https://ollama.com/

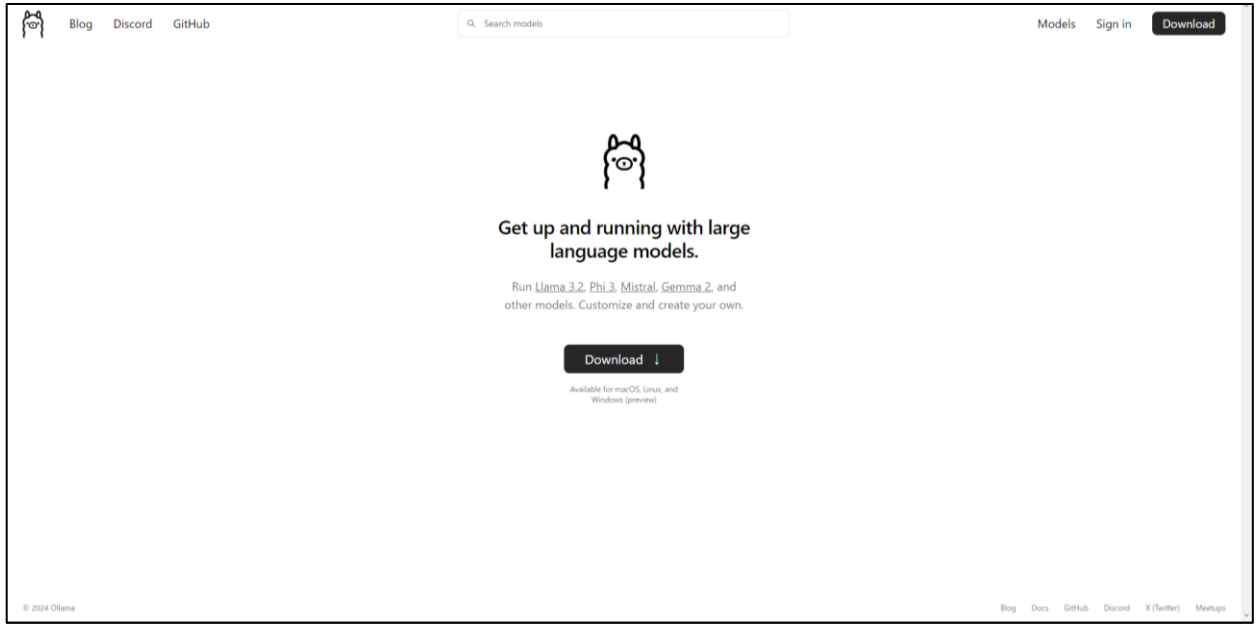
生成AIハンズオンセミナー④

Ollamaによるローカル環境での大規模言語モデルの実行

鳥井浩平
徳島大学デザイン型AI教育研究センター
（併任）徳島大学理工学部

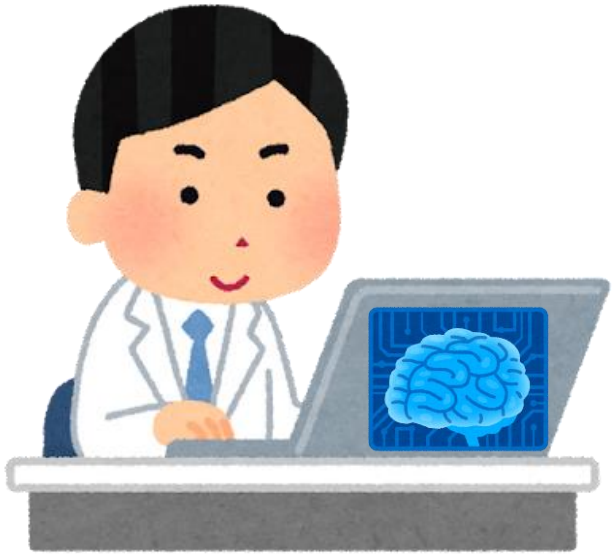
Ollama（オラマ）とは

- 大規模言語モデルをローカル環境で実行するためのソフトウェア
- Windows、macOS、Linux（Ubuntu）に対応

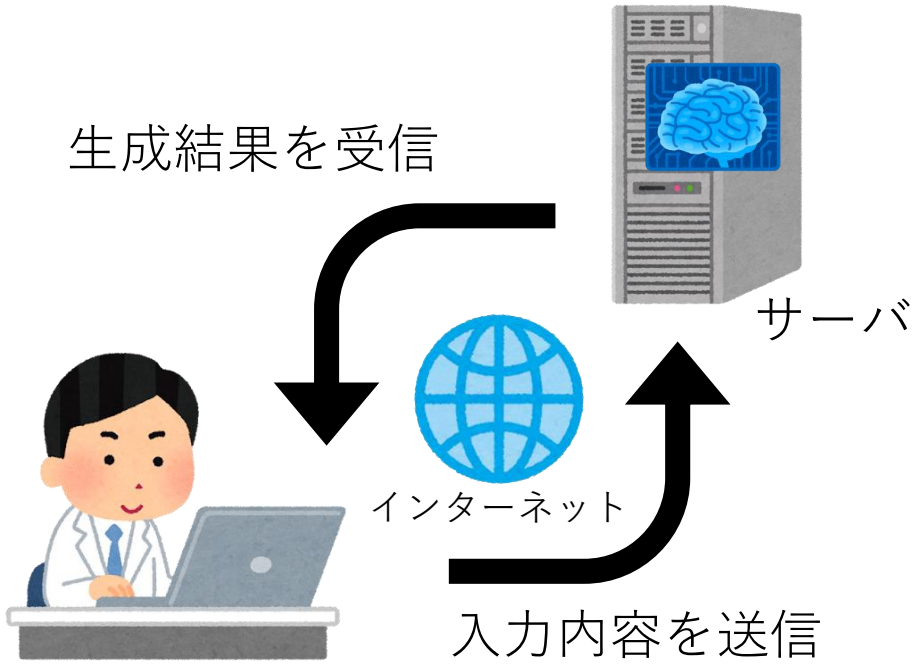


https://ollama.com/

ローカル環境とは



ローカル環境



一般的な生成AIサービス（ChatGPT等）

ローカル環境の利点・欠点

利点

- 実行時にインターネットの接続が不要
- 入力内容が外部に漏れる心配がない

欠点

- 高性能なPCでなければ、実行速度が遅い
- 最初は大規模言語モデルのダウンロードが必要
- 大規模言語モデルはサイズが大きく、PCの容量を圧迫する
  1つのモデルで数ギガバイト

