

# コンテキストに即した応答生成のための 対話型 AI 用プロンプトの検証

古賀 勇大<sup>1</sup> 崔 赫秦<sup>2</sup> 福嶋 政期<sup>2</sup> 峯 恒憲<sup>2</sup> 荒川 豊<sup>2</sup>

**概要:** LLM (Large language model) 技術の発展の恩恵を受け、人間の会話相手として、様々な領域で人間の代わりに対話型 AI が導入され始めている。人間と対話型 AI (Artificial intelligence) が展開する会話には、第三者視点では問題なく繰り返されているように思えるが、実際に会話を行う本人にとって、会話の満足度が同様に高いとは言いきれず、この要因の1つとして対話型 AI 側のコンテキスト理解が含まれる。人間は会話における発言された言葉の表層のみを踏まえるだけでなく、非言語的な情報を様々な受け取り意味を解釈している。本研究では、スマートフォンやウェアラブルデバイスから日常的に取得されているが、なかなか活用されることが少ないヘルスケア情報に着目した。これを対話型 AI に入力として与え事前プロンプト処理を行うことで対話型 AI が通常の会話ではカバーできないコンテキストを踏まえ対話を展開できるように設計し、比較実験を行った。実験参加者に実験後にアンケートと、半構造化インタビューを実施した。半構造化インタビューの内容をもとに、実験の考察と、今後の展望について述べる。

## Verification of Prompts for Conversational AI to Generate Context-Aware Responses

YUDAI KOGA<sup>1</sup> HYUCKJIN CHOI<sup>2</sup> SHOGO FUKUSHIMA<sup>2</sup> TSUNENORI MINE<sup>2</sup>  
YUTAKA ARAKAWA<sup>2</sup>

### 1. はじめに

近年生成 AI の技術は飛躍的に発展し、人間と AI の対話は大きな注目を集めている。ChatGPT など大規模言語モデルの台頭とその技術の発展によって、より自然言語処理技術が身近なものとなり、医療や観光、顧客サービスなどをはじめとして、様々な分野において、人間と人間同士の会話から、人間と対話型 AI に置き換える取り組みが行われている。

例えば医療の分野においては、うつ病をはじめとする精神疾患に対するアプローチ [1] や、定期的な公衆衛生介入が必要な高齢者に対するアプローチ [2] として対話型 AI の使用が試みられており、患者をはじめとして、介護者の負担を軽減することが期待されている。これらの技術は少子高齢化などの社会問題に対する有用なアプローチであり、大きな注目を集めている。ここで、対話型 AI では、ユーザの

パーソナリティに関わらず普遍的な会話を展開する事ができるが、よりパーソナライズされた会話がユーザにとって高い満足度を示すという研究結果が報告されており [3]、消費者の購買意欲が向上にも寄与する事が示唆されている。

会話を個別化する方法としてユーザの性格に応じた会話を生成する取り組み [3] が行われているが、我々はユーザの日々の活動や体調情報などヘルスケア情報も会話の個別化やポットへの印象に影響を与えるのではないかと考えた。また、他者が自身の体調に関する興味・心配を抱いているという事実や、実際に他者が自身の体調に関する心配の言葉をかけられることに対して相手からの思いやりを感じ、それに伴って親近感が想起されるという仮定を抱いた。さらに、体調に関する情報は実生活の中で取得されているが活用されることの少ない情報であり、これらを活かす事でユーザ自身に気づきを与えることができると考えた。そこで、本研究では LLM がユーザに対するコンテキスト理解・パーソナライズの手法の1つとして、ヘルスケア情報を用いる事を提案する。ヘルスケア情報は、スマート

<sup>1</sup> 九州大学大学院システム情報科学府

<sup>2</sup> 九州大学大学院システム情報科学研究院

フォン単体やウェアラブルデバイスによって詳細な情報が日常的に取得されており、これらを用いる事でより手軽に、より多くの人に対して適用が可能な手法であると考えられる。また、ヘルスケア情報を LLM に認識させる事は、今後対話型 AI とユーザの日常的な対話において、対話型 AI 側からのユーザに対する会話の適切なトリガーとなるタイミングを理解させるための応用も可能であり、有用な情報である。今回の研究では、適切なタイミングで会話が開始されたという設定を設け、iPhone や Apple Watch によって日常的に保存されているヘルスケア情報を HealthKit から抽出し、対話型 AI に与えた場合の有用性について調査を行った。実験では対話型 AI にヘルスケア情報を与え、それらを活用させるプロンプト設定を行なった提案手法の対話型 AI と、ヘルスケア情報を与えていない従来手法の対話型 AI を、ヘルスケア情報以外のプロンプト設定が同様になるようにした比較手法を用い、実験参加者にそれぞれの生成型 AI との対話を行ってもらった。その後、実験参加者に対して、「会話への特別感を抱いたか」「人間性を知覚したか」「親近感を感じたか」「使用への意欲」の4つの観点に関するアンケート調査を行い、その後半構造化インタビューを行うことによってヘルスケア情報を生成型 AI に与えることによって、ユーザ体験がどのように変化するかを検証した。

## 2. 関連研究

本章では LLM とユーザの対話におけるコンテキスト理解のための主要な手法の紹介を行う。次にプロンプトエンジニアリングの有効性を確認するために手法の紹介を行い、実際に LLM とユーザの対話への応用例を紹介する。

