信頼される AI の基盤技術
2020 年度採択研究者

2020 年度
年次報告書

西田 知史

情報通信研究機構脳情報通信融合研究センター
主任研究員

脳情報に基づいた AI の信頼性評価技術の開発

§1. 研究成果の概要

本研究計画では、信頼される AI の基盤技術開発に資することを目的として、認知神経科学の方法論と知見を活用した 2 つのアプローチで AI の信頼性に関する研究を進めている。1 つ目のアプローチは、人間が AI を信頼するとは何かという根本的な問いに対する認識論的説明を得ることを目的とした、AI への信頼性を生み出す脳内基盤を理解するための脳計測実験に基づく研究である。2 つ目のアプローチは、信頼される AI に不可欠な要因として人間らしい判断を行う AI を実現することを目的とした、脳情報のモデルを AI へ取り入れる技術の開発と検証である。1 つ目のアプローチに対する本年度の研究では、AI への信頼性が知覚にもたらす影響を評価する認知課題を設計し、心理実験においてその妥当性を確認するとともに、認知課題遂行中の脳活動を機能的磁気共鳴画像法（fMRI）によって計測するための実験の設計とテストを行った。2 つ目のアプローチに対する本年度の研究では、当研究者が過去に開発した技術をベースにして、脳情報モデルを取り入れた AI を実装したうえで、その AI が下す判断と脳活動から読み取った人間の判断の関係性について予備的な分析を行った。その分析の結果、脳情報を取り入れることで、AI の判断が人間の判断に近づくことが確認できた。以上の成果は、人間と AI の関係性において、人間の認知プロセスの理解と AI の人間性向上という、双方向的な探究の可能性を見出すものであり、今後得られる成果も見込んで、信頼される AI の研究開発に多大な貢献をもたらすことが期待できる。