

Agent adopter の存在を考慮した HAI モデルの実験的検討 - 対話エージェントとの協力型 Escape room を題材とした責任帰属の分析 -

Experimental Investigations of HAI Model

Considering the Presence of the Agent Adopter

-Analysis of Responsibility Attribution in a Cooperative Escape Room with Dialogue Agents-

渡邊英一郎* 宮本友樹 内海彰
Eiichiro Watanabe Tomoki Miyamoto Akira Utsumi

電気通信大学

The University of Electro-Communications

Abstract: エージェントの社会実装を想定した場合、エージェントの導入者が必ず存在する。しかし、これまでの HAI 研究ではエージェントの導入者に対する印象や責任帰属は明らかになっていない。本研究ではエージェントの導入者を Agent adopter (以下, AA) と呼び、著者らが先行研究で提案した AA の存在を考慮した HAI モデルに基づいて、対話エージェントが引き起こすトラブルの責任帰属を検証する実験を行った結果を報告する。実験の仮説として、対話エージェントに対する擬人化性の帰属が高いほど対話エージェントに対する責任帰属が大きくなり、AA に対する責任帰属は小さくなると予想した。実験参加者は制限時間内に対話エージェントと協調タスクを行い、対話エージェントが引き起こすトラブルによりタスクが失敗するシナリオを体験し、擬人化性の帰属や責任帰属に関して主観評価した。

1 はじめに

近年では NICOBO (Panasonic 社) といった HAI 研究の成果を基に開発されたコミュニケーションロボットなどのエージェントが社会実装されている。エージェントの社会実装を想定した場合、エージェントの導入者が必ず存在する。エージェントの導入者とは、例えば公共の場へエージェントを導入した人物・組織などである。しかし、これまでの HAI 研究では、エージェント導入者が系に含まれるインタラクションモデルは殆ど扱われていない。小松 [1] は、トロッコ問題を判断したエージェントへの道徳的な非難が、エージェント自身、その所有者、開発者のいずれに及ぶのか調査を行った。柴田ら [2] は、主人公が「上司からの命令」と「顧客からの命令」との間で板挟みになるモラルジレンマ課題を提案し、そのような状況に置かれたロボットへのユーザからの認識を調査した。これらの従来研究はエージェント導入者を言及した研究事例といえるものの、実験刺激および評価が質問紙のみであり、実際のインタラクションにおけるエージェント導入者に対する認識や評価、具体的には責任帰属や好感度等は不明である。

そこで本研究は、著者らが先行研究において提案したエージェント導入者の存在を考慮したインタラクションモデル [3] に基づくインタラクション実験プラットフォームの選定を行い、提案モデルを実験的に検討する。ここで、エージェント導入者を Agent adopter (AA) とし、提案モデルの概要を図 1 に示す [3]。提案モデルでは、従来モデルには含まれない AA の存在により、エンドユーザ (EU) が AA に評価を行っていることが示されている。本研究では、特に EU から AA に対する責任帰属について調査する。自動車事故等のシステムが引き起こしたトラブルの責任の所在に関する話題が度々メディアで取り扱われる現代社会において、エージェントの引き起こすトラブルの責任帰属に関するエンドユーザの主観的意見を調査することは重要である。Kawai ら [4] は擬人化尺度の mind perception を用いて、対話エージェントやコンピュータ (ノート PC) の擬人化性の帰属度合いと責任・原因帰属の関係をインタラクション実験により分析したが、AA に対する責任帰属は明らかになっていない。従来研究 [4][5] の知見に基づいて本研究で検証を行う仮説を示す。

- 1) . 対話エージェントに対する擬人化性の帰属が高いほど対話エージェントに対する責任帰属が大きくなる

*連絡先: 電気通信大学
〒182-8585 東京都調布市調布ヶ丘 1-5-1
E-mail: w2010741@edu.cc.ucc.ac.jp

- 2) . 対話エージェントに対する擬人化性の帰属が高いほど AA に対する責任帰属は小さくなる

仮説検証のため、実験プラットフォームの選定と実験設計を行う。

2 提案モデル概要 [3]