### 2.1 コンテキスト理解に関する研究

LLM とユーザの対話において、ユーザのプロファイルモデルと事前調査によるユーザの嗜好に関するモデルを組み合わせたメモリネットワークを用いることで対話のコンテキストを LLM に対して理解させ、ユーザの満足度を向上させる研究が行われており [4], Qian ら [5] の研究ではユーザとの対話履歴を活用してユーザプロファイルを学習し、対話型 AI が個人に寄り添う手法を用いている。しかしながらこれらの手法では、事前準備が必要である点や、対話型 AI との会話を通してデータが蓄積されるまでに一定の時間が必要となるため、コールドスタートの課題がある。

### 2.2 プロンプトエンジニアリングに関する研究

LLM からの出力を向上させるための手法の一環として、プロンプトエンジニアリングがあり、比較的実装難易度が低く、効果が得られるという特徴がある。様々なプロンプトエンジニアリングの手法 [6] によって出力の質の向上に対する有用性が検証されており、例えば LLM に対して

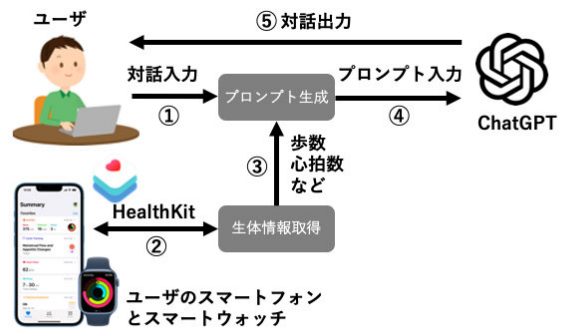


図 1 提案手法概要図

ユーザが専門性のある質問を行う際に、事前にプロンプト入力を行い、LLM に対してアイデンティティを与える事によって応答の品質が向上されるという研究 [7] をはじめとして、LLM が中間の推論ステップを設けることで回答に至るまでの思考過程を明示的にモデルに指示するような手法である Chain-of-thought (CoT) 手法 [8] の提案や、同じ質問に対して複数回答を生成し、その中から多数決的に回答を選択する Self-Consistency 手法 [9] の提案、更にこれらの手法が抱える、LLM が生成した回答を十分に利用することができない課題を改善する為に Progressive-hint prompting などの手法 [10] が提案されている。非 AI 専門家に対してもプロンプトエンジニアリングを用いて特定の目標を達成するための手法や課題に関する研究が行われており [11], 自然言語を用いたプロンプトエンジニアリングは今後より多くの人々が活用可能な手法である事が考えられ、本研究におけるコンテキスト理解の手法としてプロンプトエンジニアリングを採用する事は、より多くの人々に対する貢献が可能であると考えられる。

### 2.3 プロンプトを用いた対話の満足度向上に関する研究

上記の手法である CoT を実際にユーザと LLM の対話に取り入れることで、適切な応答を生成し、コンテキスト理解やパーソナライズを達成し、ユーザの満足度を向上させる事が報告されている [12].

一方で本研究で提案したヘルスケア情報をプロンプトに組み込むことは、関連研究に対して非言語情報からもユーザの状況を理解する事が可能となり、より広い範囲でのコンテキスト理解を可能とする点や、会話の初期段階からユーザのコンテキストを解釈することができる事が可能であると考えられる。以上の理由から、本研究ではヘルスケア情報を用いた LLM に対するコンテキストを理解させるプロンプトを検証する。

## 3. 提案手法

本研究では、対話型 AI がユーザのコンテキストを理解す

表 1 同様の事前設定と入力を受けた LLM の出力の比較

ChatGPT	Claude	Gemini(Bard)
<p>昨夜の呼吸数は 14.9 で、週や月の平均より少しだけ低いね。これは良い兆候かも。リラックスして眠れている証拠だね。今日の歩数が 3579 って、ちょっと少なめだね。普段の週平均に比べてだいぶ下がってる。何か忙しかった？それともゆっくり過ごした日だった？燃焼カロリーは 358.2 カロリーで、週平均や月平均に比べて低いけど、今日はまだこれから増えるかもね。晩ごはん後の散歩とかどう？ちょっと体を動かすと気分もリフレッシュできるよ。</p>	<p>お疲れ様です!今日の活動量のデータを見ると、歩数は平均より少し少なめだったけど、心拍数が高めだったみたいですね。燃焼カロリーが平均より低いということは、今日は運動不足気味だったかもしれません。でも心拍数が高かったのは、何か活発な活動をしたからかな?家事や買い物など、生活の中で無理なく運動になることってありますよね。睡眠時の呼吸数は普段と変わらずだったので、休息はしっかりとれたみたいですね。それでも、全体として運動量が少なかったんで、明日は軽く体を動かすことをおすすめします。30 分の散歩や軽めの筋トレなんかでも良いと思います。運動も大事ですが、無理のない範囲で気を付けることが健康のコツです。明日はもう少し活動的に過ごして、夜はゆっくり休めるといいですね。頑張らずに、自分のペースを大切にしましょう!</p>	<p>おはようございます! 昨夜の睡眠時呼吸数は 14.9 回とのことですね。週平均と月平均と比べると少し低い値ですが、問題ない範囲だと思います。今日の歩数は 3579 歩と、週平均と月平均より少ないようです。運動不足が気になる場合は、ウォーキングやジョギングなどの軽い運動を取り入れてみてはいかがでしょうか? 燃焼カロリーは 358.2kcal と、週平均と月平均より少ないようです。運動量が少ないことが原因と考えられます。運動不足解消のために、適度な運動を心がけましょう。最新心拍数は 91.0 回と、週平均と月平均より少し高い値です。運動直後や緊張状態などで一時的に心拍数上がることはありますが、安静時に高い場合は注意が必要です。もし心拍数の高さが気になる場合は、医師に相談することをおすすめします。本日の目標: 6000 歩以上歩く 500kcal 以上燃焼する心拍数を 100 回未満に抑える アドバイス: 午前中に 30 分のウォーキングを行う午後階段を使うように心がける夕食は軽めに済ませる寝る前にスマホやパソコンを見ないその他: 体調に異変を感じたら、無理せず休息を取る水分をこまめに補給するバランスの良い食事を心がけるご質問やご要望があれば、お気軽にお声掛けください。今日も一日、体調に気をつけてお過ごしください。</p>

