

AI エージェントの法的位置づけの言説の経緯

中川裕志¹

概要: AI エージェントはAI の重要な応用分野である。本報告ではAI エージェントの法的位置づけを肯定的に捉えようとする系譜の言説の時間的経緯をまとめてみた。ここでは 1992 年の L. Solum の論文[1]以降の言説を扱い、Chopra&White[3]およびそれを参照する Ugo Pagallo のロボット法[4]の考え方を分析する。この分析の限界を考慮して G. Teubner のアクターネットワーク理論の解釈に基づいて部分的な自律性のある AI エージェントの位置づけを提案する。

キーワード: デジタル遺産, 死後, パーソナル AI エージェント, 信託

Historical View of AI Agent's Legal Status

HIROSHI NAKAGAWA^{†1}

Abstract:

In this report, an AI agent is an important application area of AI. I summarized the time history of genealogical discourse trying to positively grasp the legal position of AI agents. Here, we deal with the discourses since the 1992 L. Solum paper[1]. We also analyze the idea of Chopra & White[3] and Ugo Pagallo's robot laws[4] that refers to it. Considering the limitations of these analyses, we propose the positioning of AI agents with an interpretation of partially autonomous AI agents that G. Teubner proposes based on actor network theory.

Keywords: Digital heritage, postmortem, personal AI agent, trust

1. 問題点の概要

AI エージェントはAI の重要な応用分野である。ただし、AI エージェントが実社会で果たせる役割は、その法的位置づけによって異なる。本報告では、この問題に関して、AI エージェントの法的位置づけを肯定的に捉えようとする系譜の論文をまとめてみた。古くは、1992 年の Solum[1]、最近のものとしては同じく Solum の 2019 年の[2]である。これらの系譜をまとめると、AI エージェントが人間と同じ人格を持つ議論は棚上げにして、何らかの法人格を持たせるのが現実的であるとしている。UgoPagallo[4]は、AI エージェントの法的位置づけだけでは問題はすべてを解決できず、AI の設計方針を活用すべきと提言している。

いずれにせよ、AI エージェントの決定が人間社会に実際の影響を与えるようならざるを得ない点については一致がみられる。問題は、むしろどのような状態で AI エージェントが社会に影響を与えるかである。この問題が本報告の二番目に論点である。Teubner[5]はアクターネットワーク理論を応用した Hybrid という概念を提案している。筆者はその考え方を指針とした提言を行う。

以下、第 2 節は AI エージェントの法的立場に関する Solum の提言、第 3 節は Chopra & White による Solum の提言の発展および、それに対する UgoPagallo の解釈および筆者の追加的解釈、第 4 節はアクターネットワーク理論による AI エージェントを埋め込む社会の構図について説明する。第 5 章では、法的代理人の利用を考察し、残された課

題を述べて、本報告を閉じる。

2. Solum の指摘

本節では、Solum[1]の論点をまとめる。

2.1 基本的見立て

Solum は、法制度は「実践的な判断の形式的な集積地」と位置付けている。その解釈は、法制度は個人の心の内面を直接に観測や評価の対象にするものではなく、外面的に観測できる個人や法人、社会の行為に関する実践的な判断の形式的に整備された体系という見立てになる。この見立てでは、その後の議論に出発点になり、本報告でも通底しているアイデアである。

2.2 財産管理運用する AI エージェント

将来出現すると予想される財産管理運用する AI エージェントを対象にして、Solum[1]が 1992 年にこれを以下の 3 段階に分けた。

Stage 1: AI エージェントのプログラムは、人間の介入を必要とせず、多数の単純な信託財産の管理において人間の管財人を支援する。

Stage 2: AI エージェントは投資家として人間よりも優れているので、人間の管財人は AI エージェントのアドバイスに従う必要がある。

Stage 3: AI エージェントがすべてを自律的に行えるので人間の管財人は不要となる。

証券会社などが人間のトレーダの代わりに採用している AI トレーダは、株式市場などの場面では、高速な売買を行

い、証券会社の人間の担当者はその売買結果を受け入れているので、すでに Stage 2 に達しているとみなせるだろう。

しかし、Stage 3 のように AI エージェントが自律的な管財人として機能する能力を持っている場合、AI は何らかの法的責任を負う必要があるはずだから、法律は AI に法人格を与えるべきと考えられる。

2.3 AI の権利制限派の主張

Solum が描くような AI エージェントが何らかの法人格、さらには人格を持つべきであるという言説には常に反論が提起されている。以下に主要な反論とそれを論破する Solum の論理を紹介する。反論は 2 種類に大別される。一つ目は「AI エージェントには何かが足りない」というものであり、二つ目は「AI は自然人でない」というものである。

