バーチャルエージェントによるセルフケアサポート:検索拡張生成を用いた生成 AI システム

1265063　政影 亮太
Ryota Masakage

How virtual agents can help you take care of yourself: A new generative AI system using RAG

【背景・目的】2022 年以降、ChatGPT の登場を契機に生成 AI の発展が著しく、大規模言語モデル(LLMs)の利活用が脚光を浴びている。しかし LLMs は、事実に基づかない情報の生成、時代遅れの知識の提示、ならびに後検証ができない推論プロセスを行うなどの課題を抱えている。このソリューションの一つとして、検索拡張生成(RAG)機能の導入がある。RAG を用いて、LLMs の機能と外部データベースの情報を統合することにより、モデルの精度と信頼性を向上させることができるからだ。そこで本研究では、RAG を実装した 3 次元 AI システム(AI アバターシステム)によるセルフケアサポートシステムのプロトタイピング、機能検証を行うことを目指した。
【検証方法】心身状態の指標となるバイタルデータの扱いにおいては、正確性が最重要となることを鑑み、ナレッジグラフ利用の有効性を検討した。RAG 方式のデータ検索方法において、ナレッジグラフ(KG)の効果を、主要な他の 2 つのアプローチとの比較により検証した。対象となるデータは、特定日付のバイタル情報を含む csv 形式のデータセットである。以下の 3 つの手法を採用し比較分析を行った。①ベクトル検索(Vector)：日付に基づいて分割されたバイタルデータを embedding モデルを用いてベクトル化し、コサイン類似度から検索を実施する。具体的には、OpenAI の text-embedding-ada モデルを活用して分割データと結び付けられたベクトルを生成・保管、検索時には同モデルを使用しコサイン類似度を基にデータを抽出する。②データフレーム検索(DF)：csv データを表形式とみなし、質問を gpt-3.5-turbo で Python コードに変換して実行することで検索を行う。③ナレッジグラフ検索(KG)：日付ごとに分割されたバイタルデータに基づくトリプレット[単語、関係、単語]を gpt-4-preview モデルから作成・保管。検索では質問に同モデルを用いて単語分割し、単語からトリプレットを照合する。【評価方法】異なる時間単位（一日、一週間、二週間、三週間）で必要となる情報を含む質問を GPT-4 で作成し、これらに対して実施された各 20 ずつ 80 の質問において、検索システムの性能を 3 つの指標（回答精度、検索の安定性、検索時間）で評価した。図 1 は、これら指標に関する総合的な評価結果を提示している。
【結果】異なる時間単位（一日、一週間、二週間、三週間）で必要となる情報を含む質問を GPT-4 で作成し、これらに対して実施された各 20 ずつ 80 の質問において、検索システムの性能を 3 つの指標（回答精度、検索の安定性、検索時間）で評価した。図 1 は、これら指標に関する総合的な評価結果を提示している。回答精度は、KG が 99.8%と最も優れた結果を示し、検索の安定性に関しては、Vector と KG が共に 100%という完全な安定性を実現した。一方で、検索時間においては、KG は平均 8.95 秒が必要であることが判明した。

性能評価

	Vector	DF	KG
回答精度(%)	50	44	99.8
検索の安定性(%)	100	85	100
検索時間(s)	0.24	2.58	8.95

図 1. 各 Retriever の検索精度評価

文献

1)LIU, Jerry. LlamaIndex, 11 2022. *URL https://github. com/jerryjliu/llama_index*.