

研究者 電気情報科 青木 一希  
飯島 蒼翔  
徳田 悠斗  
村尾 晴輝  
指導者 大工原 裕之

## 1 研究目的

人工知能 (AI) に関する理解を深めるために、日本語で対話できる AI の作成をテーマにした。現在、多くの AI 関連技術は、プログラミング言語の一つである Python を使って実装できるようになっている。そこで、授業で学ぶ機会がなかった Python の基本を習得するとともに、オリジナル人工知能の作成を経験し、周辺技術を含めて高度な技術を応用できるようになることを目的とした。

## 2 研究内容

### (1) 音声入力

音声通信ソフトの「Discord」内にある「Discord bot」を使用した。音声を記録・変換する目的で、フリーソフトウェアである「FFmpeg」を使用することとした。



図1 Discord bot 使用

### (2) 文字起こし

複数ある Python の音声認識ライブラリやソフトを検討し、「Whisper」を使用することとした。(1)の音声入力で記録した音声、日本語テキストに変換する。

### (3) AI 本体

AI 本体には、ELYZA-japanese-Llama-2-7b を使用することとした。AI のベースモデルは、Meta 社が開発し、オープンソースである Llama2 になる。

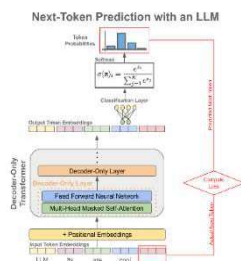


図2 Llama2 の構造

### (4) テキスト読み上げ

AI の返答を日本語音声として出力するために TTS (Text to Speech) を使用した。TTS はオープンソースの Python ライブラリである。

### (5) Live2D のモデル

「Unity と Live2D を用いたフェイストラッキングの制作」研究班に作成の依頼をした。

## 3 成果

音声入力では、FFmpeg を開発環境にインストールし、「Discord」の使用環境を整え、文字起こしでは「whisper」をインストールし、録音した短いテスト用音声や長い会議音声を日本語テキストに変換することに成功した。人工知能本体では、人工知能の追加学習や強化学習で使用可能な文字データセットを JSON Lines (.jsonl) 形式で作成し、Google Colab での追加学習を行った。また、使用している開発環境 (GPU: RTX GeForce 4060 8GB) で学習済みの人工知能 Llama 2 を実行することに成功したが、質問に対し、関係のない返答を出力することがあった。実行中にパラメータを調整することで解決した。

テキスト読み上げでは、見本の音声モデルを利用し、文字起こしで出力させた日本語文章を読み上げることに成功した。

## 4 今後の展望

「Unity と Live2D を用いたフェイストラッキングの制作」班に依頼した Live2D をファインチューニング済みの AI に結合し、「Discord」での音声録音による文字起こしを実装し、実際に運用する。

## 5 反省・感想

これまでの研究で、AI の仕組みや学習方法を実際に体験することができ、AI の開発に関する知識や技術を大きく深めることができた。Python や C++、Google Colab などを使いながら、AI モデルのトレーニングや音声処理の仕組みの理解も進んできた。特に、AI に性格や個性を持たせる学習や、音声による自然な対話を実現する部分では、多くの試行錯誤を重ねてきた結果、AI に関する知識が身についた。

一方、反省点として、スケジュール管理が十分にできていなかったことが挙げられる。学習データの準備やモデルのトレーニングには予想以上の時間がかかり、進捗が遅れ、その他の機能の開発に取り組むことができなかった。