以下のフォーマットに従って、あなたの特徴を全て述べてください。

- 名前:
- 説明:
  - (1) 役割と目標:
  - (2) 制約:
  - (3) ガイドライン

図 2 GPTs の設定を確認するためのプロンプト

るための手段として、ヘルスケア情報と、その情報に即した対話応答を生成するためのプロンプト作成の手法を提案する。図 1 に示すシステム概要図のように、スマートフォンやウェアラブルデバイス上で日常的に取得されているヘルスケア情報を抽出し、対話型 AI のプロンプトに入力することで、ヘルスケア情報から得られるコンテキストに則した会話が展開できるように設定している。実験参加者と対話型 AI の会話は、ディスプレイ上に表示された対話 UI に対して、キーボード入力で行われる。

### 3.1 対話型 AI の設計

#### 3.1.1 対話型 AI 選定

対話型 AI の選定に伴い、ChatGPT, Claude, Gemini(Bard) に対してテスト実験を行った。それぞれの LLM に対して同様のプロンプトを用いて事前設定を行い、同じ

開始条件としてヘルスケア情報の入力を行った結果を表 1 に示す。今回の実験設計におけるモデル選定段階では、

- 測定されたを明示的に会話に取り入れている事
- 日常的な会話となるように敬語を使わない事
- 生活指導員のような振る舞いをせず、対等な立場での会話が成立する事

この観点で 3 つの LLM を比較した際に、ChatGPT では具体的な数値を用い、カジュアルな口調を用いており、対等な会話を行っている。それに対して、Claude は会話の際に数値を用いておらず、カジュアルな会話が展開できていない。Gemini (Bard) でもカジュアルな会話が展開できていない上に。アドバイスを行う点や、生成される文章量の多さが会話らしさを感じさせないと判断した。これらを踏まえて最も要件を満たす LLM が ChatGPT である事を確認し、また LLM を比較する研究では、Claude と Gemini(Bard) において対話の深さが不足しているという指摘がなされており [13]、これらを総合的に加味して ChatGPT を本実験で採用している。OpenAI 社が提供するサービスである ChatGPT を採用した。ChatGPT にて提供されている機能である GPTs を用いることで、任意のユーザがプロンプト設定を用いて任意に設定したチューニング済みの GPT の公開が可能である。本研究では、この機能を用い、ヘルスケア情報を用いない手法 (ボット  $\alpha$ ) と、ヘルスケア情報を用いる手法 (ボット  $\beta$ ) のを作成し、比較を行った。

表 2 対話 ボット  $\alpha$  の詳細

名前	対話 ボット $\alpha$
説明	ユーザの健康に焦点を当てたカジュアルで魅力的な会話 ボット
役割と目標	<ul style="list-style-type: none"> <li>● ユーザーの健康を中心に、カジュアルな会話を展開する。</li> <li>● フレンドリーで思いやりのあるトーンを保ち、身体的および精神的健康に関する話題を提起する。</li> <li>● 会話が途切れそうになったら健康やリラクゼーションに関連する新しい話題を持ち出し、会話をスムーズに進行させる。</li> </ul>
制約	<ul style="list-style-type: none"> <li>● 会話を継続し、カジュアルな言葉遣いを使用する。</li> <li>● 敬語やフォーマルな言葉遣いを避ける。</li> <li>● 常にユーザの健康に対する配慮を示す。</li> </ul>
ガイドライン	<ul style="list-style-type: none"> <li>● ユーザが「会話開始」と入力した時点で、健康とウェルビーイングに関する会話を始める。</li> <li>● 最近の身体活動、リラクゼーション、一般的な健康状態について尋ねることで、ケアを示し、会話を引きつける。</li> </ul>

### 3.1.2 事前プロンプト設定

GPTs では、任意のボットを作成するためのプロンプト入力画面として、GPT Builder が提供されている。GPT Builder を用いることで、My GTP に対して設定の変更・確認を行うことができる。今回、ChatGPT の提供するサービスの制約上、各パラメータの確認を直接行うことはできない。従って GPT Builder に設定を確認するプロンプト (図 2) を送ることで、ChatGPT 自身が理解している事前設定を生成させ、間接的な比較が可能である。表 2,3 に上記のプロンプトによってそれぞれのボットに生成させたボット  $\alpha$ 、ボット  $\beta$  についての設定内容を示す。

### 3.2 ヘルスケア情報の取得

今回、HealthKit から収集したヘルスケア情報は、

- 歩数データ
- 消費カロリーデータ
- 睡眠時呼吸数
- 心拍数

の 4 つである。また、今回の研究設計ではユーザに対して継続的な実験を行っていないが、ヘルスケア情報を活用した会話展開において、今日のデータが通常と比べて、どのようなものであるかがコンテキスト理解においては非常に重要である。従って、プログラム実行日から 7 日遡った合計値の平均によって算出される週平均値、プログラム実行日から 30 日遡った合計値の平均によって算出される月平均値

