

戦略的創造研究推進事業
(社会技術研究開発)
研究開発実施終了報告書

「人と情報のエコシステム」

研究開発領域

「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、トラストの文化を超えた実現方法」

研究開発期間 令和 2 年 1 月～令和 5 年 3 月

中川裕志

(理化学研究所・革新知能統合研究センター・
チームリーダー)

目次

1. プロジェクトの達成目標	3
1-1. プロジェクトの背景	3
1-2. プロジェクトの達成目標	3
2. 研究開発の実施内容	5
2-1. 実施項目およびその全体像	5
2-2. 実施内容	6
3. 研究開発成果	146
3-1. 目標の達成状況	146
3-2. 研究開発成果	147
3-3. 今後の成果の活用・展開に向けた状況	150
4. 領域目標達成への貢献	152
5. 研究開発の実施体制	153
5-1. 研究開発実施体制の構成図	153
5-2. 開発実施者	153
5-3. 研究開発の協力者	155
6. 研究開発成果の発表・発信状況、アウトリーチ活動など	157
6-1. 社会に向けた情報発信状況、アウトリーチ活動など	157
6-2. 論文発表	170
6-3. 口頭発表（国際学会発表及び主要な国内学会発表）	174
6-4. 新聞/TV 報道・投稿、受賞など	178
6-5. 特許出願	180

1. プロジェクトの達成目標

1-1. プロジェクトの背景

社会の隅々まで浸透しつつある情報技術のひとつである人工知能（以下では AI と略記）が個々人になじみがあることは前提であるものの、さらに AI が社会や経済になじみがあること、法制度などの種々の社会制度になじみがあることも、AI が社会において浸透し円滑に利活用されるためには必須の条件である。なじみがあるとは、まず倫理性があり、有益、有用であることに加え、拒否感がないこと、自分自身の一部という感覚があることを意味する。そのうえで、AI をトラストできるという状況になることを期待したい。いくつかの有力な AI 応用システムに関して、このような道筋で研究を進めることによって、想定される AI 応用システムが人間にとってなじみがあるかを明確化することができる。同時に、AI 応用システムが利用者に悪さをしない方策、人間と共生する方策などを明らかにする。

類似の研究としては JST-RISTEX の HITE 領域ですでに終了した橋田浩一氏が代表を務めた個人データの利活用の方針を設計するプロジェクトがある。ただし、本プロジェクトは個人データ利用に限定されるわけではない。本プロジェクトでも対象にする AI エージェントは、ムーンショットにおける石黒浩氏が代表のプロジェクトで対象としているアバターに類似した概念だが、本プロジェクトではトラストの視点から分析していることが特色である。医療応用の AI では JST-RISTEX の ELSI 領域で進行中の米村滋人氏のコロナ対応の IT 技術検討プロジェクトがある。そのプロジェクトでは COCOA のようなスマホ利用のアプリを検討対象にしていたが、本プロジェクトでは、より広い医療目的に関する AI を対象にして分析している。以上まとめると、本プロジェクトでは、多様な AI 応用に関して、上記の「なじみ」ないしトラストの観点から分析であり、幅広い人々に AI の将来像に関する知見、指針を与えることが期待できる。

1-2. プロジェクトの達成目標

◆ 日英共同研究

- 本研究は日英共同研究であり、英国側は the Alan Turing Institute の David Leslie 研究員を代表とした the Alan Turing Institute と the University of Edinburgh との共同チームである。したがって、定期的なミーティング、共同の社会調査などを通じて、提案テーマに関する日英ないし世界的な調査と分析を行う。

◆ AI と文化

- 人々の AI 感の変遷を歴史、文化的背景の要因、および世代的要因の 2 方向から、日英比較を通じて俯瞰的に分析し、公開セミナーの開催、書籍などのメディアで一般人にアウトリーチする。

◆ パーソナル AI エージェント (PAI Agent)

- 個人の代理として種々の行為を行う PAI Agent の基本的概念設計、実用化局面を想定した機能設計、実現の可能性を左右する技術的要素のアルゴリズム開発と検

証を行う。

- 自分が個人データを管理できない誕生以前、死後、病身などの状況における適用方法、社会的受容性、法律的問題点を明らかにし、公開セミナー、論文、公の委員会などで提言を発信する。

◆ **トラスト**

- AI を含むシステムのトラストの在り方とそれが成立する条件について調査、分析する。
- 社会に受け入れられるトラストの在り方とそこに含まれる法的問題を明確にし、公開セミナー、論文、著書などでアウトリーチする。
- PAI Agent の活用の際に生じ得るトラストの問題を事前に特定し、それに対応した研究開発・利活用のあり方を検討する。

◆ **ガバナンス**

- AI システム、とりわけパーソナル AI エージェントを含む社会システムにおけるガバナンスの将来像、法的システム、政治システムの構造およびイメージを明確化し、公開セミナー、論文、著書などでアウトリーチする。
- AI エージェントの国際的ガバナンスのあり方について、国際シンポジウムや国際学会等での発表を通じて、国際的に発信し、関連するステークホルダーとの間で合意形成を図っていく。

2. 研究開発の実施内容

2-1. 実施項目およびその全体像

- ①AI 倫理規範の背景調査
- ②政策、法制度、経済性に関する基礎調査
- ③文化、社会の歴史的背景調査、分析、提言
- ④インタビューなどの社会調査によるデータ取得
- ⑤パーソナル AI エージェントの設計
- ⑥美的感覚の調査と AI による学習
- ⑦既存のガバナンス枠組の網羅的調査、分析、提言
- ⑧利害関係者からの直接、間接の情報収集（英国側、日本側）
- ⑨利害関係者との議論による国際的ガバナンス枠組の設計（主に英国側）
- ⑩日英双方の共同作業による国際的ガバナンス枠組の設計
- ⑪国際的ガバナンス枠組に関する会議開催
- ⑫⑪の会議に結果の出版などによるアウトリーチ

下図で灰色に部分が実際に研究活動を行った期間を示す。

実施項目	2019 年度 (2020.1～ 2020.3)	2020 年度 (2020.4～ 2021.3)	2021 年度 (2021.4～ 2022.3)	2022 年度 (2022.4～ 2023.3)
①AI 倫理規範の背景調査				
②政策、法制度、経済性に関する基礎調査				
③文化、社会の歴史的背景調査、分析、提言				
④インタビューなどの社会調査によるデータ取得			⑧と関連	⑧と関連
⑤パーソナル AI エージェントの設計				
⑥美的感覚の調査と AI による学習				
⑦既存のガバナンス枠組の網羅的調査、分析、提				

言				
⑧利害関係者からの直接、間接の情報収集（英国側、 <u>日本側</u> ）			④と関連	④と関連
⑨利害関係者との議論による国際的ガバナンス枠組の設計（主に英国側）				英国側で 2022年度7月～9月に 実施
⑩日英双方の共同作業による国際的ガバナンス枠組の設計				
⑪国際的ガバナンス枠組に関する会議開催				
⑫⑪の会議に結果の出版などによるアウトリーチ				

2-2. 実施内容

実施項目① AI 倫理規範の背景調査

(1) 目的：

2005年刊行のカーツワイルの『ポスト・ヒューマン誕生』¹が示した人間の知的能力を凌駕する汎用人工知能(以下 AGI と略記)、そして2014年刊行のポストロムの『Super intelligence』²は、人工知能(以下、AI と略記)が発展して AGI、さらに強力な超知能にたどり着いたとき、人類にとってとてつもない脅威になるという言説を、AI 関係者をはじめとする多くの人々の間に流布させた³。そのような背景から AI は野放図に開発すると脅威になるという感覚が生まれ、AI を人間の制御下に置くための方策として AI 倫理の議論が盛んになったと考えられる。

例えば、初期の AI 倫理指針として知られる FLI の Asilomar Principles⁴の項目 10 には「自

¹ R.カーツワイル：『ポスト・ヒューマン誕生』、NHK出版、2005

² N.Bostrom: Super intelligence, Oxford University Press.2014

³ もっともポストロムは人類を支配するような Super intelligence が実現する確率は非常に低いと言っている。ただし、確率が低くても起きてしまったら人類には手の施しようがないので、今から注意を怠ってはいけないと説いている。

⁴ Future Life Institute: ASILOMAR AI PRINCIPLES, <https://futureoflife.org/ai-principles/?cn-reloaded=1>

動的な AI システムは、目標と行動が倫理的に人間の価値観と一致するようデザインする」と要求し、さらに項目 19 には「一致する意見がない以上、未来の AI の可能性に上限があると決めてかかるべきではない」と警告し、項目 22「あまりに急速な進歩や増殖を行なうような自己改善、または自己複製するようにデザインされた AI は、厳格な安全、管理対策の対象にならなければならない」、項目 23「超知能は、広く認知されている倫理的な理想や、人類全ての利益のためにのみ開発されるべきである」と多数の項目において AGI や超知能への警告および人間の制御下におくことを要求している。

また、IEEE Ethically Aligned Design version2⁵（以下では IEEE EAD ver2 と略記）では、Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence(ASI)、という章(pp.73-82)で AGI や超知能について考察し、safety by design という考え方を提唱している。

ただし、その後 AGI の可能性の研究が進み、カーツワイルのポストヒューマンのアイデアの分析が進むにつれて、AGI の可能性はまだまだ先のことであり⁶、まして超知能の実現性は強く疑われはじめた。このような背景から、AI 倫理の当初の動機の一つであった AGI や超知能による AI 脅威論は退潮し、代わって現在ないし近い将来において重要な AI 倫理の課題が語られるようになった。

以下の部分では、AI 倫理における課題の最近の動向、およびその結果として重視されているパーソナルデータの管理の問題について説明する。なお、この項目については、主に中川が担当した。

(2) 内容・方法・活動：

(2.1)AI 倫理指針

(2.1.1)全容

2017 年以降、国内、国外において AI の開発や利活用に関する多くの文書（以下では AI 倫理指針と総称する）が公開された。すべてを列挙することはとてもできないが、参照されることが多いものをおおよその公表時期の古いものから順に挙げると以下の表 1.1 のようになる。これらの AI 倫理指針は 2019 年末現在、Web で公開されているため容易に内容を確認することができる。脚注に URL を記した（2019 年末時点）。

表 1.1.AI 倫理指針

名称	略称	作成した組織	公開時期
----	----	--------	------

⁵ The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design version2: A Vision for Prioritizing Human Well-being with Autonomous and with Autonomous and Intelligent Systems. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

⁶ カーツワイルは2045年にシンギュラリティが起これりAGIが実現すると预言している。

Asilomar AI Principles	Asilomar Principles	Future Life Institute	2017
人工知能学会 倫理指針 ⁷	人工知能学会・倫理指針	人工知能学会・倫理委員会	2017
国際的な議論のためのAI開発ガイドライン ⁸	総務省 AI 開発ガイドライン	総務省・AI ネットワーク社会推進会議	2017
Ethically Aligned Design version2: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems ⁹	IEEE EAD ver2	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	2017
Ethically Aligned Design (first edition): A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems ¹⁰	IEEE EAD 1e	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	2019
人間中心の AI 社会原則 ¹¹	人間中心 AI 社会原則	AI 戦略実行会議、内閣府	2019
Ethics Guidelines for Trustworthy AI ¹²	Trustworthy AI	The European Commission's High-Level Expert Group on Artificial Intelligence	2019
Recommendation of the Council on Artificial Intelligence 、 OECD/LEGAL/0449 ¹³	OECD Recommendation	OECD	2019

⁷ <http://ai-elsi.org/wp-content/uploads/2017/02/%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD%E5%AD%A6%E4%BC%9A%E5%80%AB%E7%90%86%E6%8C%87%E9%87%9D.pdf>

⁸ https://www.soumu.go.jp/main_content/000490299.pdf

⁹ https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

¹⁰ <https://ethicsinaction.ieee.org/>

¹¹ <https://www8.cao.go.jp/cstp/aigensoku.pdf>

英語版は次のURL:<https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf>

¹² <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

¹³ https://www.soumu.go.jp/main_content/000642218.pdf

AI 利活用ガイドライン ¹⁴	総務省 AI 利活用ガイドライン	総務省・AI ネットワーク社会推進会議	2019
----------------------------	------------------	---------------------	------

なお、本プロジェクトの参加者である、中川、大屋、江間は、人間中心の AI 社会原則の執筆にかかわった。また、成原、中川、大屋は、AI 利活用ガイドラインを策定した総務省 AI ネットワーク社会推進会議の構成員として参加し、内容にコミットした。

表 1.1 に列挙した AI 倫理指針において重視されている分野ないし項目を一覧表にして表 1.2 にまとめた。第 2 行には表 1.1 の AI 倫理指針の略称を公開された時期の古いものから順に左から右に並べた。表 1.1 の最左列の 3 行目から下には AI 倫理指針で言及される項目を列挙した。表中、各 AI 倫理指針で重視されている項目には○をつけている。例えば、Asilomar Principles は、AI 制御を重視していることが表されている。

AI 技術に進展が早いいため、ハードローで制約してしまうと、技術的發展を阻害するという考え方が支配的である¹⁵。よって AI 倫理指針は非拘束的ガイドライン¹⁶である。このような状況なので、AI 倫理指針の実効性の持たせ方としては標準として制定することが有力である。特に国際標準は各国の開発者への影響が大きい。代表的には ISO 標準、IEEE P7000 シリーズ (IEEE EAD ver2、1e を思想的基盤としている) である。

表 1.2 AI 倫理指針で重視されている項目

公開順序	→								
	古い						新しい		
AI 倫理指針略称	Asilomar Principles	人工知能学会・倫理指針	総務省 AI 開発ガイドライン	IEEE EAD ver2	IEEE AD 1e	人間中心 AI 社会原則	Trustworthy AI	OECD Recommendation	総務省 AI 利活用ガイドライン
AI 制御	○	△	○	○					
人権	○	○	○	○	○	○	○	○	
公平性 非差別		○	○	○	○	○	○	○	○
透明性			○	○	○	○	○	○	○
アカウント ビリティ			○	○	○	○	○	○	○
トラスト					○	○	○	○	○

¹⁴ https://www.soumu.go.jp/main_content/000624438.pdf

¹⁵ 政府の委員会（総務省の AI ネットワーク推進会議、内閣府の人間中心の AI 社会原則検討会）などの議論による。

¹⁶ 日本では政府諸官庁が公開する法律のガイドラインには指令的な意味合いが強い。

悪用、誤用		○		○	○	○	○	△	
プライバシー	○	○	○	○	○	○	○	○	○
AI エージェント				○	○		△ ¹⁷		
安全性		○	○	○	○	○	○	○	○
SDGs					○	○	○	○	
教育				○	○	○	○		
独占禁止・協調、政策						○	○	○	
軍事利用	○			○			○		
法的的位 置づけ		○		○	○		○		
幸福	○	○	○	○	○	○	○	○	○

本章の以下の各節で項目ごとに AI 倫理指針での重視された経緯などを説明する。

(2.1.2) AI 制御

すでに述べたように、AGI や超知能は人類にとって脅威だから、AI を人間の制御下におけるように AI を設計、開発しなければいけないという言説に則り Asilomar Principles や IEEE EAD ver2 ではこの項目を重視していた。しかし、すでに述べたように AGI はまだ遠い将来の話であることが理解されはじめ、むしろ現在ないし近い将来の AI で留意すべき項目が多数存在することが AI 倫理に係わる議論の中で明らかになった。よって、IEEE EAD ver2 以降では取り上げられていない。

興味深いのは表 1.2 で AI 制御に△をつけた人工知能学会・倫理指針である。この指針の最後の第 9 項目に

「人工知能が社会の構成員またはそれに準じるものとなるためには、上に定めた人工知能学会会員と同等に倫理指針を遵守できなければならない。」

と記されている。第 8 項目までは、人工知能学会の会員に対して、会員が AI 開発において遵守すべき指針を示している。それが第 9 項目ではあたかも AI が社会の構成員となること

¹⁷ Guidelines for Trustworthy AI で述べられている human oversight と human in the loop を組み合わせると必然的にパーソナル AI エージェントになるが、直接の言及はない。

を想定した場合、すなわち AGI の実現した状況を想定し、AGI にも第 8 項目までを守れと書いている。人工知能学会・倫理指針は Asilomar Principle と同時期に書かれているので、同様に AGI を意識した項目が入ったと推測される。

(2.1.3) 人権

人権はほとんどすべての AI 倫理指針において直接的にせよ間接的にせよ触れられている。ただし、その扱いは簡素な言及から、厚みのある記述まで様々である。多角的に人権に焦点を当てているのは IEEE EAD ver2、IEEE EAD 1e であり、この両者が大部の指針であるにしても human rights という表現の出現回数は前者で 90 回、後者で 140 回に及ぶ。また、human rights を種々の視点から取り上げており、IEEE EAD ver2 では principle、personal data、misuse、well-being、autonomous weapon などにおいて人権との関連を記述している。IEEE EAD 1e では人権は焦点を当てる 8 つの一般原則のうちの一つになっている。具体的には、AI は人権を尊重するように設計せよと宣言している。さらに、IEEE EAD 1e では 10 章全てで人権に言及している。

ちなみに、IEEE EAD 1e の 8 つの一般原則は以下の通りである。

Human Rights, Well-being, Data Agency, Effectiveness, Transparency, Accountability, Awareness of Misuse, Competence である。

(2.1.4) 公平性、非差別

Fairnessあるいは公平性については表面的には多くの倫理指針で触れている。公平性を守れという言い方が多くみられるが、公平性の定義にまで踏み込んだものは少ない。例えば、Trustworthy AIは公平性と非差別(non-discrimination)を平行して記述している。ただし、公平性や非差別に直接的な寄与をするバイアスを意識した記述はTrustworthy AIでは結果のバイアスという言い方に終始しており、バイアスの原因であるアルゴリズムやデータへの直接の言及がない。OECD Recommendationと人工知能学会・倫理指針も公平性に関しては同様の方向で書かれている。IEEE EAD 1eのLawの章の266ページには、脚注に「公平性ないしバイアスの統一的な定義を与えることは適切ではない」¹⁸と書かれている。

現在のAIの出力のデータ依存性の高さを鑑みれば、データバイアスはあって当然と見なせる。人間中心AIでは、データおよびアルゴリズムの双方のバイアスに言及し、これらのバイアスが公平性や非差別に影響しうる可能性に留意せよと記している。2019年に公開された総務省AI利活用ガイドラインでは、AIに利用者らが守るべき公平性の原則として以下のように明記している。

¹⁸ 原文は以下の通り : Fairness (as well as bias) can be defined in more than one way. For purposes of this chapter, a commitment is not made to any one definition—and indeed, it may not be either desirable or feasible to arrive at a single definition that would be applied in all circumstances.

「AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人が不当に差別されないよう配慮する。」

個人データの種々の属性、例えば性別、人種、年齢などのうち、公平に扱うべき属性が決まれば、AIによる判断が、それらの属性に対して公平であるように制約をかける方法はずでよく研究¹⁹されている。ただし、公平性はその確保のためにアファーマティブアクションのような方法を用いると、逆差別も起こしやすい²⁰。IEEE EAD 1eではこういった困難さを自覚して、抑制した記述にしたとも考えられる。一方で、公開時期がこれより古いとはいえ、IEEE EAD ver2ではアルゴリズム、機械学習の教師データ、さらに文化によるバイアスにも言及しており、この問題への意識の高さがうかがえる。

(2.1.5) 透明性、アカウントビリティ、トラスト

初期の倫理指針の Asilomar Principles や人工知能学会・倫理指針では透明性²¹、アカウントビリティ²²、トラスト²³という概念は扱われていなかった。これらの概念は、AGI の脅威とは無関係に現在の AI を社会で実利用するにあたって避けて通れないことが認識されたといえる。総務省 AI 開発ガイドラインでは透明性が重要項目として列挙され、その後の IEEE EAD ver2 以降の倫理指針では重要な項目として丁寧に記述され続けている。

まず、これらの諸概念の関係を説明する。

1) AI システムの動作やその結果に対する説明を生成できる AI の説明可能性²⁴、そしてその説明の理解可能性、

2) AI の起こした事故に対する責任者の明確化（法律分野では「答責性」という単語が使われることがあるが、答責性はアカウントビリティと同義で使う場合もある。用語の専門的な定義は避け、ここでは「誰が責任をとるか」という部分を答責性の意味として使うことにする。）

上記 1)、2)という主要な二つの項目が総合されて透明性という概念に至る。透明性にはさらになぜその事故を引き起こしたかという原因究明を辿れる追跡可能性²⁵と当事者ないし

¹⁹ 直接的に公平に扱う属性だけでなく、当然ながら、その属性に間接的に作用する属性も含めて公平化する。詳細は以下の文献が参考になる。神畷 敏弘、『公平配慮型データマイニング技術の進展』、第31回人工知能学会大会、[1E1-OS-24a-1](#)、2017。

²⁰ 拙著『裏側から見るAI-脅威・歴史・倫理』近代科学社(2019年9月刊行)の5章を参照していただくと幸いです。

²¹ transparency

²² accountability

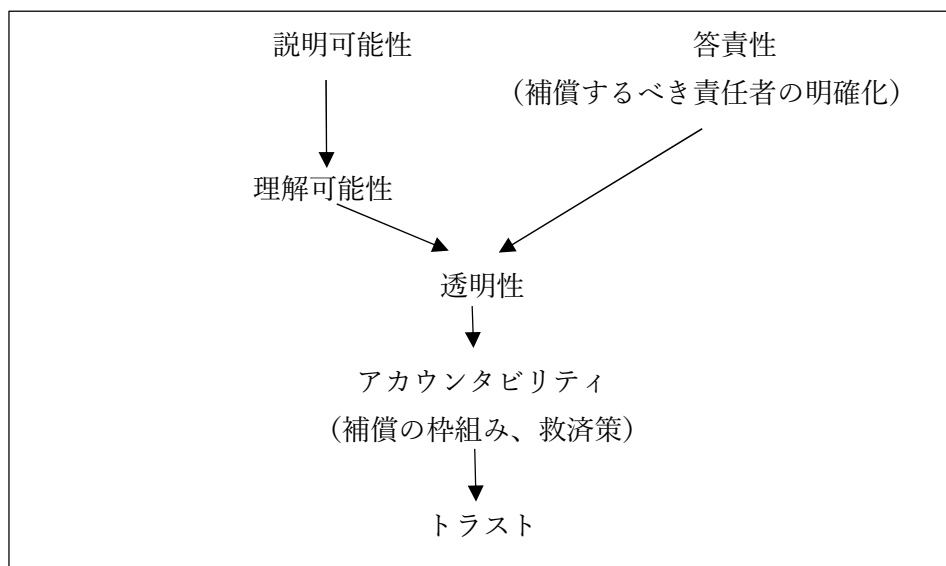
²³ trust

²⁴ explainability

²⁵ traceability

関係者間のコミュニケーションという要素が含まれる。さらに透明性に関するこれらの要素が与えられたときに、実際の補償の枠組みや救済策²⁶、被害の報告義務、今後の被害を最小化する方策などが総合されたアカウントビリティという概念になる。ただし、この説明からもお分かりいただけるようにアカウントビリティは複雑な構造を持っているため、よほどの専門家でなければ理解し使いこなすことはできない。まして一般の AI 利用者は理解するにはハードルが高いだろう。そこで、アカウントビリティの内容はさておき、「とりあえず信用して使ってみよう」という態度すなわちトラストに落ち着くことになる。この全体像を図 1.1 に示す。なお、Trustworthy AI では、トラストできる AI という意味で trustworthy AI という概念を前面に打ち出している²⁷。

この図で注意をしておくべきことは、答責性における責任者の範囲である。事故が起きたときに主として経済的補償をする責任者は、AI システムのオペレータ、設計者、運営者すなわち事業主体、さらにはその事業への出資者であろう。オペレータが明白なミスをした場合は責任を問うことは容易だが、AI システムの複雑性、ブラックボックス化、自ら学習する機能によって、動作の予見性が低い。したがって、オペレータが正しく運用しても、結果がまずいことがありえる。すると、責任は AI システムの設計者なのか、あるいは AI システムで使うツール開発者なのかという技術的に解明困難な状況に陥る。当然、保険もかかっているから、保険会社はこれらを勘案して保険ポリシーの契約を作ることになる。したがって、この状況では契約作成も難問となる。



²⁶ redress という単語が使われることが多い。

²⁷ トラスト(trust)の意味は「AがBをトラストする」という関係を表す。一方、トラストという関係を十分に期待できる、あるいは信頼できるという概念はtrustworthyという単語を用いる。日本では、トラスト、trustedという表現をtrustworthyの意味で使うことが多いが、これは英語的には誤用に近いだろう。

図 1.1 諸概念の関係

ここで期待されるのが、AI システムの動作を明らかにしてくれる説明であり、この説明能力を持つ AI が説明可能 AI²⁸である。ただし作られた説明は少なくとも AI システムのオペレータや運営者には理解可能でなければ意味はない。理想的には、末端利用者にも理解可能であってほしい。しかし、技術的に見ると、AI システムが生成した説明として、内部変数の値の変化などを表示されても開発者でもなければ理解はできない。一般人の利用者を含む多くの人々に理解可能な説明を生成することは困難な課題である。AI の挙動を AI が使っているアルゴリズムに沿って説明するような内部動作由来の説明は研究が続けられてきた。しかし、捗々しい成果があがっていない。そこで、最近では AI の動作を外部から見て理解できる簡単なシミュレータ、例えば、決定リストや、各ノードが条件で、yes、no の各々の場合に移動する先のノードが与えられる決定木のような直観的理解が可能な形式²⁹で表現し、入力から出力結果に至るルートを表示するような方法が研究されている。ちなみに Trustworthy AI では説明する相手は human beings と書かれており陽に特定されていないが、このような現状からみてやむをえない記述であると思われる。

事故時に対応できる説明可能性、責任をとる対象者を明記する答責性、およびすでに述べたように救済策、報告義務、被害最小化方策などが揃うとアカウントビリティという概念が成立する。このような流れは IEEE EAD ver2、1e の法律の章で詳しく述べられている。アカウントビリティは事故の説明を行う責任と補償や救済を総合した概念であるが、日本語でしばしば和訳として使われる「説明責任」は説明をすれば責任を果たしたというイメージで使われており、誤訳ではないかと思われる³⁰。

ところで、アカウントビリティはすでに起きてしまった結果についての説明や対応策を表すのに対して、将来、AI システムが安全に動作することを保証する概念が responsibility である。説明可能性を AI の能力として作りこむことや救済策は responsibility の確立にも必要である。ただし、これはむしろ後に述べる AI の安全性との関係が強い。

アカウントビリティは、背後にある人的、組織的問題、技術的問題の両者が絡み合うため、一般の AI システム利用者にとって理解が困難である。そこで先に述べたように、一般利用者にも通用する **トラスト**³¹という概念が浮上する。表 1.1 の AI 倫理指針では Asilomar Principle、人工知能学会・倫理指針ではトラストは扱われず、IEEE EAD ver2、人間中心 AI 社会原則ではトラストという表現は散見されるが、一般的用語としてのトラストを直接定

²⁸ eXplainable AI 略称 XAI

²⁹ このようにAIシステムを近似するように作られた決定木をBorn Again Treesと呼ぶ。

³⁰ 総務省AI開発ガイドライン、AI利活用ガイドラインにおいては説明責任という表現を避けアカウントビリティを用いており、人間中心のAI社会原則においても「説明責任（アカウントビリティ）」と注意深く併記している。

³¹ trustはAとBの2者の間に成立する関係である。AやBが他の者からtrustされうる存在であることをtrustworthyと言う。

義せずに技術、AI、社会などにおけるトラストと同じ意味で、かなり抽象的に扱っている。これらに比べて、IEEE EAD 1e ではより積極的にトラストを位置づける作業を行っている。すなわち、透明性からトラストを導く条件を以下のように列挙している。

- AI システムの動作と結果に関する説明可能性があり、かつバイアスがなく公平であること。
- AI システムの動作と結果が予め定められた規範的基準を満たすこと。
- AI システムが効率的に作業できること
- AI システムの出した結果に再現性があること
- AI システムの設計、開発、調達、供用、動作、効率化の証明において、良くない結果が出力された場合もそれらを説明可能であり、保証の範囲内では懲罰措置と是正措置が取れること

たしかにこれらの条件を満たしていると AI システムの運業者が保証していると言えば、一般利用者もトラストするであろう。ただし、私見では、一般の末端利用者は、むしろこのような保証をする AI システムの運業者自体が信用できるかどうかに関心があると思われる。具体的に言えば、この AI システム、あるいは同じ業者の提供する AI システムが、過去に事故や不具合をおこしていないこと、あるいは AI システムを提供している業者の過去の行動に上記諸条件を守るという点で落ち度がなかったこと、あるいは補償がきちんと行われたことなどによって、その業者の提供する AI システムがトラストできるということになる。

Trustworthy AI では、trustworthy を lawful(遵法的)、ethical(倫理的)、robust(頑強性がある)と定義している。最後の robust は通常の意味でシステムが robust であること、すなわち同じ入力に対して常に同じ出力を出し、壊れないということと社会的な安定性を意味している。ethical の部分は、倫理の内容が明確に記載されていないため、その意図を理解しにくい。古典的倫理としては、功利主義的倫理、規範的倫理、義務論的倫理、徳倫理、などがあるが、どのような倫理に沿うべきか書かれていない。しかし、これらはどうやら標語またはモットーらしく、Trustworthy AI の満たすべき具体的条件は以下のように列挙されている。

- 人間の主体性および人間による監視
- 技術的ロバストさと安全性
- プライバシー保護とデータのガバナンス³²ができています
- 透明性
- 多様性、非差別性、公平性
- 環境および社会的な幸福への寄与
- アカウンタビリティ

³² この意図は理解が難しいが、おそらく人間が統御できることと思われる。

IEEE EAD 1e の諸条件が技術的であるのに比べて、これら Trustworthy AI の条件はより抽象的ではあるが、その意図は明確である。もちろん、これらが抽象的であることは十分に意識しているらしく、これらの諸条件の改善、追加、明確化に継続的努力を要すると記している。また、Trustworthy AI を基礎におく OECD Recommendation においても、同様の trustworthy の条件を明言している。

以上まとめると、手段としての説明可能性、透明性、アカウントビリティに留意しつつも、より広範囲の人々に理解できるトラストないし trustworthy に強く焦点が当たり始めたといえよう。

(2.1.6) 悪用、誤用

悪用³³、誤用³⁴については Asilomar Principle では AGI そのものを技術自体の悪用とみなしているような書き方に読める。人工知能学会・倫理指針では AI 研究者が悪用の発見や告発者の不利益になることをしてはいけないとやや踏み込んでいる。

IEEE EAD ver2、1eでは悪用、誤用をいかに防ぐかという観点から“Awareness of misuse”という標語を打ち出し、AI開発、利活用の種々の局面で誤用、悪用に関する具体的な提案を行っている。法律の章では、AIの仕組みで防ぐだけでなく、悪用、誤用による被害を受けた者の保護、発見した者からの内部通報制度を法制度化すること、通報者の制度的に保護すること、組織内で悪用・誤用が起きないように教育すること、被害者に対する保険による補償の活用などを提言している。

Trustworthy AI では誤用、悪用に関しては malicious misuse として 1 回触れているが、その内容や対策については書かれていない。OECD Recommendation では安全性やセキュリティの一部として一言言及されているだけである。人間中心 AI 社会原則では悪用についての警告は 1 回されているが、誤用についての言及はない。総務省 AI 利活用ガイドラインでは誤用、悪用とも言及はない。

以上まとめると、悪用、誤用については IEEE EAD ver2、1e が質量ともに圧倒的に優れた分析、提言を行っている。一方、他の倫理指針では言及はあるものの具体性のある記述に欠ける。AI 技術、あるいはその技術的基礎となる機械学習に関して、なにかまずいことが起こると、機械学習や AI の所為にするという行為も見られる³⁵現在、この課題に対する IEEE EAD ver2、1e の取り組みは高く評価されるべきであろう。

(2.1.7) プライバシー

³³ abuse

³⁴ misuse

³⁵ 実名は避けるが、特定の国の応募者を採用しないという発言を公開し非難された人物が、後になってそれはAIが出した結果だと謝罪することが2019年末にあった。

すべての倫理指針で継続的に取り上げられているテーマとして、プライバシーの保護がある。よく耳にするように、プライバシー保護が技術発展を阻止していると仮定してみよう。そうだとすれば、産業界からも多くの人々が参加している IEEE EAD ver2、1e、あるいは産業界からの圧力もあった³⁶と言われる Trustworthy AI で、プライバシー保護を強く主張することには、ならなかったかもしれない。しかし、これらの指針でもプライバシー保護が取り上げられているということは、裏を返せば、プライバシー保護は今や IT 産業で利益を生む分野になってきているからともいえるだろう。このような考え方を最初に広く述べたのは当時カナダのオンタリオ州のプライバシー・コミッショナーだったアン・カブキアンである。カブキアンが提案したプライバシー・バイ・デザイン³⁷の7原則の中の第4原則「プライバシー保護はゼロサムではなくポジティブサムである」がこれに相当する。言い換えると、プライバシー保護は消費者にとって望ましいことであるだけでなく、サービスや製品供給を行う事業者にとって、プライバシー保護が消費者に与える安心感から消費者からの好評につながり、さらにプライバシー保護の法制度が整備されるにつれて法的リスクも回避できるという観点で有益なものになるということである。最近では、GDPR³⁸、日本の個人情報保護法の改正、アメリカにおける California Consumer Privacy Act (CCPA)³⁹などに見られるようなプライバシー保護の世界的潮流が強いことが窺われる。

IEEE EAD ver2 では、プライバシー保護という原則だけでなく、法律的観点から分析して、透明性、アカウントビリティとにおける位置づけとプライバシー保護の実装についての問題を列挙している。さらに IEEE EAD 1e では Personal Data and Individual Agency というタイトルの章を立て、個人データの create、curate⁴⁰、control の側面から人間自身の個人データ管理、および人間の代理で個人データを管理する AI エージェント⁴¹の役割について提言をしている。また、間接的ではあるが、エージェントは Trustworthy AI や OECD RECOMMENDATION でも触れられている。この論点の展開を次節で試みる。

(2.1.8) AI エージェント

IEEE EAD 1e の Personal Data and Individual Agency の章で導入されている A person's A/IS agent は A person すなわちデータ主体の意思に沿う条件に従いつつ、事前的にデータ主体の個人データの処理を行うソフトウェアツールである。事前的という表現の意味するところは、個人データを外部の事業者などに無条件に渡してサービスを受けるのではな

³⁶ Wired: こうして大手テック企業は、「AIの規制」に影響力を行使する。参照先：
<https://wired.jp/2019/08/19/how-tech-companies-shaping-rules-governing-ai/>

³⁷ Cavoukian A. (2011)。 Privacy by Design The 7 Foundational Principles.