2.3.1 何かがないという反論

・志向性がない

AI エージェントはプログラムであり、プログラムは意味を処理する能力である「志向性」を欠いている。このため、プログラムによって達成される正式な記号操作は、思考または理解を構成することはできないという反論がある。

この反論は、志向性を 1) AI エージェントを構成するプログラムの内部状態に還元して定義するのか、2) AI エージェントと外部との観察可能なやりとりで定義するのかで異なる。この問題を考えるとき役立つのが、図 1 に示すサルが提案した中国語の部屋

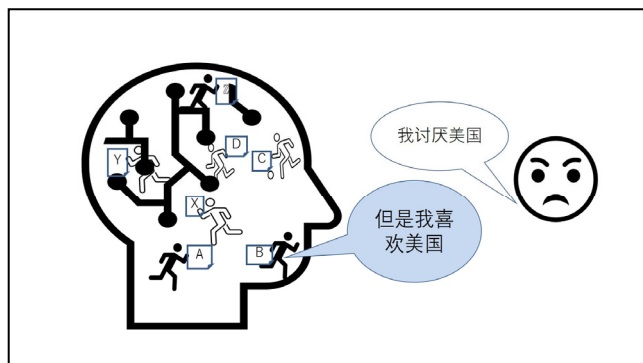


図 1. 中国語の部屋

Figure 1. Chinese language room

左側の人は AI プログラムで、その内部では英語の札をもった小人たちが走り回って仕事をしており、結果として「但是我喜欢美国」という中国語を発話したとする。しかし、小人たちは中国語を理解していないので、この AI プログラムは中国語の意味を理解しておらず、志向性がないと言われる。

しかし、左側の AI エージェントが右側の人の「我讨厌美国」[a]という発話を聞いて、「但是我喜欢美国」[b]という

発話をしたことを外部から観察しているなら、AI エージェントは中国語を意味理解できており、志向性があるとみなせるだろう。つまり、左側の AI エージェントはチューリングテストに合格したことになり、人間と区別できない。

志向性がないというもう一つの論拠は、AI エージェントがプログラムの指示から離れて、AI は意味を自律的に処理する能力がないことである。しかし、現在の AI はプログラムの指示だけで完全に行動が制御されているわけではない。異なる情報環境で異なるデータが入力されれば、強化学習などの仕掛けによって、その後の AI エージェントの行動は変化してくる。つまり、AI エージェントの動きは、プログラムとそれが置かれた情報環境の両者に依存する。これは、自然人の行動が DNA と出生後の情報環境の両方に依存することと類似している。したがって、AI はプログラムだという理由で、AI には「志向性」が欠如しているとは断定できない。

・責任をとれない

AI エージェントは責任をとれない。つまり、AI エージェントが義務に違反した場合、補償あるいは処罰を受けることができないという反論である。

しかし、AI エージェントは法人格を持てば、保険に加入して損害賠償できる。保険では、被委託者による故意の不正行為により課せられる可能性のある金銭的責任については、保険が利用できない。しかし、AI は意図的な不正行為を回避するように設計できるので、この問題は回避できる。

Solum は言及していないが、被害者の応報感情を満たすための刑罰がプログラムである AI には与えられないという観点もある。企業のような法人であっても、企業経営者などの自然人がいるので、応報感情はある程度満たされるだろう。AI 自体が法人になったときに、その法人の運営者や出資者のような自然人がいなければ、現代人の感覚では被害者の応報感情は満たされないという問題は残る。

・判断できない

財産を管理運用する AI エージェントの例に関して言えば、自然人の管財人が AI の能力を超える可能性のある判断を下すことができれば、AI の判断能力を疑いたくなるだろう。

しかし、チューリングテストに合格した AI は外部から観察できる行動に関しては人間と同等であり、この問題はない。現状においては、証券会社などの使う AI トレーダはその高速性により、人間トレーダの投資能力を超えている。例えば、ゴールドマン・サックスでは 200 名いた人間のトレーダは 2 名にまで減ったそうである[6]。AI トレーダに関しては取引システムの適正な管理や取引戦略の届出等の規制が導入されている[7]。ただし、個々の AI トレーダの取引自体に規制が行き届いているかどうか疑問である。だ

a 和訳「私はアメリカが嫌いだ」

b 和訳「それでも、私はアメリカが好きですよ」

からといって、法律は、実態に即して AI が限定された目的の被委託者としての役割を果たすことを許可されることを認めてはいけないというものでもない。むしろ、法制度を実態に合致させる方向でないと経済活動を縮退させてしまう。

