Access Token key, which you will want to copy and save in a secure location — **this will be the only chance to copy the key**. If you need the key later, you will have to invalidate and refresh the token, which will show up in the list of access tokens.

Name	Permissions
All Tokens	READ

### 3.3. Creating Linux bash commands for downloading the LLMs

For the Llama 3.x LLMs, there are multiple models to download: 1B, 3B, 8B, 70B, and 405B. The last two are quite large and will require significant amounts of storage space. The commands to download in Linux are given below as well as in the image, which will need the Hugging Face access token created above. Note that you should *not* use “sudo” on these commands or it will fail.

```
bash <(curl -sSL https://g.bodaay.io/hfd) -t [YOUR_API_KEY] download meta-llama/Llama-3.2-1B-Instruct  
bash <(curl -sSL https://g.bodaay.io/hfd) -t [YOUR_API_KEY] download meta-llama/Llama-3.2-3B-Instruct  
bash <(curl -sSL https://g.bodaay.io/hfd) -t [YOUR_API_KEY] download meta-llama/Llama-3.1-8B-Instruct  
bash <(curl -sSL https://g.bodaay.io/hfd) -t [YOUR_API_KEY] download meta-llama/Llama-3.1-70B-Instruct  
bash <(curl -sSL https://g.bodaay.io/hfd) -t [YOUR_API_KEY] download meta-llama/Llama-3.1-405B-Instruct
```

### 3.4. Downloading the LLMs

Open a terminal and paste the commands for the desired LLMs into the window, one at a time. Press enter to send the command and begin downloading. It can take several minutes to multiple hours, depending on the size of the LLM and the speed of your connection.

The screenshot shows two terminal windows. The top window displays a command to download the Llama-3.2-1B-Instruct model from huggingface (hf) using curl. The bottom window shows the progress of the download, listing files and their download status, speed, and ETA. The files listed include tokenizer\_config.json, tokenizer.json, special\_tokens\_map.json, generation\_config.json, config.json, USE\_POLICY.md, README.md, LICENSE.txt, .gitattributes, original/params.json, and original/...zer.model. The download is currently at 2% completion, with a total size of 74.2 MiB/4.6 GiB.

```

phison@phison-ThinkPad-P1-Gen-7:~$ bash <(curl -sSL https://g.bodaay.io/hfd) -t hf_(insert your API key here) download meta-llama/Llama-3.2-1B-Instruct
Repo: meta-llama/Llama-3.2-1B-Instruct  Rev: main  Dataset: false
Out: Storage  Conns: 8  MaxActive: 3  Verify: size  Retries: 4  Threshold: 32MiB
2% 74.2 MiB/4.6 GiB  0 B/s  ETA -
>Status  File  Progress  Speed  ETA
▶ downloading  original/...ed.00.pth  50.7 MiB/2.3 GiB  0 B/s  -
▶ downloading  model.safetensors  12.6 MiB/2.3 GiB  0 B/s  -
✓ done  tokenizer_config.json  53.2 KiB/53.2 KiB  0 B/s  -
✓ done  tokenizer.json  8.7 MiB/8.7 MiB  1 0 B/s  -
✓ done  special_tokens_map.json  296 B/296 B  100%  0 B/s  -
✓ done  generation_config.json  189 B/189 B  100%  0 B/s  -
✓ done  config.json  877 B/877 B  100%  0 B/s  -
✓ done  USE_POLICY.md  5.9 KiB/5.9 KiB  1 0 B/s  -
✓ done  README.md  40.8 KiB/40.8 KiB  0 B/s  -
✓ done  LICENSE.txt  7.5 KiB/7.5 KiB  1 0 B/s  -
✓ done  .gitattributes  1.5 KiB/1.5 KiB  1 0 B/s  -
✓ done  original/params.json  220 B/220 B  100%  0 B/s  -
✓ done  original/...zer.model  2.1 MiB/2.1 MiB  1 0 B/s  -
Press Ctrl+C to cancel • linux amd64

```

### 3.5. LLM Preparation

Once the model has been downloaded, you will need to package the data into a ZIP file, or copy and paste it into the correct folder for Pro Suite to utilize. The download command given above puts all the files into the directory: Home/Storage/[model]. There are two methods for preparing the model for use with aiDAPTIV+, with the first being the recommended approach.

