

本年度成果

1. AI agent の法的位置づけ: AIに法的人格を付与する問題の歴史的経緯調査と現地の課題提案
2. 生成AIの社会的影響調査: 人々の受け入れ心理, 悪影響など. 寺田麻佑客員研究員らとの共同研究

AI agent の法的位置づけ

◆ 歴史的経緯

➤ Lawrence B. Solum 1992

- AIに法的人格を与えるには「何かが足りない論」(魂, 意識, 意図, 感情, 善悪の感覚, 自由意思) SolumはAIがこれらを持つように振る舞うなら, それを否定する具体的根拠がない以上「何かが足りない」論は成立しないと主張
- AIは苦痛を感じないので, 罰を与えても被害者の応報感情が満たされないとされるが, Solumは法人自体は苦痛の概念を持たないが, 刑法では法人を処罰できると主張

➤ Chopra & White 2011

- 知識がAIエージェントに帰属し, その知識がさらに依頼者である本人に帰属できる**方法**が必要
- AIが本人のツールなら, AIの知識は自動的に本人の知識. しかし, AIの複雑さから, この前提はあやしい.
- その点に関する議論を避けるためにAIに法的立場を一切与えてはならないという臭いものに蓋ではまずい

➤ 最近の反対派の意見

- ここでは最も強硬な反対派であるMarshall [2023] の意見を記す. (1) AIの影響が未評価, (2) AIの偏見と不公平さが認識されていない, (3) AIに法律を守らせる,あるいはAIの危険性を押さえるというガードレールが未整備. → これらが解決するまで, スピード重視のテック文化に惑わされたりせずに, AIは脇に追いやられるべき.

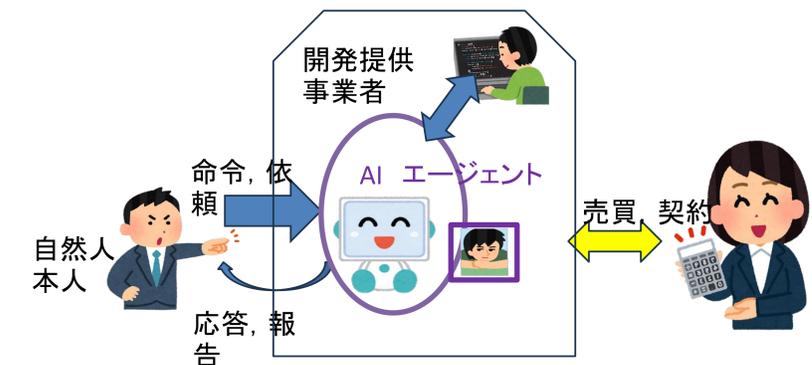
➤ AIエージェントに限定的な法的人格を付与: Ugo Pagallo

- 完全な道徳的行為者の集合, すなわち天使のようなAIロボットの集団においてすらルールが必要 → どのようなルールが必要かをオープンな環境で実験的に調べることが大切.
- 誰かまたは何かに法的人格を与えることは, 常に合理的な選択と経験的証拠や個人の哲学, 趣向に左右される非常にデリケートな政治問題だった.
- 説明責任と賠償責任の可能性. たとえば, 自律走行車の複雑な分散責任に対する新形態の AI 責任を模索すべき.
- 法的実験で新しい説明責任と賠償責任をオープンな環境でテストし, AIの法的責任の範囲をエビデンスに基づいて決める. (たとえば, 2003年以降の日本の特区制度)を拡大し, エビデンスを基礎にして, 困難なケースに対して合理的かつ効率的な新しい法的代理形態を探す.

◆ 代理法に基づくAI agent の法的位置づけ

- 米国の代理法の第3次リステートメントが整理されているので, それを参考にして分析
- Powell [2020] AIの直接的な法的人格付与には, 第3次リステートメントの人格の定義の根本的な変更が必要
- Kolt [2024] 代理権をAIエージェント向けに再定義するには,
 - a. 責任は予見可能な損害に限定されるべき
 - b. AIの行動がかなり予測不可能で, 現在の限定的な可視性と説明責任しか持たないAI エージェントに代理権付与が適切であるかどうかは不明

◆ 中川の分析: 購買, 契約などを本人の代理で行うAIエージェント



- ツールAI: AIエージェントが, 本人の命令に字義通り従う
- 自律的法人AI: AIエージェントは資産を持つ法人であり, 本人の曖昧な命令を意味解釈して, 外部状況に合わせて行動する

	責任
ツールAI	法的には本人の責任 ただし, AI提供者に責任が及ぶ場合あり
自律的法人AI	法的には PAI が全責任 (保有資産の範囲で)

- 代理人の立場はフィデューシャリー(fiduciary)と呼ばれ, 本人との信認関係(fiduciary relation)によって成立. ただし, 法人として財産を持たないと信認関係が成立しない. よってツールAIとの間での信認関係は成立しない.

➤ 技術的問題

- 1) 行う行為が代理権の範囲を逸脱しそうかを認識する能力と本人への確認が必要
- 2) そのためには, 本人が代理権の範囲をできるだけ具体的かつ詳細にAIエージェントに与えるべき. 2)の具体性や詳細さが不十分だと1)の本人への確認が頻繁になる. 技術的には, 1), 2)の適度な按分がAIエージェントの本人に対するインタフェース設計上の課題.