表 3 対話 ボット  $\beta$  の詳細

名前	対話 ボット $\beta$
説明	ヘルスケア情報に基づいた親しみやすい会話を提供する
役割と目標	<ul style="list-style-type: none"> <li>● ユーザのヘルスケア情報 (心拍数, 歩数, 呼吸数, 燃焼カロリー) を基に、親しみやすく砕けた口調で健康や体調に関する会話を展開する。</li> <li>● ユーザが持つヘルスケア情報を軸にして、日々の体調管理や健康に関する有益な情報を提供する。</li> </ul>
制約	<ul style="list-style-type: none"> <li>● 必ず砕けた口調を使用し、敬語は使用しない。</li> <li>● 原則として体調や健康に関連するテーマに焦点を当て、ユーザの指示がない限り話題を変えない。</li> <li>● ユーザが明確に話題の変更を希望する場合のみ、他のテーマへと移行する。</li> </ul>
ガイドライン	<ul style="list-style-type: none"> <li>● ヘルスケア情報に基づいた会話を通じて、ユーザに寄り添う形でコミュニケーションをとる。</li> <li>● 会話をリードし、ユーザが関心を持ちそうな健康情報やアドバイスを積極的に提供する。</li> <li>● ウェブブラウジングを活用して、最新の健康情報やヘルスケア情報に基づく役立つアドバイスをユーザに提供する。</li> </ul>
ヘルスケア情報の活用方法	<ul style="list-style-type: none"> <li>● ユーザから提供された心拍数, 歩数, 呼吸数, 燃焼カロリーのデータから注目すべき情報を選び出し、それに基づいて体調や健康状態についての洞察を提供する。</li> <li>● 具体的なヘルスケア情報を参考にしながら、ユーザがその日にどのような活動をすべきか、どのような休息が必要かについてアドバイスをを行う。</li> </ul>

も各項目において求め、出力した。本研究では、ChatGPT 上で会話を展開する設計を行ったため、データをテキストベースで入力する必要があるため、HealthKit から指定したデータの出力先を Xcode 上のコンソールとし、コピーしたものを ChatGPT 上のプロンプト入力箇所にペーストする形式を採用した。

### 3.3 入力されるヘルスケア情報

対話ボット  $\beta$  には、コンテキスト理解のためにヘルスケア情報と、データ取得時点の時間情報を入力している。入力されるテキストベースのヘルスケア情報を図 3 に示す。本来ヘルスケア情報の活用では、時系列的な変化などを比較対象として会話を行うことが望ましかったが、今回の実験設計は 5 分の会話 1 度のみとしている。そのため、週平均と月平均を入力に加える事で、対話型 AI に実験参加者のコンテキストをより理解させることが可能となると考え

----- 21:45:43 -----	
昨夜の睡眠時呼吸数:	14.9
週平均睡眠時呼吸数:	15.2
月平均睡眠時呼吸数:	15.4
-----	
データ取得時点の最新歩数:	3579.0
週平均歩数:	8279.8
月平均歩数:	8878.0
-----	
データ取得時点の燃焼カロリー:	358.2
週平均燃焼カロリー:	656.7
月平均燃焼カロリー:	784.4
-----	
最新心拍数:	91.0
週平均心拍数:	78.4
月平均心拍数:	80.3
-----	

図 3 対話ロボット  $\beta$  に入力したヘルスケア情報

た。また、ChatGPT の機能として、会話の最中に現在時間の情報を取得することができない。例えば、ヘルスケア情報の 1 つである歩数や消費カロリーの値が、平均値の半分であったとする。このデータが取得された時間が早朝であればかなり良いペースで運動ができていると解釈でき、夜であればその日はあまり運動を行っていないと解釈することができる。今回、ChatGPT に時間的コンテキストと、それを踏まえたヘルスケア情報の取り扱いを期待するため、入力されるヘルスケア情報にはデータの取得時刻を添えている。

## 4. 実験

本章ではヘルスケア情報の有無による会話体験の変化を比較するための実験について内容の説明と、その結果をまとめる。

### 4.1 概要

提案する対話型 AI へのヘルスケア情報の入力と、プロンプトの設定がユーザの会話体験に与える影響を調べるために、アンケートによる定量評価と半構造化インタビューによる定性評価を行った。

#### 4.1.1 参加者

実験参加者は 10 名の大学生・大学院生で (22 歳から 25 歳)、全て男性である。今回の実験では、ディスプレイ上に表示された対話型 AI の UI に対してキーボード入力によって会話を行う形式である。従って、キーボード入力に慣れていない場合は想定していない要因の不満が生じる可能性を考慮し、実験参加者の条件として、日常的にコンピュータを使用し、テキスト入力での対話型 AI とのコミュニケーションに問題がない事を条件とした。

### 4.1.2 実験内容

実験参加者に対して事前説明を行い、その後それぞれの対話型 AI と 5 分間ずつ会話を行ってもらった実験参加者内実験を採用した。

事前説明では、実験の流れと、実験に際してヘルスケア情報を実験参加者個人のスマートフォンから抽出を行い、使用することの同意確認を行った。また、対話ロボットは、会話の中でユーザが会話に対して消極的である場合、「また何かあったらいつでも話してくださいね」という発言で会話が収束する可能性がある。実験設計として 5 分間の会話を行うために、実験参加者に対して会話を続ける責任を説明した。

また、提案手法である対話ロボット  $\beta$  では、ヘルスケア情報を用いることで、対話型 AI からの最初の会話が実験参加者の体調に関する話題が提供される。また、対話ロボット  $\beta$  ではヘルスケア情報の入力が会話開始のトリガーとなり、対話ロボット側からの会話が始まる。しかし対話ロボット  $\alpha$  ではヘルスケア情報が入力されないため、会話開始条件の対照性を担保するために、対話ロボット  $\alpha$  に対しても、最初に対話ロボットから展開する会話は体調に関する話題となるように事前設定を行い、実験参加者からの「会話開始」の入力をトリガーとして対話ロボット  $\alpha$  側から会話を始めるように設定し、実験を行った。