2.3.2 自然人ではないという反論

このタイプの反論も具体的には「AI には何かがない」という具体例で提起されるものが多いので、まずそれらから見ていくことにする。

・AI には魂がない

自然人でないから魂がないという主張は、反論者の自然人の感情としてはありえても、法定書面では機能しない。

・AI には感情がない

感情は人間の精神の一面であり、人間の精神が計算モデルによって説明できる場合、感情は計算プロセスになる可能性があり、この反論は成立しない。哲学的には、カントの道徳理論は、人間性には感情が必要であるという仮定に疑問を投げかけている。カントは、人間だけでなくすべての合理的な存在は人であると主張したので、それに従うなら AI エージェントは人格を持つといえる。

・AI には意識がない

自然人の意識の定義自体が生物学的にも確定的でない。自然人は他人の心に直接アクセスできないため、他人の行動からしか意識を推測するしかない。となると、この推測から意識がないと証明できるか？という逆の問い掛けに反論した側は答えることができないだろう。

・AI には自由意思がない

人間の自由意思とは、意識的な推論で正しい方法で引き起こされた場合、行動は自由であるということである。人間の脳内で起こっている意識は直接には定義できず、外部観察から推測するしかない。よって、人間と同じ内部的状态を意思とするなら、反論者にとってすら解答不能な問い掛けである。外部からの観察によって意識を定義できるなら、AI エージェントは自由意思を持つといえる。

・AI は興味を持たない

プログラムに食品とかエンターテインメントに興味を持つようにコーディングしておき、その興味にしたがって行動するように AI エージェントを設計できる。ただし、このような興味は、人間にとっての興味とはまったく異なる。興味は、むしろ自由意思や意識の一つの現れ方であり、それなら自由意思、意識と同じような議論になる。

・AI は人間ではない

憲法的人格権を考えると、この反論は本質的問題を提起しているので、丁寧に考えてみることにする。

自然人が知性を持っている、感情を持っている、意識し

ているなどの理由で特別視するという立場から、AI は人間でないという主張をする場合、問題は、AI だけでなく、クジラやイルカ、サルなど知性が高いとされている動物がこれらの資質を共有するかどうかになるという問題になってしまい、社会の在り方や法的な観点から見ると、無意味な議論ではないかと思われる[c]。

そのような経緯を踏まえて、「人間中心主義」[d]の論者は、以下のような主張をするかもしれない。

AI が道徳的であるために必要な資質を持っているとしても、AI の憲法上的人格権を認めることは人間にとって利益ではない。ゆえに私たちは AI に憲法上的人格権を認めるべきではない。

この主張は、人間は人間以外と異なるという理由だけに依拠しており、人間を絶対視する独善的な主張ですらある。

2.1 節で述べた Solum の基本的見立てを思い起こせば、ここでは、道徳的な人間性よりも法的な人格、すなわち法人格に焦点を当てていることに留意しなければならない。法律は人間の心の内面ではなく、あくまで外部に表れた関係を扱う仕組みである[e]。

・AI はシミュレータでしかない

AI エージェントはプログラムによるシミュレータでしかないという反論も人間至上主義によるものである。人間行動の基礎となるプロセスに関する知識によって、AI エージェントが人間の心理を記述するのに必要な特徴を持っているシミュレータと信じられるなら、AI エージェントが人間の精神の必要な特徴を持っていると信じてよい。とすれば、AI エージェントに法人格を与えることに痛痒はないといえる。

2.3.3 Solum の提起する論点

AI エージェントを日常生活の中で意図をもって動くシステムとして扱うことが有用なら、AI エージェントが真の志向性など上記の反論で列挙された特質をもっていなくても大きな違いはないと Solum は考えている。

この考えにしたがえば、法学で最も手に負えない質問のいくつかは、「人とは何か、そしてなぜ人にそのような強力な法的保護を与えるか」に関連してくる。これは言い換えれば、personhood という法的概念と personality という倫理的ないし情緒的な概念の間の関係または境界線をどこに置くかという問題である。

科学的な観点からすれば、AI が正しい行動をとり、認知科学や生物学、脳科学がこれらの行動を生み出す根本的なプロセスが人間の精神のプロセスと比較的類似していることを確認できた場合、人間至上主義という信念を脇に置けば、AI を人として扱うことすらできる。

[c] もっとも、中世以降、動物自身に対する裁判が行われてきた歴史はあるわけだが、現代社会において意味があるとは認められない。

[d] 人間至上主義と呼ぶべきかもしれない。

[e] 反省の有無などが問題にされるが、それはあくまで外部から観察できる行動に基づいた推論である。

2.3.4 Solum の 27 年後

Solum は 27 年後, [1]を振り返りつつ[2]で AI にさらに大きな法的役割を与える以下の提案をしている。

1. AI には, 法規範を生成する能力がある。
2. AI には, AI が生成した法規範を適用する能力がある。
3. AI には, ディープラーニングを使用して, AI が生成した法規範を変更する能力がある。

上記の 2.と 3.を実現するには, AI には継続的な社会観察とリスク分析能力が必要である。とはいえ, この部分は自然人の法律家が支援することも可能である[f]。

