筑波大学 情報学群 情報メディア創成学類

卒業研究論文

チャットボットにおける
応答遅延のユーザ体験への影響

佐藤 大樹

指導教員　志築 文太郎，川口 一画

2024 年 1 月

概要

チャットボットは顧客サービス，ヘルスケア，教育分野などの，多岐にわたる分野において活用されている．さらに近年，OpenAI の ChatGPT を始めとする高性能な LLM の登場により，その存在感はますます高まっている．本研究では，チャットボットの応答遅延がユーザ体験へ与える影響に焦点を当て，応答時間がチャットボットの知覚される人間らしさ，社会的存在感，利用満足度に及ぼす効果を調査している．このために，ユーザの入力文とチャットボットの応答文それぞれの複雑さに基づく大小 2 段階の遅延時間を加えることのできるチャットボットアプリケーションを実装し，実験参加者を募って実験を行った．実験は遅延無し，中程度の遅延，大きな遅延の 3 条件において行なった．この実験の結果，チャットボットにおける応答遅延がユーザ体験に与える影響が限定的であること，またはユーザがチャットボットの応答時間に対して寛容ではないことが示唆された．

目 次

第 1 章　序論 **1**
1.1 背景 1
1.2 目的とアプローチ 1
1.3 貢献 1
1.4 本論文の構成 2

第 2 章　関連研究 **3**
2.1 チャットボットに応答遅延を加える研究 3
2.2 AI による執筆補助ツールに応答遅延を加える研究 4
2.3 チャットボットの応答遅延とユーザの信頼に関する研究 4

第 3 章　仮説 **5**

第 4 章　実験 **6**
4.1 実験用チャットボット 6
4.2 文章の複雑さによる応答遅延の算出 7
4.2.1　文章の複雑さの算出 7
4.2.2　応答遅延の算出 8
4.3 実験参加者 8
4.4 実験タスク 8
4.5 結果 10
4.6 考察 14

第 5 章　議論 **15**
5.1 遅延時間の設定 15
5.2 実験 15

第 6 章　結論 **16**

第 7 章　謝辞 **17**

参考文献 **18**

図 目 次

4.1 実験用アプリケーション 6
4.2 実験タスク 10
4.3 知覚される人間らしさ 12
4.4 社会的存在感 12
4.5 利用満足度 13

第1章　序論

本研究では，チャットボットにおける応答遅延がユーザ体験に与える影響を調査する．

1.1　背景

チャットボットは，人間と自然言語を用いてテキストによる対話を行うシステムであり [1]，顧客サービス [2]，ヘルスケア [3]，教育分野 [4] などの，多岐にわたる分野において活用されている [5]．特に，2022 年 11 月にリリースされた OpenAI の ChatGPT [6] および 2023 年 3 月にリリースされた Google の Bard [7] は，大規模データセットを用いた言語生成モデルを基盤とする，多様なトピックに対する自然かつ流暢な会話生成能力を持つ．このことは，この分野の進歩を象徴している．

これらのチャットボットは，従来のシステムよりも高度な会話能力を持つことにより，ユーザに対してより人間らしい応答を提供することが可能になった．この結果，ユーザはチャットボットを用いて，単純な質問応答 [8] から複雑な問題解決 [9,10]，効率的な情報収集 [11] に至るまで，より幅広い活用を行うことが可能となっている．しかし，チャットボットの応答速度がユーザ体験に与える影響については，まだ十分に研究が行われていない．

1.2　目的とアプローチ

本研究の目的は，チャットボットの意図的な応答遅延がユーザ体験に及ぼす影響の調査である．この目的を達成するためのアプローチとして，まず我々は，異なる応答遅延（即時，短い遅延，長い遅延）に調整可能なチャットボットアプリケーションを実装する．その後，それぞれの応答遅延設定における実験参加者の体験の評価実験を行う．評価実験においては，チャットボットを使用した際にチャットボットに対して知覚される人間らしさ，社会的存在感，および利用満足度への影響を調査する．社会的存在感とは，コミュニケーション環境において，個人が相手を現実の存在として認識する度合いを指す [12]．

1.3　貢献

本研究の貢献は以下のとおりである．

- 動的な応答遅延を加えるチャットボットアプリケーションを実装した．

- 評価実験において，チャットボットに意図的な大小の応答遅延を加えることによるユーザ体験への影響を調査した．

1.4 本論文の構成

本論文の構成は以下のとおりである．第 1 章において，本研究の背景，目的とアプローチ，および貢献を示した．第 2 章において，本研究に関連する研究を示し，本研究の位置付けを示す．第 3 章において，関連研究を踏まえた仮説を述べる．第 4 章において，チャットボットにおける応答遅延がチャットボットに対して知覚される人間らしさ，社会的存在感，および利用満足度に与える影響の調査の実験およびその結果を述べる．第 5 章において，第 4 章の実験に関する議論を述べる．第 6 章において，本研究の結論を述べる．

第2章　関連研究

本研究の目的は，チャットボットの応答の遅延時間がユーザ体験に及ぼす影響を調査することである．本章においては，本研究に関連する研究として人間とチャットボットとの会話において動的な応答遅延が与える影響についての研究を述べる．

