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
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

2023 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・過不信の予測モデル構築と認知的 XAI デザイン**：信頼に影響する要因を実験的に解明する研究に加え，Dynamic SEM をベースに過不信を直接予測する方法論および SEM や Transformer ベースのリライアンスモデルを使って選択的に説明提示できる認知的 XAI デザインの枠組みを開発した．
2. **認知・AI 性能モデルの開発**：AI に対する信頼に影響を与えるアルゴリズム嫌悪と裏切り嫌悪を定量化定量化し，過信バイアスを発見した．また，深層画像生成神経回路内の特徴空間上に個人の認知性能モデルを構築可能であることを示した．
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に加え，将来的な拡張現実技術の利用を考慮し，MR ナッジによる較正キューを実現した．さらに AI に対するアルゴリズム嫌悪の原因を実験により明らかにした．
4. **人間-AI 協調読影における信頼較正の有効性の実験的検証**：セマンティックセグメンテーションと画像分類を技術とするタスク AI の実装の結果，190 症例中 79 例に信頼較正の効果が期待できることを実験的に検証した．この実証のために，医療用モニターを利用した専用システムを構築し，実証実験を開始した．
5. **人間-AI 協調スクリーニングのためのタスク AI の開発**：北水会記念病院，佐々町健康相談センターでタスク AI の訓練データ収集を終了した．Prosody ベースの医療タスク AI，三角描画タスク AI を開発し，論文投稿した．また，片足立ちでスクリーニングを行う医療タスク AI も開発中である．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．

追加予算を用いて読影者の謝金支払いが可能となり，国内の4つの医療施設で経験ある医師の観察者実験を企画・実施した（現在 13 名を実施済み，最終的には 30 名の実施予定）．なお，この実験は 2024 年度にも継続して実施する．

【代表的な原著論文情報】

1) Takanori Komatsu, Chenxi Xie, Seiji Yamada (2024). Waiting Time Perceptions for Faster Count-downs/ups Are More Sensitive Than Slower Ones: Experimental Investigation and Its Application, In Proceedings of the ACM CHI conference on Human Factors in Computing System (CHI 2024), 624, 1-13. https://doi.org/10.1145/3613904.3641942

2) Yosuke Fukuchi, Seiji Yamada (2023). Selective Presentation of AI Object Detection Results While Maintaining Human Reliance In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2023), 3527-3532. DOI:10.1109/IROS55552.2023.10341684,

3) Motoaki Sato, Kazunori Terada, Jonathan Gratch (2023). Teaching Reverse Appraisal to Improve Negotiation Skills, IEEE Transactions on Affective Computing, 1-14, DOI:10.1109/taffc.2023.3285931

4) Maino Shinohara, Daisuke Sakamoto, James Everett Young, Tetsuo Ono (2023). Understanding Privacy-friendly Design of Robot Eyes, In Proceedings of the Eleventh International Conference on Human-Agent Interaction (HAI 2023), 133-141. DOI:https://doi.org/10.1145/3623809.3623829

5) Shintaro Sukegawa, Sawako Ono, Futa Tanaka, Yuta Inoue, Takeshi Hara, Kazumasa Yoshii, Keisuke Nakano, Kiyofumi Takabatake, Hotaka Kawai, Shimada Katsumitsu, Fumi Nakai, Yasuhiro Nakai, Ryo Miyazaki, Satoshi Murakami, Hitoshi Nagatsuka, Minoru Miyake (2023). Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists, Scientific Reports, 13, 1, 11676-11676. DOI:10.1038/s41598-023-38343-y