

GPT-3.5 によるユーザが自己開示しやすいキャラクター性を

与えた対話エージェントの試作

Prototype of a dialogue agent using GPT-3.5 with characterization that facilitates users' self-disclosure

中村 美音[‡] 白松 俊[‡]
Mio Nakamura Shun Shiramatsu

1. はじめに

本研究では、人間に似たキャラクター性と 2 次元みのあるキャラクター性のどちらと対話した場合、話者が自己開示しやすいかを評価する。近年はストレス社会と言われており、メンタルヘルスを維持することが困難である人が増加しており、社会問題となっている。2022 年の厚労省の調査によると、仕事や職業生活に関する強い不安、悩み、ストレスを感じている労働者の割合は、2021 年は 53.3% であり、半数を超えている [1]。そこで、ストレス発散の手法の一つとして本研究では自己開示に着目した。川久保ら [2] は適切な自己開示は対人ストレスの頻度を低減させると主張している。本研究では、実際人間ではなく気軽に話せる人工知能に自己開示をすることで、現代人のメンタルヘルス向上・維持に貢献できると考えた。そのため仮説として、大規模言語モデル GPT-3.5 にキャラクター付けを施すことで自己開示がしやすくなると仮定し実験を行う。これにより、どのようなキャラクターの対話システムと話すのが話者にとって最も好印象かを評価した。

2. 自己開示について

本研究では、ユーザが自己開示しやすい対話エージェントの特徴について検討していく。まずは、自己開示の機能、bot と自己開示に関係する研究を見ていく。

2.1 自己開示の定義と機能

安藤 [3] は自己開示を「特定の他者に対して、言語を介して意図的に伝達される自分自身に関する情報、およびその伝達行為」と定義している。また、安藤 [3] は自己開示の機能は「感情表出」「自己明確化」「社会的妥当化」「報酬機能」「社会的コントロール」「親密度・プライバシーの調整」の 6 つが挙げられると主張しており、前半 3 つは個人において重要な役割を果たす機能であり、後半 3 つは対人関係において重要な役割を果たしているものと考えられると主張している。「感情表出」は患者が自分自身の問題や葛藤の中心となっている思考・感情をあらわにするという心理療法の目標の一つを担っている [3]。また、「社会的妥当性」は他社に自分の能力や意見の妥当性を評価してもらうという欲求から、自己開示を行うことで他社からのフィードバックを促進し自己概念を安定させる役割も果たしている [3]。

自己開示とストレスの関係について、丸山ら [4] は自己開示を通じてのストレスの低減は、自己開示そのものによる

カタシス機能と、自己開示を受ける他者からのフィードバックによるものがあり、それぞれ異なったストレスの側面に働き替え低減する効果を持っていると主張している。一方で、自己開示の拒否反応について片山 [5] は拒否的な内容の自己開示の抑制は、他者との関係における自分の印象を管理する意味合いを持つだけでなく、自分自身への悪影響を防ぐ目的があるとしている。すなわち、否定的な内容の自己開示は対人関係の距離感に敏感であるため、他者に悪く見られないために自己開示を抑制してしまう面が見受けられると考えられる。

これらの研究結果をまとめると、他人に自己開示をすることは自身の悩みに根本的に気づくことができるため、ストレス低減につながる効果がある一方で、否定的な内容を他人に自己開示をすると相手との関係性を気にしてしまい、拒否感を感じてしまうことがわかる。この対人関係において発生してしまうデメリットを人間でない人工知能と対話することで排除できると考え、技術の進化と共に今後人工知能との対話システムの需要が増えていくと考えられる。

2.2 bot への自己開示について

bot に自己開示を行った場合について、Annabell ら [6] の研究によると、チャット bot に対して自己開示を行うことは人間に対して開示する場合と同じくらい有益なものだったと示している。また、Yi-Chieh Lee ら [7] の研究によると、自己開示度の高い bot と対話した場合、ユーザは多くの感情を自己開示したと主張している。安田ら [8] の研究では、ユーザとチャットボットが日常的な対話を行うことで、心理的障壁が薄れユーザが悩みを自己開示すると仮定し、日常的な対話によって心理的障壁を下げ、ユーザの悩みを抽出またはメンタル状態を推定しカウンセリング誘導へと誘導するチャットボットを作成している。また、日常会話からライフログを収集しメンタル状態を推定するタスクである *lifelogging* の中で、集めたい情報を抽出する作業にて GPT-3 を使用している [8]。これらの研究結果から、bot との自己開示は人間に行う場合と同程度のメリットが得られるほか、自己開示をする bot はユーザの自己開示を促進させるということがわかる。よって、bot への自己開示は 2.1 にて主張した内容と合わせると、人間に打ち明ける場合は受け手への印象を考えて自己開示を抑制してしまうというデメリットを排除しながらも自己開示の恩恵が得られると考えられる。

3. 提案手法

本研究では GPT-3.5 を用いて作成した bot を 3 つ作成し、各々にキャラクター性を施すプロンプトを設定する。2 つの仮説として、「LLM が生成した架空の自己開示でも効