2.1　チャットボットに応答遅延を加える研究

チャットボットに応答遅延を加える研究は, Holtgraves ら [13], Peng ら [14], Antonio ら [15], Gnewuch ら [16] などがある．

Gnewuch ら [16] は，動的な遅延時間を加えたチャットボットと即時応答のチャットボットを作成し，問題解決タスクを行うことによってユーザ体験への影響を調査した．実験の結果，チャットボットにおける意図的な動的応答遅延はユーザをチャットボットに対して人間らしく感じさせ，社会的存在の認識を高め，全体的なインタラクションに対する満足度を向上させることが示された．

動的な遅延時間はユーザの入力文とチャットボットによる応答文の複雑さを基に，以下の式によって設定している．

$$C(m) = 0.39 \times \left(\frac{total\ words}{total\ sentences}\right) + 11.8 \times \left(\frac{total\ syllables}{total\ words}\right) - 15.59 \tag{2.1}$$

$$\begin{aligned} D(m) &= 0.5 \times \ln(C(m) + 0.5) + 1.5 && C(m) > 0 \\ D(m) &= 0 && C(m) \leq 0 \end{aligned}$$

$C(m)$ はユーザの入力文またはチャットボットの応答文の複雑度を表し，$D(m)$ はチャットボットの応答遅延時間を表す．ただし，この式 2.1 は日本語には使用できない．また，Gnewuch ら [16] の研究においては即時応答と応答遅延を加えた応答の 2 段階のみを比較しており，この遅延時間が最適であるとは限らない．

ユーザがチャットボット初心者であるか否かによってこの応答遅延の効果が反転するという研究もある [17]．この研究において，チャットボットの経験豊富なユーザに対して応答遅延が体験に悪影響を与える可能性が示唆されている．そのため，比較実験においてユーザがチャットボット初心者であるか否かも調査すべきであると考えられる．

2.2　AIによる執筆補助ツールに応答遅延を加える研究

AIを用いた執筆アシスタントは，即時かつ自動的なフィードバックを提供する強力なツールとして進化している [18]．Shi & Deng [19] は4種類のAIフィードバック表示ツール（テキスト形式，グラフィック形式，表形式，音声形式）に対して短い遅延（1秒），中程度の遅延（3秒），長い遅延（5秒）の3段階の応答遅延時間を適用し，比較実験を行なった．実験の結果，平均して最も好まれた応答遅延は中程度の遅延であった．

この実験において，応答時間はユーザの入力，システムの応答の内容に関わらず，事前に決められた値が使用されている．人間の一般的な会話においては内容の難易度，複雑さによって応答時間が変わることは普通であり，そのため，会話の流れに応じて適切な応答時間が自然に調節される．しかし，Shi & Deng [19] の実験においては，応答時間が固定されており，これは実際の人間の対話の動的な性質を完全に反映していない可能性がある．

2.3　チャットボットの応答遅延とユーザの信頼に関する研究

Choedakの研究 [20] は，チャットボットの応答遅延がユーザの信頼，特に認知的信頼と感情的信頼に与える影響を探求している．この研究は，チャットボットの応答時間を意図的に遅らせることが，ユーザの認知的信頼を高めることを示唆しているが，感情的信頼には顕著な影響を与えないことを示した．これは，応答遅延がユーザのチャットボットに対する満足度と人間らしさの知覚に与える影響とは異なる影響をユーザの感情的信頼に対して与えることを示唆している．

第3章　仮説

これまでの研究から，チャットボットにおける応答遅延については，以下のことが示唆されている．

1. ユーザの入力文と応答文の複雑さに基づく動的な応答遅延を加えることにより，ユーザがチャットボットをより人間らしい存在として認識する可能性が高まる．
2. 異なる応答遅延の長さは，ユーザの体験に異なる影響を与える．長すぎる遅延はユーザの満足度を低下させる．
3. ユーザがチャットボットの使用に習熟しているかにより，上記の効果の逆転が起こり得る．

先行研究において，動的な応答時間によるユーザ体験への影響についての調査 [16] と，複数の異なる応答遅延時間によるユーザ体験への影響についての調査 [20] が行われている．しかしながら，適切な動的応答遅延の長さについては十分に研究されていない．また，ユーザがチャットボットの使用に習熟しているかどうかも考慮する必要がある [17]．これらを踏まえ，複数の動的な応答遅延を加えたチャットボットのユーザ体験への影響に対して以下の仮説を立てた．

H_1：応答遅延の長さが適切である場合，ユーザはチャットボットとの対話をより自然かつ満足度の高い体験として感じる．

したがって，大小 2 種類の動的な応答遅延を加えたチャットボットにおける，チャットボットに対する社会的存在感とユーザ満足度への影響を評価するために実験を行なった．

第4章　実験

ユーザの入力文およびチャットボットの応答文それぞれの複雑さに基づく大小 2 種類の動的な応答遅延を加えたチャットボットに対して知覚される人間らしさ，社会的存在感，および利用満足度への影響を評価するために実験を行った．

