

# How to Use Open-Source LLMs

## Instructor

Dipanjan Sarkar

Head of Community & Principal AI Scientist at Analytics Vidhya

Google Developer Expert - ML & Cloud Champion Innovator

Published Author



## Key Focus

- How to use open-source LLMs with **HuggingFace**
- How to use open-source LLMs with **Groq**
- How to use open-source LLMs with **LangChain**

# Platform Covered



Groq

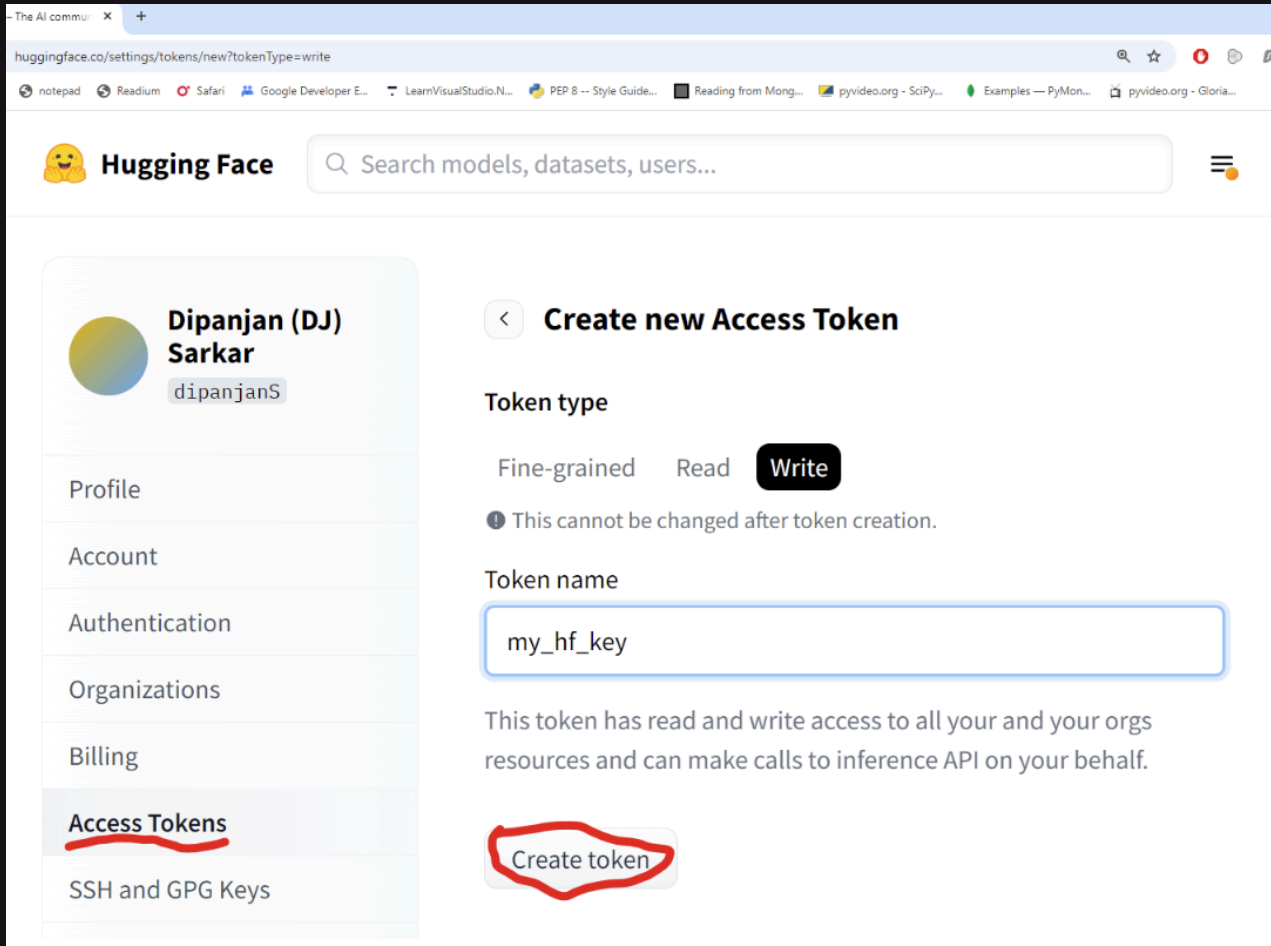


HuggingFace



LangChain

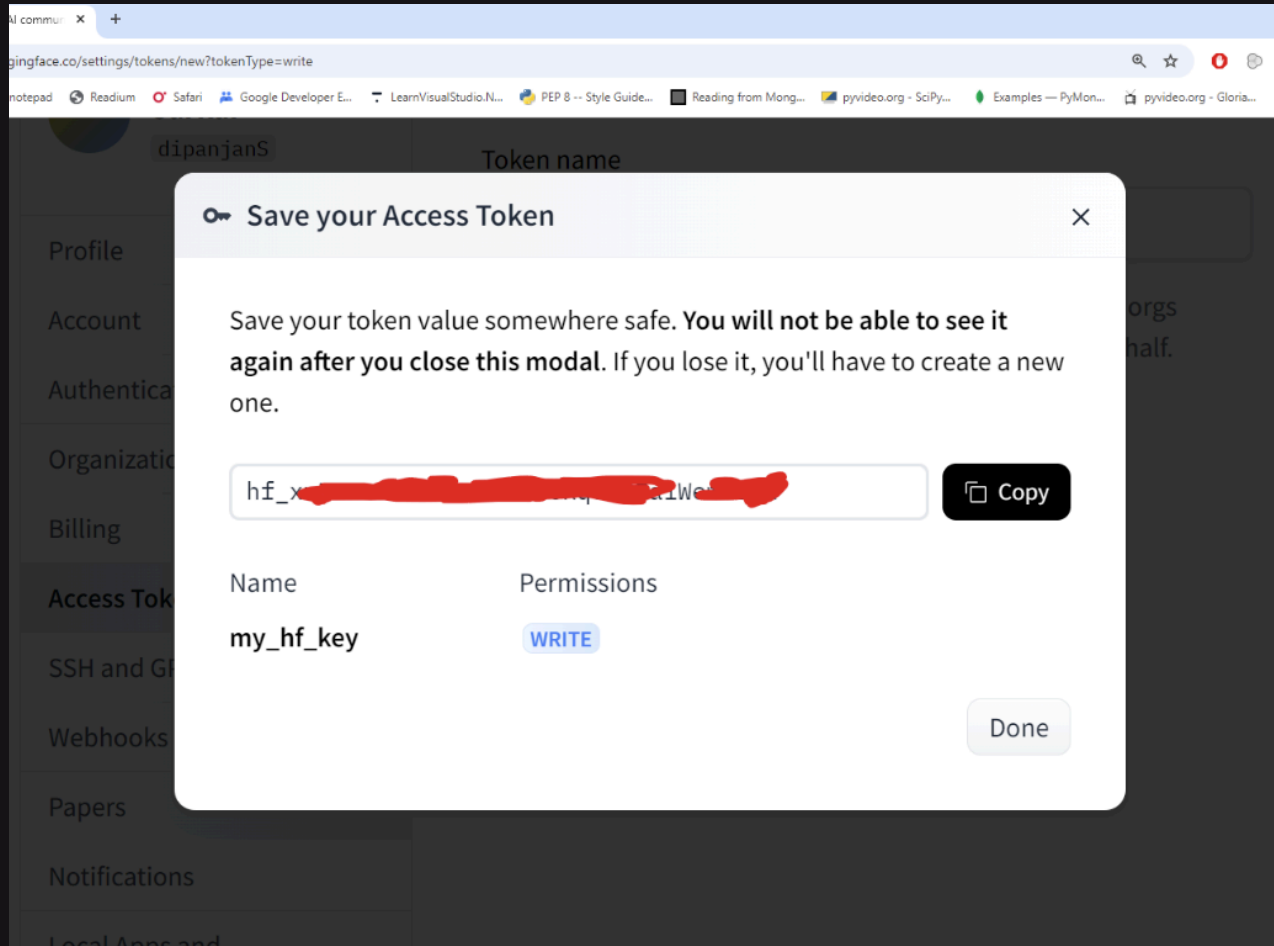
# Get your Hugging Face Access Token



The screenshot shows the Hugging Face website interface. On the left, a sidebar contains the user's profile information and a list of navigation links: Profile, Account, Authentication, Organizations, Billing, **Access Tokens** (highlighted with a red underline), and SSH and GPG Keys. The main content area is titled 'Create new Access Token'. It features a 'Token type' section with three options: 'Fine-grained', 'Read', and 'Write' (which is selected and highlighted with a black background). Below this, a note states: 'This cannot be changed after token creation.' The 'Token name' section has a text input field containing 'my\_hf\_key'. At the bottom, a 'Create token' button is circled in red. The page also includes a search bar at the top and a browser tab indicating the URL 'huggingface.co/settings/tokens/new?tokenType=write'.