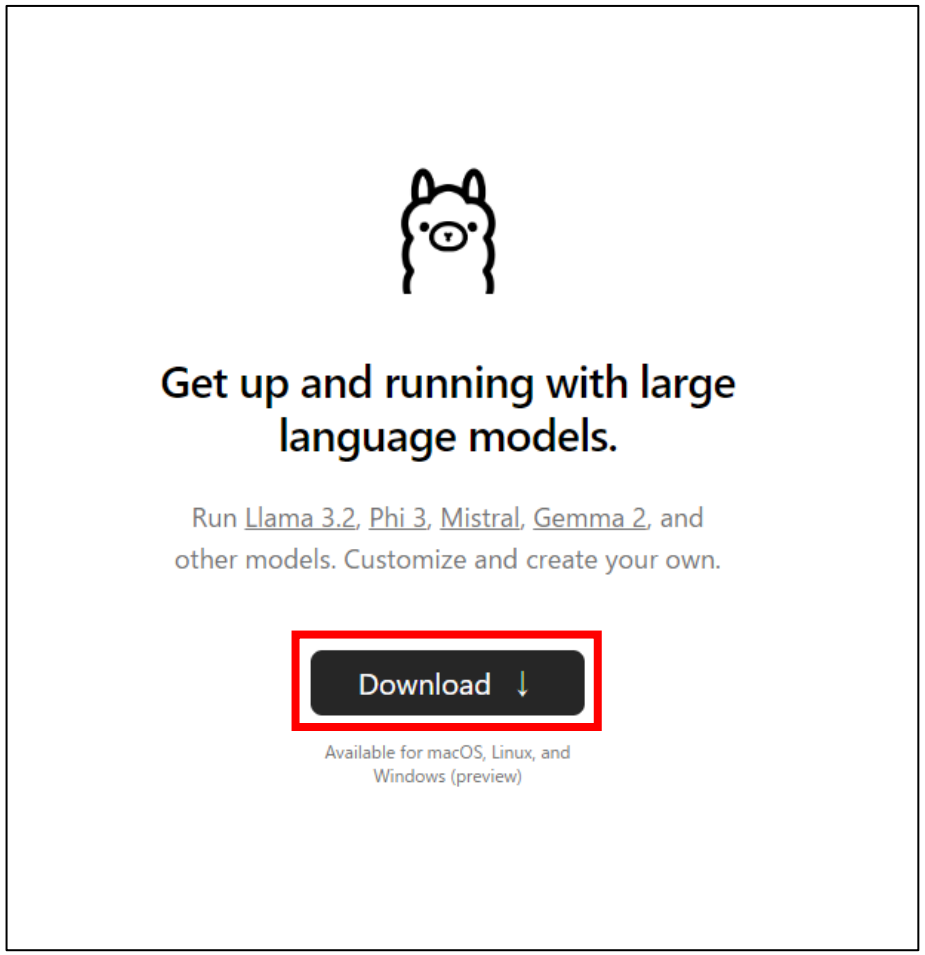
Ollamaをインストールする

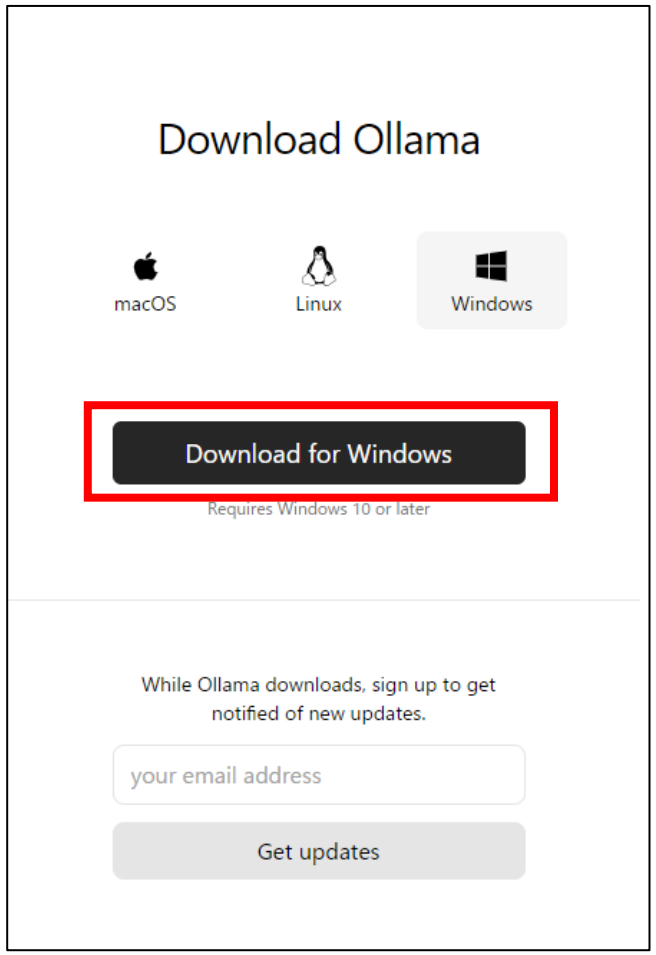
※ここからはWindowsを前提として例示する

1　ウェブサイトにアクセス

https://ollama.com/

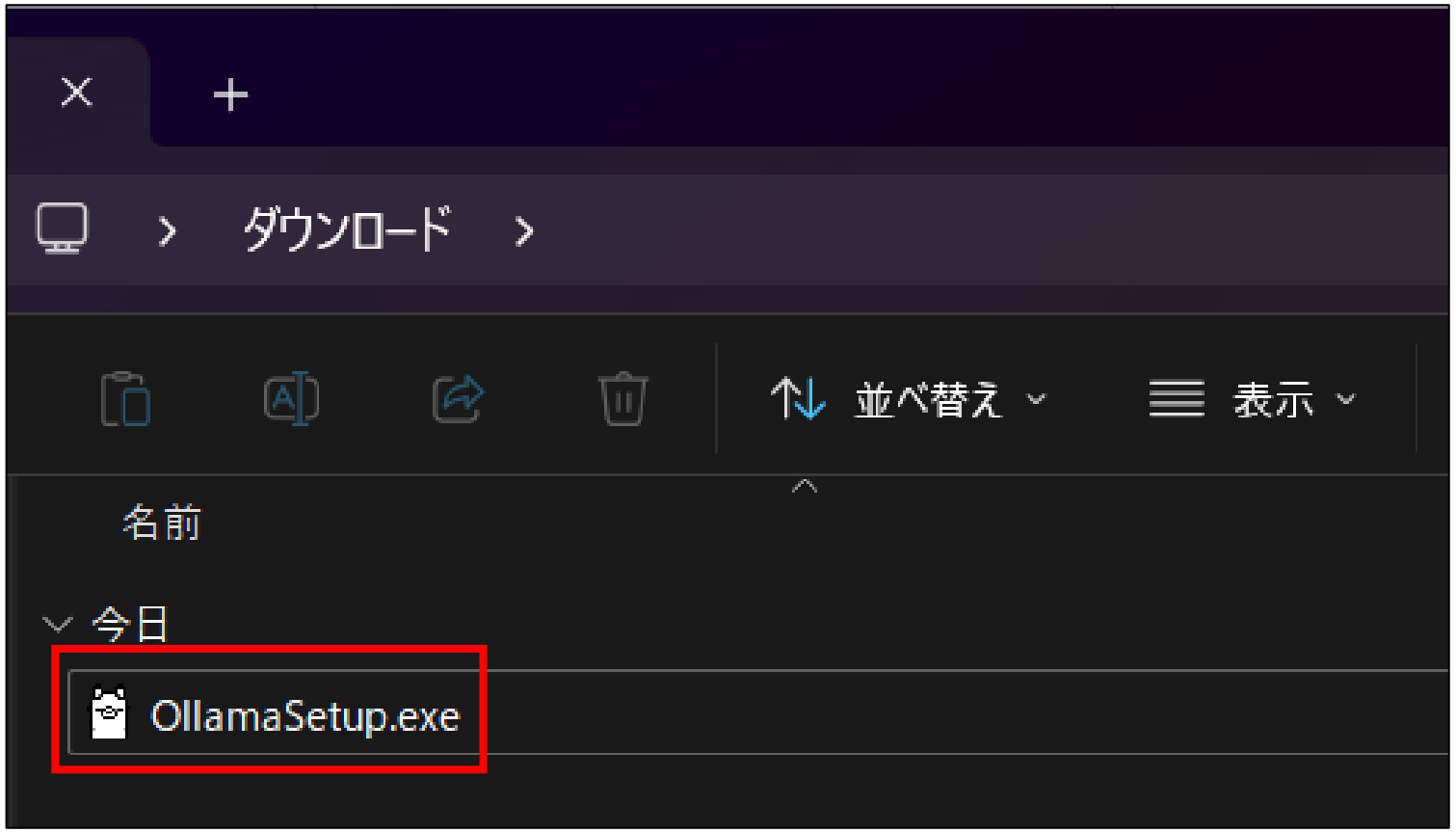
2　ダウンロード


Get up and running with large language models.
Run Llama 3.2, Phi 3, Mistral, Gemma 2, and other models. Customize and create your own.
Download ↓
Available for macOS, Linux, and Windows (preview)

