

受け答え式 AI マシュマロマンの開発

岸本 龍羽 大森 稜斗
東 叶愛

1. 研究概要

本研究では、近年急速に発展する人工知能技術、特に大規模言語モデル (LLM) の仕組みと応用を理解することを目的として、対話型キャラクター「AI マシュマロマン」を開発した。

ユーザが自然言語で問いかけると、キャラクターが適切な応答を返す仕組みを構築し、AI の受け答えの仕組みやデバイス連携の技術的要素を実践的に学ぶことを狙いとした。

本システムは IoT デバイス (例: M5Stack 等) を用いて音声やテキスト入力を受け取り、クラウド上の AI モデル (ChatGPT API) に送信して応答を取得する構成をとる。

2. 開発環境

(1) M5Stack Core2 の概要

M5Stack Core2 は、ESP32 を搭載したオールインワン型のマイコン開発モジュールであり、以下の特徴を持つ。

- ESP32 (デュアルコア、Wi-Fi / Bluetooth 対応) 搭載
- 2.0 インチ IPS カラー液晶 (タッチパネル対応)
- 内蔵スピーカ・マイク
- 物理ボタンおよび静電容量式タッチボタン
- 加速度センサ、RTC、バッテリー管理機能
- USB Type-C による給電および書き込み

これらを活用することで、外部回路を最小限に抑えた UI・UX 設計が可能である。

活用できる環境も幅広く以下のようなものがある。

開発言語

- C++ (Arduino フレームワーク)
- 必要に応じて Micro Python も検討可

統合開発環境 (IDE)

- Arduino IDE
- Platform IO (VS Code 拡張)

主要ライブラリ

- M5Unified / M5Core2
- WiFiClientSecure
- HTTP Client
- Arduino Json (OpenAI API レスポンス解析用)

(2) 利用した開発環境

M5Stack Core2 を用いて、ChatGPT と連携する対話型 AI デバイスを構築した。Arduino IDE と Platform IO を用いて環境を整え、ディスプレイやタッチパネル、スピーカを統合的に活用できる仕組みとした。Wi-Fi を介して OpenAI と通信し、取得した応答を画面に表示することで、ユーザとの継続的な対話を実現している。タッチ操作による入力とディスプレイ表示を組み合わせ、単体で完結する AI インタフェースとして構築した。

3. 研究の具体的内容

本研究では、M5Stack Core2 (以下 M5Stack) と ChatGPT API を組み合わせ、会話に応じて反応する「受け答え式 AI マシュマロマン」を開発した。

マシュマロマン本体は、M5Stack のディスプレイにキャラクターの顔を表示し、音声入力・応答生成・発話の一連の流れを自動で行

うインタラクティブデバイスとして設計した。

本システムは以下の3つのモジュールから構成されている。

(1) 全体構成

(ア) 音声入力モジュール (M5Stack マイク)

M5Stack Core2 に搭載されたマイクから人の声を取得し、音声データを解析してテキスト化する。

音声認識は外部サービスを利用して行い、認識結果を M5Stack から ChatGPT に送信する。

(イ) 応答生成モジュール (ChatGPT API)

音声認識で得られたテキストを ChatGPT に送信し、適切な返答を生成させる。

返答は JSON 形式のデータとして M5Stack に返される。

(ウ) 表情表示・発話モジュール

(M5Stack ディスプレイ & スピーカ)

ChatGPT の返答テキストを M5Stack のスピーカで音声出力する。

返答内容に応じてキャラクタの表情を変える仕組みを導入予定であったが、現段階では、図1のようなサンプルプログラムの表情パターンを利用して代用している。

ディスプレイには目や口のアニメーションを表示し、会話しているような演出を行う。



図1 通常時の表情

(エ) キャラクタ性制御モジュール

(ChatGPT プロンプト設計)

本研究では特に ChatGPT の話し方・性格の研究および調整を重点的に行った。

キャラクタとして一貫した個性を持たせるため、プロンプト設計を複数パターン試行し、応答の口調・語尾・テンション・優しさなどが変化する仕組みを研究した。

(オ) 外装ボディ製作 (3D プリンタ)