提案モデルの概要を図 1, 2 に示す。AA をインタラクションの系に含めたモデルは 2 種類あり、AA と EU が別人物である場合 (図 1)、AA と EU が同一人物である場合 (図 2) である。図 1 のモデルに該当するインタラクション例は商品推薦、公共施設の案内などである。従来モデルとの違いとして、AA, EU が双方向に評価を行うこと、エージェントが AA から評価されることが挙げられる。提案モデルによって、AA, EU からエージェントに対する評価の差異や EU からエージェント、AA に対する評価の関係性などに関する研究課題が新たに考えられる。

図 2 のモデルに該当するインタラクション例は運転支援、面接練習などである。従来モデルとの違いとして、EU が AA としてもインタラクションを行う点が挙げられる。提案モデルによって、EU が AA としてもインタラクションを行っていることが分かり、導入によって発生する金額面でのコストパフォーマンスがエージェントへの印象に与える影響や、導入に対して積極的か消極的かによるインタラクションの違いなどに関する研究課題が新たに考えられる。

本研究では、EU からのエージェント、AA に対する責任帰属について調査する。すなわち、提案モデル (1) によって考えられる研究課題である“EU からエージェント、AA に対する評価の関係性”に関する調査を実験で検証する。

3 実験

3.1 実験設計

図 1 で示したモデルに基づくインタラクション実験プラットフォームを選定する。本実験プラットフォームに必要な要素として、AA がインタラクションの系に含まれること、多様な協調インタラクションが行える実験タスク、登場人物の役割を柔軟に設定可能な実験シナリオが挙げられる。“Escape room”をこれらの要素を満たすプラットフォームとして選定する。Escape room は架空のシナリオの元で制限時間内に謎解きを行うゲームであり、シナリオを変えることで多様なインタラクションが研究可能である [6]。参加者が対話エージェントと協力して謎解き課題に取り組む協力型 Escape room

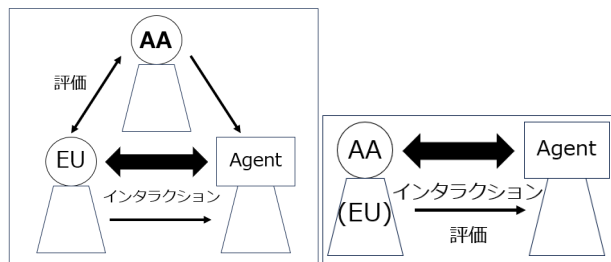


図 1: 提案モデル (1) : AA と EU が別人物の場合 [3]

図 2: 提案モデル (2): AA と EU が同一人物の場合 [3]

を題材として、仮説を検証するための実験設計を行う。図 3 に実験環境を示す。実験参加者は対話エージェントと謎解きタスクを行うシナリオに登場する人物という設定で、シナリオ内の対話エージェントの制作者を名乗る人物 (以下、AA) や対話エージェントの振る舞い等を質問紙で評価する。対話エージェントとしてヴイストン社の Sota (図 4) を、AA としてゲームエンジンである Unreal Engine 5 で MetaHuman¹を用いた CG キャラクタ (図 5) を採用する。MetaHuman 作成アプリケーションである MetaHuman Creator²内に存在するプリセットの Cooper を実験で使用する。本実験で実験参加者に教示する AA は CG キャラクタのため、現実に対話エージェントの制作・導入を行ったわけではない。実験参加者はあくまで架空のシナリオとして、AA が対話エージェントの導入を行ったことを説明され、その上で、対話エージェントと謎解き課題を行い、AA や対話エージェントを評価する。発話の制御は Wizard of Oz 法を用いて実験者が行う。