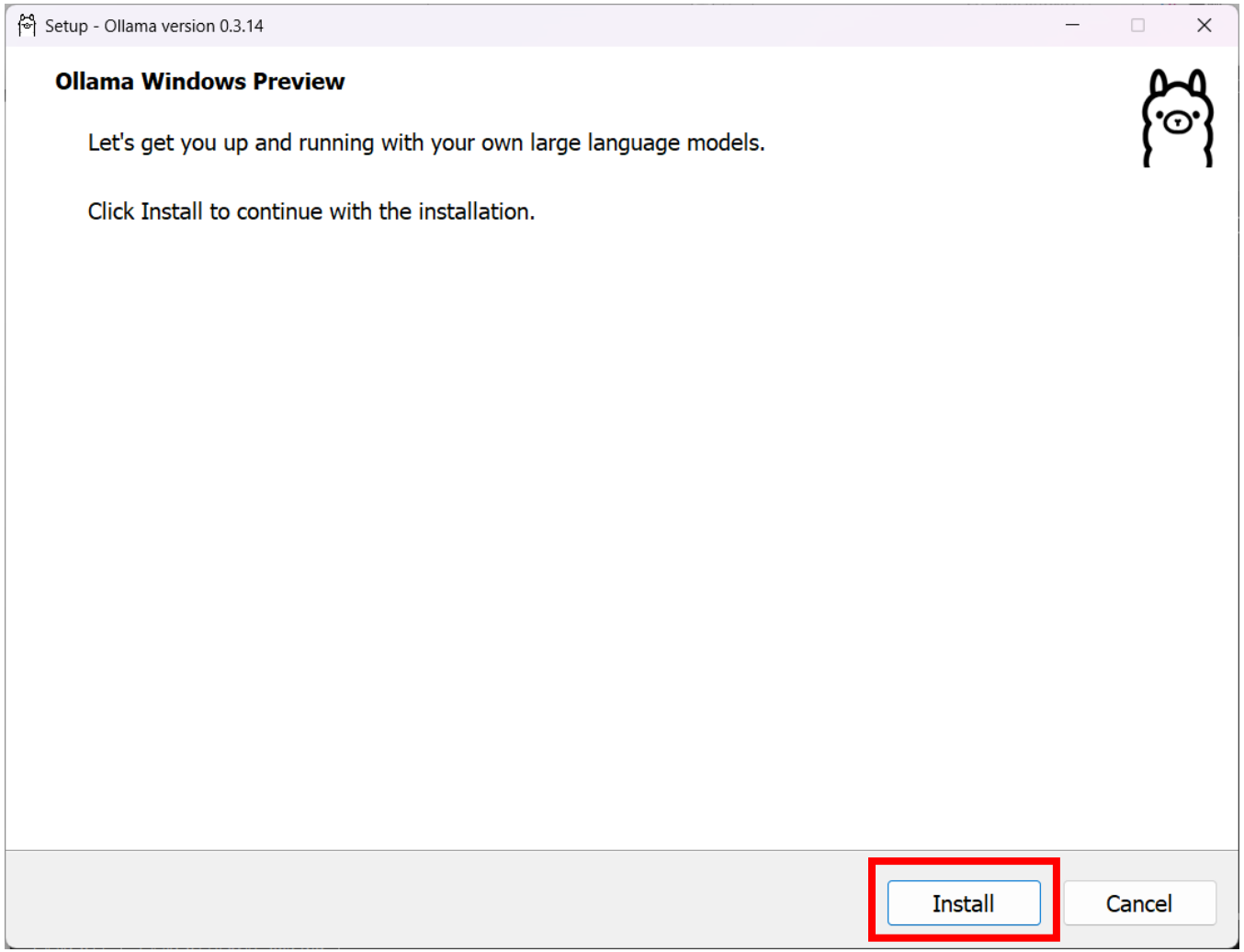

Download Ollama
macOS
Linux
Windows
Download for Windows
Requires Windows 10 or later
While Ollama downloads, sign up to get notified of new updates.
your email address
Get updates

