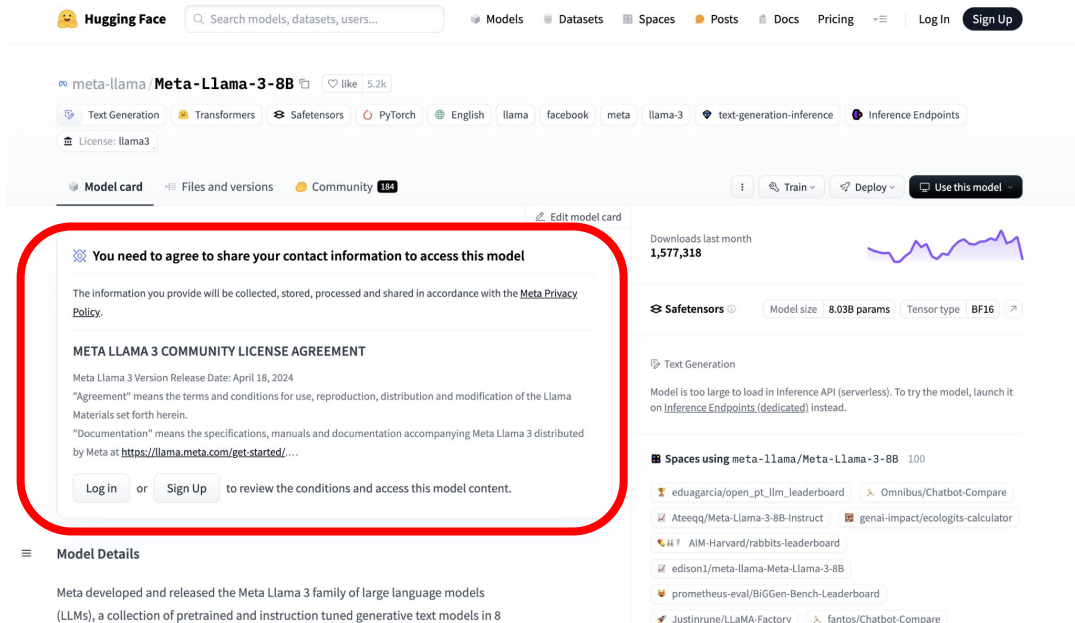


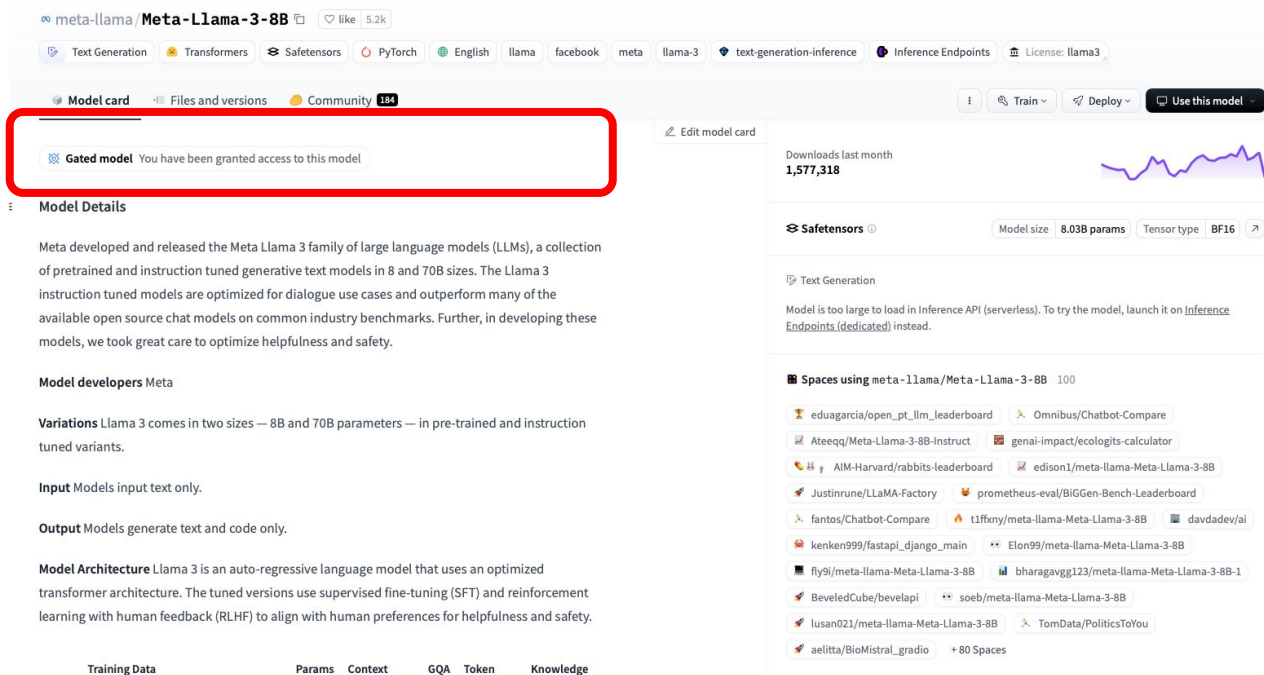
課前準備 - 申請Llama3權限

- (1) 註冊HuggingFace帳號 (若已有請跳過):
前往 <https://huggingface.co/>, 點擊Sign up輸入信箱密碼並驗證
- (2) 前往<https://huggingface.co/meta-llama/Meta-Llama-3-8B>
紅框處填表申請Llama3的權限
大約一天內通過



課前準備 - 申請Llama3權限

(3) 申請通過後 “You have been granted access to this model” 代表申請成功



The screenshot shows the Hugging Face model card for **meta-llama/Meta-Llama-3-8B**. The page features a navigation bar with tabs for 'Model card', 'Files and versions', and 'Community'. The 'Model card' tab is selected, and a red box highlights the 'Gated model' status, which includes the text 'You have been granted access to this model'. Below this, the 'Model Details' section provides information about the Meta Llama 3 family, including its purpose, developers, variations, input/output capabilities, and architecture. On the right side, there is a 'Downloads last month' chart showing 1,577,318 downloads, a 'Safetensors' section with model size and tensor type, and a 'Spaces using meta-llama/Meta-Llama-3-8B' section listing various applications.

meta-llama/Meta-Llama-3-8B like 5.2k

Text Generation Transformers Safetensors PyTorch English llama facebook meta llama-3 text-generation-inference Inference Endpoints License: llama3

Model card Files and versions Community 184

Gated model You have been granted access to this model

Model Details

Meta developed and released the Meta Llama 3 family of large language models (LLMs), a collection of pretrained and instruction tuned generative text models in 8 and 70B sizes. The Llama 3 instruction tuned models are optimized for dialogue use cases and outperform many of the available open source chat models on common industry benchmarks. Further, in developing these models, we took great care to optimize helpfulness and safety.