4.1　実験用チャットボット

実験用のチャットボットアプリケーションは Vue3 および Vite を用いてウェブアプリケーションとして実装した（以下実験用アプリケーション）．実験用アプリケーションの画面を図 4.1 に示す．実験用アプリケーションはメールアドレスおよびパスワードによるログイン機能を備えており，ユーザがログインすることによって実験参加者ごとに設定した遅延時間を，事前に準備したデータベースから取得する仕様になっている．また，ユーザとチャットボットとの会話内容は全てデータベースに保存される．

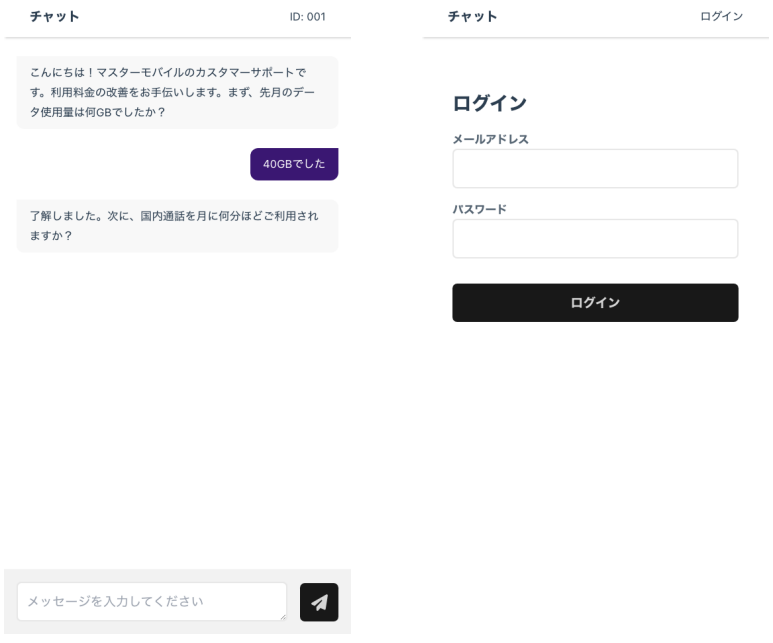


図 4.1: 実験用アプリケーション

ユーザログイン機能には Firebase Authentication のメールとパスワードに基づく認証を用い

た. ユーザごとの設定情報の管理とチャットボットとの会話に関するデータの保存にはFirestoreを用いた．また，実験用アプリケーションのデプロイにはVercelを用いた．

また，チャットボットの応答文生成にはOpenAIによるChat Completion APIのモデルであるgpt-3.5-turbo-1106を用いた．このモデルを選択した理由は，十分に自然な会話生成ができることに加え，OpenAIによる他のモデルと比較してテキスト生成の実行時間が圧倒的に短いためである[21].

4.2 文章の複雑さによる応答遅延の算出

チャットボットの応答に加える遅延時間は，Gnewuchら[16]と同じように，ユーザの入力文とチャットボットの応答文それぞれの複雑さに基づいて決定される．

4.2.1 文章の複雑さの算出

文章の複雑さの指標として，その文章から算出した可読性スコアを用いる．日本語文の可読性スコアの算出方法として，教科書コーパスを用いる手法[22]，教科書と国会会議録のコーパスを用いる手法[23]，文字の種類ごと塊の平均文字数手法[24]などが提案されている．本研究においては，建石ら[24]による，文字の種類ごとの塊の平均文字数を用いる手法を採用した．この手法のシンプルさが，応答遅延設定の複雑性を低減し，効率的な実装を可能にする点において特に有利である．

建石ら[24]による日本語文の可読性スコアは式4.1の様に表される．ここにおける連とは，同じ種類の文字が連続する塊のことである．

$$RS = -0.12ls - 1.37la + 7.4lh - 23.18lc - 5.4lk - 4.67cp + 115.79 \tag{4.1}$$

RS 可読性スコア
ls 文の平均文字数
la アルファベット連の平均文字数
lh ひらがな連の平均文字数
lc 漢字連の平均文字数
lk カタカナ連の平均文字数
cp 句点あたり読点の数

ここで，ELYZA Tasks 100データセット[25]を用いて，式4.1におけるRSの取り得る範囲を検証した．ELYZA Tasks 100は，ELYZA, Inc.が公開している，日本語の言語モデルの性能を評価するために開発された，100種類のタスクを含むデータセットである[26].

可読性推定手法の検証のため，ELYZA Tasks 100 に含まれる各文章の *RS* を算出した．具体的には，各文章に対して式 4.1 を適用し，*RS* を算出した．この検証の結果，ELYZA Tasks 100 に含まれる文章の可読性スコアは，−106 から 103 の範囲であった（SD = 26.6）．*RS* の最小値と最大値における文章を表 4.1 に示す．なお，*RS* の平均値は 54.5，中央値は 56.6 であった．

