

日本 カナダ 国際共同研究 「Well Being な高齢化のための AI 技術」 2023 年度 年次報告書	
研究課題名（和文）	FureAI: 社会的会話エージェントとスマート行動モニタリングによる高齢者生活支援プラットフォームの実現
研究課題名（英文）	FureAI: An Implementation of An Elders' Living Support System based on Social Conversational Agents and Smart Activity Monitoring
日本側研究代表者氏名	伊藤孝行
所属・役職	京都大学 情報学研究科・教授
研究期間	2023 年 11 月 1 日 ~ 2026 年 10 月 30 日

1. 日本側の研究実施体制

氏名	所属機関・部局・役職	役割
伊藤孝行	京都大学・大学院情報学研究科・教授	研究開発の推進
桑原英人	Agreebit 株式会社	研究開発の補助と検証

2. 日本側研究チームの研究目標及び計画概要

独居の高齢者との会話エージェントが実際に会話を実現するための会話ソフトウェアモジュールを実現する。会話の中でも「相談」のタイプの会話を実現することを考える。相談では、単に発言するのではなく、傾聴することも重要であり、その点にも着目する。

会話ソフトウェアモジュールは、大規模言語モデル(LLM)を用いる。基本的には、LLM を用いて、ファインチューニングやプロンプトによって高齢者むけの会話ができるようなソフトウェアモジュールを実現する。ファインチューニングとしては、OpenAI の GPT やオープンソースの LLM(例えば、LLaMa2)を用い、高齢者向けの会話のためのファインチューニングを行う。

プロンプトについては、上記の LLM に対して、ロールプレイのプロンプトを利用しながら、Chain of Thought、Graph of Thought といった技術に基づいてプロンプトを作成する。訓練には、すでに伊藤研究室に装備している A100 GPU もしくは AWS などのクラウドで提供される GPU モジュールによって行う。

予測される研究推進上の困難としては、ファインチューニングの精度が十分に得られるかという課題があるが、プロンプトエンジニアリングで対応する。

これらの LLM による会話モジュールについては、すでに現在進行中の CREST プロジェクトで技術を持っているので実現上問題はないと想定しているが、最低限ルールベースの会話モジュールも実現できる。

英語と日本語で実装し、翻訳モジュールを利用する。

スケジュールは以下の通りである。

2023/11-12 インストラクションの作成とファインチューニング

2024/1-3(及び 2024/4-9) プロンプトの作成と会話モジュールの実現(基本的には 2024 年度で終了予定であるが、ソフトウェア開発などで想定外の遅延が起こることも考え、2024 年度にもスケジュールを組み込んでいる)

3. 日本側研究チームの実施概要

2023 年度は、初年度であることもあり、プロジェクトの立ち上げ、および、カナダ側との書類の調整等にエフォートを割いたが、滞りなく進んだ。

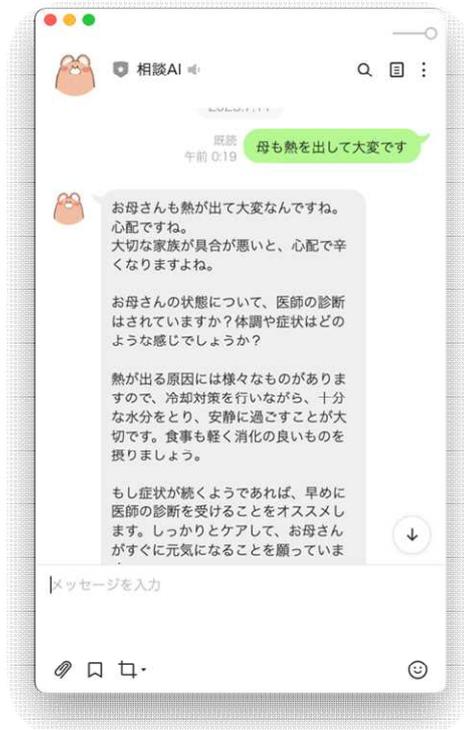
研究では、独居の高齢者との会話エージェントが実際に会話を実現するための会話ソフトウェアモジュールを試作した。会話の中でも「相談」のタイプの会話を実現することを考える。相談では、単に発言するのではなく、傾聴することも重要であり、その点にも着目する。

会話ソフトウェアモジュールは、大規模言語モデル(LLM)を用いる。基本的には、LLM を用いて、ファインチューニングやプロンプトによって高齢者むけの会話ができるようなソフトウェアモジュールを実現した。

計画で述べたとおり、プロンプトの作成やファインチューニングについてさまざまな手法を試みた。期間中に新しい LLM が次々に公開され、その周辺技術も次々にアップデートされたが、フォローアップはできる限り行っている。以下の 2 点について具体的な成果として挙げられる。

【プロンプトによる寄り添い相談エージェント】
プロンプトを調整したエージェントをいくつか試作した。インターフェースに LINE を用いて、ユーザと簡単にインタラクションをするエージェントをいくつか試作した。プロンプトを調整し、OpenAI の GPT3.5 や GPT4 を利用して、ユーザの相談に対して受け答えする簡単なエージェントを作成した。数名の被験者とやりとりをしたが、真剣にやりとりをするユーザもみられた（右図）。ただし、エージェントの回答をうまくまとめる方法や、過去の話をもっとまとめる機能がない点、さらには、「傾聴」をうまくする方法が必要と認識している。特に、発言と傾聴を調和させることがまだ十分には実現できておらず、これらの機能を実現するためには、LLM だけではなく、それを制御するプログラムが必要と考えており、LangChain もしくは同等のフレームワークを実装中である。

【LlaMa2 の日本語ファインチューニング】
ファインチューニングとしてはオープンソースの LLM(例えば、LLaMa2)を用い、高齢者向けに、まずは日本語によるファインチューニングを行った。52000 のインストラクションセットによって LlaMa2 をファインチューニングを行った。簡単なテストでは、予想以上に良い回答を得られている。プロジェクトで作成するソフトウェアシステ



ムに導入することができる。最終的なシステムでは個人情報扱う必要があるため、ローカルな LLM を実現することは大変有用性が高い。ただし、現有設備では、一度ファインチューニングするのに 1 ヶ月程度かかるため、今後さらに設備を増やすのと、AWS などのクラウドを利用することを考えている。

以上の主な 2 つの実装以外にも、同時翻訳エージェントも DeepL の API を用いて実装し、意見交換プラットフォーム（掲示板システム）に接続などを行っている。

カナダ側との協力活動については以下を行なった。

2023 年 11 月 29 日に、カナダ側のチームと日本側のチームが zoom で一同に会し、自己紹介と今後の計画について話し合った。2024 年 3 月 12 日にも、各チームが一同に会し、研究進捗や方向性について話し合った。また、その後メール交換により、2024 年度に、カナダチームの日本訪問、および日本チームのカナダ訪問を行い、夏から秋の期間にそれぞれシンポジウムを開催する日程調整も行なっている。