Model developers Meta

Variations Llama 3 comes in two sizes — 8B and 70B parameters — in pre-trained and instruction tuned variants.

Input Models input text only.

Output Models generate text and code only.

Model Architecture Llama 3 is an auto-regressive language model that uses an optimized transformer architecture. The tuned versions use supervised fine-tuning (SFT) and reinforcement learning with human feedback (RLHF) to align with human preferences for helpfulness and safety.

Downloads last month
1,577,318

Safetensors Model size 8.03B params Tensor type BF16

Text Generation

Model is too large to load in Inference API (serverless). To try the model, launch it on [Inference Endpoints \(dedicated\)](#) instead.

Spaces using meta-llama/Meta-Llama-3-8B 100

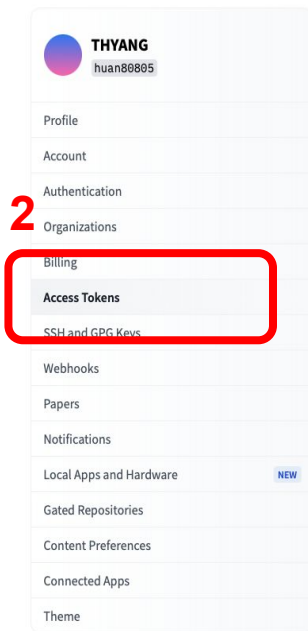
- edugarcia/open_pt_llm_leaderboard
- Omnibus/Chatbot-Compare
- Ateeq/Meta-Llama-3-8B-Instruct
- genai-impact/ecologits-calculator
- AIM-Harvard/rabbits-leaderboard
- edison1/meta-llama-Meta-Llama-3-8B
- Justinrune/LLaMA-Factory
- prometheus-eval/BigGen-Bench-Leaderboard
- fantos/Chatbot-Compare
- tiffkny/meta-llama-Meta-Llama-3-8B
- davdave/ai
- kenken999/fastapi_django_main
- Elon99/meta-llama-Meta-Llama-3-8B
- fly91/meta-llama-Meta-Llama-3-8B
- bharagav123/meta-llama-Meta-Llama-3-8B-1
- BeveledCube/bevelapi
- soeb/meta-llama-Meta-Llama-3-8B
- lusan021/meta-llama-Meta-Llama-3-8B
- TomData/PoliticsToYou
- aelitta/BioMistral_gradio
- + 80 Spaces