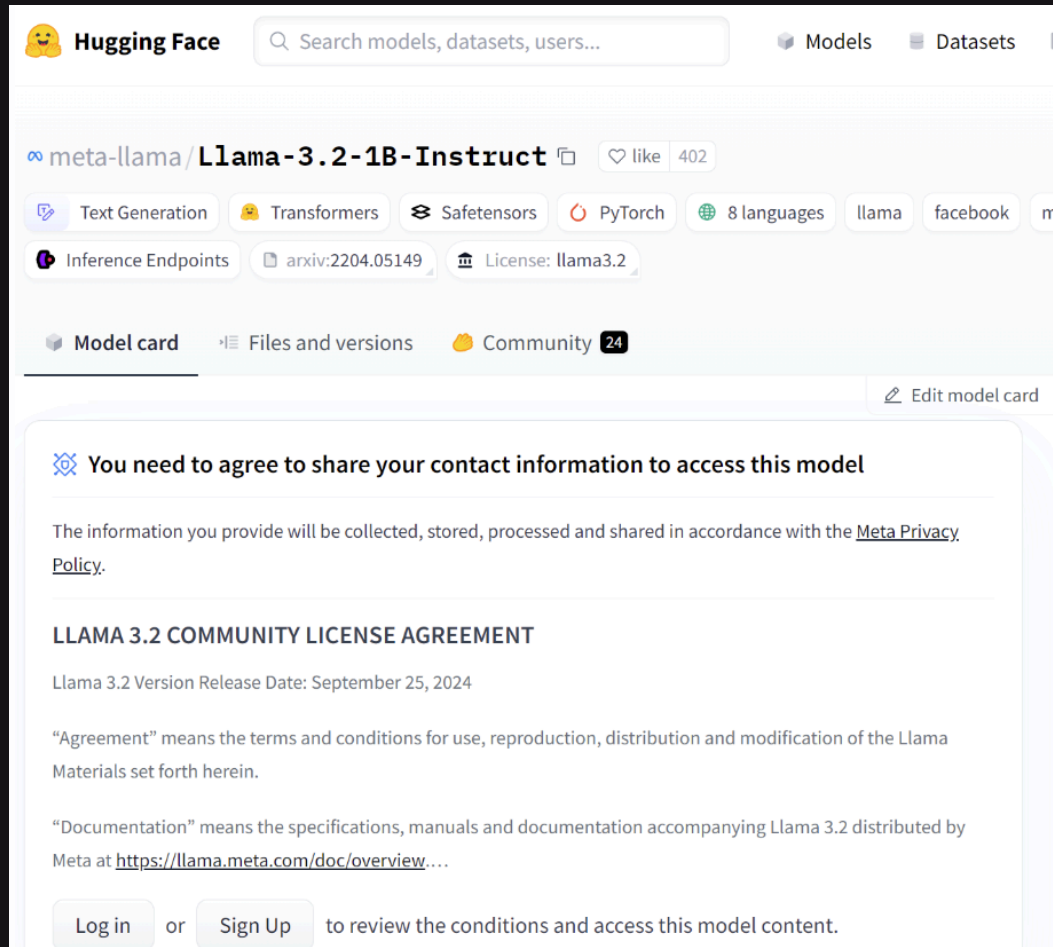
- Create and sign in to your account on <https://huggingface.co/>
- Go to Settings -> Access Tokens
- Create a new token with Write permissions

# Save your Hugging Face Access Token



- Save your secret key in a secure location as it will be just shown once
- If you forget your key just generate a new one
- **Do NOT** paste your key publicly on the internet