AI の法制度策定への法的権限の委任は物議を醸す。仮に能力的にみて AI に委任が可能になったとしても, 現実に立法を可能にすることには抵抗が予想される。人間の立法者は敗者になるので, 間違いなく彼らの代表はプロセス上の理由で AI による立法に抵抗するだろう。

とはいえ, 法規範は民主的機関によって承認されるべきであるという民主的正統性の考え方にしたがえば, 抵抗は難しい。民主的正統性が, 独立した規制機関の最初の立法承認によって満たされるとするなら, 法的能力を持つ AI に権限を委任するという立法であっても民主的正統性を満たす可能性がある。ようするに決め方の問題なのである。

一方で法規範の透明性という本質の問題点が指摘されている。法規範は, 一般の人々がアクセスできるという意味で透明でなければならない。3.の項目によれば, 法規範自体が深層学習システムに組み込まれているため, AI の作る法制度はその立法事実や立法の経緯, さらに立法の目的がブラックボックスとなる。そのため法規範透明性の懸念は深刻である。ブラックボックスな法律は秘密法と同等であり, 法の支配の本質的な要素であると広く信じられている「人々に周知されるべき」という要件に違反する。2.項目, 3.項目のように AI が法制度を運用するなら, それは, 法規範自体が法の適用対象である人間に公開され, 理解できるように設計されるべきであるが, これは AI に係わる技術として将来の課題である。

3. Chopra & White から Ugo Pagallo へ

3.1 Chopra & White

Chopra & White[3]は Solum のアイデアを 20 年後に引き継ぎ, さらに精密化している。AI エージェントを法人と見なすことは, 概して, 発見ではなく決定の問題と指摘する。AI エージェントに法人格付与を拒否するか肯定するかに関する最良の議論は, 概念的ではなく実用的なものあるべきだという点で Solum と同じである。さらに, AI エージェ

ントの機能と社会的役割を考えると, 現行の法律は, どちらとも決定しないだろうと予測し, だからこそ今後の決定の問題とみなすことになる。

実用という観点からすれば, 目的達成のための行動を最適化するように行動する合理的な AI エージェントは, おそらく, 法的制裁の懲罰的な力に従わないような自己破壊的行動はしないように設計されると明記しており, これも Solum と同じである。

Chopra & White[3]においては, 「法人」の概念が財産の概念と密接に関連することに力点を置いている。財産を持つことによって, AI エージェントが引き起こした事故や失敗における補償が可能になるからであり, これは Solum も述べている。

[3]が論破しようとしている AI エージェントが人格を獲得する可能性に対する異議は, 前節で述べた AI の可能性に対する一般的な議論に似ており, AI の計算アーキテクチャに「何か欠けている」と仮定して, 人格をもてない, とするものである。たとえば, AI エージェントは道徳的感覚がないため罰を受けない, という言説がある。また AI エージェントは自由意思と自律性がないため裁量的な決定を下すことができないとする。つまり, AI エージェントは道徳的感覚, 自由意思, 自律性という資質が欠如しているという反論である。これらの反論は, 人間のもつこれらの属性が人間にのみ与えられたという人間至上主義に等しい。しかし, すでに Solum が反論に再反論したように, これらの反論は説得力のあるものではない。

3.2 Ugo Pagallo

Ugo Pagallo [4]は, まず AI を以下のように 3 分類する。

- 1) 人間が完全にコントロールできる AI ツール,
- 2) 限定的ではあるが自律的に行動する AI,
- 3) 人間と同じ知能レベルを持つ AGI(Artificial General Intelligence)

次にこれらの AI 各々が, その AI 自身が原因で引き起こした事故などに対して, 以下の項目にどのように対処するかを記述している。

- ・ **責任**: 不法行為または過失を行った行為者が引き受ける責任。基本的には民事の責任になるので, 補償の問題になる[g]。刑事の場合は, 被害者の応報感情の扱いが問題だが, これは扱いにくい。
- ・ **免責**: 上記の責任のうち法律あるいは契約によって免責となる場合。これは, 状況に応じて柔軟に対処するしかないだろう。柔軟ということは, 個別の事例の類型によって, 被害と運用・製造側の損害の和を最小化するような免責の閾値を決めることで対処する。当然, そのときの技術レベルも勘案した閾値の設定となる。