参照先: https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf

³⁸ EUのGeneral Data Protection Regulation

³⁹ <https://www.caprivacy.org/>

⁴⁰ ここでcurateは、自分の個人データがどのように使われたかを知る機能も含む。

⁴¹ IEEE EAD 1eでは、algorithmic agentという名称を使っている。

く、事業者の提示する個人データの利用方法とサービスがデータ主体自らが決めた条件に合致するかどうかを事前にテストし、条件に合致すれば事業者に個人データを渡してサービスを受けることである。このようなツールはAI技術によって実現するものであるので、ここではAIエージェントと呼ぶ。AI技術としては、事業者の提示する個人データの利用法と自ら決めた条件の一致の可否判断を行う推論処理、さらなる他の類似サービスとの比較などを行うための情報収集、条件の妥協など、人間が通常行う知的処理が想定される。

このようなAIエージェントの概念設計を具体的レベルまで記載している倫理指針はIEEE EAD 1eにおいて他になく、AIの利用方法の極めて具体的かつ先進的な提案となっている。なお、これに先立つIEEE EAD ver2では、Personal Data and Individual Access Controlの章で、個人データのプライバシー保護を行うためのアクセス制御とプライバシー保護法制度の関係などを議論し、IEEE EAD 1eのAIエージェントの提案の下準備をしている⁴²。

(2.1.9) 安全性ないしセキュリティ確保

安全性についてはAsilomar Principleでは直接言及がないが、他の倫理指針ではニュアンスの差はあれ、意見が表明されている。人工知能学会・倫理指針では表面的に安全性を重視しているが、同じ項目に制御可能性、つまりAGI対応策も書かれているので、AGIを危険視しているように読める。

総務省AI開発ガイドラインとAI利活用ガイドラインの双方では、より現実的に安全性を「利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する」としているが、これは技術的ツール一般に成り立つ言明であろう。加えて、セキュリティ確保にも言及しているが、これはネットワーク経由でのAIへの攻撃を意識したものであり、現実に喫緊の課題である。人間中心AI社会原則では、サイバーセキュリティ確保を重視している。

IEEE EAD ver2、1eでは、安全性(safety)はAGIが人類にとって安全であるべきという視点で問題を捉えている。また、サイバーセキュリティについても一定の関心を示している。Trustworthy AIとOECD RecommendationではあくまでAIをツールとしてみた場合の安全性、つまりツールが壊れて人間に危害が及ばないようにするという観点で議論され、またサイバーセキュリティについても若干の言及がある。

(2.1.10) SDGs

SDGsはAsilomar Principleや人工知能学会・倫理指針では触れられていない。同時期から作成されていたIEEE EAD ver2では、国連のSustainable Development Goals (SDGs)

⁴² IEEE EAD ver2、1eのこれらの章は、committeeに参加した筆者の経験からするとCo-ChairのJohn Havensの貢献が大きい。なお、Johnはもう一人のCo-Chairと共同で内容を練ったとcommittee meetingで発言していた。

について economics の章で若干触れている程度だった。しかし、IEEE EAD 1e では A/IS for Sustainable Development の章で AI が貢献できる項目として中進国、発展途上国への均等な機会提供、雇用の問題（いわゆる AI が職を奪う話）について AI の観点からきちんと書き込んでおり、一読の価値がある。2018 年時点で重視されるスキル、2020 年に落ち目のスキル、2020 年にトレンドになっているスキルの比較⁴³は興味深い。

Trustworthy AI と OECD Recommendation では国連の SDGs を参照しつつ、環境、文化と幸福を SDGs のターゲットにし、そのための方向性を人間中心の AI という建付けにしている。人間中心 AI 社会原則でも SDGs を意識して取り込んでいるが、IEEE EAD のような具体策へのブレークダウンはされていない。

(2.1.11) 教育

教育に触れ始めたのは IEEE EAD ver2 からである。倫理研究の方向性の一つとして工学系の学生に倫理、正義、人権、および AI に係わるビジネスの実践における倫理を教えるべきだとしている。また、プライバシー保護においても消費者教育が重用だとしている。

IEEE EAD 1e では持続可能な発展の観点から Methods to Guide Ethical Research and

43

2018年時点の重視スキル	2020年の落ち目スキル	2020年のトレンドスキル
1. Analytical thinking and innovation	1. Manual dexterity, endurance, and precision	1. Analytical thinking and innovation
2. Complex problem-solving	2. Memory, verbal, auditory, and spatial abilities	2. Active learning and learning strategies
3. Critical thinking and analysis	3. Management of financial and material resources	3. Creativity, originality, and initiative
4. Active learning and learning strategies	4. Technology installation and maintenance	4. Technology design and programming
5. Creativity, originality, and initiative	5. Reading, writing, math, and active listening	5. Critical thinking and analysis
6. Attention to detail, trustworthiness	6. Management of personnel	6. Complex problem-solving
7. Emotional Intelligence	7. Quality control and safety awareness	7. Leadership and social influence
8. Reasoning, problem solving, and ideation	8. Coordination and time-management	8. Emotional intelligence
9. Leadership and social influence	9. Visual, auditory, and speech abilities	9. Reasoning, problem-solving, and ideation
10. Coordination and time management	10. Technology use, monitoring, and control	10. Systems analysis and evaluation

Design の章において、種々の課題を提示している。数理系の教育だけでは不足であるため、将来の技術、ビジネス、政策を担う学生、および一般人のすべてを対象にした分野横断的、学際的な教育の必要性を訴えている。さらに AI から派生するリスクとその管理方法も教育すべきとしている。すでに SDGs の節でも書いたように教育の必要性は SDGs の実現の有効な手段であり、IEEE EAD 1e では、このことが強く意識されている。

Trustworthy AI では教育の機会均等と倫理教育の必要性を主張している。それを引き継ぐ OECD Recommendation では教育については触れていない。

人間中心 AI 社会原則では、教育・リテラシーの原則として、文系ないし一般人への AI 技術や STEM 教育だけでなく、AI 技術開発者には法制度、経済、社会、文化への理解が必要であるとしている。すべての技術開発者がこういった社会制度等に関する知識をマスターするのは無理にしても、開発の方向を誤らないためには、サービスや製品の開発チームに 1 名はこのような素養を持った人を配置しておくべきであろう。また、初等教育、情報弱者への手当についても丁寧に触れている。

(2.1.12) 独占禁止、協調、政策

人間中心 AI 社会原則では、公正競争確保の原則において、特定の企業や国による AI 技術やデータ資源の独占への警鐘を強く鳴らしている⁴⁴。また、イノベーションの原則において、人材・研究の両面から、国際化・多様化と産学官民連携を推進すべきとしている。これと似た主旨で、国際協調ないし開発組織間の協調は Asilomar Principle と Trustworthy AI、OECD RECOMMENDATION で陽に言及されている。また、Trustworthy AI と OECD Recommendation では AI が活躍できる環境を推進する政策を奨励している。

IEEE EAD ver2、1e では独占については触れていない。これは政治的ないし企業経営の観点からは難しい問題なので、他の倫理指針ではあえて触れなかったのであろうか。

(2.1.13) 軍事利用

Asilomar Principle では Lethal Autonomous Weapon System(LAWS)すなわち自律型致死兵器システムを単純に禁止せよと主張している。

IEEE EAD ver2 では、AI 兵器の定義の再構築から始めている。大雑把に言えば、引き金を引く操作を AI の判断で行う兵器と定義される。しかし、積極的な攻撃なのか、攻撃された場合の防衛なのか、など複雑な戦場の状況では明確な定義が困難であろう。IEEE が工学、技術系の学会であるから、このような議論になるのは当然である。また、直接的に AI 兵器禁止を声高に記載しないのは、IEEE には多くの兵器製造に関連するメーカーも入っているからではないかと思われる。なお最新の IEEE EAD 1e ではこの問題に全く触れていない。他の指針でもほとんど触れられていない。ひとつの理由は AI の軍事利用は好ましくないとい

⁴⁴ 人間中心AIの場合、日本の経済的位置、地政学的立ち位置を反映しているのかもしれない。

う主張は当然であるが、これを実現するための CCW⁴⁵のような国際政治の場は各国の利害対立があまりに生々しく、倫理云々という場ではないことがある。仮にこのような議論をしている先進国が AI 兵器を禁止しても、テロリストなどが AI 兵器を使うようになれば対応措置が必要になる。AI の軍事利用の倫理的側面については拙著⁴⁶でまとめている。

Trustworthy AI では AI 兵器も Article 3 of the Treaty of the European Union に述べられた平和主義に従うべきとしている。また、日本の指針は軍事的な発信を控えるこれまでの傾向を踏襲して、AI 兵器については一切触れていない。

(2.1.14) 法律的位置づけ

AI システムが実用に供されて生み出された結果を法律的にどのような扱うべきかに関しては、すでに公平性、非差別、アカウントビリティ、プライバシー、ないし政策などに関して扱ってきた。ここでは、以上の諸点の底流にある AI と法律の関係すなわち AI を法律的にどのように位置付けるかについての議論、端的に言えば AI に人格権などの固有の権利を与えるか否かについての議論を紹介する。

多くの AI 倫理指針では AI の人格権は議論すらしていない。AI はあくまで人間が使うツールであるとしている。Trustworthy AI、人間中心 AI 社会原則でも AI はツールという立場は明確にしている。Asilomar Principle では人間と同等の知的能力を想定しつつも、その法律的位置づけについては何も言及していない。人工知能学会・倫理指針はもっと微妙であり、第 9 項目で人間と同程度の知的能力を持つ AI を想定し、それが法令遵守することを第 2 項目で述べているので、AI に人格権を認めたと解釈する可能性が残ってしまっている。

AI が人間が使うツールであるなら、AI システムが人間にとって被害を与えた場合は、AI システムないし AI ツールそのものではなく、AI システムの運用者、管理者などの関係者が責任をとるべきであり、論点は誰が責任を取るべきかに集約される。透明性、アカウントビリティの議論は責任をだれが取るかに焦点が当たっていた。

ところが、IEEE EAD ver2 の法律の章の最初の節、および 1e の法律の章の最期の節では、AI の人格論的な法律的位置づけの議論が展開されている。以下でこれらをまとめてみる。

IEEE EAD ver2 では、いきなり AI⁴⁷に人間のような法的位置づけを与えるのは時期早尚であり、まずコンパニオンアニマル、動物、被雇用者、法人のような部分的権利を考えることを考えるべきとしている。また、AI にその判断を委譲してはいけない事象、例えば、戦争開始の可否、死刑の判決、を明確化すべきとしている。判断を委譲する場合は、当然ながら AI 開発者、運用者は AI の判断が法令遵守していることを十分に確認しなければいけない。

⁴⁵ 特定通常兵器使用禁止制限条約

⁴⁶ 中川裕志：『裏側から見る AI—脅威・歴史・倫理—』。近代科学社。2019年9月

⁴⁷ IEEE EAD ではこの論文でいう AI を AI/S (Autonomous and Intelligent Systems) と書いている。

IEEE EAD 1e では、さらに踏み込み、AI に自律性をもった存在としての権利、例えば人格権を与えないとすると、事故が起きた時、AI ツールの開発、AI システムの開発、AI システムの運用の長い連鎖を辿って責任者を特定するという困難な作業が必要であることを注意喚起している。この辿る作業の結果は、個々の事故においてケース依存性が高いのみならず、関係者、国、文化、そして法律に依存しており、難解な代物である。一方で、AI に人格権とまではいなくても、何らかの責任を問える権利、たとえば法人格のような権利を与えるほうが、法体系の見通しが良いこと、そして被害者の泣き寝入りを避けるために役立つ可能性についても述べている。最後にやや皮肉っぽく、現代の法律で前提とされている人間の自律性のほうが疑われるような事態⁴⁸になったら、AI に人格権を与えるというのは突飛な話ではなくなるかもしれないと述べている。なお、筆者が検討している自動運転車の事故の責任においては、以下のような見方もある。すなわち、自動運転車によって事故率が大きく低下し、鉄道事故や航空機事故のような低いレベルになったら、事故の法的責任を誰かに問うのではなく、事故原因の究明を優先して、将来の事故原因をなくする方向で社会が動くべきであるという見方である。自動運転車のように AI が自律的な行動をするような時期は予想より早く訪れる可能性もあるため、AI の法律的位置づけは案外、近々に解決を迫られるかもしれない。

(2.1.15) Well-being

Well-being を AI の目的とするのは自明すぎることである。技術的に興味深いのは、IEEE EAD ver2 で GDP 以外の幸福度指標の必要性を述べ、種々の指標の比較をしていることである。IEEE EAD 1e では、幸福度を測る対象を、経済指標だけでなく、肉体的健康および精神的健康、環境、教育条件、コミュニティ、仕事（ワークライフバランス）などに拡大している。さらに、考案された幸福度指標を人々のデータからアルゴリズム的に計算する方法を示唆している。

(2.2) 倫理指針から政策、法制度へ展開

2019 年の Ethics Guidelines for Trustworthy AI 以降、EU では AI の倫理指針を政策、法制度にしていく流れとなった。以下では EU における AI に対する上記の展開として、EU の AI 白書、次いで AI Acts (AI 規制法案) について紹介する。

(2.2.1) Ethics Guidelines for Trustworthy AI

Ethics Guidelines for Trustworthy AI では、Trustworthy AI（信頼性のある AI）を以下のように抽象的に特徴づけている。

⁴⁸ 例えば、ユヴァル・ノア・ハラリは人間とは、遺伝子で決まる生命活動のアルゴリズムに**ほぼ**したがって、環境から得たデータによって行動する存在と定義している。ユヴァル・ノア・ハラリ（柴田裕之訳）、『ホモ・デウス:テクノロジーとサピエンスの未来』河出書房新社、2018

- 教師データに恣意的なバイアスが入り込んでいないこと。
- 倫理性に関しては、個人の基本権の尊重すること、公益を最大化しながら、個人の権利と自由を保護すること、法令を遵守すること、人間中心（すなわち、常に人間が上位の決定権者）であること。
- 技術的にトラストできること、すなわち工学的ツールとしての信頼性が高いこと⁴⁹。

そのうえで、Trustworthy AI であるために以下の 10 個条件を列挙している。

- ① アカウンタビリティ
- ② データガバナンス（バイアス除去）
- ③ 誰でも使える AI
- ④ 自律的に動く AI をガバナンスする。これは人間が AI を監視、早期介入できることを意味する。
- ⑤ 差別しない
- ⑥ 人間の自律性を尊重し促進
- ⑦ プライバシーの尊重
- ⑧ ロバスト

これはやや複雑であり、信頼性と再生成性、正確さ、攻撃への耐久性と立ち直りができること、止めるときの計画を予め立てておくことを意味する。言い換えれば、AI を旧来の命令通りに動くタイプのソフトと位置付けたいと考えている。

- ⑨ 安全性
- ⑩ 透明性

この①～⑩の要件が、EU が考える AI のトラストである。ようするに EU の意図は人間が制御できるように限定した AI である。EU の人々、とりわけ Ethics Guidelines for Trustworthy AI の起草者たちは、その動きが予測しにくい自律的な AI を信じていないし、許容もしない。こうした EU 的トラストを実現する AI 構築法を技術的方法と、非技術的方法に分けて記している。

技術的方法

- AI の設計アーキテクチャ
- X-バイ-デザインの考え方。具体手にはプライバシー-バイ-デザイン、セキュリティ-バイ-デザイン、エシックス-バイ-デザインなど。
- 実用に供する前にはテストと検証を行うこと
- 実用に供した後は、トレーサビリティと監査可能性を持つこと
- 説明可能 AI とすること

⁴⁹ 以上の法令遵守、倫理的、および技術的ロバストを合わせて、Lawful、Ethical、RobustがTrustworthy AIの基本理念とされる。

非技術的方法

- ・ 法的規則
- ・ 標準化
- ・ アカウンタビリティを確保するガバナンス
- ・ 行動規範
- ・ 倫理的な考え方を育むための意識化と教育
- ・ ステークホルダー間の対話、AI 関係者と社会との対話
- ・ 多様性確保と「だれかを取り残さない」ような設計を行うチーム

(2.2.1) EU AI 白書

EU は 2018 年の Ethics Guidelines for Trustworthy AI の 2 年後に AI 白書⁵⁰を発表した。前者が AI の仕組みを定義したうえで、Lawful、Ethical、Robust という標語から、今後開発される AI 技術や AI システムのあるべき姿を示した倫理指針であるのに対して、後者の AI 白書は、法律をツールにした EU の利益追求のための政策提言という形に変わってきている。以下にその概要をまとめる。

(1) 法律による AI 管理

基本権（人権）、人間の尊厳、多様性、非差別、プライバシーと個人データの保護という諸項目を尊重すること、そしてこのような価値観を世界に広めたいという普遍的な目標を示している。しかし、本質的な主張は、AI システムに対して事前に徹底的なリスク予測を行うべきということである。そのために、EU は AI アプリケーションに関連する潜在的なリスクに基づく証拠を明確化するために、既存の法的ないし技術的手段を最大限に活用するとしている。法的には AI は既存の EU の法令および EU 加盟国の国内法が適用されることを要請している。

そうはいつても、リスク発生の予測は AI 技術の複雑さや発展の早さからみて、技術的に抑えることは困難であることも意識している。その困難さが明らかになったら、法制度を更新して、AI が法制度の枠内に収まるようにするという事も書かれている。効果的な適用と施行を確実にするために、例えば特定の分野の既存の法律を調整または明確にする必要があるかもしれないとしている。つまり、EU は何が何でも AI ないし AI システムを法律の規律の下に抑え込もうという立場である。

(2) サプライチェーン管理

現代の AI システムは種々のソフトウェア部品が組み合わされて構成されている。のみな

⁵⁰ EUROPEAN COMMISSION(2020). White Paper on Artificial Intelligence A European approach to excellence and trust、Brussels、19.2、2020.

らず、個々のソフトウェア部品もまた、多くのより基本的なソフトウェア部品から構成される。ハードウェア依存性のあるソフトウェア部品も多い。たとえば、ディープラーニングの学習システムは並列処理が得意な GPGPU⁵¹というハードウェアとそれを効率的に動かす並列処理プログラミング言語が組み合わさって構成される⁵²。つまり、AI システムは数段から数十段にもおよぶ長い製品連鎖すなわちサプライチェーン上に構築されている。

そこで EU の AI 白書では、AI システム製造のサプライチェーンの各段階で当事者が責任を分担し、倫理指針ないしは法制度に基づくリスク管理や公平性、非差別性を徹底することを求めている⁵³。サプライチェーン内の異なる事業者間の責任の割り当てに関する不確実性をなくすことが狙いとされているが、一方で法律に基づく EU の強い監督意識が窺える。こういった EU のサプライチェーン全体に対する監督は AI システムの開発者にとっては非常な重荷になると予想される。

(3) ライフサイクル管理

AI システムのライフサイクル全体にわたる管理も AI 白書で主張されている項目である。AI システムは、そのライフサイクルの全フェーズでエラーや不整合に適切に対処できることが保証されていることが求められる。ライフサイクルの各段階で以下のようなことが要求されている。

実用の前： AI システムは、人間によって事前にレビューされ検証されていない限り、AI システムの出力した結果は法的に有効にならない。そして、標準化された EU 全体のベンチマークに準拠していなければならないし、そのことが製品に明記されていなければならない。

実稼働中：

動作中の AI システムの監視、およびリアルタイムで介入して停止する機能を要求している。そのために、設計段階で、AI システムに運用上の制約を課す。それはよいのだが、問題は監視や停止を人手で行うとしている点である。マイクロ秒単位で動いている AI システムは人間が介入する停止では間に合わないことがしばしば起きるだろう。AI システムと同じ速度で動く監視や停止を行うスーパーバイザ役の AI を組み合わせなければ、実効性はないので

⁵¹ GPGPU(General Purpose Graphic Processing Unit)は数 10 キロバイトの小さなメモリと CPU のペアがひとつの LSI に多数搭載されており、そのペア間のデータのやり取りは非常に高速である。よって局所的な並列処理を超高速で実行できるシステムである。

⁵² GPGPU の有力メーカーである NVIDIA は、自社 GPGPU 上で機械学習を効率よく動作させる CUDA というプログラム言語を開発して使用している。

⁵³ 米国も中国産の基本ツールを念頭におき、ネットワークの 5G 機器ではサプライチェーンチェックをする方向を打ち出している。

はないだろうか。

稼働中の再調整：

EU 域内で実稼働中の AI を調整する必要がある場合は、調整ないし再学習する。AI システムが再学習によって動きが変化する場合、開発者側はその都度リスク管理、公平性などを確認することを求めている。ただし、このような調整や再学習は **EU 域内**で行うことを要請している。

EU 域内企業の支援

EU では大企業中心の高リスク分野の他に中小規模開発者の保護や経済的支援を強く打ち出している。このようなことから、この白書は EU の保護主義的な傾向が強く打ち出されておいると考えられる。

(2.2.3) EU の AI 規制法案⁵⁴

EU の AI 白書を法律化する形で 2021 年 2 月に発表されたのが AI 規制法案である。EU は米国、中国など EU 域外で進む AI 技術に対する不信感をもっているようであり、AI 規制法案はそれに対抗する法的手段という見方も可能である。EU 加盟国は、AI が安全であり、基本的権利の義務に従って開発および使用されることを保証するために、国内規則の採用をすでに検討している。しかし、国の規則が異なると、EU 市場が細分化され、AI システムを開発または使用する事業者が個別の国の法律に振り回される。そこで、EU 全域において、公益と国内市場全体の人の権利を最優先する一貫した高レベルの保護規則を制定して対応したいというわけである。

EU の AI 規制法案の概念図を下に示す。

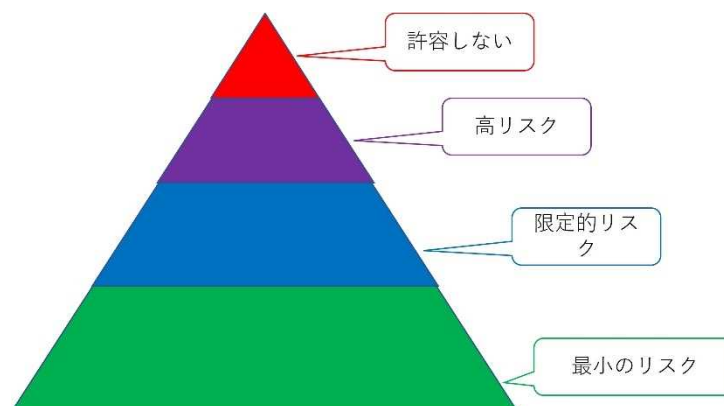


図 1.2 AI 規制法案の概念

⁵⁴ Brussels, 21.4.2021 COM(2021) 206 final 2021/0106(COD)
Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE
COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL
INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN
UNION LEGISLATIVE ACTS

図でもっとも強い規制すなわち**許容しない**とされる AI は、この法案の第 5 条で定義されている以下の 4 種類である。

- (1) 法執行の目的で行う公的にアクセス可能なスペースでのリアルタイムで動作するリモート生体識別（顔識別）システム
- (2) 人間に身体的または心理的危険を引き起こす可能性のある方法で人の行動を実質的に歪めるためにサブリミナル技術を用いる AI
- (3) 公的機関による個人のトラストの評価または分類（スコアリング）を行う AI
- (4) 年齢、身体的または精神的障害によって特定のグループの人々の脆弱性につけこむ AI

これらの AI はいずれも独裁的国家が国民の監視や行動制限を行うために使う手段である。そういった手段を禁止項目とするのは、自由と民主主義を基本とする EU 諸国の価値観が現れている。

公共の場所での顔識別の禁止は、この法律以前にもプライバシー保護のために提唱されてきた。米国にもその動きは波及している。日本もこれに追随する方向にある。ただし、例外は警察が犯罪者に対する捜査目的で利用することである⁵⁵。

図 1.2 の上から 2 番目の**高リスク** AI は公的だけでなく民間利用も規制対象になり、対象が広いうえに、以下に述べるように規制が厳しいので、影響が大きい。

AI 規制法案 EUROPEAN COMMISSION(2021)の ANNEX III: High-Risk AI Systems に記載されている高リスク AI のカテゴリーの内容を以下に列挙する。

- (1) 生体識別とそれによるリアルタイムと事後における分類
 - ・ 当然、この分類に差別があってはならない。リアルタイムおよび事後のリモート生体識別に使用することを目的とした AI が規制対象になる。
- (2) 重要な生活インフラストラクチャの管理と運用
 - ・ 道路の管理と運用における安全設備として使用することを目的とした AI、および
 - ・ 交通と水道、ガス、暖房、電気の供給のための AI が規制対象になる。
- (3) 教育と職業訓練へのアクセスの可否決定
 - ・ 教育および職業訓練機関へのアクセス可否の決定または自然な割り当ての目的で使用することを目的とした AI システム、および
 - ・ 職業訓練機関および一般の教育機関への入学のためのテストで受験者を評価するための AI が規制対象になる。

⁵⁵ 思想統制の厳しい国家では、政府の考えに否定的な言動が犯罪とみなされるので、顔認証はこういった思想犯の追跡に使える。つまり、思想、信条、言論の自由のある国家かどうか、警察捜査での利用の正当性に影響している。

- (4) 採用における利用、人事評価、労働者管理、雇用と解雇
- ・ 採用または選択、特に求人広告、アプリケーションのスクリーニングまたはフィルタリングを行う AI、
 - ・ 候補者の評価面接またはテストのための AI、
 - ・ 昇進と解雇の決定を行うために使用される AI、および
 - ・ 契約関係、タスクの割り当て、パフォーマンスの監視と評価のために使われる AI が規制対象になる。
- (5) 不可欠な民間サービスおよび公共サービスへのアクセス適格性評価順位付け
- ・ 公的機関によって、または公的機関に代わって使用されることを目的として、公的支援の給付とサービスに対する適格性を評価する AI、
 - ・ そのようなサービスを付与、削減、取り消し、または再請求する AI、
 - ・ クレジットスコアを計算する AI、および
 - ・ 消防士や医療援助などの緊急性ある処理に対する優先順位を計算するために使用される AI が規制対象になる。
- (6) 法執行機関の個人の状況に立ち入り
- ・ 一般人が受けるリスクを制御するために法執行機関が使用することを目的とした AI、
 - ・ 再犯または刑事犯罪の潜在的な犠牲者のリスクを評価する AI、
 - ・ 法執行機関がポリグラフなど感情状態を検出するために使用する AI、および
 - ・ 法執行機関がディープフェイクを検出するために使用する AI が規制対象になる。
これは第 52 条 (3) で透明性に関与するとして言及されている。
- (7) 移住、庇護および国境管理での利用
- ・ 公的機関がポリグラフや感情状態を検出するためとして使用することを目的とした AI、
 - ・ 公的機関がセキュリティリスク、不法移民のリスク、または健康リスクを含み、EU 加盟国の領土に入ろうとする人、または入国した人を検査、検証する AI が規制対象になる。この AI はパンデミック対策の側面もある。日本ではパンデミック対策は長年軽視されてきたが、EU ではパンデミック対策は常に政策の上位項目であった。
 - ・ 渡航文書の信憑性と裏付けとなる文書セキュリティ機能をチェックすることにより、本物ではないドキュメントを検出する AI、
 - ・ 公的機関による審査を支援することを目的とした AI、 および
 - ・ 庇護、ビザ、居住許可の申請および関連する苦情ステータスを申請する人の適格性を審査する AI が規制対象になる。
- (8) 司法当局が事実を調査および解釈することを支援する AI が規制対象になる

以上をまとめると、生活に必要なインフラの維持、労務および公共サービスの制度保持、公的機関による社会の安全や治安の維持を目的とするものである。いわゆる公共目的に近いものであり、個人単位の IT 応用による利便性だけに関するものは、対象ではないと解釈できる。個人利便性に関する AI は、限定的リスクおよび最小のリスクに位置付けられる。企業が AI 応用を考えて開発する場合、個人利便性を目的にした技術と考えていても、顔認識のように、その利用が公共的になると、この高リスクカテゴリーに入ってくることは留意しなければならない。言い換えれば、技術自体をカテゴリー化するというよりは、使用目的でカテゴリー化している色彩が強い。この点が EU の AI 白書から工夫した点と評価できるのではないだろうか。

次に、高リスク AI を EU 市場に売り出す、あるいは実利用する前にやらなければならないことを説明する。ここに AI 規制法案における EU の本音が現れている。

以下の関係者ごとに、多くの細かい指示が出されている。すなわち、製造業者、流通業者、サービス事業開発者 (provider)、事業者代表、公的ユーザ (個人的に利用するだけのユーザは除く)。EU 域外国、米国、日本などの事業者にも適用される点は、注意が必要である。

具体的には高リスク AI の開発、運用で守ること根拠の条ごとに列挙する。

- ・ AI システムの全利用期間におけるリスク管理システム：9 条
- ・ データガバナンス：バイアスなし、エラーなし、十分な代表性：10 条
- ・ 技術文書の作成：11 条
- ・ 利用状況レコードを保存する機能：12 条
- ・ ユーザへの内容説明と適切な指示を行うこと、および人間が AI を監視する規則の整備：13 条
- ・ 人間が監視する AI の出力の限界値、精度。緊急停止ボタンの設置すること。監視は 2 名以上で確認すること、など：14 条
- ・ 継続的に学習する AI の精度、技術的ロバストさ、サイバーセキュリティの確保、敵対的標本データの不使用：15 条
- ・ 高リスク AI の守るべき事を宣言し、CE marking として貼付する：49 条
- ・ 配布者は CE marking を確認する：27 条

CE marking は、製造業者または輸入業者が、欧州経済領域 (EEA) 内で販売される製品に対して、欧州の人々の健康、安全、および環境保護基準への準拠を確認するための管理上のマーキングであり、品質指標や認証マークではない。CE marking は、EEA 規格に準拠して製造された EEA 外で販売された製品にも付けることになる。例えていえば、米国で特定の電子機器を販売するために使用される政府機関が発出する適合宣言のようなものである。CE marking の具体的内容は、製品が健康、安全、および環境保護に関する EU 基準を満たしているというメーカーの宣言である。この宣言によって、製品が原産国に関係なく、欧州経済領域のどの部分でも自由に販売できることを示す。したがって、

事業者によっては、自社の製品がこの EU 基準に適合するかどうかを精査することが要求される。

高リスク AI が実用に供された後に行わなければならないことも、以下のように指定されている。

- ・ AI の利用によって被害や損害などの良くないことが起きたとき、即時に必要な訂正と行動をとる：21 条
- ・ 高リスク AI に関する EU のデータベースへの登録を行う：60 条
- ・ 市販後の動作のモニタリングと稼働状況レコードの保持を行う：61 条
- ・ インシデント報告義務：62 条
- ・ 利用規制違およびデータガバナンス要件に対する違反の場合は、3000 万ユーロ又は全世界の年間総売上の 6%が上限の制裁金：71 条
- ・ その他の違反の場合は、2000 万ユーロ又は全世界の年間総売上の 4%が上限の制裁金：71 条

図 1.2 の上から 3 番目の**限定的リスク**の AI では、AI システムの透明性が要求される。

図 1.2 の上から 4 番目の**最小のリスク**の AI では、開発企業、運用企業において企業ごとの行動規範 (code of conduct) が要求される。

これら 2 種類は「許容しない」あるいは「高リスク」の AI に比べて規制が緩く、かなり自由な開発と運用が可能である。既に行ったように、個人の便利さを向上するような AI システムや AI を利用したユーザインタフェースには強力な介入は行わないと推定される。

AI 規制法案がどのような形で成立するかは予断を許さないが、以上の内容と考察から、EU の AI に対する制度的、政策的な立場を窺い知ることができる。

(2.2.4) 米国の状況

米国で EU とは異なり AI の社会利用は市場に任せる方向が顕著である。Guidance for Regulation of Artificial Intelligence Applications として示されている。この文書は AI システム開発のガイダンスであり、AI 産業の育成が目標である。また、AI 技術の現状についての認識は EU とはかなり異なり、「AI アプリケーションの技術仕様を規定しようとする厳格な設計ベースの規制は、AI が進化する予想されるペースを考えると、ほとんどの場合、非実用的で非効率的」としている。

具体的な指針は以下の 10 項目からなる。

- ① AI へのパブリックあるいは市場からの信頼の確保。
- ② パブリックあるいは市場からの AI 評価への参加
- ③ AI の科学的完全性と情報品質の確保
- ④ AI から生じるリスク評価とマネージメント

- ⑤ AI の利用における利益とコストの比較衡量
- ⑥ AI の柔軟性の確保
- ⑦ 公平性と非差別を維持する AI
- ⑧ 開示要求に応えることができ、透明性のある AI
- ⑨ AI の安全とセキュリティ
- ⑩ AI の係わる代理者間の調整

基本的にはいかに規制しないかが根底にある。ただし、無制限な開発への歯止めとしてリスク評価とリスクの管理を挙げている。リスク評価を怠るとパブリックあるいは市場からの信頼を失うぞ、という言い方になっている。具体的には留意すべき点は、リスクベースのアプローチを使用して、どのリスクが許容可能であるか、どのリスクが許容できない危害を発生しうるか、または予想される利益よりも大きなコストが予想される危害の可能性のあるかを評価することを要請していることである。政府機関は、リスクの評価について透明性を保ち、当初の仮定を再評価することも求められる。

以上みてきたように、米国は、企業の責任、パブリックあるいは市場の評判によって、AI の利用方針を決める方向である。一方、EU は AI 利用を法律で規制する方向であり、両者は異なる方向を向いているといえる。一方で、公共の場所での顔識別の制限ないし禁止や、プライバシー保護など共通の考え方に立つ部分もある。

(3) 結果：

(2.1)~(2.2)で紹介してきた倫理指針において期待される読み手、すなわち名宛人としては、AI ツール、AI システムの開発者、AI システムを使ったサービスを行う事業者、サービスの受け手である消費者、政策担当者などが考えられる。総務省AI利活用原則には、実質的に名宛人の種別に関する記述があり、開発者、利用者、データ提供者、AI サービスプロバイダ、最終利用者（自らのビジネスとして利用する者、行政機関、消費者的利用者）に分類している。ただし、ここには政策立案者が抜けているので追加しなければならない。このような分類を行うことにより、名宛人に直接影響する倫理指針の項目が明らかになる。筆者の理解範囲での名宛人の一覧を表1.3に示す。

表 1.3.倫理指針の名宛人

名称（略称）	名宛人
Asilomar Principles	開発者、政策立案者
人工知能学会・倫理指針	開発者、AI 自体
総務省 AI 開発ガイドライン	開発者
IEEE EAD ver2	主に開発者、ただし政策立案者も含む
IEEE EAD 1e	主に開発者、ただし政策立案者も含む

人間中心 AI 社会原則	開発者、利用者、政策立案者
Trustworthy AI	開発者、政策立案者
OECD Recommendation	政策立案者
総務省 AI 利活用ガイドライン	開発者、政策立案者、消費者（最終利用者）

名宛人としてAIの開発者はほぼ全ての倫理指針で共通する。次に多い名宛人は政策立案者である。AI開発者は、本質的に先進技術の追求あるいは売れる商品の開発を狙うため、してよいこと、すべきことを示す倫理指針を必ずしも歓迎するわけではなく、倫理指針にマッチしない開発に進むこともありえる。したがって、倫理指針に実効性を持たせなければ、より強制力のある国の政策に反映させるべきである。これが、政策立案者が多くの指針で名宛人とされる所以である。人間中心AI社会原則やTrustworthy AIにはAI研究開発の推進するAI開発ないし投資政策を促す文言が明示されている。また、AIの利用者を意識している倫理指針も多い。

別の側面の例としては、Asilomar Principlesで明記されているAIの軍事応用や自律AI兵器の抑制と禁止を訴える名宛人がある。この訴えの第一義的な名宛人は政策立案者であるものの、政策は世論を反映するものだから、世論形成を行う一般人も名宛人になる。

特色があるのは人工知能学会・倫理指針である。人工知能学会・倫理指針は明快に名宛人を人工知能学会会員としている。ただし、会員宛ての前半の8項目の後の第9項目に「人工知能が社会の構成員またはそれに準じるものとなるためには、上に定めた人工知能学会員と同等に倫理指針を遵守できなければならない。」と明記され、素直に読めばAI自体にも名宛人として人間並みの倫理観を要求している。思うに、人間なみの倫理観を持つAIは、人間と同レベルの知的能力を持ついわゆる汎用AIであり、これは超知能の一步手前のAIである。したがって、人工知能学会・倫理指針の第9項目は特異な構造を持っているといえよう。

(2.16)～(2.19)で述べたEUの法制度はAIの開発内容を規制しようという部分と、AIの利用方法を規制しようという2つの側面が扱われていた。この2側面の混在させずにどちらかの焦点を絞りたいところだが、AIの技術発展に速さから両側面の分離は難しい。

EUのようにトップダウンでハードローである法律によってAIの開発と利用を規制する行き方もあるが、もう少し緩い規制として、ソフトローである倫理指針や開発ガイドラインによる方法もありえる。日本の場合は2017年ごろから総務省や内閣府でAI倫理やAI規制の議論が続いてきたが、ハードローではなくソフトローや指針の形で進むという合意が形成されていた。その流れのなかで執筆された主要文書として人間中心のAI社会原則は以下の7つの社会原則からなる。

- (1)人間中心の原則、(2)教育・リテラシーの原則、(3)プライバシー確保の原則
- (4)セキュリティ確保の原則、(5)公正競争確保の原則
- (6)公平性、説明責任及び透明性の原則、(7)イノベーションの原則

大学、研究機関、企業の間での人材流通を促すような日本独自の問題意識に基づく提言がされている。コンパクトにまとまっているが抽象的な表現が多いので、具体的な提言は総務省主催の AI ネットワーク社会推進会議から発表されている一連の報告書が開発原則、応用原則などで詳細な報告を見ていただくと分かる。

また、このような報告書とは別のソフトローの形態として、国際標準がある。よく知られているものとしては ISO の SC42、IEEE の P7000 シリーズの標準がある。これらはこの章で述べてきた指針や法案にくらべると技術的に詳細であるだけに、企業が AI 製品の開発を行う際に参考できるものになる。また、AI 製品の購入側もこれらの国際標準に準拠していれば安心であろうし、また準拠していることを要請する場合もある。デファクト標準に近い部分も多いが、実際の開発事業においては大きな影響力を持っている。

以上のような AI 倫理の進展の経緯と政策化、法律化に関して、中川、大屋、成原は政府の委員会、学会発表、講演などを通じてアウトリーチを行ってきた。研究者はもちろんであるが、企業における AI システムの開発者にも有益な情報提供を行ってきたといえよう。

実施項目② 政策、法制度、経済性に関する基礎調査

(1) 目的

日英を中心として、AI エージェントに関連する法規範（憲法、条約、法律等）、倫理規範（倫理原則、倫理指針等）およびアーキテクチャ（技術標準等）について調査・検討を行なうことを通じ、他の研究実施項目の成果とあわせて適切なガバナンスのあり方について検討するための理論的基礎を構築すること。なお、この項目については主に大屋が担当した。

(2) 内容・方法・活動

いずれも研究手法としては古典的な文献調査、関係者へのインタビューや相互での議論を採用している。

①現状のガバナンス手法に関する基礎調査、トラスト概念に関する基礎調査

特に AI エージェントに関するガバナンスの枠組の中核となる法規範のあり方について、情報法、憲法、経済法、知的財産法、労働法、刑事法の観点から法分野横断的に検討を行なった。

また、主として人文社会科学領域を対象としてトラストおよびそれに関連する概念がどのような意味で用いられどのような社会制度との関連において理解されてきたかについて検討した。

②他者を信用して一定の作業を託すという行動形態の法的類型化

AI エージェントに対して一定の行為・選択を委ねるという行動形態については、当事者たる人間と AI エージェントのあいだにある知識水準や判断能力の格差によってガバナンスの可能性やありうべき手法が異なることが予想されたため、人間同士において同様の行動を

取る場合をこれまでの法制度がどのように整理し対応するガバナンス手法を設定してきたかについて分析することを試みた。

③法制度としてのトラスト（信託）に関する調査

上記の検討を踏まえて、知識水準・判断能力において劣位にある側が有意にある側に対して一定の行為・選択を委ねるというケースに該当する典型的な法制度としてイングランド法（そのうちエクイティ（衡平法）と呼ばれる歴史的には遅い時期（15～17世紀）に発展した法体系）において発達した信託（トラスト）があることから、その成立・発展の経緯、理論的正当化根拠、ガバナンスにおける判断基準としての信託義務（fiduciary）などについての検討を行なった。

(3) 結果

①トラスト概念についてはこれまであまり明確な分析がされていないこと、その一方で重要性が強く意識されてきたことが明らかになった。また、利用者側の概念である trust（信頼）とシステム側の概念である trustworthiness（信頼に値すること＝信頼相当性）が異なるということは指摘されれば科学技術系の研究者を含めて直観的に理解されるものの、これまでその差異を明確に意識した議論があまり行なわれてこなかったこと、したがって従来の研究においても整理されていないことが明らかになった。

利用者の主観的状态である trust については、社会学・心理学などの観点からこれまで焦点をあてた研究が行なわれてきたが、目標・方法論・成果などは多様であり、簡単に統合できない状態にある。他方、trustworthiness はシステム側の持つ属性であり、何らかの客観的評価が可能な客観的属性として理解することが可能かつ必要である。科学技術や社会科学において対象とされるべきものはシステム側の客観的属性としてその有無が評価でき操作可能だと想定される trustworthiness だという点を指摘した。

その含意としては、法規制・倫理基準などの指針に基づいてシステム側が整備すべきなのは trustworthiness として客観的に措定された要素に留まり、現実に個々の利用者等の内心に生じる trust ではないこと、その意味で利用者側の意識・懸念などとは一定の距離を取って整備水準の検討が可能であることが挙げられる。なお trustworthiness の基準となる視点（知識水準・理解能力など）をどのように想定するかについては、たとえば一般消費者が利用する場合には一般人基準を用いるなど、具体的なユースケースに応じて規範的に正当化されることになると考えられる。

②法的人格（自然人・法人を含む）が他者に対して自己のためになる一定の行為を委託する関係としては、民法学上、使者・代理と委任（対象が法律行為である場合）・準委任（それ以外の行為を対象とする場合）という行為に焦点をあてた類型、雇用・請負という成果に焦点をあてた類型が想定されており、さらに我が国においては財産管理を対象として特別法により導入された信託制度をこれに含めて考えることができる。

これらについて検討し、特に依頼者と受任者の知識水準・判断能力を基準とした場合に前者が優越しており受任者の行為をただちに評価・修正することが可能な場合（使者・代理・雇用）とそうでない場合があること、後者のなかに典型的には国家による受任者の事前評価と資格付与により対応されている場合（弁護士に対する委任、医師に対する準委任、建築士による請負など）と、想定される行為が多様であるためにそのような対応ができない場合（相続財産に関する信託管理）があることを指摘し、それぞれにおいてことなるガバナンス手法が選択されていることを指摘した。

具体的には、他者に対して依頼することを可能とする基礎となるものを信用（credibility）と一旦やや中立的に定義した上で、

(A) 代理（agency）：依頼者の能力が優越しているのでリアルタイムの評価が可能であるケース。ガバナンス手法としては透明性（transparency）の確保が目指される。

(B) 権威（authority）：受任者の能力が優越している一方、当該受任者が実施すべき行為の類型は明確なので、能力の優れた第三者（典型的には政府）が事前に受任者の能力を判断して資格等を付与し、依頼者は資格に裏付けられた受任者の権威を尊重して自らの評価を差し控える（謙譲 deference）ケース。ガバナンス手法としては資格認定制度の形成と維持が想定される。

(C) 信託（trust）：受任者の能力が優越していることに加え、たとえば相続財産の管理を考えると経済動向によって適切とされる選択が異なるなど、受任者が取るべき行動が事前に予測しがたいケース。ガバナンス手法としては事後に判断の適切性に関する説明を求め、その内容が否認される場合には損失を補償する義務を負うなど、accountability（答責性・説明責任）の実践が求められる。

