信頼される AI システムを支える基盤技術　　2022 年度
2021 年度採択研究代表者　　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
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）
熊崎 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 准教授）
原 武史（岐阜大学 工学部 教授）

研究成果の概要

2022 年度は，信頼モデルの精緻化を中心に，認知・AI 性能モデル構築，較正キューデザイン，そしてタスク AI の開発とデータ収集を行った．

1. **信頼モデルの精緻化と信頼できる XAI**：AI のアピアランス，ビープ音，説明の解釈可能性，タスクの難易度，タスク構造などの要因が信頼に与える影響を参加者実験と統計的分析，SEM（構造方程式モデリング）により解明した．また，Dynamic SEM をベースに過不信を直接予測する方法論を開発した．さらに，AI への依存度指標を Transformer で予測し，信頼を維持しつつ XAI の説明提示を制御できる信頼できる XAI を開発し，CAPTCHA と自動運転 UI で評価した．
2. **認知性能モデルと AI 性能モデルの開発**：心理物理とガウス過程回帰を基に人間の性能モデルである認知性能モデルの構築方法を開発した．一方，同一の学習アルゴリズムを用いて両モデルを構築できること，当初のタスク成功確率の代わりにより柔軟な confidence を単位として導入することなどを議論し，研究を計画した．
3. **較正キューのインタラクションデザイン**：拡張現実 AR ベースのナッジ・ブースト技術を利用して，較正キューのプロトタイピングを行った．また，AI に対する拒否反応である algorithm aversion について，特に医師や一般患者の algorithm aversion を抑制するためのインタラクションデザインの方法論を参加者実験により構築した．
4. **人間−AI 協調診断のためのタスク AI の開発**：胸部 X 線画像と発達障害のスクリーニングを行うタスク AI 開発とデータ取得を開始し，部分的に完了した．また，人間-AI 協調読影のプロトタイピングを行い，参加者実験による実験計画を立てた．発達障害スクリーニングでは，三角描画等のデータ収集を一部完了し，時系列分析によってスクリーニングを実現した．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ている．

追加予算を用いて購入した高性能 PC を用いて，リライアンス予測モデルの Trasnformer の学習および強化学習を効率的におこなうことができた．

【代表的な原著論文情報】

1) Tsumura, Takahiro, Yamada, Seiji. (2023). Influence of anthropomorphic agent on human empathy through games. IEEE Access, pp.40412-40429, 10.1109/ACCESS.2023.3269301
2) Tsumura, Takahiro, Yamada, Seiji (2023). Influence of agent's self-disclosure on human empathy. PLOS ONE, 18(5): e0283955. https://doi.org/10.1371/journal.pone.0283955
3) Sukegawa, Shintaro, Tanaka, Futa, Hara, Takeshi, Yoshii, Kazumasa, Yamashita, Katsusuke, Nakano, Keisuke, Takabatake, Kiyofumi, Kawai, Hotaka, Nagatsuka, Hitoshi, Furuki, Yoshihiko. (2022). Deep learning model for analyzing the relationship between mandibular third molar and inferior alveolar nerve in panoramic radiography. Scientific Reports, 12(1), 16925-16925, doi.org/10.1038/s41598-022-21408-9
4) Sukegawa, Shintaro, Tanaka, Futa, Nakano, Keisuke, Hara, Takeshi, Yoshii, Kazumasa, Yamashita, Katsusuke, Furuki, Yoshihiko. (2022). Effective deep learning for oral exfoliative

cytology classification. Scientific Reports, 12(1), 13281. doi:10.1038/s41598-022-17602-4

5) Sukegawa, Shintaro, Yoshii, Kazumasa, Hara, Takeshi, Tanaka, Futa, Yamashita, Katsusuke, Kagaya, Tutaro, Nakano, Keisuke, Takabatake, Kiyofumi, Kawai, Hotaka, Nagatsuka, Hiroshi, Furuki, Yoshihiko. (2022). Is attention branch network effective in classifying dental implants from panoramic radiograph images by deep learning? PLOS ONE, 17(7), 1–15. doi:10.1371/journal.pone.0269016