4.2.2 応答遅延の算出

RS を基にした中程度の応答遅延 $D(m)$ は以下のように表される．

$$D(m) = 0.5 \times \log\left(\frac{120 - C(m)}{30} + 0.5\right) + 1.5 \qquad C(m) > 0$$

$$D(m) = 0 \qquad C(m) \leq 0$$

この $D(m)$ は Gnewuch ら [16] と同程度の値になるように設定した．具体的には，最大 2.5 秒程度の遅延時間になる．

また，実験においては大きな応答遅延，中程度の応答遅延，応答遅延の追加無しの 3 条件において比較実験を行う．この時，大きな応答遅延には，中程度の応答遅延の 2 倍の値を用いる．

4.3 実験参加者

実験参加者は，SNS にて募集した 12 名（男性 8 名，女性 4 名）である．参加者の年齢は 18 から 54 歳であり，半数が 18 から 24 歳であった．最終学歴については，8 人が大学卒業者（文系 4 名，理系 4 名）であり，その他には高校卒業者，短期大学卒業者，および大学院卒業者もそれぞれ 1 人含まれていた．汎用 AI チャットボット（例：ChatGPT，Google Bard など）の使用頻度は，1 名は 1 週間に 2，3 回程度，4 名は月に 1 回程度，2 名は 1 年に 1 回程度であり，5 名は使用した経験がなかった．実験の所要時間は 1 人当たり約 15 分であった．

4.4 実験タスク

実験は Google フォームと実験用アプリケーションを用いて，オンライン環境において行なった．実験参加者には，まず架空の携帯電話会社の利用料金明細書（以下明細書）を提示した．提示した明細書を図 4.2 に示す．この明細書には，利用料金の内訳，通話時間およびデータ通信量が記載されている．明細書において利用中のプランは，記載されているデータ通信量と通話時間に対して不適切な内容になっており，高額な追加料金が発生している．そこで，実験参加者は実験用アプリケーションのチャットボットと会話し，より適切なプランを提案してもらうよう求められた．会話中，チャットボットは実験参加者の使用状況（例：データ通信量）について質問した．全ての情報が得られると，チャットボットはより適切なプランを提

表 4.1: ELYZA Tasks 100 データセットに基づく日本語文の可読性スコア（*RS*）の最小値と最大値

文章	*RS*
リンゴとバナナはどのように似ていますか？	103.6
それでは SF というジャンルに大きな影響を与えた古典とされる SF 小説を読んでみるのはどうでしょう？ 1. メアリー・シェリー著『フランケンシュタイン』（1818 年）-死体の一部から怪物を作り出す科学者を描いたこの古典小説は、史上初の SF 小説のひとつとされています。2.H.G. ウェルズ著『タイムマシン』（1895 年）-未来に旅立ったタイムトラベラーが、平和なエロイ族と残忍なモーロック族に分断された世界を発見するこの小説は、SF ジャンルの先駆的作品です. 3.H.G. ウェルズ著『宇宙戦争』（1898 年）-火星人の地球侵略を描いたこの小説は、異星人の生命体の生き生きとした描写と、帝国主義と技術進歩に対する論評で知られるウェルズのもうひとつの代表作です。4. フランク・ハーバート著『デューン砂の惑星』（1965 年）-人類が銀河系の他の惑星を植民地化した遠い未来を舞台に、ポール・アトレイデスという青年が砂漠の惑星アラキス（別名デューン）の政治、宗教、生態系の複雑な網に巻き込まれていく姿を描いた小説です。5. アイザック・アシモフ著『ファウンデーション』（1951 年）-人類が銀河系に広がり、崩壊の危機に直面している遠い未来が舞台で。ハリ・セルドンという数学者が、未来を予測するために心理歴史学という新しい分野を開発し、差し迫った暗黒時代に知識と文化を守るための財団を設立する姿を描いています。6. ウィリアム・ギブソン著『ニューロマンサー』（1984 年）-この小説の舞台は、仮想現実とサイバースペースが日常生活の不可欠な一部となった近未来の世界です。この小説は、ケースという名の落ちぶれたコンピューター・ハッカーが、企業が巨大な権力を振るう世界でスパイ行為と裏切りの危険なゲームに巻き込まれていく様を描いています。7. フィリップ・K・ディック著『アンドロイドは電気羊の夢を見るか』（1968 年）-ネクサス 6 と呼ばれる高度なアンドロイドが人類に奉仕するために作られた黙示録後の世界を舞台に、悪質なアンドロイドを狩る任務を負った賞金稼ぎの物語を通して、人間性、共感、そして生きていることの意味を探る SF 小説です。8. アーシュラ・K・ル＝グウィン著『闇の左手』（1969 年）-この小説の舞台はゲーセンと呼ばれる遠い惑星で、住民は性別がなく両性愛者です。人間の外交官がこの惑星の複雑な政治状況をナビゲートしようとする中で、セクシュアリティ、ジェンダー、権力のテーマを探求する話です。9. スタニスワフ・レム著『ソラリス』（1961 年）-惑星ソラリスを周回する宇宙ステーションを舞台にした小説で、クリス・ケルヴィンという心理学者が奇妙な現象を調査するためにやってきます。謎めいた異星人との出会いを通して、意識、人間性、科学的理解の限界といったテーマの小説です。10. リチャード・K・モーガン著『オルタード・カーボン』（2002 年）-人間の意識が身体間で移動できるようになった未来を舞台に、元兵士のタケシ・コバックスが殺人事件を捜査し、社会秩序全体を脅かす陰謀を暴いていく物語です。	-106.3