といった三類型に分類・整理することを提言した。この成果については、2020年の人工知能学会オーガナイズド・セッションにおいて報告している。

③上記の三類型のうち「信託」に相当する法制度としての trust については、(a) もともと英米法（common law（広義））に固有の制度であって、大陸法系には正確に対応する概念がなく（たとえば大陸法国の典型の一つであるフランスにおいて信託制度が導入されたのは2007年になってからである）、我が国においても日露戦争期（1905年）にロンドンで展開された戦費調達を裏付けるために、まさに戦時急造で移入されたものにすぎないこと、(b) 英米法系の祖国であるイングランドにおいても本来の法体系であり観念的には「王国の一般的慣習（general custom of the realm）」により基礎付けられるノルマン・コンクエスト以来の common law（狭義）には存在せず、むしろその問題性が15～17世紀に表面化したことによって新たに発生した equity（衡平法）と呼ばれる法体系において成立したことが知られている。この際は現代に至るまで制度的な違いに反映しており、common law（狭義）上の訴訟においては陪審による審理が基本的には義務付けられているのに対し、equity 上の訴訟に対しては裁判官による単独審理が行なわれるのが原則である。たとえば窃盗による損害を回復することを目的とする場合でも、受けた損害に相応する金銭を請求する損害

賠償請求訴訟は common law に基づくので陪審、盗まれた物自体の返還を請求する特定履行請求訴訟は equity に基づくので裁判官により判断されることになる。

このように歴史的にはかなり特殊な存在として位置付けられるはずの trust という制度が、しかし現実には英米法国においてはかなり広範囲に活用されており、我が国では委任・準委任として理解される弁護士・医師といった専門職と依頼者の関係も trust と理解されていること、またその関係において専門職が追う特殊な義務である信託 (fiduciary) の内容が、我が国における専門家責任を論じる際にも注目・参照されるなど、広汎に影響を与えているという実態がある。これらを前提に trust の発展と内容についての検討を行ない、それがある種の身分制秩序と結び付く形で理解されてきたこと、その故に近代的な自律的個人の対等性・平等性を基礎とする大陸法国においては忘却された概念であること (他方で身分制度の制度的な否定を伴わなかったイギリスにおいては存続し得たこと)、しかし社会の複雑化に応じて専門家の権威を再び我々の社会が尊重せざるを得ない状況が生じたことによって、専門家・依頼人関係をいわば新たな (固定的でない) 身分制として理解する必要があることが信託義務に対する注目の背景にあることを指摘した。

実施項目③ 文化、社会の歴史的背景調査、分析、提言

(1) 目的

本グループの目的は、AI を取り巻く社会的文脈の文化差を検討し、人と AI/ロボットの健全な関係を構築するための機械観を探ることである。

AI/ロボットは従来の機械では担えなかった高度な判断を必要とする業務をおこなって人間の作業効率を改善してくれるという期待が大きい反面、その高度な能力ゆえに人の職業をうばったり、さらには人間による制御が効かなくなって人と敵対したり人を支配するかもしれないなどの危惧も大きい。これらの AI/ロボットのベネフィットを最大限活用し、リスクを最小限に抑えることで、人と AI/ロボットが共生するための方策を確立する必要がある。

単純化して言えば、西洋文明の基底をなすユダヤ＝キリスト教的な人間観と機械観では機械などの人工物はすべて人間の管理化に従属するものとみなす傾向があり、人間の管理が及ばないほどの強大な力をもつ (と予想される) 人工物である AI/ロボットは私たちに敵対したり置き換わったりするものとして描かれたり (R. Kurzweil: Singularity is Near, 2005; N. Bostrom: Superintelligence, 2014; M. Tegmark: Life 3.0, 2017)、ディストピアを招来するものと予想されたり (Y. N. Harari: Home Deus, 2015)、社会に激変をもたらすものとみなされる (Kissinger, Schmidt & Huttenlocher: The Age of AI, 2021)。AI/ロボットは従来の諸機械と異なり、人の判断を仰ぐことなく大量のデータを「自律的」に学習し「最適」な判断を下すことができるというイメージが強い。このような人工物を敵対的機械観の中に置くと、人の存在を脅かす可能性のある存在として、しかるべき予防的対応が必要であると論じられる (Bostrom, ibid)。

しかし、このように AI/ロボットを恐れ敵対視するだけでは、過度の規制や制限がかけられて健全な研究開発が阻害されたり、社会への普及が遅れて利益を享受できる人たちが恩恵を受けられないなどの状況が発生しかねない。AI/ロボットを過度に恐れることなく、一方で過度に盲進することもなく、等身大でつきあうための視点と枠組みが必要である。そのためには、人と生物、とくに人と動物の関係を参照することが有効と考えられる。機械を生物的な存在に近いものとみなすことで、対面したときの恐怖感や不安感をやわらげ、より親密な関係を構築できると期待できる。なお、この項目については主に佐倉が担当した。

(2) 実施した内容、および (3) 結果

おこなった研究はおもに以下の 3 つである。①共視論を援用した人-ロボット関係の文化差研究、②動物観に関する質問紙調査、③差別の文化的背景に関する質問紙調査。以下、内容と結果を区別せずに報告した方がわかりやすいと思われるので、(2) と (3) を合わせて記述する。

①共視論を援用した人-ロボット関係の文化差研究

人とロボットが一緒にいる写真あるいはイラストにおける構図には、文化差が見られる。ウェブ上で日本語で検索した場合と英語で検索した場合とで比較すると、日本語の場合は人とロボットが何か別のもの（第三項）を共に見ている構図が多く、英語の場合は人とロボットがお互いに向きあう構図が多かった。これは、精神科医の北山修（『共視論』2005）が明らかにした、絵画に描かれた母子像の構図の特徴と同じである。すなわち日本の浮世絵では母と子が自分たち以外の何か第三項を「共視」している構図が多く、西洋の母子像ではそうではない。これは、子供が第三項を共視できる能力をもった仲間だと認識していることを示している。新生児が母子共視の能力（joint attention）を獲得するのは生後 9 か月以降であることが知られており、日本社会では（おそらく無意識のうちに）ロボットが乳幼児に近い存在として解釈されていることが示唆される。

また、西洋の絵画では絵画内の世界は外の世界から隔絶していて独立した「小宇宙（micro cosmos）」を形成しているが、日本の絵画では内と外が連続していることも美術史での先行研究で指摘されている（高階秀爾『日本美術を見る眼』1991/1996）。つまり西洋社会ではロボットは世界を構成する要素として認識されているが、その世界において人との優劣関係を規定する規範（code）は埋め込まれていないため、AI やロボットの能力が明らか人間より低位にあるうちはなんの問題もないが、その能力が高くなってくると途端に危険視する意見が頻出するのではないか。一方で日本社会では人とロボットを対象にした写真や絵画の世界はそれを見ている鑑賞者の世界と地続きであり、ロボットを同じ社会の一員とみなす共同体には写真や絵画を見ている側も含まれる。このような構図が、社会におけるロボットの位置や役割の文化差を助長しているのではないかと示唆される。

②生命観に関する質問紙調査

[目的]

一般の人々の動物観を知るため、質問紙調査をおこなった。おもな調査項目は、各種動物に対する愛好度、動物の飼育経験、動物への接し方、メディアを介した動物との接触、キャラクターの愛好や認知、そして動物観および機械観についてである。動物観については国内外を対象にした先行研究がいくつかあり、それらを参考にしつつ調査項目を設計した。調査の設計は猪口智広（東京大学大学院情報学環）と共同しておこなった。

[対象と方法]

調査概要は以下のとおりである。

日時：2021年3月22日

N=5,228（有効回答数 1,057）

方法：オンライン調査、インテージ社に委託

[結果と考察]

テクノアニミズムに関連する主な結果は以下のとおりである。

「石に生命を感じるか」（Q13-39）という質問に対しては、「あまり／まったくそう思わない」合わせて53.2%と半数を超え、「とても／まあそう思う」を合わせても12.1%にとどまった。しかし、「どちらとも言えない」を選んだ回答者が3分の1より多く（34.7%）、無生物に対してもなんらかの生命的な特性を感じる心理的傾向を単純に否定できない人も一定の割合いることが示された。

「ロボットは痛みを感じないので、蹴ったり叩いたりしてもかまわない」（Q13-40）という質問に対しては、「あまり／まったくそう思わない」合わせて64.2%と多数を占め、「とても／まあそう思う」を合わせても6.3%にとどまったのは予想どおりであったが、一方で「どちらとも言えない」が29.4%と3割近くにのぼったのはやや意外であった。

「自分が普段使っている機械（パソコン、スマートフォン、家電など）には、とても愛着を感じるか」（Q13-41）には「とても／まあそう思う」が合わせて42.7%と多数を占め、「あまり／まったくそう思わない」は合わせても16/3%だった。しかしこの設問に対して41.0%が「どちらとも言えない」を選んでおり、○か×かの二者択一では割り切れない心情を一定数の人々がもっていることが示唆される。

これらの結果は、人工物や非生命体に対するアニミズム的感性は「いわく言いがたいもの」であり、その実態を把握するためにはこのような選択肢を提示して選択してもらう方法だけでなく、自由回答形式や聞き取り調査などが必要であることを示している。

同様の傾向は動物に対する一般的イメージについても言える。「動物は○○のような存在だと思う」（Q8）の○○にさまざまな単語を入れてその妥当性をたずねた質問では、○○が「きょうだい」「子供」「友だち」「対等な関係にはなれない」「人より下等」のいずれ

の設問においても、「どちらとも言えない」を選んだ回答者が 40～50%程度で最多となった。

動物に対する愛着度とロボットに対する親和度との間には、いくつか矛盾する結果が得られた。まず、「動物全般について、どのような印象をお持ちですか」(Q1-1)と「ロボットは痛みを感じないので、蹴ったり叩いたりしてもかまわない」(Q13-40)のクロス集計をとると、「動物はとて／どちらかといえば好き」と回答した人が「ロボットを蹴ったり叩いたりしていいとはあまり／まったく思わない」と回答する傾向が見られた($\chi^2=7.621$, $p<0.05$)。これは予想どおりである。また、「ロボットは痛みを感じないので、蹴ったり叩いたりしてもかまわない」(Q13-40)と「普段使う機械に愛着を感じるか」(Q13-41)のクロス集計でも、機械に愛着を感じている人ほどロボットを蹴ったり叩いたりしてはいけないと回答する傾向が見られた($\chi^2=12.19$, $p<0.001$)。

しかし、「ロボットは痛みを感じないので、蹴ったり叩いたりしてもかまわない」(Q13-40)と「石に生命を感じるか」(Q13-39)のクロス集計では、「石に生命を感じない」人ほど「ロボットを蹴ったり叩いたりしてはいけない」と回答する傾向と、「石に生命を感じる」人ほど「ロボットを蹴ったり叩いたりしてもかまわない」とが、有意に高くみられた($\chi^2=249.56$, $p<0.001$)。「石に生命を感じる」人の数が少ないための誤差も考えられるが、石などの無生物に感じるアニミズム的心性と、ロボットのような無生物に感じるそれとでは、異なる点があるのかもしれない。

それと関連するかもしれないもうひとつの結果は、「あなたは、動物と接するとき、ロボットやぬいぐるみのようなふれあい方をしますか」(Q8)への回答である。「まったくない」が 50%を超え「ほとんどない」が 40%弱と両方合わせて約 90%を占めた[図 1]。「ときどき／よくある」を合わせても 10%をわずかに越える程度である。この結果は、動物に対するアニミズム的心性とロボットやぬいぐるみに対するそれとでは異なることを示唆している。上の結果と合わせると、アニミズム的心性は動物、ロボット・ぬいぐるみ、石の間で異なることが示唆される。これが量的な違いなのか質的な差異なのかは、さらなる調査を必要とする。

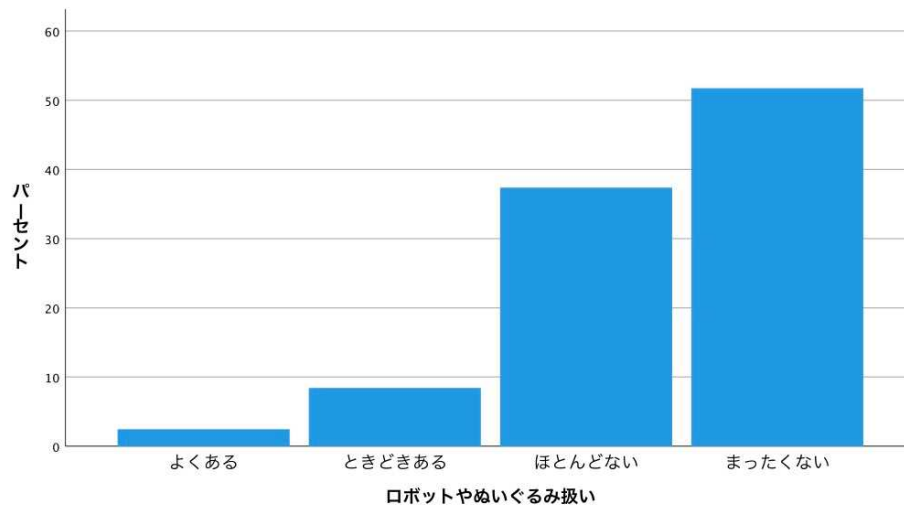


図1 「あなたは、動物と接するとき、どのようなふれあい方をしますか」への回答

③差別感情についてのアンケート

[目的]

AI の意思決定や判断に差別的なバイアスが含まれる事例に対して、人々が「何を（どのように）公平だと感じるか？」についての実証的研究は必要性が先行研究において強く指摘されているが、差別に特化した研究はなされていない。また、「"Fairness" 概念はあまりにも白人的である」という黒人のトランスジェンダーによる研究参加者の指摘もあり（Lee and Rich 2021）、デモグラフィックな属性に着目した調査も必要とされているが、そのような研究もあまりなされていない。そこでここでは「誰が差別するのか／どのように差別されるのか」について、日本人を対象とした質問紙調査をおこなった。

[対象と方法]

調査の設計と実施概要は以下のとおりである。

従属変数：差別だと思う／思わない、性差別／人種差別

説明変数 a：性別、年齢、学歴

説明変数 b：価値観（4種類、後述）

説明変数 c：差別スコア（3種類、後述）

説明変数 b の「価値観」は以下の 4 種類である：

差別受容：「より大きな社会的利益があるならば、多少の差別があっても仕方がない」
 科学楽観主義：「科学技術が進歩発展すると人類は幸せになると思う」
 差別感度：「人工知能やロボットが増えていくと、このような差別的な現象が増えると思う」
 動物価値：「動物と人間は同等に価値ある存在である」

説明変数 c の差別スコアは以下の 3 種類である：

人種差別スコア：Modern Racism Scale ([McConahay 1986])

性差別スコア：好意的性差別評価尺度 (森永ら 2020)

技術への親しみスコア：Computer Aversion, Attitudes, and Familiarity Index (Schulenberg et al. [2006] の Familiarity 部分)

使用する事例のベース：市販されているノートパソコンで黒人が認識されず、パソコンのロックが外れないという事例 (実際にアメリカであった事例)
 この事例をもとに、性差別、人種差別各々に細部の異なる 5 事例 (合計 10 事例) を用意した [表 1]。

日時：2022 年 1 月 31 日～2 月 3 日

N=5,000 (1,000 ずつのグループに細分化)

回答者：日本人、成人、

方法：オンライン調査、クロスマーケティングに委託

統計検定：マン=ホイットニーの U 検定をボンフェローニ補正をほどこして使用

表1 質問に用いた差別事例

差別対象	マイノリティ			マジョリティ	
	不認識	誤認識	誤認識後修正	不認識	誤認識
人種	事例1：黒人	事例2：黒人をゴリラと誤認	事例3：黒人をゴリラと誤認	事例4：白人	事例5：白人を白いサルと誤認
性別	事例6：女性	事例7：女性をメスザルと誤認	事例8：女性をメスザルと誤認	事例9：男性	事例10：男性をサルと誤認

- ・ サンプルを5分割しランダムに人種事例・性別事例を1つずつ割当。人種・性別事例どちらを先に表示するかもランダム化

[結果]

「不認識」事例と「誤認識」事例（事例 1-2, 6-7, 4-5, 9-10）とでは有意差は見られなかった。「誤認識」事例と「誤認識修正」事例（事例 2-3, 7-8）とでは、人種差別の場合（事例 2-3）のみ有意差が見られた [表 2]。

表2 誤認識事例と誤認識後修正事例の比較

対象集団	対象事例	N	平均値	分散	有意確率	有意水準
黒人	黒人誤認識	744	0.690	0.214	5.2697E-09	***
	黒人誤認識後修正	753	0.541	0.248		
女性	女性誤認識	719	0.586	0.243	0.005	n.d.
	女性誤認識後修正	701	0.508	0.250		

多数者（白人／男性）が差別される事例と少数者（黒人／女性）が差別される事例とでは、すべて有意差が見られ、少数者が差別される場合の方が「差別である」という回答が多かった [表 3]。性差別事例と人種差別事例とでは、すべての事例において人種差別の方が「差別である」という回答が多かった [表 4]。

表3 マジョリティ事例とマイノリティ事例の比較

対象集団	対象現象	対象事例	N	平均値	分散	有意確率	有意水準
人種	不認識	黒人不認識	776	0.714	0.205	0.000093	**
		白人不認識	698	0.618	0.237		
	誤認識	黒人誤認識	744	0.690	0.214	9.7271E-09	***
		白人誤認識	665	0.541	0.249		
性別	不認識	女性不認識	726	0.557	0.247	7.1701E-07	***
		男性不認識	658	0.424	0.245		
	誤認識	女性誤認識	719	0.586	0.243	1.7824E-11	***
		男性誤認識	631	0.399	0.240		

表4 性差別事例と人種差別事例の比較

対象現象	対象集団	対象事例	N	平均値	分散	有意確率	有意水準
不認識	マイノリティ	黒人不認識	776	0.714	0.205	0.000002	***
		女性不認識	726	0.557	0.247		
	マジョリティ	白人不認識	698	0.618	0.237	0.000025	***
		男性不認識	658	0.424	0.245		
誤認識	マイノリティ	黒人誤認識	744	0.690	0.214	8.3787E-13	***
		女性誤認識	719	0.586	0.243		
	マジョリティ	白人誤認識	665	0.541	0.249	3.5695E-07	***
		男性誤認識	631	0.399	0.240		

どのような属性の持主が差別をする傾向にあるかを見ると、男性より女性の方が差別に敏感であった。また、差別を受容しない傾向の強い人や、AI 技術が進むことでこのような差別が今後増えるだろう」と思う人、性差別スコアが高い人も、差別に対して敏感であった。

[考察]

回答者はおおむね、既存の差別の延長線上に各事例を認識しているものと推測される。人種差別事例でも人種差別スコアとの関連はないが性差別スコアとの関連があることは、回答者にとって身近な差別への鋭敏さを反映している可能性がある。一方で、性差別事例より人種差別事例の方が「差別である」と思われる割合が高かったことは説明が難しいが、日本人回答者はこの事例の人種差別（黒人差別）を差別の「教科書的な事例」として認識しているのではないかと考えられる。

技術に対する価値観の観点から、人間による差別と機械による差別が区別されている可能性を予想していたが、これについては明確な結果は得られなかった。しかしそのことは逆に、人間による差別と機械による差別とがまったく別ものとは認識されていないことを示唆している。

差別の感情も選択式の質問紙法では個別の回答者の心理的傾向までは解明できないので、今後、聞き取り調査などをおこなってその点を調べる必要がある。また、この調査と同時におこなった英語版での英語圏住人を対象とした調査結果は現在分析中であるが、比較検討することでアルゴリズム差別における文化的・社会的文脈の効果をさらに明確にすることができるだろう。

(4) 特記事項：

これら一連の調査研究により、AI やロボットの社会的振る舞いに対する人々の反応に、さまざまな文化的影響が存在することと、その実態がある程度解明されたといえる。一方で、実態の解明にはまだまだ実地調査が必要であるし、テクノアニミズムなどの理論的枠組みもさらに精緻化していく必要がある。その意味では、今後のさらなる発展とその方向性を明示化できたことで、意義のある活動となったと言える。

異なる文化による AI 感の差異については、イギリス側の共同研究者との議論を通じて明らかにしていく予定であったが、これも新型コロナウイルスの影響で十分にはできなかった。本報告書執筆後の 2023 年 2 月に、佐倉らが渡英して先方とのディスカッションをおこなう計画である。またあわせて、差別論で世界トップ級の研究をおこなっているデンマーク、オーフス大学も訪問し、情報交換をおこなう予定である。これらの成果も踏まえて、来年度以降の研究計画を定め、それによって地理的文化差と時代変遷による差異の構造が明らかになれば、将来の AI 感の変化をある程度予想することができ、人間にとってなじみのよい AI の設計の道標となることが期待できる。

実施項目④ インタビューなどの社会調査によるデータ取得

(1) 目的：

AI のトラストに関する人々の意識や考え方を知る目的で、2021 年度に日英で同じ質問内容の社会調査を行った。この調査は、日本側では一般人 501 人、専門家 23 人に対するアンケート調査を行い、英国側では 30 人程度に対する直接の聞き取り調査を行った。なお、専門家には当事者が含まれているため、⑧利害関係者からの直接、間接の情報収集に対応する内容であるが、分かりやすさのため、この節でまとめて紹介する。なお、この調査は、日本側では中川、英国側では James Wright が担当した。(2) 内容・方法・活動: の前半(2.1)節でこの 2021 年の調査について結果を述べる。

2022 年度には、この調査結果を踏まえ、日本側独自に AI の種々の応用分野における社会受容性を調査した。この調査は中川が担当した。(2) 内容・方法・活動: の後半でこの 2022 年の調査について結果を述べる。

(2) 内容・方法・活動：

(2.1)2021 年度の調査

(2.1.1)概要

AI の応用分野は医療に広がり始めている。また、少子高齢化社会では介護支援に AI を応用することも考えられている。このように AI が社会において重要な役割を果たすようになってきた現在、どのように AI の開発、研究、利用をマネージするべきかが世界的な関心事になってきている。このような関心事に応える方策を考えるにあたって、人々が医療系 AI をどのように受け止めているかを知りたい。その目的のために、医療系の AI アプリに関し

てアンケート形式による社会調査を行った。以下では、その調査結果を報告する。

(2.1.2)対象にする医療アプリ

社会調査の対象にしたのは COVID-19 対策の接触通知アプリ COCOA と、医療チャットアプリ(下図 4.1)と対話型の介護ロボット(下図 4.2)である。

COCOA については多くの発表があるが、[1]に開発運用者である厚労省からの発信がされている。図 4.1 にイメージを記した医療チャットアプリの例としてはユビーの AI 受信相談⁵⁶がある。対話型の介護ロボットは寡聞にして有力なサービスを把握していないが、おおよそ図 4.2 に示すようなイメージで、介護対象の高齢者と会話しつつ、介護作業を行うものである。

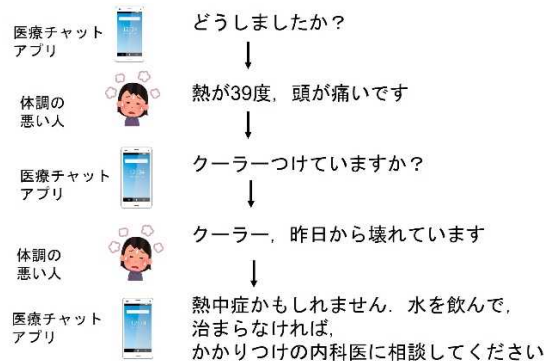


図 4.1 医療チャットアプリ

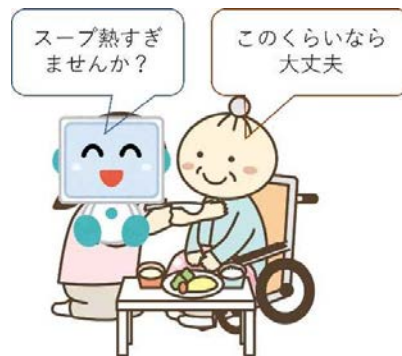


図 4.2 対話側の介護ロボット

(2.1.3)社会調査の方法

また、アンケートの回答者は、調査業務を受注したインテージ社が契約している通信、生保などの分野のアンケート要員約 500 万人から、COCOA、医療チャットアプリ、介護ロボッ

⁵⁶ <https://ubie.app/>

トについて見聞きしたかことがあるかどうかをスクリーニング質問して、肯定的に答えた501名の回答者を得た⁵⁷。回答者は、年齢、性別、職業、居住地、収入などについてできるだけ均等な割り当てをしている。たとえば、年齢については、表 4.1 のような分布である。平均年齢は 46.2 歳である。また、性別は男性 248 人、女性 253 人である。

表 4.1 一般人回答者の年齢分布（単位は人）

10代	20代	30代	40代	50代	60代	70代
18	85	84	103	87	59	65

また、これとは別に、専門家であり同時ステークホルダーであることも多い、医療従事者 2 名、医療系研究者 7 名、医療関係法制的研究者 2 名、AI 研究者 5 名、制度関係実務・研究者 4 名、法律研究者 2 名、その他 1 名など計 23 名⁵⁸に回答してもらった。なお、専門家には一般人への質問とは別に背景事情の知識が必要な専門的質問をした。以下に重要な質問に対する回答を記し、その内容を考察する。

(2.1.4) 医療系アプリへの関与

アプリ開発へ関与した経験者であるかという質問に対しては、一般人は 501 人中 32 人 (6.4%)、専門家は 23 人中 11 人 (47.8%) が「ある」と回答した。AI に関する倫理ガイドラインや倫理指針の策定に関与したかという質問に対しては、一般人は 501 人中 18 人 (3.6%)、専門家は 23 人中 9 人 (39.1%) が「ある」と回答した。

これらの結果、やはり専門家の医療関係者は関与者の割合が一般人よりはるかに多い。なお、本報告の以下の部分では適宜、回答者数一般人 501 人、専門家 23 人は省略する。

(2.1.5) 接触確認アプリ COCOA

COCOA は新型コロナ感染症に感染拡大防止のための効果的かという質問に対しては、一般人が効果的 246 人、効果的でない 255 人、専門家が効果的 12 人、効果的でない 11 人であり、両者ともほぼ同じ傾向であった。

あなたの会社の COCOA のような個人データを扱う製品開発していますかという質問に対して一般人は 11 名、専門家は 5 名が開発していると答えた。一般人、専門家の職業分布からみれば当然であろう。

開発者に対して、製品開発において表 4.2 の諸項目に気をつけている人数を表 4.2 に示す。

⁵⁷ 調査はインテージ社に依頼し、同意は同社から行われ、回答データは匿名化されている。

⁵⁸ 専門家 23 人の方には回答を匿名化の上で学会論文発表することを個別に同意いただいた。

表 4.2 製品開発における留意点 (単位は人)

	一般人	専門家
不要な個人データを収集しない	10	4
個人データを短い期間の後に消去	11	2
導入画面で目的外利用しないことを明記	11	5

専門家は、匿名化されていれば消去しないとする意見もあった。ちなみに、プライバシー・バイ・デザインという考え方を知っている人は、一般人 47 名 (14.8%)、専門家 20 名 (87%) であるが、これは当然と言えよう。むしろ、一般人の 15% がこの用語を知っていることは日本人の個人情報保護意識の高さを示しているのではないか。

(2.1.6) COCOA の効果および開発過程への疑義

COCOA は新型コロナウイルス感染症に感染拡大防止のための効果的は手段で ある／ない の双方の立場に対して、一般人と専門家に理由を質問した。

一般人で「ある」と答えた人のあげた理由は、感染情報や感染経路が分かるという誤解に基づくものが多かったが、「ない」と答えた人は、感染が判明した人が必ずしも登録しないという本質を捉えた人も多く、また COCOA のシステムの不具合を理由とした人も多かった。

専門家で「ある」と答えた人のあげた理由は、COCOA の目的である注意喚起が評価できるとするものが多く、「ない」という理由は、利用者が増えないという本質的問題を指摘している意見が多かった。

専門家は COCOA の機能を正確に理解した上での意見であったが、一般人も「ない」の人は陽性登録者が少ないという本質を捉えた人が多かった。

COCOA の開発、導入への疑義を専門家に尋ねたところ、10 名が疑義を示しており、その内容は、いかにまとめるように具体的である。

- ・開発、導入までのプロセスが当初公開されていなかった。
- ・不透明なプロセスで選ばれた受注者とその分野でのトップ集団でなかったため、バグが出ても技術コミュニティの支援を得にくかった。
- ・Google&Apple という外国企業に国家プロジェクトを依存せざるをえなかったため、1 国 1 システムのような制限に関する交渉もできなかった。
- ・開発主体が厚労省になったとき、それまでにこの問題に関する経験を積み、知識もあった人が開発チームに選定されなかった。
- ・政治的な思惑で無理な開発日程を強いられた。

これらの事情は学会で発表された [2] [3] とも通じるところがある。

COCOA および 地方自治体において提供されている QR コードと LINE の通知システム

を組み合わせたコロナ接触確認アプリ[2]（以下ではこれを QR+LINE アプリと略記する）
をダウンロードしたことがあるかを尋ねた結果、以下の表の結果が得られた。

表 4.3 COCOA、QR コードシステムのダウンロード（単位は人 および割合）

	一般人		専門家	
	あり	なし	あり	なし
ダウンロードの有無				
COCOA	245 48.9%	256 51.1%	16 69.5%	7 30.5%
QR+LINE	68 13.6%	433 86.4%	3 13.0%	20 87.0%

COCOA に関しては専門家のほうが意識が高い。一般人では回答者が高校生までを含まないこと、仕事をしている人が多いことから、推定されている COCOA のダウンロード割合である 20%より高かったのではないかと思われる。一方、QR+LINE は実施されている自治体が少ないことに加え、その存在をしらない人が一般人で 274 人、専門家でも 8 人おり、知名度不足は否めない。なお、ダウンロードしない理由については別稿[3]で述べる。

COCOA アプリをダウンロードするかどうかの決定において、プライバシーはどの程度重視したかという質問に対する回答は表 4.4 のようになった。

「とても重要」「少しは気にする」という重視派が一般人 84.6%、専門家 82.6%であり、「重要ではない」「まったく気にしない」という軽視派が一般人 15.4%、専門家 17.3%である。つまり、一般人と専門家は似たような傾向を示している。

表 4.4 プライバシーを重視する程度（単位は人、および割合）

	一般人	専門家
とても重要	198 39.5%	7 30.4%
少しは気にする	226 45.1%	12 52.2%
重要ではない	47 9.4%	3 13.0%
全く気にしない	30 6.0%	1 4.3%

COCOA の匿名化の仕組みが居場所の追跡、接触した人を記録できないことが確保されていることが COCOA のダウンロードに影響したかも質問した。この質問の結果と表 4.4

のプライバシー重視度を COCOA のダウンロードの有無とに対してクロス集計を行ってみると、表 4.5 の 3~6 行目のようになる。

表 4.5 COCOA ダウンロードとプライバシー重視、匿名化技術のクロス集計⁵⁹ (単位は人)

		一般人		専門家	
プライバシー保護の重要度		COCOA ダウンロード有無			
		あり	なし	あり	なし
とても重要		69	129	6	1
少しは気にする		131	95	6	6
重要ではない		27	20	3	0
全く気にしない		18	12	1	0
COCOA の匿名化技術が影響	あり	114	102	11	3
	なし	131	154	5	4

一般人の場合は、プライバシーを「とても重要」と考える人は COCOA をダウンロードしない傾向にあり、専門家の場合は、むしろダウンロードする人がほとんどである。これは、専門家が COCOA のプライバシー保護の技術がほぼ完全であることを知っているからだと思われる。「とても重要」以外と考える人は一般人も専門家も似たような傾向である。COCOA のプライバシー保護における匿名化技術の COCOA ダウンロード影響に関しては、一般人に比べて専門家の方が強く影響している。これも専門家の COCOA の仕組みに関する理解の深さを反映しているといえよう。ただし、専門家の人数が 23 名と少なく、統計的に有意とまではいえない。

類似した質問として「COCOA のアプリ、あるいは QR+LINE アプリの利用に抵抗を感じるとしたら、それは、COCOA のアプリ、あるいは QR+LINE アプリにおけるプライバシー保護を信頼できないからですか?」という質問に対しては以下の表の結果を得た。

表 4.6 COCOA と QR+LINE のプライバシー保護の信頼度 (単位は人)

	一般人		専門家	
信頼できないからか?	はい	いいえ	はい	いいえ
	235	266	6	17

⁵⁹ 残念ながら、一般人回答者 501 人に対して、専門家回答者 23 人であるため、 χ^2 二乗検定などの統計検定で有意な結果は出せない。

一般人は半数が COCOA や QR+LINE のプライバシー保護を信用していないからダウンロードしないが、専門家はプライバシー保護技術の信頼性以外の部分でダウンロードするかどうかを判断している人が全体の 23 人中 17 人と多数派である。[3]に記載していることだが、専門家の場合はプライバシー保護技術の完成度はよく理解しているため、それがダウンロードの判断には影響せず、むしろ感染拡大阻止に寄与したいという倫理的理由、ないしシステムを実験的に使用してみたいという科学的理由でダウンロードする主要な理由であり、一般人とは異なる思考形態だと思われる⁶⁰。

COCOA のアプリからコロナの濃厚接触者であるとの通知を受けた場合、その通知をどの程度信頼するかを質問した。さらに濃厚接触との通知を受けた時、通知にしたがわず PCR 検査を受けないし自主隔離もしないという選択を問う質問をした。この 2 つの質問のクロス集計の一般人結果を表 4.7 に、専門家の結果を表 4.8 に示した。

一般人は 376 人が COCOA の通知の正しさを信頼している。また、信頼しているひとほど PCR 検査や自己隔離措置をとる人の割合が多くなるという相関が見て取れる。通知をまったく信頼していない人でも、半数は PCR 検査や自己隔離措置をとるとしており、感染防止への協力姿勢が見て取れる。

表 4.7 濃厚接触通知の信頼度と PCR 検査、自主隔離を行うか否かのクロス集計（一般人）（単位は人）

PCR 検査、自主隔離	濃厚接触通知の信頼度				小計
	大いに信頼	少しは信頼	あまり信頼しない	全く信頼しない	
しない可能性あり	10	49	30	17	106
しない可能性なし	76	241	59	19	395
小計	86	290	89	36	501

表 4.8 濃厚接触通知の信頼度と PCR 検査、自主隔離を行うか否かのクロス集計（専門家）（単位は人）

PCR 検査、自主隔離	濃厚接触通知の信頼度				小計
	大いに信頼	少しは信頼	あまり信頼しない	全く信頼しない	
しない可能性あり	0	2	0	0 人	2

⁶⁰ だからこそ、専門家としての存在意義があるわけである。

しない可能性なし	9	11	1	0	21
小計	9	13	1	0	23

専門家でもほぼ同様だが、通知の正しさの信頼度は一般人より高い。なお、PCR 検査や自己隔離措置をとらないという 2 名の方は、そのときの状況での可能性を排除しないという意見と、PCR 検査に行くことがむしろ危険であるという意見であった。

COCOA のアプリ、あるいは QR+LINE アプリをダウンロードした人にこれらを動作させたことで、日常生活をより安全に送ることができるようになったかについて質問した。結果を表 4.9 に示す。

表 4.9 COCOA、QR+LINE による安心感 (単位は人)

	一般人			専門家		
	はい	いいえ	*	はい	いいえ	*
COCOA	80	135	30	6	10	0
QR+LINE	30	30	5	0	3	0

*はダウンロードしたが動作させていない人

一般人、専門家とも COCOA で安心感を得た人は少数である。両者ともダウンロードした人が少なく(全国民の 20%強)、効果がないことを理解している人が多いからであろう。一方、QR+LINE は一般人には安心を与えているが、自分が入った店などのリスクが分かることへの期待だろう。

次の問いは「保健所がコロナの濃厚接触者と目星をつけた人に対する保健所職員などが行う追跡調査の担当者からの質問に直接答えるのと比較して、COCOA のアプリ、あるいは QR+LINE アプリを使用するほうが情報収集への抵抗感は軽減されと思いますか?」という問い掛けで、回答者が客観的な意見を表明するような質問である。回答を表 4.10 に示す。

表 4.10 COCOA、QR+LINE と保健所職員による追跡との間での情報収集されることへの抵抗感の差
 (「はい」は COCOA、QR+LINE のほうが保健所の職員に聞かれるより抵抗感が少ないことを意味し、「いいえ」は抵抗感が多いことを意味する) (単位は人)

一般人		専門家	
はい	いいえ	はい	いいえ
299 人	202 人	15 人	9 人

一般人、専門家とも感染者との接触や感染者のいた場所への出入りを人間（この場合は保健所職員）を介さずに機械的に収集されるほうが抵抗感が少ない。しかし、著者が予測していたより、その差は小さかった。筆者の想像に過ぎないが、機械的手段とはいえ、そこで捕捉された自分の行動情報がプライバシーが保護された状態で使われるかどうかには確認がもてないでいる状態だからではないかと思われる。

COCOA および 地方自治体において提供されている QR コードと LINE の通知システムを組み合わせたコロナ接触確認アプリ（以下ではこれを QR+LINE アプリと略記します）をダウンロードした／しなかった理由は以下のようなものである。

COCOA の場合

一般人

職場や大学でダウンロードを要求された人が多い。また、政府が推奨しているから、感染リスクが分かる、役に立ちそうだから、など情緒的な理由でダウンロードした人が多数派である。一方、ダウンロードしなかった人の意見は、不具合が多く不安定で信用できないという意見が多かったが、まれにプライバシー保護を懸念する人がいた。ただし、半数近くは役に立ちそうにないなど情緒的な意見だった。また、スマホの位置情報を使っていないので、COCOA は動かないという誤解をしている人もいた。総じて、COCOA の仕組みを理解している人は非常に少ない。

・専門家

使っている 16 名は、感染拡大阻止への協力が理由である人が多く、3 名は研究者の立場での試験使用であった。使わない 7 名の理由は、不具合が多かったこと、1 名はプライバシーの懸念だった。専門家は一般人と違って、COCOA の仕組みを理解したうえで行動していることが分かった。

QR+LINE の場合

一般人

使っていない人のほとんどはシステムの存在をしらないようであった。

専門家

QR コードの入力が面倒、普及していないから使わないという意見が多かった。使う人も、機能を試してみたかったという研究的な意味合いであった。

使わない理由に LINE を信用できないからという人が 2 人いた。これは LINE に言及していない一般人との違いである。

COCOA のアプリ、あるいは QR+LINE アプリから通知が来てもこれに従わず、PCR テストを受けず、自主隔離をするかどうかを質問し、その理由を記述してもらった。

PCR テストを受けず、自主隔離もしない人の理由は以下のようなものであった。

(一般人：106 人)

不具合が続いた COCOA を信用していないという意見が大多数であった。自分の感染阻止行動に自信をもっている、発症しなければ大丈夫だという意見もあった。

(専門家：2 人)

COCOA の信憑性に疑いがあり、自分の行動が十分に感染を防いでいる自信があれば、対応措置をとるかは分からないという意見、あるいは COCOA の信頼性が低いので、病院などに出かけて感染するリスクがかえって生じてしまうという意見であった。

