

Agent adopter・エージェント・エンドユーザの 三項関係における信頼モデルの検討

Investigation of Trust Model in The Ternary Relationship between Agent Adopter, Agent, and End User

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: HAI の研究分野においては基本的にエンドユーザとエージェント間のインタラクションが研究対象となるが、エージェントが社会実装される際、そのエージェントを導入した人物・組織 (Agent adopter: AA) が必ず存在し、著者らは AA をインタラクションの系に含めた新たな研究の枠組みが必要であると考えている。著者らはこれまで、AA の存在を考慮した HAI モデルを提案し、対話エージェントがトラブルを起こした際に AA に対して責任が帰属されることや対話エージェントに対する擬人化性の帰属度合いが AA への責任帰属に影響を及ぼすことをインタラクション実験により示した。本研究では、この枠組みを発展させ、EU のエージェントおよび AA に対する信頼の関係性に着目した研究課題を探索する。先行研究では、EU のエージェント・AA に対する信頼がもう一方に対する信頼に影響を及ぼす可能性が示唆されているが、その影響がどのようなプロセスで生じるのかは未解明である。提案した枠組みにおける信頼に関する研究課題を明らかにし今後の研究の方向性を示す。

1 はじめに

近年、対話エージェントの社会実装が進んでいる。例えばコミュニケーションロボットには、Panasonic 社の NICOBO や Jibo 社の JIBO があり、バーチャルエージェントシステムには、接客や案内に活用されるティファナ。ドットコム社の AI さくらさんがある。エージェントが社会実装され、エンドユーザ (EU) とエージェントがインタラクションを行う際には、必ずそのエージェントの設計・導入を行った人物・組織 (Agent adopter: AA) が存在する。例えば、街頭でティッシュ配りをする擬人化エージェント [1] が存在するならば、広告・宣伝を目的にエージェントを設置した企業や、そのエージェントを開発した企業が存在する。商品を推薦するエージェント [2] やカスタマーサポートのために開発されたエージェント [3] 等も同様に導入者が存在する。

これまでの HAI 研究では AA が系に含まれるインタラクションは殆ど扱われていない。しかし、EU のエージェントに対する評価が AA に対する評価に影響を及ぼす知見が先行研究から得られている。Ayad ら [4] は、エージェントの行動が人の死を引き起こすシナリオに

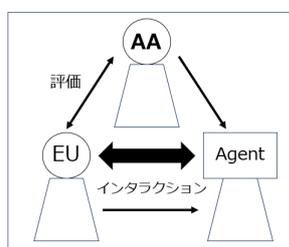


図 1: Agent adopter・エージェント・エンドユーザの三項関係 [6]

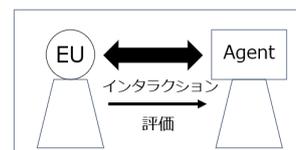


図 2: 従来モデル

において、エージェントの道徳的価値観が低いことで、観察者はエージェントだけでなくエージェントのプログラマーに対する責任も大きくなることを示した。さらに、社会的つながりのあるエージェントは、エージェントの意図とプログラマーの意図が切り離されにくい存在であると捉えられることを示した。また著者らは、EU のエージェントに対する擬人化性の帰属度合いと AA に対する責任帰属の間に負の相関があることをインタラクション実験により示した [5]。これらの研究は、エージェントに対する評価 (道徳的価値観や擬人化性の帰属度合い) が AA に対する責任帰属度合いに影響

*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: w2430145@gl.cc.ucc.ac.jp

を及ぼすことを示しており、EUとエージェント間だけでなく、AA・エージェント・EU間のインタラクションを理解することの重要性を示唆している。

著者らはAgent adopter・エージェント・エンドユーザの三項関係モデル(図1)を提案し、提案モデルによって新たに考えられる研究課題として「EUのエージェント、AAに対する評価の関係性に関する研究」を挙げた[6]。これは、エージェントに対する評価がAAに対する評価に影響を及ぼす関係、逆にAAに対する評価がエージェントに対する評価に影響を及ぼす関係を捉えるというものである。前述した先行研究はいずれもエージェントに対する評価がAAに対する評価に影響を及ぼす関係を捉える研究という位置づけとなる。

本インタラクションモデルに基づいた研究において、EUのAAに対する評価がエージェントに対する評価に影響を及ぼす関係は検証されていない。こうした関係が見られる可能性のあるインタラクションとして、例えばEUが、自身がインタラクションを行うエージェントを開発・導入した人間のことを、開発・導入のプロと評価している場合と素人と評価している場合で、Agentに対して異なる評価が下されることが予想される。開発・導入のプロ(素人)であるとみなされていた場合、信頼度が高く(低く)評価されることでエージェントに対するインタラクション前の期待感が高く(低く)なり、実際にインタラクションを行うことでその期待感との差に基づくエージェントに対する評価がなされると考えられるからである。

