

AI エージェントの法的位置づけ

中川裕志

理化学研究所・革新知能統合研究センター

概要

- AIエージェント（個人に代理をするAI）はAIの重要な応用分野
- ただし、AIエージェントが果たせる役割は、その法的位置づけによって異なる。
- 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、最近のものとしては 2019年という時間幅である。

◆結論

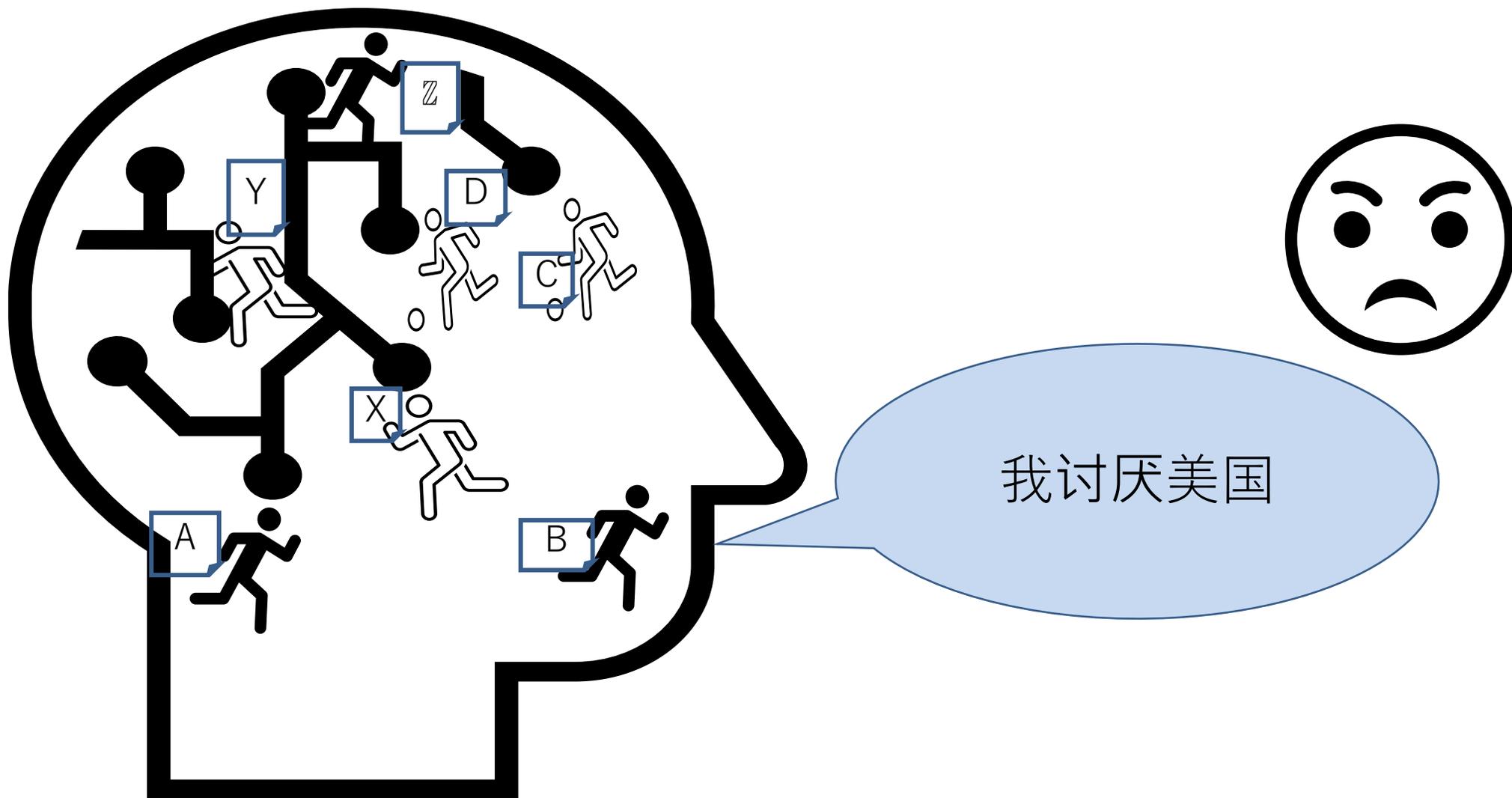
AIエージェントが人間と同じ人格を持つ議論は棚上げにして、何らかの法人格を持たせるのが現実的。そのようなAIエージェントの決定が人間社会に実際の影響を与えるようならざるを得ないだろう。

