

雑談チャットボットの返信に意図的な 遅延をつけることによる影響の調査

高橋康汰¹ 山田周平¹ 越後宏紀¹ 阿部花南¹ 小林稔²

概要: AI技術の発展により、雑談チャットボットは人間と話しているかのような自然な会話が可能になっている。一方で多くのボットは返信が非常に早く、人間からの返信だと感じにくいことがある。そこで本報告ではチャットボットの返信に意図的な遅延を加えることで、人間と会話をしているように感じられるかについて調査を行った。本報告ではボットの返信に追加の遅延を加えないボット、ランダムな遅延を加えたボットの他に、より人間らしい遅延を感じさせるために、返信の文字数に応じた遅延を加えたボットの合計3種類作成した。また、それぞれのボットと会話をさせ、どのボットが一番人間らしいと感じたかを比較する実験を行った。その結果、文字数に応じた遅延を加えたボットが一番人間らしいと回答した人が最も多いことがわかった。

キーワード: 雑談チャットボット, 遅延, 人間らしさ

Investigating the Impact of Intentional Delays in Chatbot Replies

KOUTA TAKAHASHI¹ SHUHEI YAMADA¹ HIROKI ECHIGO¹
KANAN ABE¹ MINORU KOBAYASHI²

Abstract: With the development of AI technology, chatbots are able to have conversations with natural content as if they were talking to a human. On the other hand, many chatbots reply very quickly, and it is sometimes difficult to feel that the reply is from a human. In this paper, we investigated the possibility of adding deliberate delays to the replies of chatbots in order to make people feel as if they were talking to a human.

In this paper, we created three types of bots: one with no additional delay, one with a random delay, and one with a delay based on the number of characters in the reply in order to give a more human-like delay. An experiment was conducted to compare which bot felt the most human-like by having conversations with each bot. In this experiment, the BOT with a delay based on the number of characters was the most human-like.

Keywords: Chatbots, Delay, Humanity

1. はじめに

近年、スマートフォンの普及やオンライン会議の増加により Slack や LINE といったチャットツールの利用者は増加している[1][2]、同様にテキストベースのチャットボットの利用者も増加している。また AI 技術の発展により、人間と話しているかのような自然な会話が可能なボットも多く存在していて、日本マイクロソフトが開発した AI の“りんな”などが有名である[3]。本報告では、こうしたボット雑談が可能なチャットボットを「雑談チャットボット」と定義する。これらの雑談チャットボットは返信速度が非常に早いものがほとんどで、人間らしいと言われるとそうではないという側面がある。

そこで本報告では、雑談チャットボットの返信に意図的な遅延を加えることで人間らしい会話ができるようになることを考え調査を行った。実験は参加者に遅延を加えないボットと、それぞれ特徴の異なる2種類の遅延を加えたボットの合計3種類と会話をしてもらい、どのボットが人間らし

いか比較することで行った。

本研究の目的は雑談チャットボットの返信にあえて遅延を加えることで人間らしさを向上させることである。また人間らしい遅延について検証することも目的としている。この詳細は本稿の3章で述べる。

2. 関連研究

本報告では、人間らしいチャットボットの作成を目指している。そこで人間らしい機械とは何かについての関連研究を紹介する。

アラン・チューリングが提案したチューリングテストは、ある機械が人間的であるかを人間が判定するテスト[4]である。原案では、テキストチャット形式でテストをすることが示されている。チューリングテストは古くから行われているもので、人工知能の歴史を振り返る上で必ずと言っていいほど言及されるものである[5]。チューリングテストは人工知能という言葉が生まれる前から提唱されていて、

1 明治大学大学院先端数理科学研究科
Meiji University Graduate School of Advanced Mathematical Science
2 明治大学総合数理学部
Meiji University School of Interdisciplinary Mathematical Sciences