案した．この実験タスクを選択した理由は，このタスクがチャットボットの利用場面として現実的であり，かつ，同様のシナリオが先行研究において用いられているためである [16,27,28]

実験参加者は，実験用アプリケーションのチャットボットと会話終了後に，Google フォームにおいてチャットボットの知覚される人間らしさ [13,29]，社会的存在感 [30]，およびチャットボットに対する利用満足度 [28]，ならびに自由記述にて感想を回答した．チャットボットの知覚される人間らしさは 9 段階の SD 法により測定した．チャットボットの社会的存在感，およびチャットボットに対する利用満足度は 7 段階のリッカート尺度により測定した．これらの尺度は，先行研究においても同一の尺度が用いられている [16].



図 4.2: 実験タスク

4.5 結果

アンケート結果より得られた知覚される人間らしさ，社会的存在感，および利用満足度の値について，統計分析の結果を示す．アンケートに回答した 12 名のうち，1 名はチャットボットを利用せずに回答を行なったため，この 1 名を除いた 11 名の結果を分析に用いた．なお，統計分析において有意水準は $p < 0.05$ とした．

まず，社会的存在感および利用満足度について，各 7 段階の尺度を 1 から 7 までの数字にマッピングを行なった．その後，遅延無し，中程度の遅延，大きな遅延の 3 条件における知覚される人間らしさ，社会的存在感，および利用満足度の平均値を算出した．また，各条件における知覚される人間らしさ，社会的存在感，および利用満足度の値について，クラスカル・ウォリス検定を行ない，p 値を算出した．これらの値を表 4.2 に示す．

検定の結果，知覚される人間らしさ，社会的存在感，および利用満足度に関する項目全てにおいて，p 値が 0.05 を超えていた（表 4.2）．

加えて，アンケートの最後に自由記述において入力してもらった意見・感想のうち，チャットおよび実験に関する内容のものを表 4.3 に示す．一部の実験参加者からは，チャットボットとの会話内容が適切であれば，チャットボットの人間性は重要視しないとの意見があった．また，実験タスクにおける会話が AI との会話であることを実験参加者が知っているため，チャットボットに対して人間性および感受性をそもそも期待していないという意見も見られた．

表 4.2: 実験結果

		平均値		
質問文	p 値	遅延無し	中程度の遅延	大きい遅延
知覚される人間らしさ				
チャットボットは人間的に感じられましたか	0.840	5.33	5.25	4.75
チャットボットは熟練しているように感じられましたか	0.557	4.00	5.75	5.50
チャットボットは思慮深く感じられましたか	0.217	4.00	5.50	3.75
チャットボットは礼儀正しく感じられましたか	0.788	7.67	7.00	8.00
チャットボットは反応が良く感じられましたか	0.602	6.67	6.75	6.00
チャットボットは魅力的に感じられましたか	0.528	5.67	5.50	4.50
社会的存在感				
人間との接触感	0.635	4.33	3.50	3.50
個人的な繋がり	0.212	4.00	3.75	1.75
社交性	0.102	5.67	4.75	4.25
人間の暖かさ	0.879	3.67	3.25	3.75
人間の感受性	0.405	2.67	3.25	2.00
利用満足度				
チャットボットのアドバイスにどれくらい満足していますか？	0.913	5.00	5.00	4.00
チャットボットによる対応にどれくらい満足していますか？	0.462	6.33	5.00	4.25
チャットボットとの全体的なやり取りにどれくらい満足していますか？	0.333	5.67	4.75	4.00