本研究では、著者らが挙げた「EUのエージェント、AAに対する評価の関係性に関する研究」のさらなる深掘りとして、EUのエージェント・AAに対する信頼が相互に影響を及ぼすインタラクションに関する検討を行う。検討を行う評価の関係性の構成要素を明示するために、提案モデルの定式化を行う。そして、EUのエージェント・AAに対する信頼に関する従来研究の知見から、三項関係における信頼モデルの作成とリサーチクエストの設定を行う。

2 AAの存在を考慮したインタラクションモデル

2.1 モデルの概要[6]

提案したモデルの概要について述べる。提案モデルに該当するインタラクション例として、商品推薦[2]、公共施設の案内[7]など、企業や組織がエージェントを導入した際のインタラクションが挙げられる。ここで、作成したモデルと従来モデル(図2)の差異と、作成したモデルにより新たに考えられる研究課題について述べる。従来モデルとの違いとして、EUがAAに対して

評価を行うこと、エージェントがAAからも評価されることが挙げられる。提案モデルによって、AA、EUからエージェントへの評価の差異やEUからエージェント、AAへの評価の関係性などの研究課題が新たに考えられる。

2.2 モデルの定式化

本研究ではEUからのエージェントおよびAAに対する評価の関係性について議論を行うため、これらについて定式化を行う。Agentの特徴を $F_i(i = \text{見た目, 声, 発話様式, 動作...})$ 、AAの特性を $C_j(j = \text{性別, 年齢, 職業, ...})$ とする。エージェントに対する評価指標 $X_k(k = \text{好感度, 責任帰属, 信頼...})$ について、EUからAgentに対する評価関数を $R_{X_k}^{EU \rightarrow Agent}$ とする。さらに、AAに対する評価指標 $Y_l(l = \text{好感度, 責任帰属, 信頼...})$ について、EUからAAに対する評価関数を $R_{Y_l}^{EU \rightarrow AA}$ とする。

ここでこれらの評価関数に関わる独立変数を定義する。Agentに対する評価は、Agentの特徴が関わっている。また、Agentに対する評価にはAAへの評価が含まれると考えられる。前述した通り、EUが、自身がインタラクションを行うエージェントを開発・導入した人間のことを、開発・導入のプロと評価している場合と素人と評価している場合で、Agentに対して異なる評価が下される可能性がある。したがって、ある評価指標 X_k について、評価関数 $R_{X_k}^{EU \rightarrow Agent}$ はAgentの特徴と、AAへの評価を独立変数として、式(1)が成り立つ。

$$R_{X_k}^{EU \rightarrow Agent}(F_{\text{見た目}}, \dots, F_i, \dots, F_{\text{動作}}, R_{Y_l}^{EU \rightarrow AA}, \dots, R_{Y_l}^{EU \rightarrow AA}, \dots, R_{Y_l}^{EU \rightarrow AA}) \quad (1)$$

同様にEUからAAに対する評価が行われる際に、ある評価指標 Y_l について評価関数 $R_{Y_l}^{EU \rightarrow AA}$ はAAの特性と、Agentへの評価を独立変数とする[5]ため、式(2)が成り立つ。

$$R_{Y_l}^{EU \rightarrow AA}(R_{X_k}^{EU \rightarrow Agent}, \dots, R_{X_k}^{EU \rightarrow Agent}, \dots, R_{X_k}^{EU \rightarrow Agent}, C_{\text{性別}}, \dots, C_j, \dots, C_{\text{職業}}) \quad (2)$$

3 AA・エージェント・EUの三項関係における信頼モデルの検討

3.1 エージェントに対する信頼がAAに対する信頼に与える影響

本研究の目的はEUのエージェント・AAに対する信頼の関係を検討することである。図3にAA・エージェント・EUの三項関係における信頼モデルを示す。また、図4で示すように、EUのAA・エージェントに対する信頼は相互に影響を及ぼす。

一般に信頼の構築は動的なプロセスであり [8], EU とエージェントが継続的にインタラクションを行うことで信頼がより強固なものになる可能性や信頼が不信へと変化する可能性がある. Olson ら [9] は人間のエージェント (またはスマートオブジェクト) に対する信頼・不信が長期的なインタラクションの中でどのように発展していくのか, 信頼・不信を構築する要因がどのような役割を果たしていくのかを探るための質的研究を行った. その結果, EU が長期的にエージェントに対して信頼を構築し続けた際に, 信頼をエージェントだけでなくそのブランドやメーカーに対して拡大し, さらに同じブランドによって開発・導入されたエージェントに対する初期の信頼が高まることを示した. つまり, EU のエージェントに対する信頼 $R_{X_{信頼}}^{EU \rightarrow Agent}$ が AA に対する信頼 $R_{Y_{信頼}}^{EU \rightarrow AA}$ に帰属される可能性が示唆された. しかし, 先行研究は自由記述のアンケートによる質的な調査であり, 定量的にこのインタラクションを捉えたわけではない. そこで, リサーチクエスチョン 1 が設定される.

