

AI エージェントによって分断された民族間の偏見と不安を軽減

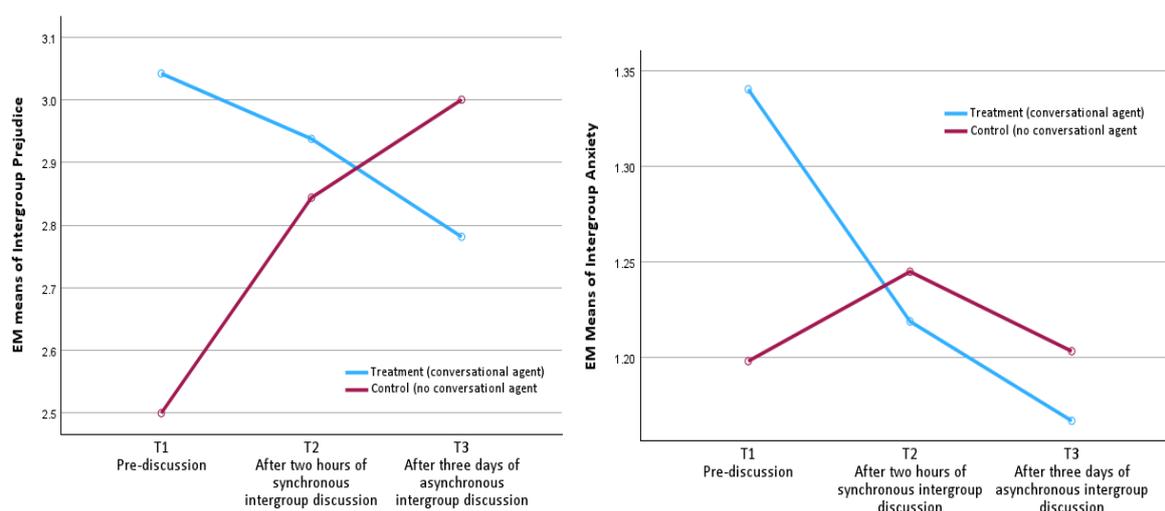
—対立グループの E-Contact における AI エージェントの効果—

概要

本研究では、会話型の AI エージェントを用いることで、対立状況にあるグループの偏見と不安を軽減できることを明らかにしました。実世界での対立状況として、アフガニスタンで歴史的に分断され対立し続けている民族グループを対象としました。さらに、会話型 AI エージェントと、偏見と不安の軽減効果の因果関係を、ランダム化比較実験 (RCT) を用いて明らかにしました。ここでは、オンライン議論支援システム D-Agree を用い、議論をファシリテートする会話型 AI エージェントを用いました。実験では、3つの民族出身の参加者を会話型 AI エージェントがあるグループ (介入グループ) と、会話型 AI エージェントのないグループ (対照グループ) に分け、効果を確認しました。介入グループの参加者は、対照グループの参加者よりも意見をより多く出し、より長い意見を述べ、外部グループへの偏見と不安がより大きく減少しました。本結果は、会話型 AI エージェントに基づく電子的な接触が、歴史的に分断し対立したグループ間でさえ、偏見や不安を改善できることを示しています。

当研究チームは、京都大学大学院 情報学研究科 社会情報学専攻の Sofia SAHAB 特定助教、Jawad HAQBEEN 特定助教、Rafik HADFI 特定准教授、伊藤孝行教授、名古屋大学大学院国際開発研究科国際協力専攻博士課程学生 Richard Eke IMADE (研究当時)、北海道大学 文学研究院 行動科学分野/社会科学実験研究センター 大沼進教授、名古屋工業大学 大学院工学研究科 情報工学専攻の学生 長谷川拓也 (研究当時) から構成されています。

研究結果は、2024年3月21日に国際学術誌「*Communications Psychology (Nature Portfolio)*」でオンライン公開されました。



