

立教大学学術推進特別重点資金（立教 S F R）
大学院学生研究
2024年度研究成果報告書

研究科名	立教大学大学院	人工知能科学研究科	人工知能科学専攻
研究代表者 (2025年3月現在 のものを記入)	在籍課程・学年	氏名	
	<input type="checkbox"/> 博士前期課程 年 <input checked="" type="checkbox"/> 博士後期課程 2年	浦東 聡介	
指導教員	所属部局・職名	氏名	
	人工知能科学研究科 教授	村上 祐子	
自然・人文 ・社会の別	自然	個人・共同の別	個人
研究課題	法的意思決定プロセスの透明性と信頼性向上に向けた研究		
研究組織 (研究代表者 ・共同研究者) ※2025年3月現 在のものを記入	在籍研究科・専攻・課程・学年	氏名	
	人工知能科学研究科・人工知能化学 専攻・博士課程後期課程 2年	浦東 聡介	
研究期間	2024 年度		
研究経費 (1円単位)	(支出金額) 249,864円 / (採択金額) 250,000円		

研究の概要 (200~300字で記入、図・グラフ等は使用しないこと。)

当該研究の研究目的を含むこと。

本研究は、AIを司法分野に導入する可能性を探究し、その透明性・信頼性の向上を目指すものである。AIの裁判補助や自律的な判決の可能性を考察し、司法制度の効率化と公平性向上に寄与することを目的とした。本年度の研究では、特にAIによる法的意思決定の社会受容に着目した。そのために、哲学カフェでの議論分析とYouTubeコメント分析を用い、社会のAI裁判官に対する期待と懸念を明らかにした。哲学カフェでは、AI裁判官の公平性や一貫性が高く評価される一方、被告の心情理解の欠如や責任の所在の不明確さが懸念された。特に、参加者が自身を当事者と想定すると受容度が低下し、AI裁判官の導入には社会的合意が不可欠であることが示唆された。また、技術的課題として学習データのバイアスが、倫理的課題として裁判官としての責任と正当性の確保が指摘された。YouTubeコメント分析では、裁判官に求められる公平性と共感性のバランスが求められることが明らかとなった。公平な判断が重視される一方で、被告への共感が視聴者の評価に大きく影響することが示された。SNS分析が一般市民の意識を捉える有効な手段であることも確認された。本研究は、AI裁判官の社会的受容度の検証と技術的・倫理的課題の特定に貢献した。

キーワード (研究内容をよく表しているものを3項目以内で記入。)

[AI裁判官] [法とAI] [法的意思決定]

研究成果の概要 (図・グラフ等は使用しないこと。)

以下の視点を含めて記載のこと。

- ・当該研究は何をどこまで明らかにできたのか(できなかったのか)。
- ・何をもちて研究成果(経過)を達成できた(できなかった)と考えられるのか。
自身が設定した研究目的・目標に照らして、その根拠がわかるよう記載のこと。

本研究は、意思決定科学、AI によって法的な意思決定プロセスの透明性と信頼性を向上させ、社会における公正さ、平等さ、そして信頼性の向上に貢献することを目的とする。このことを通じて、法律家及び一般市民の法令及び判例理解を補助し、司法制度の更なる民主化の実現に貢献することを目指すものである。また、AI を利用した裁判補助の可能性を探求し、判決の一貫性及び正確性、司法制度の効率性向上へ貢献することを目指している。

司法分野における AI 利用は、2000 年代以降本格的に実務に利用され始めた。米国で早期保釈決定に利用される COMPAS[1]、判例検索に利用できる ROSS Intelligence[2]、世界初のロボット弁護士と言われる DoNotPay[3]などは実用化されている代表例である。このような AI の司法利用の期待は 2023 年の GPT-4[4]の公開によって加速した。GPT-4 は初期の AGI の兆候を示唆するなど[4]、従来の予測モデルなどの便利ツールの域を大きく超える汎用的な能力が期待される。司法分野においても GPT-4 は米国の司法試験模試やロースクール試験において合格レベルのスコアを記録し[5、6]、英国司法省では利用のためのガイダンスが作成された[7]。

一方で、司法分野における AI の利用についてはブラックボックス性などの懸念も指摘されている[8]。このような AI の技術上の問題に加え、AI を司法、とりわけ裁判官の補助、あるいは裁判官として利用する場合には AI が司法機関や一般市民に受け入れられるかどうか重要となる[9]。そして、AI の受容度を高めるためには、司法や裁判官が社会においてどのような役割を期待されているのかを検討し、その期待に沿うように AI を適合させる必要がある。

今年度の研究では、AI を司法分野で活用するため、司法・裁判官が社会においてどのような期待を受けているのかを明らかにした。そのための手法として、哲学カフェの利用とソーシャルメディア上のコメント分析の手法を採用した。哲学カフェでは深度ある意見を把握するため議論を促し、そこから得られた意見を観察した。また、ソーシャルメディア上のコメント分析では、裁判に関する YouTube の動画に対して投稿されたコメントについて、機械学習手法を活用して分析することで、大規模なデータを定量的に評価した。それぞれの成果について以下に述べる。

1 哲学カフェの議論分析

本研究は、AI を司法分野に導入する可能性を哲学カフェを通じて検討し、特に AI 裁判官の社会的受容度と技術的・倫理的課題について議論することを目的とした。司法における AI 活用がもたらす公平性や効率性向上の可能性を明らかにするとともに、AI と人間裁判官の役割分担についての洞察を得ることを目指した。この研究で明らかになったことについては以下のとおり。

(1) 研究で明らかになったこと

イ AI 裁判官の可能性

哲学カフェの議論では、AI 裁判官が提供する公平性や一貫性が高く評価された。特に、全ての裁判を統一した基準で裁くことで個人差を排除できる点や、感情に左右されない客観性への期待が示された。