3　インストーラを起動する

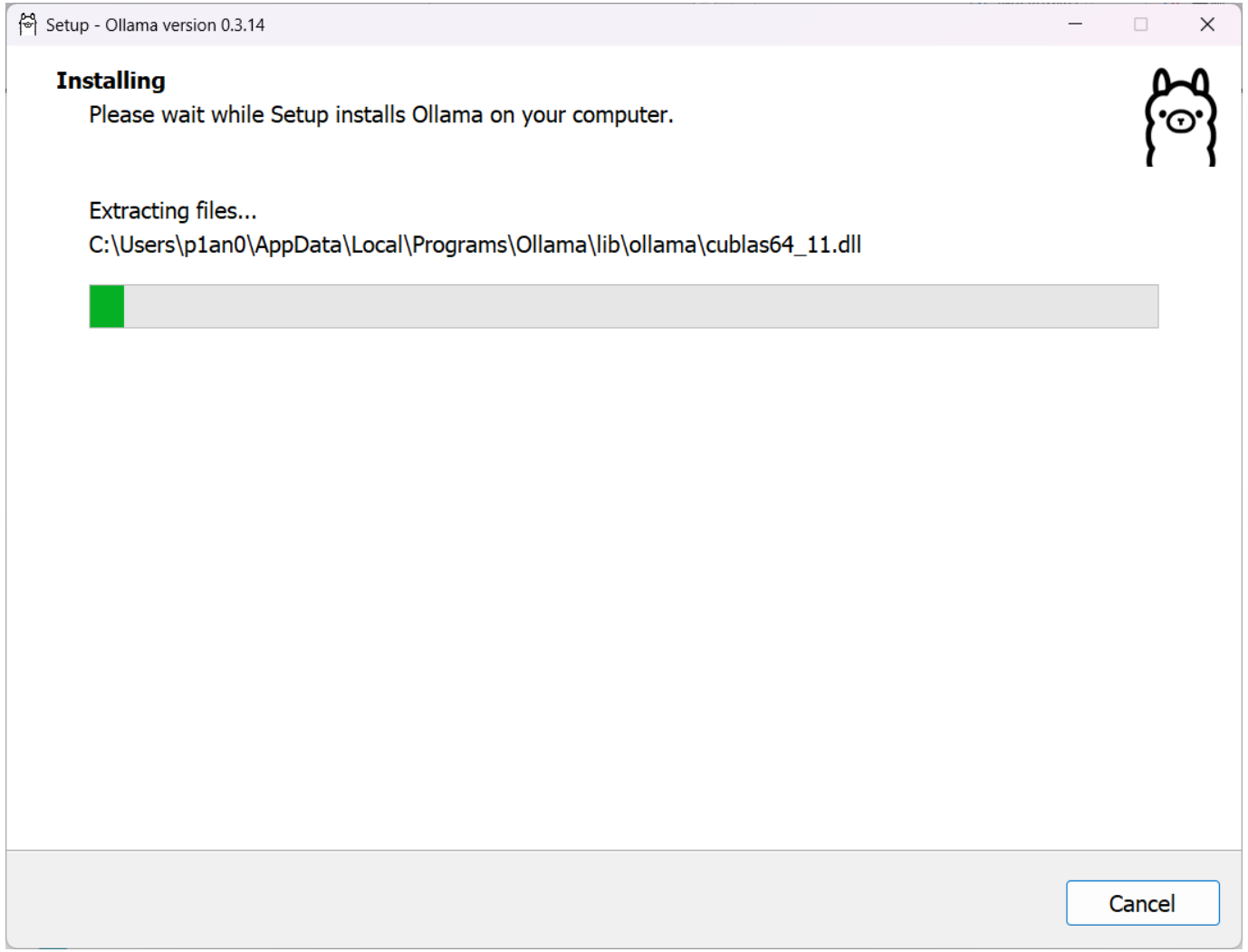
※ダブルクリックで実行



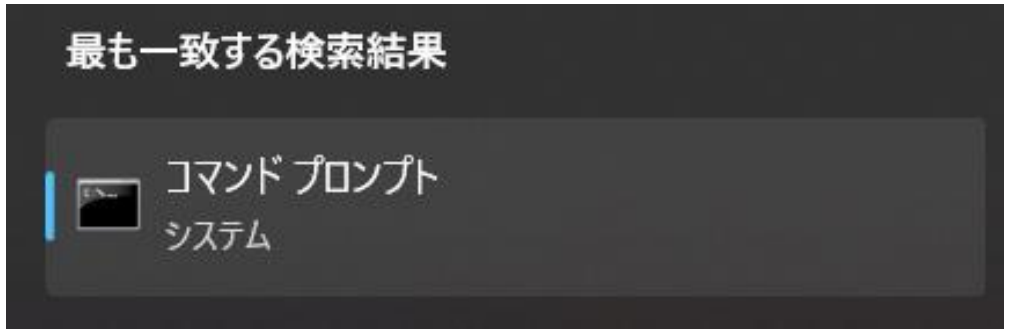
4　インストール①

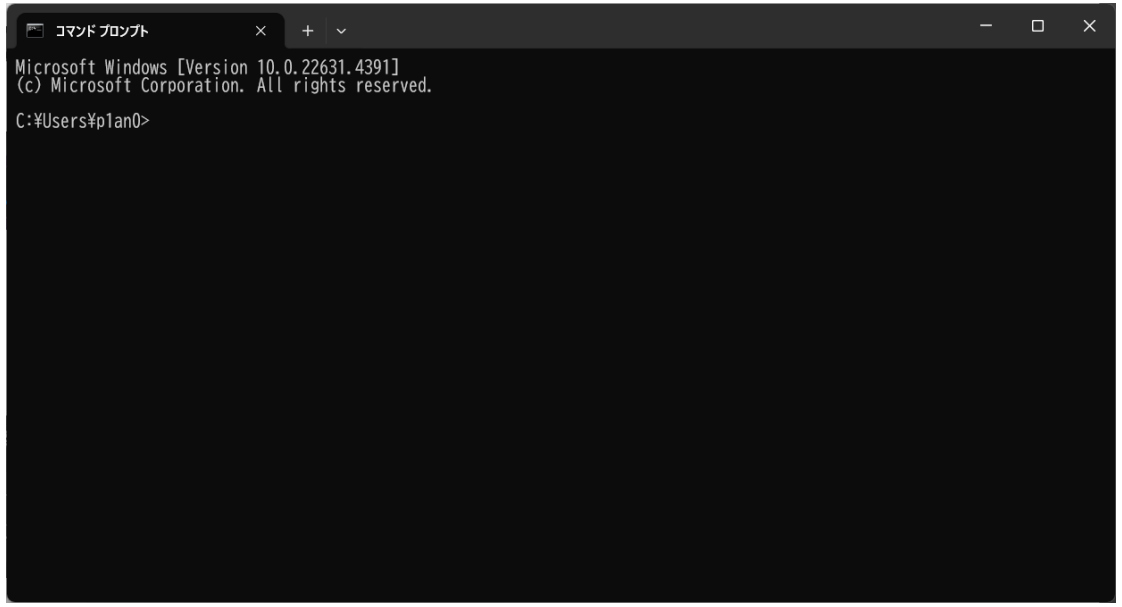

Setup - Ollama version 0.3.14
Ollama Windows Preview
Let's get you up and running with your own large language models.
Click Install to continue with the installation.
Install
Cancel

5　インストール②


Setup - Ollama version 0.3.14
Installing
Please wait while Setup installs Ollama on your computer.
Extracting files...
C:\Users\p1an0\AppData\Local\Programs\Ollama\lib\ollama\cublas64_11.dll
Cancel

6　コマンドプロンプトを開く





※「タスクバーにピン留め」しておくと後々便利

7　動作確認

① C:¥User¥you>　の右側に　**ollama**　と入力してEnterを押す

② 下のような画面が出ればOK

```
C:¥Users¥p1an0>ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve       Start ollama
  create      Create a model from a Modelfile
  show        Show information for a model
  run         Run a model
  stop        Stop a running model
```