M5Stack が内部にぴったり収まるマシュマロマン型の外装を 3D モデリングし、STL 方式の 3D プリンタで出力して組み立てた。

外装はディスプレイ部分に M5Stack の画面が見える穴を設け、キャラクタが“顔を出している”ように見えるデザインとした。

(2) ChatGPT の話し方・性格の調整

マシュマロマンが「ただ質問に答える AI」にならず、キャラクタとして魅力を持つよう以下の調整を行った。

(ア) ベースキャラ設定の作成

- ・やわらかい雰囲気
- ・語尾に軽い伸ばし語を付ける
- ・なるべくユーザの気持ちを肯定するなど、マシュマロマンらしい性格をプロンプト内で定義。

(イ) ユーザの発言から感情を推定し、軽く反応させるテスト

「楽しそうだね」「大変そうだね」など、ユーザの気持ちに寄り添う応答を生成できるよう、システムプロンプト内に配慮の方針を設定。

(ウ) 誤認識に強い会話設計

音声認識で多少誤差があっても成立するように、ChatGPT 側の理解力を活かして文脈から補完するようなプロンプトを設計した。

(エ) サンプルプログラムによる表情設計

M5StackBurner (図2) に提供されているサンプルプログラム (図3) を活用し、表示される顔や表情の種類を拡張して視覚表現の多様性を設計した。

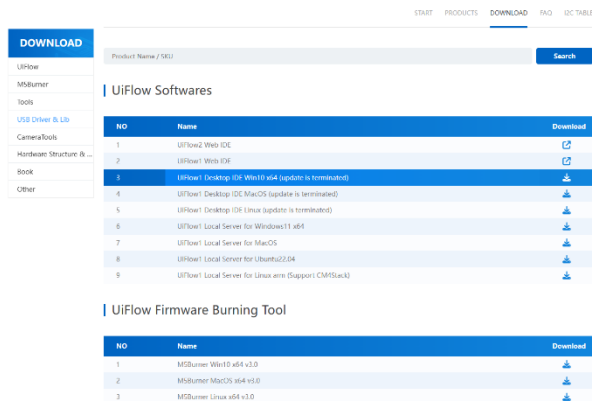


図 2 M5stackBurner

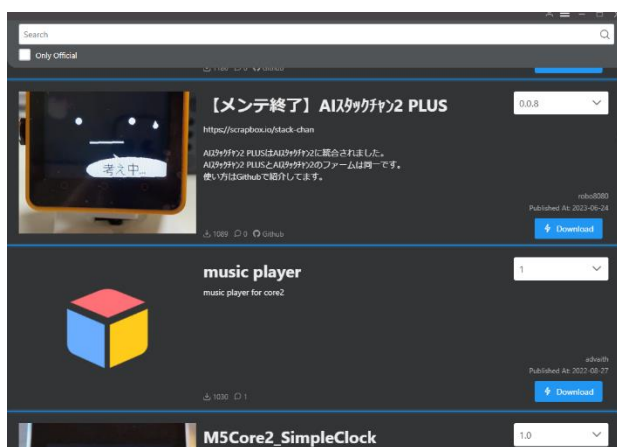


図 3 サンプルプログラムのダウンロード欄

(3) 3D プリント外装（ボディ）の構造

AI マシュマロマンの外装（図 5）は、Fusion360 を用いて設計した独自の 3D モデルで構成されている。外装は上下 2 つの主要パーツに分割されており、この内部に M5Stack Core2 が収納できるように設計した。

(ア) 全体デザイン

ボディは丸みのあるキャラクター形状で、頭部・胴体・腕・脚を一体感のあるフォルムでデザインしている。

マシュマロのように柔らかい印象を出すため、円柱や球体をベースとした滑らかな曲面を多用している。

(イ) M5Stack 収納部（頭部）

M5Stack Core2 は頭部の内部空間に収まるよう寸法調整を行った。

前面には M5Stack のディスプレイ部分が

見える 顔用の開口部（図 4）を設け、表示される表情がそのままキャラクターの顔として見える構造にしている。

(ウ) 上下分割構造

ヘッド部分は上部と下部の 2 つのパーツに分かれており、これにより以下の利点が得られる。

- ・ M5Stack を簡単に取り外し・交換できる
- ・ 内部の調整がしやすい

- ・ 3D プリント時のサポート材を軽減できる

(エ) 組み立て後の特徴

- ・ 頭部の窓から M5Stack の表情が見え、キャラクターに生命感が出る
- ・ 座りポーズの脚により自立しやすく、安定した展示が可能
- ・ 表情アニメーションと 3D ボディが合わさり、インタラクティブで愛嬌のある見た目となる