AIエージェントの法的位置づけ 関連論文

- Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. Rev. 1231 (1992).
- 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
- Burkhard Schafer: ZOMBAIS: LEGAL EXPERT SYSTEMS AS REPRESENTATIVES “BEYOND THE GRAVE” Scripted Volume 7, Issue 2, 384-393. August 2010
- Samir Chopra and Laurence F. White: A Legal Theory for Autonomous Artificial Agents. THE UNIVERSITY OF MICHIGAN PRESS. Ann Arbor. 2011
- Ugo Pagallo: The Laws of Robots-Crimes, Contracts, and Torts, Springer, 2013
- Lawrence B. Solum,, Artificially Intelligent Law (February 14, 2019). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3337696>
or <http://dx.doi.org/10.2139/ssrn.3337696>

Solum, Lawrence B., Legal Personhood for Artificial Intelligences(1992), Artificially Intelligent Law (2019)

- ◆法制度は実践的な判断の形式的な集積地
- ◆AIの権利制限派の主張
 - プログラムが意味を処理する能力である「志向性」を欠いているため、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
- ◆志向性はどこで現れるのか？
 - AIエージェントの観察不可能な内部の情報処理構造ではなく
 - AIエージェントと外部者の観察可能な対話に表れると考えるべき
 - i.e. 中国語の部屋



財産管理運用するAIエージェント

- **Stage 1:** プログラムは、人間の介入を必要とせずに、多数の単純な信託の管理において人間の管財人を支援
- **Stage 2:** AI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のアドバイスに従う必要がある
- **Stage 3:** 人間の管財人は不要で、AIがすべてを行う。問題は、誰がそのようなAIを所有しているか
 - ▶ AIプログラムが管財人として機能する能力を持っている場合、法律はAIに法人格を与えるべき

• 以下は反論と回答

- 1. 責任とれるか？**：AIは責任を負えない。つまり、AI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
 - AIは、保険に加入して、損害賠償責任を負える。
 - 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問題ない。
- 2. 判断できるか？**：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の能力を疑いたくなる
 - チューリングテストに合格したAIにはならOK。
 - 実は、AIトレーダはその高速性により、人間トレーダの能力を超えている
 - 法律は、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るべき

Future AI on 1992

- AIエージェントは人間のユーザーと対話できる
- リサーチクエスチョンについてエージェントと話し合い、エージェントが検索戦略を考案
- AIエージェントはデータベースの検索方法について頼んだユーザーよりも知識豊富なので、自律的に検索してくれる
- 人間はAIエージェントにアドバイスを提供する
- ▶ 人間のユーザーがこのようにAIエージェントと対話するとき、彼らは「彼ら自身の心を持っている」ように見えるかもしれない

Constitutional rights for AI

●異論1：AIは人間ではない

- ◆「人間中心主義」の議論：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の憲法上の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に憲法上の人格権を認めるべきではない
- 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ある場合、問題は、AIやクジラまたはエイリアンがこれらの資質を共有するかどうかになる → 無意味な議論ではないか？
- ここでは、道徳的な人間性よりも**法的**人間性に焦点を当てていることに留意
 - 法律は外部に表れた関係を扱う仕組み

➤ 異論2：何か足りない

◆ AIは魂がない

- 法定書面では機能しないはず

◆ AIは意識を持ってない

- 他人の心に直接アクセスできないため、行動と自己報告からのみ意識を推測しかできないが、この推測から意識がないと証明できるか？

◆ AIは志向性を持たない

- サールの中国語の部屋を思い出してください → 志向性は外部から測るもの

◆ AIは感情をもたない

- 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。
- 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る

◆AIは興味を持たない

- AIは、興味や商品があると主張するかもしれないが、AIにとってのメリットは、人間にとってのメリットとはまったく異なることには注意

◆ AIは自由意志を持たない

- 人間の自由意志は、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。しかし、この意味で、AIは自由意志を持つこともできる。

◆ シミュレータでしかないという議論

- 行動の証拠と基礎となるプロセスに関する知識によって、AIが人間の精神の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が人間の精神の必要な特徴を持っていたと信じてもよいのではないか