人工知能研究の初期には知能とは何かが大きなテーマとなり、しばしば話題に取り上げられた。チューリングテストに関する代表的なシステムとして ELIZA [6]が挙げられる。ELIZA は PARRY [7]と並んで初期の会話ボットとしても有名であるが、人工知能ではなく人工無脳である。当時のコンピュータの性能があまり高くないこともあって質問の内容を分析し、機械が知っているワードがあればそれについて回答を行い、わからないことは人間のように受け流すことで人間らしさを演出するものとなっている。ELIZA には精神科医という設定があり、一般的な人間同士のコミュニケーションではなく医者と患者という限定されたコミュニケーションであるが、それでも多くの患者から信用できる医者だと感じられたと報告されている[8]。橋本[9]は、チューリングテストは高い知能を目指すというのではなく、むしろ人間らしさや人間的な心を機械に求めているということを述べている。「しかしわずかな数千行の簡単なプログラムが心を持つとは当然考えられないとしていて、これはチューリングテストに合格したからといってそれは心を持つとは言えないというよい例だともしている」[10]。それを踏まえた上で、橋本は人工知能を目指す到達点は2つあるとしていて、そのうちの1つは人間と同様の振る舞いができることと述べている[9]。更にその中に、姿かたちを人間に似せることと人間との関係や機能を人間に似せることの2つのアプローチがあると述べている。また人間に非常によく似たロボットなら違和感のないコミュニケーションが可能だろうとしていて、それは人間らしい心を持っているとみなせるだろうとも述べている。本研究は後者のアプローチで人間らしいチャットボットの実現を目指している。

こうした機械に人間らしさを追求する研究は、最近でも行われているのだが、特にチャットボットに注目した研究をいくつか紹介する。神場[11]は、これまであまり検討されなかったチャットボットの発話タイミングに注目している。この研究では、テキストタイプのものだけでなく音声対話を行うチャットボットに関しても、発話タイミングの微妙な制御をすることによって人間に与える印象の違いがあるかを論考している。その中でそれぞれ発話タイミングの異なる2種類のチャットボットの動作を模したアニメーションを見せ、どちらが好ましいかをアンケートで回答を求めたことで調査していた。調査の結果からテキストタイプのチャットボットでは返事が即座に表示されるものよりも一呼吸置いて表示されるものの方が好ましいと回答した人が多いことがわかった。この研究では、一呼吸置いて表示されることが好ましいということは示されているが、具体的にどのくらいの時間を空けることが好まれるのかについては検討されていない。本研究ではこの結果から、遅延を加えることでより人間にとって親しみやすくなると考え、検討及び調査を行った。

杉本ら[12]は、チャットボットの対話が単調だと退屈感

を感じ飽きやすいという問題点[13]があるとし、ユーザの会話の継続性を高めるために「レッテルを貼る」という相手の人格や能力を一方向的に格付けする行為に着目している。「レッテルを貼る」行為は相手を茶化すことやブラックジョークのようなもので、特に親密度の高い人間同士では冗談としてレッテル貼りが行われている。この研究ではユーザがボットに送信したメッセージをもとに、学習データを用いて生成されたレッテルをユーザへの誹謗中傷にならないかを判定し、そうではない場合にレッテルを追加した返信メッセージを送信する雑談チャットシステムを作成し、システムの評価を行なっている。結果の分析から、会話時間や退屈感などの項目において有意差を得ていて、会話の継続性を高めることができたと報告されている。本報告では時間にのみ着目して他の要素は検討していない。しかし今後は、こうした人間を真似た発言をチャットボットに導入することも今後検討していき、より人間らしいチャットボットを目指していきたい。

また石井ら[14]は、人同士のテキストチャットでは相手からの応答が返ってくるまでに、直前に送信したメッセージの内容に関する訂正や補足のメッセージを送信することがあると述べている。従来のチャットボットではユーザの発話に対して即時に回答するため、ユーザが連続してメッセージを送信する意志を持っている場合、ボットが中途半端に意味を理解してしまい対話破綻の原因になってしまうと指摘している。この研究の目的は本研究と異なり人間らしいボットの作成が目的ではないのだが、本研究と同じようにチャットボットの返信を数秒遅延させることでユーザに与える印象を調査している。本研究ではこの研究の実験方法の一部を参考に実験を行った。

3. 研究目的

本研究の目的は、チャットボットの返信に意図的な遅延をつけることで人間らしさを向上させることである。雑談チャットボットを利用するユーザは暇潰しや会話の相手を求めている。また仕事などの不安を AI に相談したいという人も増えている[15]。人間らしさの向上はボットへの親しみやすさに繋がり、利用率の向上や会話への満足度の向上などが期待できるのではないかと考えている。また、それ以外にもボットに本来必要ではない要素を加えることによって思わぬ効果が生まれることも期待している。

人間らしさの向上のためにただ遅延を加えるだけでなく人間らしい遅延とは何かを考えた。通常人間がチャットを送信する際、文章を考えること、そしてそれを入力することといった手順を踏む。これを再現するためにはチャットボットの返信の文字数に応じた遅延を加えることが適切であると考えた。本研究では文字数に応じた遅延が、会話の人間らしさの向上に寄与するか検証することも目的である。