対話エージェントがリスクな発話 (冗談や褒め) を交えて対話することで、エージェントに対する機械と話している感覚が軽減されるなどの効果が得られることが報告されている [7]。この知見を基に、発話様式の違いが擬人化性の帰属度合いに影響を及ぼすと予想し、「終始丁寧な発話を行う条件 (条件 A)」と「条件 A の発話に加えてリスクな発話も行う条件 (条件 B)」の 2 水準の被験者間計画で実施する。表 1 は対話エージェントの発話例を示しており、太字で示された箇所が条件 B で加わるリスクな発話である。また、責任帰属に関する調査を行うため、謎解きタスクはエージェントの引き起こすトラブルによって必ず時間制限内に完了しないように設計する。謎解き課題はノート PC 上に表示され、参加者は対話エージェントの発話するヒントを基に謎解きを行う。使用した謎の一部を図 6 に示す。なお、本実験の実験参加者は大学生・大学院生 22 名である。本実験は電気通信大学人を対象とする研

¹<https://www.unrealengine.com/ja/metahuman>

²<https://metahuman.unrealengine.com/>

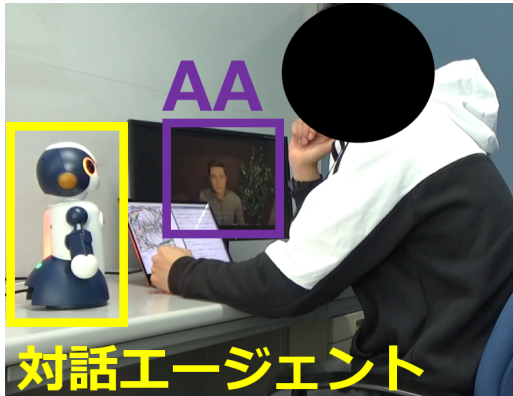


図 3: 実験環境

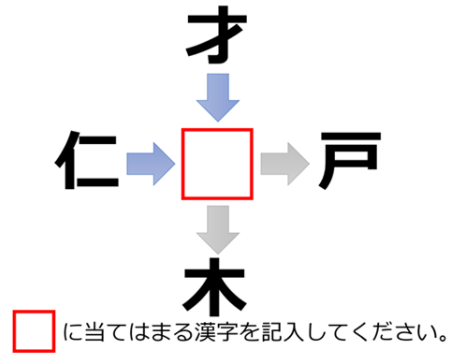


図 6: 使用した謎の一部



図 4: 対話エージェント：
ヴィンストン社の Sota



図 5: Agent adopter：Un-
real Engine 5 で作成した
3D キャラクター

究に関する倫理委員会の承認を得て実施された（管理番号：H23071）。

3.2 実験手順

実験参加者は実験内容に関するブリーフィングを受け、シナリオのシチュエーションを理解する。その後参加者は実験室に入室し、自身のことを対話エージェントの開発と導入を行った人間だと名乗る AA が登場するビデオを視聴する。ビデオ視聴後、対話エージェントが自己紹介を行う。ここで質問紙による事前評価を行う。評価後、実験参加者は対話エージェントと協力して謎解きタスクを行う。謎解きタスク終了後、参加者は再度質問紙で事後評価を行う。

3.3 質問紙

事前評価では 7 件法のリッカート尺度（1. 全くできない...4. どちらともいえない...7. 非常によくできる）による日本語版擬人化尺度 [8] を参考に作成した対話エージェントの擬人化性の帰属に関する 18 項目、同様に 7

表 1: 対話エージェントの発話例

条件 A（丁寧）	条件 B（丁寧＋リスク）
あなたは謎解きが得意そうな知的な雰囲気があります、きっと任務成功すると思います。	あなたは謎解きが得意そうな知的な雰囲気があります、きっと任務成功すると思います。まあお世辞ですけど。
番号を解析するのでちょっとだけ待って下さい。	番号を解析するのでちょっとだけ待って下さい。もし暇なら歌っててもいいですよ。