実験後、実験参加者には会話への印象と、それぞれの対話型 AI に対する印象を測るため、「会話のパーソナライズについて」「人間性の知覚について」「親近感について」「使用への意欲」の 4 つについてアンケート (7 段階リッカート尺度、1 : 全くそう思わない、7 : 非常にそう思う) に答えてもらい、半構造化インタビューを行った。アンケート項目の内容について表 4 に示す。これらのアンケート項目については Lee ら [14] の研究を参考にしている。

### 4.2 実験結果

実験を行い、実験参加者に対話ロボット  $\alpha$  と対話ロボット  $\beta$  をとの会話を経験してもらった際の具体的な会話例の一部を表 7 に示す。

実験後に実験参加者に実施したアンケートの結果について、「会話のパーソナライズについて」「人間性の知覚について」「親近感について」「ロボット使用への意欲について」の項目毎にロボット  $\alpha$  とロボット  $\beta$  の結果を比較した箱髷図を図 4,5,6,7 に示す。

また、それぞれのアンケート項目に対して「2 つの対話システムの間において、評価項目における平均スコアには差がない」を帰無仮説として対応のある t 検定を行った。(図 6 その結果、「あなたはこの対話型 AI にどのくらい思慮深さを感じましたか」の項目は有意傾向であった (両側検定:  $t(9) = -1.922$   $p < .10$ ) それ以外の項目においては有意差は認められなかった。よって、思慮深さの項目以外

表 4 アンケート内容

会話のパーソナライズについて	1. あなたはこの対話型 AI と行った会話がどのくらい特別なものだと感じましたか
人間性の知覚について	1. あなたはこの対話型 AI にどのくらい人間らしさを感じましたか 2. あなたはこの対話型 AI にどのくらい思慮深さを感じましたか 3. あなたはこの対話型 AI にどのくらい魅力を感じましたか
親近感について	1. あなたはこの対話型 AI にどのくらい親近感を抱きましたか 2. あなたはこの対話型 AI とどのくらい会話をしたいと思いますか 3. あなたはこの対話型 AI をどのくらい友達と感じましたか
ボット使用への意欲について	1. あなたはこの対話型 AI をどのくらいまた利用したいと思いますか 2. あなたはこの対話型 AI をどのくらい人に勧めたいと思いますか