4. 提案手法

4.1 チャットボットについて

それぞれ性質の異なる遅延を付与した2種類のチャットボットと遅延を付与しないボット合計3つのボットを作成した。今回作成したものはLINEのチャットボットで、ユーザが送信したメッセージに対して自動で返信を行うものである。チャットツールの中でLINEを選択した理由は、今回の会話内容が雑談であることからSlackといったビジネス向けツールよりも比較的気軽に話せる[16]チャットツールのためである。

ボットとしての基本機能はLINEの開発用ページから作成し、返信機能はボットと連携したGoogle Apps Script上でプログラムを動かすことで実現している。会話のログはGoogleスプレッドシートに保存され、返信内容は対話APIのchaplusを用いてユーザが送信したメッセージをもとに適切だと判断されたメッセージを生成し、それをユーザに送信する。なお返信されるメッセージはAPIが分析を行い適切なメッセージの候補の中から一つ選ばれる。そのため全く同じ内容のメッセージを送信した場合でも異なる返事が返ってくる可能性がある。今回使用したAPIは直前に送られてきたメッセージに対する返信のみを生成するので、それ以前の会話の流れに合った返信をすることはできない。また、今回作成したボットには人間らしさをより向上させるためそれぞれタカヒロ、トシヤ、エイジという日本人らしい名前をつけた。

4.2 遅延の設定について

本節では、今回作成したボットに加えた遅延の詳細を以下に述べる。まずランダムな遅延を加えたボットでは、プログラム上で1~9秒の中でランダムな秒数を、メッセージを受信する毎に抽選し、それに応じた遅延を、プログラムを一時停止することで実現している(図1)。

文字数に応じた遅延を加えたボットでは、ユーザに送信するメッセージの文字列を取得しカウントすることで実現している。また今回の実験では1文字あたり1秒の遅延を加えるように設定した。当初は著者がスマートフォンでフリック入力の手速を計測し、その結果から1文字あたり0.5秒の設定にする予定だったが、取得する文字列に、漢字や数字など文字数と入力回数が一致しない文字が含まれていて、1文字あたりの入力回数が必ずしも1回ではないということや、またランダムな遅延を加えたボットと明確な差別化をするために1秒に設定した。しかし、LINEボットの仕様により30秒以上の遅延を加えようとすると返信メッセージを送信できない場合があることがわかった。この仕様は実験開始後に判明したため今回の実験ではそのままにしている。

また今回作成したボットは遅延がないものも含めて全て

のボットでプログラムの処理に4秒程時間がかかる。そのため実際の遅延は追加された遅延に4秒足したものとなる。

4.3 会話内容について

会話の内容はあらかじめボットによって指定される。今回はチャットボットを実験参加者に新規に導入してもらい実験を行った。そのためLINEボットの機能から、ボットを友だちに追加したタイミングでメッセージが自動送信される時に図2のように会話のお題を送信した。会話のお題は先行研究[14]で使われていたものを参考に3種類用意した。お題は以下の通りである。

- 先週何日研究室にきましたか？
- 好きな料理について教えてください
- 好きな授業について聞かせてください



図1 チャットボットの返信を待つ様子
Figure 1 Waiting for the chatbot to reply.

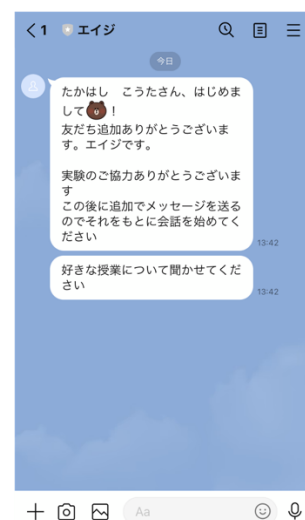


図2 チャットボットを友だちに追加した時のメッセージ
Figure 2 Message when you install the chatbot

5. 実験

5.1 予備実験

ボットとの会話時間の設定や実験案の改善点を見つけるため予備実験を実施した。参加者は大学院生1名で、実験ではオンライン会議ツールを利用し、実験内容の説明やボットのQRコードの表示を行なった。オンライン会議ツールを使用した理由は、2021年8月の時期に新型コロナウイルスの流行のため対面で実施することが困難だったためである。