Ollamaを使う

大規模言語モデル llama3.2 のダウンロードと実行

1　llama3.2 をダウンロード・実行する

① **ollama run llama3.2**　と入力してEnterを押す

② 下のような画面が出ればOK

※ 最初はモデルのダウンロードが入るため、時間がかかります

```
C:¥Users¥p1an0>ollama run llama3.2
pulling manifest
pulling dde5aa3fc5ff... 100%  2.0 GB
pulling 966de95ca8a6... 100%  1.4 KB
pulling fcc5a6bec9da... 100%  7.7 KB
pulling a70ff7e570d9... 100%  6.0 KB
pulling 56bb8bd477a5... 100%   96 B
pulling 34bb5ab01051... 100%  561 B
verifying sha256 digest
writing manifest
success
>>> Send a message (/? for help)
```

2　llama3.2 と会話する

※ここはChatGPT等と使い方は同じ

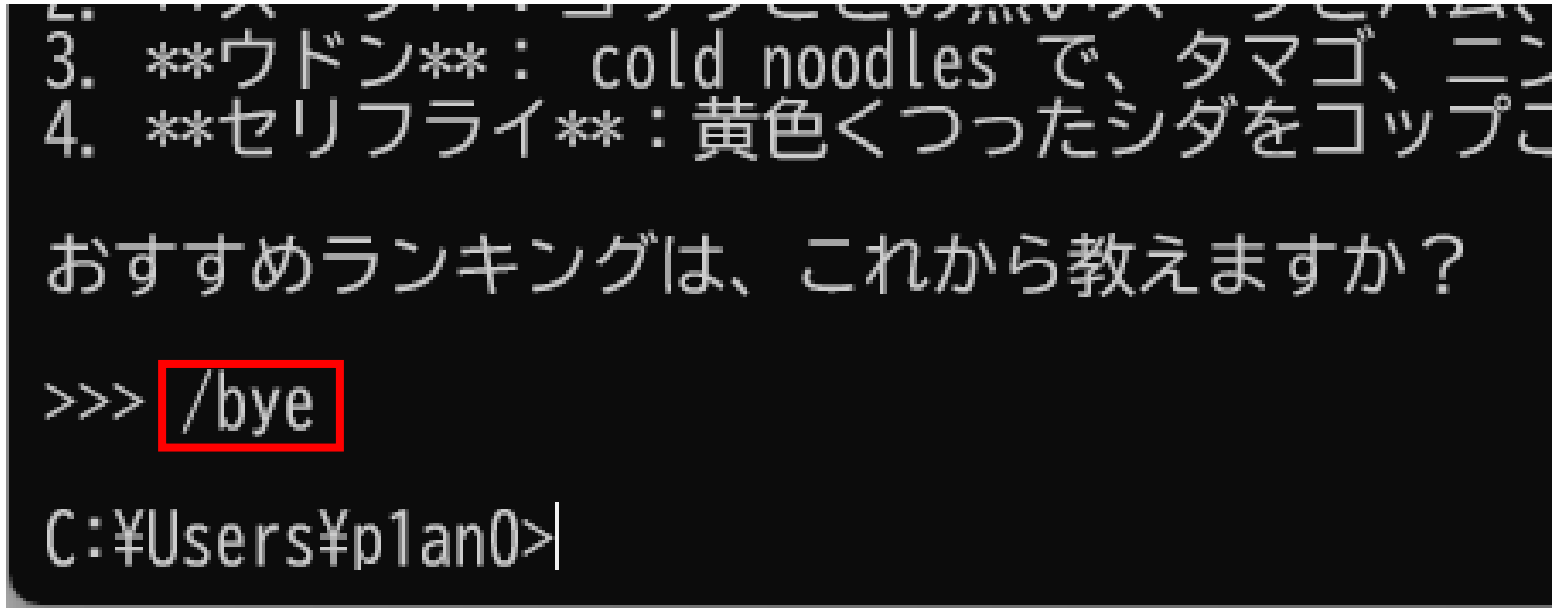
① >>>　の右側に何か入力してEnterを押す

② しばらくすると llama3.2 から返答が返ってくる

```
>>> こんにちは！
こんにちは！どういたしまして？何か相談したいことがありますか？

>>> 今日の晩御飯は何がいいかな？
今日の晩御飯を選ぶのはとても難しいことですね！

以下是一些人気の料理を紹介します。

1. **ラーメン**：たんぱくや牛肉、ひもの種子などを加えた味わいが良いでしょう。
2. **スープ**：コップごとの熱いスープとハム、トマトなどを楽しむのも良いでしょう。
3. **ウドン**： cold noodles で、タマゴ、ニンニク、pêsの種などを加えたみそ汁がอร่อยです。
4. **セリフライ**：黄色くつったシダをコップごとのスープに染めたもので、甘味が強いです。

おすすめランキングは、これから教えますか？

>>> Send a message (/? for help)
```