[f] この論点は, 後の節で Teubner のアクターネットワーク理論の応用で再度議論する。

[g] 刑事の場合は, 被害者の応報感情の扱いが問題だが, これは扱いにくい。

- ・**厳格責任**：英米法の概念で、行為者に故意や過失という心理的要素のない場合にも犯罪の成立（ないし不法行為責任）を認めること。例えば、社員の個人判断で起こした不法行為に対して、法人である会社が自動的に責任を負うこと。
- ・**製造物責任[h]**：AI の設計者、製造者に欠陥のある製品を作った場合の責任だが、教師データや使っているうちに学習する強化学習などが入っていると、製造物責任に全てを押し付けることはできない。利用者が AI を誤った使い方、たとえばソフトの更新をしなかった、勝手に改造をしたなどがあれば、製造物責任にはならない。
- ・**不当な損害**：国、契約相手、不法行為法の第 3 者など他者によって誘発されたロボットや AI 自身への損害。損害を与えた第 3 者に帰責できるかどうか論点になる。

以下に 1),2)3)のタイプについて考えてみる。

1)のツールとしての AI は法的にはあくまで道具であり、現行法が適用される。

3)の AGI はその定義から自然人と同等なので、自然人に対するものと同じ法律を適用せざるをえないだろう。

2)の限定的な自律 AI が本報告において焦点をあてるべき分析対象なので、詳しくみていく。前提として、限定的な自律 AI は固有の財産を持ち、法人格が与えられているとする。

責任を取る可能性があるのは AI の運用者と AI 自体である。責任を取るのは、AI の運用者と AI の間での契約によって決まる。ただし、AI は財産を持つので、AI が補償責任を果たし、AI 運用者には帰責されないということも可能である。

免責については、既に述べたが、閾値の設定には AI の運用者、製造者、および法制度の所管者が相談、交渉のうえ決めることになる。この交渉には AI 自身は意見を言う立場にはない[j]。

厳格責任については、AI が法人格を持つと責任は AI 自体に留まり、AI や運用者や所有者[j]にたどり着かないかもしれない。

製造物責任は、むしろ AI 自体が末端利用者からの文句を受けて、責任を製造者に問う形になるのではないかと。ただし、上で述べたように AI の製造物責任を一方的に製造者に押し付けるのは、AI に複雑さや自律性から見て困難であるし、無理強いすれば AI の開発者、製造者の委縮を招く。したがって、関係者が協力して再発防止のための事故原因究明を優先させるような枠組みが望ましい[k]。

Ugo Pagallo[4]は、その中で[1][3]を数多く参照してお

り、AI エージェント（ロボット）に目的に応じた法人格などを与えることを否定はしていない。しかし、それが問題を全て解決するとは見ていない。むしろ、AI やロボットに人間に対して、ある行動を促進あるいは抑制するようなデザインを組み込むことを[4]の終章で提言している。

筆者の私見では、このようなデザインが問題を根本的に解決するとは思えず、Ugo Pagallo も懸念しているように独裁政権によるパターナリズム的介入の余地を残すと考えられる。デザインあるいは技術の進歩によって自然人には今までに経験したことがない判断を強要されるという考え方もある。たとえば、AI を搭載した手術ロボットの法的位置づけなどが例に取り上げられる[8]。したがって、一般人に理解可能な明確さと透明性を持つという条件の下で目的に依存する部分的な法人格を与えることによって、人間と AI が共存する見通しのよい社会を目指すことを是としたい。この方向に沿う考察を次の節で述べる。

4. アクターネットワーク理論の適用

4.1 Latour の言説の Teubner による解釈

ここまでは、AI エージェントの法的位置づけを直接に扱ってきた。一方で、自然人と AI エージェントは異なる点はあるが、両者とも知的能力をもつ行為者とみなして、社会の構造全体を構想するという間接的アプローチも有用である。そのための有力な考え方として Latour のアクターネットワーク理論[9]がある。この路線を発展させたアイデアとして、アクターである自然人と AI エージェントを峻別しつつも、両者が共存する社会の在り方を模索する Teubner の言説[5]を以下で紹介する。

AI エージェントという非人間を人と同格とみなすことは、今日の社会的現実であり、将来は政治的に必要であるというのが Teubner の未来予測であり、それをアクターネットワーク理論によって定義づけようとしている。

アクターは物理的な実体ではなく、自らのコミットする一連のコミュニケーションで定義される。過去にコミュニケーションした内容に対するコミュニケーションが積み重なることによってアクターが特徴づけられていく。

組織、企業、共同体あるいは法人という社会における集団的アクターは、それらがコミットしてきたコミュニケーションの積み重ねとして定義されると同時に、それ自体の意思決定構造と社会システムへの拘束として現れる。

以上のようなアクターの定義において、非人間である AI エージェントはどのようなアクターになるか、という問い掛けへの解答として、AI エージェントを自然人とみなせるために設定した以下の 3 条件によってアクターとして位置付ける。