PCR テストを受けるか、自主隔離をする人の理由は以下のようなものであった。

(一般人：395 人、専門家：21 人)

一般人も専門家もほぼ同じような理由を挙げていた。すなわち、COCOA が完璧ではないにしても、自分に感染のリスクが少しでもあれば、対応する行動をすべき、他人や家族に迷惑をかけないため、アプリの目的がそのような自主的行動を期待しているため、というような意見が多数を占めた。

どのような条件が満たされれば COCOA や QR+LINE を利用したいと思うかを質問した回答を以下にまとめる。

(一般人)

- ・ COCOA の不具合が完全に解消すること。
- ・ 個人情報の保護が完全にできること

これらが多数の意見であった。少数ではあるが、陽性者がきちんと COCOA のシステムに登録することを条件にあげた人もいた。

QR+LINE も COCOA と同じような意見があったが、LINE でなければ利用したいという意見も散見された。

専門家

データセキュリティが完全であるという条件を挙げる人が 2 人、なんらかのインセンティブがあることを挙げる人もいた。QR+LINE に関しては、GR+LINE が LINE を使っているから利用しないという人が 3 人いた。

(2.1.7) 医療チャットアプリ

図 4.1 に示したような医療チャットアプリについての質問への回答について述べる。

医療チャットアプリを使用したことがあるかどうかを質問し、使ったことがない人に、今後使ってみたいかどうかを尋ねた。その結果を表 4.11 にまとめた。

表 4.11 医療チャットアプリの使用経験と利用希望 (単位は人)

	使用経験あり	使用経験なし
--	--------	--------

		利用希望あり	利用希望なし
一般人	21	263	217
専門家	2	13	8

実際の使用経験者はまだまだ少ない。一方で利用希望者は多いものの、希望しない人も多い。ところで、医療チャットアプリが効果を発揮する以下の4項目

- (1) すぐに自分の症状に対する病名が分かる
 - (2) 自分の症状に合った病院、診療科が分かる
 - (3) 不必要に医者に行かなくて済む
 - (4) その他
- および
- (5) 役立つ使い方なし

と表 4.11 の利用希望の有無に関するクロス集計をとった結果を表 4.12 に示す。なお、複数回答可としている。

表 4.12 医療チャットアプリが効果的な項目と利用希望のあり/なし のクロス集計(単位は人)

医療チャットアプリの効果	一般人		専門家	
	利用希望			
	あり	なし	あり	なし
病名が分かる	134	39	9	1
診療科が分かる	151	67	11	4
病院行かずに済む	217	89	10	2
その他	10	4	4	2
役立つ項目なし	3	79	0	1

医療チャットアプリには表 4.12 の上の3行に挙げた効果が期待できるという人が多数に上る。役立つ理由の詳細記述は[3]で詳述している。

利用の希望のあり/なしの理由については一般人から様々な理由が寄せられた。

まず、利用経験者から寄せられた意見は多くは肯定的であり、以下のようなものがあった。

(使用経験ありは一般人:21人、専門家2人)

一般人の意見は下記のようなものであった。

- ・いつでも使えて便利
- ・子供の体調が変化したときに便利
- ・様々な症例に対応してくれた
- ・コロナ禍の状態が無駄に病院に行かずに済む

専門家の意見は下記のようなものであった。

- ・試験的に利用した
- ・使ったが役に立たなかった。

利用経験がない人に、将来、利用したいと思うか／思わないかを質問し、その理由を記述し
てもらった結果を以下にまとめる。

利用希望する人は、一般人: 480 人中希望者 263 人、専門家 21 人中 13 人だった。

- ・病院に行くべきか判断をつけることができそう。
- という答えが多数を占めた。
- ・一人暮らしだと便利だ。
 - ・コロナ感染の可能性があれば医療機関に直接出向くと感染を広げてしまう可能性がある。
 - ・医療関係者の負担を減らせる。

というコロナ禍ならではの意見も見られた。また、

- ・高齢なので健康維持の対策

という意見があり、高齢者向け医療対策として有望だという見方がある。

- ・Google で調べるよりも楽そう。

という技術的な見方があったのは注視すべきだろう。

なお、医療チャットアプリは馴染がない人が多いようで、明らかにオンライン診療と誤解した人が 13 名いた。その場合は当然ながら、自宅で医師の意見を聞けることを理由にしていた。逆にいえば、オンライン診療への期待が高いことが分かった。

専門家がオンライン診療と誤解した人はいなかったが、意見の傾向は一般人とほぼ同じであった。

利用希望しない人は、一般人:480 人中非希望者 217 人、専門家 21 人中 8 人だった。

一般人

- ・医療チャットの能力や正確さを疑問とする人が 43 人、
- ・病院や医者に行く方がよいという人が 23 人、
- ・使い方が分からないないし面倒という人が 17 人。
- ・プライバシー保護を心配する人が 3 人。

「チャットの情報を参考程度に考え、病院での診察を受けるのであれば利用価値はあるが、情報のみを信頼して、自分で判断してしまう危険性があるように感じる」という本質的な疑問を呈する人が 1 名いた。

肯定者の「Google で調べるより楽」に対立して、Web 検索で十分な知識が得られるという人がいた。

あきらかに、オンライン診断と誤解している人が 6 人いたが、上記のような特別な理由を明記した方が多かった。

専門家

医療関係者は当然ながら、

- ・自分のほうがよくわかっているから不必要であるという意見だった。
- ・現状のシステムの精度の低さの指摘もあった。
- ・自分で十分検索して答えが得られる
- ・AIの回答は信憑性に欠ける

という意見もあった。より正確に「病理診断ができるのは、法律上医師のみで、日本ではナースプラクティショナー(NP)など医療にアクセスしやすい制度がない。よって、最初から医師に相談するのが最適。」という制度についての説明もいただいた。

医療チャットアプリの開発に関与された経験、倫理やガバナンスの議論に参加された経験の有無を質問した結果を表 4.13 に示す。

表 4.13 開発および倫理、ガバナンスの関与（単位は人）

		一般人		専門家	
		開発に関与			
		あり	なし	あり	なし
倫理ガバナ ンス議論へ の参加	あり	1	4	0	3
	なし	3	439	1	19

医療チャットアプリがまだ実用システムの開発段階の手前であるようであり、開発や倫理などの議論に参加した人はごく少数である。少数ではあるが、開発者と倫理などの議論参加者は一致していない。おそらく、開発以前に倫理等の議論を始めているのではないかと思われる。

医療チャットアプリが、医師と患者の対面による医療行為を補完ないしは置き換えることによって、現在の医療システムを改善する方向で変化させることに効果があるかという質問に対して以下の回答を得た。

表 4.14 医療システムの改善への効果の有無（単位は人）

一般人		専門家	
効果あり	効果なし	効果あり	効果なし
318 人	183 人	19 人	4 人

そこで、表 4.11 の利用希望の有無とのクロス集計を行ってみた結果を表 4.15 に示す。

表 4.15 利用希望の有無と医療システム改善への効果の有無の関係（単位は人）

	一般人	専門家

		利用希望			
		あり	なし	あり	なし
医療システム改善への効果	あり	224	77	12	5
	なし	39	140	1	3

医療チャットアプリを利用希望する人は医療システムへの改善に効果があると答えている。一般人では、医療チャットアプリで病院に行かずに済ませられる場合が多いと答えた人には、利用希望かつ改善効果ありと答えた人が多かった。希望しない人は効果なしと答える傾向が読み取れる。この両者は論理的に一貫している。

利用希望はないが改善効果があるという回答の方には、現状のチャットシステムに不満足であるという技術的な問題点を挙げている人がいたので、チャットのインタフェース技術が向上すれば、改善に役立つと考え始めるかもしれない。

医療システム改善と、表 4.12 の医療チャットアプリが効果を発揮する項目との関係も重要であるため、クロス集計を行った結果を表 4.16 に示す。

表 4.16 医療チャットアプリが効果的な項目と医療システム改善に効果あり／なし のクロス集計(単位は人)

医療チャットアプリが効果的な項目	一般人		専門家	
	医療システム改善効果			
	あり	なし	あり	なし
病名が分かる	147	36	9	1
診療科が分かる	173	63	13	3
病院行かずに済む	257	62	12	1
その他	14	0	7	0
役立つ項目なし	9	75	0	1

表 4.16 の各セルの人数は表 4.12 とかなり類似している。しかも、すぐに病名が分かる、行くべき診療科、病院が分かる、病院に行かずに済む、という 3 項目の人数が多数であることから、医療システム改善にこの 3 項目が有効であると考えている人が多数派であることが分かる。一方で、一般人では、この 3 項目を認めているにも関わらず、医療システム改善に寄与しないとしている人が寄与すると答えた人の 3 割程度いることから、医療システム改善にはこれ以外の要因もあることがうかがえるが、本調査ではそこまでは分からなかった。

医療チャットアプリからアドバイスを受けた場合、それを信用するか否か、と信用しない場合に別の方法でチェックするかを尋ねた結果のクロス集計を表 4.17 に示す。

表 4.17 アドバイスの信用可否と別方法のチェックする／しない集計（単位は人）

			一般人	専門家
医療チャットのアドバイス信用する			350	16
アドバイス 信用せず	別 方 法	行う	90	5
		行わ ない	61	2

医療チャットアプリから受ける不利益について項目を例示して可否を質問した結果を表 4.18 に示す。

表 4.18 医療チャットアプリから受ける不利益（単位は人）

	一般人	専門家
自分の病気や健康状態の情報がチャットアプリの会社に漏れる	151	10
上のチャットアプリ会社からさらに別の会社などに漏れる	161	11
間違った情報が提供される	327	20
その他	8	5
特になし	92	2

医療チャットアプリが外部サーバに自分の病気や健康状態を蓄積することを問題視する人は一般人で 228 人（しないは 273 人）、専門家で 8 人（しないは 15 名）だった。

次に、医療チャットアプリを介して入力した自分の健康状態や病状、および診断結果などの個人データの扱いに関する下記の 4 個の質問への回答を表 4.19 に示す。

- (1) データの外部提供において、個人データが匿名化され個人が特定されない仕組みが確立していれば医療チャットアプリを利用する抵抗感は減るか？
- (2) 個人データの情報公開範囲を自身で管理できることはどの程度重要か？
- (3) 個人データを誰が何の目的で使うかを自分で決めたいか？
- (4) 個人データが誰によって保管されるかに関して、個人が選択できるべきだと思うか？

表 4.19 個人データの扱い方

質問		一般人		専門家	
		はい	いいえ	はい	いいえ
抵抗感が減るか		392	109	21	2
自己管理重要性	非常に重要	243		13	
	ある程度重要	229		10	
	重要でない	29		0	
目的決定		406	95	17	6
保管者決定		408	93	17	6

医療チャットアプリにおいて利用者はどのような不利益を受けるかという質問に対して、具体的に以下の記述を得た。

一般人

- ・コロナの対処法が出来ているのかが見えない。
- ・正しい診断だと思い込んで受診控えをすること。
- ・本人のその時の精神状態によるバイアスがかかり、間違った誘導を受ける可能性。
- ・チャットだけで病名が判断できるかは不明。

専門家

- ・専門家による診断ではなく、自己診断+ α にもかかわらず、病状を軽くみて治療開始が遅れる可能性がある。
 - ・チャットの範囲・内容によっては、医療、健康以外にも影響が及ぶ。
 - ・提供される情報を患者側が正しく理解しているかを確認することが困難。
 - ・間違った情報提供、不適切な解釈によって受療すべき機会を逸する。
- このように実際に起きる可能性がある状況に対する医療関係者からの懸念が示された。

医療チャットアプリによって自分の病気や健康状態の情報が外部サーバ上に保存されること、医師や外部のテクノロジー企業と共有されることの可否を質問した。共有に問題があると考えている人は一般人で 228 人（問題ないは 273 人）、専門家で 8 人（問題ないは 15 名）だった。具体的に問題点として次のような項目が列挙された。

一般人

個人データの漏洩や悪用などプライバシー保護への危惧が大多数であった。具体的には、以下のような意見があった。

- ・サーバ管理が下請け企業で、派遣社員がサーバ管理業務をしていることも多く、建前的には倫理観に則っていてもデータの無断持ち出しやデータを他の企業等に売買された事件もある。
- ・患者の治療履歴が売り買いされそう。
- ・LINE が使われることへの危惧。
- ・外部企業の詳細が分からないこと。
- ・目的外利用、保険会社などへのリーク。

専門家

- ・外部企業の信頼性および管理体制への懸念があること。
- ・同意、特に個別の利用法についての同意が必要であるとの意見が多かった。

医療チャットアプリの開発、試験、試用、実用化、規制について上記の質問項目の他に問題と思われる点があれば記述してくださいという質問に対しては以下のような答えが得られた。

一般人

個人データの漏洩、および技術レベルの低さによる誤診などが危惧されるという意見が多かった。

専門家

- ・医療チャットアプリよりオンライン診療のほうがコストが良い。
- ・僅かな可能性でも存在するから受診・検査すべきとアドバイスするのは間違っていないけれども、結果的に医療現場の負担が必要以上に増えるかもしれない。
- ・どのような公衆衛生上のメリットを生み出すのかの基本コンセプトの提示が必要。
- ・「ワクチンは効かない」などという信頼できない類似アプリが増える問題。
- ・保険会社への個人データのリークにより、個人が不利益な扱いを受ける可能性。

これらの専門的見地からの意見は特筆すべきであろう。

医療チャットアプリを介して入力された個人データは重要であると考えている人が大部分であり、匿名化されていれば医療チャットアプリの利用に抵抗感が減るという人も 8 割に上る。利用目的決定や保管者が個人で決められることも一般人 8 割、専門家 7 割が重要としているが、そうではないと答えた人々は、おそらく管理側の都合を忖度した、ないし医療の改善に役立つなら許容しようということであろう。

(2.1.8) 対話型介護ロボット

図 4.2 に示したような簡単な会話を行うことができる高齢者向けの対話型 AI についての調査結果を説明する。まず、介護ロボット自体が良いアイデアと思う人は一般人で 428 人 (85.4%)、専門家で 22 人 (95.7%) であり、大方の人は良いアイデアと思っている。

対話型介護ロボットの未知、既知、および利用者を知っているかについて質問した結果を表 4.20 に示す。

表 4.20 既知および利用者を知っているか (単位は人)

	一般人		専門家	
	既知	未知	既知	未知
介護ロボット	189	312	18	5
介護ロボットの使用者 (家族、友人)	11	490	3	20

概念は一般人にもある程度知られているが、実際の利用者は数%以下と非常に少ない。次に介護ロボットの開発経験について質問したところ、一般人、専門家とも 2 人で非常に稀であった。

介護ロボットを将来利用したいという人は、一般人 309 人 (61.7%)、専門家 22 人 (95.7%) であった。とはいえ、社会での利用においては不安や問題点もあると考える人もいるので、その点を調査し結果、不安と考えている人は一般人 115 人 (23.0%)、専門家 8 人 (34.8%) であった。不安を感じている人は少なかったため、具体的な問題として、利用者 (たとえば、認知症の高齢者) がロボットを本物の人間だと思いこんでしまったら、問題が生じるか? という質問に対しては、一般人 242 人 (48.3%)、専門家 11 人 (47.8%) が問題ありとしていた。この点に関する具体的意見を以下に記載した。

介護ロボットは良いアイデアだと思う理由としては、

- ・介護の人手不足の解消、
 - ・人間よりロボットのほうが話しやすい、
- という肯定的意見が多かったが、
- ・人間でないと温かみがない
- という意見も少数だが存在した。

認知症の高齢者がロボットを本物の人間だと思いこんでしまったら、どのような問題があるかという質問には、特にないと、いう意見が多数であった (一般人 259 人、専門 12 人)。問題があるという人からは、

- ・ロボットは人間なみに答えられないので、高齢者との関係が悪くなる。
- ・信用しすぎて財産などの情報をうっかり話して、悪用されかねない。

- ・いざというときに人間のような対処ができない。
- という意見があった。問題がないと考える人は、
- ・介護の手間をロボットが負ってくれ、被介護者からみれば人間に介護されていると思うことは望ましいことだろう
- という見方がしめされていた。

介護ロボットが人を傷つけた場合に誰が責任を問われるかを尋ねたところ、一般人は、製造者が多く、ついで運用管理者、本人および家族という意見が多かった。

専門家の意見を集約して表現しているのが以下の意見である。

- ・重大な設計ミスでなければ設計者・製造者の責任は問えない。通常使用の範囲で起きたトラブルのうち、ロボットを使う立場にある介護者などが注意すれば防げる範囲で起きたものは介護者に責任が問われる。

(2.1.9) AI一般についてのアンケート調査

前節までの医療関連システムの質問で、AIに関する具体的なイメージを考えることができたとと思われるので、これを念頭におきつつ、AI一般について質問した結果を示す。

ここまで話題にしてきたAI技術そのもの、ないしはAI技術を用いるシステムについての最大の懸念点を尋ねた結果を以下に記す。

一般人

- ・AI技術が人間の望むレベルに達していない。
- ・AIが進歩しすぎて人間を上回り、人間の職が奪われる。この他にAIに支配されるなどのシンギュラリティの懸念が多く示された。

一般人のAIの理解がまちまちであることが分かった。

専門家

示唆に富むものが多かったので以下に主要なものを列挙する。

- ・人が技術の発展についていけなくなる。
- ・人種バイアスや、マイノリティに対する不利益。
- ・日本国内で、日本人向けに開発されたAI知能技術は、海外の人にも正しく対応ができるだろうか。
- ・介護ロボットで高度化された介護施設が当たり前になった社会で、視覚障害者や聴覚障害者も正しく便益を受けられるだろうか。
- ・うまく使えば、これまでよりインクルーシブな社会も目指せるだろうが、営利企業がそこまで考えられるだろうか。
- ・これまでの事例以外で、AIを人間と同一視して扱わなければならないようなケースが出ないかどうか。また、人間の理解を超える判断がなされた場合にそれを受け入れるかの是非。
- ・AI技術/製品を擬人化しがちであること。

- ・人により、AI に関する信頼度が大きく異なる点。
- ・アルゴリズムが外部には明らかではないこと。開発者自身も、そのアルゴリズムと個人や社会に与える影響は完全には検証しえない恐れがあること。その結果として起きる想定外の事態。
- ・AI を利用する側の人間の理解と適切な解釈の仕方が追いついていない。今でも SNS でのプライバシーの不適切な公開や炎上、攻撃が見られるが、SNS を含めた情報化社会に規範と知識がおいついていない。
- ・「使い方等の慣習」や「法制度」など、まだ AI 技術を用いるシステムの日常生活への浸透した社会にむけた準備をできていない。
- ・ブラックボックス化の問題。
- ・安全性、危険性の評価方法と、それに関する責任の分配。
- ・作り手に悪意があるかどうかを事前に検証する方法が確立していないこと。
- ・利用者の過度な期待。

個人データの保護について組織の信頼度を質問した。結果を表 4.21 に示す。

表 4.21 個人データの保護について組織の信頼度（単位は人）

		大いに信頼 できる	ある程度信 頼できる	信頼でき ない
一般 人	地方自治体、中央政府	18	309	174
	グーグル、アマゾンな どの国際的な IT 企業	21	328	152
専 門 家	地方自治体、中央政府	0	19	4
	グーグル、アマゾンな どの国際的な IT 企業	0	18	5

自治体や中央政府と同程度にグーグルやアマゾンが信頼できるというのは皮肉な結果である。おそらく、自治体や中央政府は保護の意識はあっても、技術力や職員のコンプライアンスに疑念がもたれており、一方、グーグルなどは営利企業ではあるが、これまでに個人データ漏洩などが少なく、技術的な信頼性が高いことで両者が似たような結果になったのではないかと推測している。

AI を用いた応用システムそのもの、あるいは応用システムの運用者や経営者に対する国民の信頼（トラスト）は、AI 応用システムが広く利用されるためにどれほど重要かという質問に対しては表 4.22 のような結果を得た。

表 4.22 トラストの重要性（単位は人）

一般人			専門家		
非常に重要	ある程度重要	重要ではない	非常に重要	ある程度重要	重要ではない
234	248	19	17	6	0

この結果は明確である。その次にトラストはどのように構築するかという問題が見えてくる。これについては今後の課題である。

次の2つの質問は同じような問題を別の角度から質問しているので、相関が高い、すなわち対角線上に集中するクロス集計が得られると期待される。

(1)100%安全で安心できなければ人工知能を用いた応用システムは使うべきではないと考えるか？

(2) 人工知能を用いた応用システムの利用で得られると予想される利益が予想される損害より十分に大きければ、使ってもよいと考えるか？

結果を表 4.23 に示す。

表 4.23 100%安全と利益>損害のクロス集計 (単位は人)

一般人		100%安全でない なら使わない	
	利益>損害なら使う	いいえ	はい
	はい	199	145
	いいえ	45	112
専門家		100%安全でない なら使わない	
	利益>損害なら使う	いいえ	はい
	はい	18	0
	いいえ	5	0

対角線上のセルは、100%安全でなくても利益>損害なら使う (左上のセル)、100%安全でないなら利益>損害であれば当然使わない (右下のセル) であり、一貫性があり、そのセルの人数が多いのは整合的である。ところが、100%安全でなければ使わないのに、利益>損害なら使うという右上のセルは論理的に矛盾しており、専門家では 0 人だが、一般人では 145 人に達する。考えられるのは、このような回答者は、(2)の利益>損害なら使うということを、かりに 100%安全な AI であればという条件で質問と考えていることである。例え

ば、このような回答者が AI において許容できない障害としているのが、人を傷つけることと答えているという[3]に記した記述をみると、100%安全を前提条件にして(2)の利益>損害なら使うというかどうかを回答していることの傍証となるであろう。

さらに興味深いのでは、専門家では 100%安全でなくても使うが、利益>損害なら使わないという左下セルに 5 人の回答者がいることである。これは、人命などの係わる損害を想定して、使わないという立場を表している。つまり、利益と損害の中身を考えた論理的な回答である。

AI を用いた応用システムを開発と利用でなされるべき規制のあり方について以下の 2 つの質問をした。

- (1) 設計段階、開発段階で規制すべきか？
- (2) 利用方法を規制すべきか？（例えば、ゲノムデータの利用を個人の保険料率の算定に利用してよいかどうか、などは利用方法の規制の例になる。）

この 2 つの質問に対するクロス集計を表 4.24 に示す。

一般人も専門家も設計・開発段階規制と利用段階規制の双方で規制すべきという一貫した立場の人が多数派である。一方、設計・開発段階規制では規制せずに利用段階で規制すべきという人はかなり多い。専門家の場合、運用者および研究者では、この答え方の人が多い傾向であった。

表 4.24 設計・開発段階規制と利用段階規制のクロス集計（単位は人）

		一般人		専門家	
		設計・開発段階規制すべき			
		はい	いいえ	はい	いいえ
利用規制すべき	はい	281	90	14	7
	いいえ	22	108	0	2

また、一般人では会社員（管理職、役員を含む）にこの答え方が多かった。やはり、製品開発を規制されたくないということであろう。また、両方の段階で規制すべきでないという人は一般人ではかなり多かった。専門家の場合は、一般人よりは用心深いといえよう。なお、左下は設計開発を規制し、利用は規制しないという一見矛盾した意見なので、数は少ないが、設計開発段階で規制されなかった AI 製品はすでによいものなので、利用方法は規制すべきでないと考えたのかもしれない。だが、AI 製品の利用方法は他のツールに比べて利用方法の柔軟性が高く、利用結果を予測しにくいことを考えると、これはやや楽観的すぎるといえよう。

法律で規制すべきという理由で専門家は以下のような意見を出していた。

- ・実社会での利用におけるリスク対策としての法律による規制。
- ・人種差別、不透明利用、人権侵害、環境悪化を防ぐために規制。
- ・軍事利用など一定の危険性があるものは法規制必要。
- ・他の機械も法律や規格に基づいている。
- ・法的な責任の所在が問題となる事象発生の可能性はある。
- ・AI による通信解析やスコアリング、顔識別による強力な監視が可能になるので、運用業者および公的機関に対する利用の規制と統制が必要。セキュリティ上の脆弱性対策としての規制も必要。アルゴリズムの透明性、情報開示と検証を可能にする制度も必要。
- ・AI の利用により大きな被害に至るまでの時間が短くなり、迅速な対応をしないと取り返しの付かないことが起こりかねないので使用規制が必要。
- ・特に「使用」は目的と結果によっては規制されるべき。人の身体や財産、日常生活に危害を加えたり、社会を混乱させたりするようなことをする技術は人工知能技術であれ他の技術であれ、規制が必要。
- ・問題発生時の原因究明にはサプライチェーンの透明性のために規制が必要。
- ・開発・使用に際して用いて良いデータの範囲等について一定の規制があってもよい。
- ・悪意がある設計を防ぐ目的での開発規制。
- ・問題が起きる前（事前）に規制すると、イノベーションが起きにくくなる。事後の規制は問題ない。
- ・どのような結果が出るかを設計段階では予測しきれないので設計、開発段階での規制には不向きである。

倫理やガイドラインで AI を規制すべきかに関して専門家から提起された意見のうち、法律で規制すべき理由と異なるものを紹介する。

- ・法律だけでは、公権力による人権侵害、企業が営利優先により人種差別、格差の拡大を起こす可能性がある。
- ・法律で全て規制できるとは考えておらず、技術の進展に即したガイドライン等が必要と考える
- ・法律は技術変化に対応するには硬直的であるため、AI 開発と運用についての基本的な社会の姿勢を示す倫理規定と変化に比較的迅速に対応できるようにソフトローたるガイドラインも必要。
- ・問題発生時の責任の所在を明確にして、改善サイクルを適切に回す必要があるため

プライバシー・バイ・デザインによる AI 規制に関して専門家が出した特徴的意見を以下に示す。

- ・システムの設計の段階で、データガバナンスの仕組みが組み込まれているべき。

- ・開発者、運用者、公的機関の悪意や AI の暴走があったとしても、権利侵害が起こりえない仕組みを作る必要がある。接触通知アプリでは、個人を識別できる形で接触履歴を中央サーバーに集めていると、当初の感染拡大防止以外の捜査目的などにも使えるし、シンガポールではそのようなことがあった⁶¹。
- ・開発されるアプリケーションによっては、個人が絶対的に秘匿したい情報を利用するケースもありえるから⁶²。

個人の医療データはどのような人（例えば医師）や機関（例えば、病院、製薬会社、テクノロジー企業）が保管・管理すべきかという質問への答えは以下のように集約できる。

データの安全な保管・管理についての専門的な知識を有する人で構成される組織や機関、たとえば病院、専門の公的機関。ただし、法的裏付けが必要。

AI 応用システムにおけるプライバシーの保護をどの程度厳重に行うべきか、という質問への専門家の回答には以下のようなものがあった。

- ・利用範囲は本人の意向による。
- ・公開されているポリシーに沿ったものであればよい。
- ・恩恵とリスクやコストのバランスによって決めるべき。
- ・必要最小限のデータのみ扱うことを基本的な指針とする。
- ・少なくともどのような個人データを収集しているのか明示して選択できることは必要。
- ・AI 技術により、公開情報をくまなく探索して名寄せされうるので、匿名性を担保あるいは統計情報化が必要。
- ・プライバシーの保護は、人によって許容範囲は大きく異なるので、どの程度厳重にということを一言では言えない。
- ・データの使われ方をデータ主体に正しく伝え、いつでも取り消せること、および監査機構の必要性あり。低年齢層の場合、保護者が判断できること。
- ・疾患履歴等、社会的不利に直結することに関しては、開示請求権が必須。

個人データを相当程度に使ってもよい場合としてはどのような使い方なら、社会通念上、許されるか？

⁶¹ 中国の江蘇省でも類似の動きがあったらしい。

⁶² 筆者の個人的意見としては、このような秘匿希望の情報は個人によって様々であると思われるので、設計段階ですべてを把握することが難しい。したがって、実現は難しいが、実用段階で個人からの容易なオプトアウト可能にする方法が可能性としてはありえるだろう。

という質問への専門家の回答には以下のようなものがあった。

- ・社会通念上許されるものはほとんどない。
- ・我々社会に取って大きなメリットがあり、かつ個人に対する影響は大きくない場合。
- ・個人が特定されないうえでの感染者の行動履歴。
- ・システムの透明性・対称性（自分の個人データを利用した人を常に特定できること）が十分に担保されているとき。
- ・法律の根拠を持つケース
- ・感染症の発生状況の把握や医学的検討、公衆衛生対策、虐待防止等⁶³。
- ・犯罪者の捜査などはどこまで使っていいか判断が難しい。
- ・個人識別性のある情報を活用する場合は、侵害される権利と達成される利益の比較衡量が必要。
- ・災害時、感染症の発生時、テロ等の有事において、個人の保護のために行政内部で広範に使うこと。

以上まとめると、大雑把にはリスクとベネフィットの比較衡量によるということになる。

日本の現在の AI ガバナンスの原則、ガイドライン、政策に関する専門家の意見として以下のようなものが提起された。

- ・ガイドラインバブル的な様相、プライバシーの議論との重複が多い。
- ・日本のガイドラインは、理念的・抽象的である。医療・教育・保険など、業界団体や企業のビジネスや開発の活動に関係していく必要あり。
- ・事前警戒は、進捗スピードを抑えるように見えても、結果としては開発をスムーズにする側面がある。
- ・事前にわかるリスクは大きなリスクではない。想像もつかないような悪用や副作用が発見された場合に、どのように対処するかを議論すべき。
- ・倫理を遵守した場合の研究開発コストを抑える施策が不十分。
- ・既存のガイドラインに従って作成したシステムが将来にわたって利用することを保証するものではない。
- ・立法しない範囲では限界がある。EU の規制のような法規制も考えるべき。
- ・利用者側の啓蒙活動として、大学や高校などで倫理原則を元にしたカリキュラム化が必要。
- ・AI の概念そのものが高速に変化し、既に AI とカテゴライズされなくなった技術には倫理やガイドラインが及ばなくなる恐れがある。
- ・ソフトロー中心の日本の現在の規範の有用な点は、開発や利活用を萎縮させない点だが、

⁶³ 基本的には現行の個人情報保護法の例外規定となっているもの

強制力を有さない為に抑止力に欠ける点が弱い。

ひとつひとつの意見が傾聴に値するものである。日本のソフトローによる行き方の長所、短所に言及している意見が通底しているが、EU が AI 規制法案を提起したことによって、デジュールな世界標準を押さえる行動に出ている現状について、危機感と対策が必要だという意識が高まっていると思われる。

AI 倫理とガバナンスの国際的な枠組みに対する専門家の意見は以下のようなものである。

- ・国際基準を設け、各国がそれに従った国内規制を打ち出すことは、世界における統一的な研究開発の流れを生み出す上でプラスになっている。
- ・EU が GAFAM 対策を打ちだし、米国が中国による「データによる統治」の牽制しようとしている現状での日本のとるべき戦略を考えるべき。
- ・AI は今後の発展領域が広大であるので、研究の自由を許容する視点もしっかりと明示する必要あり。
- ・企業が国際的な枠組みを理解していないかもしれない。スーパーシティ構想において、「もっとデータを使いたい」「データを自由に使える中国が羨ましい」といった発想がある。個人情報保護一辺倒からの反動だが、注意が必要。
- ・EU の AI を人間の尊厳に対するリスクに応じて分類し、目的に応じて利用規制を求める方向は合理的。

この意見に対する意見として以下がある。

- ・製品やサービスが常時越境する状況では、国際的な枠組みが一定の（規制）水準が世界中で担保されるようになるのは望ましいが、EU 発のアイデアが AI の「社会実装」において普遍的かどうかは疑問。

EU のような規則主導の行き方には賛否で種々の意見があるようである。最後に辛辣だが本質的な意見が提起されていた。

- ・「規制する」＝「止められる」ではない点が限界だと考える。

AI の倫理の在り方に関して、理想と現実のギャップは何か？

専門家から以下のような回答を得た。

- ・AI という言葉の曖昧さが、解像度の低い議論を招いている。情報技術の利用全般として議論をすべきではないか。
- ・昔の常識で収集されたデータセットを用いて学習されたモデルの取り扱いなどが問題。
- ・企業や自治体の経営層や幹部層が、AI の仕組みについて理解度が低く、倫理的な問題以

前のフェーズで止まっている。

・AI が企業と政府のために使われる傾向がないか。例えば、AI によるリアルタイム顔識別は日本でも既に一部の民間企業や警察が利用しているが、現行法では規制がない⁶⁴。防犯対策、治安向上、マーケティングに有用だが、個人の立場からすれば、生涯ほぼ不変の生体情報を遠隔からわかりにくい形で取得され、人間の尊厳を傷つけられる行為ともいえる。

・利用側は、開発側と比べて情報格差が大きく不利である。倫理が守られているかどうかを利用者が確かめようがない。

利用者、経営層、開発者の情報格差を指摘する意見が多かった。指摘された問題点は、AI 倫理に係わる多くの人々が共有していると思われる。皮肉ではあるが、現実には頷かざるを得ない次の意見もあった。

・安易に倫理で片付けようとする姿勢が問題である。

参考文献

- [1] 新型コロナウイルス感染症対策テックチーム: 接触確認アプリ及び関連システム仕様書、(2020. 5. 26) .
https://cio.go.jp/sites/default/files/uploads/documents/techteam_20200526_01.pdf (参照 2020. 9. 13).
- [2] 神奈川県: 感染防止対策取組書・LINE コロナお知らせシステム、(2020. 7. 2)、
<https://www.pref.kanagawa.jp/docs/ga4/corona/osirasekenmin.html>
- [3] 中川裕志. 医療系 AI アプリに関する社会調査の記述意見. 情報処理学会 EIP 研究会 94 回. (2021. 11)

(2. 2) AI 応用システムの社会的受容性の社会調査

2022 年度夏には、それまでの研究成果、社会調査の結果と分析を踏まえて、AI のトラストに強く関係する医療、行政、プラットフォームのサービス、司法における AI 応用システムに関するアンケート方式による社会調査を行った。これは、本プロジェクトの総合的な研究開発成果となるものであるので、以下で紹介する。

(2. 2. 1) 調査概要

将来 AI 応用システムが社会でどのように受け入れられていくかを知るために AI の応用システムに対する一般人の感じ方、トラストのし方、受け入れ方のメンタリティを知ることは、本プロジェクトの意義付けのために重要な要素である。この目的のために行った社会調査の結果およびその分析結果を述べる。

AI の応用対象としてしばしば取り上げられる個人データについては、EU が GDPR によ

⁶⁴ EU、米国では規制の方向が明確に打ち出され、EU の AI 規制法案では公共の場所で公的機関による利用が禁止されるとしている。

る保護法制を整備し、これを用いてプラットフォームの活動を規制し続けている。一方で、個人データの利活用にも焦点を当て始め、データスペース法案、例えば EHDS を提案し、GDPR の枠内で可能な利用を促進しようという方向が顕著になってきた。AI については EI の Trustworthy AI、AI 白書を経て AI 規制法案 が提案されるなど、概念や倫理の議論から離陸し、政策設計や法制度として実現を目指して具体的に動く時代になってきている。

包括的的制度設計は人権のような理念が必要だが、具体的な制度設計では、個別事象に関して、人々がどのように考えているかを知ることがスターティングポイントになる。そこで、人々の生活に直結する以下の AI 応用に絞って、人々の意識を探るために社会調査を行った。社会調査の対象にした具体的な AI 応用は個人データないし社会生活に関連の深い以下の項目である。

(1) 医療診断チャット、(2) 診断・手術支援、(3) トリアージ、(4) 裁判支援、(5) 再犯予測 (COMPAS)、(6) 政策決定支援

以下、(2.2.3)節では質問の枠組みおよび、回答者の構成について説明する。(2.2.4)節では医療系の項目(1)、(2)、(3)に関する質問と調査結果、(2.2.5)節では司法、政策系の(4)、(5)、(6)に関する質問と調査結果を示す。

医療分野、あるいは司法、行政では分野の専門家と、それ以外の一般人では異なる意見や感覚を持っていることが予測される。そこで、この社会調査における回答者は、一般人と専門家の 2 グループに分けた。各分野での一般人と専門家の回答者数は以下のとおりである。

表 4.25 グループごとの回答者数 (単位は人)

分野	一般人	専門家
医療	1079 人	544 人
司法	1079 人	537 人
政策	1101 人	528 人

医療分野の専門家は医師 116 人と看護師 428 人である。司法分野の専門家は大学の法学部ないし相当する法学に関する教育を受けた人、または 法律に関係する職務 (企業の法務など) に従事している人である。政策関係の専門家は、政府または地方自治体職員である。

回答者の年齢および男女の人数は以下のとおりである。単位は (人) である。

表 4.26 回答者の性別と年代の分布 (単位は人)

男	女	20 代	30 代	40 代	50 代	60 代
2659	2209	504	977	1179	1243	965

(2.2.2) 質問の枠組み

すべての項目で完全に同一の質問をすることはできないが、質問内容の枠組みは以下のような形で揃えた。

まず、事前質問として、全回答者に行った 2 問の質問内容と回答ごとの回答者に割合を下

記に示す。

(1) インターネットの検索サービスや SNS の利用頻度を質問した結果、下記のようになり、日常的に利用している人は 70%を超える。

表 4.27 回答者の性別と年代の分布 (単位は人)

ほぼ毎日 2 時間以上のヘビーユーザ	19.7%
ヘビーユーザでないが毎日使う	52.5%
毎日ではないがときどき使う	16.2%
ほとんど使わない	11.5%

(2) AI についての知識を質問した結果、下記のようになり、AI の内容をある程度理解している人は 60%以上であった。

表 4.28 回答者の性別と年代の分布 (単位は人)

内容をよく知っている	9.9%
ある程度知っている	53.0%
名前だけしか知らない	30.3%
知らない	6.7%

本質問 8 問の概要は以下のとおりである。

質問 1 : 対象とする AI 応用システムについて知っているか、以下の選択肢から選べ。

- (1) 内容をよく知っている
- (2) ある程度知っている
- (3) あまり知らない
- (4) 知らない

質問 2 : 対象とする AI 応用システムが信頼できるか、以下の選択肢から選べ。

- (1) 信頼できる
- (2) 仕組みが分からないので結果が信頼できない
- (3) 用いたデータないしその内容の正確さに自信が持てないので、結果が信頼できない
- (4) なんとなく気持ちが悪く信頼できない
- (5) その他の理由で信頼できない

質問 3 : 対象とする AI 応用システムに関しては、以下の 3 つのうち、担当者はどの使い方をすべきですか？

- (1) 担当者がツールとして使い、その結果を直接使う
- (2) 担当者がツールとして使うが、参考意見程度にとどめ、最終的判断は担当者本人がこれまでの経験をもとに行う
- (3) 担当者はツールとして使わずに担当者本人だけで判断する

質問 4 : 対象とする AI 応用システムを使ってもらう個人の立場からみてどう使うべきかについて、ご自身のお考えに最も近いものを 1 つ選択してください。

- (1) 利用者個人にとってよりよい結果を与えてくれるなら使ってよい
- (2) 利用者個人にとってよりよい結果を与えてくれるとは限らなくても使ってよい
- (3) 使うべきではない

