2024 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
2021 年度採択研究代表者

山田 誠二

国立情報学研究所 コンテンツ科学研究系
教授

納得感のある人間-AI 協調意思決定を目指す信頼インタラクションデザインの基盤構築と社会浸透

主たる共同研究者:
小野 哲雄（北海道大学 大学院情報科学研究院 教授）
熊﨑 博一（長崎大学 生命医科学域 教授）
寺田 和憲（岐阜大学 工学部 教授）
原 武史（岐阜大学 工学部 教授）

研究成果の概要

2024 年度は，信頼較正 AI 実装のための要素技術の確立を研究目的として，認知・AI 性能モデルの開発，過不信の予測モデル構築を中心とした，下記の研究を遂行した．

1. **信頼ダイナミクスに影響する要因分析と過不信の予測モデル構築**：信頼ダイナミクスに影響する様々な要因を解明する研究を行った．また，ダイナミック構造方程式モデリング DSEM をベースに過不信を予測する方法論の開発と参加者実験による評価を行った．
2. **認知・AI 性能モデルの開発**：感情・協力・説明・進化・身体性の 5 観点から，AI に対する信頼の認知メカニズムを精緻にモデル化し，協力率・説明適合性・最小許容確率などの社会的性能指標の設計と定量評価手法を拡充した．また，統制実験によってモデルの妥当性を検証し，医療・教育など高リスク領域での応用指針を提示した．
3. **較正キューの機構の解明と実装**：信頼較正を促すナッジ＆ブーストエージェントの実装と評価に加え，拡張現実技術を用いた MR ナッジによる較正キューを実現し，その有効性を参加者実験により確認した．さらに，社会的ロボットの眼のセキュリティ効果についても検証実験により確認した．
4. **人間－AI 協調読影における信頼較正の実証と専門医レベルの能力を持つ視覚－言語モデルの活用**：他のグループと連携し，肺がん検診専用の人間－AI 協調読影ワークステーションを試作し，その実証実験を行った．この試作方法に基づき，熊崎 G におけるタスク AI の実装方法について，寺田 G と連携して検討を開始した．さらに，説明の言語化を試みるため，追加予算で導入した GPU サーバーを用いて胸部画像，病理画像を対象とした視覚-言語モデルを実装し，実際の所見文との比較を行った．加えて，追加予算で読影実験設備を増設し，医師を対象とした観察者実験を企画・実施した．
5. **医療タスク AI の開発と医療施設でのデータ収集**：片足立ちベースタスク AI，三角描画タスク AI を開発した．また，プロソディでスクリーニングを行う医療タスク AI も開発中である．さらに，現在は発達障害の協調スクリーニングにおける信頼較正実験の準備を進めている．なお，これらの研究はタウンミーティングにより，実験倫理的に問題なく実施された．

【代表的な原著論文情報】

1) Akihiro Maehigashi, Takahiro Tsumura, Seiji Yamada (2024). Impacts of Robot Beep Timings on Trust Dynamics in Human-Robot Interaction. International Journal of Social Robotics. 18 pages. doi:10.1007/s12369-024-01181-7
2) Hisashi Takagi, Yang Li, Masashi Komori and Kazunori Terada (2024). Measuring Algorithm Aversion and Betrayal Aversion to Humans and AI using Trust Games, Proceedings of the 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN2024), 357-364. doi: 10.1109/RO-MAN60168.2024.10731215
3) Yuki Abe, Hikaru Tsujiguchi, Daisuke Sakamoto, Tetsuo Ono (2024). Temaneki: Map-Based Collaboration Tool for Consensus-Building in Student-Run Festival Management Teams, In Proceedings of the 2024 ACM CHI Conference on Human Factors in Computing Systems (CHI 2024), 311, 1-8. https://doi.org/10.1145/3613905.3651013

4) Shintaro Sukegawa, Futa Tanaka, Keisuke Nakano, Takeshi Hara, Takanaga Ochiai, Katsumitsu Shimada, Yuta Inoue, Yoshihiro Taki, Fumi Nakai, Yasuhiro Nakai, Takanori Ishihama, Ryo Miyazaki, Satoshi Murakami, Hitoshi Nagatsuka & Minoru Miyake (2024). Training high-performance deep learning classifier for diagnosis in oral cytology using diverse annotations, Scientific Reports, 14, 1, 17591-17591. DOI:10.1038/s41598-024-67879-w
5) Yoshimasa Ohmoto, Kazunori Terada, Hitomi Shimizu, Hiroko Kawahara, Ryoichiro Iwanaga, Hirokazu Kumazaki (2024). Machine Learning's Effectiveness in Evaluating Movement in One-Legged Standing Test for predicting high autistic trait. Frontiers in Psychiatry. 15:1464285. https://doi.org/10.3389/fpsyt.2024.1464285