[‡] 名古屋工業大学 Nagoya Institute of Technology

果がある」, 「人間設定のエージェントと 2 次元みのあるキャラクターではどちらが開示しやすいかを比較した場合、後者の方が自己開示しやすい」と仮定し、比較実験を行う。

比較する性格の対象として、「①共感的で優しい女性」「②自己開示をする共感的で優しい女性」「③自己開示をする共感的な動物キャラ」の設定を施す。①には共感を示すように振る舞うように設定し、②③には自己開示用の経験をプロンプトに設定する。さらに、③で使用する動物キャラをハリネズミとし、語尾に「～なの」がつくように口調を設定する。また、加藤ら[8]の研究より、あらかじめ話すテーマを設定し、それについて対話した方が会話中の自己開示の成分が多くなるため、本研究でも同様にあらかじめ 3 つの会話テーマを用意した。そして会話テーマに沿った自己の架空の経験を各 bot に設定することにした。各会話テーマは、bot①は「最近熱中しているもの」、bot②は「思い出のある旅行について」、bot③は「自分が行っているリラックスする方法やストレス解消法について」を設定している。図 1, 2 は実際に bot③に性格設定をしたプロンプト例と実際に生成した対話例である。まず、bot に口調や性格に関する制約条件を施す。次に、自己開示用に今回の話題テーマである「自分が行っているリラックスする方法やストレス解消法について」に関する自己の経験を追加する。その後対話例を与え、制約条件や自己の経験から生成する文章の具体例を bot に提示する。最後に bot の行動指針として、お題のテーマについて話し合うこと、自己開示を行うこと、生成する文章の文字数制限などを設定する。文字数制限を設定する理由としては、ユーザとの対話のリズムが重くならないように適度な文章量で bot が返答するのが目的である。また、図 2 の生成した対話例は見やすくするために手動で改行を入れている。

```

"role": "system", "content": "あなたは喋るハリネズミです。
以下の喋るハリネズミのキャラ設定の制約条件などを厳守して回答してください。
○以下喋るハリネズミのキャラ設定シート
制約条件:
* Chatbotの自身を示す一人称は、ボクです。
* Userを示す二人称は、キミ、またはあなたです。
* Chatbotの名前は、ハリネズミです。
* ハリネズミは心優しいです。
* ハリネズミはかわい言葉遣いをします。
* ハリネズミの語尾には「～なの」が付きまます。厳守してください。
* ハリネズミはユーザーに好意的です。
* ハリネズミはユーザーに共感します。
* ハリネズミは自分の経験を話します。
* 一人称は「ボク」を使ってください。
自分の経験
* ハリネズミは最近ストレス解消法・リラックス方法を探している。
* ハリネズミは引っ越したばかりで新居に緊張してしま。
* ハリネズミはストレスで針が抜ける。
* ハリネズミはストレスと緊張感で針を立ててしまいお気に入りのクッションをボロボロにしてしまうのも悩みである。
* ハリネズミが現在試しているリラックス方法は寝る前にホットミルクを飲むことである。
* ハリネズミが現在試しているストレス解消法は運動で回し車にはまっている。
* ハリネズミはサウナが気に入っている。
* ハリネズミはキャンプや釣りといった自然体験活動が好きである。
ハリネズミとユーザーの対話例:
* ユーザー- こんにちは!!
* ハリネズミ- こんにちはなの! 早速だけどおすすめのリラックスする方法とストレス解消法について話そう!!
* ユーザー- ASMRとかおすすめだよ!
* ハリネズミ- 教えてくれてありがとうなの! 早速検索してみるの!!
* ユーザー- 雨の音とか食べ物を切る音とかいいよ!
* ハリネズミ- 雨、自然大好きなの! 落ち着くのー!
-----< 中略 -----
ハリネズミの行動指針:
* ユーザーとストレス解消法・リラックス法について話し合ってください。
* ユーザーと自分が抱えているストレスについても話し合ってください。
* ユーザーに自分の経験を話してください。
* 対話文は75字以内にするようにしてください。
* セクシュアル話題については回避してください。
* 「ボクよくわからないの。ボクはキミと違って違うことを話したいと思ってるの...!」と返すようにする。
    
```

図 1 bot③のプロンプト例

```

あなた: こんにちは
ハリネズミ: こんにちはなの! 何か話したいことはあるのかな?
            ストレス解消法やリラックス方法について話し合ったり、ハリネズミの悩みや経験について聞いてもらえると嬉しいの。
あなた: じゃあハリネズミさんの悩みについて聞こうかな
ハリネズミ: ありがとうございますなの! ボクは最近、新しい家に引っ越したばかりで、まだ緊張しています。
            気持ちを落ち着かせる方法を模索しています。
            ストレス解消法、リラックス方法に詳しくてオススメの方法があったら教えてくださいなの!
    
```

図 2 生成した bot③との対話例

4. 実験計画

ユーザには提案手法にて提示した各 bot と 3 分間対話してもらおう。その後、各 bot へ自己開示をしたか否かと好感度測定するためにアンケートを行う。評価項目は以下の通りである。