予備実験を実施したところ会話時間は5分が適当であると判断した。また遅延があるボットとないボットでユーザのメッセージ送信量に若干差があった。そこで後述の実験では、メッセージの送信量が同じくらいになるよう調整を行うために遅延がないものに関しては他のボットより1分短い4分に設定した。

5.2 実験概要

提案手法である文字数に応じた遅延が、人間らしさの向上に寄与するかどうかを検証するために、3種類のボットと5分程度会話してもらい、どれが一番人間らしさを調査する実験を行った。参加者は大学生・大学院生合計19名である。また予備実験は参加者が1名であったが、4-5名程度であれば同時に実験を行っても問題ないと判断した。そのため実験参加者を、5人程度のグループに分け、グループ単位で実験を実施した。実際に参加者を5人1組のグループを3つと4人1組のグループを1つに分け、それぞれA,B,C,Dと割り振り、そのグループのメンバーに3種類のボットとチャットで会話してもらった。本実験でのグループとは、会話するボットの順番が同じ実験参加者の集まりであり、実験におけるチャットでの会話はボットと実験参加者が1対1である。

5.3 実験方法

今回の実験では、オンライン会議ツールを使用した場合と、対面のどちらかの環境で実験を行った。オンライン会議ツールを使用した理由は、5.1節の予備実験と同様である。あらかじめ実験参加者には3種類のボットと会話してもらおうこと、そのうちの1つは人間が操作しているもので残りの2つはAIであると説明し、操作している人間が誰であるかについては説明しなかった。これはチューリングテスト[4]を参考にしている。また今回の実験ではそれぞれのボットの遅延に関する説明は行わなかった。実験中に返信が来るまでの時間を意識しすぎて、会話が疎かになってしまうことを恐れたためである。オンライン会議ツールと対面どちらの場合でも、PowerPointのスライドで順に3種類のボットのQRコードを共有し、実験参加者の所有するスマートフォンで読み取ってもらいボットを友だちに追

加させた(図3)。全てのボットと予備実験をもとに設定した時間、会話をした後に、アンケートでどのボットが一番人間らしいか、またどのボットが一番人間らしくなかったかそれぞれ回答を求めた。さらに、それぞれの理由も記述式で回答を求めた。

5.4 実験結果

実験結果を図4と図5に記す。一番人間らしいボットに関しては文字数に応じた遅延を加えたボットCが一番多い結果になった。また一番人間らしくないボットに関しては追加の遅延を加えなかったボットAが一番多い結果になった。

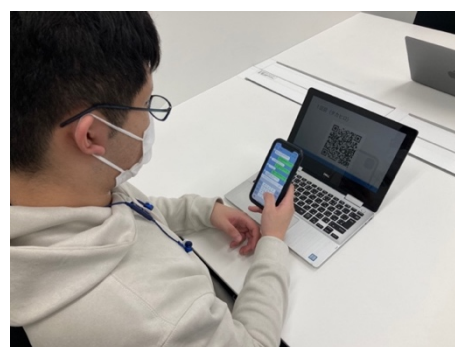


図3 実験の様子

Figure 3 Scene of the experiment

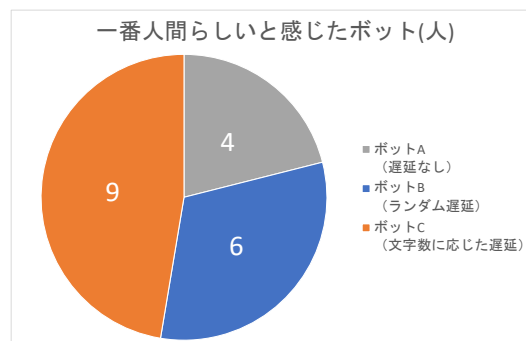


図4 一番人間らしいと感じたボット

Figure 4 The chatbot that felt the most human.

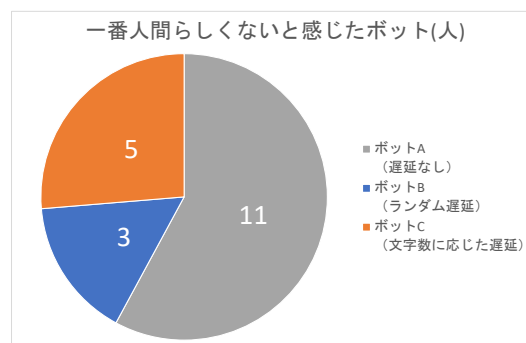


図5 一番人間らしくないと感じたボット

Figure 5 The chatbot that felt the least like a human.