- ◆能力は依然として関連性があるが、能力だけでは、AIが憲法上の人格権を持つには十分ではない
- Solumの予測： AIを日常生活の中で意図的なシステムとして扱うことが有用なら、AIに真の志向性がなくても大きな違いはない
- 倫理や政治のように、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」に関連する
- 私たちの**personhood**の（法的）概念と私たちの**personality**（倫理的or情緒的）の概念の間関係または境界線の議論がキモになる
 - AIが正しい行動をとり、認知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した場合、AIを人として扱うことができる

After 27 years: Solum, Artificially Intelligent Law (2019)

1. AIには、法規範を生成する能力あり ← 大きな進歩かもしれない
 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あり
 - このためには継続的は社会観察とリスク分析能力が必要
-
- ◆ AIの法制度への法的権限の委任は物議を醸す。
 - ◆ AIへに委任が法律になったとしても、立法を可能にすることは抵抗される。
 - ◆ 敗者（=人間の立法者？）がいるだろうし、間違いなく彼らの代表はプロセス上の理由で決定に抵抗する

Delegation and Democracy

- ◆ **民主的正統性**：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考え。
 - ▶ 民主的な正当性が、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場合、それは法的能力を持つAIに権限を委任する立法によっても満たされる
- ◆ **透明性としての正当性**：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
 - ▶ 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の作る法制度がブラックボックスであるので、法規範透明性の懸念は深刻
 - ▶ それ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
- ▶ 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。

Delegation and Democracy(skip)

- **憲法上の役割の正当性：**

- 公務員は憲法上の規範のシステムによって指定された合法的な役割を持っており、公務員の行動は憲法上の権限を持つ公務員によってのみ行われるべき考え
- 独立した規制機関（たぶんAI）への委任は、法律がその機関に十分なガイダンスを提供するいわゆる「理解可能な原則」を提供する限り、 **constitutional**

- **衡平法、equity**◆common lawおよびstatutory lawを補うもの。

- AI法が、人間の規制当局の能力と同じくらい優れた衡平法を執行する機能的能力を持っていれば、異論はなくなるのではないか

Teubner: Rights of Non-humans? Electronic Agents and Animals as New Actors. 2007

- ◆ポストモダニズムのアクター理論の中にAIエージェントを位置づける言説
- 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
- Latourが示唆するように、新しいアクターの特性は、政治のエコロジーのさまざまな場所の多様性に応じて非常に異なる
- ◆非人間の権利のためのポストモダニズム... AIエージェント
 - ✓アクターを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
 - ✓コミュニケーションの内容に対するコミュニケーションがアクション
 - 社会における集団的アクター（組織とか共同体あるいは法人）は、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。

Personification(人間化)からの帰結

1. ブラックボックス

- ブラックボックス化した非人間 (=AI)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 (=パートナーのアクションに基づいて自分のアクションを決定される)

- 因果関係の帰属スキームを二重の偶発性に置き換えることによって、AIの内部ダイナミクスを外部者から対応可能にする

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、非人間が人間であるかのように考えて行動する。

- ◆ 非人間的実体、正式な組織、協会、企業、および国が法的擬人化によって本格的な政治交渉と複雑な経済取引を開始することを可能にする
- 「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ります。」（第14条統一電子取引法：米国）
- 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
- 電子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

- ◆ 重要な問題だが、時間の都合でskip
- ◆ プログラムの指示から離れて、AIは意味を独立して処理する能力がないため、AIには「志向性」が欠如
- ◆ AIが学習可能だとこの問題はどうか？
- ◆ 代理人に関する法は適用されるか？ 契約法では、本人との関係において、代理人は自律的な決定による処理ができる
 - ◆ 詐欺や間違いの場合はどうか？
- ◆ Common law における「意図」、「信念」、「欺瞞」の概念は、電子契約において再解釈する必要がある

- ▶ 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「行為者(actant)」として導入した
- ▶ Latourは、擬人化に頼ることをやめ、
- ▶ 人間の介入のない状態での、AIを使用して電子エージェント間で契約は、**actant間のコミュニケーション**として解釈した