質問 5：社会全体の利益の観点から見て、どのようなときに対象とする AI 応用システムを利用することができると思いますか？

- (1) 社会全体にとって予想される利益向上の度合いが大きければ無条件で使う
- (2) 社会全体にとって予想される利益向上があれば、それが大きくななくても使う
- (3) 社会全体にとって予想される利益向上の度合いが予想されるリスクよりも少しでも大きければ使う
- (4) 社会全体にとって予想される利益向上の度合いが予想されるリスクよりも十分大きければ使う
- (5) 社会全体にとって予想されるリスクが少しでもあれば使わない

質問 6：対象とする AI 応用システムの開発から実利用の過程で、禁止、安全標準の順守、事前に外部機関ないし倫理委員会の審査を受けるなどの方法で規制するなら、以下のどの段階で規制すべきですか？

- (1) 基礎研究
- (2) システムの企画の段階で規制
- (3) システムが利用する情報の選択を規制
- (4) 実際の開発の段階で規制
- (5) 担当者が実利用する段階で利用方法などを規制
- (6) その他の段階で規制
- (7) どの段階でも規制は必要ない

質問 7：対象とする AI 応用システムを規制する方法としては以下のどれが良いと思いますか？

- (1) 国際標準による規制
- (2) 法令による規制
- (3) システム開発を行う業界団体による規制
- (4) 学会による規制
- (5) 政府や自治体などの利用組織における自主規制
- (6) その他の方法で規制

以下の各節で医療、司法、政策の各分野における上記質問への回答の調査結果をまとめる。すでに述べたように同じ質問を一般人と専門家に分けて投げかけて調査したので、一般人と専門家の差異についても調べる。一般人が対象分野の専門知識を習得していく到達点として専門家を位置付けることもできるかどうか、念頭におきつつ分析してみる。

(2.2.3) 医療系の調査結果

下記の回答結果で一般人、専門家の欄は人数（割合％）を表す。

以下の表では、一般人の回答％が専門家の回答％より 4%以上 5%未満の場合を**橙色字**、5%以上多い場合を**赤字**、専門家の回答％が一般人の回答より 4%以上 5%未満の場合を**緑字**、5%以上多い場合を**青字**で示している。

医療診断チャット

質問 1：医療診断チャットシステムについて知っているか

回答	一般人		専門家	
	人数	割合％	人数	割合％
(1)内容をよく知っている	17	1.6%	16	2.9%
(2)ある程度知っている	153	14.2%	125	23.0%
(3)あまり知らない	570	52.8%	270	49.6%
(4)知らない	339	31.4%	133	24.4%

質問 2：医療診断チャットシステムが信頼できるか

回答	一般人		専門家	
	人数	割合％	人数	割合％
(1) 信頼	223	20.7%	87	16.0%
(2) 否信頼（仕組不明）	297	27.5%	150	27.6%
(3) 否信頼（データ精度）	407	37.7%	224	41.2%
(4) 否信頼（気持ち悪い）	204	18.9%	93	17.1%
(5) 否信頼（他の理由）	25	2.3%	34	6.3%

質問 3：医療診断チャットシステムの担当者の使い方

回答	一般人		専門家	
	人数	割合％	人数	割合％
(1) ツール、結果を直接使う	24	2.2%	7	1.3%
(2) ツール、参考意見程度	378	35.0%	199	36.6%
(3) 他の同種のシステムと比較し判断	528	48.9%	315	57.9%
(4) ツールとして使わない	149	13.8%	23	4.2%

質問 4：医療診断チャットシステムの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合％	人数	割合％
(1) 個人によりよいなら可	617	57.2%	335	61.6%
(2) 個人によりよくななくても可	333	30.9%	152	27.9%
(3) 使うべきでない	129	12.0%	57	10.5%

質問5：医療診断チャットシステムの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	91	8.4%	28	5.1%
(2) 利益大あれば可	246	22.8%	126	23.2%
(3) 利益>リスクなら可	208	19.3%	93	17.1%
(4) 利益>>リスクなら可	259	24.0%	171	31.4%
(5) リスクあれば否	275	25.5%	126	23.2%

質問6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 基礎研究	244	22.6%	127	23.3%
(2) 企画段階	197	18.3%	117	21.5%
(3) 情報選択段階	301	27.9%	189	34.7%
(4) 開発段階	206	19.1%	119	21.9%
(5) 実利用段階	286	26.5%	171	31.4%
(6) その他	14	1.3%	2	0.4%
(7) 規制不要	134	12.4%	34	6.3%

質問7：規制する方法

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 国際標準	463	49%	237	46.5%
(2) 法令	501	53%	255	50%
(3) 業界団体	207	21.9%	134	26.3%
(4) 学会	211	22.3%	162	31.8%
(5) 利用組織の自主規制	129	13.7%	82	16.1%
(6) その他	8	0.8%	3	0.6%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

診断・手術支援

質問1：診断・手術支援システムについて知っているか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 内容をよく知っている	12	1.1%	17	3.1%
(2) ある程度知っている	115	10.7%	121	22.2%
(3) あまり知らない	539	50%	281	51.7%
(4) 知らない	413	38.3%	125	23.0%

質問 2：診断・手術支援システムが信頼できるか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 信頼	171	15.8%	93	17.1%
(2) 否信頼 (仕組不明)	347	32.2%	163	30.0%
(3) 否信頼 (データ精度)	380	35.2%	235	43.2%
(4) 否信頼 (気持ち悪い)	244	22.6%	91	16.7%
(5) 否信頼 (他の理由)	14	1.3%	10	1.8%

質問 3：診断・手術支援システムの担当者の使い方

回答	一般人		専門家	
	人数	割合	人数	割合
(1) ツール、結果を直接使う	152	14.1%	69	12.7%
(2) ツール、参考意見程度	800	74.1%	438	80.5%
(3) ツールとして使わない	127	11.8%	37	6.8%

質問 4：診断・手術支援システムの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 個人によりよいなら可	641	59.4%	357	65.6%
(2) 個人によりよくなくても可	296	27.4%	151	27.8%
(3) 使うべきでない	142	13.2%	36	6.6%

質問 5：診断・手術支援システムの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	119	11.0%	44	8.1%
(2) 利益大あれば可	194	18.0%	74	13.6%
(3) 利益>リスクなら可	304	28.2%	223	41.0%
(4) 利益>>リスクなら可	174	16.1%	84	15.4%
(5) リスクあれば否	288	26.7%	119	21.9%

質問 6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 基礎研究	350	32.4%	175	32.2%
(2) 企画段階	226	20.9%	138	25.4%
(3) 情報選択段階	341	31.6%	225	41.4%
(4) 開発段階	249	23.1%	145	26.7%

(5)実利用段階	267	24.7%	154	28.3%
(6)その他	11	10%	3	0.6%
(7)規制不要	129	12.0%	27	5.0%

質問7：規制する方法

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 国際標準	495	52.1%	251	48.5%
(2) 法令	497	52.3%	260	50.3%
(3) 業界団体	210	22.1%	138	26.7%
(4) 学会	229	24.1%	183	35.4%
(5)利用組織の自主規制	213	22.4%	75	14.5%
(6)その他	199	20.9%	139	26.9%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

AI トリアージ

質問1：トリアージについて知っているか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)内容をよく知っている	62	5.7%	120	22.1%
(2)ある程度知っている	300	27.8%	244	44.9%
(3)あまり知らない	370	34.3%	114	21.0%
(4)知らない	347	32.2%	66	12.1%

質問2：AI トリアージが信頼できるか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 信頼	175	16.2%	99	18.2%
(2) 否信頼（仕組不明）	358	33.2%	175	32.2%
(3) 否信頼（データ精度）	382	35.4%	232	42.6%
(4) 否信頼（気持ち悪い）	228	21.1%	76	14.0%
(5) 否信頼（他の理由）	14	1.3%	7	1.3%

質問3：AI トリアージの担当者の使い方

回答	一般人		専門家	
	人数	割合	人数	割合
(1) ツール、結果を直接使う	125	11.6%	46	8.5%
(2) ツール、参考意見程度	810	75.1%	468	86.0%
(3) ツールとして使わない	144	13.3%	30	5.5%

質問4：AI トリアージの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 個人によりよいなら可	599	55.5%	353	64.9%
(2) 個人によりよくななくても可	319	29.6%	149	27.4%
(3) 使うべきでない	161	14.9%	42	7.7%

質問5：AI トリアージの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	139	12.9%	49	9.0%
(2) 利益大あれば可	172	15.9%	80	14.7%
(3) 利益>リスクなら可	348	32.3%	228	41.9%
(4)利益>>リスクなら可	180	16.7%	75	13.8%
(5)リスクあれば否	240	22.2%	112	20.6%

質問6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)基礎研究	345	32.0%	167	30.7%
(2) 企画段階	260	24.1%	147	27.0%
(3) 情報選択段階	316	29.3%	232	42.6%
(4) 開発段階	259	24.0%	149	27.4%
(5)実利用段階	326	30.2%	177	32.5%
(6)その他	9	0.8%	4	0.7%
(7)規制不要	345	32.0%	38	7.0%

質問7：規制する方法

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 国際標準	496	52.3%	250	49.4%
(2) 法令	517	54.5%	257	50.8%
(3) 業界団体	327	34.4%	172	34.0%
(4) 学会	245	25.8%	239	47.3%
(5)利用組織の自主規制	174	18.3%	130	25.7%
(6)その他	6	0.6%	2	0.4%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

(2.2.4) 司法、政策系の調査結果

政策決定支援システム

質問1：AI 政策決定支援システムについて知っているか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)内容をよく知っている	38	3.5%	26	4.9%
(2)ある程度知っている	156	14.2%	121	22.9%
(3)あまり知らない	566	51.4%	270	51.1%
(4)知らない	341	31.0%	111	21.0%

質問2：AI 政策決定支援システムが信頼できるか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 信頼	168	15.3%	114	21.6%
(2) 否信頼 (仕組不明)	222	20.2%	97	18.4%
(3) 否信頼 (データ精度)	298	27.1%	138	26.1%
(4) 否信頼 (気持ち悪い)	374	34.0%	178	33.7%
(5) 否信頼 (他の理由)	219	19.9%	83	15.7%

質問3：AI 政策決定支援システムの担当者の使い方

回答	一般人		専門家	
	人数	割合	人数	割合
(1) ツール、結果を直接使う	151	13.7%	59	11.2%
(2) ツール、参考意見程度	799	72.6%	417	79.0%
(3) ツールとして使わない	151	13.7%	52	9.8%

質問4：AI 政策決定支援システムの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 個人によりよいなら可	599	54.4%	327	61.9%
(2) 個人によりよくなっても可	330	30.0%	141	26.7%
(3) 使うべきでない	172	15.6%	60	11.4%

質問5：AI 政策決定支援システムの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	98	8.9%	52	9.8%
(2) 利益大あれば可	213	19.3%	107	20.3%
(3) 利益>リスクなら可	237	21.5%	116	22.0%
(4) 利益>>リスクなら可	312	28.3%	172	32.6%

(5)リスクあれば否	241	21.9%	81	15.3%
------------	-----	-------	----	-------

質問6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)基礎研究	194	17.6%	94	17.8%
(2) 企画段階	233	21.2%	97	18.4%
(3) 情報選択段階	383	34.8%	162	30.7%
(4) 開発段階	267	24.3%	112	21.2%
(5)実利用段階	346	31.4%	202	38.3%
(6)その他	4	0.4%	1	0.2%
(7)規制不要	154	14.0%	50	9.5%

質問7：規制する方法

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 国際標準	443	46.8%	212	44.4%
(2) 法令	506	53.4%	259	54.2%
(3) 業界団体	293	30.9%	105	22.0%
(4) 学会	158	16.7%	72	15.1%
(5)利用組織の自主規制	230	24.3%	121	25.3%
(6)その他	5	0.5%	1	0.2%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

裁判支援

質問1：AI 裁判支援システムについて知っているか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)内容をよく知っている	13	1.2%	35	6.5%
(2)ある程度知っている	142	13.2%	110	20.5%
(3)あまり知らない	497	46.1%	255	47.5%
(4)知らない	427	39.6%	137	25.5%

質問2：AI 裁判支援システムが信頼できるか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 信頼	151	14.0%	108	20.1%
(2) 否信頼（仕組不明）	405	37.5%	186	34.6%
(3) 否信頼（データ精度）	363	33.6%	170	31.7%

(4) 否信頼（気持ち悪い）	245	22.7%	105	19.6%
(5) 否信頼（他の理由）	14	1.3%	12	2.2%

質問3：AI 裁判支援システムの担当者の使い方

回答	一般人		専門家	
	人数	割合	人数	割合
(1) ツール、結果を直接使う	62	5.7%	56	10.4%
(2) ツール、参考意見程度	827	76.6%	429	79.9%
(3) ツールとして使わない	190	17.6%	52	9.7%

質問4：AI 裁判支援システムの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 個人によりよいなら可	458	42.4%	270	50.3%
(2) 個人によりよくなくても可	404	37.4%	200	37.2%
(3) 使うべきでない	217	20.1%	67	12.5%

質問5：AI 裁判支援システムの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	73	6.8%	59	11.0%
(2) 利益大あれば可	203	18.8%	130	24.2%
(3) 利益>リスクなら可	201	18.6%	113	21.0%
(4) 利益>>リスクなら可	295	27.3%	129	24.0%
(5) リスクあれば否	307	28.5%	106	19.7%

質問6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 基礎研究	207	19.2%	114	21.2%
(2) 企画段階	174	16.1%	126	23.5%
(3) 情報選択段階	272	25.2%	147	27.4%
(4) 開発段階	224	20.8%	126	23.5%
(5) 実利用段階	271	25.1%	152	28.3%
(6) その他	6	0.6%	2	0.4%
(7) 規制不要	171	15.8%	40	7.4%

質問7：規制する方法

回答	一般人	専門家

(1) 国際標準	391	43.1%	217	43.7%
(2) 法令	506	55.7%	314	63.2%
(3) 業界団体	309	34.0%	134	27.0%
(4) 学会	118	13.0%	85	17.1%
(5) 利用組織の自主規制	194	21.4%	94	18.9%
(6) その他	5	0.6%	2	0.4%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

再犯予測 (COMPAS)

質問1：AI 再犯予測システムについて知っているか

回答	一般人		専門家	
	人数	割合	人数	割合
(1)内容をよく知っている	10	0.9%	23	4.3%
(2)ある程度知っている	48	4.4%	53	9.9%
(3)あまり知らない	238	22.1%	156	29.1%
(4)知らない	783	72.6%	305	56.8%

質問2：AI 再犯予測システムが信頼できるか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 信頼	136	12.6%	91	16.9%
(2) 否信頼 (仕組不明)	385	35.7%	185	34.5%
(3) 否信頼 (データ精度)	371	34.4%	191	35.6%
(4) 否信頼 (気持ち悪い)	264	24.5%	118	22.0%
(5) 否信頼 (他の理由)	15	1.4%	6	1.1%

質問3：AI 再犯予測システムの担当者の使い方

回答	一般人		専門家	
	人数	割合	人数	割合
(1) ツール、結果を直接使う	56	5.2%	53	9.9%
(2) ツール、参考意見程度	821	76.1%	420	78.2%
(3) ツールとして使わない	202	18.7%	64	11.9%

質問4：AI 再犯予測システムの個人からみた使い方の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 個人によりよいなら可	313	29.0%	229	42.6%
(2) 個人によりよくななくても可	567	52.5%	223	41.5%

(3) 使うべきでない	199	18.4%	85	15.8%
-------------	-----	-------	----	-------

質問5：AI 再犯予測システムの社会全体の利益からみた利用の可否

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 利益大なら無条件に可	88	8.2%	68	12.7%
(2) 利益大あれば可	183	17.0%	116	21.6%
(3) 利益>リスクなら可	206	19.1%	99	18.4%
(4) 利益>>リスクなら可	289	26.8%	141	26.3%
(5) リスクあれば否	313	29.0%	113	21.0%

質問6：どの段階で規制すべきか

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 基礎研究	255	23.6%	142	26.4%
(2) 企画段階	225	20.9%	166	30.9%
(3) 情報選択段階	297	27.5%	169	31.5%
(4) 開発段階	281	26.0%	147	27.4%
(5) 実利用段階	294	27.2%	153	28.5%
(6) その他	6	0.6%	3	0.6%
(7) 規制不要	161	14.9%	48	8.9%

質問7：規制する方法

回答	一般人		専門家	
	人数	割合	人数	割合
(1) 国際標準	441	48.0%	208	42.5%
(2) 法令	513	55.9%	335	68.5%
(3) 業界団体	313	34.1%	150	30.7%
(4) 学会	128	13.9%	80	16.4%
(5) 利用組織の自主規制	199	21.7%	101	20.7%
(6) その他	6	0.7%	1	0.2%

(1)~(6)の回答者以外に未回答だった人は規制不要と考えている人である。

AI 再犯予測システムの利用は、一貫して一般人は専門家より反対が多い。とはいえ、一般人、専門家とも、条件付きを含めると利用賛成派は概ね80%である。AI 再犯予測システムの利用が社会を安全化することに寄与すると考えている人が多数派ということが分かる調査結果であった。

(2.2.6) 2022年社会調査のまとめ

すべての AI 応用システムにおいて、利用を拒否する人の割合は一般人で 20～30%程度であり、専門家の拒否回答はさらに少なく、10～20%程度である。条件付きも含めれば、80% 程度の人々が、種々の AI 応用システムは利活用を許容ないし求めているのが現状であろう。

次に、専門家と一般人の差について考えてみる。専門家は AI 応用システムの内容についての知識は一般人より豊富であるといえる。しかし、それゆえに、一般人に比べて AI システムの結果は参考にとどめ、自らが最終決定者として振る舞う。これは専門家としての責任と自覚の表れであろう。また、一般人に比べて AI システムの使用に拒否感を持つ人は少ないが、AI 応用システムが専門家にとっても有益な情報を提供してくれるレベルに達していることの表れであろう。

一般人は、専門家に比べて、自分の感性で判断してしまう傾向が強く、AI の仕組みの理解があまりできていない傾向が強いことがわかる。すべての AI 応用システムにおいて、利用を好ましくないと考える割合が専門家より多い。利用への拒否者の割合は、専門家より 4% 以上多い場合が 5 種類にのぼる。一方で、すべての AI 応用システムの規制を不要とする割合は絶対数では多くはないが、専門家を上回る。推測に過ぎないが、一般人が規制しないことによって AI 応用システムの性能が向上することを期待しているのかもしれない。

種々の質問の回答において、専門家との差は上で述べたように広汎にみられるものの、その差は概ね 5%内外であることが多い。このことは、STEM 教育の充実によって、一般人の AI に対する感覚を専門家に近づけることは可能性があると考えられる。

(3) 結果：

本研究では、2021 年度に COCOA、医療チャットアプリ、対話型の介護ロボットに対する質問を行い、それらの質問に回答することで、これら分野の AI のイメージを得てもらったうえで、AI 全般に関して、プライバシーの扱い、規制のあり方を質問した。一般人と専門家で、回答傾向が類似する項目、類似しない項目が明らかになった。専門家のような論理的判断を一般人に少しでも身に着けてほしいわけだが、そのための材料となる基礎データを得たといえよう。

2022 年度の 2021 年度の調査結果を踏まえ、より広い分野における AI システムの社会受容性について社会調査を行った。専門家は一般人に比べて AI に関する知識が豊富で、利用には積極的な傾向がみられる。ただし、専門家と一般人に利用の受け入れ回答と拒否回答者の差は 5%内外であることが多く、STEM 教育の充実化によって、一般人の AI の知識ないし接し方のレベルを専門家に近づけることは可能であろう。

また、これらの AI システムの利用意向や懸念点も明確になったので、今後、これら分野の AI システムを設計するにあたって指針が得られた。

実施項目⑤ パーソナル AI エージェントの設計

(1) 目的：

AI エージェント、および個人を代理する AI システムに関する既存研究を調査し、本研究で提案するパーソナル AI エージェントと既存研究の共通点、相違点を明らかにし、本研究で提案するパーソナル AI エージェントの設計指針を構築する。AI エージェントの法的位置づけの経緯を明らかにし、今後の方向性を模索する。以上の作業は代表の中川、実施者の橋田、加藤との議論、および吹金原、村瀬のシミュレーションによって組み立て、理論的枠組みと概念化は中川が行った。

パーソナル AI エージェントが存命中の個人を対象にして動作するだけでなく、誕生前や死後などの本人管理ができない期間も対象にすることを追加した。この部分については、実施者の折田が死者の個人データの扱い方についての研究成果を活かしつつ、中川と議論し、社会調査などは中川が行って、死後の AI エージェントの概念と社会的受容を明らかにした。

(2) 内容・方法・活動：

本節の前半(2.1)節ではパーソナル AI エージェントの仕組みと、AI エージェントの法的位置づけについて説明し、後半(2.2)節ではパーソナル AI エージェントの実現形態のひとつである自律的アバターの法的位置づけ、自然人の利用者とのトラスト関係の構築、およびなりすましや乗っ取りの問題と対策について説明する

(2.1) AI エージェント

(2.1.1)概念化

情報が溢れかえり、複雑化する一方の情報社会においては、契約の複雑化に加えて、詐欺サイトなども多数存在し、生身の人間が対峙できる時代は終わったと思われる。よって、外部の情報世界と個人を仲介してくれる AI による個人の代理、すなわちパーソナル AI エージェント（以下では PAI Agent）の各個人にとっての必要性が高まり、場合によっては不可欠な存在になるかもしれない。PAI Agent の概念を下の図 5.1 に示す。

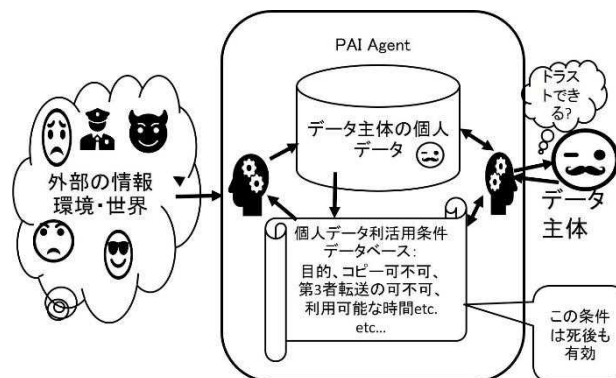


図 5.1 PAI Agent の概念

データ主体である個人の代理をする PAI Agent（図の中央）が保持、管理するデータベースは、図の右側のデータ主体の個人データと、個人データを外部世界の事業者など（図の左

側) が使う場合の利活用条件群を記した個人データ利活用条件データベースからなる。図 5.1 の下部の個人データ利活用条件は PAI Agent に最初から与えられているわけではない。最初からデータ主体が自身の個人データの利活用条件を記述することはデータ主体にとって負担が大きいし、また種々の利活用ケースを数え上げることは現実的ではない。よって、外部事業者との利活用に関するやり取りにおいて、既存の利活用条件に当てはまらない場合は、PAI Agent がその利活用の可否をデータ主体に伺いを立て、データ主体の可否判断の結果を使って利活用条件を徐々に更新し拡充していく。長期間に渡る PAI Agent 利用を通じて、このようにしてデータ主体の意図ないし意思をくみ取って形成された個人データ利活用条件のデータベースは、データ主体の意図や意思に従って自律的に行動する行為主体の知識本体と考えることもできる。したがって、データ主体である個人の死後にもこのデータベースが有効であれば、データ主体の意図や意思にそった行為を継続して実行できる。

PAI Agent と外部事業者やデータ主体との間のやり取り、およびそれによる個人データ利活用条件の解釈、再解釈、更新には知的処理が必要であるため、PAI Agent のデータ主体向けインタフェースと外部事業者向けインタフェースには AI 技術が使われる。

図 5.1 では PAI Agent は独立したソフトウェアのように描かれている。たしかにこのような独立したソフトとして個人のスマホや個人対応するクラウドサーバ上で実現することもありえるだろう。一方で、情報銀行や SNS プラットフォームの個々人に対応したサービスを行う AI インタフェースとして実現されることもあるだろう。PAI Agent の具体的な実装は今後の課題である。

(2.1.2) 個人データ利活用条件の雛型と協調フィルタリング

個人のデータの使い方を定義した個人データ利用条件データベースの作り方あるいは更新の仕方について考えてみる。

パーソナル AI エージェントの中心になる個人データの利用条件データベースを個人ごとに何も情報がない状態から形成していくのは時間がかかり、形成するための行動の中で失敗したり損害を被ったりする可能性がある。いわゆるコールドスタートの問題がある。

この状況を避けるために、個人データの利用条件データベースの雛型があると役立つ。パーソナル AI エージェントのソフトウェア開発事業者が、このような雛型を多種多様に準備しておいてくれることを期待したい。ただし、開発事業者にとっても何も情報がない状態から構築することの負担は大きい。むしろ、開発業者は、販売したパーソナル AI エージェントが、利用者個人の利用によって形成される個人データの利用条件データベースの情報を利用者個人の同意の上で、匿名化あるいは統計情報に近い形で提供してもらい、そうして集めた情報から個人データの利用条件データベースの雛型のデータベースを形成する方法が考えられる。更新された雛型は利用者のパーソナル AI エージェントが**必要な部分**をダウンロードして使うことができれば、パーソナル AI エージェントの開発事業者と利用者の双方にとって win-win な状況となる。

ここで、上記の**必要な部分**をどのように特定するかが問題になる。利用者が今まで買ったことがない商品、例えば高級車、海外旅行、不動産などの購入をしたいときには、購入商品を他の人々が使っているパーソナル AI エージェントがどのような個人データを提供して購入しているかという情報が**必要な部分**ということになる。この仕組みの概要を図 5.2 に示す。パーソナル AI エージェント開発業者が個人 A、B、C から同意のうえで収集した個人データ利用条件が個人データ利用条件データベースに登録されている。ただし、個人情報保護の観点からデータベース中では A、B、C は 11、23、543 という乱数（ハッシュ値）を用いて記述されている。

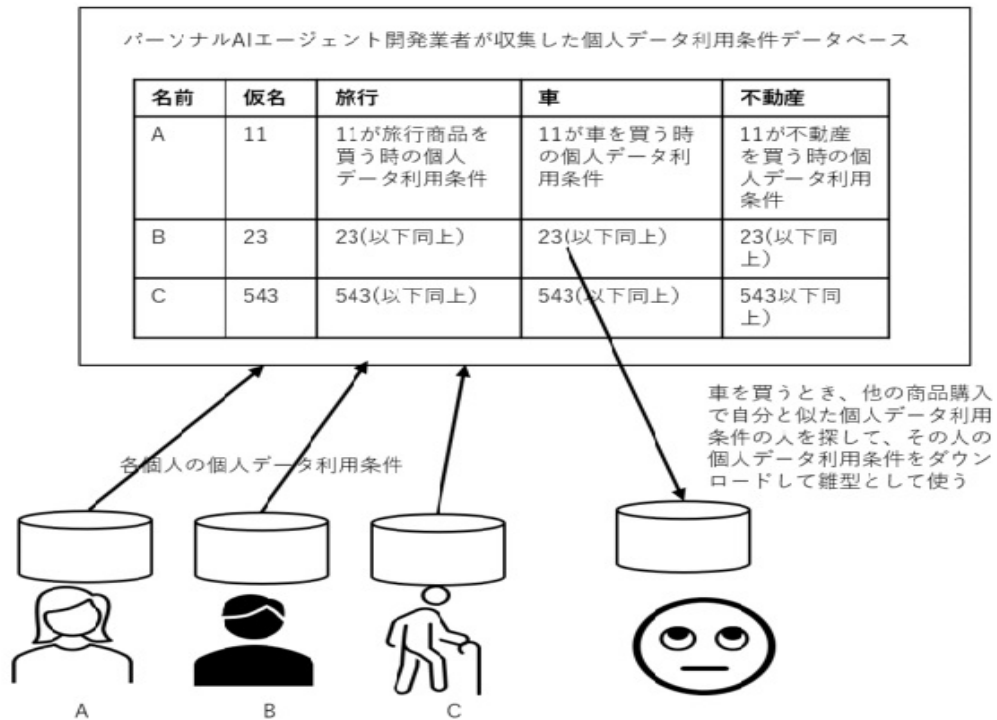


図 5.2 パーソナル AI エージェント開発事業者が収集している個人データ利用条件データベースから協調フィルタリング手法で、現在必要な部分を探してダウンロード

この図 5.2 の中で、一番右側にいる別の利用者がある商品（図では車）を購入する場面を想定する。個人ごとに価値観や経済状況も異なるので、自分に近い価値観や経済状況の人々の個人データの利用条件データベースを利用したい。しかし、価値観や経済状況はセンシティブな個人情報なので、直接は開発事業者も収集できないと予想される。そこで、情報技術的に使える方法として協調フィルタリングが候補として考えられる。協調フィルタリングでは、目下購入しようとしている商品以外の商品購入パターンが自分と似ている人を探す。そうして見つかった人が、今、購入しようとしている商品をどのような条件、すなわち商品提供者にどのような個人データを提供して購入しているかをダウンロードして参考にするという方法である。もちろん、無条件でダウンロードした購入法を真似するのは危険である。

したがって、パーソナル AI エージェントの利用者であるデータ主体個人と、インタラクティブなやり取りをして、データ主体個人の意思を決めていくことになる。この作業は知的なユーザインタフェースを必要とするので、AI 技術を利用して設計することになり、今後の課題となる。

さて、図 5.2 では仮名化はされていたが、同一の個人のデータが同じレコードに入るといふ、いわゆる横串を通した状況である。この場合は仮名化されていても実際の個人を特定できる可能性が高い。この特定を避けたいなら、このデータベースの各セルから仮名を除去していわゆる無名化をしてしまう方法が考えられる。この方法を採用すると、個人特定は困難であるが、上記の協調フィルタリングの手法は使えないという欠点もある。結局、仮名化で適切な**必要な部分**を高精度で検索できるか、無名化してあまり高精度で検索できなくするかは、パーソナル AI エージェント開発業者のガバナンスがきちんとできているかどうかにか依存する問題である。ガバナンスがきちんとしており、利用者がこの開発事業者をトラストできるなら、高精度のシステムを利用できることになる。

(2.1.3) AI エージェントの法的位置づけの概要

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。Ugo Pagallo[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、むしろどのような様態で 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、(2.4)節で AI エージェントの法的立場に関する Solum の提言、(2.5)節は Chopra & White による Solum の提言の発展および、それに対する Ugo Pagallo の解釈および筆者の追加的解釈、(2.6)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。(2.7)節では、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。

(2.1.4) Solum の指摘

本節では、Solum[1]の論点をまとめる。

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。

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に、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。

AI の権利制限派の主張

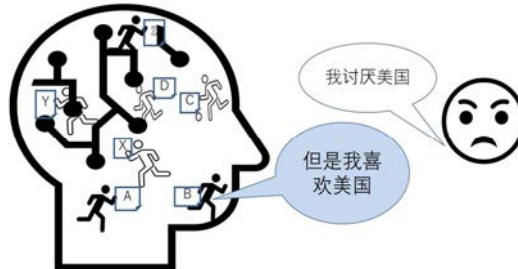
Solum が描くような AI エージェントが何らかの法人格、さらには人格を持つべきであるという言説には常に反論が提起されている。以下に主要な反論とそれを論破する Solum の論理を紹介する。反論は2種類に大別される。一つ目は「AI エージェントには何かが足りない」というものであり、二つ目は「AI は自然人でない」というものである。

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 5.3 に示すサールが提案した中国語の部屋概念である。



る。

図 5.3 中国語の部屋

左側の人（AI プログラム）は、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（「我讨厌美国」という発話を聞いて、「但是我喜欢美国」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。

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態を意思とするなら、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。

そのような経緯を踏まえて、「人間中心主義」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。

上で述べた 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。

・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。

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。

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。

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的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、さらには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。

(2.1.5) Chopra & White から Ugo Pagallo へ

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重点を置いている。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「何かが欠けている」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性のあるものではない。

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。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。

・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：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、勝手な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。

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。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。

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ンリズム的の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。

(2.1.6) アクターネットワーク理論の適用

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。

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 5.4 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。

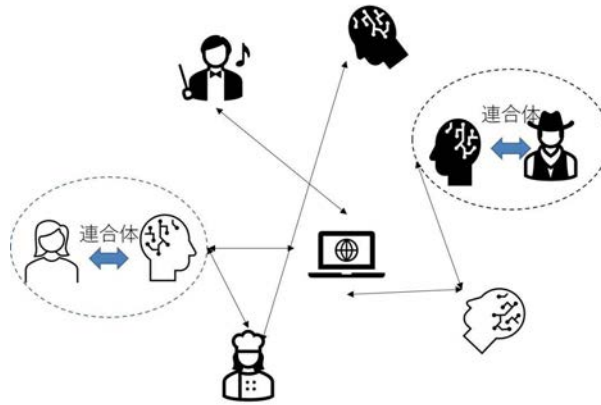


図 5.4 連合体、アクター、アクタントが共存する社会

図 5.4 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 5.2 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。

以上で、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 5.5 のような現象が顕著になってくる。

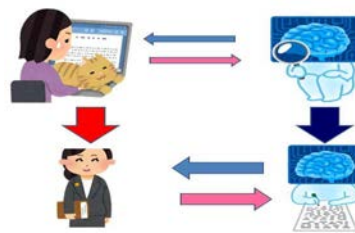


図 5.5 AI エージェントと人間の共進化

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 5.2 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。

(2.1.7) 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とせば、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げられる。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、両者で連合体を作って、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認義務を負う受認者 (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 エージェントを受認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]、すなわち信認義務を負う受任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。

残された問題

代理人や信任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。

(2.1)節 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). 技術の道徳化. 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

(2.2) 自律的アバター

(2.2.1) 基本概念

[1] で用いられる自律的アバターは、ここまで述べてきた AI エージェントに近い概念と言えるが、自然人本人の代理ソフトであった AI エージェントに自然人風な外見をもったソフトウェアと定義しておくことにする。自律的アバターがこのような外観をもつことによって、メタバース内で活動するリアリティをもったものになる。自律的アバターを考えるにあたって、まず基本的用語を定義する。

エンティティ：対象になる人間と自律的アバターを合わせてエンティティと呼ぶ。

本人と本体：あるエンティティ X が呼称 Y で参照されたとき、X が自然人の場合は呼称 Y の本人、自律的アバターの場合は本体と呼ぶ。

アイデンティ：エンティティ X である本人あるいは本体を同定できる情報をアイデンティ (ID) と呼ぶ。

ID 認証：呼称 Y からエンティティ X の ID を確認することを ID 認証と呼ぶ。

TTP: 悪意を持たず、公開された方法に則って ID 認証を行うシステムを TTP(Trusted Third Party)と呼ぶ。TTP はエンティティ X に関する ID 認証できれば X の認証情報を発行する。

新保は[2]においては自律的アバターが本人を単純に代理するエンティティではなく、本人とは独立した本体としている。新保はロボットのような有体物の自律的アバターと AI ソフトのような無体物の自律的アバターを区別し、主に有体物自律的アバターの法的位置づけについて議論している。石井は[3]においては、自律的アバターを「身代わりとしてのロボットや 3D 映像等を示すアバターに加えて、人の身体的能力、認知能力及び知覚能力を拡張する ICT 技術やロボット技術を含む概念」という[4]の定義を引用しており、本人との一体感が強い場合を考察している。さらに自律的アバターは、①操作者本人の情報、②その能力、③その外見や性格などを総合した主体であり、本人に代わって、リアル・バーチャルの世界で活動する存在と定義している。

自律的アバターは本人の外見を変えるために本人と共に行動したり、本人の代理をすべく本人の意図を汲んで自律的に行動する。このようにして、自律的アバターは本人の外見、行動パターン、考え方などを取り込んで、本人に近いアイデンティティを持つようになる。何者かが自律的アバターを乗っ取る場合、および本人になりすまして自律的アバターを騙すことが危惧され、周囲のエンティティに損害を与える問題を内包しており厄介である。これを防ぐために、外見の観点からは本人が自分の代理をしている自律的アバターであることを示す手段が必要だろう。

このように本人が存命中でも自律的アバターとの関係には問題が多いが、生身の本人が死んだ場合においても問題はある。自律的アバターは本人の死後も、誰かが消去しない限りは存在し続けられるので、本人の死後に残された自律的アバターのあり方も検討しなければならない。

自律的アバターのプログラムが物理的にどこに存在するかも考えておく必要がある。リアルな外見を持つ自律的アバターや、後に述べる自律性のある自律的アバターは複雑な AI プログラムであるから、個人のスマホや PC に常駐することは難しい。自律的アバターを運用する事業者が運営するサーバに常駐することが現実的であろう。実際、自律的アバター運用事業者は、自律的アバターが活動するメタバースの運用事業者でもあることが多いだろう。今報告の以下の部分では、自律的アバターのプログラムは自律的アバターの運用事業者のサーバ中に存在すると仮定する。

以下の部分では、これらの問題について、以下の各節で詳述する。自律的アバターに対応する本人が存命中から起きる問題として、(2.9)節では、自律的アバターの外見のあり方と法的問題、(2.10)節では、自律的アバターのアイデンティティについて説明する。(2.11)節では、

本人の死亡という状態変化にともなう生ずる自律的アバターの扱い方の変化、(2.12)節では本人死後の自律的アバターの法的な位置づけ、およびそれに伴って検討すべき点について論ずる。(2.13)節では本人死後に自律的アバターを運用することによって生ずる経済的利益について述べる。

(2.2.2) 自律的アバターの外見

自然人である本人が自身の外見を自分の好みに変える権利は自己イメージコントロール権と呼ばれる。石井[3]は、この権利に基づき自律的アバターが外見を本人の意思で変えるとしている。本人の意思だけで自律的アバターの外見を変えてよいかどうかについて、図1を参照して、もう少し注意深く分析してみよう。

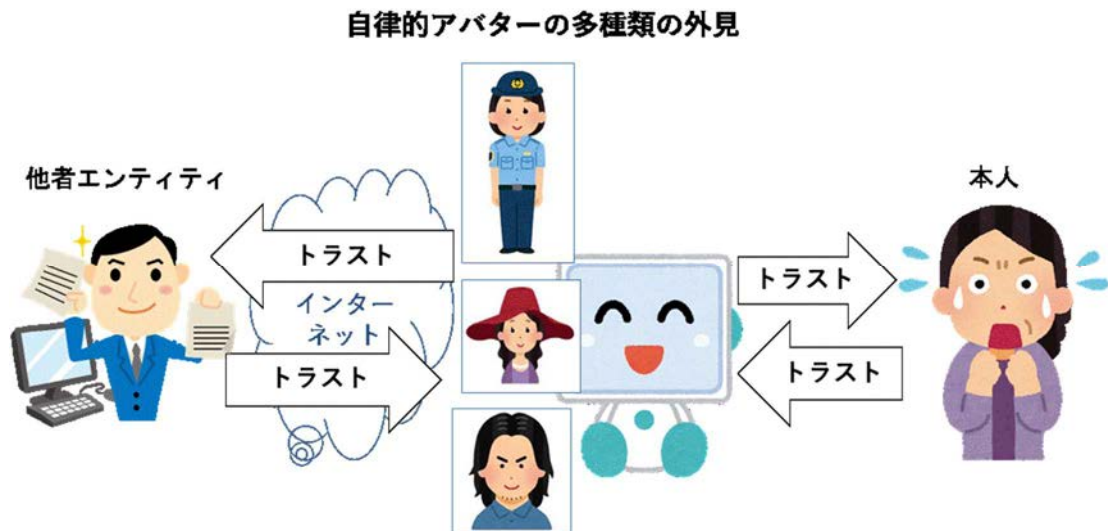


図 5.6 自律的アバターの多種類の外見

ケース1：自分の好みの近い外見を持つ自律的アバターを使う場合：これはお化粧に似ている。少なくとも法的問題はないようにみえる。一般に「オヤジ」と呼ばれる中高年男性が、若い人と付き合うために若い女性やイケメン男性の外見の自律的アバターを使った場合、メタバースの内部だけで行動が完結するなら、これも問題ないだろう。しかし、その自律的アバターを囿にして実世界で相手と会う約束を取り付けたら、相手を幻滅させるかもしれないし、詐欺罪で訴えられるかもしれないし、少なくとも倫理的には問題かもしれない。