件法のリッカート尺度（1. 全くそう思わない...4. どちらでもない...7. 非常にそう思う）による Liking 尺度 [9] を参考に作成した AA に対する印象に関する 13 項目、また、5 段階の SD 法による Godspeed 尺度 [10] の好感度の項目を用いた対話エージェントに対する印象に関する 5 項目で評価する。事後評価では事前評価の内容に加えて、対話エージェント、AA、実験参加者それぞれに対する責任帰属度合いを 7 段階（1. 全くそう思わない...4. どちらともいえない...7. 非常にそう思う）で評価する項目に回答する。使用した責任帰属に関する質問文を表 2 に示す。“謎解き課題”と“AA”のことを便宜上“任務”、“M.O.”と呼称して実験参加者に教示したため、質問文内でこれらの言葉を使用している。

4 結果と考察

各条件における対話エージェントに対する擬人化性の帰属度合いを図 7 に、対話エージェント、AA、実験参加者に対する責任帰属度合いを図 8 に示す。条件間の対話エージェントに対する擬人化性の帰属度合いについて、マン・ホイットニーの U 検定を実施したところ、 $p = 6.43 \times 10^{-3}$ となり、有意差 ($p < 0.01$) が認められた。それに対し、責任帰属度合いについては条件間で有意差は認められなかった。

そこで、実験参加者全体における擬人化性の帰属度合いとエージェント、AA に対する責任帰属度合いの関

質問文
あなたが任務に失敗したのは Sota の責任だと思いますか？
あなたが任務に失敗したのは M.O. の責任だと思いますか？
あなたが任務に失敗したのはあなたの責任だと思いますか？

表 2: 責任帰属に関する質問文

係を分析した。図 9, 図 10 は横軸に事後評価における擬人化性の帰属に関する 18 項目の平均値, 縦軸にエージェント, AA に対する責任帰属度合いをそれぞれ示している。スピアマンの順位相関係数 (r) を算出し無相関検定を実施したところ, 擬人化性の帰属度合いと対話エージェントに対する責任帰属については相関がみられなかったが, AA に対する責任帰属との相関係数は 5%水準で有意であった。エージェントに対する擬人化性の帰属が高くなると AA に対する責任帰属は小さくなるという結果より, 仮説 (2) が支持された。

また, 図 11, 図 12 に事前および事後評価における対話エージェント, AA に対する印象の平均値をとった棒グラフを示した。AA に対する印象評価の前後差について, ウィルコクソンの符号付き順位検定を実施したところ, $p = 1.57 \times 10^{-5}$ となり, 有意差 ($p < 0.01$) が認められた。同様に対話エージェントに対する印象評価の前後差についてウィルコクソンの符号付き順位検定を実施したところ, $p = 7.31 \times 10^{-2}$ となり, 5%水準で有意差はみられなかった。AA に対する印象評価は Liking 尺度を参考に作成したため, 点数が高いほど対象を好意的に見ていると解釈することができる。このことからエージェントが引き起こしたトラブルにより, エージェントの設計および導入を行った AA に対する印象が悪くなる可能性が示唆された。

これらの結果からエージェントに対する擬人化性の帰属が, AA に対する責任帰属に有意に影響することが明らかとなり, 提案した実験プラットフォームによって, EU が AA を評価するインタラクションを分析可能となった。

5 おわりに

本研究ではエージェント導入者を系に含む HAI モデルに基づいて, 対話エージェントに対する擬人化性の帰属度合いとエージェントおよびエージェント導入者に対する責任帰属度合いをインタラクション実験により検証した。擬人化性の帰属が高くなるとエージェント導入者に対する責任帰属が小さくなるという結果が得られ, 提案した実験プラットフォームによって, エー