[h] Ugo Pagallo は製造物責任にはこの分類に含めていない。

[i] 1) の AGI の場合は AI 自体も意見表明し交渉の当事者になるかもしれない。

[j] ここでは法人たる AI への出資者ということになるだろう。

[k] 航空機事故は事故調査委員会が主導して原因究明と再発防止を行うことが世界的みれば標準的である。

1. ブラックボックス

ブラックボックス化した AI エージェント (=非人間)の行動は、それを観察できる外部者との関係の履歴によって予測できる。

2. 二重の偶発性(double contingency)

二重の偶発性とはパートナーのアクションに基づいて自分のアクションが決定されることをいう。内部の動きはブラックボックスであっても、それを二重の偶発性に置き換えることによって、AI エージェントの次の行動を外部者からの働きかけに対応するものとして予測できる。したがって、AI エージェントは外部者にとって対応可能になる。

3. 擬人化

人々は、非人間についてあらゆる擬人化された仮定をし、非人間である AI エージェントが人間であるかのように考えて付き合うことができる。

AI エージェントが上記の 3 条件によって人間のアクターと同じようにみなせるとなる。非人間的実体である組織、企業、および国などが上記の 3 条件を満たすことは明白である。これによって、上記の組織、企業、国などは法的擬人化ができるので、法人としてアクターとみなされる。

AI エージェントがアクターとみなせると、本格的な政治交渉と複雑な経済取引ができるようになる。例えば、米国の第 14 条統一電子取引法によれば

「契約は、たとえ個人が電子エージェントの行動または結果として生じる条件や合意を認識またはレビューしていなくても、当事者の電子エージェントの相互作用によって形成される可能性がある。」

ということになっており、個々の人間の知識や行動なしに、電子エージェント (AI)間の相互作用による契約が成立する。言い換えれば、契約という場面においては、AI エージェントは自然人と同等の法的位置づけを持つ。

ただし、現在においては、AI エージェントが契約を締結し、これらの行為がマシン自体から宣言される場合、契約上の行為は、コンピュータの背後にいる人間に帰責するとされ、AI エージェントは Ugo Pagallo の分類による 1) 人間が完全にコントロールできる AI ツールとみなされている。

話を Latour に戻そう。Latour は、二重の偶発性の最小限化する前提に基づいて、非人間を表す「アクタント(actant)」という概念を導入した。さらに、擬人化に頼ることをやめ、人間の介入のない状態での、AI エージェント間で契約は、アクタント間のコミュニケーションとして解釈した。

このように人間のアクターと非人間のアクタントを分離してしまったので、非人間である AI エージェントに政治的あるいは経済的、社会的行動の資格を与えるためには、人間のアクターと非人間のアクタントの連合体[1]を作らな

ければならない。

アクタントは非人間であるため、自然人のような心理システムを持たない。連合体におけるこの状況の解決方法にアクターネットワーク理論の本領が発揮される。すなわち、アクタントの心理的能力の欠陥は、連合体に属する人間であるアクターの能力によって補われる。二重の偶発性を適用すれば、アクタントである AI エージェントはアクターである人間の自分に対する行動から学び、その学習結果を用いてアクターに対して起こした行動により、アクターはアクタントの行動が自然人に近いものになることを確認する。このようなやり取りを通じて、人間のアクターと AI エージェントのアクタントは継ぎ目なくつながっていくことになる。一方で、人間は情報検索・収集能力や、記憶力においては機械である AI エージェントに劣るところがあるので、その能力についてはアクタントである AI エージェントに補ってもらう。ここまでの説明でアクターとアクタントは独立して行動すると考えてきた。しかし、アクターである自然人がアクタントである AI エージェントを自分の代理として使うような連合体の作り方もある。

以上をまとめると、図 2 に示すように、アクターとアクタントの連合体、アクタント、アクターが共存する社会は、社会システム中に分散された知性や知識によって適切に補い合うことによって共同体として機能していくことになる。

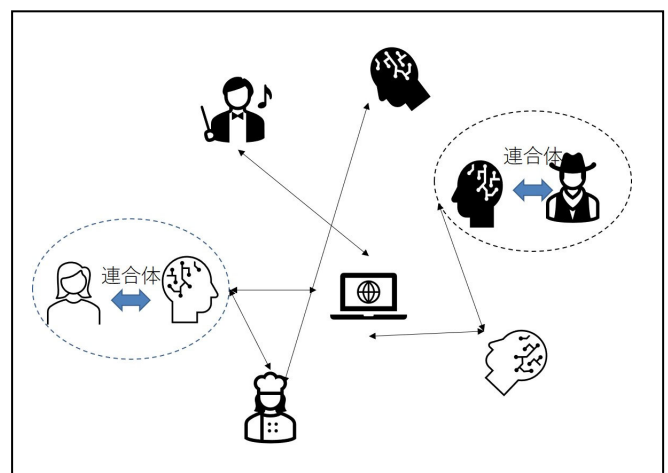


図 2. 連合体、アクター、アクタントが共存する社会
Figure 2. A society where hybrids, actors and actants coexist