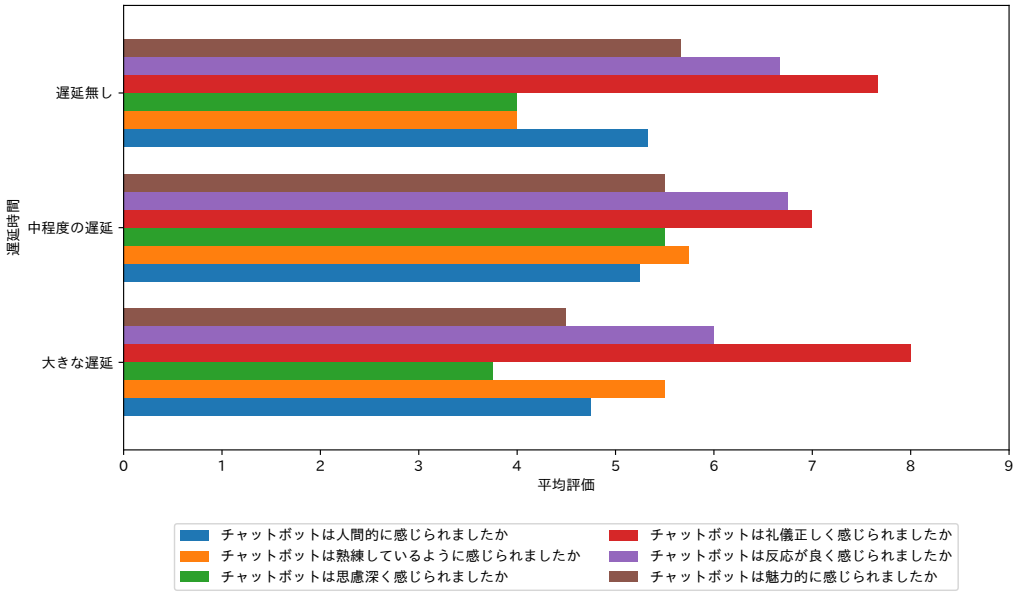


図 4.3: 知覚される人間らしさ

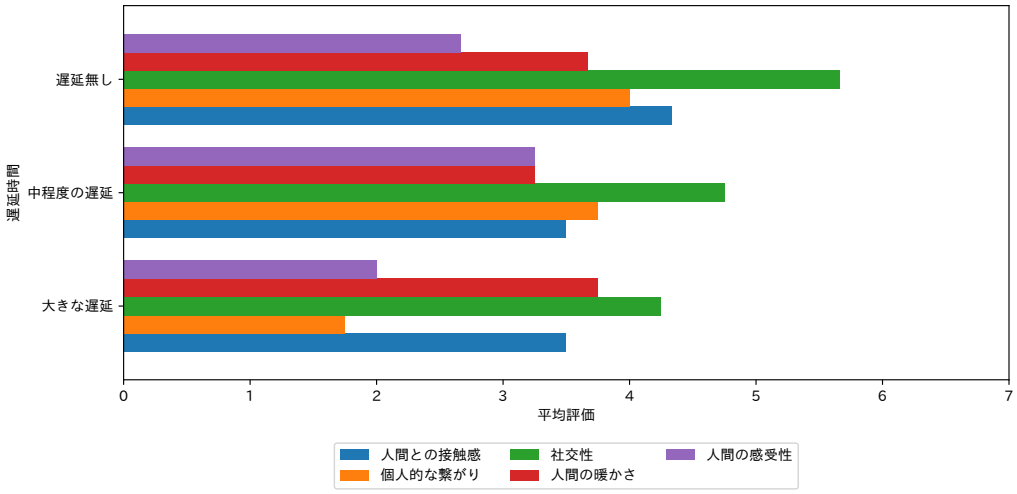


図 4.4: 社会的存在感

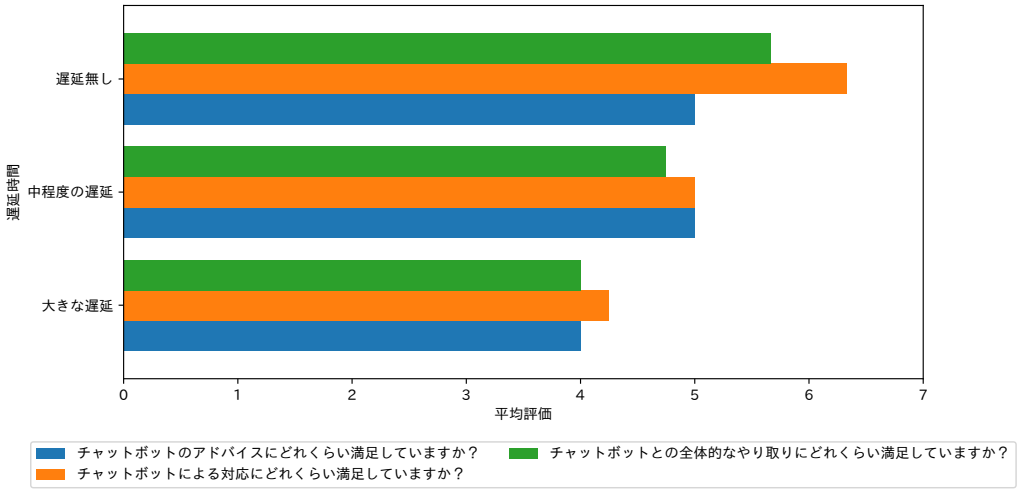


図 4.5: 利用満足度

表 4.3: 意見・感想

回答文
今回のやりとりの内容だといかに適切なプランを提案するかが主となると思うので、会話が噛み合っていれば別に人間性はそこまで高くなくてもいいと個人的に思った。
堅苦しすぎず、自然なやり取りが出来ていると思います。ただ、あらゆる年代の人が利用することを考えると「了解しました」よりは「承知しました」「かしこまりました」の方が違和感を感じられないかと思います。
通常であれば、ウェブサイトなどである程度自分で情報を知っているうえで Bot を使用するため、何も下情報無しに Bot に会話を行いプラン提案されたためそのプランが本当に適切かどうか判断出来なかったこと、また説明の際に発生した疑問点が結局解決出来なかったのが今回このように感じた結果に繋がっていると個人的に考えている。人間性や感受性に関しては、最初から AI と会話していると知っているため、最初から無いものと考えているところがあり、無いからといって不満に繋がっているとも感じていない。

4.6 考察

以上の実験結果から，チャットボットの応答遅延の大小がユーザの知覚される人間らしさ，社会的存在感，および利用満足度に有意な影響を与えるとは言えない．このことから，仮説 H_1 は棄却された．これは，チャットボットにおける応答遅延がユーザ体験に与える影響が限定的であること，またはユーザがチャットボットの応答時間に対して寛容ではないことを示唆している．

