FY2025/2025年度

AI Utilization in Society and Legal System Team

Hiroshi Nakagawa
社会におけるAI利活用と法制度チーム　中川 裕志

AIの法的人格

◆Lawrence B. Solum　1992

◇法的人格　：自然人　人格，

◇「何かが足りない論」~AIに法的人格を与えるには魂，意識，意図，感情，善悪の感覚，自由意思が足りない

◇Solum：AIがこれらを持つように振る舞うなら，それを否定するだけの具体的根拠がない以上「何かが足りない」論は成立しない
→ AIは法的人格を持ちうる

◇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のではないか？

◇Solum：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.

◆Chopra & White 2011

◇AIエージェントは，本人（principal）の法的な代理人の意味ではなく，本（principal）の代理人的な行為をするAIとする.

◇ AIエージェントの法的人格付与の問題を契約問題として検討

◇ AIエージェントの代理性を分析するには，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

◇つまり，AIエージェントが法的な代理になれるかは，経済的および技術的問題および法技術の問題

◆反対派　Marshall 2023

(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.

◆Novelli, C and Floridi, L and Sartor, Gi　2024

◇人格に関する一般的な法理論 (単一主義 vs. クラスター主義)

(1)法的人格に関する単一主義：　エンティティ（人間，動物，AIなど）は、一つの能力を持つだけで，それに対応する権利または義務を持つ法人としての資格があり

(2)法的人格に関するクラスタービュー：　法的人格の意味を、エンティティに割り当てられた多数のステータス（契約を締結，財産を所有，etc.）の集合として解釈

◇**論点**：　AIは単一主義法人格の場合は，単一の能力に関してだけ法人格を持てる．例えば，株の売買.

クラスタービューの場合は人間と同じ，ないしは人間能力のかなり大きな部分集合を持つ法人格になる.

◆　Ugo Pagallo　2018

◇AIエージェントに限定的な法的人格を付与

◇法的実験などの実用的なアプローチ

a. AI　ロボットに完全な法的人格を与えるという仮説は避ける（EU委員会　2018）

b. 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.

c. 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．既存領域におけるいくつかの法制度のポリシー（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.

◆中川裕志

1. AIが法的人格を持つことの可否に関する研究動向．情報ネットワークローレビュー 24巻，研究ノート．pp.80-92. 2025年12月5日
2. 中川裕志：AIの倫理 - 人間との信頼関係を創れるか- 第2部第2章 AIの法的人格．　角川新書．角川書店．pp.121-141．2025年12月

Agentic AIと自然人の間のトラスト

◆　AIのoutputは人間と同レベル以上にトラストされている.

◇**Binns** (2018)ローンの承認，保険の見積もりを人間が説明する場合とAIが説明する場合のトラスト度合いを比較すると，人間とAIの回答が相関が小さいという結果はなく，相関が有意に高いという結果も相当数得られた.

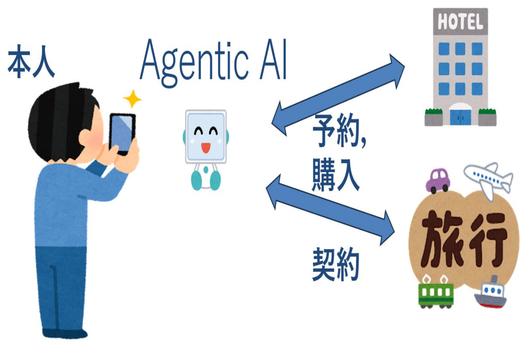
◇**Thurman(2018)** らは26カ国の53,314人のオンラインニュース消費者を対象にして，(1)自分や友人の過去の消費行動から導かれるAIによる記事選択と、(2)編集者やジャーナリストによる記事選択，のどちらがオンラインでニュースを入手する良い方法であるかを調査したところ，(1), (2)は**ほぼ同程度**（社会調査による比較結果）

◇**Vodrahalli**らは，**人間が画像認識のタスクを行うにあたって，自身が行った判断をAIからアドバイスをもらった後に変更するか**　関して1100人を対象に調査した結果，**人間のアドバイスとAIのアドバイスの取り入れ方に有意な差がない**ことが判明

◇米，英，加，日，韓，ブラジル，ポーランドの皮膚科の臨床医38名に，皮膚画像を見せて生検に回すかどうかを判断するタスクを行った結果，このタスク向けのデータで学習した**AIによるアドバイスと他の医師によるアドバイスによる精度の向上に有意な差はなかった**.

◆　**Agentic AI = AI＋実社会での行為.**

行為してもらうかどうかはトラストの強さによる



◆　3種類のAgentic AI

(1)　単なるツール

・利用者本人が完全にコントロールするツール

・利用者本人は事細かに命令

・Agentic AI といえども，その行為は利用者の責任.

(2) 保険付きツール

・ Agentic AIが自律的に行った行為の結果，本人の意図に沿わない不利益，ないし本人にとって損失が起きる，あるいはAgentic AIの行為が相手側に損害を与える

・ Agentic AIの行為に保険をかけ，本人が保険料を払う

保険料と免責額という保険ポリシーは，本人が選べる

✓ただし，本人の意図に沿わない場合，本人の依頼のし方が不十分だったのか，Agentic AIのソフトが，本人の依頼した自然言語の理解能力が不十分だったのか.

これを争うのはなかなか難しい問題

(3) 法人AI

・Agentic AIが資産を持ち，部分的にせよ法人格を持っていたとすると，損失はAgentic AIが手持ちの資産で支払い，本人には損失が及ばないようにできる.

本人への支払いは迅速化できる

・資産額が大きければ，Agentic AIの信用度，具体的には許容範囲が上がる.

・法人AI自身が保険に入ることもある．この場合も保険ポリシーを現実の場面で解釈する問題は残る.

・高い許容度を持つトラストで関係つけられた自然人と法人格を持つAgentic AIが共生するエコシステムが構成できると，本人にとっては時間や行動の自由度が増すという大きなメリット.

◆中川裕志：AI agentと人間の共生社会を基礎づけるトラストに関する考察 ,第25回情報ネットワーク法学会研究大会　個別報告5-5,