ロ 社会的受容の限界

具体的事例を設定し、自らが当事者になった場合を想定した議論では、AI 裁判官の受容度が低下した。特に、被告の心情や背景を理解する能力や、責任の所在が不明確であることへの懸念が顕著に示された。具体的事例を設定することで、哲学カフェ参加者の AI 裁判官に対する受容度がそうでない場合と比較して低下することは、一般市民の意識調査における深度ある議論の必要性を示唆している。

ハ 課題の特定

技術面では、学習データのバイアスが指摘された。また、社会的課題としては AI が裁判官としての責任をどのように果たすべきか、倫理的な正当性をどう確保するかが重要であることが示された。

(2) 研究目的との比較と成果の評価

本研究は、当初の目的である「AI 裁判官の社会的受容度の検証」と「技術的・倫理的課題の特定」に大きく貢献した。一方で、AI 裁判官の具体的な導入条件や、社会的合意形成の方法論については、今後さらなる研究が必要である。

研究成果の概要 (つづき)

2 YouTube コメント分析

本研究は、哲学カフェでは得られない広範な意見の収集と分析を目的とした。YouTube 動画へ投稿されたコメントを通じて裁判官に対する社会的期待を分析し、特に AI 裁判官の社会受容に役立つ洞察を得ることを目的とした。具体的には、YouTube 上の裁判シーンを描いた動画のコメントをトピックモデリング手法である Latent Dirichlet Allocation (LDA) を用いて分析し、裁判官に求められる公平性や共感性といった人間的な要素を明らかにした。これにより、AI が司法分野において果たし得る役割を検討するための基盤を構築した。

(1) 研究で明らかになったこと

イ 公平性と人間性の期待

- ・コメント分析の結果、公平性と共感・人間性が裁判官に求められる重要な要素であることが示唆された。特に、裁判官が被告人に対して示す共感や思いやりが、視聴者に感動を与えた一方で、個人的関係が公平性を損なう可能性が指摘された。
- ・公平性と共感とは本質的に矛盾しないものの、それぞれが異なる状況で強調される傾向が見られた。

ロ SNS 分析の有効性

ソーシャルメディアコメントの分析は、一般市民の意見を直接観察する手段として有効であり、裁判官への期待を把握するための手段として役立つことが示唆された。

3 研究目的との比較と成果の評価

本研究は、「AI 裁判官の社会受容に役立つ要素を広範な意見の収集と分析によって明らかにする」という目的に対して大きな成果を挙げた。具体的には、YouTube という巨大ソーシャルメディアから情報を収集し、AI の社会受容検討に向けて公平性と共感性という二つの重要な要素を特定した。一方で、今年度の研究では分析対象が一つの動画に限定されていたため、他文化圏や多様な裁判状況における一般市民の期待分析するためには、更なる情報の収集と分析が必要となる。

[1] Reiling, A. Dory. "Courts and artificial intelligence." IJCA. Vol. 11. 2020.

[2] Semmler, Sean, and Zeeve Rose. "Artificial Intelligence: Application today and implications tomorrow." Duke L. & Tech. Rev. 16 (2017): 85.

[3] Julie Fishbach, Coder, 19, Builds Chatbot That Fights Parking Tickets, NBC NEWS (Jul. 21, 2016),

<https://www.nbcnews.com/feature/college-game-plan/coder-19-builds-chatbot-fights-parking-tickets-n612326> [<https://perma.cc/UEN7-DBAM>].

[4] Bubeck, Sébastien, et al. "Sparks of artificial general intelligence: Early experiments with gpt-4." arXiv preprint arXiv:2303.12712 (2023).

[5] OpenAI. 2023., GPT-4 technical report.

<https://cdn.openai.com/papers/gpt-4.pdf>

[6] Choi, Jonathan H. and Hickman, Kristin E. and Monahan, Amy and Schwarcz, Daniel B., ChatGPT Goes to Law School (January 23, 2023). Journal of Legal Education (Forthcoming), Available at SSRN: <https://ssrn.com/abstract=4335905>

[7] Courts and Tribunals Judiciary

<https://www.judiciary.uk/guidance-and-resources/artificial-intelligence-ai-judicial-guidance/>

[8] Deeks, Ashley. "The judicial demand for explainable artificial intelligence." Columbia Law Review 119.7 (2019): 1829-1850.

[9] 弥永真生, 宍戸常寿, 編 "ロボット・AI と法【初版】", 有斐閣(2018), p. 237.

研究発表 (研究によって得られた研究成果を発表した①～④について、該当するものを記入してください。該当するものが多い場合は主要なものを抜粋してください。なお、成果発表を確認できる資料を合わせて研究成果報告書提出フォームより提出してください(紙媒体等、研究成果報告書提出フォームから提出できない場合は、別途リサーチ・イニシアティブセンターへ提出してください)。

- ①雑誌論文 (著者名、論文標題、雑誌名、巻号、発行年、ページ)
- ②図書 (著者名、出版社、書名、発行年、総ページ数)
- ③シンポジウム・公開講演会等の開催 (会名、開催日、開催場所)
- ④その他 (学会発表、研究報告書の印刷等)

※修士論文・博士論文は含みません。

④ 学会発表

- ・ Mapping the ideal judge: Public expectations of judges Analyzed via LDA on social media for AI design
Ninth International Conference on Robot Ethics and Standards 2024年7月29日
- ・ 裁判官への期待: LDA を用いた YouTube コメント分析手法の提案
法と心理学会第25回大会 2024年10月27日
- ・ AI 裁判官の社会的受容に向けた責任の追求と処罰の意義
電子情報通信学会 合意と共創研究会 2025年1月15日