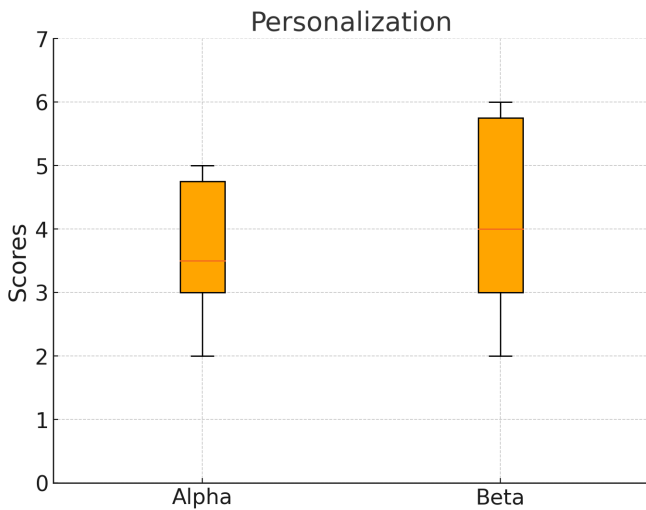


図 4 会話のパーソナライズについて

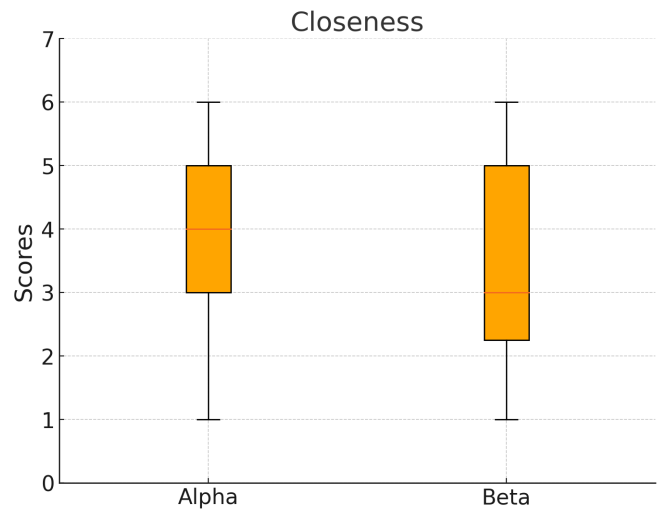


図 6 親近感について

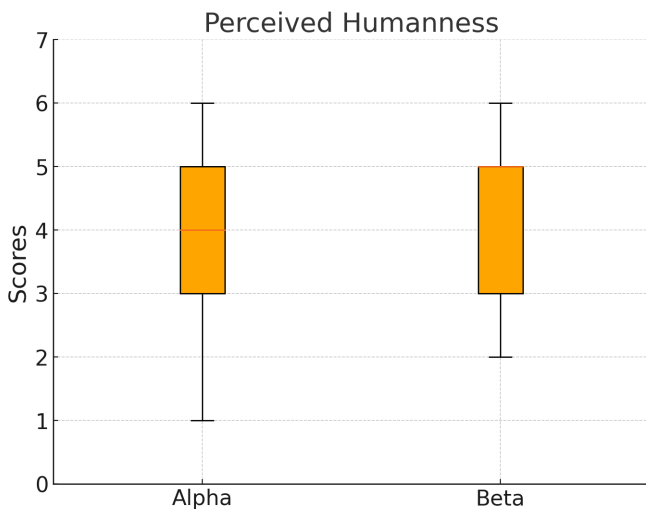


図 5 人間性の知覚について

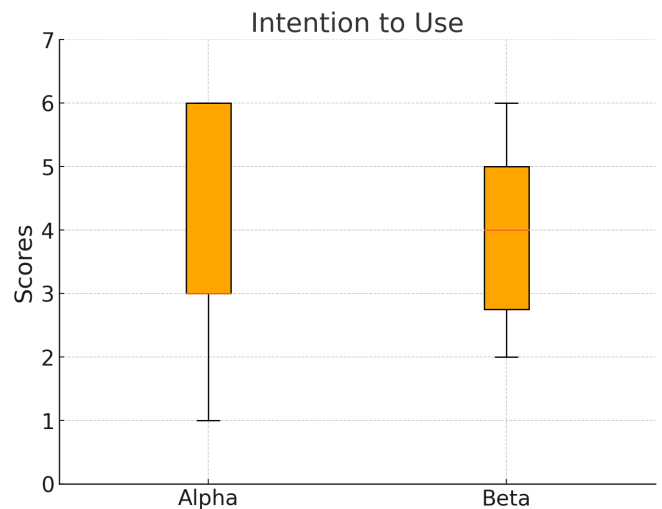


図 7 ボット使用への意欲について

に関しては帰無仮説が支持された。

また、半構造化インタビューを行った結果、

- ヘルスケア情報を用いることに関する意見
- 人間らしさの知覚について
- 実験設計に関する課題
- 今後の展望

についての意見を中心として、意見を得ることができた。

これらを用いて、次のセクションでアンケート結果についての洞察を行う。

## 5. 考察

本章では、実験によって得られた定量的結果を、半構造

表 5 人間性の知覚に関する半構造化インタビューの結果

ボットタイプ	インタビューからの感想
ボット $\alpha$	<ul style="list-style-type: none"> <li>唐突に体調に関する話題を開始するので業務的な会話を展開されている感覚を覚えた。</li> </ul>
ボット $\beta$	<ul style="list-style-type: none"> <li>ヘルスケア情報を用いて会話を開始することで、会話のきっかけとしてすぐ現実味を感じた。</li> <li>ヘルスケア情報を参考にした上で話しかけてくることに対して、自身に対して興味を持ってきているような感覚を覚えた。</li> <li>ヘルスケア情報を具体的に用いたときに、普段看過してしまう内容に気付かされるような感覚を得た。</li> </ul>

化インタビューによって実験参加者から得られた定性的評価をもとに考察する。

### 5.1 ボットの想定外の挙動による影響

半構造化インタビューを経て得られた知見をもとに、結果の考察を行う。まず、実験を通しての対話ボット  $\alpha$ 、対話ボット  $\beta$  双方の挙動として、ヘルスケア情報を踏まえた体調等に関する会話から、会話の早い段階で簡単に逸れてしまい、新たな話題を展開するケースが非常に多く見られた。事前設定においては、健康や体調に関するテーマを常に展開するように指定しており、これは図 2 と 3 で示されるように ChatGPT によっても解釈されている項目である事が確認できる。本来、5分間の会話を通して実験参加者にはそれぞれの対話型 AI と健康や体調に関する会話を行ってもらい、抱いた印象を調査する予定であったが、半構造化インタビューにおいても

- ボット  $\alpha$  とボット  $\beta$  では異なるテーマでの会話となってしまうので、比較が難しい。
- ヘルスケア情報を踏まえた会話の展開を期待していたが、すぐにテーマが変更されてしまった。
- もっとヘルスケア情報を押し出して無理矢理にでもパーソナライズ感を貫いてほしかった。

などの意見があり、会話における早い段階での話題の転換が要因となって適切な実験参加者からの評価が得られなかった。しかしながら、インタビューの中では、「普通の友人との会話をイメージすれば、目まぐるしく話題が変化していく事は自然」という意見もあり、これらを踏まえると実験参加者にとっては「実験を通して対話型 AI との会話を行っている」という強い先入観を有している蓋然性が高い。従って、本研究では実験参加者が対話型 AI に対して人間性を知覚する事を評価観点として取り入れているが、先入観との対立によって有意差が得られなかったと考察できる。またこの点を踏まえて実験を行う際には対話型 AI が指定した挙動を行う可能性を向上させる事前設定と、パ

イロットスタディなどを積極的に取り入れる事で想定外の挙動に対する理解を深める必要がある。

### 5.2 ChatGPT 固有の挙動による影響

実験後に得られたユーザからのインタビュー結果より

- 片方のボットは敬語を使用していたが、もう片方のボットは常に丁寧語を使用していた。
- 片方のボットからは箇条書きを頻繁に用いられたので、人間らしさ・親近感を感じられなかった。
- アニメの話題になった際に登場しない人物に関する発言が見られ、ハルシネーションを感じた。

という望まれない ChatGPT の挙動が多く見られた。これは対照性の喪失による、結果への信頼度の低下や、LLM 自体の挙動に対する印象がアンケート結果に反映されてしまう要因になったと考え、「親近感」や、「再使用への意欲」に関する項目が、前述の原因も重なって適切に評価できない状況を招いてしまったと考えられる。