- RQ1: あるエージェントに対する信頼は, どの程度 AA に対する信頼に影響を与えるのか?

また, 信頼には認知的要素・感情的要素に基づく 2 種類が存在する [13]. 先行研究ではこの枠組みを用いた分析が行われたものの, エージェントに対する信頼がブランドに対して影響を及ぼすプロセスについて, それが認知的信頼に基づくものか感情的信頼に基づくものを明確にはしていない. そこでリサーチクエスチョン 2 が設定される.

- RQ2: エージェントへの信頼が AA に対する信頼に影響を与える際, そのプロセスは認知的信頼と感情的信頼のどちらに基づくのか?

RQ1, RQ2 は, エージェントへの信頼が AA への信頼に影響を及ぼす関係に関するものであり, AA に対する評価関数 $R_{Y_{信頼}}^{EU \rightarrow AA}$ の独立変数としてエージェントへの評価 $R_{X_{信頼}}^{EU \rightarrow Agent}$ が含まれることを示す式 (2) と対応する. EU がエージェントを信頼することで, その信頼が AA に対する信頼に影響を与えるメカニズムを検討することが本研究の主要な課題の一つである.

3.2 AA に対する信頼がエージェントに対する信頼に与える影響

Olson ら [9] は, 性能などのエージェントの特徴 F_i だけでなく, AA (ブランド・メーカー) の倫理観や社会的な価値観に基づく信頼 $R_{Y_{信頼}}^{EU \rightarrow AA}$ が, EU のエージェントに対する信頼 $R_{X_{信頼}}^{EU \rightarrow Agent}$ に影響を与える可能性を示した. 例えば, EU が AA の悪い評判を聞き, AA に対す

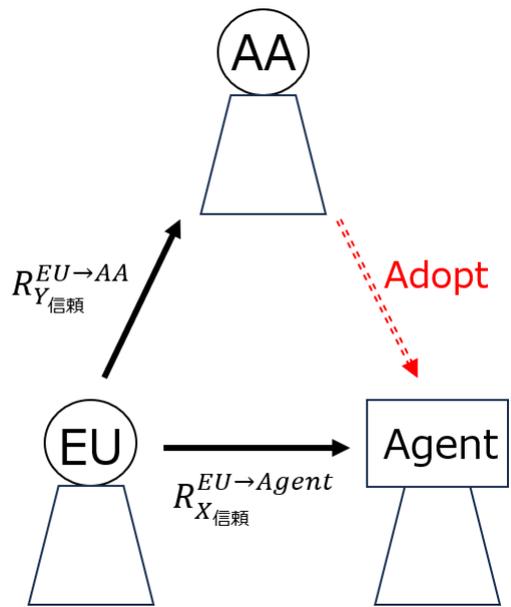


図 3: 三項関係における信頼モデル



図 4: EU の AA・エージェントに対する信頼が相互に影響を及ぼす関係

る信頼が下がった場合はエージェントに対しても不信感を抱く. しかし, その悪評の種類による違い (例えば, 企業の不正行為等の倫理的な問題や, 同社の別プロダクトにおけるバグの多発など技術的な問題等) による影響は調査されていない. 例えば EU がエージェントを性能面での評価に基づいて信頼していた場合, AA の技術的問題に関する悪評は, その EU にとって重要な意味合いを持つ可能性がある. そこでリサーチクエスチョン 3 が設定される.

- RQ3: AA の悪評による信頼の低下は, その悪評の種類によって, エージェントへの信頼に異なる影響を与えるのか?