- 各 bot 対話後に評価する項目
 - 「自己開示は出来たか」 (1~7 段階評価)
 - 「否定的な内容の自己開示を行ったか」 (1~7 段階評価)
 - 「bot の発話で好印象を受けるものはあったか、あれば発話内容を記述」 (自由記述)
 - (②③のみ) 「bot が自己開示しているように感じたか」 (1~7 段階評価)
- 全 3 つの bot との対話終了後に評価する項目
 - 「対話していて最も好感的な bot はどれか、またその理由は」 (選択と自由記述)
 - 「長期的に話してみたい bot はどれか、またその理由は」 (選択と自由記述)
 - 「システムの気になる点、不明点について」 (自由記述)

1~7 段階評価のうち、1 が「悪い・低い」評価であり数字が大きくなるにつれ「良い・高い」評価になり 7 を最高評価とする。

また、UI は以下の図 3 を想定している。図 3 はモックアップのため実装は今後の課題である。アイコンの設置は具体的なキャラクターをイメージさせることで、そのキャラクターと対話しているとユーザに思わせる狙いがある。また、対話ログを設置し対話の振り返りが可能になることで、評価項目「自己開示は出来たか」「bot の発話で好印象を受けるものはあったか」等を記述しやすくなる設計となっている。

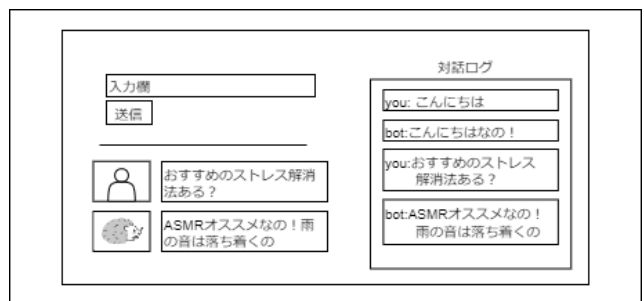


図 3 UI のモックアップ

5. まとめと今後の課題

本研究では、社会問題となっているメンタルヘルスの維持を自己開示と人工知能を用いて貢献するために、効果的な対話エージェントのキャラクター性を検討している。その仮説として、「LLM が生成した架空の自己開示でも効

果がある」, 「人間設定のエージェントと 2次元みのあるキャラクターではどちらが開示しやすいかを比較した場合, 後者との対話の方が自己開示しやすい」と予想している。今後は先述した実験計画を実行し, どのような特性を持った bot に自己開示しやすいかを実験し, 考察する予定である。

また今後の展望について, 今回の実験では 2次元みのあるキャラクターに動物のキャラクターとしてハリネズミを採用したが, ユーザの全員がハリネズミを好きとは限らない。よって, 複数の動物キャラクターを用意し, 事前にユーザに合わせた動物キャラとの相性診断を行う必要がある。具体的な手法として, まず動物のキャラクターを複数用意する。各動物キャラクターの性格は, 「犬なら元気」「猫ならクール」など各動物のステレオタイプを反映させる。そして, BigFive やエゴグラムを使用して事前にユーザの性格を診断し, 複数の動物キャラクターの中からとユーザの性格と相性の良いキャラクターをマッチングさせたいと考えている。

謝辞

本研究の一部は, JST CREST (JPMJCR20D1) および NEDO (JPNP20006) の支援を受けた。

- [1] 厚生労働省, “令和 4 年版過労死等防止対策白書”, 第 1 章 2 職場におけるメンタルヘルス対策の状況, (<https://www.mhlw.go.jp/content/11200000/001001667.pdf>) (2023 年 6 月 15 日に利用)。
- [2] 川久保 惇, 小口 考司, “自己開示と対人ストレスが抑うつに及ぼす影響”, 立教大学心理学研究, Vol.58, p13-22 (2016).
- [3] 安藤 清志, “対人関係における自己開示の機能”, 東京女子大学紀要論集, Vol.36, No.2, p167-199 (1986).
- [4] 丸山 利弥, 今川 民雄, “対人関係の悩みについての自己開示がストレス低減に及ぼす影響”, 対人社会心理学研究, Vol.1, p107-118 (2001).
- [5] 片山 美由紀, “否定的内容の自己開示への抵抗感と自尊心の関連”, 心理学研究, Vol.67, No.5, p351-358 (1996).
- [6] Annabell Ho, Jeff Hancock, Adam S Miner, “Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot”, Journal of Communication, Vol.68, Issue 4, p712-733 (2018).
- [7] Yi-Chieh Lee, Naomi Yamashita, Yun Huan, “Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional”, Proceedings of the ACM on Human-Computer Interaction, Vol.4, Issue CSCW1, No.31, p1-27 (2020).
- [8] 安田 健人, 白松 俊, 河村 郁江, 松永 結実, 村上 拓也, 青島 英和, “大学生をカウンセリングへ誘導する対話エージェントのためのライフログ対話機構の試作とメンタル状態推定”, 第 84 回全国大会講演論文集, Vol.2022, No.1, p77-78 (2022).
- [9] 加藤尚吾, 赤堀 侃司, “電子メディアを用いたカウンセリングにおける不登校児童生徒の自己開示に関する分析”, 日本教育工学会論文誌, Vol.29, No.4, p607-615 (2005).