

オリジナル質問応答システムの
構築

館野 昌一



写真1 Dockerを利用してラズベリー・パイ上にGPT質問応答システムを構築した

画面は左のPCのものでWindows 11 Proのデスクトップ画面。画面左のリモート・デスクトップにラズベリー・パイの画面があり、その中でターミナルが黒く表示されている。右のブラウザにラズベリー・パイで動いているStreamlitのGPTのアプリケーションが表示されている

本稿では、ラズベリー・パイとGPT-4 Turboを使って、API経由で操作するStreamlitアプリケーション(質問応答システム)を作ります(写真1, 図1)。また、そのアプリケーションをDockerコンテナを用いて起動する方法についても説明します。なお、Dockerは、Windows, Mac, Linuxなど複数のOSで利用可能です。

使用するサービスやツール

● GPT-4 Turbo

OpenAIによって2024年1月25日にリリースされた最新の言語モデルであり、gpt-4-0125-previewとしても知られています。このモデルは、従来のプレビュー・モデルよりもコード生成などのタスクをより効率的に完了し、作業の遅延を減らす特徴があります。また、UTF-8で発生するバグ修正も含まれており、多言語対応が強化されています。このモデルは、特に開発者や研究者にとって、より高度な自然言語処理タスクを実行する上で強力なツールとなるでしょう^{注1}。

注1: <https://forest.watch.impress.co.jp/docs/news/1564658.html>

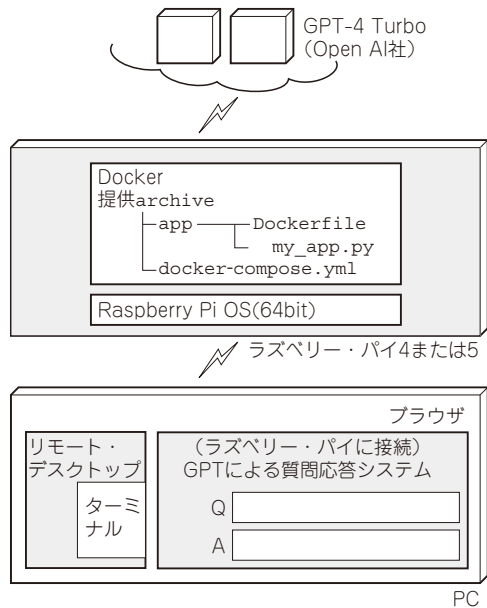


図1 システムの全体構成

● Streamlit

Pythonで記述されたウェブ・アプリケーションを簡単に作成できるオープンソースのライブラリです。データサイエンスや機械学習のプロジェクトでよく利用され、コーディングの知識があれば、複雑でインタラクティブなウェブ・アプリケーションを迅速に開発できます。データの可視化、ユーザ入力を受け付け、リアルタイムのデータ処理など、ウェブ・アプリケーションに求められる多くの機能を短いコードで実装できます。このツールを活用して、GPT-4 Turboを操作するアプリケーションを構築していきます。

ラズベリー・パイのセットアップ

● 準備するもの

あらかじめ準備するものを次に示します。

- ラズベリー・パイ本体(ラズベリー・パイ Pico/Pico Wを除く全てのモデル、本稿ではラズベ