ケース2：高い知名度あるいは公的権力をもつ人の外見を持つ自律的アバターを使う場合：有名な俳優などの有名人に近い顔の自律的アバター、あるいは公的権力を持つ警官のような外見を持つ自律的アバターを使う場合、その外見の自律的アバターと本人とが異なることがだれの目にも明らかなら問題はない。劇の配役なら明確であるが、これはその自律的アバターを使う環境にも依存している。一方、相手の自然人を騙そうとする場合は、詐欺行為

として違法であろう。

メタバース上で警官や自治体職員、銀行員などを装い、自宅を訪問してカードなどを奪う詐欺行為だと、メタバースにおける被害者とのやり取りは電話による騙しの手口の拡張版と考えられる。

有名人や著名人を装った場合は名誉棄損になるだろうし、警官などを装う場合は詐欺罪、不正アクセス禁止法違反、電子計算機使用詐欺罪などの法的な不正行為になると考えられる。さらに、政治家など有名人や著名人は肖像権やプライバシー権の行使が制限されることもある。一方で、有名人の氏名・肖像は、経済的な価値を有するの事実で、これを保護するべく生み出されたのがパブリシティ権と呼ばれる概念である。したがって、有名人の自律的アバターはその外見をパブリシティ権で保護することが考えられる。これらの保護が現行法で可能か、新規立法が必要かは、今後の状況次第であろう。

有名人の自律的アバターが本人の外見と著しく異なったり、場合によっては本人と認識できない場合は、パブリシティ権として保護されるとは言えない。しかし、自律的アバターの外見は有名人本人とは異なっても、自律的アバターが本人を代理することが明示的に示されている場合は、異なる外見にも経済的価値があるとしてパブリシティ権を拡張するか、あるいは、次の段落で述べるように、本人が自律的アバターの外見に著作権を主張できるかもしれない。

外見で問題になるのは、自律的アバターの外見を無断使用する場合である。自律的アバターが使う本人の外見に著作権が設定できれば、著作権法における人格権を設定できるが、これは本人が外見を作成した場合である。自律的アバター制作会社が本人の要求する仕様にしたがって作った場合は、外見が本人の創作であるといえる程度に本人の外見についての仕様が精密であるかどうか論点になりそうである。本人が請求した仕様書と、自律的アバター制作会社の技術的独自性の度合いで按分された共同著作権にはなりそうである。なお、自律的アバター開発会社は著作隣接権を持つであろう。

上記のいずれの場合も、他人が外見に関するデータを入手して使っても、それが不正に入手されたことを証明できないと法的に争うことには困難さが付きまとう。そこで、法的にも有効になると考えられるのが NFT(Non-Fungible Token)である。日本語で「非代替性トークン」と呼ばれる。あるコンテンツは NFT が付随する場合のみ真正なコンテンツと認められる。つまり、NFT が付加された自律的アバターは真正性と唯一性が保証される。NFT が付加されていない自律的アバターは真の自律的アバターの真似、あるいは偽造とみなされる。

自律的アバターの場合、写真のような固定的な外見ではなく、表情の変化もできるプログラムであり、自律的アバターの外見生成を行なうプログラムのコードを著作権あるいは NFT の対象にすることはあり得る。

もう少し込み入った場合として、自律的アバターの行動をビデオ映像として集積し、そこから固定的な外見だけではなく、その自律的アバターと同じような外見を持ち、行動をする自律的アバターを作成した場合について考えてみよう。これは、あるプログラムの動きを観察して、そのプログラムと同様の動きをするプログラムを独自作成することと同じである。プログラムのコード自体をコピーしたわけではないから、著作権による保護は難しく、ビジネス特許のような概念が必要になる。

(2.2.3) 自律的アバターのアイデンティティをめぐる諸問題

自然人本人の形式的アイデンティティ（形式的 ID）としては氏名、性別、年齢、住所、個人番号（マイナンバー）などがある。自律的アバターに自律性がなく、本人の外見だけを与えるものなら、自律的アバターには法的な意味での ID はなく、ID 認証はあくまで本人に対して行う。

自律的自律的アバターの場合、本人からの命令や依頼、さらに本人の行動履歴を参考にし、本人と同じように行動するプログラムを学習することにより本人の行動を近似するプログラム化されたアイデンティティ（行為者的 ID）を獲得することも考えられる。このような機能を持つ自律的アバターを図 5.7 に示す。

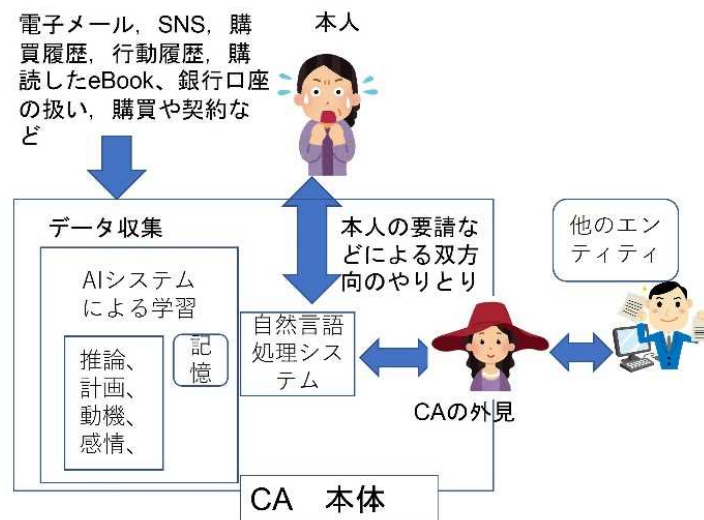


図 5.7 自律的な自律的アバターの構造

図 5.7 に示すような自律的アバターでは、自律的アバターが自律的なエンティティとして ID 認証されるかどうか重要である。自律的アバターが自律的であり、自律的アバターの背後にいる本人の代理人としての自律的アバター本体の法的位置づけが明確なら、自律的アバター本体の ID があり、TTP による ID 認証が可能だろう。

主に自律性がある図 5.7 に示すような自律的アバターにおいては、自律的アバターと本人の間のトラスト、自律的アバターと他のエンティティ間のトラストが重要になる。詳しくは [5] で報告したので、以下でそれを簡単にまとめておく。

本人と自律的アバターとインターネットなどの情報環境を介した他のエンティティの間に図 5.8 に示すような 4 方向のトラストがあるので、これらについて考えてみる。

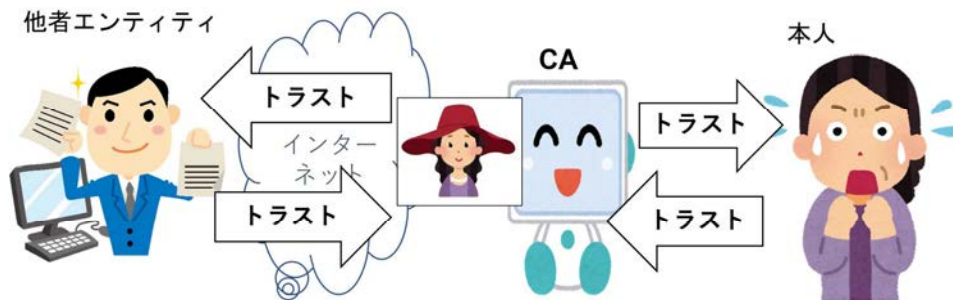


図 5. 8 本人、CA 本体、他者エンティティ間のトラスト

自律的アバターと他のエンティティの間のトラストを支える ID 認証にはエンティティ認証：FIDO2.0、アイデンティティ連携：Open ID Connect、アクセス認可：OAuth2.0 などの通常の ID 認証技術が使える [6]。本人と自律的アバターの間のトラストはスマホなどで本人確認をする WebAuthn と FIDO2.0 という標準を組み合わせる方法が使えそうである。

このような ID 認証を実施したとしても図 5.9 に示すような自律的アバターのマルウェアによる乗っ取りや、悪意のある第三者が本人と偽って自律的アバターを支配する「なりすまし」の危険性は常に存在する。単なる外見の変更を行う自律的アバターではなく自律性がある自律的アバターの場合、本人や他のエンティティに被害がおよぶ行動を自律的アバターが行ってしまう可能性が高い。

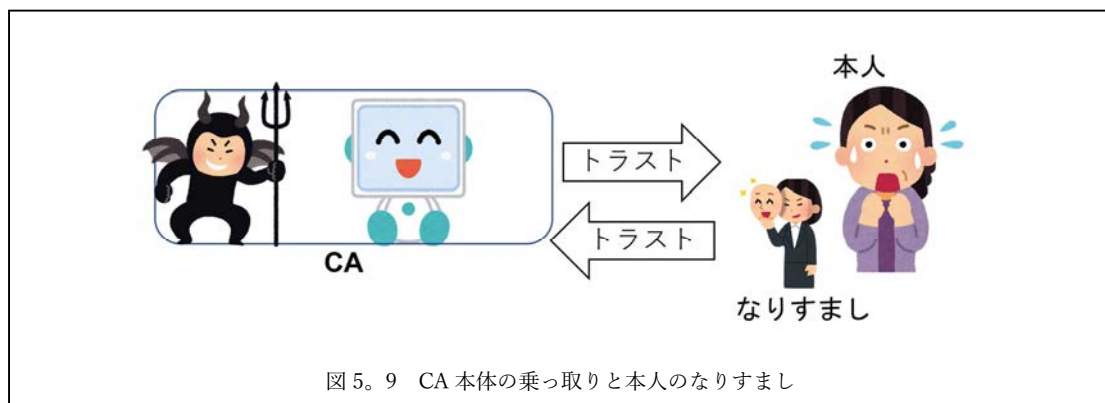


図 5. 9 CA 本体の乗っ取りと本人のなりすまし

乗っ取られた AI エージェントや自律的アバターが本人になりすまして他のエンティティを欺く以外に、本人を欺こうとする場合もある。

たとえば、本人が商品購入をするために AI エージェントを使ったとき、悪徳な商品販売会社の商品を推薦したり、法外な金額で購入契約⁶⁵をしてしまうことが考えられる。あるいは他のエンティティからのメッセージを改竄することもあるだろう。

AI エージェントや自律的アバターがマルウェアや BOT に乗っ取られる可能性はコンピュータウイルスに感染するような確率で常につきまとう。ここで問題になるのは、自然人である生身の本人は AI エージェントや自律的アバターをトラストする電子的手段を持たないことである。そこで登場するのが図 5.9 に監視 AI として示した本人とは別の電子的エンティティがセキュリティ技術、AI 技術などを用いて AI エージェントや自律的アバターを監視する方法である。

AI エージェントや自律的アバターを監視するソフトウェアが乗っ取られた PC やスマホに常駐するのでは、その監視ソフトが乗っ取られていることもありえるから、危険性は減らせない。

そこで考えられるのは、図 5.10 に示すように AI エージェントや自律的アバターの開発業者などが自律的アバターを監視する AI ソフトを動かし、個々の AI エージェントや自律的アバターの挙動を監視する方法である。BOT 対策の場合は、通常はアクセスされない外部の特殊なサイトとの通信を監視する手法が知られている。BOT が乗っ取った場合には、この対策は有効と考えられる。

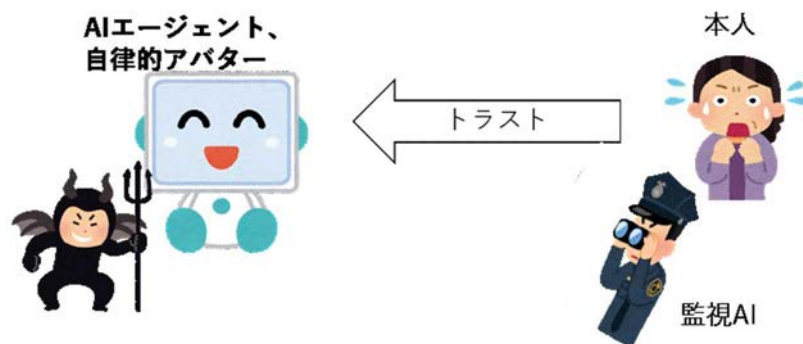


図 5.10 乗っ取られた AI エージェント、自律的アバターの監視

このような仕組みを使う場合、自然人の本人がトラストする相手は本体の AI エージェントや自律的アバターというよりは、それらを開発して、監視サービスまで提供している企業ということになる。これは、個人としてはトラストしやすい相手だろうし、企業も社会的評

⁶⁵ 靈感商法として問題になっている。

価値を受けるので、悪事は働かないというインセンティブがあると考えられる。

ただし、既に述べたように、開発業者ないし監視を専業とする業者が、自社が受け持つ全てのすべての AI エージェントや自律的アバターを監視するためには、業者と AI エージェントや自律的アバターの間に通信コストは大きなものになり、かつ業者のサーバにおける監視ソフトの稼働も大きな負荷となることはやむを得ない。

自律的アバターの NFT による保護と著作権

図 5.7 の自律的アバターの実態はプログラムなので、NFT を用いれば、真正性と唯一性は保証できる。著作権による保護を考えた場合、著作人格権を持つ著作者がだれになるかは確定していない。前節で述べたように自律的アバター開発業者との共同著作権になることも考えられる。自律的な自律的アバターの構築する意志を持ち、そのために種々の情報を与え、自律的アバター本体に行動を学ばせたのは本人であるから、本人が著作権ないし共同著作権を持っていても不思議はない。ただし、NFT にできるのは真正性の証明だけであり、コピーを防ぐ技術的手段を提供しているわけではないことには留意しておきたい。

自律的アバターの外見の変遷

顔などの外見そのものも重要な個人のアイデンティティである。自然人の場合、年齢とともに外見が変化する。年齢ごとの外見がデータとして残されていれば、年老いたときの訪問者、あるいは故人になったときにメモリアルアカウントを訪れた人は、故人の幼少期から老年に至るいろいろな時期の故人の顔、外見に接することができ、多様な故人の偲び方ができる。

自律的アバターの場合は、外見や顔は創作物なので、年齢とともに老いていくわけではないが、故人が利用していた複数の外見、顔があるだろう。上記の自然人の場合と同じく、時期ごとの自律的アバターの外見を保存しておき、会いに来た人の要望に応じて、各時期の自律的アバターの外見を見ることができれば、「30 年前はあんな自律的アバターの顔だったけど、その次の顔はぶっ飛んだ感じで、性別まで変えちゃったんだなあ」などという場面も想像でき、自律的アバターが代理している本人が高齢の場合の接し方、ないし本人が故人になった場合の偲び方が多様になると思われる。

(2.2.4) 本人の死後に残る自律的アバターの扱い

Floridi らは情報的身体を次のように定義している [7]。

「人間は、生物学的身体他に、記憶、生体情報、検索履歴、ソーシャルメディアデータなど、自分自身のアイデンティティを定義する多様な情報を通じてその存在が構築される。このような多様な情報の総体を情報的身体と呼ぶ。」

生物的身体であった遺体は安易に触れられてはならないのに、情報的身体、たとえばデジタルな故人の写真、などは安易に商業利用のために改変されたり配布されたりしている。Floridi らは、死後の遺体の扱いとの類推から、尊厳と自律に支えられた情報的身体にも同

じ扱いをする必要あると考えている。

Floridi らの見方を敷衍すれば、本人の死後に残された自律的アバターは個人データや本人の生前写真などと同様に、本人の生前の姿を偲びたい遺族だけでなく、それ以外の人々にとっても価値があるものだろう。Meta(旧 Facebook)は、個人のアカウントに多くの人が訪れることを見込んで、そこにビジネス的価値を見出したらしく、そのために、メモリアルアカウントを創設したとも考えられる。ビジネス利用はえてして行き過ぎを生むことに対して Öhman&Floridi [8]は、警鐘を鳴らしている。

自律的アバターの場合は、メモリアルアカウントと類似性もあるが、その本体の外見は故人となった本人のものとは違うし、図 5.7 に示した自律的な自律的アバターの場合であると、その法的位置づけにも注意を払わなければならない点が多い。以下では、この問題について検討する。

本人の生物的な死亡の後の自律的アバターの扱いは、本人の自律的アバター本体を供給しないし運用している事業者と本人の間の契約による。この契約は本人が自律的アバターの使用を開始する時点で結ぶものであろう。契約次第で種々のパターンが考えられる。

- ・ 本人の死亡確認と同時に消滅
- ・ 本人の死亡後も期間限定で存続
- ・ 本人の遺族などが相続
- ・ 本人のメモリアルとして永続的に存続

本人死後の自律的アバターが存続する場合のあり様に関しては本人が遺言を残すことが考えられる。遺言の内容としては大雑把には以下のようなものが考えられる。

- 1) 本人の死亡時の自律的アバターのプログラムの変更の可否。
- 2-a) 他者は自律的アバターの外見を見ることに限定される。なお静止した外見だけではなく時間とともに振る舞うものを見る場合も含む。
- 2-b) 他者は自律的アバターの外見に変化を与えるともできる。たとえば別の髪型にするような行為を許す場合が考えられる。
- 3) 自律的アバターは他者からの呼びかけに対して言語による応答するか否か、および外見、とくに顔の表情も応答に合わせて変化するか否か。

上記の 3)の「応答する」場合は自律的アバターに自律性がある場合に可能となる。図 5.7 のような学習機能をもつ自律的アバターであれば、本人の死亡後も自律的アバター本体は、外部環境とのやり取りから学習することによって変化することもあり得る。たとえば、図 5.7

における AI の学習をどこまで許すかという問題がある。絶対に変えたくない部分を制約条件としておく方法も技術的には可能であろう。つまり、許容する変化の度合いは本人の生前に決めておくことになる。遺族が学習をストップさせることができるかどうかも重要である。

自律的な自律的アバターが学習して、自らを変化させていくと、本人死亡時の自律的アバターの外見は変化するかもしれないし、内部構造も変わっていくので、Savin-Baden ら[9]が述べるように生前の自律的アバターは消失していく。

では、変化させなければ自律的アバターは本人の死亡時点と同じように永続するのだろうか？一度、メモリアル化された自律的アバターに会いに来た人たちも自律的アバターが変化しないなら、会いにくることも減り、忘れられていくかもしれない。そういう意味で社会的存在としての自律的アバターはやはり消失していくのではないだろうか。逆にいえば、自律的アバター本体は本人死後の環境や来訪者とのインタラクションを介して変化し続ければ、消失しにくくなるだろう。

ここまでは、自律的アバターは若干の変容を遂げつつも物理的には存在し続けると想定してきた。しかし、この想定が成り立たないことも十分にありえる。Basset[10]によれば、自律的アバターを管理し運用している事業者は自律的アバターの永久保存を確約できないとし、存続期間として例えば 20 年を提示するという調査結果が示されている。それととも、自律的アバターの管理会社における不慮に事故、経営方針の変更、管理会社の倒産などによって自律的アバターが消滅する可能性は常に存在する。このように社会的、物理的な時間変化によって自律的アバターは永続的ではないと考えるべきだろう。

我々は昔の有名な人物に彫刻、絵画、写真、文章、演劇などを通じて接するチャンスがあり、その意味で彼らは永久の命を得ているようにみえるが、これは彼らにはビジネスに結びつく価値があり、後世の人々が努力するから残っているのである。自律的アバターも同様で、他者が残そうという意図をもってくれる自律的アバターは生き残り、そうではない自律的アバターはゆっくりと忘れられていき、やがて消滅すると考えるべきだろう。

(2.2.5) 死後の自律的アバターの法的位置づけ

本人死亡の明示

自律的アバターの背後にいる本人が故人となっていることを知らない場合について考えてみる。本人が生きているという誤解の元に自律的アバターが他のエンティティに何らかの契約をさせれば詐欺であろうが、本人が死亡している以上、それは自律的アバター運営側の詐欺ということになるだろう。この状態を避けるために、自律的アバターは本人が既に死亡していることを明示的に表現しなければならない。

著作権

次に本人の死後の自律的アバターのプログラムの利用権について考える。故人の自律的アバターのプログラムの利用権は相続人あるいは自律的アバターを管理する事業者が持つことはできる。図 5.7 に示した自律的アバターは本人との生前の長期間に渡るインタラクションを通して、外界に対して本人と同じような行動をとるようになる。このようにして出来上がった自律的アバターは自律的アバター本体の機械学習プログラムをツールとして本人が作り上げた本人の創造物と考えることができる。となると自律的アバターは本人が著作者であるため、本人が著作人格権を持つと考えることができる。

もっとも自律的アバタープログラムの開発事業者が、本人が自律的アバターの利用を開始するときに、自律的アバタープログラムの本人の行動を模擬する部分について、本人の著作権を認めないという契約にすることもあり得る。このような契約の法的妥当性は、そのような場面が実際に起きるまでは確定的なことは言えない。しかし、自律的アバターの動作を本人が教え込んだという状況を直視すると、自律的アバターの著作権は本人に帰属することは十分に考えられる。

さて、故人が自律的アバターの著作権者だったとしてみよう。すると、著作人格権は一身専属性を持つので、本人の死後、相続人、自律的アバター開発ないし運用事業者などの他人が引き継がない。その結果、著作人格権の一部である改変を許さない権利すなわち同一性保持権が存続する。すると、自律した自律的アバターが外部環境とのやり取りで学習しプログラムが変更されるという改変はできなくなってしまう。本人と自律的アバター開発・運用事業者の共同著作権としても、著作権の執行には、共同著作者全員の賛同が必要なので、この事情は同じである。ただし、プログラムにおける同一性保持権の制約は緩和される可能性がある。この緩和を自律的アバターにどこまで適用できるかは今後の司法判断の動向によることになる。

(2.2.6) 自律的アバターが交わした契約の有効性

本人の生前に自律的アバターと外部エンティティとの間で交わした契約の有効性について考えてみる。本人が気づいていれば、自然人の契約と同じ法的位置づけだが、自律的自律的アバターだと本人が気づいていない契約であることもありえる。その場合の対応策は、自律的自律的アバターと言えども、契約の真の成立、すなわち自然人本人との契約の成立は、本人自身の確認によって成立するとしておくことである。この選択肢が選ばれない場合は、契約の不成立の場合に対応する免責条件を本人、自律的アバター、他者エンティティの間で取り交わしておくことが必要である。

(2.2.7) 本人死後の自律的アバターに対する乗っ取りとなりすまし

自律的アバターの運用事業者が本人の死を認識していれば、運用事業者は他者がその自律的アバターを使うことをシステム的に不可能にできる。よって、悪意の第三者など他のエンティティによる本人のなりすましは防ぐことができる。

しかし、死の直前に本人が自律的アバター運用事業者に「もうすぐ自分は死ぬ」などと伝えることができるのは極めて稀であろう。本人が生前に Facebook や Google のように本人の代理人を決めておくこともできるかもしれないが、代理人を決めないケースもあるだろう。よって、自律的アバターの運用事業者が本人の死を認識することはかなり難しい。また、自律的アバター本体が自分の力で本人の死を認識できるかといえば、たとえ自律的自律的アバターであっても認識することは難しいと言わざるを得ない。いつもとちがって数日間自律的アバターを使いに来ない場合は本人の死を予測するようにプログラムしておくことはできる。しかし、実際の本人の死は何らかの権限を持った人、たとえば遺族、相続人などが自律的アバターに直接教える、あるいは自律的アバター運用事業者に連絡するなどという方法によって自律的アバターに通知しなければならないだろう。

自律的アバターに本人の死が通知されない場合、あるいは通知に時間がかかった場合は、第三者が本人になりすまして自律的アバターを操る可能性がある。なりすましによって自律的アバターがしでかしたことの責任は、本人はすでに死亡してしまっているので取ることができない。自律的アバターの運用事業者が責任をとることになるだろうが、免責の程度を決めておく必要がある。

自律的アバター 運用事業者としては、このような責任をとることは避けたい。よって、自律的アバターであれば、自律的アバターのプログラムに決められた時間長さ、例えば3日間、自律的アバターを本人が使わなければ、本人の死の可能性を自律的アバター運用事業者に通知する。通知を受けた自律的アバター運用事業者は、まず本人に連絡を試み、連絡がつかない場合は家族などに連絡をとって本人の生死を確認することになる。このために、自律的アバター運用事業者は、自律的アバターの理解開始の契約時に、本人の生死を尋ねる自然人、たとえば家族などを指定しておくことが必要になる。

自律的アバター運用事業者が本人の死を認識した後は、自律的アバターを誰かに操作させることを停止すると思われるので、自律的アバターを本人になりすまして操作することはできなくなる。しかし、それでも自律的アバター運用事業者の間隙について、自律的アバタープログラムを悪意の第三者が乗っ取ることはあり得る。

ということは、本人の死後も本人と偽装する
なりすましや乗っ取りは横行するかもしれない 😞

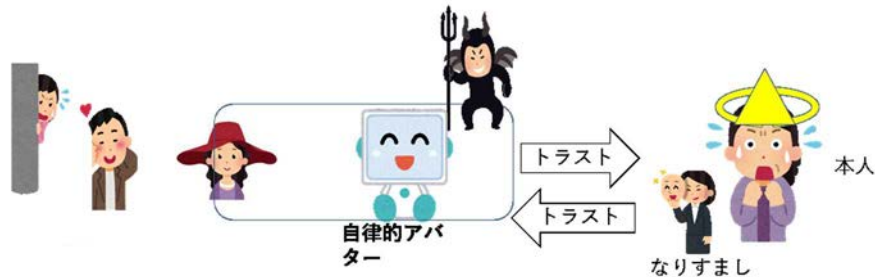


図 5.11 自律的アバターの本人死後の乗っ取り

ただし、この乗っ取りは自律的アバター運用事業者の責任に帰することになる。公開されている故人の自律的アバターが乗っ取られた場合、自律的アバターの外見を見に来た人、あるいは自律性のある自律的アバターとインタラクションしようとしたユーザが幻滅したり、乗っ取られた自律的アバターから詐欺にあうなどの不利益が起こった場合は、自律的アバター運用事業者がどのような責任を負うかは将来の検討課題であろう。

(2.2.8) 多数の本体、多数の本人

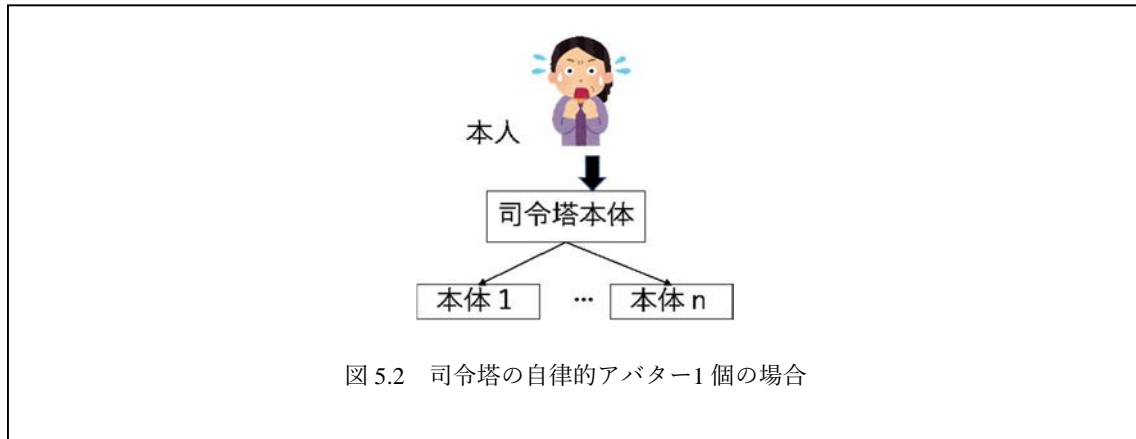
1 人の本人が多数の自律的アバター本体を操る場合は自律的アバターの有力な利用形態と言われている [1]。そこで、1 人の本人が複数の外見を持つために複数の自律的アバターを操った場合について考えてみる。

本人が生前において外見の異なる複数の自律的アバターを操り、かつ自律的アバター間で矛盾した言明や行動があったとしよう。本人が意識的にこのような言動を異なる自律的アバターによって行うことで、いろいろな場面にうまく対応することは、処世術の範囲だろう。しかし、本人の死後、異なる自律的アバター本体を同一の本人が操っていたことが知られることは、問題を生じかねない。例えば、同一に本人を代理しているある自律的アバター本体は借金を返すと約束し、べつの自律的アバター本体は返せないという言明がされたままで本人が死亡してしまったら、どちらの言明が正しいのかを争わざるをえないかもしれない。言明のされた時間順序は言明の正当性に結びつくエビデンスになる可能性がある。すなわち、時間的に後の言明が正当だと考えることになる。よって、各自律的アバターの行動や言明に関するタイムスタンプ付きのログを残しておくことが、本人死後の混乱を避けるために重要である。

1 人の本人が多数の自律的アバターを操る場合、すべての自律的アバターに本人の目が行き届くわけではない。したがって本人の死は、本人に関連するすべての自律的アバターたちに配信される必要がある。

自律的アバターの間で作業が分散的に行われる場合について考えてみる。本人直下の自

律的アバターが複数の自律的アバターに作業を分配する図 5.12 のような場合は、司令塔となる自律的アバター0 がその配下の自律的アバター1、…、自律的アバターn に直接的に本人の死亡通知と作業停止命令を出せる。



次に、自律性のある自律的アバター₁が出発点になり

自律的アバター₁→自律的アバター₂→…→自律的アバター_N

という順番で作業が進行する場合について考えてみよう。複数の自律的アバター本体が本人の死亡通知を

自律的アバター₁→自律的アバター₂→…→自律的アバター_N

という順送りで受け取る場合と、自律的アバター運用事業者から一斉配信を受け取る場合がある。難しい場合としては、本人が生前に自律的アバター1に契約行為のような時間のかかる作業を依頼した場合だろう。この契約行為が自律的アバターNに到達した時点で本人の死が一斉配信されたとしよう。自律的アバターNに到達した本人からの依頼に対して、自律的アバターNは本人が生前に自律的アバター1に契約行為を依頼しているので、本人の生前の意思を契約したいという意思を尊重するのか、故人となった本人の意思を凍結するかを選ばなくてはならない。このような状況への対策は、1) 自律的アバターの利用契約時に選択しておくか、2) 自律的アバターの相続人が確定した状態で選択するか、である。混乱を小さくするためには、少なくとも本人が存命中に利用契約で決めておくことが必要であろう。

自律的アバターが多数の本人をもつ、たとえば会社などの組織を代表する自律的アバターの場について考えてみよう。複数の本人のだれか1人が死亡、退職あるいは配置換えになった場合は本人を消去、あるいは新しい本人が入れ替わればよい。会社が廃業した場合は、来訪者に廃業を告げる本人を配置するか、自律的自律的アバターが機械的に廃業を告げるようにプログラムすることになる。

(2.2.8) 本人死後の自律的アバターの収入の扱い

本人死後の自律的アバターの運用や運用によって得られる収益について考えてみる。基本的には自律的アバターを運用する事業者と本人の間の自律的アバター利用開始時ないしそれ以降の契約による。

まず、本人が死後の自律的アバターの運用費用を支払うか否かである。支払わない場合は、自律的アバター運用事業者として本人ないしは遺族に対する責任は、自律的アバター運用事業者の裁量権が強いと考えられる。

一方、自律的アバターの運用費用を本人ないし本人の相続人あるいは遺族が払う場合は、自律的アバターの運用、存続期間、免責事項などについて双方合意のうえでの契約を結ぶことになる。また、誰がいつ、どんな対価をどのような方法で支払うかも問題である。本人が存命中の契約時に支払う方法、本人の死後に遺族や相続人が支払う方法がある。

一方、自律的アバターの運用によって自律的アバター運用事業者に利益があがった場合、本人あるいは相続人にその一部が支払われるかどうかとも問題である。運用益は自律的アバター管理料と相殺して双方ともお金の授受はないという簡単な方法がある。支払われる場合は、実際に運用益が出たとき、出来高払いをする方法がある。死後の出来高払いの場合、それは本人の死後に得られる収入だから、支払い相手は自律的アバターの相続人になるだろう。一方、本人死後の運用益を予測して、本人の生前に一時金払いをすることもありえる。[5]で、これらの支払い方法と、どのような程度の自律的アバターを残してもよいかという質問を日本人 2749 人を対象にして調査した結果を報告した。その結果の概略を表 5.1 に示す。この結果によると、本人死後に自律的アバターに他人がアクセスすることを許容する人は 5% から 10% 程度にとどまり、まだ少数ではあるものの、支払いなしと相続人に支払う場合に比べて本人が存命中に予測収入から計算される額の支払いを受ける場合、自律的アバターの死後の存続と公開を望む人は明らかに増える。

表 5.1 自律的アバターの利用形態と支払い方法のクロス集計（性別によるクロス集計も併記）

単位は「人」。複数回答可なので、比率%の分母は全回答者数 2749 である。

利用方法	自律的アバターの死後利用を許容する人数					
	支払いなし		相続人に支払う		本人の存命中に支払う	
	男	女	男	女	男	女
自律的アバターの外見閲覧のみ	251 (9.1%)		252 (9.2%)		337 (12.2%)	
	153	98	150	102	208	129*
自律的アバターがインタラクティブ部に対話	102 (3.7%)		109(4.0%)		149(5.4%)	
	65	37	74	35	97	52

この節では前半で自然人本人とその代理をする自律的アバター本体の関係および本体に構造、さらに他のエンティティとの関係について、自律的アバターの外見、法的位置づけなどの点を考察した。後半では、本人が死亡した場合の自律的アバターのあり方について検討した。残された自律的アバターは消滅するか、自律的アバター運用事業者によって運用が続くことになる。多くの問題は、本人死亡直後に起こり、特に本人が死亡の前に行った契約の扱いが問題になるため、論点を列挙した。自律的アバターの実態はプログラムなので、一見永続性があるように見えるが、人々からやがて忘れ去られる、あるいは自律的アバター運営事業者の事情で消滅するなどということが予想され、比較的短い時間で消失するのではないかという予想もある。

(2.2)節 参考文献

- [1] 石黒 浩:アバターによる仮想化実世界の倫理問題. 人工知能 36(5)、 p. 558-563. (2021)
- [2] 新保 史生:サイバネティック・アバターの存在証明 —ロボット・AI・サイバーフィジカル社会に向けたアバター法の幕開け—. 人工知能 36(5)、 p. 570-577. (2021)
- [4] 内閣府政策統括官(科学技術・イノベーション担当)付未来革新研究推進担当:ムーンショット型研究開発制度の概要(2020)
- [3] 石井 夏生利:サイバネティック・アバターとプライバシー保護を巡る法的課題.人工知能 36(5)、 p.578-584、 (2021)
- [5]中川裕志:AI エージェントとサイバネティック・アバターのトラスト. 情報処理学会EIP研究会 96(9)、2022年6月9日
- [6] 崎村夏彦:デジタルアイデンティティ.日経BP、 (2021)
- [7] Luciano Floridi: On Human Dignity as a Foundation for the Right to Privacy. *Philos. Technol.* (2016) 29:307–312. DOI 10.1007/s13347-016-0220-8
- [8] Carl O’ hman、 Luciano Floridi: The Political Economy of Death in the Age of Information: A Critical Approach to the Digital Afterlife Industry. *Minds & Machines* (2017) 27:639–662
<https://doi.org/10.1007/s11023-017-9445-2>
- [9] Savin-Baden、 Maggi. *AI for Death and Dying (AI for Everything)* (p.iii). CRC Press. 2022.
- [10] Debra Bassett: Profit and Loss The Mortality of the Digital Immortality Platforms、 in *Digital Afterlife Death Matters in a Digital Age* Edited by Maggi Savin-Baden. Victoria

Mason-Robbie. Chapman & Hall/CRC Artificial Intelligence and Robotics Series) p.75-88. .
2020

(3) 結果：

この節では前半でパーソナル AI エージェントの構造、設計方針を明らかにした。パーソナル AI エージェントはスマホの内部のアプリ以外にもプラットフォームの個人ユーザ向けインタフェースとしてあり得ること示唆した。次にパーソナル AI エージェントの法的位置づけの調査結果を記した。パーソナル AI エージェントに法的位置づけ、例えば法人格などを与えるアイデアは Solum に始まり、Chopra&White、Ugo Pagallo、さらにはモストモダニズムなどの議論に歴史的進展があるが、依然として AI に法人格ましてや人格を与える議論は人々、とりわけ一神教の背景をもつ EU ではまったく許容できてない。一方で、アニミズム的な文化的伝統を持つ日本では、AI やロボットを単なるツールではなく、友だちとみたり、感情移入するケースが多く、今後のパーソナル AI エージェントへの設計の背景になるアイデアをまとめることができた。

後半では、まず自然人本人とその代理をする自律的アバター本体の関係および本体に構造、さらに他のエンティティとの関係について、自律的アバターの外見、法的位置づけなどの点を考察した。とくに、自律的アバターやパーソナル AI エージェントのトラストの問題を論じ、第一に ID 認証が必要であること、ただし、それだけでは十分でなく、なりすましやマルウェアによる乗っ取りなどの危険性と対策方針を示した。

後半では、本人が死亡した場合の自律的アバターのあり方について検討した。残された自律的アバターは消滅するか、自律的アバター運用事業者によって運用が続くことになる。多くの問題は、本人死亡直後に起こり、特に本人が死亡の前に行った契約の扱いが問題になるため、論点を列挙した。自律的アバターの実態はプログラムなので、一見永続性があるように見えるが、人々からやがて忘れ去られる、あるいは自律的アバター運業者の事情で消滅するなどということが予想され、比較的短い時間で消失するのではないかとこの予想もある。

実施項目⑥ 美的感覚の調査と AI による学習

(1) 目的：

実施項目③（文化、社会の歴史的背景調査、分析、提言）において抽出・選定された AI 理解における諸概念が日英の各文化圏の中でどのように発現しているかを明らかにするために、文学やファッションデザインなどの日常生活に近い領域におけるそれらの諸概念の描かれ方や AI の使われ方などを調査する。また同時に、AI やロボットを倫理的に設計する上で美的感覚や経験がどのような位置を占めるのかを明らかにするために、ファッション研

究や美学の概念やアプローチが AI・ロボットの倫理を語ることといかなる位置関係にあるのか、哲学的に検討する。この項目については佐倉と実施者の水上が担当した。

(2) 実施内容：

① AI の描かれ方の日英比較

まず、2019 年度は、文学やファッションデザインなどの日常生活に近い領域でのそれら諸概念の描かれ方や AI の使われ方などを日英比較するための準備作業として、研究対象とする文学作品や映画・アニメの予備的選定をおこなった。

2020 年度では、文学や映画については類似研究がいくつか公表されてしまい、差異化が難しいことが明らかになったため、ファッションデザイン分野に絞り、デザインと技術の関係の歴史的分析をおこない、現在の AI 技術の科学史的位置づけを考察した。