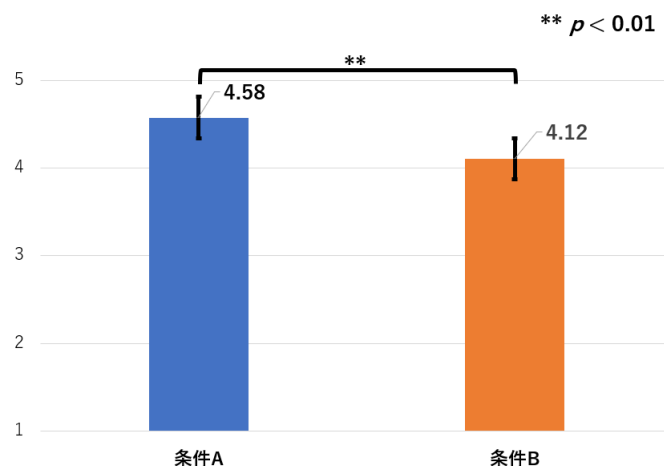


図 7: 対話エージェントに対する擬人化性の帰属度合い

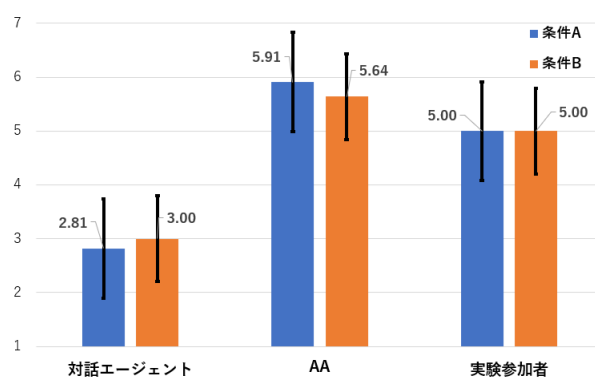


図 8: 対話エージェント, AA, 実験参加者に対する責任帰属度合い

ジェント導入者の存在を考慮に入れた分析を行うことが可能なることを示した。今後の課題として, タスクが失敗せず成功した際の賞賛の帰属に関する調査等が挙げられる。

謝辞

本研究は JSPS 科研費 (JP23K16923) の助成を受けたものです。記して感謝します。

参考文献

- [1] Komatsu, T. Owners of Artificial Intelligence Systems are More Easily Blamed Compared with System Designers in Moral Dilemma Tasks. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 169-170, 2017.