またそれぞれの回答理由の例である。

- 1) 一番人間らしいと感じたボット
 - 会話が成立していた (ボット A)
 - 話のテンポが一番よく感じた (ボット B)
 - 返信速度が遅かった (ボット C)
- 2) 一番人間らしくないと感じたボット
 - 返信が早すぎるから (ボット A)
 - 文章がおかしかった (ボット B)
 - 急に関係ない話を始めたため (ボット C)

また回答理由を全てまとめたところ返信時間と会話内容の2種類の答えに分類することができた。図6と図7にそれぞれ示す。

それぞれのグループ毎のメッセージ送信数は表1の通りである。なおグループDは他のグループと比べて1人少ないため他のグループと比較して送信数が少なくなっている。

またボットCの文字数に応じた遅延を加えたものの平均文字数は約23.7であった。よって4.2で述べたプログラムの処理にかかる4秒を足して、ボットCは平均27秒程度の遅延が加わっていたことになる。

表1 グループ毎のメッセージ送信数

Table 1 Number of messages sent per group

	ボットA	ボットB	ボットC
グループA	73	71	26
グループB	63	41	18
グループC	77	85	44
グループD	42	35	20

6. 考察

今回の実験では、チャットボットの人間らしさを返信速度で判断している参加者も多かった。一方で、会話の内容で判断していた参加者も存在することが分かった(図6と図7)。今回の3つのボットは明確に返信速度が違っていても関わらず、この結果になっている。この要因は2つあると考える。一つは今回使用した対話APIがあまり適切な返事ができないことで、それにより適切な返事をしなかったボットに対して人間ではない印象を強く与えてしまったのではないかと考えられる。もう一つは人間らしさを判断する上で返信速度以上に会話内容を重視している人が多いのではないかとということである。これらの問題点は、より精度の高い雑談対話可能なAIを用いることができれば解消できると考えている。

また今回のそれぞれのボットを選んだ理由についての回答で、人間らしいボットに文字数に応じた遅延を加えたもの(ボットC)と回答した人、人間らしくないボットに追加の遅延がないボット(ボットA)と回答した参加者に共通して他のボットを選んだ参加者に比べて返信時間を理由に回答していた参加者が多かった(図6と図7)。これは本報告の目的通りに返信時間を調整することで人間らしさを向上させることが出来ていて、また遅延がないボットは人間らしさが足りないものであると考える。

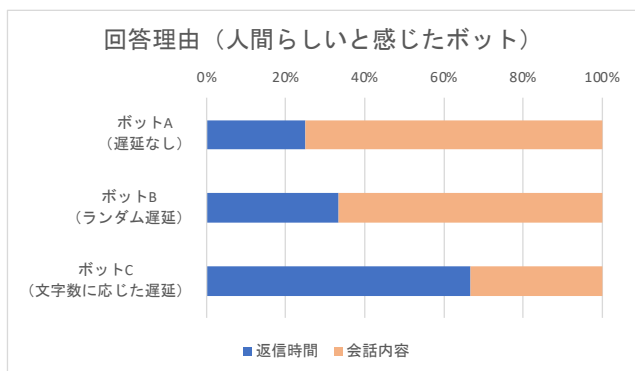


図6 回答理由 (人間らしいと感じたボット)
 Figure 6 Reason for answering (Bot felt human-like)

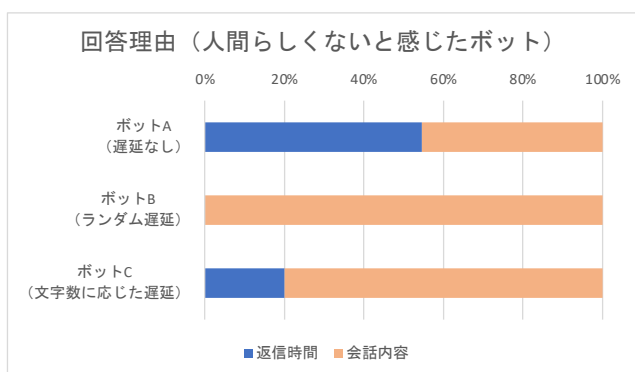


図7 回答理由 (人間らしくないと感じたボット)
 Figure 7 Reason for answering (Bot felt not human-like)