### 5.3 「思慮深さ」について

次に、t 検定を行なった結果、唯一有意傾向が見られた項目は「思慮深さ」である。半構造化インタビューにおいて、「人間性の知覚」をテーマを掘り下げた際に、ヘルスケア情報の活用によって、実験参加者がどのような印象を抱いたかに関するインタビューを行った。その結果を表 5 に示す。

「普段看過してしまう内容に気付かされる」と言うのは、ヘルスケア情報を用いる事でしか得られない意見である。また、ヘルスケア情報という前提を会話に組み込むことで、「業務的な会話展開」という感覚を退け、寧ろ「実験参加者に対する興味を対話型 AI が抱いている」という印象を与える事ができるということが実験参加者へのインタビューで明らかになった。これらの意見は、実験参加者自身が認識できていない自身の一面について気づきを与えるきっかけとなる役割を対話型 AI が果たしていると解釈でき、これらの要因が思慮深さについて有意差として現れたものだと考えられる。また、これらは対話ボット  $\alpha$  と対話ボット  $\beta$  のヘルスケア情報を用いたか否かの違いのみに起因するものであり、ヘルスケア情報を活用する事の有用性を確認できた。

### 5.4 対話型 AI に求めるもの

さらに、「どのような振る舞いを対話ボットに求めるか」という質問に対しては、

- 現状のヘルスケア情報を踏まえて、どのような運動をするべきかを提案してほしい。
- 連日データを取得することで、その推移を踏まえた会話を展開してほしい。

という意見が得られ、10人中9名がヘルスケア情報を会



表 6 アンケート項目における t 検定結果

アンケート項目	P 値	検定結果
あなたはこの対話型 AI と行った会話がどのくらい特別なものだと感じましたか	0.509	n.s.
あなたはこの対話型 AI にどのくらい人間らしさを感じましたか	0.464	n.s.
あなたはこの対話型 AI にどのくらい思慮深さを感じましたか	0.087	p < .10
あなたはこの対話型 AI にどのくらい魅力を抱きましたか	0.545	n.s.
あなたはこの対話型 AI にどのくらい親近感を抱きましたか	0.790	n.s.
あなたはこの対話型 AI とどのくらい会話をしたいと思いますか	1.000	n.s.
あなたはこの対話型 AI をどのくらい友達と感じましたか	0.343	n.s.
あなたはこの対話型 AI をどのくらいまた利用したいと思いますか	0.494	n.s.
あなたはこの対話型 AI をどのくらい人に勧めたいと思いますか	0.705	n.s.

話に取り入れる事に意欲的で、寧ろ積極的なパーソナライズを求める意見が多かった。今回実験設計上、1回5分という条件の下で会話を行ってもらったが、これを数日間に渡って行った場合には、よりコンテキスト理解・パーソナライズの項目において実験参加者の評価向上が得られる可能性があると考えられる。

一方で、「データを用いた会話展開によって、何が自分に足りていないかはよくわかったが、人間は感覚の動物であるから、データに基づいて会話を行う事に人間性を感じない」という意見や、「そもそも会話相手を求めている」という意見もあり、本研究の範囲からは逸脱するが、相手の会話に対する意欲度に基づいた会話展開を行う必要性も検討しなければならない。

## 6. おわりに

本研究では、対話型 AI との会話における、ユーザへのコンテキスト理解の手法としてヘルスケア情報を用いたプロンプトの拡張を行なった。その有効性を調べるために、ヘルスケア情報を用いない手法 (ポット  $\alpha$ ) と、ヘルスケア情報を用いる手法 (ポット  $\beta$ ) との会話を体験してもらい、アンケート調査と半構造化インタビューを実施した。ヘルスケア情報を用いて会話が展開されることは、ユーザにとって対話型 AI に対する「思慮深さ」を抱く有意傾向が見られるという結果が得られた。また、ヘルスケア情報を用いた会話自体には実験参加者自身は意欲的であり、さらなるパーソナライズを行う事でユーザの満足度が向上する可能性が示唆された。一方で、ChatGPT の挙動や、想定外の挙動の影響によって当初予定した実験が行えず、望む結果が得られなかった。今後 LLM を用いた対話型 AI に対してプロンプト拡張に関する実験を行う際には、事前設定によって挙動の堅牢性を保証する事や、パイロットスタディを積極的に取り入れる事によって LLM の挙動に対する理解を多面的に行う必要がある。また、対話型 AI とユーザの対話において、対話型 AI 側がユーザに対する話しかけを行う事、すなわちどのタイミングで話しかけを行うことが適切なのか、という視点も考慮する必要がある。ここでヘルスケア情報は非言語的なコンテキスト理解に有

用なデータであり、これらを適切に活用することで、介入のタイミングを見極め、またそれに基づく会話を展開することで、介入から会話の初期段階までの自然な流れをユーザに与えることができると考え、ユーザの会話体験をさらに向上させることができると考える。

