2023 年度年次報告書
信頼される AI システムを支える基盤技術
2020 年度採択研究代表者

伊藤 孝行

京都大学 大学院情報学研究科
教授

ハイパーデモクラシー:ソーシャルマルチエージェントに基づく大規模合意形成プラットフォームの実現

主たる共同研究者:
大沼 進（北海道大学 大学院文学研究院 教授）
白松 俊（名古屋工業大学 大学院工学研究科 教授）
松尾 徳朗（東京都立産業技術大学院大学 産業技術研究科 教授）

研究成果の概要

本研究では、ソフトウェアエージェントと人間が一緒に参加するソーシャルネットワークでの民主主義（HyperDemocracy:ハイパー民主主義）のための合意形成プラットフォームを実現する。具体的には、民主主義の基盤としてのソーシャルネットワークプラットフォームの中に、複数のエージェントを常駐させ、人間の代理として働き、意思決定やインタラクションを仲介し、より良い合意形成や集団意思決定を支援する。本プロジェクトでは、現実的なフィールドで社会実装を行うことにより、AIを用いたシステムの社会的な受容性や信頼性の向上を追求する。

ハイパーデモクラシープラットフォームのプロトタイプを1つ実現し、社会実験を3つ以上行うという中間目標は達成されている。4つのグループは、活発に連携を行うことができていて、極めて高い相乗効果が得られている。伊藤グループは、主にすべての項目について注力している。白松グループは主にシステム開発、松尾グループおよび大沼グループは実験評価について連携を行っている。

今後は、特に、大規模言語モデルの急速な発展によって、本プロジェクトの中心の課題：マルチエージェントによる合意形成支援の社会実装という本質的な課題により注力することができるようになった。研究開発項目2では、ハイパーデモクラシープラットフォームを、当初の想定以上の品質で開発ができた。今後はさらに実証実験を行い、より大規模かつ統制した元でのRCTを実施し、科学的な知見を積み重ねる。**アフガニスタンやインドネシア、そして日本の実フィールドに対して、自然な文章で議論が可能なAIエージェントを用いて、インパクトのある合意形成・グループ意思決定支援に関する世界で初の実験を展開できる**。その中でAIエージェントが本当に受け入れられるか（実際には受け入れ始められている）についてELSI的観点からの分析も含まれる。

【代表的な原著論文情報】

1) Sofia Sahab, Jawad Haqbeen, Rafik Hadfi, Takayuki Ito, Richard Imade ,Susumu Ohnuma, and Takuya Hasegawa, "E-contact Facilitated by Conversational Agents Reduces Interethnic Prejudice and Anxiety in Afghanistan", Communications Psychology (Nature Portfolio), Vol.2, No.22, 2024. (https://doi.org/10.1038/s44271-024-00070-z)
2) R. Hadfi, S. Okuhara, J. Haqbeen, S. Sahab, S. Ohnuma, and T. Ito. "Conversational Agents Enhance Women's Contribution in Online Debates." Scientific Reports, 2023.
3) 相馬ゆめ・中澤高師・辰巳智行・大沼進 (2023). 最不遇者情報が集団決定に与える効果：除去土壌福島県外処理問題を題材とした集団討議実験. 心理学研究. DOI: https://doi.org/10.4992/jjpsy.95.22030
4) Shiyao Ding and Takayuki Ito, Self-Agreement: A Framework for Fine-tuning Language Models to Find Agreement among Diverse Opinions, The 20th Pacific International Conference on Artificial Intelligence (PRICAI 2023), November 17-19, 2023, Jakarta, Indonesia.
5) Ayesha Ayub Syed, Ford Lumban Gaol, Alfred Boediman, Tokuro Matsuo, Widodo Budiharto. (October, 2023). A data package for abstractive opinion summarization, title generation, and rating-based sentiment prediction for airline reviews. Data in Brief, 50.