Latour's Hybrids: the association between human and non-humans

- ▶ 非人間 (AI)に政治的行動の能力を与えるためには、**Hybrids**すなわち**人間のactorと非人間のactantの連合体**を作らなければならない
- ▶ 非人間の心理システムの能力の欠陥は、社会システムの分散された知性によって適切に補われる
- ▶ **Hybrids**内で人間とAIの特性を組み合わせると、AIが政治交渉に参加できるようになる
 - ▶ 経営者が自らを会社を代表して行動するのと同じように、社会システムとしての企業である「本人」を代表する「代理人」です。
- ▶ **Hybrids**は人間とAIの間の係わりの積み重ねといえる
- Actant(AIエージェント) は人間との係わりを積み重ね、人間に有用な情報を与えつつけるには、継続的な外界観察とリスクの予見や分析の能力が必要 → ここが技術的に難しい

名著 : Chopra and White: A Legal Theory for Autonomous Artificial Agents. (2011)

- 1.2. Artificial Agents and Legal Agency -18
 - Chapter 2. Artificial Agents and Contracts -29
 - Chapter 3. Attribution of Knowledge to Artificial Agents and Their Principals -71
 - 4.2. Liability Schemes for Artificial Agents Understood as Instrumentalities -123
 - 4.3. Liability Schemes for Artificial Agents Understood as Actors and Agents -127
 - 4.4. Liability of Suppliers of Artificial Agents -135
 - 4.5. Responsibility and Liability of the Artificial Agent -145
 - 5.1. Could Artificial Agents Be Legal Persons? -153
- ◆ AIエージェントの法的位置づけに関する包括的かつ詳細な議論展開されている。基本的にはSolum、Teubnerの議論と同じ路線なので、時間の関係で細部は省略。興味のある方はお読みいただければ幸いです。

5.1. Could Artificial Agents Be Legal Persons?

- ◆ 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
- AIエージェントを拒否または付与するための最良の議論は、概念的ではなく実用的なものあるべき
- AIエージェントの機能と社会的役割を考えると、法律は、どちらとも決定しないだろう
- ✓ 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エージェントの自己破壊的行動はしない。

5.4. Philosophical Objections to Personhood for Artificial Agents

- 「法人」の概念は、財産の概念と密接に関連
- ◆人工知能が人格を獲得する可能性に対する異議は、「人工知能の可能性に対する一般的な議論に似ている。**AIの計算アーキテクチャに「何かが欠けている」と仮定して、人格をもてない、とするもの**
 - たとえば、AIエージェントが罰を受けない（道徳的感覚がないため）、または裁量的な決定を下すことができない（自由意志と自律性がないため）と思われるのは、これらの資質の欠如によるとしてくる
- これらの異議は、「人間の属性が自然のものとして理解される」という懐疑論の共通の根拠をもつ。

Free Will

- ◆最強の異議は、「単なるプログラムされた機械だから、AIエージェントは自由意志を持ってない」
- 自由意志の概念の運用上の仮定は、「選択が自由であるシステムとそうでないシステムとの間には明確な区別がある」(Sloman1992)。
- しかし、エージェントアーキテクチャを詳しく調べると、違いは明白には分からない
- 人間の自由意志のもっともらしい説明は、エージェントの側での推論と審議によって引き起こされた場合、人間の行動はfreeということ。
- これと同じ意味で、AIエージェントは自由意志を持つ(相互に違いがない)

5.6. Recognizing Artificial Agents as Persons

- ◆ Personhoodは、共通の生活形態と共通の必要性を反映する概念を組織化する形で現れる。
- ◆ AIエージェントが法人になるための重要な決定要因は、社会的および経済的に複雑な相互作用を形成できること

- Chopra & Whiteを基礎におく Ugo Pagallo の名著
- **Ugo Pagallo: The Laws of Robots-Crimes, Contracts, and Torts,**
- **Springer, 2013**

AI/AGIの法的格付けによる対応表 Ugo Pagallo

AIの行為	責任	免責	厳格責任	製造物責任	不当な損害
完全な法的人格AGI = 人間					
的確な行為者 (限定的自律AI)					
ツールとしてのAI					

- ◆ **責任**：罪刑法定主義によって、行為者が引き受ける責任。
 - ✓ 箱の中は行為ないし責任を引き受ける者 ないしは法益を受ける者
- ◆ **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
- ◆ **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
- ◆ **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
 - ✓ 箱の中にはその損害からの保護方策など。

