

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

2021 年度 年次報告書

山田 誠二

情報・システム研究機構 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

§ 1. 研究成果の概要

今年度は半年間の初年度であり、本研究の中心テーマである信頼インタラクションデザインの研究を進めるための一番基本的な準備研究を中心に研究を遂行した。大まかには、本チームのコア技術である適応的信頼較正を実現するための要素技術・要素システムの開発の基礎的な研究を実施した。また一方、信頼較正 AI の応用領域である、胸部 X 線画像の人間-AI 協調読影と 5 歳児健診のタスク AI 開発とデータ取得環境の整備を進めた。これらの研究テーマについて、具体的に遂行した研究内容を以下に示す。

適応的信頼較正理論を具現化するためには、過信／不信を自動判定し、較正キューを適切なタイミングで表出できる「信頼較正 AI」と実際に人間に代わって領域タスクを処理する「タスク AI」を開発する必要がある。今年度は、信頼較正 AI 開発の基礎研究として、人間がどのように AI に対して信頼を構成するのかという信頼メカニズムを仮説・実験的検証のアプローチで解明することを目的とした研究を行った。合わせて、一般的な信頼モデルについてサーベイを行い、本チームがターゲットとする認知信頼モデルの構成要件を明らかにした。また、較正キューを人間に表出する方法論であるメタサジェクションを開発する信頼インタラクションデザインのための基礎研究を行った。このテーマに関しては、複数ロボットによる感情表出、較正キューのためのナッジ技術の開発を行った。

以上の信頼較正 AI 開発に平行して、その応用分野である人間-AI 協調肺ガン検診と 5 歳児健診のタスク AI を開発するための準備研究を遂行した。具体的には、読影タスク AI の開発に必要な機能要件の確認、開発のための基礎調査を実施した。また、5 歳児健診に必要な 3 つのタスク AI (問診 AI, 視線分析 AI, 行動分析 AI) のうち、機械学習をベースにした行動分析 AI のプロトタイピングを行い、さらに実際に現場の医療施設でデータを取得するための IT 環境の整備を行った。

§ 2. 研究実施体制

(1) 山田グループ

- ① 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)
- ② 研究項目
 - ・適応的信頼較正理論のコア確立
 - ・信頼概念の期待効用への拡張
 - ・原 G・熊崎 G と連携した医療健診のタスク AI・信頼較正 AI の開発

(2) 寺田グループ

- ① 主たる共同研究者: 寺田 和憲 (岐阜大学工学部 准教授)
- ② 研究項目
 - ・認知信頼モデル構築
 - ・学会連携による倫理的課題の検討

(3) 小野グループ

- ① 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)
- ② 研究項目
 - ・信頼較正 AI とタスク AI の相互作用モデル
 - ・較正キューの探索的デザイン
 - ・ツイン AI のエージェントデザイン

(4) 原グループ

- ① 主たる共同研究者: 原 武史 (岐阜大学工学部 教授)
- ② 研究項目
 - ・人間-AI 協調肺がん検診システムの実装と大規模実験による性能評価
 - ・協調肺がん検診システムの実稼働

(5) 熊崎グループ (研究機関別)

- ① 主たる共同研究者: 熊崎 博一 (国立精神・神経医療研究センター精神保健研究所 室長)
- ② 研究項目
 - ・5 歳児健診のデータ収集の計画と実施
 - ・協調健診のための信頼較正 AI によるツイン AI 構築と 5 歳児協調健診の実稼働

【代表的な原著論文情報】

- 1) Choi, Myunguen, Sakamoto, Daisuke, and Ono, Tetsuo (2022). Kuiper Belt: Utilizing the “Out-of-natural Angle” Region in the Eye-gaze Interaction for Virtual Reality. Conference on Human Factors in Computing Systems (CHI '22). Article 357, 1-17.

<https://doi.org/10.1145/3491102.3517725>.

- 2) Jérôme, Dinet, Hanna, Verdel, Hirokazu, Kumazaki, Kazunori, Terada, Sylvie, Vitel, Eloïse, Zehnder (2021). A Hybrid Robotic System for Children with Autism: A longitudinal Study of Social Interaction in Classroom. ERGO'IA 2021.
- 3) Goto, Kento, Mizumaru, Kazuki, Sakamoto, Daisuke, Ono, Tetsuo (2022). ADioS: Angel and Devil on the Shoulder for Encouraging Human Decision Making, ACM/IEEE International Conference on Human-Robot Interaction (HRI'22), Video, 1192-1193.
- 4) Sano, Shin, Yamada, Seiji (2022). AI-Assisted Design Concept Exploration Through Character Space Construction. *Frontiers in Psychology*, 12. doi:10.3389/fpsyg.2021.819237
- 5) Sukegawa, Shintaro, Matsuyama, Tamamo, Tanaka, Futa, Hara, Takeshi, Yoshii, Kazumasa, Yamashita, Katsusuke, Nakano, Keisuke, Takabatake, Kiyofumi, Kawai, Hotaka, Nagatsuka, Hitoshi, Furuki, Yoshihiko (2022). Evaluation of multi-task learning in deep learning-based positioning classification of mandibular third molars, *Scientific Reports*, 12, (in press).
<https://doi.org/10.1038/s41598-021-04603-y>