また, 長期間のインタラクションによって信頼の構築だけでなく, EU とエージェントの間に深い関係が形成されることがある. Laban ら [10] はロボットとの対話インタラクションを 5 週間にわたって 10 回実施した結果, EU のロボットに対する自己開示行動が増加し, EU は対話を行ったロボットをより社会的な存在であると認識した. さらに, skjuve ら [11] は, 自己開示の深さと幅の拡大という観点から関係が構築されると捉

える社会的浸透理論 [12] に基づいて、人間とチャットボット間の関係が長期的なインタラクションからどのように発展するのかを分析した。社会的浸透理論では、自己開示が関係の構築に強く影響を及ぼすとしており、自己開示は関係を築く者同士が信頼していることが前提となるため、関係の構築は強い信頼に関連していると考えられる。EU がエージェントと関係を築くほどに強固な信頼が生まれている場合に、AA の悪評等の外部情報によって AA に対する信頼が下がることで、EU のエージェントに対する信頼にどのような影響が見られるかは検証されていない。例えば、EU がエージェントとの長期的なインタラクションを通じて強い感情的信頼を形成している場合、AA の悪評がエージェントへの信頼に与える影響は限定的である可能性がある。そこで、リサーチクエスト 4 が設定される。

- RQ4: 長期間のインタラクションによって形成されたエージェントへの強固な信頼は、AA の悪評に基づく信頼の低下の影響を受けるのか？

RQ3, RQ4 は、AA への信頼がエージェントへの信頼に影響を及ぼす関係に関するものであり、エージェントに対する評価関数 $R_{X_k}^{EU \rightarrow Agent}$ の独立変数として AA に対する信頼の評価 $R_{Y_l}^{EU \rightarrow AA}$ が含まれることを示す式 (1) と対応する。すなわち、EU が AA に対する信頼をどのように形成し、それがエージェントへの評価にどのように影響するのかを検討することが、もう 1 つの重要な研究課題となる。

4 おわりに

本研究では、著者らが提案した AA の存在を考慮したインタラクションモデルに基づいて、EU のエージェント・AA に対する信頼に関する研究課題の探索を行い、リサーチクエストの設定を行った。また、EU の AA・エージェントに対する評価の定式化を行い、本研究で設定したリサーチクエストの立ち位置を明確にした。これにより、個別に論じられていたエージェントに対する信頼と AA に対する信頼のインタラクションを体系的に整理し、新たな研究の方向性を示した。

今後として、設定したリサーチクエストを検証するためのインタラクション実験の設計を行い、実験を実施する。本研究の成果により、エージェントに対する信頼と AA に対する信頼が相互に影響を与え合うプロセスの詳細が明らかとなり、HAI の設計論やブランド戦略への示唆を提供することが期待される。

参考文献

- [1] Fujishima D, Miyamoto T, Katagami D (2022) Would you be stressed if anthropomorphic agents were handing out pocket tissues around town? Evaluation using a VR study. In *Proc of the 10th Int Conf on Human-Agent Interaction*, pp 284–286.
- [2] Iwamoto T, Baba J, Nishi K, Unokuchi T, Endo D, Nakanishi J, Yoshikawa Y, Ishiguro H (2021) The Effectiveness of self-recommending agents in advancing purchase behavior steps in retail marketing. In *Proc of the 9th Int Conf on Human-Agent Interaction*, pp 209–217.
- [3] Kuramoto I, Baba J, Ogawa K, Yoshikawa Y, Kawabata T, Ishiguro H (2018) Conversational agents to suppress customer anger in text-based customer-support conversations. In *Proc of the 6th Int Conf on Human-Agent Interaction*, pp 114–121.
- [4] Ayad R, Plaks JE (2025) Attributions of intent and moral responsibility to AI agents. *Computers in Human Behavior: Artificial Humans*, 3.
- [5] 渡邊英一郎, 宮本友樹, 内海彰 (2024) 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. HAI シンポジウム 2024.
- [6] 渡邊英一郎, 宮本友樹, 内海彰 (2023) 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.
- [7] Saito M (2023) Effects of presentation modalities in virtual museum guides on agent impressions and painting evaluations. In *Proc of the 11th Int Conf on Human-Agent Interaction*, pp 446–448.
- [8] Hoffman RR, Johnson M, Bradshaw JM, Underbrink A (2013) Trust in automation. *IEEE Intelligent Systems*, 28, pp 84–88.
- [9] Olson DM, Xu Y (2021) Building Trust Over Time in Human-Agent Relationships. In *Proc of the 9th Int Conf on Human-Agent Interaction*, pp193–201.
- [10] Laban G, Kappas A, Morrison V, et al. (2023) Building Long-Term Human–Robot Relationships: Examining Disclosure, Perception and Well-Being Across Time. *Int J Social Robot*, 16, pp1–27.

- [11] Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB (2021) My Chatbot Companion - a Study of Human-Chatbot Relationships. *Int J Human-Computer Studies*, 149.
- [12] Altman I, Taylor D (1973) Social Penetration Theory. Holt, Rinehart and Winston, New York.
- [13] Madsen M, Gregor S (2000) Measuring human-computer trust. In *11th Australasian Conference on Information Systems*, Vol. 53. Citeseer, pp 6—8.