

客観的または感情的な判決理由が AI 裁判官への態度に及ぼす影響

LIU YICHEN

人工知能(AI)技術の急速な発展に伴い、司法分野における AI の活用は、補助的なツールとしての役割から、より中核的な判断過程へと拡大しつつある。実際に、一部の国や地域では、AI は証拠の識別などを担う司法支援システムとしての位置づけを超え、より中核的な判断過程へと関与の範囲を拡大しつつある。しかしながら、AI が判断主体として刑事司法の決定に直接関与することを、一般市民がどの程度受容し得るのかについては、依然として議論が続いている。先行研究の多くは、「AI への信頼」という単一の次元から人々の態度を検討してきたが、判断主体そのものに対する評価と、具体的な判決結果に対する評価とを明確に区別して検討した研究は限られている。本研究はこの点に着目し、刑事司法判断における AI 裁判官に対する一般市民の態度構造を体系的に検討するとともに、「裁判官への信頼」と「判決結果の受容度」という二つの評価次元を区別して分析することを目的とした。

本研究は、研究 1 および研究 2 の二部構成からなり、日本の成人を対象としたオンライン実験を計 4 回実施した。研究 1 では、3 つの実験(研究 1a~1c)を通じて、刑事司法の判決場面を想定し、裁判官の種類(AI 裁判官 対 人間裁判官)および判決理由の種類(事実に基づく理由 対 感情に基づく理由)を操作した。その上で、裁判官に対する信頼および判決受容度における評価の差異を検討するとともに、裁判官の提示方法や注意チェック項目を段階的に調整することにより、主要な結果の再現性および安定性を検証した。

研究 1 では、実験材料として過失運転致死事件を用いた。当該事件は、事実関係の構造が比較的明確であり、量刑判断の相場が確立されており、日本において発生頻度が高いことから、刑事司法文脈における AI 裁判官の受容を検討する上で適切な事例であると判断された。研究 1 の結果から、3 つの実験(N = 2024)すべてにおいて、「裁判官への信頼」を従属変数とした場合、裁判官の属性が一貫して有意な主効果を示し、参加者は総じて AI 裁判官よりも人間裁判官を高く信頼する傾向を示した。これは、アルゴリズム忌避の傾向が安定して観察されたことを意味する。この効果は、判決理由の種類やその提示方法によって左右されることはなく、信頼判断においては、「裁判官が AI であるか否か」という属性情報が優先的な手がかりとして機能していることが示唆された。これに対し、「判決結果の受容度」を従属変数とした場合には、裁判官の属性の影響は相対的に弱く、参加者は主として判決理由の内容そのものに基づいて評価を行う傾向を示した。研究 1a および研究 1b では、感情に基づく判決理由よりも、事実に基づく判決理由の方が高い受容度を示し、刑事司法判断において「事実を重視する」規範的期待が存在することが示された。しかしながら、研究 1c ではこの効果が完全には再現されず、判決受容度は特定の条件下において、より高次の手がかりの影響を受ける可能性が示唆された。この点については、結果の安定性を含め、今後の検討が必要である。

以上の結果を踏まえ、研究 2(N = 906)では、裁判官への信頼と判決受容度との構造的関係をさらに検討するとともに、両者が AI 裁判官の導入に対する態度をどのように予測するかを検証した。研究 2 では、研究 1 とは異なる刑事事件類型として窃盗罪(万引き事件)を用い、異なる刑事司法文脈における結果の再現可能性を検討した。また、研究 2 では、裁判官への信頼、判決受容度、ならびに AI 裁判官の導入に対する態度を、それぞれ複数項目からなる尺度によって測定した。その結果、裁判官への信頼と判決受容度は高い相関を示しつつも、統計的には区別可能な二つの態度次元であり、両者がそれぞれ独立して AI 裁判官の実際の導入に対する態度を予測することが明らかとなった。この知見は、AI 裁判官に対

する一般市民の態度を多次元的構造として捉える必要性を一層裏づけるものである。

総じて、本研究は、刑事司法の判決場面において、AI 裁判官に対する一般市民の態度が、単一の信頼判断に還元されるものではなく、少なくとも判断主体に対する情緒的信頼と、判決結果に対する機能的評価という二つの側面から構成されていることを示した。前者は裁判官の属性情報の影響を受けやすい一方で、後者は判決理由の合理性により強く依存する傾向が認められた。本研究は、同一の実験枠組みにおいて AI 裁判官と人間裁判官を直接比較することにより、司法 AI 文脈における対 AI 信頼と対人信頼の差異を理解するための体系的な実証的知見を提供するとともに、今後、より精緻な司法 AI 態度モデルの構築および社会実装に関する議論に向けた基盤を提示するものである。(社会心理学)