In the same terminal as before, change to the appropriate directory:

```
cd ~/Storage/meta-llama
```

Next, copy the files and directory into the correct system directory with the following command:

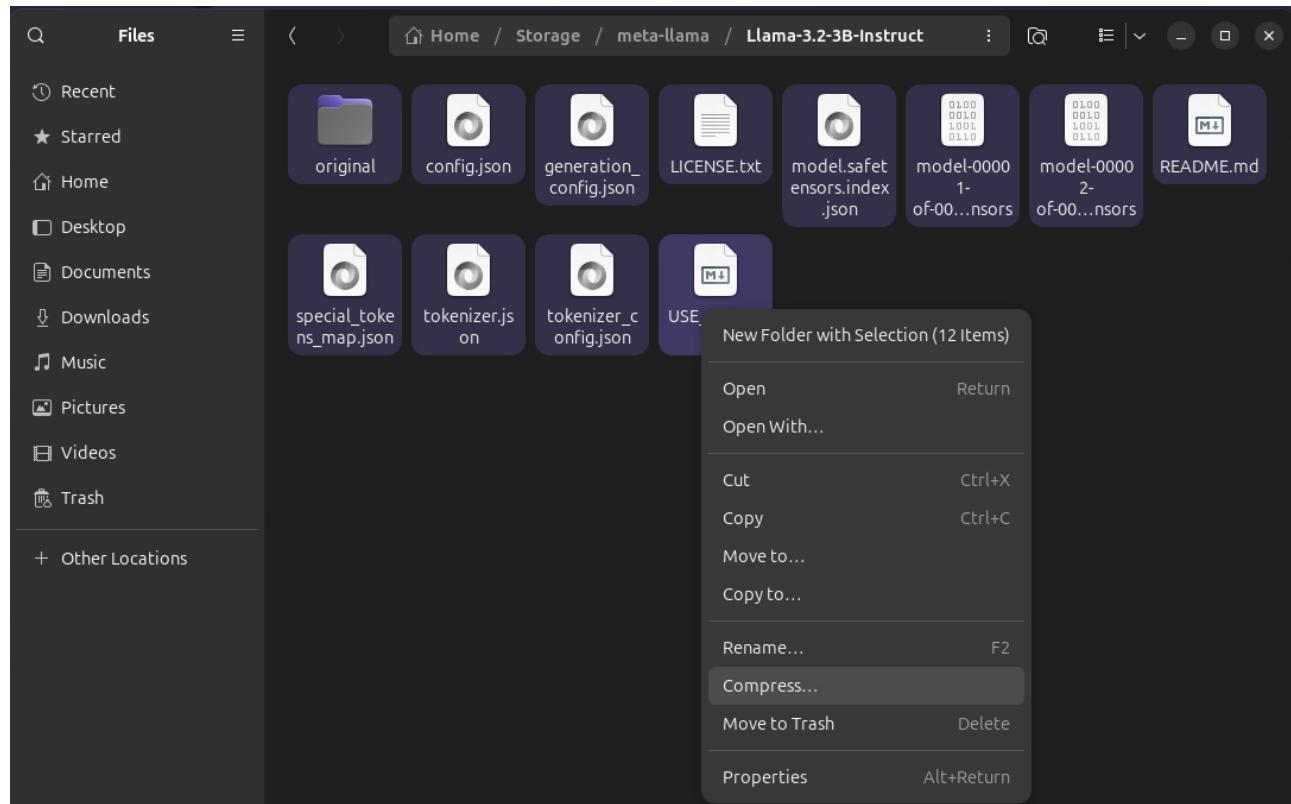
```
sudo cp Llama-3.1-8B-Instruct/ /usr/local/models/
```

The screenshot shows a terminal window with a password prompt for the sudo command. The command entered is sudo cp -r Llama-3.1-8B-Instruct/ /usr/local/models/. The user is prompted for a password.

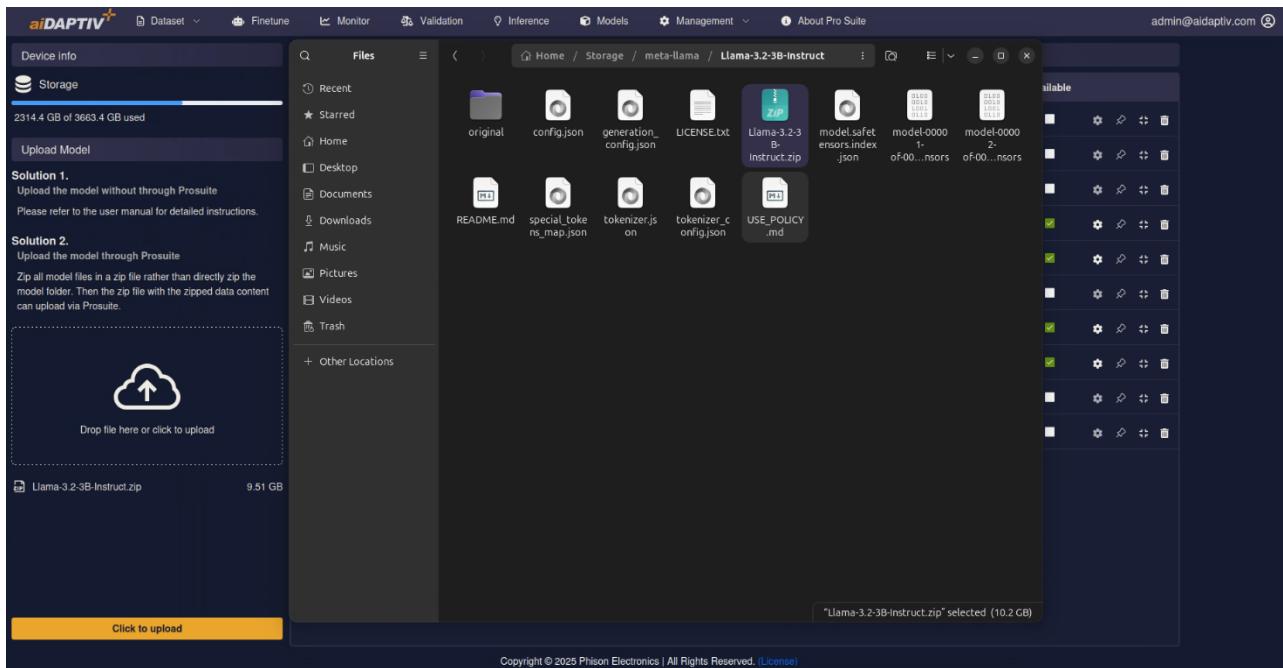
Note that you will need the root password for the sudo command to work.