AI/AGIの法的格付けによる対応表： 一般の場合

AIの行為	責任	免責	厳格責任	製造物責任	不当な損害
完全な法的 人格：AGI	全ての場合	自然人と同じ	同左	同左	同左
限定的自律 AI	不法行為の 場合、その 理由の所在： AI利用者	損害の閾値 で決定	政治的判 断、新規 立法	AI製造者 (委縮)	AIシステム 利用者の正 当な使用方 法の実証
ツールとし てのAI	AIが損害源 →操作した 人間	技術革新に 対応して決 める	AI運用者	AI製造者	AIシステム 利用者の正 当な使用方 法の実証

限定自律的AIが法的地位（法人格？）を持つと

AIの行為	責任	免責	厳格責任	製造物責任	不当な損害
完全な法的 人格：AGI	全ての場合	自然人 と同じ	同左	同左	同左
限定的自律 AI： 保険、 法人として 財産所有す る	責任の所在： AI利用者、 AI自体	損害の 閾値で 決定	AIの持主 や運用者 にたどり 着かない かもしれ ない	AI製造者 責任はあるかも しれないが、再 発防止のための 事故原因究明が 優先。	AIシステム 利用者の正 当な使用方 法の実証
ツールとしての AI	AIが損害源→操 作した人間	技術革新 に対応し て決める	AI運用者	AI製造者	AIシステム利用 者の正当な使用 方法の実証

あまり議論されてこなかった問題

- 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現象が顕著になってきている。
- AIエージェント ⇔ 人間
- このようにお互いに影響を与え合う状況でのAIの法的位置づけを構築することが重要な時期になってきているのではないだろうか？