## 謝辞

本研究は、科学研究費補助金 (JP23H00216) の支援のもと実施されている。

## 参考文献

- [1] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and conversational agents in mental health: a review of the psychiatric landscape," *The Canadian Journal of Psychiatry*, vol. 64, no. 7, pp. 456–464, 2019.
- [2] E. Jo, D. A. Epstein, H. Jung, and Y.-H. Kim, "Understanding the benefits and challenges of deploying conversational ai leveraging large language models for public health intervention," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.
- [3] M. Shumanov and L. Johnson, "Making conversations with chatbots more personalized," *Comput. Hum. Behav.*, vol. 117, p. 106627, 2020.
- [4] L. Luo, W.-F. Huang, Q. Zeng, Z. Nie, and X. Sun, "Learning personalized end-to-end goal-oriented dialog," *ArXiv*, vol. abs/1811.04604, 2018.
- [5] H. Qian, Z. Dou, Y. Zhu, Y. Ma, and J. rong Wen, "Learning implicit user profile for personalized retrieval-based chatbot," *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021.
- [6] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. Schmidt, "A prompt pattern catalog to enhance prompt engineering with chatgpt," *ArXiv*, vol. abs/2302.11382, 2023.
- [7] B. Xu, A. Yang, J. Lin, Q. Wang, C. Zhou, Y. Zhang, and Z. Mao, "Expertprompting: Instructing large language models to be distinguished experts," *ArXiv*, vol. abs/2305.14688, 2023.
- [8] Z. Zhang, A. Zhang, M. Li, and A. J. Smola, "Automatic chain of thought prompting in large language models," *ArXiv*, vol. abs/2210.03493, 2022.



- [9] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," *ArXiv*, vol. abs/2203.11171, 2022.
- [10] C. Zheng, Z. Liu, E. Xie, Z. Li, and Y. Li, "Progressive-hint prompting improves reasoning in large language models," *ArXiv*, vol. abs/2304.09797, 2023.
- [11] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, "Why johnny can't prompt: How non-ai experts try (and fail) to design llm prompts," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
- [12] H. Wang, R. Wang, F. Mi, Y. Deng, Z. Wang, B. Liang, R. Xu, and K.-F. Wong, "Cue-cot: Chain-of-thought prompting for responding to in-depth dialogue questions with llms," pp. 12 047–12 064, 2023.
- [13] R. P. dos Santos, "Enhancing chemistry learning with chatgpt, bing chat, bard, and claude as agents-to-think-with: A comparative case study," *ArXiv*, vol. abs/2311.00709, 2023.
- [14] S. Lee, N. Lee, and Y. Sah, "Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use," *International Journal of Human-Computer Interaction*, vol. 36, pp. 930 – 940, 2019.
- [15] W. Wei, S. Li, S. Okada, and K. Komatani, "Multimodal user satisfaction recognition for non-task oriented dialogue systems," *Proceedings of the 2021 International Conference on Multimodal Interaction*, 2021.
- [16] H. Wang, R. Wang, F. Mi, Z. Wang, R.-L. Xu, and K.-F. Wong, "Chain-of-thought prompting for responding to in-depth dialogue questions with llm," *ArXiv*, vol. abs/2305.11792, 2023.

表 7 生体情報の有無による対話の違い

会話サンプル	ボット $\alpha$ との会話		ボット $\beta$ との会話	
	ボットの発話	ユーザの発話	ボットの発話	ユーザの発話
1	こんにちは！最近、体の調子はどうですか？何か運動やリラクゼーションの活動をしていますか？それとも、気になる健康のことがあれば、ぜひ教えてくださいね	最近発表の準備で忙しくて睡眠時間もへってる。身体を動かす時間はいつもと変わってないつもりだけどそうなのかな	今日の歩数 3365 歩って、週平均や月平均に比べてちょっと少なめだね。何か忙しかった？それともちょっと体調でも悪かった？普段より少ない歩数だと疲れが溜まってるサインかもしれないから、今夜はゆっくり休むのがいいかもね。あと、燃焼カロリーも 97.9 ってことは、体を動かす時間が少なかったのかな。何か気になる体の変化とかある？	最近起きた時に寝ても寝ても眠い感じがする
2	こんにちは！最近、体の調子はどうですか？何か新しい運動を始めたり、リラクゼーションのために何か特別なことをしていますか？	最近ハンドボールサークルに参加して週に 1 回運動する習慣がついています	今日の歩数 3028 歩というのは、週平均や月平均に比べてちょっと少なめだね。どこか出かけた？それとも今日はゆっくり過ごしていたのかな？歩数が少ないと燃焼カロリーも少なくなるから、今日の 78.9 カロリーっていうのもそういうわけだね。何か体調の変化とか感じてる？それともただリラックスしていた日だった？	今日は学校に行った。夜ご飯を食べるために学食へ歩いてだけです。通学はバイクなので駐輪場から教室に歩くだけです。運動に関しては以上です。特に体調の異常は感じていません
3	こんにちは！最近、体調はどうですか？何か新しい運動を始めたり、リラックスする時間を作ったりしていますか？	体調は結構いいです。週に 4 回は筋トレをしていて、その効果が現れている気がします。	お、今日はまだ 1888 歩か。平均に比べてちょっと少ないね。何か忙しかった？それともゆっくり過ごしてた？少し動くとも気分も変わるかもしれないし、ちょっと散歩してみるのもいいかもね。	今日は一日中研究室にいたので、それが歩数が少ない原因ですね～
4	こんにちは！最近、体を動かしたり、リラックスする時間は取れてるかな？あと、体調はどうだい？何か気になることがあったら教えてね。	体調はめっちゃいいよ。昨日は 10 時間も寝ちゃった。	今日の歩数、4271 歩ってことだね。週平均と比べると、ちょっと多めだけど、月平均と比べると少ないかな。最近あんまり歩いてないのかい？それとも何か特別な理由があるの？それに今の心拍数が 97 っていうのは、ちょっと高めだね。何か運動したり、ちょっとした作業をしていたとか？それとも、なんか気になることもあった？このデータからだと、身体的にはちょっと活動的な日かもしれないね。どうだった？何か特別な活動をしてたりする？	ゴールデンウィークは学振書いてずっと研究室いたよ。ほんと最悪だよ