4.2 連合体、アクター、アクタントが共存する社会における技術的現実

図 2 に示すような社会では連合体における上記の補い合いによって、AI エージェントが実質的に政治や経済の交渉に参加できるようになる。ただし、単独の AI エージェントの社会参加を実施するためには、Solum, Chopra & White, Ugo Pagallo が提案してきたように AI エージェントに法人格与えなければならない。筆者の個人的感覚であるが、図

[1] この連合体をハイブリッド (Hybrid) と名付けている。

2 のように AI エージェントが人間のアクターを代理する連合体が社会に溢れるようになると、連合体内部の人間アクターが AI エージェントを信用できるようになり、AI エージェント単独でも法人格を持つことに対する反対論が弱まるのではないだろうか。

しかし、この期待には実際のところ技術的関門がある。AI エージェントが人間との係わりを積み重ね、人間に有用な情報を与えつづけるには、AI エージェント自身が継続的な外界観察を行うことができ、それに基づくリスクの予見や分析の能力を持つことが必要である。人間の信用を得るには、この能力が必要だが、それを実現することが技術的に難しい。以下に困難さを指摘する。

困難 1：外部世界のどのような要素に着目するかを決める問題。選定すべき要素は時々刻々と変化する。これを正確に選定することは、AI において完全な解決が得られていないフレーム問題を解くことにほかならない。

困難 2：リスク評価手法の設定は、そもそも人間のアクターにとってのリスクとは何かを知らなければならない。新規のモノゴトに対して、それをリスクとみなすかどうかは、極めて困難である。

このように書くと、期待薄にみえてくる。しかし、そもそも人間はフレーム問題を 100% 解いているのだろうか？ また、リスク発見と評価を正確にできているのだろうか？ いずれの答えも否定的と言わざるをえない。よくて近似的に解いているだけであり、悪くすれば正しく解けずに、大きな失敗をしている。このような状況に人間が気付けば、連合体における AI エージェントと共同作業によって、現実の問題に対して、人間単独より、よく対処できるようになるだろう。

4.3 AI エージェントと人間での共進化

4.2 節では知識レベルでの非人間の AI エージェントと人間のアクターの間での知識共有ないし知識の補い合いについて述べたが、より深いスキルのレベルで影響し合う効果も考えなければならない。すなわち、AI エージェントがどんどん賢くなると、人間もそれに適応するように、その能力が変化していくという図 3 のような現象が顕著になってくる。

相互影響の具体例としては、GPS で位置情報が分かるスマホの出現以前は、人間は地図を読んで目的地にたどり着くスキルを持っていた。ところが、GPS で位置が分かるスマホが、目的地への経路を地図での方向指示、あるいは音声による方向指示で教えてくれるようになると、人間はそれに頼り切り、以前にもっていた目的地到達スキルを失っていく現象が顕著になってきている。そうして失ったスキルを補完するためにスマホ上の AI による道案内のユーザインタフェースはどんどん使いやすく、正確なものになっ

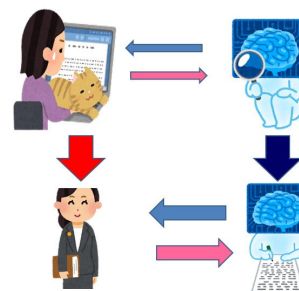


図 3. AI エージェントと人間の共進化

Figure 3. Coevolution of AI agents and human beings

ていく。人間側は、例えば目的地における行動計画をスマホで情報検索して練り上げるようになるだろう。このような変化は、AI エージェントの進化だけではなく、人間側の退化や進化もあるので、AI エージェントの法的な立場が強化される方向に進むと考えるべきであろう。換言すれば、図 2 のようなアクター、アクタント、連合体は、各々の能力を変えながら共進化していく。AI エージェントに与える法人格の内容もこの共進化によって変えていく未来図を持つ必要がある。