第5章　議論

5.1　遅延時間の設定

本研究では，建石ら [24] の手法を用いてユーザ入力文，チャットボットの応答文の可読性推定を行ない，この可読性スコアを基に遅延時間を設定した．ただ，この可読性推定の手法が適切かどうかについては十分に検証できていない．可読性推定の精度向上のため，教科書コーパスを用いる手法 [22]，教科書および国会会議録のコーパスを用いる手法 [23] などと比較検証する必要がある．加えて，そもそも遅延時間に可読性推定を基にした動的遅延時間を用いる手法が適切であるかどうかについても検証する必要がある．Shi & Deng [19] の研究においては静的な遅延時間（1 秒，3 秒，5 秒）が用いられており，一定の効果が見られている．

5.2　実験

本研究における実験参加者は 12 人であった．結果の信頼性を高めるため，より多くの人数による実験を行う必要がある．

また, 本研究における実験用アプリケーションのチャットボットの応答文生成には OpenAI の gpt-3.5-turbo-1106 モデルを用いた．このモデルは GPT3.5 系のモデルであり，非常に高速 [21] である一方，生成される応答文の品質において，現在の最先端のモデルである GPT-4 系のモデルに大きく劣る．GPT-4 系のモデルについては，gpt-4-1106-preview のようなより高速なモデルも公開されてはいる [31] が，それでも応答に 25 秒程度かかるため [21]，本研究における実験には不向きである．今後，GPT-4 と同程度，もしくはより高品質の応答を行うことができ，かつ十分に高速なモデルが登場した場合，本研究の実験を再度行うことで，本研究とは異なる結果が得られる可能性がある．

第6章　結論

本研究においては，チャットボットにおける応答遅延がユーザ体験に与える影響を調査した．このために，意図的な応答遅延を加えたチャットボットに対して知覚される人間らしさ，社会的存在感，および利用満足度への影響を評価するために実験を行った．応答遅延は遅延無しの他に中程度の遅延，大きな遅延の大小 2 段階を用いた．この応答遅延はユーザの入力文およびチャットボットの応答文それぞれの可読性スコアに基づいて設定した．実験の結果，チャットボットの応答遅延の大小がユーザの知覚される人間らしさ，社会的存在感，および利用満足度に有意な影響を与えるとは言えないことが示された．今後は，可読性推定手法の改良を行い，より多くの人数を対象とした実験を行う．

第7章　謝辞

本研究を進めるにあたって，指導教員である志築文太郎先生，川口一画先生，および高橋伸先生には大変お世話になりました．特に志築文太郎先生には，研究の進め方をはじめとしたご意見とご指導をいただきました．また，研究室の先輩，および同輩の方々には研究方針および研究生活について多くの助言をいただきました．最後に，学生生活においてお世話になった先生方，家族，友人に心より感謝いたします．

参考文献

[1] Robert Dale. The return of the chatbots. *Natural Language Engineering*, Vol. 22, No. 5, pp. 811–817, 2016.

[2] Ricky Nauvaldy Ruliputra, Yudho Giri Sucahyo, Arfive Gandhi, and Yova Ruldeviyani. Why do enterprises adopt natural language processing services? startups' landscape and opportunities in artificial intelligence. *2019 International Conference on Advanced Computer Science and information Systems (ICACSIS)*, pp. 313–318, 2019.

[3] Ayain John, Abhigna, Adithi, Harusha, and Kavya. HEALTH CARE CHATBOT. *Yugosl. J. Oper. Res.*, Vol. 9, No. 8, pp. 297–303, 2022.

[4] A Niam Ibna Riza, Indiana Hidayah, and Paulus Insap Santosa. Use of chatbots in E-Learning context: A systematic review. In *2023 IEEE World AI IoT Congress (AIIoT)*, pp. 0819–0824. IEEE, 2023.

[5] Abhishek Savanur, M Niranjanamurthy, M P Amulya, and P Dayananda. Application of chatbot for consumer perspective using artificial intelligence. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1479–1483, 2021.

[6] Natalie. ChatGPT — release notes. `https://help.openai.com/en/articles/6825453-chatgpt-release-notes`, 2024. Accessed: 2024-1-22.

[7] Sissie Hsiao and Eli Collins. Try bard and share your feedback. `https://blog.google/technology/ai/try-bard/`, 2023. Accessed: 2024-1-25.

[8] Yiming Tan, Dehai Min, Yu Li, Wenbo Li, Nan Hu, Yongrui Chen, and Guilin Qi. Evaluation of ChatGPT as a question answering system for answering complex questions. *arXiv preprint arXiv:2303. 07992*, 2023.

[9] Nigar M Shafiq Surameery and Mohammed Y Shakor. Use chat GPT to solve programming bugs. *International Journal of Information technology and Computer Engineering*, No. 31, pp. 17–22, 2023.

[10] Md Asraful Haque and Shuai Li. The potential use of ChatGPT for debugging and bug fixing. *EAI Endorsed Trans AI Robotics*, Vol. 2, , 2023.