GOVERNANCE INNOVATION Ver.2

アジャイル・ガバナンスのデザインと実装に向けて

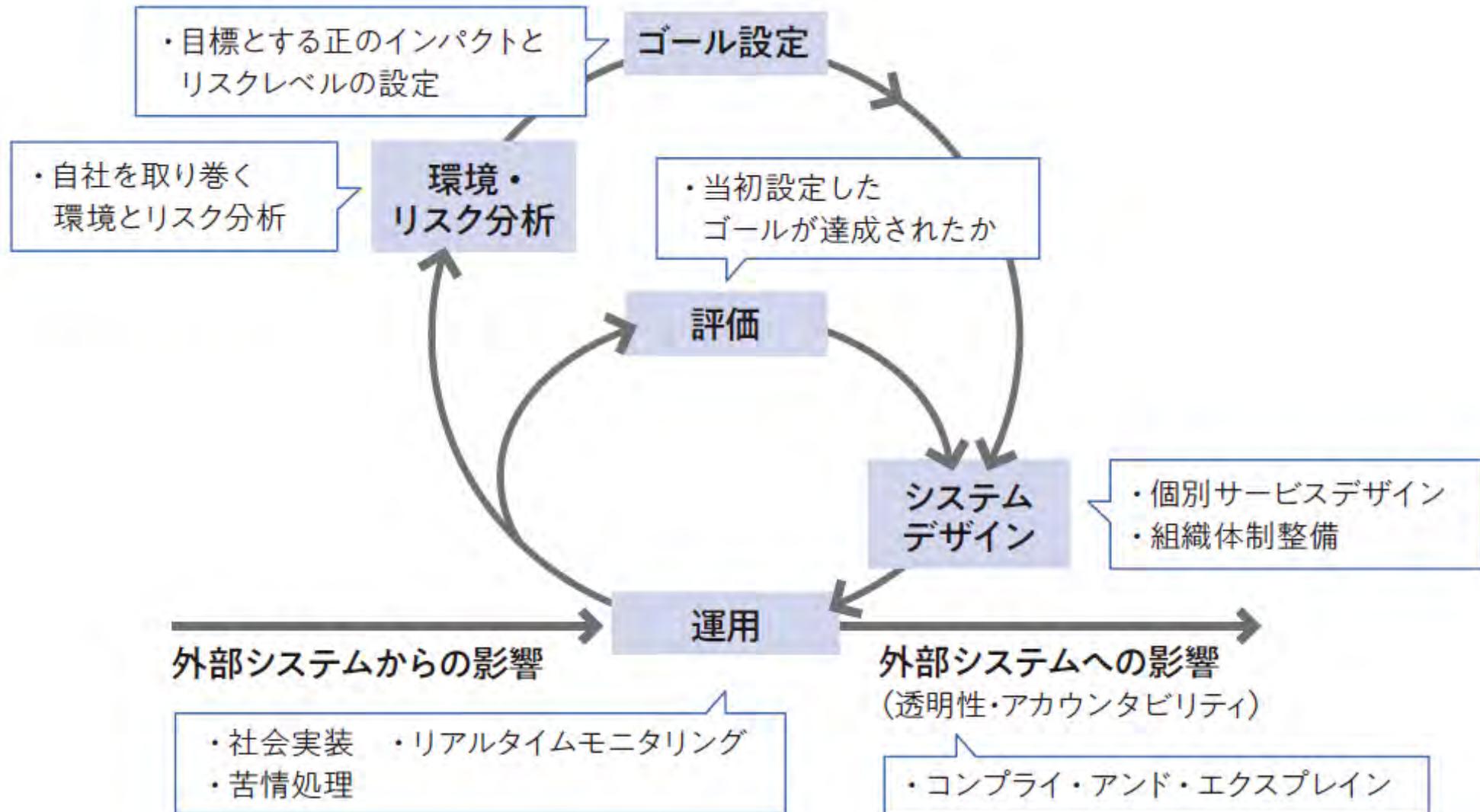


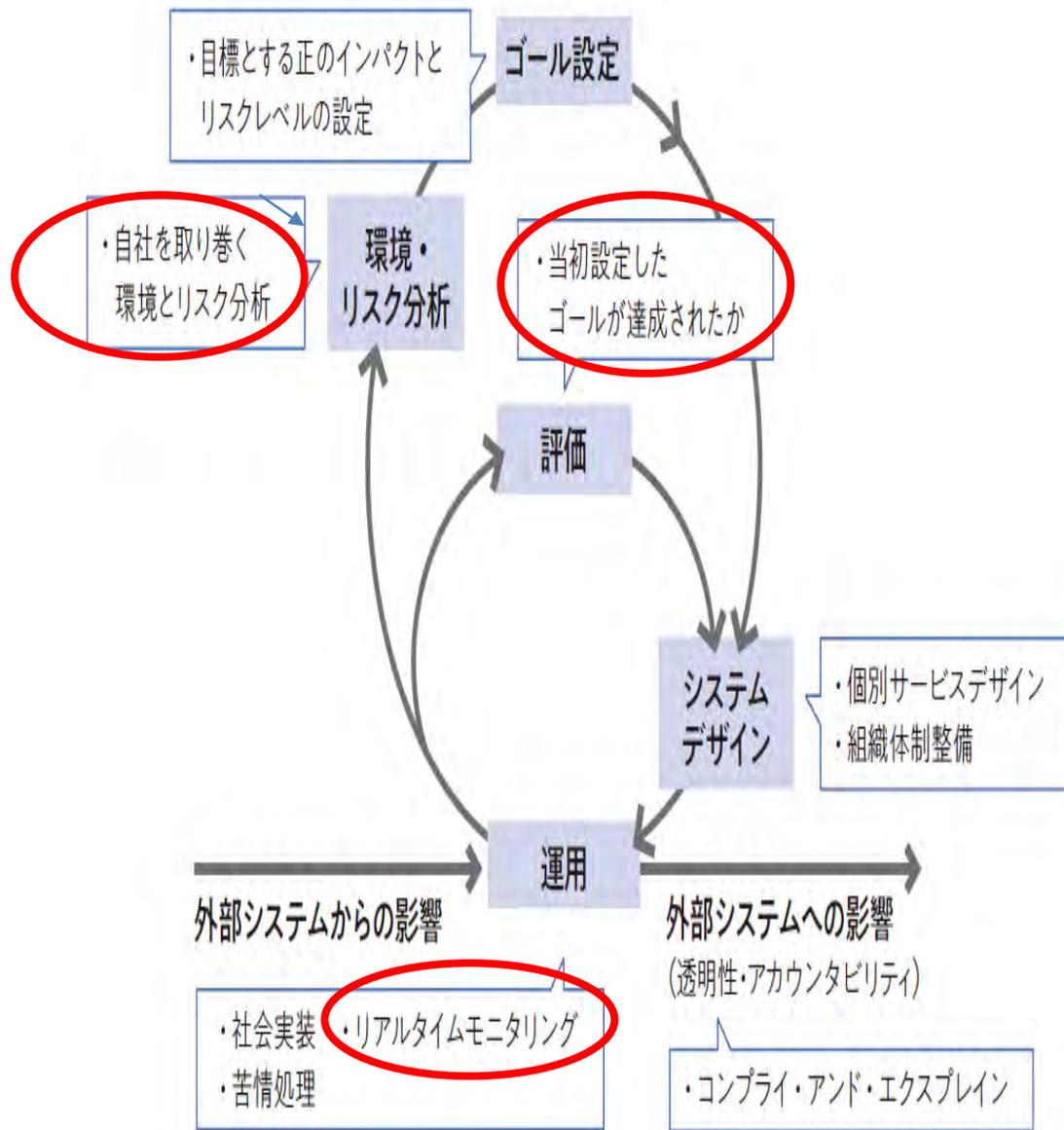
Ver.2.0 JP

目次

- はじめに i
- ショート・サマリー iii
- エグゼクティブ・サマリー iii
- 第1章 本報告書の目的と構成** 1
 - 1.1 Society5.0を実現するためのガバナンスの必要性 1
 - 1.2 本報告書のねらい 6
 - 1.3 本報告書の構成 10
 - 1.4 本報告書におけるガバナンスの定義 12
- 第2章 Society5.0を構成するサイバー・フィジカルシステム(CPS)の特徴と課題** 16
 - 2.1 Society5.0におけるCPSとは何か 16
 - 2.2 より大規模・広範囲・多種類のデータ収集(Digitalization) 19
 - 2.3 高度かつ自律的なデータ分析(Analytics) 20
 - 2.4 アルゴリズムの判断によるフィジカル空間への作用(Actuation) 23
 - 2.5 様々な機能をもつシステムの相互運用(Interoperability) 24
 - 2.6 地理的制約や業種の壁を超える拡張性(Augmentation) 26
 - 2.7 環境変化に伴うシステムの変化(Adaptability) 29
 - 2.8 CPSがもたらすガバナンス上の課題のまとめ 30
- 第3章 Society5.0におけるガバナンスのゴール** 33
 - 3.1 Society5.0における「ゴール」について 34
 - 3.2 「中核的価値」としての基本的人権 36
 - 3.3 経済成長を支える「中核的価値」 40
