2023 年度年次報告書
文理融合による人と社会の変革基盤技術の共創
2023 年度採択研究代表者

山田 寛章

東京工業大学 情報理工学院
助教

立場と規範を反映した言語モデルによる法議論シミュレーション

研究成果の概要

本研究では自然言語処理を用いたアプローチによって、日本の民事領域の法的判断形成過程をシミュレーションするシステムの開発を目標としている。

研究初期の実施内容として、法議論シミュレーションにおけるエージェント向けのモデル開発の基盤およびベースラインの構築が必要となる。本年度（2023 年 10 月～2024 年 3 月）においては、①以後の実験に用いる大規模言語モデル Fine-tuning 用のデータ作成の事前準備として、少数のサンプルデータ作成、②各エージェントモデル単体性能のベースラインを確立するため、既存の大規模言語モデルの備える推論・議論の能力が法分野においてどの程度に達しているかの評価を試みた。具体的には、司法試験の記述式問題の解答の生成を行い、ロースクール（法科大学院）の教員との連携を行いながらその出力の評価を実施した。このうち、②については規模を拡大して第二年度でも継続して実施し、既存の大規模言語モデルが法議論シミュレーションにおけるエージェントモデルとして振る舞うに十分なドメイン知識と性能を有するかを検証する。

また、当初の計画にはなかった成果として、法ドメインにおけるデータセットの構築と共有に関する法的・倫理的課題の整理を行った。具体的には、判決書を用いて構築されたデータセットについて、そのデータ収集源に由来するバイアス・判決書という特殊なデータに由来する機密やプライバシーへの配慮・データのステークホルダーの権利や利益とデータセットの利用可能性のバランスについて整理した。