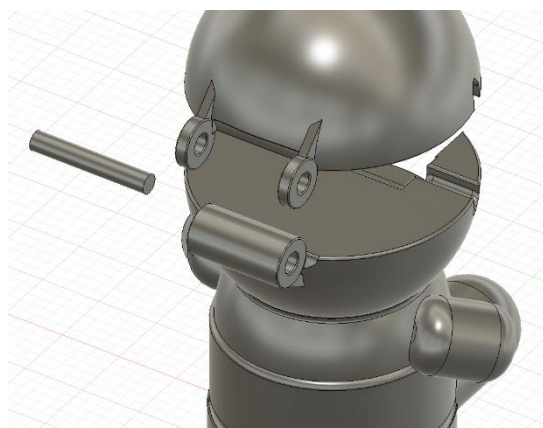


図 4 顔用の開口部

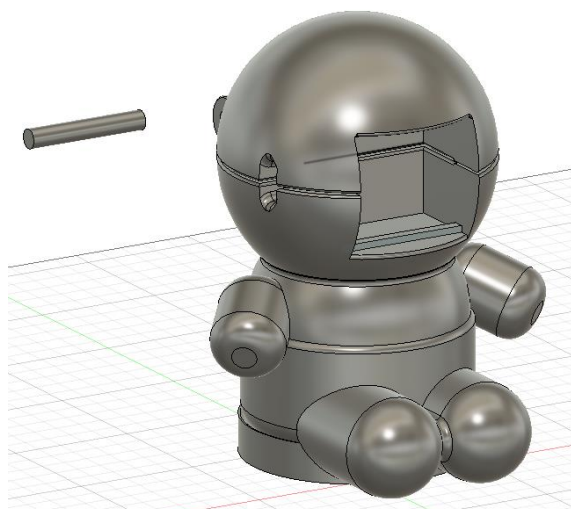


図 5 マシュマロマンの外装

4. 研究のまとめ

本研究では、M5Stack Core2 と ChatGPT を連携させることで、音声やテキストを用いた対話が可能な会話システムの開発を目的として取り組んだ。近年、対話型 AI の発展が進んでいる一方で、それらは主にスマートフォンやパソコン上で利用されることが多い。本研究では、小型の組み込みデバイスである M5 Stack Core2 を用いることで、身近な環境でも対話型 AI を活用できる可能性を検証した。

研究の初期段階では、M5Stack Core2 の基本的な機能を理解するために、M5Stack Burner に用意されているサンプルプログラムを実行し、画面上に顔を表示させる動作確認を行った。この作業により、ディスプレイ表示、タッチ操作、デバイスの安定した動作を確認することができ、今後の開発に必要な基礎知識を得ることができた。また、顔表示のサンプルは、対話内容に応じて表情を変化させるといった拡張の可能性も示していた。

次に、M5Stack Core2 を Wi-Fi に接続し、インターネットを介して ChatGPT の API と通信する仕組みについて検討した。入力したテキストを外部サーバへ送信し、返却された応答を画面表示や音声出力として利用することで、人と自然に近い会話が成立することを確認した。この結果から、M5Stack Core2 のような計算資源が限られたデバイスであっても、クラウド上の AI を活用することで高度な対話処理が可能であることが分かった。

一方で、研究を進める中でいくつかの課題も明らかになった。通信環境によっては応答までに時間がかかることや、音声認識の精度が周囲の環境音に影響されやすい点が挙げられる。また、API の利用回数やセキュリティ面への配慮も、実用化を考える上で重要な課題である。

本研究を通して、ハードウェアとクラウド AI を組み合わせることで、単体では実現が難しい高度な機能を実装できることを学んだ。

今後は、会話の文脈を保持する機能の追加や、AI の応答内容に応じて顔の表情を変化させる機能の実装、さらに操作性の向上を図ることで、より実用性の高い対話型デバイスへ発展させていきたいと考えている。

5. 参考文献

M5Core2 Arduino ライブラリ (公式 GitHub)
Core2 のハードウェア仕様やライブラリ情報
<https://github.com/m5stack/M5Core2GitHub>

OpenAI ESP32 向け Arduino ライブラリ例 (GitHub) - ESP32 で OpenAI API を使うためのライブラリ

<https://github.com/me-no-dev/OpenAI-ESP32> GitHub

M5Stack で ChatGPT API を使ってみた！

<https://www.techlife-hacking.com/?p=2532>