Alternatively, you can package the files into a ZIP archive. This will require significantly more time, both to create the ZIP archive and to decompress the archive, but it can be useful if you intend to deploy the model files on multiple systems.

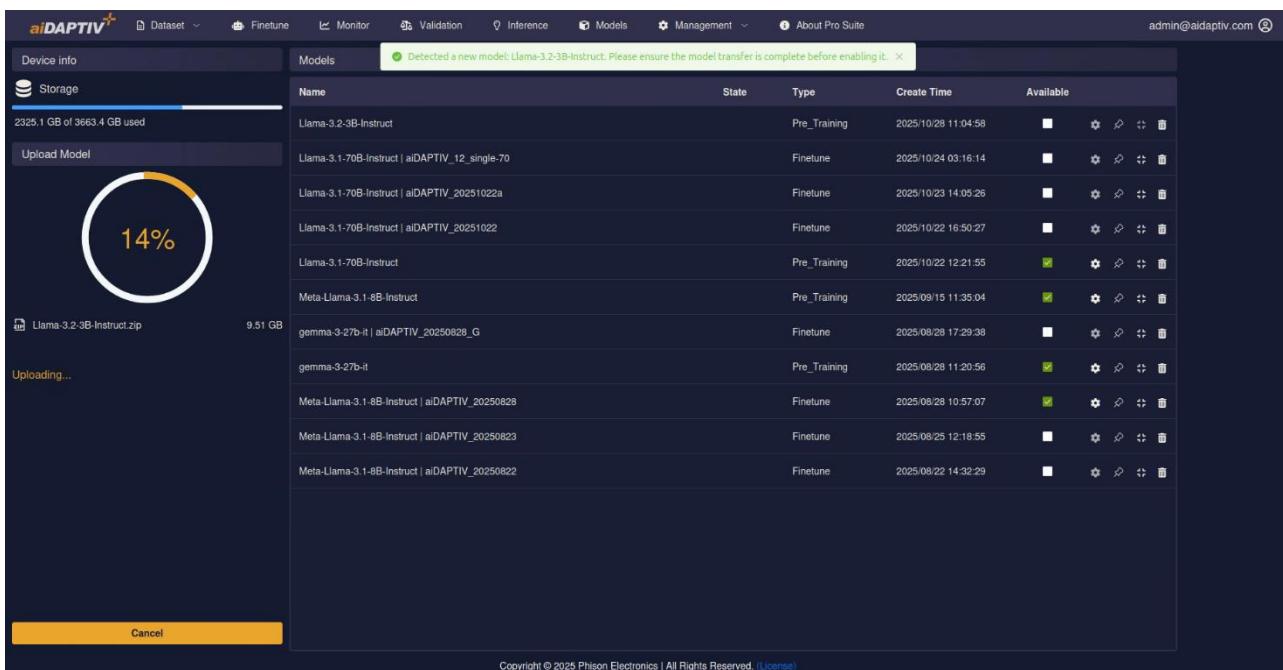
Open File Manager and navigate to Home/Storage/meta-llama, and then choose the appropriate subdirectory (e.g. Llama-3.2-3B-Instruct). Select all the files in the directory (Ctrl+A), right-click and choose “Compress...” then give the archive the appropriate name (Llama-3.2-3B-Instruct). It will take several minutes for the system to create the compressed archive.



Once the ZIP file has been created, open aiADAPTIV+ in the browser, navigate to the Models tab, and then drag and drop the ZIP file in the area on the left of the page that says “Drag files here or click to upload.” (You can also click and navigate to the ZIP file.)



You will now see the model name and size listed, and the “Click to upload” button will appear in the bottom-left corner of the window. Click that and it will copy the files to the necessary folder and decompress the files to the appropriate directory. Again, this can take several minutes.



When the process finishes, you should now see the newly uploaded model in the list of models within aiDAPTIV+. To make the model available for use, check the “Available” box, which will load the model into memory.