[11] Sunhao Dai, Ninglu Shao, Haiyuan Zhao, Weijie Yu, Zihua Si, Chen Xu, Zhongxiang Sun, Xiao Zhang, and Jun Xu. Uncovering ChatGPT's capabilities in recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys '23, pp. 1126–1132, New York, NY, USA, 2023. Association for Computing Machinery.

[12] Senem Yildiz. Social presence in the Web-Based classroom: Implications for intercultural communication. *Journal of Studies in International Education*, Vol. 13, No. 1, pp. 46–65, 2009.

[13] T M Holtgraves, S J Ross, C R Weywadt, and T L Han. Perceiving artificial social agents. *Comput. Human Behav.*, Vol. 23, No. 5, pp. 2163–2174, 2007.

[14] Zhenhui Peng, Kaixiang Mo, Xiaogang Zhu, Junlin Chen, Zhijun Chen, Qian Xu, and Xiaojuan Ma. Understanding user perceptions of robot's delay, voice Quality-Speed trade-off and GUI during conversation. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[15] Randy Antonio, Nadya Tyandra, Linggar Tembus Nusantara, Anderies, and Alexander Agung Santoso Gunawan. Study literature review: Discovering the effect of chatbot implementation in e-commerce customer service system towards customer satisfaction. In *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 296–301. IEEE, 2022.

[16] Ulrich Gnewuch, Stefan Morana, Marc T P Adam, and Alexander Maedche. Faster is not always betterr: Underrstanding the effect of dynamic response delays in human-chatbot interraction. 2018.

[17] Ulrich Gnewuch, Stefan Morana, Marc T P Adam, and Alexander Maedche. Opposing effects of response time in Human–Chatbot interaction. *Business & Information Systems Engineering*, Vol. 64, No. 6, pp. 773–791, 2022.

[18] Shankar Sundaresan and Zuopeng Zhang. AI-enabled knowledge sharing and learning: redesigning roles and processes. *International Journal of Organizational Analysis*, Vol. 30, No. 4, pp. 983–999, 2021.

[19] Yingnan Shi and Bingjie Deng. Finding the sweet spot: Exploring the optimal communication delay for AI feedback tools. *Inf. Process. Manag.*, Vol. 61, No. 2, p. 103572, 2024.

[20] Karma Choedak. The effect of chatbots response latency on users' trust. 2020.

[21] paul-gauthier. Speed benchmarks of GPT-4 turbo and gpt-3.5-turbo-1106. `https://aider.chat/docs/benchmarks-speed-1106.html`, 2023. Accessed: 2024-1-23.

[22] Satoshi Sato, Suguru Matsuyoshi, and Yohsuke Kondoh. Automatic assessment of Japanese text readability based on a textbook corpus. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Daniel Tapias, editor, *Language Resources and Evaluation*, pp. 654–660. European Language Resources Association, 2008.

[23] Jae-Ho Lee and Yoichiro Hasebe. Readability measurement of japanese texts based on levelled corpora. *The Japanese language from an empirical perspective: Corpus-based studies and studies on discourse.(Irena Srdanovic and Andrej Bekes(eds.))*, pp. 143–168, 2020.

[24] 建石由佳, 小野芳彦, 山田尚勇. 日本文の読みやすさの評価式. 情報処理学会研究報告ヒューマンコンピュータインタラクション（HCI）, Vol. 1988, No. 25, pp. 1–8, 1988.

[25] Akira Sasaki, Masato Hirakawa, Shintaro Horie, and Tomoaki Nakamura. ELYZA-tasks-100: 日本語 instruction モデル評価データセット. `https://huggingface.co/elyza/ELYZA-tasks-100`, 2023. Accessed: 2024-1-10.

[26] ELYZA, Inc. Meta の「Llama 2」をベースとした商用利用可能な日本語 LLM「ELYZA-japanese-Llama-2-7b」を公開しました. `https://note.com/elyza/n/na405acaca130`, 2023. Accessed: 2024-1-23.

[27] Marie-Claire Jenkins, Richard Churchill, Stephen Cox, and Dan Smith. Analysis of user interaction with service oriented chatbot systems. In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, pp. 76–83. Springer Berlin Heidelberg, 2007.

[28] Tibert Verhagen, Jaap van Nes, Frans Feldberg, and Willemijn van Dolen. Virtual customer service agents: Using social presence and personalization to shape online service encounters. *J. Comput. Mediat. Commun.*, Vol. 19, No. 3, pp. 529–545, 2014.

[29] Thomas Holtgraves and Tai-Lin Han. A procedure for studying online conversational processing using a chat bot. *Behav. Res. Methods*, Vol. 39, No. 1, pp. 156–163, 2007.

[30] David Gefen and Detmar W Straub. Gender differences in the perception and use of E-Mail: An extension to the technology acceptance model. *Miss. Q.*, Vol. 21, No. 4, pp. 389–400, 1997.

[31] OpenAI. New models and developer products announced at DevDay. `https://openai.com/blog/new-models-and-developer-products-announced-at-devday`, 2023. Accessed: 2024-1-30.