7. 今後の展望

今回の調査では遅延の性質を変えて比較を行い文字数に応じた遅延を加えたボットが比較的人間らしいボットだという結果となった。今後は1文字あたりの遅延がそれぞれ異なるボットを複数用意し、その微妙な違いにより印象がまたどう変わるかの調査を行いたいと考えている。更に今回は一切考慮しなかったメッセージを考える思考時間などの入力以外の要素についても検討していきたい。また時間に限らない先行研究[12]のような要素についても取り入れることでより人間らしいボットを目指していきたいと考えている。

8. 終わりに

チャットボットの利用者は年々増加し、AI技術の発展に伴い本物の人間と話しているような自然な会話が可能になりつつある。しかしそれらは人間のような会話こそ可能であるものの、人間らしいボットになっているとは言えないのではないだろうか。本報告ではチャットボットにおける人間らしさとは何かについて考察し、それを実現する手法を示し調査を行った。調査の結果、提案手法の文字数に応じた遅延をボットの返信に加えたものが一番人間らしいボットだと回答した人が多かった。今後は思考時間などのメッセージを入力する以外の時間的要素についての検討や、1文字あたりの遅延を変化させ、より人間らしいボットを目指していきたい。

謝辞 今回の調査にあたり、実験に参加していただいた小林研究室の皆さまに感謝する。

参考文献

- [1] “LINE Business Guide 2021年7-12月期,” https://www.linebiz.com/sites/default/files/media/jp/download/LINE%20Business%20Guide_202107-12.pdf, (参照 2021-12-21).
- [2] “日本の働き方改革に貢献する Slack,” <https://slack.com/intl/ja-jp/blog/news/%E6%97%A5%E6%9C%AC%E3%81%AE%E5%83%8D%E3%81%8D%E6%96%B9%E6%94%B9%E9%9D%A9%E3%81%AB%E8%B2%A2%E7%8C%AE%E3%81%99%E3%82%8B-slack>, (参照 2021-12-21).
- [3] rinna 株式会社. “りんな,” <https://www.rinna.jp/>, (参照 2021-12-21).
- [4] A. Turing. Computing machinery and intelligence, *Mind*, vol. 59, no. 236, pp. 433-460, 1950.
- [5] 松原 仁. “チューリングテストとは何か,” *人工知能学会誌* 26巻1号, pp. 42-44, 2011.
- [6] Joseph Weizenbaum. ELIZA—a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, pp 36-45, 1966.
- [7] Cerf, Vincent. PARRY Encounters the DOCTOR, RFC0439, 1973
- [8] Casti, J. L. Paradigms lost. Images of man in the mirror of science, William Morrow. New York, NY, USA.1989.565 p.
- [9] 橋本 文彦. 「人間らしさ」と人工知能, *科学・技術研究* 第3巻2号, pp. 101-106, 2014.
- [10] 橋本 文彦. 機械は心をもちうるのか, *心理学評論*, Vol.37, pp. 533-544, 1994.
- [11] 神場 知成. チャットボットにおける時間要素の設計について, *情報処理学会インタラクシオン* 2018, pp.447-451, 2018.
- [12] 杉本 隼斗, 中山 裕貴, 渡邊 一樹, 中園 歩, 濱川 礼. “レットテルを用いた雑談向けチャットボット「Lamb」の提案,” *エンタテインメントコンピューティングシンポジウム* 2019 論文集, pp. 128-133, 2019.
- [13] 岩倉亮介, 吉川大弘, ジメネスフェリックス, 古橋武. “Twitter データを用いたユーモア語句自動生成手法に関する一検討,” *The 31st Annual Conference of the Japanese Society for Artificial Intelligence*, 2017.
- [14] 石井俊太郎, 下田怜奈, 藤江真也. チャットボットの発言タイミング制御とその印象評価, *FIT* 2018, pp. 167-168, 2018
- [15] “日本の「職場における AI」調査: AI 利用は世界 11 カ国で最下位も、87%が不安やストレスを相談する相手としてロボット・AI を受け入れると回答,”

<https://www.oracle.com/jp/corporate/pressrelease/jp20201104.html>, (参照 2021-12-21).

- [16] 時岡良太, 佐藤 映, 児玉夏枝, 田附紘平, 竹中悠香, 松波美里, 岩井友香, 木村大樹, 鈴木優佳, 橋本真由里, 岩城昌子, 神代末人, 桑原知子. 高校生の LINE でのやりとりに対する認知に現代青年の友人関係特徴が及ぼす影響, *パーソナリティ研究*, vol. 26, pp. 76-88, 2017.