Training Data Params Context GQA Token Knowledge

課前準備 - 建立HuggingFace token

點擊頭像 -> Settings

點擊Access Tokens-> Create new token

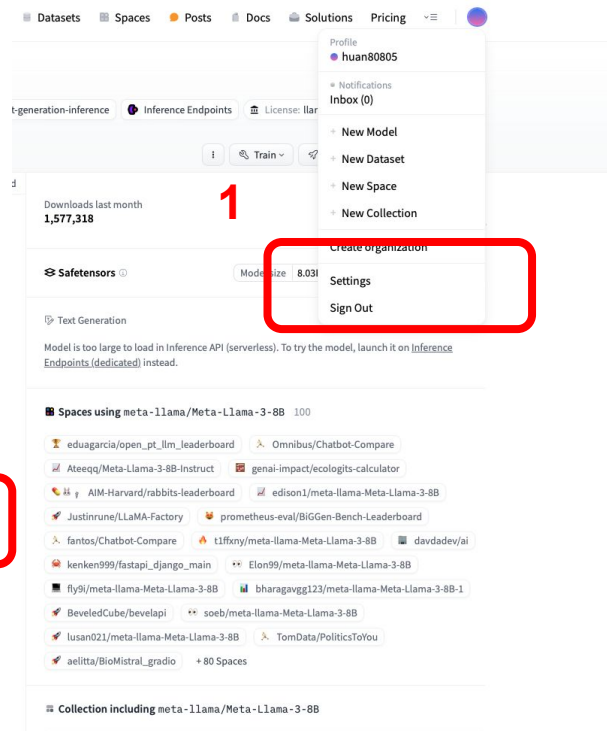


Access Tokens

User Access Tokens

Access tokens authenticate your identity to the Hugging Face Hub and allow applications to perform actions based on token permissions.
Do not share your Access Tokens with anyone; we regularly check for leaked Access Tokens and remove them immediately.

Name	Value	Last Refreshed Date	Last Used Date	Permissions
llama2	hf_...UpFX	3 days ago	-	READ
gemma	hf_...	Jun 5	-	READ
toxiGen	hf_...	Sep 20, 2023	-	WRITE



課前準備 - 建立HuggingFace token

Token type選擇write 並輸入任意名字-> Create a token
完成後點選copy 複製token, lab時將會用到這個token

若忘記token, 會需要更新token(見下頁)

Token type

Fine-grained Read **Write**

ⓘ This cannot be changed after token creation

Token name

lab

This token has read and write access to all your and your orgs resources and can make call

5 Create token

Save your Access Token

Save your token value somewhere safe. **You will not be able to see it again after you close this modal.** If you lose it, you'll have to create a new one.

6

Copy

Name	Permissions
lab	WRITE

Done

更新Token

invalidate and refresh -> Copy

Access Tokens

User Access Tokens

+ Create new token

Access tokens authenticate your identity to the Hugging Face Hub and allow applications to perform actions based on token permissions. **Do not share your Access Tokens with anyone**; we regularly check for leaked Access Tokens and remove them immediately.

Name	Value	Last Refreshed Date	Last Used Date	Permissions
lab	hf_...Z0nc	12 minutes ago	-	WRITE
llama2	hf_...UpFX	3 days ago	-	Invalidate and refresh
gemma	hf_...	1 day ago	-	DELETE

Save your Access Token

Save your token value somewhere safe. **You will not be able to see it again after you close this modal.** If you lose it, you'll have to create a new one.

Copy

Name

Permissions

MTK

FINEGRAINED

Done

Torch Tutorial

實作課將會運用到常用的torch function

請大家follow colab走過一次, 熟悉tutorial裡教到的function

連結:

<https://github.com/Huan80805/MediaTek2024/blob/main/TorchTutorial.ipynb>