3　会話を終了する

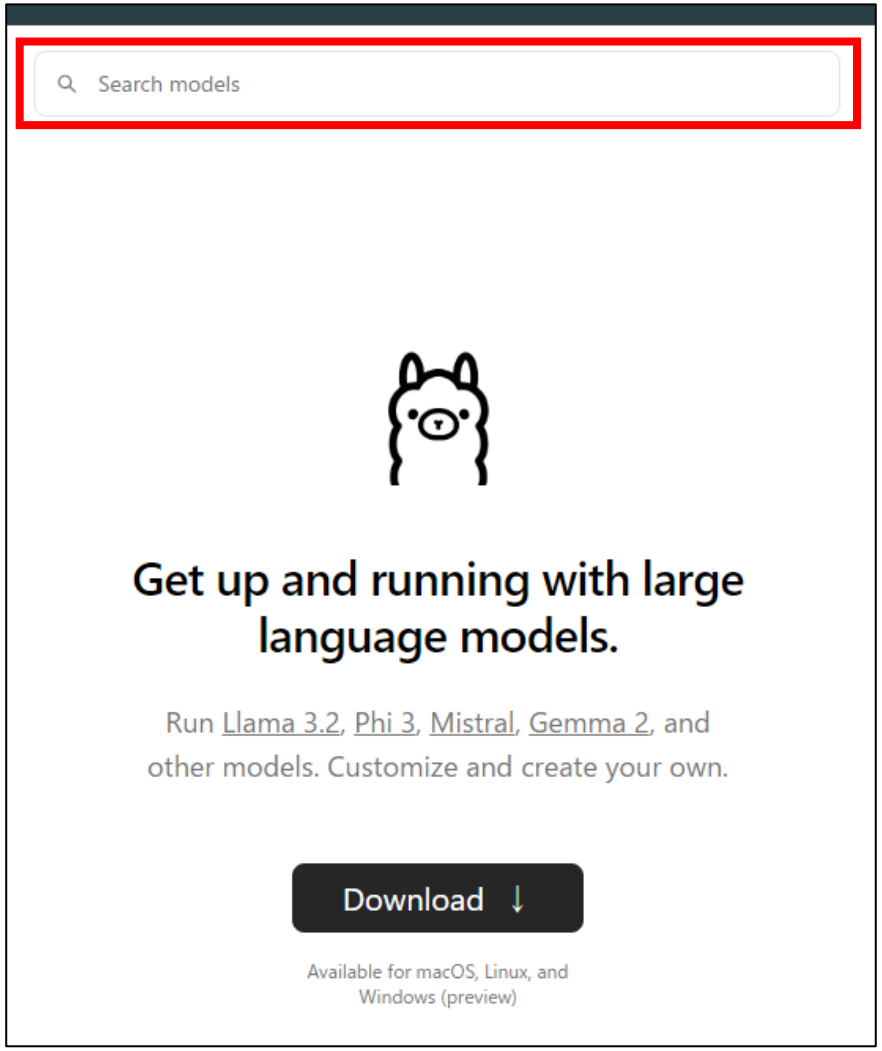
①　**/bye**　と入力（またはCtrlとDを同時に押す）と会話が終了する



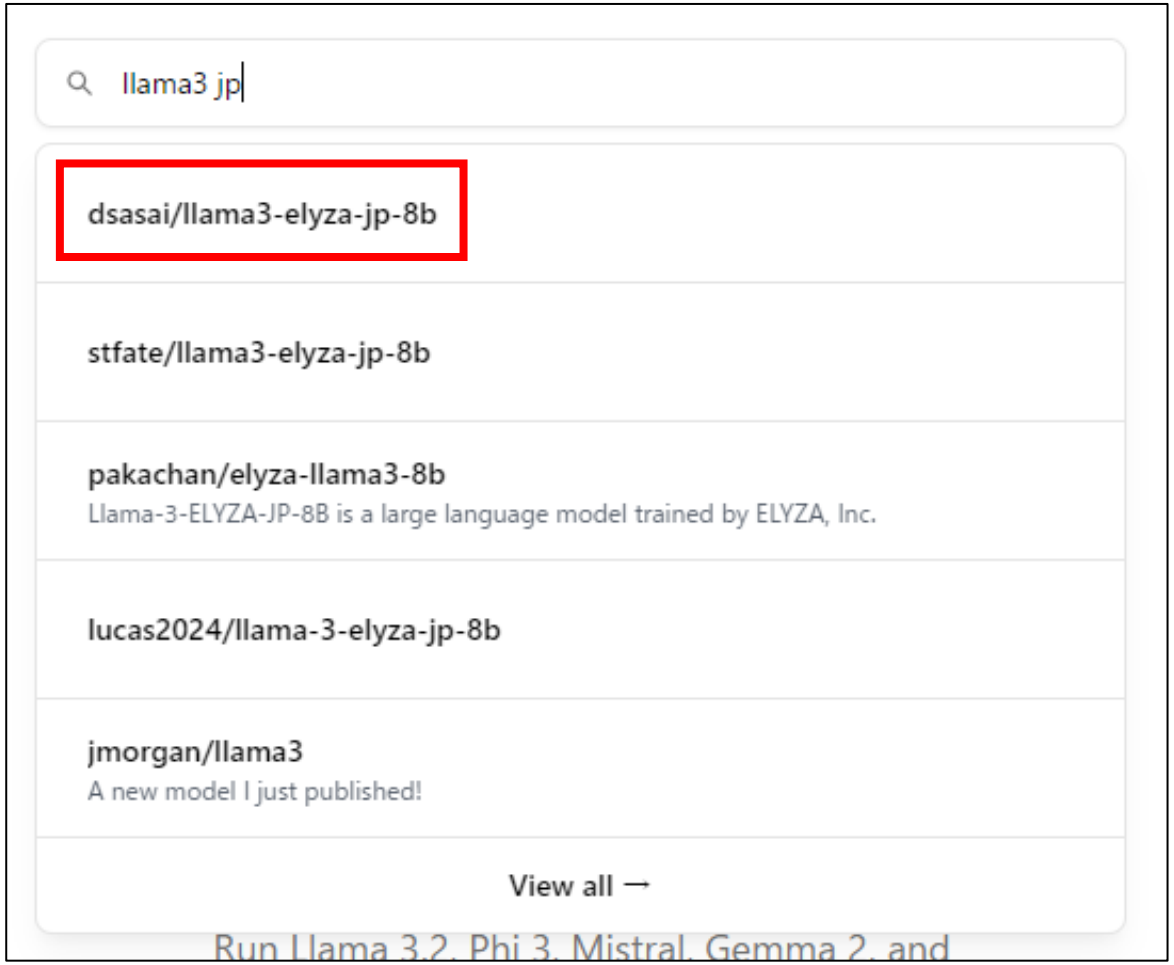
いろんなモデルを使う

日本語特化モデルである llama3-elyza-jp-8b を使ってみる

1　モデルを検索する



今回は　llama3 jp　と入力して検索



2　コマンドをコピーする

3　コマンドプロンプトに張り付けてEnter

※貼り付けはコマンドプロンプト上で右クリック

```
C:¥Users¥p1an0>ollama run dsasai/llama3-elyza-jp-8b
pulling manifest
pulling 91553c45080b... 100%                                    4.9 GB
pulling 7737014881a9... 100%                                    147 B
pulling 8ab4849b038c... 100%                                    254 B
pulling c0aac7c7f00d... 100%                                    128 B
pulling f704a8a590af... 100%                                    487 B
verifying sha256 digest
writing manifest
success
>>> Send a message (/? for help)
```

モデルの一覧表示と削除

PCの容量がいっぱいになる前に、いらないモデルは削除しよう

1　ダウンロードしたモデルの一覧表示

① **ollama ls**　と入力してEnterを押す