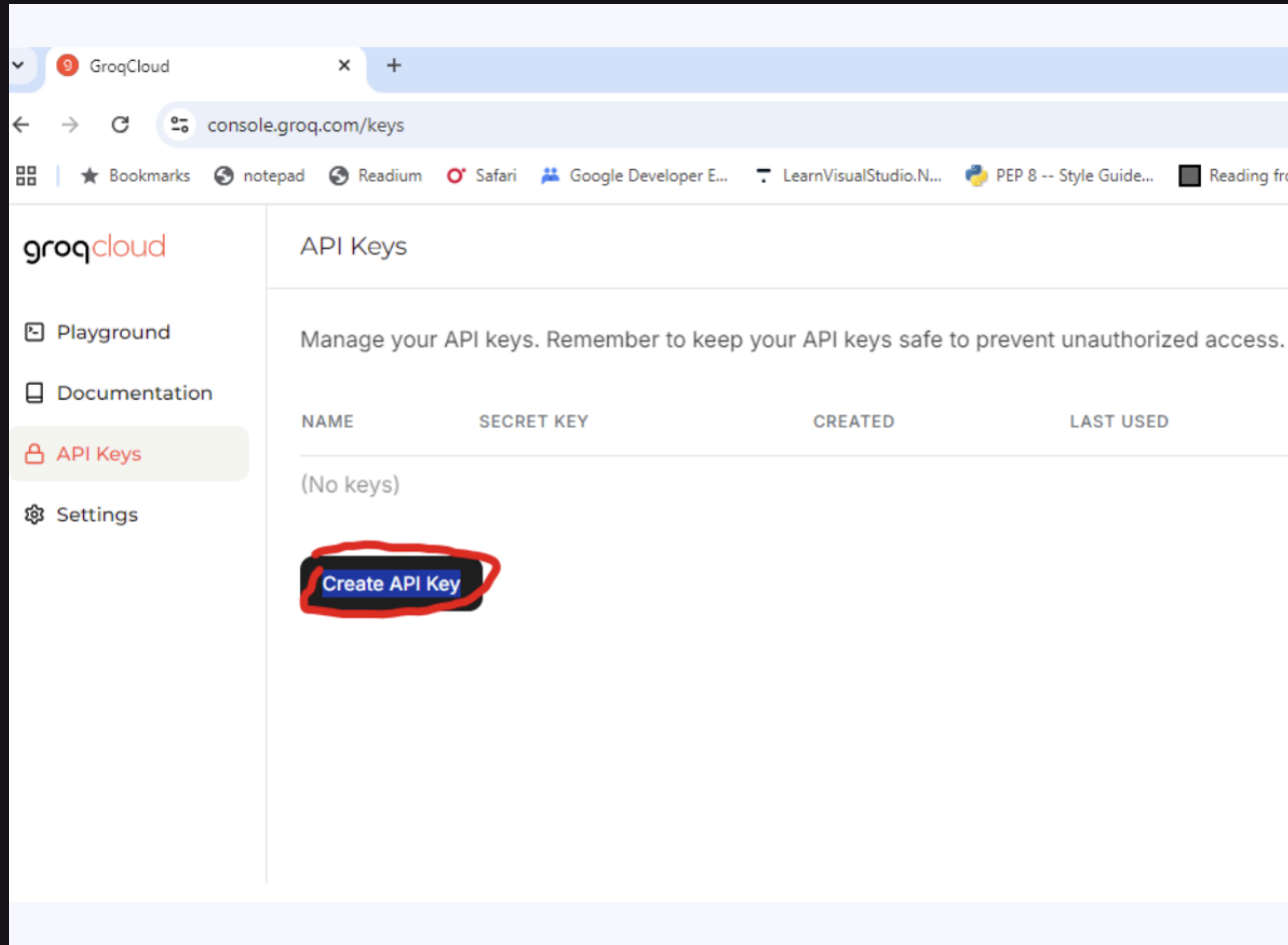
# Apply for Access to Gated Models



- Several new LLMs have gated access
- If you want to use these LLMs like Llama 3.2 Instruct you need to apply for access
- Go to the relevant model page for models you want to use and access the terms and conditions
- You can then start using the model via Hugging Face transformers