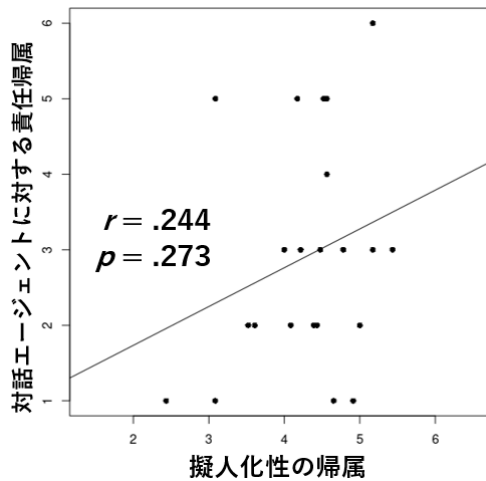


図 9: 対話エージェントに対する擬人化性の帰属と対話エージェントに対する責任帰属の関係

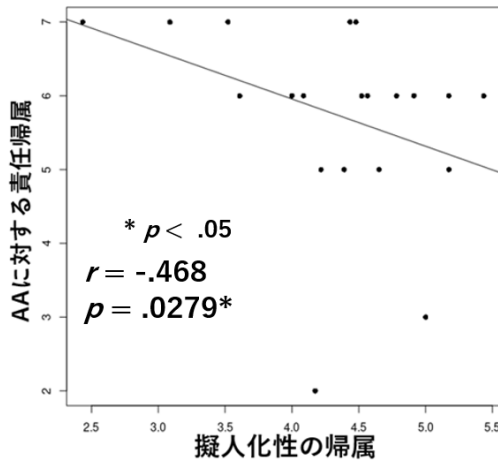


図 10: 対話エージェントに対する擬人化性の帰属と AA に対する責任帰属の関係

- [2] 柴田夏蓮, 小松孝徳. ロボットへの本音を把握するための板挟み型日常的モラルジレンマ課題の提案, HAI シンポジウム 2021, 2021
- [3] 渡邊英一郎, 宮本友樹, 内海彰. Agent adopter の存在を考慮したインタラクションモデルの提案, 第 35 回人間共生システム研究会, pp. 10-11, 2023.
- [4] Kawai, Y., Miyake, T., Park, J., Shimaya, J., Takahashi, H., Asada M. Anthropomorphism-based causal and responsibility attributions to robots. *Scientific Reports* 13, 2023.
- [5] Gray, H. M., Gray, K., Wegner, D. M. Dimensions of mind perception. *Science*, Vol. 315, p. 619, 2007.

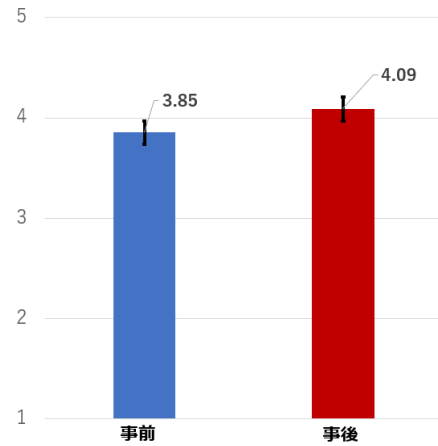


図 11: 対話エージェントに対する印象の前後差

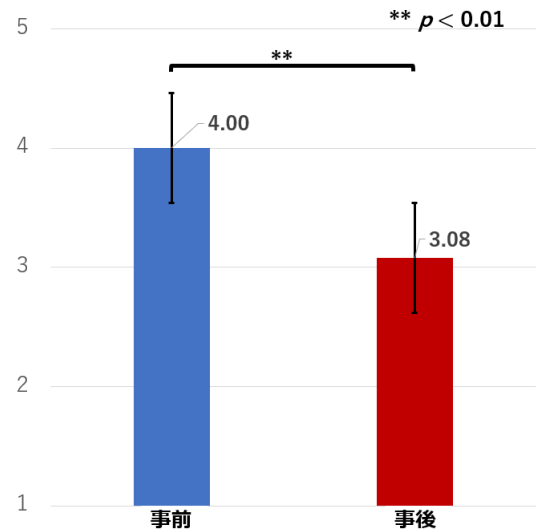


図 12: AA に対する印象の前後差

- [6] Gao, Y., Sibirtseva, E., Castellano, G. and Kragic, D. Fast Adaptation with MetaReinforcement Learning for Trust Modelling in Human-Robot Interaction. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 305-312, 2019.
- [7] 宮本友樹, 片上大輔, 重光由加, 宇佐美まゆみ, 田中貴紘, 金森等. ポライトネスストラテジーに基づく会話エージェントの言語的な振る舞いの違いが人との関係性構築にもたらす効果～初対面における冗談の心理効果～, 知能と情報, Vol.30, No.5, pp753-765, 2018.
- [8] 上出寛子, 高嶋和毅, 新井健生. 日本語版擬人化尺度の作成, パーソナリティ研究, Vol. 25, No. 3, pp. 218-225, 2017.

- [9] 藤原武弘, 黒川正流, 秋月左都士. 日本版 Love-Liking 尺度の検討, 広島大学総合科学部紀要, Vol. 7, pp. 265-273, 1983.
- [10] Bartneck, C., Kulić, D., Croft, E., Zoghbi, S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*. Vol 1, pp. 71-81, 2009.