② ダウンロードしたモデルの一覧と情報が表示される
※ NAMEの列に書かれているのがモデルの名前

```
C:¥Users¥p1an0>ollama ls
NAME                                ID              SIZE      MODIFIED
dsasai/llama3-elyza-jp-8b:latest    ecfdd92e89f6    4.9 GB    6 minutes ago
llama3.2:latest                     a80c4f17acd5    2.0 GB    46 minutes ago
```

2　モデルを削除する

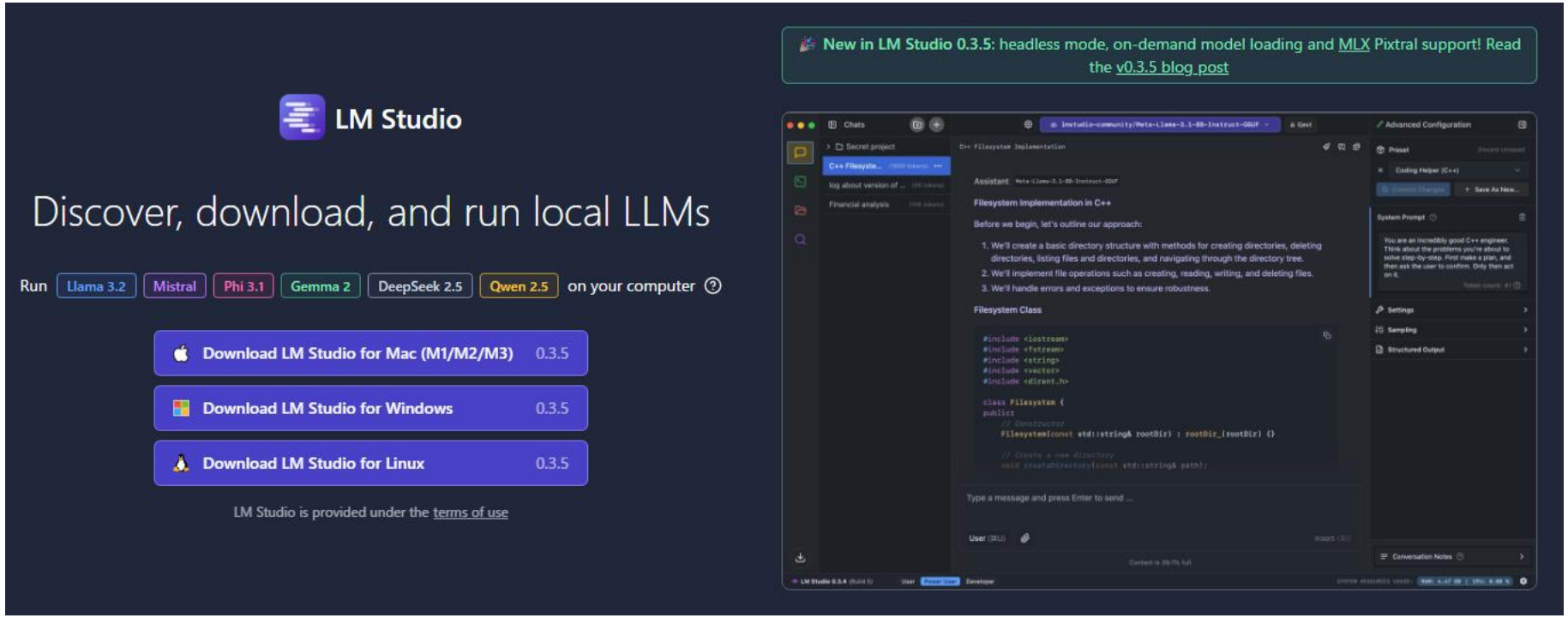
① **ollama rm モデルの名前**　と入力してEnterを押す
② deleted モデルの名前　が表示されたら成功

```
C:¥Users¥p1an0>ollama ls
NAME                                ID              SIZE      MODIFIED
dsasai/llama3-elyza-jp-8b:latest    ecfdd92e89f6    4.9 GB    6 minutes ago
llama3.2:latest                     a80c4f17acd5    2.0 GB    46 minutes ago

C:¥Users¥p1an0>ollama rm llama3.2:latest
deleted 'llama3.2:latest'
```

さいごに

（参考）もう少しリッチな画面が好きな方は

LM Studio



https://lmstudio.ai/

※使い方は下記ウェブサイトなどを参考に
https://note.com/masayuki_abe/n/nd65ed694eec0