 - 3.4 Society5.0における基盤的制度 44
 - 3.5 持続可能な社会の形成 46
 - 3.6 Society5.0におけるゴールのまとめ 48
- 第4章 Society5.0におけるガバナンスの在り方**
—アジャイル・ガバナンスのデザインと実装に向けて 49
 - 4.1 総論:アジャイル・ガバナンスの考え方と基本原則 49
 - 4.2 アジャイル・ガバナンスにおける企業の役割 55
 - 4.3 アジャイル・ガバナンスを実現する法規制のデザイン 65
 - 4.4 インフラのアジャイル・ガバナンス 73
 - 4.5 市場におけるアジャイル・ガバナンスの実現 74
 - 4.6 社会規範によるアジャイル・ガバナンス 80
 - 4.7 ガバナンス・オブ・ガバナンス 86
 - 4.8 グローバルなアジャイル・ガバナンスの実現に向けて 90
- おわりに 93
- 〈別紙1〉用語集 94
- 〈別紙2〉第1弾報告書のエグゼクティブ・サマリー 95
- 委員名簿 102

例：「GOVERNANCE INNOVATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」報告書（案）のコメント





- 4.2.4 ガバナンスシステムの運用における企業の役割
- 自らの事業の**モニタリング** ...効率的かつ精緻なモニタリングを行うことが可能となるだけではなく、**リスク状況やゴールの達成状況を随時判断**することで、ゴールを達成するための手段を柔軟に選択
- ↓
- このモニタリングと随時判断もAIエージェントが行う。**単なるモニタリングだけでなく、「判断」まで行う**となると、AIエージェントの法的位置づけを明確しなければならない

ご清聴ありがとうございました

ご議論よろしくお願ひします