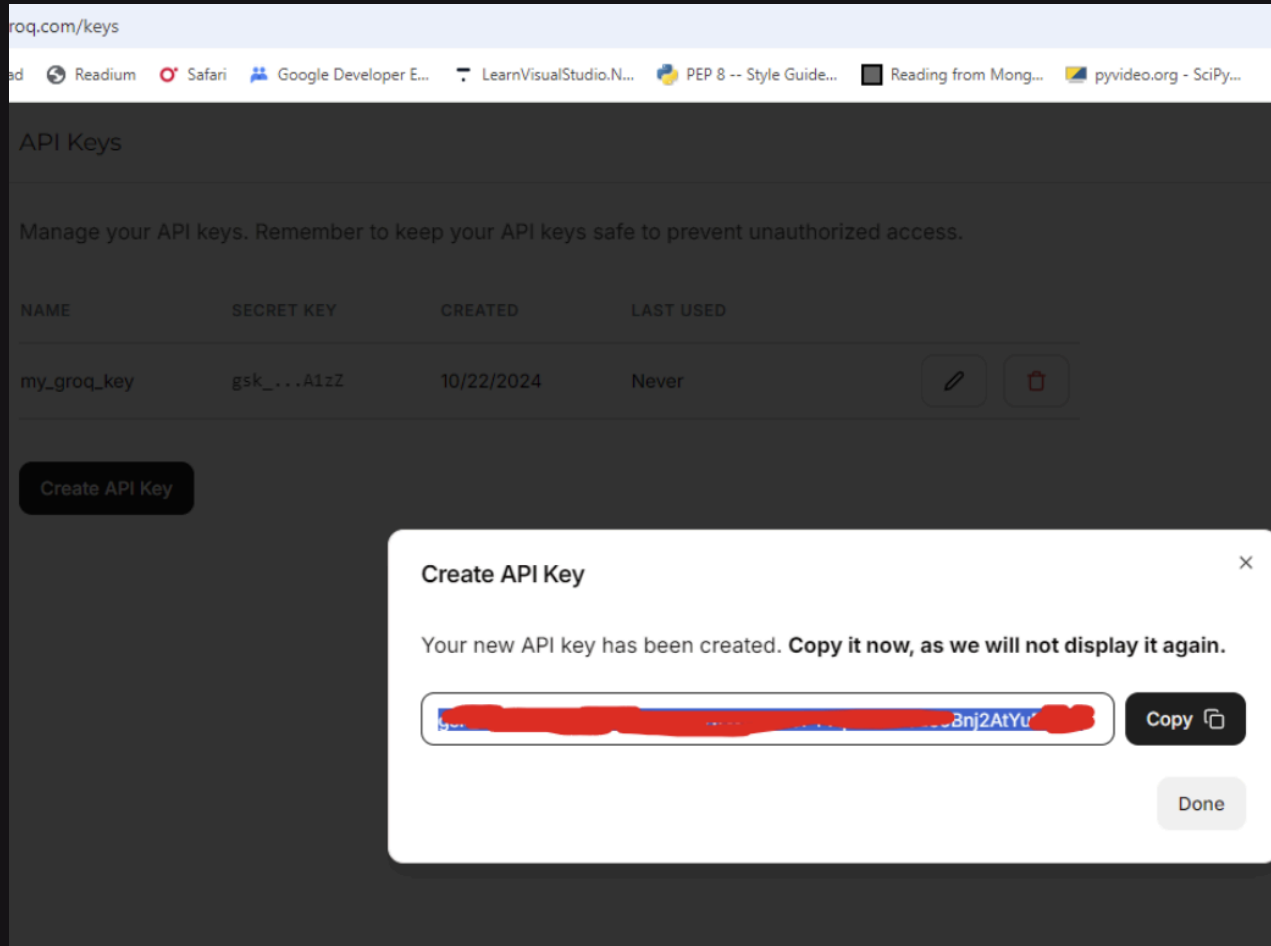


# Get your Groq Cloud API Key



- Create an account on <https://console.groq.com/keys>
- Go to **Groq Cloud** -> **Create API Key** to create your API key
- Several open source models have a free tier access

# Save your Groq Cloud API Key



- Save your secret key in a secure location as it will be shown just once
- Can create more keys if you lose it
- Do NOT paste your key publicly on the internet



# Groq Cloud Free Tier Limits (Late 2024)

ID	Requests per Minute	Requests per Day	Tokens per Minute	Tokens per Day
gemma-7b-it	30	14400	15,000	500,000
gemma2-9b-it	30	14400	15,000	500,000
llama-3.1-70b-versatile	30	14400	18,000	500,000
llama-3.1-8b-instant	30	14400	20,000	500,000
llama-3.2-11b-text-preview	30	7000	7,000	500,000
llama-3.2-11b-vision-preview	30	7000	7,000	500,000
llama-3.2-1b-preview	30	7000	7,000	500,000
llama-3.2-2b-preview	30	7000	7,000	500,000
llama-3.2-90b-text-preview	30	7000	7,000	500,000
llama-3.2-90b-vision-preview	30	7000	7,000	500,000
llama-guard-3-8b	15	3500	7,000	250,000
llama3-70b-8192	30	14400	15,000	500,000
llama3-8b-8192	30	14400	6,000	500,000
llama3-groq-70b-8192-tool-preview	30	14400	30,000	500,000
llama3-groq-8b-8192-tool-preview	30	14400	30,000	500,000
llava-v1.5-7b-4096-preview	30	14400	15,000	500,000
mixtral-8x7b-32768	30	14400	5,000	500,000

Check out [pricing details here for free API](#) and [here for paid API](#)

# Thank You

---