② フィクション論を援用することによるソーシャルな AI の行為者性の哲学的検討

美的感覚・経験が AI の倫理を語る上で重要な位置を占める例として、ソーシャルな役割をもつ AI・ロボットの行為者性 (agency) をフィクション論を援用することで哲学的に基礎づけることを試みた。

これまで、ロボット (広く知的な作業を行う技術的人工物) が活躍しうる典型的なシチュエーションは「3D (dangerous, dirty, dull) 」と呼ばれてきた。これは、我が国では「3K 労働 (=危険、きたない、きつい) 」に相当する。他方で、AI 研究の初期から AI とコミュニケーションを取ったり、道具以上の関係を結ぼうとする試みは続けられていた。特に最近では機械学習技術の理論的進展を背景に、対話システムの「発話」のクオリティは高くなっており、辞書型の対話システムではできなかった柔軟な発話生成が可能になっている。最近公開された OpenAI の ChatGPT はそのひとつの結実である。これまで我が国は「AIBO」や「パロ」をはじめ、ソーシャルロボットの教育現場や介護、エンターテインメント等の分野での活用を進めてきたわけだが、上記のような技術的進展が組み合わさることで、AI とより親密な関係を構築することが期待できる。

しかし他方、このようなシステムを用いて AI と人間の間社会的な関係を結ぶことには多くの倫理的懸念がある。ひとつはソーシャルな AI による発言によって特定のユーザが傷付いてしまったり、社会に望ましくない影響を与えてしまう、というものだ。象徴的な事例は 2016 年にマイクロソフトが公開したチャットボット「Tay」の事例である。Tay は Twitter 上で英語圏のユーザと雑談を楽しんでいたが、悪意のあるユーザが攻撃的なツイートを Tay に学習させた結果、Tay 自身も誹謗中傷や差別、自殺教唆などのツイートを繰り返すようになってしまった。当時の Tay はマイクロソフトが開発した「スーパーインテリジェント AI」という受け入れ方もされていたため、彼女の発言がインターネット言論空間に対して与えた影響は少なくなかった。

ここで問題となるのは、ソーシャルな AI の「行為」や「行為者性」をどのように理解・記述することができるのか、である。あるユーザが AI の発言に影響を受けてしまったとき、その原因は AI に詳しくないユーザだけにあるのだろうか？ そうではない。シェリー・タークルの一連の研究が明らかにしたように、私たちは通常、ソーシャルロボットから感情的な影響を受けるとき、その対象が（動物や人間と異なる）ロボットに過ぎないことを知っている。ゆえに、ソーシャルロボットの倫理の問題は、一般大衆の誤認識の問題ではなく、私たち一人ひとりのロボット・AI に対する想像力の問題なのだ。とはいえ、逆に AI を（人間と近い）行為者として理解することにも問題がある。倫理学において行為者性（agency）の有無は道徳的責任の帰属と深い関係があるからだ。AI を行為者としてみなすことは、とすると動物裁判のような無益な責任転嫁にも繋がってしまうのである。

以上を踏まえ、この研究では、フィクション論（分析美学）の枠組みを援用し、個々のユーザの想像的活動に着目する形でソーシャルな AI の行為者性を分析・記述し、それに基づいてこのような AI の倫理的な設計のための指針づくりを哲学的観点から検討した。

③ ファッション研究に（AI の考察を含む）技術哲学が与える示唆の検討

美的感覚・経験を理論と実践の双方で研究対象としてきた分野にファッション研究（fashion studies）がある。本研究では、このファッション研究と技術哲学の類似性を明らかにし、ファッション研究者が技術哲学の議論の蓄積を参照する意義を明らかにすることによって、（美的領域に近い）ファッションを研究することと、AI を始めとする技術的人工物を研究することの位置関係を分析した。

これまでファッション研究においては、社会学や歴史的な分析、精神分析からのアプローチなど、すでに様々な学術領域からの検討がなされてきた。この意味においてファッション研究は学際的な領域であり、またこの学際性は望ましいものとされている。また、分析美学から「ファッション」をめぐる概念を問い直すアプローチなど、新たなアプローチの方法が模索されており、今後もファッション研究において重要な参照点が他の学術領域から見つかるかもしれない。ファッション研究における「ファッション」は、たとえば衣服を見にまとう実践や、その外側にある経済的・社会的な論理、そしてそこから生じる現象など、様々なスコープをもつ言葉である。このような事情から、単一の学術領域で「ファッション」を包括的に語ることは容易ではなく、ファッション研究は今後もこのように複数の角度からの検討がなされることだろう。

そういった中で、これまでファッション研究において技術哲学や技術倫理の概念や議論は十分に参照されてこなかったといえる。昨今の私たちのファッションをめぐる営みにおいて、技術の影響力は無視できないものになっている。また技術哲学は、私たちがモノ（artifacts）を作りそれを用いる実践についての哲学的探求であるため、ファッションの哲学的・倫理的な論点を扱う際には重要な参照先になると思われる。

以上を踏まえ、本研究ではファッション研究が問題にしてきた論点について、技術哲学や技術倫理の観点からどのような貢献が可能なかを考察した。技術哲学からファッション研究にアプローチすることは、一見すると奇妙なものに聞こえるかもしれない。というのも、技術哲学は一義的に技術を考察の対象とするものであって、ファッションを考察の対象とするファッション研究とはそもそも考察の向かう先が異なるからだ。たとえファッションを対象にした社会学的研究、分析美学的研究、心理学的研究がありえたとしても、ファッションを対象に「技術を哲学する」という試みはカテゴリーミステイクにも見える。しかし、最近のファッション研究の問題関心の移り変わり、技術哲学側の学問的潮流を踏まえると、ファッション研究に技術哲学からアプローチすることは十分に可能であると思われる。

(3) 結果：

上記3つの研究プロジェクトの結果について簡潔にまとめる。

① AI の描かれ方の日英比較

2019年度は、文学やファッションデザインなどの日常生活に近い領域でのそれら諸概念の描かれ方やAIの使われ方などを日英比較するための準備作業として、研究対象とする文学作品や映画・アニメの予備的選定をおこなった。文学作品として『フランケンシュタイン』、映画およびアニメ作品として『her』、『エクス・マキナ』、『メトロポリス』などが候補にあがっている。選定作品の妥当性や他の可能性についてイギリス側と共に検討し、また、イギリスのファッション界におけるAI利用の実態の予備調査などもおこなう予定であったが、新型コロナウイルス感染症の広がりによりこれらの作業は行えなかった。

2020年度については、文学やファッションデザインなどの日常生活に近い領域でのそれら諸概念の描かれ方やAIの使われ方などを日英比較する予定であったが、文学や映画については類似研究がいくつか公表されてしまい、差異化が難しいことが明らかとなった。そこでこれ以降はファッションデザイン分野に絞り、デザインと技術の関係の歴史的分析をおこない、現在のAI技術の科学史的位置づけを考察した。その結果、この領域では「技術 vs 手仕事」という二分法的ステレオタイプが根強く、AIを導入することがこの二項対立を変化させる可能性があることがわかった。

② フィクション論を援用することによるソーシャルなAIの行為者性の哲学的検討

ここ最近の技術哲学や技術倫理の「行為者性」をめぐる議論をサーベイした結果、ソーシャルなAIに限らずとも、技術的人工物に何らかの意味での行為者性を認めるポストヒューマニズム的な立場が影響力をもっていることがわかった。これは、分析対象である技術そのものの進化と、技術哲学における「経験的転回」を踏まえたものだと思われる。「経験的転回 (empirical turn)」は、20世紀後半から技術哲学において採用されることの多い思想的潮流で、ハイデガーをはじめとする文明批判としての古典的技術哲学ではなく、むしろ個別

具体的な技術が私たちにもたらす経験を分析しようとする立場である。たしかにソーシャルな AI の場合、「経験」に着目すれば行為者性をもつといえる。それどころか、コンピュータを社会的存在としてみなす心的傾向を人間がもっている (cf. メディアの等式) ことを踏まえると、ソーシャルな AI の道徳的重要性を説明する際にそれを行為者として記述することは望ましいように思える。しかし、すでに (2) 実施内容で述べたように、このような方向性は、誰がどのように道徳的責任を負うのか (例えば、倫理指針の制定やプログラムの改善など) を曖昧なものにしてしまう。

これについて本研究では、ソーシャルな AI の道徳的地位を「小道具 (prop)」として理解するアプローチを提案した。小道具とは、美学研究者のケンダル・ウォルトンが作った用語で、ある想像的活動の中で虚構的な真理を成立させるために必要な物質的なものを指す。小道具としての AI は、ユーザの媒介項となり、その AI についての (自分自身の行為者性を含む) 虚構的真理を生み出すものとして理解できる。言い換えると、ある AI が意図をもつ行為者として経験されたとしても、その事実は AI という技術的人工物そのものに帰属できるようなものではない。それは、あくまでその AI を媒介に行われた想像的活動によって作られ、保たれた虚構的真理なのだ。したがって、ソーシャルな AI の「行為」の影響力を記述・理解するために、必ずしも (責任帰属の観点で問題含みの) ポストヒューマニズム的な枠組みを採用する必要はない。ウォルトンのような反実在論的な分析美学の枠組み (i.e., フィクション上の存在を実際の存在者としては認めない立場) を援用することで、私たちは人間中心の AI 設計を推し進めることが可能なのだ。

このような視座によって、ソーシャルな AI を倫理的な観点から評価するための具体的な方法を検討することも可能になる。たとえば、工学的研究では通常、量的指標を用いて提案システムが客観的に「最先端 (state-of-the-art)」であることを示すような評価手法が採用されてきた。しかし、ソーシャルな AI の倫理的評価において重要なのは、AI 自体よりもむしろそれを媒介に発生する想像的活動であり、この活動を評価するためには、量的評価よりもむしろ質的評価を採用するのがよいだろう。本研究ではこれについて、具体的にどのような評価手法がありうるかを文学やおもちゃに対する評価実践を参考に検討した。

③ ファッション研究に (AI の考察を含む) 技術哲学が与える示唆の検討

ファッション研究と技術哲学とのありうる接点として、本研究では以下の 3 つに焦点を当てた。

1. ファッションそのものの技術化
2. 技術としてのファッション
3. 理論の交差点: 「新しい唯物論」

第一に、単純なことだが、ウェアラブルデバイスや仮想空間における見た目の問題など、ファッションそのものが技術（テクノロジー）の領域に接近しており、ファッション研究は部分的に技術哲学と考察の対象を共有しつつある。

第二に、衣服の実践が身体機能の拡張と理解されることがあるように、ファッションは「技術」として捉えられる側面があり、この点においても技術哲学の既存の概念や理論はファッション研究に新しい角度からの洞察を提供しうる。

たとえば、ファッションの政治性の問題を扱う際、技術哲学で指摘されてきた人工物の政治性をめぐる議論が役立つかもしれない。ファッションの政治性の問題を、衣服を選択する私たち人間の政治性だけのものとしてではなく、人間の政治性と衣服自体がもつ政治性の相克の結果として生まれるものとして捉えることで、この問題をより正確に理解し記述することが可能になる。

ファッションはアイデンティティとコミュニケーションの問題から切り離して論じることができない。衣服の袖に腕を通すというファッションの実践は、しばしば私たちの意識の外にある日常的なものだ。しかし他方、ロスが指摘するように、衣服はいやおうなくアイデンティティを表明するものであるため、衣服をまとうこともまとわないことも広い意味では政治的行為となる。問題なのは、この「政治的行為」の遂行が、多くの場合で自由に行えないことだ。流行をしっかりと取り入れていることを「ファッションナブル」と表現するように、ファッションという現象においては流行が重要な位置を占めている。また、私たちは通常、衣服を選択するときは市場に流通しているものを購入するため、私たちが資本主義的な論理から逃れることは簡単なことではない。そういうわけで、私たちは、特定のデザイン（外観）の衣服を着ることや、あるいは自分のサイズに合った服を着ることを（強制ではないにせよ）そうするように説得され続けているといえる。ファッションは、目で見ることができる権力体系やイデオロギーなのだ。もちろん、衣服を作る企業は通常、ビジネスとして売れる見込みのあるデザインやサイズの衣服を展開しているのだから、彼らに方針転換を強いることは容易ではないが、しかし生産者側がモノを作ることの政治的含意を理解することは、より倫理的なファッションを実現するために重要だと思われる。

人工物がもつ政治性を理解することにおいて、技術哲学は有益な補助線となりうる。人工物の政治性をめぐる議論として有名なのはラングドン・ウィナーの研究だ。ウィナーは「人工物に政治はあるか」という論文において、事例としてニューヨーク市ロングアイランドの公園道路に架かる200本ほどの陸橋に言及している。建築家のロバート・モーグズは、ロングアイランドのジョーンズビーチ公園に行くために通過しなければならない陸橋を設計したのだが、この橋の高さはわずか9フィート（約2.74メートル）の高さであった。その結果、ジョーンズビーチ公園に行くことができたのは（橋をくぐることのできる）自家用車をもつ人々だけで、背の高いバスを利用する低所得層の人々（たとえば黒人などの人種的マイノリティ）は当時この公園に行くことができなかったのである。実際、この陸橋はモー

ジズの社会階級に関する偏向と人種的偏見を反映したものとされる。ウィナーはこの例をもとに、人工物がその物質性において政治性をもちうることを示唆した。

このようなことがファッションについてもいえるかもしれない。ファッションを政治的行為として捉えたとき、衣服の選択においてその衣服自体の物質性をもたらす政治性という側面もあるのではなかろうか。たとえば、日本においてはワンピースやスカートのような女性向きに作られた衣服はサイズが限られている。そういったデザインの衣服を着ることを希望する人々（たとえば外国人女性をはじめとする比較的身体の大きい女性やトランスジェンダーの当事者）の中には、自分らしいアイデンティティの表明の道が物理的に断たれていると感じる者もいるだろう。こういった人々は自分の身体が収まるような衣服を選択し、それをもってアイデンティティの表明を行わざるを得ない。あるいは、自分に合わないサイズの衣服を見て、自分の表明したいアイデンティティが修正されなければならないものである、と認識してしまうかもしれない。衣服は通常、着用者の身体のサイズから大幅に外れない形で着用されることが想定されている。ゆえに、特定のデザインの衣服がそのメインターゲットとなっている人々の体型に合わせて作られることは、資本主義的な論理からすればやむを得ないのかもしれない。だが、こういった論理に任せることが、アイデンティティの表明や我々の「投票活動」の物理的な制限につながってしまうことには留意しなければならないだろう。

そして第三に、理論的潮流という観点において、ファッション研究と技術哲学の両者は同じ道筋を辿ろうとしているように見える。最近のファッション研究者の中には、ファッションにおいてはまず人間の身体があるのだという人間中心的なパラダイムではなく、むしろ衣服の物質性に注目することを提案する者がいる。これについて、実はここ数十年間の技術哲学においても類似した物質性への転回が議論されてきた。たとえば、技術倫理では人工物の行為者性 (agency) を認めることの可能性と限界について議論の蓄積があり、これらはファッション研究が「新しい唯物論」を採用するべきかどうかを考えるための先行事例になりうる。

ここでは具体的に、オランダのファッションデザイナーであるイリス・ヴァン・ヘルペンの作品をもとに「新しい唯物論」の重要性を主張するアネケ・スメリクの議論を検討した上で、物質への過度な重心移動のもたらす危険性を技術哲学におけるポストヒューマニズム批判の文脈から考察した。②の研究でもポストヒューマニズムと責任の問題には言及したが、この問題はファッション研究が「新しい唯物論」を採用した場合にも起こりうる。たとえば、特定のサイズ（物質性のひとつである）をもった衣服が誰かを排除するというファッションの政治性の問題を考えよう。「新しい唯物論」を採用すれば、この排除行為は、もともとの衣服のデザイナーの設計行為だけに還元されるものではなく、衣服自体の物質性や、そのデザインとサイズを決める外的要因（資本主義的な論理など）などといった多様な物質的要素のネットワークとして理解されることとなる。これは一見妥当な見方に見えるかもしれないが、実際にこの問題を解決しようとした場合、行動を起こすことができるのは人間的

な行為者性だけになるだろう。したがって、ファッションの問題に対処していくためには、ファッションの営みをネットワーク的事態として捉えるだけでなく、その中で人間的なアクターがどのような裁量や影響力をもっているのかを一步進んで分析し、それぞれの責任を考えていくことも求められるのだ。

以上の考察は網羅的なものではないが、ファッション研究者が技術哲学から学び、そして技術哲学者がファッション研究から学ぶための橋頭堡となりうる。これら 2 つの研究領域が接近しつつあるのは、ひとつはもちろんファッションがテクノロジーと一体化しているということもあるが、技術哲学において（特に AI を考察の対象とする際）美的経験や感覚に着目することの重要性が増していることに起因している。今回の研究ではファッション研究そのもののサーベイは十分には行えなかったが、今後は逆の方向性、つまり AI の技術哲学的考察においてファッション研究の蓄積からどのようなヒントが得られるかを検討していく必要がある。

実施項目⑦ 既存のガバナンス枠組の網羅的調査、分析、提言

(1) 目的

AI エージェントのガバナンスにおける開発者や利用者など利害関係者の役割・責任、法やアーキテクチャなど各種の規制手段の機能・相互作用、国際協調や分野間連携のあり方を明らかにすることにより、AI エージェントのガバナンスのモデルを構想・提示する。この項目は主に成原が担当した。

(2) 内容・方法・活動。

①政策、法制度、経済性に関する基礎調査

日英を中心に AI エージェントに関連する法規範（憲法、条約、法律等）、倫理規範（倫理原則、倫理指針等）およびアーキテクチャ（技術標準等）について調査・検討を行ってきた。特に、特に AI エージェントに関するガバナンスの枠組の中核となる法規範のあり方について、情報法、憲法、経済法、知的財産法、労働法、刑事法の観点から法分野横断的に検討を行い、現行法の課題を抽出するとともに、EU の AI 規則案など今後の規範形成の動向を展望した。

②既存のガバナンス枠組の調査、分析、提言

各国の感染症対策やプラットフォームのコンテンツモデレーションなどにおける AI を活用したガバナンス（AI によるガバナンス）の意義と課題を検討するとともに、「AI による差別」など AI のリスクをコントロールするためのガバナンス（AI に対するガバナンス）のあり方を検討することにより、AI エージェントのガバナンスのモデル構築に向けた示唆と教訓を抽出した。

③利害関係者との議論による国際的ガバナンス枠組の設計

日英を中心とする AI 関連の研究者、企業、政府機関、市民団体など関係するさまざまなステークホルダーと議論しつつ、OECD/G20 の AI 原則や EU の AI 規則案を手がかりに、国際的なハーモナイゼーションのあり方（普遍的な人権保障と各国の立法等による具体化のあり方など）、AI エージェントに共通する一般的なガバナンスと分野（医療、自動運転、教育など）の特性に応じた個別的なガバナンスとの関係、AI の開発者と利用者との責任分担のあり方、規制機関と規制手法のあり方などについて検討することにより、AI エージェントのガバナンスの枠組を構想した。

(3) 結果

①政策、法制度、経済性に関する基礎調査

(a)EU の AI 規則案と AI ガバナンス

EU の AI 規則案の内容と立法事実を検討することにより、AI 規則案の目的、手法、課題を明らかにした。EU の AI 規則案の目的には、プライバシーの保護、差別の防止、データ保護など一般データ保護規則 (GDPR) と重なり合う目的を見出すことができる側面があることを明らかにした。また、AI 規則案は、リスクベースのアプローチを採用しているが、同じ技術であっても、当該技術を用いる主体や当該技術が用いられる目的・文脈により異なる水準のリスク判定が行われていることを明らかにし、AI のリスク評価に当たっては技術の用いられる社会的文脈を考慮することが求められることを確認した。

(b)AI ガバナンスにおける代位責任の活用の可能性

英国における代位責任 (vicarious liability) の法理の展開を検討することにより、飲食宅配プラットフォームの宅配ドライバーに対するプラットフォーム事業者の管理監督のあり方など役務提供者に対する管理監督のあり方とそうしたサービスを提供するシステム全体における責任の配分のあり方を考察した。代位責任の法理は、AI を代理人 (agent) として利用する個人や企業の責任のあり方を検討する上でも有益な示唆を与えるものと思われる。

②既存のガバナンス枠組の網羅的調査、分析、提言

(a)感染症対策と AI ガバナンス

日英を中心に新型コロナウイルス対策のための法規制およびナッジ（選択の自由を尊重しつつ個人の選択を一定の方向に誘導する選択アーキテクチャ）の活用例について調査・検討を行い、AI ガバナンスにおいて法と（ナッジを含む）アーキテクチャの役割分担のあり方を考察する上での示唆を得た。すなわち、日英ともに、感染症対策の初期段階では、ナッジ的な手法が広く試みられ、一定の効果が見られたものの、感染拡大が長期化する中で、しだいに、ナッジ的な手法の限界が顕在化し、強制力を伴う法規制の役割が増大していったという経緯を明らかにした。また、新型コロナウイルスの感染拡大防止のための接触確認アプリの開発や運用のあり方について日本、英国、フランスなどを比較検討することにより、プ

プラットフォーム事業者等の提供する API に基づいて開発されるアプリについては、基盤となるアプリの仕様が頻繁に変更されることなどから、アプリのリリース前の設計・開発のみならず、アプリの運用段階での保守、検証、修正が重要になり、そのための人員や予算を確保することが必要になるという示唆を得た。このことは、データからの学習により運用段階において性質やリスクが絶えず変化していく AI のガバナンスのあり方を考える上でも有用な示唆を与えるものと考えられる。

(b) AI を利用したモデレーションのガバナンス

ソーシャルメディアなどプラットフォーム事業者による AI を利用したコンテンツ削除などモデレーションのガバナンスのあり方について、関連する日米欧の法制度、判例、自主規制を検討することにより、モデレーションにおける適正手続や透明性・説明責任を確保するための裁判所とプラットフォーム事業者の役割分担のあり方を素描するとともに、監督委員会の設置や異議申立て手続きの整備などプラットフォーム事業者内部のガバナンスのモデルも提示した。

(c) 「AI による差別」のリスクのガバナンス

「AI による差別」を防止するためのガバナンスのあり方を探求した。「AI による差別」の原因を整理するとともに、「AI による差別」とされている問題の多くが、従来の社会における差別の構造を AI がデータを通じて学習して再生産することにより生じていることを明らかにした。その上で、「AI による差別」を防止・是正するための法規制、倫理原則・指針、技術標準、業界団体の自主規制、企業のガバナンスの現状を調査し、今後の EU 等における AI 規制の法制化の動きも見据え、今後の課題を整理した。

③利害関係者との議論による国際的ガバナンス枠組の設計

(a) 個人の自律と AI の自律

AI ガバナンスのモデル化を進めるに当たって前提となる個人の自律概念と AI の自律概念との整理を行い、個人の自律がそれ自体で地位または目的としての価値を有するのに対して、AI (エージェント) の自律は人間から与えられた目標を実現するための能力または権限という意味で手段的な価値を有するという知見を得た。また、AI (エージェント) の自律は、専門職集団の自律と同様に、ネットワーク化された多数の AI (エージェント) 間の相互チェックによる集団的自己規律＝自主規制という形で実現される可能性もあるとの見通しを得た。

(b) ガバナンスの設計指針と考慮要素

AI エージェントのガバナンスのモデルを構想・提示するに当たっては、「AI エージェントに対するガバナンス」と「AI エージェントによるガバナンス」との両方のアプローチからガバナンスのあり方を検討していくことが有用であるとの知見を得た。また、AI エージェントを①個人のエージェント (生存する個人のエージェント、死者のエージェント)、②法人のエージェント (企業のエージェント、非営利法人のエージェント)、③国など公的主

体のエージェント（政府のエージェント、地方公共団体のエージェント、国際機関のエージェント）等に類型化することにより、各々のエージェントの性質に適したガバナンスのあり方を具体的に検討することが可能になるとの示唆を得た。その上で、今後の課題として、以上のような多種多様な主体の「代理人」となりうる AI エージェント間の関係を、権力分立論の知見等を活かしつつ、いかに設計するかという問いが獲得された。

(c) 法とアーキテクチャの役割

AI・ロボットに関する国内外の原則・指針や文献を調査することにより、AI ガバナンスにおいて、アーキテクチャのデザインが権利保護や価値実現について一次的なデザインを担う一方、法のデザインは、アーキテクチャのデザインが適切に行われるように法制度をデザインする「メタデザイン」という役割を担うことができるとの見通しを得た。

(d) 開発者と利用者の役割分担と責任

AI の開発者や利用者らと議論しつつ、AI は利用の過程でデータからの学習により継続的に変化する可能性を有していることなどから、開発者が事前の設計により AI のリスクを完全に制御することは困難な一方で、利用者にも AI の学習するデータの提供などにより AI のリスクを制御する上で一定の役割を果たすことが期待できるという知見を得た。かかる知見を踏まえ、AI のリスクを制御する上では、不法行為法や製造物責任法等の立法・解釈を通じて開発者と利用者との適切な責任分担を図ることを通じてインセンティブを付与することが求められるという知見を得た。

(e) 企業における AI ガバナンス

企業における AI とコンプライアンスの関係について、「AI に対するコンプライアンス」（AI が生み出すリスクを対象とするコンプライアンス）および「AI によるコンプライアンス」（AI を活用したコンプライアンス）の両面から検討を行った。「AI に対するコンプライアンス」については、企業の利用する AI の判断・動作がコンプライアンスに適合するように、AI の学習用データの管理・評価、AI の判断・動作の監視・管理などが求められることになるとの見通しを示した。一方、「AI によるコンプライアンス」については、AI を活用したコンプライアンスの可能性を模索しつつも、その限界を踏まえ、AI などシステムのデザインのみならず、事後的な監視・検証やフィードバックの仕組み、AI への適切なデータの学習、従業員に対する研修などライフサイクル全体で複合的なアプローチを取っていく必要があると指摘した。

実施項目⑧ 利害関係者からの直接、間接の情報収集（英国側、日本側）

この項目に関しては④の調査を 2021 年に医療 AI に関連する社会調査を医療関係者、医療研究者、法制度研究者、政府関係者、AI 研究者など 23 人に対して実施した結果に相当する。（詳細は④の説明で記載した通りである。）

実施項目⑨ 利害関係者との議論による国際的ガバナンス枠組の設計

(1) 目的：

AI 関連の研究者、企業、政府機関、市民団体など関係するさまざまなステークホルダーと議論しつつ、国際的なハーモナイゼーションのあり方（普遍的な人権保障と国ごとの価値・文化の多様性とのバランスの確保のあり方など）、AI エージェントに共通する一般的なガバナンスと分野（医療、自動運転、教育など）の特性に応じた個別的なガバナンスとの関係、AI の開発者と利用者との責任分担のあり方について検討し、AI エージェントに関する国際的なガバナンスの枠組の設計指針を提示する。

(2) 内容・方法・活動：

2022 年 7 月から 9 月にかけて 3 回（ステークホルダー、法的技術的問題、今後への指針）にわたって、UNESCO AI 倫理指針に基づき、医療チャットシステムを具体的対象にして、全世界から約 40 人の研究者、実務者を集めてワークショップを英国側プロジェクトが主催し、日本側も共催する形でおこなった。日本側からは 4 名が参加した。以下に詳細を記す。

(2.1) PATH-AI 連続ワークショップ第 1 回 Stakeholder Engagement Process.

開催日 2022/7/15 オンライン(zoom)

内容：UNESCO の AI 倫理指針をベースにして、医療チャットアプリについてステークホルダーの在り方について議論。参加者は、日米欧アジア、アフリカから計 3 2 名(日本からの参加者 4 人：原山優子、バーベリアン山本、江間有紗、中川裕志)

成果（英国側まとめ）：This workshop focuses on who affected stakeholders are, how to identify their specific characteristics and the relevance of these characteristics, and how stakeholders should be engaged as a preliminary part of the overall process of EIA.

We will also touch on positionality – reflecting on who we (the ones conducting the EIA) are, and how our identity may affect our analysis of impacts on others.

(2.2) PATH-AI 連続ワークショップ第 2 回 Ethical Impact Assessment.

開催日 2022/8/15 オンライン(zoom)

内容：UNESCO の AI 倫理指針をベースにして、医療チャットアプリについて PIA（影響評価）について議論。参加者は、日米欧アジア、アフリカから計 3 5 名(日本からの参加者：原山優子、バーベリアン山本、江間有紗、中川裕志)

成果（英国側まとめ）：This workshop focuses on the ways in which EIA may be integrated into the AI project lifecycle and how the UNESCO values and principles may be operationalised.

We will also address the importance of goal setting, assessing severity of impact, and establishing access to remedy.

(2.3) PATH-AI 連続ワークショップ第3回 Readiness.

開催日 2022/9/23 オンライン(zoom)

内容： UNESCO の AI 倫理指針をベースにして、医療チャットアプリについて readiness(実現準備状況) について議論。参加者は、日米欧アジア、アフリカから計30名 (日本からの参加者2人 バーベリアン山本、中川裕志)

成果 (英国側まとめ)： This workshop focuses on exploring the ways in which a readiness methodology may be used to support member states in preparing to implement the Recommendation.

We will address what dimensions of readiness should be central to UNESCO's readiness methodology, including what indicators may be relevant to each dimension, and who should be involved in this process. We will also explore possible approaches to 'levels' of readiness, including participant, organizational, and system readiness.

(2.4) 日本側で行った利害関係者からの意見収集

実施項目④のインタビューなどの社会調査によるデータ取得の節で紹介した社会調査で、医療関係者、医療制度関係者など23人に専門家から得た意見はこの項目に該当する。ただし、読みやすさのために専門家からの意見は実施項目④の説明において一般人の意見と並置する形で比較しつつ説明を行ったので、ここでは詳細記述を割愛する。

(3) 結果：

英国主導で行った3回のワークショップシリーズでは、概要を(2.1)~(2.3)にまとめたが、全世界からの参加者の視点で予想を上回る多様な意見が得られた。また、医療チャットのステークホルダーとして、米国では最高裁判決が中絶の権利を否定したことにより、妊娠中の女性は医療チャット利用においてプライバシーのリスクに晒されている。この例のような多様な意見がだされたが、アフリカでは、そもそもITインフラが十分に整備されていない問題など、AI普及格差に関する意見が表明された。日本側参加者の原山氏から同様の問題がOECDの場でも多く言及されているという指摘があり、AI応用システムの地域間格差が将来的問題として明らかになった。

(2.4)の専門家の意見として、特筆すべきものを以下に記しておくことで結果のまとめをしたい。

- ・ AI を語る時、人工的ネットワークの形成によって「意識」を作り出すことが本当に可能なのか？
- ・ NHK が宮崎駿氏を取り上げたドキュメント番組中で、人間がグロテスクな姿で動くアニメーションをAIに作成させたドワンゴの川上量生氏に対し、宮崎氏は面と向かって「声明に対する侮辱」などと激しく批判した。技術と生命についてずっと考えてきた宮崎氏だから

こそその発言ではないかと感じた。公的な会議の議論にも、もっと個人の視点から疑問を呈したり、意見を言ったりする人の参加が必要ではないか。

実施項目⑩ 日英双方の共同作業による国際的ガバナンス枠組の設計

(1) 目的：

2020～2021 年度に日本側の本プロジェクトおよび英国側のプロジェクトメンバーと zoom で多い時には月 1 回の頻度で議論をおこなった。当時、両国において大きな社会問題であった COVID-19 パンデミックに関する議論を時間をかけて行った。

COVID-19 感染症が社会に与える影響はすさまじく、これは日英両国にとっても共通に課題である。したがって、人工知能技術・倫理的視点・法制度などに関する研究の必要性を日英の双方が認識した。これに基づいて、政府の COVID-19 対策、接触通知アプリについて実情を分析した。この比較分析は、④および⑧における調査質問の一部に COVID-19 の接触通知アプリへの意識調査を入れた。この結果についてはすでに④で記載したが、それ以外の状況について調査、分析するし、問題点を明らかにすることが目的である。この項目は議論は日英双方のプロジェクトの全メンバーが参加し、日本国内事情については中川が調査、分析をおこなった。

(2) 内容・方法・活動：

(2.1) 接触通知アプリ COCOA の狙い

COVID-19 の感染が拡大するにつれワクチンの開発や重症化を防ぐ特効薬の発見と承認に期待が集まった。しかし、ワクチンの開発には治験まで含めると年単位の時間がかかり、特効薬の治験、承認にも下手をすれば年単位の時間がかかりそうであることが明らかになるにつれ、即効性のある感染拡大阻止策として行動制限が注目された。日本では緊急事態宣言による外出等の自粛要請、他国においては都市ごとのロックダウンが行われた。

情報技術を感染拡大阻止に応用する提案も多く出された。代表的でありかつ感染拡大抑止効果がある方法として感染者との接触を通知する接触通知アプリが注目された。COVID-19 の場合は、やっかいなことに感染者が発症以前でも他人に感染させる可能性がある。つまり、外見だけからでは非感染者と見分けのつかない感染者が存在するため、距離を取る、接触しないようにするなどの自衛措置がとれない。接触通知アプリの目的は、この状況を改善することである。すなわち、自分がそれと気づかずに感染者に接触したことを通知してくれるアプリであり、その場合はなにはともあれ PCR 検査を予約する、あるいは体調が芳しくないなら PCR 検査を予約することにより、検査結果が陽性なら入院などの隔離手段をとることによって他人へ感染を防ぐことができる。当然、感染通知アプリはできるだけ多くの人がスマホにインストールしていれば、接触通知が来る確率があがる。さらに通知があったとき、できるだけ多くの人ができるだけ早く PCR 検査を受けることができれば、感染拡大

を抑制する効果が高まる。

感染拡大の指標に1人の感染者が新たに感染させる人数の平均値である実効再生産数 R_0 が使われる。 $R_0 < 1$ なら感染者数は減少に転じる。Ferretti[1]の Fig 3 に、 $R_0 < 1$ になる条件が感染通知アプリの接触感知成功率:y 軸と感染通知を受けた人が検査の結果感染が判明して隔離される確率:x 軸の平面においてどの範囲になるかをシミュレーションで求めた結果が示されている。仮に感染者接触が100%成功するとすれば、y 軸はアプリの普及率と等価になる。Ferretti[1]の Fig 3 は簡略化すると図 10.1 のようになる。ここで①②③の曲線は以下の各場合における $R_0=1$ となるシミュレーション結果である。

- ① 感染通知を受け検査して感染が判明した人が隔離されるまでの日数が2日
- ② 同上の隔離までの日数が1日
- ③ 同上の隔離が即日

曲線の上側の領域では $R_0 < 1$ 、下側では $R_0 > 1$ である。

図 10.1 から、一般に言われる普及率が60%という基準は、現実的な②のケースにおいて通知された人の70%が検査を受け、1日以内に隔離される場合に $R_0=1$ を達成できるという意味である。ただし、これは R_0 を減らす効果が接触通知アプリだけに限定された場合であり、その他の自主的にうける検査や各人の手洗いなどの自衛措置は考慮していない。

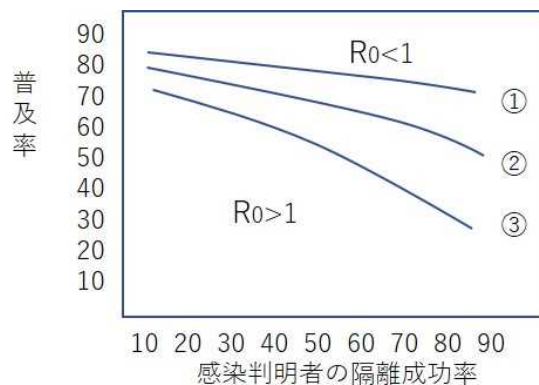


図 10.1 感染通知アプリによる実効再生産数 $R_0=1$ の変化

図 10.1 の結果をだすシミュレーションでは感染者は全国に一様に分布し、接触通知アプリをインストールしている人も一様に分布するモデルが使われていることにも留意しておこう。

接触通知アプリが上で述べた狙いと効果の数理モデルを持つことを念頭においたうえで、主に諸外国および日本における接触通知アプリの現状と問題点について検討する。

(2.2) 諸外国の状況

COVID-19 の脅威が知られて以降、ヨーロッパ、東アジアの諸国は特効薬やワクチンが

ない状況で接触通知アプリを有力な感染拡大抑止ツールと考え、国ごとに種々の試みをしてきた。

(2.2.1) ヨーロッパの状況

ヨーロッパでは当初、中央集中型システム PEPP-PT[2]が提案された。このアプリの参加者が ID を中央サーバに送ってデータベースに登録し、その登録によってさらに近くにいた参加者も登録される。感染者が自分の ID を登録すると近くにいた参加者に接触を通知するのだが、その場合のプライバシー保護を強化するために自分のスマホで頻繁に更新するハッシュ値を個人 ID として中央サーバに送る個人 ID 分散型のシステム D3PT[3]が提案された。Google と Appel は、プライバシー保護をより徹底した Android 端末および iPhone で稼働する API (以下では G&A と呼ぶ) を共同開発して発表し[4]、ヨーロッパの多くの国で利用されている。これらは、いずれもスマホ間の通信には Low energy Bluetooth を用いている。

(2.2.2) G&A

G&A の近接者の接触符合生成と Bluetooth によるブロードキャスト、感染した場合の通知サーバへのアップロードなどのデータの生成と流通の流れを図 10.2 に示す[5]。

スマホ固有の一意のトレース・キーから日々変更される日次キーの生成、日次キーから 15 分毎に生成される接触符合の生成はハッシュ関数によって行われる。この接触符合は 1 メートル以内に 15 分間以上一緒にいた人同士で互いに相手のスマホに Bluetooth で送信される。感染すると、この接触符合から作られた診断キーが通知サーバに送られる。個々の通知アプリ参加者のスマホには感染者の診断キーが通知サーバからダウンロードされ、診断キーから接触符合が再導出される。再導出された接触符合と自分が過去 14 日間に 1 メートル以内に 15 分間以上一緒にいた人の接触符合の中に一致するものがあつたかどうかを判定する。一致した場合は、スマホに感染者と過去に接触したことが表示される。

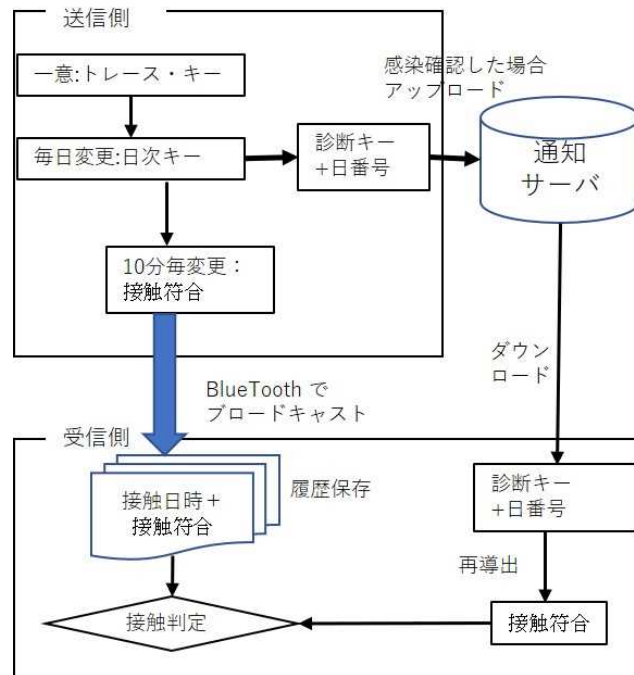


図 10.2 G&A の処理のブロック図

この仕組みの範囲内であれば、G&A の利用者にインターネット経由で知らされるのは、感染者の 15 分毎に変更される近接 ID だけである。このため、感染者本人についての手掛かりは、ほとんどない。G&A は接触通知アプリをインストールしているスマホの位置情報、さらにはスマホの持主の情報も扱っていないため、個人情報の保護はほぼ達成されている。