対照グループと介入グループにおける3つの時間ポイントでの対集団偏見と不安の推定周辺平均値。

1. 背景

紛争が続く世界で、分断された民族やグループの偏見と不安を減らすことは喫緊の課題です。紛争や分断を解決し社会を安定化させるためにはグループ間の相互作用が重要です。しかし、グループ間の相互作用は、利益、価値観、文化の違いから、さらなる紛争を引き起こすことがあります。G. W. Allport のグループ間の接触 (Intergroup contact) 仮説によれば、特定の条件下での Contact (接触) により、偏見と緊張を減少できることが分かっています (Contact は、一般に「接触」と翻訳され、Interpersonal interaction (対人相互作用) のことです)。さらにアフガニスタンのように紛争が続き分断がある社会では、直接的な接触は困難です。このように分断された対立的な状況では、オンラインプラットフォームを使った E-Contact (電子的な接触) による、間接的な接触の効果が期待できます。課題は、分断されたグループ間の E-Contact のプロセスを適切に管理し仲介できる人がいないです。管理と仲介を行う人間のいない、非構造的なグループ間の接触は、偏見や不安を緩和するどころかエスカレートさせる可能性もあります。管理や仲介の能力のある人間がいたとしても、各グループから信頼できる人間を探すのは極めて困難です。本課題を解決するために、我々は会話型 AI エージェント (CA: Conversational AI Agent) を利用して E-Contact を促進する方法を提案しています。CA は、人間らしい会話を行う AI プログラムであり、E-Contact での相互作用を改善するために、費用対効果の高い、偏りのない仲介戦略を提供できます。我々の CA は、E-Contact の目標を達成するために、参加者の会話を適切にガイドし仲介することで、リスクの軽減を目指しています。また、CA は特定の民族グループに属さず、中立的なので、多様なグループ間での調和の促進が期待できます。さらに Allport の権威支援 (Authority Support) の概念を応用し、公平かつ相互に尊重した会話を実現するための CA の新しい利用法を示します。

2. 研究手法・成果

本研究では、無作為化比較試験 (RCT: Randomized Controlled Trial) デザインを用いて、会話型 AI エージェント (CA) の支援が、グループ間の偏見と不安に及ぼす影響を調査しました。我々は、参加者を「CA 支援あり」と「CA 支援なし」の 2 つの設定のグループにランダムに割り当てました。E-Contact として、オンラインのテキストベースの議論支援プラットフォーム D-Agree 上で、以下の (1) と (2) の 2 つの議論セッションを行いました。(1) まず、2 時間の同期議論を行い、その後、(2) 3 日間の非同期で議論を行いました。参加者は、4 人ずつのグループとして編成され、与えられたテーマについて議論を行いました。そして、オンラインでのアンケート調査を次の (A)、(B)、および (D) の 3 つのタイミングで実施しました。(A) 議論の前、(B) 同期議論の後、および (C) 非同期議論の後。結果は、CA による支援を受けたグループでは、集団間の偏見と不安が有意に低下しました。さらに、CA による議論を支援する介入グループの参加者は、対照グループよりも、議論のコンテンツへの貢献において社会的エンゲージメントのレベルが有意に高くなりました。すなわち、介入グループの参加者は、統計的に有意な差のアイデア数を共有し、長い意見を述べました。

3. 波及効果、今後の予定

本研究の結果は、アフガニスタンの分断されたグループ間の偏見と不安を軽減する点で、他の紛争地域での類似の介入の可能性を示唆しています。CA によるグループ間の E-Contact の促進は、世界的にポジティブな相互作用を促進する可能性があります。また、人間のファシリテータと CA のファシリテータの比較研究も計画しています。ここでは、人間のファシリテータと AI のファシリテータがそれぞれ、どのように E-Contact に貢献するかという知見を得られることが期待されます。さらに長期的には、CA ファシリテーションに基づく人間の相互作用に関して、AI による介入がポジティブな社会変化を促進し、社会の持続可能性を高める可能性

を明らかにしたいと思います。

4. 研究プロジェクトについて

本研究は、JST CREST 基金 (Grant Number: JPMJCR20D1, Japan) および JSPS KAKENHI (Grant Number: 22K17948, Japan) の一部から支援を受けました。

<研究者のコメント>

我々は多様な人間のグループ間の相互作用に理解を深め、ポジティブな相互作用を促進することに深く関心を寄せています。本研究は、特に紛争の影響を受け分断が起きている地域において、偏見や不安の軽減に会話型 AI エージェントが役立つ可能性を示しています。本研究が、社会における平和と理解を促進する可能性に、私たちは強い興味を持っています。現在世界中で起こっている社会の分断など、我々が直面しているのは極めて困難な課題です。本研究の示唆するものは、技術が隔たりを埋め、より包括的な世界を創造する力があることを再確認するものです。今後の研究でこれらの可能性をさらに探求することを楽しみにしています。

<論文タイトルと著者>

タイトル：E-contact Facilitated by Conversational Agents Reduces Interethnic Prejudice and Anxiety in Afghanistan (アフガニスタンにおける会話エージェントによる E コンタクトが異民族間偏見と不安を軽減する)

著者：Sofia SAHAB, Jawad HAQBEEN, Rafik HADFI, Takayuki ITO, Richard Eke IMADE, Susumu OHNUMA, and Takuya HASEGAWA.

掲載誌： *Communications Psychology (Nature Portfolio)*

DOI： <https://doi.org/10.1038/s44271-024-00070-z>