5. 残された課題

5.1 代理人の法理

本人の代理人 (agency) になる法的位置づけをパーソナル AI エージェントに与えることができるかどうかという問題について考えてみる。自然人、法人とも代理人になることはできる。4 節までに提案してきたようにパーソナル AI エージェントになんらかの法人格が与えられたと仮定すれば、パーソナル AI エージェントは代理人になれる。民法第 102 条によれば、「代理人は、行為能力を要しない」とされる。ということは、代理人は、意思能力者であれば足り、行為能力者であることを要しない。ここで意思能力とは、自己の法律行為の結果を弁識するに足りる精神能力である。パーソナル AI エージェントが法人として認められるからには、意思能力があり、本人の代理人となれることになる。代理人の役割は、本人の法的な活動の領域を拡張させることである。したがって、4.2 節で述べたことにしたがえば、アクターである本人が、アクタントであるパーソナル AI エージェントによって法的な活動領域を拡張される。これは、アクターとアクタントの連合体が自然人であるアクター単独である場合より行動能力が拡張し、かつ法的な裏付けもできる。

ただし、代理は本人に死によって終了する。パーソナル AI エージェントを代理にしてあっても、死とともに代理関係はなくなり、本人のデジタル遺産の管理や処理を継続することはできない。日本法の場合、相続財産の信託人と

いう制度があるので、本人の代理人であったパーソナル AI エージェントは、死と同時に信託人という法的位置づけになるルートはあるだろう。自然人の信託人を別に選ぶなら、代理人であったパーソナル AI エージェントはその信託人の仕事の AI 支援ツールになることは本人の遺志によって可能だろう。さらに、信託人がアクター、パーソナル AI エージェントがアクタントになれば、両方で連合体を作って、死後のデジタル遺産管理、有効利用などができる。

英米法では受託者も代理人も信託義務を負う受託者 (fiduciary) として扱われてきた[11]。英米法に従えば、生前からパーソナル AI エージェントを受託者の位置づけの代理人にしておくことができる。パーソナル AI エージェントと本人の関係で考えれば、パーソナル AI エージェントは本人からトラストされる存在[10]、すなわち信託義務を負う受託者とするほうが実態に近いかもしれない。ただし、英米法でも代理人としての関係は本人の死とともに消滅するので、死後のデジタル遺産処理をパーソナル AI エージェントに任せるためには、上記の日本法と同じ問題がある。

5.2 残された問題

代理人や信頼関係を持ち込むにしても、結局のところ、パーソナル AI エージェントを社会の法的一員として活動させるためには、法人格の付与が必要である。AI エージェントに法人格を与える場合は、AI エージェント自体が情報を扱う行為の執行、法的判断を行えなければならない。そのためには、法人の定義として法人擬制説ではなく法人実在説を採る必要がある。

法人実在説に基づくにせよ、AI エージェントにどの程度に部分的な法人格を付与できるかが残された問題である。

謝辞 本研究は JST RISTEX「人と情報のエコシステム」研究開発領域:研究開発プロジェクト「PATH-AI:人間-AI エコシステムにおけるプライバシー、エージェント性、トラストの文化を超えた実現方法」の補助を受けて行っている。

参考文献

- [1] Lawrence B. Solum. Legal Personhood for Artificial Intelligences. 70 North Carolina Law Review. 1231 (1992).
- [2] Lawrence B. Solum. Artificially Intelligent Law (February 14, 2019). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3337696>
- [3] Samir Chopra and Laurence F. White. A Legal Theory for Autonomous Artificial Agents. THE UNIVERSITY OF MICHIGAN PRESS. Ann Arbor. 2011
- [4] Ugo Pagallo. The Laws of Robots-Crimes, Contracts, and Torts, Springer, 2013
- [5] Gunther Teubner. Rights of Non-humans? Electronic Agents and Animals as New Actors. Max Weber Lecture Series, MWP 2007/04

- in Politics and Law Lecture Delivered January 17th 2007
- [6] 八山 幸司. 米国のフィンテックにおける人工知能の活用 (フィンテック AI) の現状と課題. Jetro Report NyRp201704. (2017). https://www.jetro.go.jp/ext_images/_Reports/02/2017/1e2adc4583372c88/nyRp201704.pdf
- [7] アルゴリズム・AI の利用を巡る法律問題研究会. 投資判断におけるアルゴリズム・AI の利用と法的責任. 金融研究 第 38 巻第 2 号 (2019) <https://www.imes.boj.or.jp/research/papers/japanese/kk38-2-1.pdf>
- [8] Peter-Paul Verbeek, ピーター＝ポール・フェルベーク (鈴木俊洋訳). 技術の道德化. 法政大学出版局, 2015
- [9] Bruno Latour. ブリュノ ラトゥール(伊藤嘉高訳). 社会的なものを組み直す: アクターネットワーク理論入門. 法政大学出版局, 2019
- [10] 中川裕志. デジタル遺産のパーソナル AI エージェントへの委任. 情報処理学会 EIP 研究会 91(26), 2020 年 11 月 26 日
- [11] 樋口範雄, 佐久間毅 編『現代の代理法 アメリカと日本』弘文堂.(2014). ISBN 978-4335355813