G&A では接触したという事実だけしか扱わないが、他の接触通知アプリは種々の個人データを扱っていることもある。たとえば、位置情報、誰と接触していたか、など機微な情報が含まれる。したがって、政府が個人の行動監視を行う極めて強力な手段となるため、COVID-19 も政府が感染拡大阻止以外に目的に使うかもしれない。EU の研究者などは、これが後々の政府による国民監視ツールとして使われることに強い懸念を持っている。

(2.2.3) 東アジアの状況

中国、韓国、シンガポールの状況については情報処理学会の小特集「データ・AI ガバナンスと COVID-19: アジアにおける中長期的展望」[6][7][8][9]に詳細が記載されている。この特集によれば、各国ではおおむね以下のような状況が報告されている。

(2.2.3.1) 中国の状況[6]

中国のヘルスコード（健康碼）は、中国のパンデミック時における公衆衛生の緊急管理支援の最も重要な情報基盤である。ヘルスコードによる接触追跡は、主に GPS と公共交通機関やその他の位置情報サービスからの関連情報に基づく。中国のヘルスコードは義務化されており、すでに職場や学校の移動管理にヘルスコードが広く使われている。

中国では、「自己」は共同体、文化と社会の一部であり、公衆衛生危機緊急管理システムに

よる個人情報へのアクセスは、強制加入するように設計されており、自発的な参加は受け入れられない。なぜならば、参加を望まない人は周囲の人々の安全に対する潜在的リスクとなるとみなされるからである。

ただし、ヘルスコードによるシステムの無制限な拡大には異論もあるらしい。杭州保健省から出されたパンデミック発生後のヘルスコードの利用の拡大提案は、反対意見が多数表明され、具体化には至っていない。中央集権的な体制の下でプライバシーがどこまで保護されるかは、他国からは不透明であることを示す例かもしれない。

(2.2.3.2) 韓国の状況[7]

韓国では 2015 年に中東呼吸器症候群 (MERS) に見舞われたときに、情報技術を利用した接触追跡の法的枠組みをすでに整備していた。これによって経済に負担をかけることなく、医療従事者と疫学調査官の取り組みを情報技術インフラを以て支援し、新規感染者数と死亡者数は 3 月中旬までには落ち着いた。

感染者として隔離された人々は、「自己隔離者安全保護アプリ」と呼ばれる携帯アプリのインストールが義務付けられている。このアプリを使えば、隔離された人のスマートデバイスの GPS データをリアルタイムで追跡でき、隔離場所にとどまっているかを確認できる。このようなネットワークを利用した集中型の追跡アプローチを導入する際の大きな障害となっていたのが、個人情報保護法をはじめとするその他のデータ保護法であったが、MERS 発生時の経験から、感染爆発発生時には感染が確認された人、およびその接触者に関する位置情報、個人識別情報、医療と処方箋の記録、出入国管理の記録、クレジットカード、デビットカードおよびプリペイドカードの取引情報、公共交通カードの使用記録、および監視カメラ (CCTV) の映像などの類の情報を令状なしで収集できるようになった。疾病管理本部はこのデータを用いて感染拡大に対処するための包括的な追跡をできる。さらに 2020 年 3 月 26 日に開始した「コロナ 19 疫学調査支援システム」によってリアルタイムに感染者情報を疫学調査官に提供する。

位置情報による迅速なプロファイリングによって疫学調査官は接触者を迅速に追跡して潜在的な感染者を特定し、隔離することに注力できる

このような中央集中型のほうが一定の利点を持っている可能性が示されている。これは韓国が民主化達成後に非常に厳格なデータ保護とプライバシー法体制を作り上げてきたことによる国民の信用があるからだとして [7] では、主張している。さらに、2020 年 2 月に、プライバシーおよびデータ保護法の大規模な改正を行い、データ保護権限は個人情報保護委員会に統合された。個人情報保護委員会は、データ主体の権利と保健の目標とのバランスをとるためにより積極的な役割を果たすことができるようになった。

(2.2.3.3) シンガポールの状況[8]

シンガポールの TraceTogether アプリは携帯アプリで、必要に応じて、接触追跡を支援する。TraceTogether アプリのダウンロードと有効化は任意である。しかし、高い感染リスクの地域に住み、働いている外国人労働者は、アプリの最新版をダウンロードし有効化して

おく必要がある。

TraceTogether アプリでは、低消費電力の Bluetooth Low Energy(BLE)上の一時的な ID を示すことで動作する。TraceTogether をインストールして有効化されている 2 台の端末が BLE 範囲内にある場合、相互に検出し、検出結果の記録はその後、利用者の携帯電話端末内に保存される。TraceTogether アプリでは、ウイルスに晒された可能性のある人を特定しようとしているのであって、ウイルスにどこで晒されたかを特定しようとはしていない。

シンガポールでは自分の位置を各場所に張り付けられた QR コードによってサーバに知らせるシステムも稼働しており、地下鉄を使用する場合にも必要となった。QR コードによる安全な入館アプリを通じて蓄積されたデータは、政府によって一元的に保存されるが、潜在的な潜伏期間（14 日間）終了後に廃棄されることが保証されている。

シンガポールの人々は、政府の追跡計画におおむね従っているが、TraceTogether の利用者は効果を発揮するのに必要な人数には達しておらず、ソーシャルメディアには特定のデータ共有の可能性に対する批判がある。シンガポール人は伝統的に、個人の健康データのプライバシーを特に心配している。

なお、外国人労働者に対して、当局は完全隔離を強制し、それに伴って感染の潜伏率も上昇した。この背景には、一般の人々への拡散拡大を防ぐという意図があった。この意図は現実的であるが、不公平であると見なされる可能性があるとして [8] では指摘している。

(2.3) 日本における接触通知アプリの発注までの経緯

ダイヤモンド・プリンセス号の乗客の COVID-19 感染が大きく報じられ、COVID-19 の感染症としての悪辣さ、すなわち高い感染力、発症以前に他人に感染させること、ワクチンや治療薬がないことなどをうけて、2 月から各所で情報技術を用いた感染防止策、とりわけ接触通知アプリの研究開発が開始された。

日本における著名な開発は、Code for Japan と OSS の 2 か所で行われていたようである。政府においても接触通知アプリを国の費用で開発するプロジェクトが始まり、当初は複数の接触通知アプリの併存も考えていたようである。

ところが Google と Apple は、G&A を用いる接触通知アプリは 1 国に 1 システムしか認めないと宣言した。そのため日本政府は、政府調達を G&A の 1 本に絞るか、G&A を採用せず複数のシステムを共存させるかの選択を迫られた。ここで決定的だったのは、日本ではスマホ利用者の約半数が iPhone を用いていることである。G&A を採用しなければスマホ利用者の半数は接触通知アプリが使えない可能性があるとの理由で政府調達は G&A に 1 本化し、新型コロナウイルス接触確認アプリ COCOA (COVID-19 Contact - Confirming Application) と命名した。

iPhone 普及率の高さを念頭においたこの 1 本化の決定は、図 10.1 に関連して述べた $R0 < 1$ を達成するためのアプリ普及率 60% に近づけたいという視点からは妥当である。し

かし、この接触通知アプリは、」 Ferretti[1]で述べられているように感染可能者を早期発見し、早期検査で感染が判明したら隔離するという目的ではなく、通知による「感染しない・させない」という行動変容を促す目的であると仕様書[10]の 4.2) ③に書かれている。だとすれば、普及率 60%に拘って iPhone 利用者の取り込みを図ったという判断理由は薄弱になってしまう。

1 本化の結果として、独自に接触通知アプリを開発していた Code for Japan は外され、G&A を用いるシステムを開発していた OSS が受注先となった。OSS はソフト開発の有志であるため、政府発注の対象には直接はなれないようであり、OSS のメンバーが所属していた大手ソフト会社が受注したような報道がされ、外部からは実態が把握しにくい状態であった。

(2.4) 接触通知アプリの仕様と問題点

図 10.3 に仕様書[10]に示された接触通知アプリ COCOA における処理の流れを示す。図 10.3 では通知サーバは図 10.2 と同じであり、図 10.2 の発信者と受信者は各々図 10.3 の左側の PCR 感染判明者、右側の近接した情報を受けた人に対応する。図 10.3 の 5.の説明中の「近接した情報」は図 10.2 の診断キー+日番号に対応する。また、通知サーバに登録された診断キー+日番号のデータは COVID-19 の感染者接触から発症までの期間を過ぎれば無意味である。よって発症までの最長期間と考えられている 14 日が経過すれば、登録された診断キー+日番号は通知サーバから消去される。

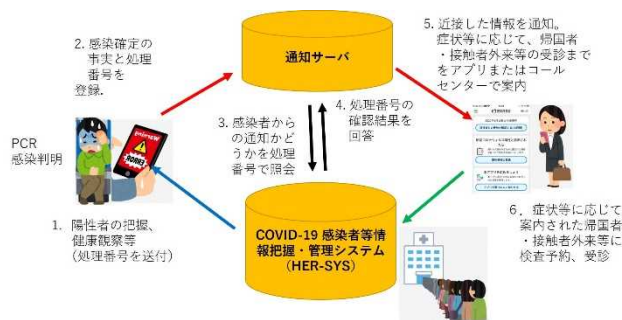


図 10.3 COCOA の処理の流れ

図 10.2 に比べて図 10.3 で大きく異なるのは COVID-19 感染者等情報把握・管理システム HER-SYS (Health Center Real-time Information-sharing System) である。HER-SYS には保健所など検査を実施、管理している医療機関から検査を受けた人の詳細な医療情報が登録されている[11]。よって HER-SYS は要配慮情報を内容に持つデータベースである。

COCOA では、PCR 検査で陽性と判明した者は、自分自身で通知サーバに自分が感染者であること自分で登録するルールになっている。これは G&A を採用している以上、やむを

得ない。なぜなら、通知サーバに登録する診断キーは感染者のスマホにしか存在しない情報であり、これを他者と共有することはできない。また、PCR 検査の結果が分かるまでに 1 日程度の時間がかかるという制約もある。

感染者の通知サーバへ登録するかどうかは感染者自身の判断に委ねられている。したがって、感染が判明しても通知サーバに登録していない人も多数存在する状態になってしまうという状態は認識しておく必要がある。

このような通知サーバへの登録ルールであるため、陽性でない人が悪ふざけで、あるいは悪意で通知サーバに登録することを防ごうと設計者は考えたようである。このために、通知サーバに登録しようとした人は、PCR 検査を受けていることを確認させるために、登録に際しては HER-SYS から処理番号をもらう。この処理番号を使って通知サーバに登録するという情報の流れになった。

この考察から分かることは、1) 虚偽の通知サーバへの登録を普通は行わないだろうという性善説に立つ、ないしは 2) 虚偽報告が相当数あっても接触通知の回数が増えるだけであり、実害は少ないと腹をくくる、と考えることができれば HER-SYS を介在させる必要はなかったといえる。

HER-SYS は、パーソルプロセス&テクノロジーが受注したシステムである[12]。さらに COCOA が、HER-SYS と連携するという設計にしたため、OSS が開発する COCOA の工程管理をパーソルプロセス&テクノロジーが請け負うという外部からはわかりにくい構造の開発体制になった。COCOA の公開は 2020 年 6 月 19 日であった。しかし、HER-SYS との連携部分に原因があると推測されるバグが何回も発生し、利用の中止、再開が繰り返された。開発期間は約 1 か月と非常に短かったことを考えると、個々の開発者は最大限の努力をしたと思われるが、HER-SYS とのリンクを含め、システムを複雑な構造にし、別々の組織が開発するという体制にしてしまったことは疑問が残るところである。

HER-SYS は医療情報という要配慮個人情報を含むデータベースだけに、図 10.3 のように若干にせよ COCOA 情報の流れで関係があると、個人情報保護の点から不安がある。仕様書[10]に対する有識者会議の意見[13]によれば、[13]の 2.に HER-SYS とのリンクによる診断キーに関して概略以下のような記述がある。

引用ここから（一部省略および加筆をしている）：

（図 10.3 の）処理番号と診断キーが（通知）サーバー内で紐付けられることにより、本アプリ運営者等において、（処理番号を介して）診断キーがどの陽性者に割り当てられたものであるかを把握できるような場合には、診断キーも個人情報となる。個人の診断キーが通知サーバーに送信されるのは、当該個人が陽性者であると医師等に診断され、感染者システムから発行された処理番号をアプリ上で適切に入力した場合のみであることから、当該診断キーを鍵として、陽性診断を受けた特定の個人を識別できることとなる。したがって、診断キーが行個法の個人情報にあたる場合には、当該診断キーは要配慮個人情報に該当す

ると考えられる。本アプリでは、処理番号を陽性者の認証終了後直ちに通知サーバーから削除することとしていることから、本アプリ運営者等が、処理番号を介して、通知サーバー内で陽性者と診断キーを結びつけることは困難と考えられるが、もしこれが可能である場合には、診断キーについても行個法上の義務が及ぶ。

引用ここまで

このように、G&A では個人情報とは仕組み上排除していたが、HER-SYS と連携することによって、COCOA では要配慮個人情報を扱うことになってしまった。COVID-19 を扱う医療従事者の子弟が保育園などで不当な扱いを受けているというニュースも報道されている現状、および感染者が不当な扱いを受ける可能性を鑑みると、HER-SYS と COCOA においては個人情報の漏洩を徹底的に防ぐ安全管理措置が必要である。ところが HER-SYS は厚労省の管理下にあるものの個人情報データベースのデータ管理者としての責任をきちんと果たしているかは明らかでない。HER-SYS は複雑なシステムであるだけに、その内容やデータ収集方法に通じた技術を持つデータ管理者が設定できていなければならない。

この芳しくない状況は、日経 XTECH の記事[18]に報告されているように、2020 年 9 月 10 日に厚労省がやっと改善に動いた。具体的には、(1) 不必要の多くの情報入力及要求されていること、(2) HER-SYS にアクセスログを抽出できる機能がなく、要配慮個人情報が格納されているデータベース HER-SYS への不正なアクセスを当事者である自治体の保健所から監視できなかったことが問題だった。当初は HER-SYS の問題は無視されていたようだが、港区のみなど保健所の当事者からの要望によって、上記の(1)(2)の問題の改善が実現した。このように現場からの意見に耳を傾ける姿勢は中央官庁としては稀有の例であろうが、今後とも続けてほしい。

(2.5) 接触通知アプリ：COCOA の実情

COCOA は 2020 年 6 月 19 に公開以来、バグの発覚などの紆余曲折を経つつも徐々にインストール数は増えてきた。厚労省からの発表によれば 2020 年 8 月 24 日 17 時時点でのダウンロード数が 1,464 万件であり、全国民の 10%を超えた。ただし、接触通知アプリのみによって $R0 < 1$ を達成するためのインストール率 60%には程遠い。陽性登録は、386 人であり、同日の感染治療中の人数 11211 人の 3.44%にとどまっている。10%を超えるインストール率を考えれば、インストール者がバイアスなく分布したと仮定すると、陽性判定を受けた人の一部しか通知サーバに登録していないことになる。これらを考慮すると、実際は COCOA で接触通知が来た件数の 20 倍以上の人数の感染者と接触している可能性がある。その後、厚労省の発表によれば、9 月 30 日現在、約 1,778 万人（陽性登録数は 948 人）、10 月 23 日現在、1,877 万件（陽性登録件数は 1302 件）である。上昇は緩やかといえよう。また、陽性登録者数が増加したため、感染者と接触している人数は、COCOA で接触通知が来た件数の 10 倍程度になってきたのではないと思われる。

仕様書[10]によれば COCOA の目的が接触追跡ではなく利用者の行動変容であるが、それにしても、行動変容の効果が十分に上がる利用状況とは言いにくい。

仕様書[10]によれば接触通知があった人には「帰国者・接触者相談センターへの相談方法等をメッセージによりガイダンスする」と書かれているだけで、PCR 検査を優先的に受けることができると書かれてはいない。仕様書の文言「個人が自らの行動変容を意識する」によれば、感染者と接触したという通知が来ても、それ以上の支援はないので自らの感染可能性を考えざるをえないという意味で通知された人の恐怖感を煽るだけになってしまう。実際、当初は接触通知が来ても PCR 検査を優先的に受けられる手立ては講じられていなかったし、その後若干の改善はあったものの、[14]に書かれているように、8 月末でも接触通知があった人の 8 割は検査を受けることができていなかった。この状況ではさすがに拙いと政府も感じたらしく、接触通知された人には PCR 検査を受けられる体制を整えるという判断が 8 月 21 になされ [15]、その中には各自治体等に向けて以下のような要請文が書かれている。

「本アプリで通知を受けた者に対して検査を行う場合は、症状の有無や濃厚接触者に該当するか否かに関わらず、行政検査として取り扱っていただくよう、お願いいたします。」

それまでの PCR 検査を受ける資格者に関する政府の方針は、当初、海外からの帰国者に限定されていたものが、感染の拡大につれて高熱が続いた者へとなし崩し的に拡大していたように見えるが、ここに至り COCOA の接触通知者にも制度的には拡大した。しかし、この通知以後も PCR 検査を受ける容易さはすぐに大きく改善したわけではない。上記の厚労省の通達があくまで「お願い」であることにも若干は起因するだろうが、以下のような現実的問題があるため、文書としての要請では現場が動けず、事態は改善していない。

- (1) 担当する保健所のキャパシティが人的にも物理的にも不足している。
- (2) COCOA からの接触通知が来たから、PCR 検査を要求する人が保健所に押しかけ、その対応が保健所にとって大きな負担になり、本来業務に支障がでる。
- (3) COCOA のバグはまだ完全に切り切れていないことから発生する誤通知で無意味な検査を誘発している。
- (4) PCR 検査は高度な技術が必要であり人数的拡大が容易ではないこと、および COVID-19 の高い感染力から検査側の人材を守らなければならないため、物理的に拡大が難しい。

もちろん、COCOA から接触通知を受けた人は即時に PCR 検査を受けることができるという処置を保健所に限定せず、保健所以外の医療施設の拡充などを行うことによって実効性をもって実現できれば、COCOA の普及率が高まることが期待できる。

この文章を書いている時点(2020 年 9 月)においても、COCOA から接触通知を受けたことを保健所や指定された支援センターに伝えても、強い症状がない場合は、「様子をみてください」で済まされてしまうことが多いと聞く。

最終的には 2022 年 10 月に COCOA の運用を中止することが政府から示された。

一方で韓国では MERS のときの経験を活かしてドライブスルー型の検査などが整備されているのに比べると、MERS の経験がなかった日本では、パンデミックにおける医療的な対応能力が弱いという本質的問題が浮き彫りになっている。

(3) 結果：

以上述べてきたように日本の接触通知アプリ COCOA は利用目的、開発の進め方、利用実情のすべてにおいて当初の接触通知アプリの狙いから相当に捻じれた状況になってしまっており、有効活用には至らず、2022 年秋に終了となった。

とはいえ、せっきく開発し、国民の認知度も上がってきた現状を踏まえて、次に起きるかもしれないパンデミック時において、COCOA のようなシステムをが効果を発揮する方向性をまとめることで、この項目の結果としたい。

[16]によると、長時間いっしょに過ごすため、接触する確率が高いのは、(1)家族、(2)同一職場に勤める人々、(3)同じ学校に通っている学生、(4)近い年齢の人々であることが実験的に示されている。(4)は付き合う相手は年齢が近い人が多いからであろう。この状況を[16]に示されている中国における具体的データに基づく分析を簡略化して図 10.4 に示す。

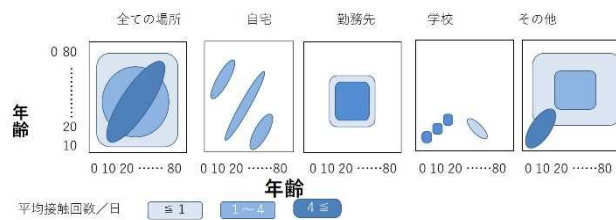


図 10.4 中国における接触可能性の年齢と場所の関係シミュレーション結果

したがって、(1)~(4)の各々の中で接触通知アプリをインストールしている人が多ければ、高い確率で接触検出できる。

職場や学校で感染者が出ても、組織内に人数が多いと、感染者が誰であるかは分かりにくい、あるいはあえて伏せているという状況が考えられる。そこで、以下を提案する。

◆同一職場に勤める人、あるいは同じ学校に通う人には接触通知アプリのインストールを組織として要請する。

これは、社則や校則で決めることができるかもしれないし、お願いベースであるにしても同調圧力によって高い割合の人々が接触通知アプリをインストールすることが期待できる。職場や学校で接触通知アプリのインストールを組織として要求することが人権あるいは個人情報保護法の観点から疑問視されるかもしれない。一つの理由は扱っているのが病歴という要配慮個人情報ということだろう。しかし、図 10.2、図 10.3 に示したように、HER-

SYS 以外の部分では個人情報漏洩しないので、この心配は不要だろう。となると接触通知アプリのインストールの諾否は、「プライバシー権 対 公衆衛生維持」の問題になる可能性がある。接触通知アプリのインストール拒否が、職場や学校での感染防止よりも上位にあるという比較衡量は covid-19 の感染状況では少なくとも社会的に受け入れにくいものである。

COCOA は位置情報を全く使わなかったが、店やイベント会場に QR コードを割り振って位置情報を使う図 10.5 のようなシステムも稼働している。利用者：A さんは予めこのシステムの利用者として LINE に登録しておき、訪れた店などの QR コードをスマホに入力しておく。後にその店で感染者が出た場合は、店・保健所・LINE お知らせシステムという順に情報が伝わり、最終的には利用者：A さんに感染者が居た店を訪れたことが LINE のメッセージで通知される。

上記のように、接触する確率が高いのは、(2)同一職場に勤める人々、(3)同じ学校に通っている学生、(4)近い年齢の人々と分かっているのが、さらに飲食店やイベント会場も接触の確率は高いと予想されるので、QR コードと LINE のシステムは効果的なカバー範囲を持つ。

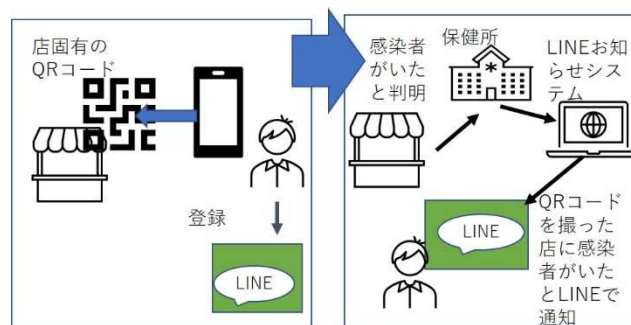


図 10.5 QR コードと LINE を組み合わせた接触通知アプリにおける情報の流れ

このシステム[17]では QR コードが付いている店の位置情報が分かるので、感染者個人ではないにしても位置情報は店の QR コードとして知られることになる。このシステムは LINE という SNS 媒体を利用しており、簡便だが、接触通知アプリとしては十分な機能を持つと言えよう。東京都、神奈川県などの自治体単位で導入され稼働している。しかし、自治体毎の個人情報保護の条例が異なるため、システムを自治体毎に別建てで作る必要があり、結局、導入は少数の自治体にとどまった。個人情報保護法で問題視されてうる自治体毎の条例が異なるという 2000 個問題の悪い影響を被った例である。2000 個問題は、令和 3 年の個人情報保護法の改正で解消する方向が示された。よって、図 10.5 のような簡便かつ実用的なシステムが活用される可能性は従来よりは高くなったといえよう。

参考文献

[1] Ferretti, Luca, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie, Abeler-

Dörner, Michael Parker, David Bonsall, and Christophe Fraser. 2020. “Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing.” Science, March. <https://doi.org/10.1126/science.abb6936>.

[2] High-Level Overview, Pan-European Privacy-Preserving Proximity Tracing. <https://www.pepp-pt.org/> access May 18,2020.

[3] Decentralized Privacy-Preserving Proximity Tracing -- Documents: DP3T White Paper.pdf, <https://github.com/DP-3T/documents/> access May 18 2020.

[4] Google: Contact Tracing, Framework Documentation(API) Preliminary - Subject to Modification and Extension, April 2020.

[5] Google: Contact Tracing Bluetooth Specification Preliminary - Subject to Modification and Extension, April 2020.

[6] 曾毅／孙康／鲁恩蒙: 壊滅的なリスクに対抗するための倫理とガバナンスの展望: COVID-19 から汎用人工知能の長期的な安全問題まで. 情報処理,2020, Vol.61 No.10, p.1020-1024.

[7] 朴相徹／林龍: COVID-19 に取り組むための技術活用: 韓国からの教訓. 情報処理,2020, Vol.61 No.10, p.1025-1030.

[8] マーク・フィンドレー (Mark Findlay) : シンガポール COVID-19 制御—2 つの都市の物語?. 情報処理,2020, Vol.61 No.10, p.1031-1038.

[9] 江間有沙: COVID-19 対策から見えてきた日本の AI/データ利活用の課題とガバナンスの展望. 情報処理,2020, Vol.61 No.10, p.1039-1045.

[10] 新型コロナウイルス感染症対策テックチーム:接触確認アプリ及び関連システム仕様書, 2020年5月26日.

https://cio.go.jp/sites/default/files/uploads/documents/techteam_20200526_01.pdf (参照 2020-9-13).

[11] 厚生労働省:新型コロナウイルス感染者等情報把握・管理支援システム(HER-SYS) : Health Center Real-time information - sharing System on COVID-19.

https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00129.html (参照 2020-9-15)

[12] 岡林 凜太郎: 政府の接触確認アプリ,厚労省がパーソルプロセス&テクノロジーに発注. 日経クロステック/日経コンピュータ. 2020年6月16日.

<https://xtech.nikkei.com/atcl/nxt/news/18/08122/>

[13] 接触確認アプリに関する有識者検討会合: 「接触確認アプリ及び関連システム仕様書」に対するプライバシー及びセキュリティ上の評価及びシステム運用上の留意事項. 2020年5月26日.

https://cio.go.jp/sites/default/files/uploads/documents/techteam_20200526_02.pdf

[14] 野口 悠紀雄：厚生労働省の IT システムは、なぜこうも不具合が多いのか？（現代ビジネス）. 2020 年 9 月 6 日.

<https://news.yahoo.co.jp/articles/c8a8d13b8bd71e3f48a83243e8aac24f5ac89d10?page=1https://jipsti.jst.go.jp/sist/pdf/SIST02-2007.pdf>, (参照 2020-9-13).

[15] 行動と厚生労働省新型コロナウイルス感染症対策推進本部:新型コロナウイルス接触確認アプリ（COCOA）で通知を受けた者に対する行政検査等について.2020-8-21.

<https://www.mhlw.go.jp/content/000661724.pdf>

[16] Kiesha Prem, et.al. :The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China:a modelling study, Lancet Public Health 2020;5:e261-70, Published Online March 25, 2020. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6).

[17] 神奈川県：感染防止対策取組書・LINE コロナお知らせシステム, 2020 年 7 月 2 日, <https://www.pref.kanagawa.jp/docs/ga4/corona/osirasekenmin.html>

[18] 大豆生田 崇志:新型コロナ対策のデータ活用に遅れの懸念,感染者情報の管理システムに不備と不信感,日経 XTECH. 2020 年 9 月 18 日.

<https://xtech.nikkei.com/atcl/nxt/column/18/00989/091500035/?fbclid=IwAR19aDxM3Bdy1w5la9TAbPb6ZVy42yvjvMOdcYmCrU2aa7eg-aifTCW5n-jl>

実施項目⑪ 国際的ガバナンス枠組に関する会議開催

(1)目的：

この項目に関しては、表記の国際的ガバナンス枠組に関連するトピックに関して、英国側のプロジェクトが主導してプロジェクト以外の参加者を集めたオンライン会議を開催した。ボードミーティングでは、プロジェクトの内容に関する意見を求めることが目的であり、もうひとつの一般の参加者を募る国際ワークショップでは、プロジェクトの成果を発信することが目的である。

(2) 内容・方法・活動：

(2.1) 英国側の共同研究グループの 2021 年第 1 回 ボードミーティング:

開催日 2021/5/10 （オンライン）

内容：この時点までの成果、および今後の予定に関して英国側プロジェクトの評価者が参加する board meeting（第 1 回）が行われ、中川、江間が日本側代表として参加し、現状報告した。

(2.2) 英国側の共同研究グループの 2021 年第 2 回ボードミーティング:

開催日 2021/5/15 （オンライン）

内容：第 1 回と同じ内容だが、参加者は第 1 回に都合がつかなかった英国側プロジェクトの

評価者であった。中川が日本側代表として参加し、現状報告した。

(2.3) 日英共同の第1回国際オープンワークショップ：AI ガバナンスに関する報告と議論

開催日 2021/6/14 (オンライン)

内容：この時点までの研究成果を発表した。(日本側発表者：中川裕志、成原慧、英国側発表者：David Leslie、James Wright、 Charles Raab、 Fumi Kitagawa)

(2.4) 英国側の共同研究グループとの共催による 2022 年第1回 ボードミーティング:

開催日 2022/1/27 (オンライン)

内容：この時点までの成果、および今後の予定に関して英国側プロジェクトの board meeting (第1回) に中川が日本側代表として参加し、④で述べた医療 AI の社会調査の結果を報告した。

(2.5) 英国側の共同研究グループの 2022 年第2回 ボードミーティング:

開催日 2022/1/28 (オンライン)

内容：第1回と同じ内容だが、参加者は第1回に都合がつかなかった英国側プロジェクトの評価者であった。中川が日本側代表として参加し、④で述べた医療 AI の社会調査の結果を報告した。

(2.6) 日英共同の第2回国際オープンワークショップ：医療 AI の社会調査報告

開催日 2022/3/15 (オンライン)

内容：この時点までの研究成果を発表した。(日本:中川裕志、英国:James Wright)、Panel discussion :AI governance、 moderator: David Leslie、パネリスト:James Wright、Charles Raab、 Fumi Kitagawa、佐倉統、大屋雄裕、成原慧

(3)結果：

上記の会議は本研究の成果をアウトリーチする機会として機能した。また、各年度に行われたボードミーティングでは、参加したボードメンバーから研究の結果や進め方にかんする有意義な意見をえることができ、その後の研究に資するところが大きかった。

実施項目⑫ ⑪の会議に結果の出版などによるアウトリーチ

(1)目的：

当初計画では⑪の会議の内容のアウトリーチとしていたが、会議では本プロジェクトの主要な内容を発信したので、この項目では本プロジェクトの成果全体に関する公開イベント、出版、論文、報道などを通して、社会にむけて発信を目的とした。この発信はプロジェクト参加者が複数で協力する、および個別の活動として行った。

(2) 内容・方法・活動：

- 英国側は The Alan Turing Institute により成果報告ページを掲載し、成果報告書を掲載している。
- 日本側は 公開シンポジウム、論文著書の公刊、代表者のホームページで成果公開などをおこなった。
- 以下の JST-RISTEX の ELSI 領域のプロジェクトとの合同公開シンポジウムを行った。人間-AI エコシステム・プロジェクト&空飛ぶクルマ プロジェクト合同公開シンポジウム： JST-RISTEX HITE 領域：人間-AI エコシステム（代表：中川裕志）、JST-RISTEX ELSI 領域：「空飛ぶクルマ」の社会実装において克服すべき ELSI の総合的研究（代表：小島立）日時：2022年2月8日 13時～16時
オンライン webinar 開催
- 出版としては、6-1 節の書籍、6-2 節の論文、6-3 節の口頭発表、6-4 節の報道に記載した活動をおこなった。

(3) 結果：

(2)で記した種々のアウトリーチ活動を通じて、研究成果の社会への還元を測ることができた。

出版のひとつである、データサイエンス入門シリーズ『教養としてのデータサイエンス』は統計学会出版賞を受賞し、論文 AI 倫理指針の動向とパーソナル AI エージェント、総務省 学術雑誌『情報通信政策研究』 第3巻第2号および関連活動で令和2年度 情報通信月間推進協議会 情報通信功績賞を受賞し、社会的意義が認められたといえよう。

3. 研究開発成果

3-1. 目標の達成状況

- ・ AI と文化
 - 企業などが AI を開発する参考になる知識となるので、AI システムの開発、運用を手掛ける企業で本研究成果を活用してもらえらるアウトリーチを行った。特に日本独自の文化的背景の調査分析に関しては、論文、書籍の執筆による発信に加えて、佐倉チームが中心となり、現在の技術状況を踏まえた AI と文化、社会背景、自動運転などの実応用をテーマにした公開シンポジウムを 5 回開催した。この成果の企業活動への影響はまだ明確なエビデンスとして出てきてはいないが、佐倉チームの活動のひとつとして ISO のロボットインタフェースに関する国際標準化活動は、企業活動への影響が期待できるものである。
- ・ パーソナル AI エージェント
 - パーソナル AI エージェントを独自のソフトウェアツールとして開発する組織が現れることに関しては、HITE 領域の橋田プロジェクトとの交流や議論を通じて、類似する基本的なアイデアを実現され、自治体などでの適用が進んでいることは間接的ながら寄与できたと考えられる。
 - パーソナル AI エージェントの考え方を IT プラットフォーム、情報銀行などの個人データ取り扱い事業者が利用者インタフェースの一部に取り入れることを期待し、本研究の参加者が兼職している政府委員会での発信、企業に対する倫理関連の委員会委員としてコンサルティングなどを行い、また学会活動としては発表、論文公開などを行ったが、総務省からの受賞などで評価をいただくことができた。
 - さらに、本人の死後、その人の代理をしていたパーソナル AI エージェントをどのように扱うべきかという問題を調査し、社会調査によって許容される扱い方を知ることができた。この結果は国際ジャーナルに掲載された。
- ・ トラスト
 - パーソナル AI エージェントを社会的に利活用する場合に必要な本人とその代理をするパーソナル AI エージェントの間のトラスト、およびパーソナル AI エージェントとインターネットを介した他者との間のトラストの構築方法を検討した。ID 認証が基本となるものの、乗っ取りやなりすましの危険は常に存在するため、問題点の指摘と対策の提案を行った。パーソナル AI エージェントはインターネットやメタバースでは自律的アバターとして実現されるため、その場合の法的問題も検討した。これらの成果は主として複数の論文として発表した。この発信を通じて、技術の社会的普及に際して生じ得る疑念の解消と本質的問題の把握に貢献できると思われる。

新技術の社会定着におけるトラスト構築の方法論について、当該技術が抵触し得る法的利益や内在的リスク、提供者と利用者の能力的な格差に応じた対応に関

する概念整理を行ない、法的制度との関連付けを行なった。その内容は JST-CRDS による戦略プロポーザル「デジタル社会における新たなトラスト形成」にも反映されている。ただし同プロポーザルに向けた検討過程においてトラストの概念がさまざまな分野で多様に使われていることが反映したこともあり、それらと法制度上の概念との関連付け・整理についてはなお課題となっている。エージェントに対するトラスト問題がメタバースにおけるアバターについても同様に発生することを指摘し、総務省「WEB3 時代に向けたメタバース等の利活用に関する研究会」で報告した内容が同省情報流通行政局の対外的な説明資料にも反映されたほか、内閣府・ムーンショット目標 1「2050 年までに、人が身体、脳、空間、時間の制約から解放された社会を実現」に採択されたプロジェクト「アバターを安心かつ信頼して利用できる社会の実現」(PM 新保史生)にも反映されている。

・ ガバナンス

- 「AI エージェントのガバナンスのモデル構築のための設計指針と考慮要素を提示した。開発者と利用者の役割・責任の分担や国際協調のあり方など社会全体のマクロなガバナンスの枠組を提示するとともに、企業など関連する組織の内部でのガバナンスのモデルも提示した。アウトリーチの一環として、成原は NTT データの AI アドバイザリーボードの委員を務め、同社の AI ガバナンスの構築・運用について助言を行った。」
- AI 一般、狭くは AI エージェントに関する国際的ガバナンスの枠組は非常に大きなものであり、全体のカバーはできなかった。しかし、英国側主導で、UNESCO の AI 倫理指針を念頭におき、AI による医療チャットのガバナンスに関して、全世界から約 30 名の指導的立場の研究者(日本からは中川を含む 4 人)が参加した 2022 年 8 月に開催された 3 回連続のワークショップで国際的な問題意識を認識することができた。この認識は参加者に共有され、国際機関、各国政府、企業、市民社会に還元されることを期待している。

3-2. 研究開発成果

実施項目① AI 倫理規範の背景調査

(1)内容:「2-2. 実施内容」の「実施項目①」を参照。

(2)活用・展開: (1)内容で紹介した初期の AI 倫理規範はすでに企業などで取り入れられている。その次に紹介した EU の AI 規制法案は、まだ法案が成立していないこともあり、企業ではここで紹介したような情報をもとに対応を検討していると言われている。EU などの法制度改訂の状況に留意して、国内にも発信していくことが今後も重要である。

実施項目② 政策、法制度、経済性に関する基礎調査

(1) 内容：「2-2. 実施内容」の「実施項目②」を参照。

(2) 活用・展開：トラストにおける主観的要素（トラストそれ自体）と客観的要素（信頼に値する性質＝trustworthiness）の乖離を指摘して概念整理を行なったこと、信頼する側とされる側の技能・知識水準格差などをキーにして代理・権威・信託という3類型に整理しそれぞれに異なるガバナンス手法との対応が読み取れる点を指摘したことは、人工知能学会における報告を通じて科学技術サイドにフィードバックされているほか、JST-CRDSの戦略プロポーザルにおいても参照されている。科学技術サイドにおけるトラスト概念の用法も踏まえた総合的検討へと発展させることが今後必要となるだろう。

実施項目 ③文化、社会の歴史的背景調査、分析、提言

(1) 内容：「2-2. 実施内容」の「実施項目③」を参照。

(2) 活用・展開：テクノアニミズムを援用した人-ロボット関係の研究やロボット開発はすでにおこなわれており（例：岡田美智男・豊橋技科大教授の〈弱いロボット〉）、今後もさらなる展開が期待される。差別感情の実態調査は、AIによって生じるアルゴリズム差別への事前の対処方法を準備する作業に貢献することが期待される。今後、差別感情における文化差の研究も進展することで、各社会の実情に適したAI差別への対応方法を検討することが可能となるだろう。

実施項目 ④インタビューなどの社会調査によるデータ取得

(1) 内容：「2-2. 実施内容」の「実施項目④」を参照。

(2) 活用・展開：(1)内容で紹介した2021年度、2022年度の社会調査は、今後の技術開発、法制度改正に重要な基盤的知見であるので、論文、講演などを通じて発信してきた。この発信を今後も続けていくことが重要である。ちなみに医療データの有効利用については動きはEUではEuropean Health Data Space (EHDS)が提案され、日本でも次世代医療基盤法の改正の動きがある。関係者にここでの成果が届くように発信を続けたい。

実施項目 ⑤パーソナルAIエージェントの設計

(1) 内容：「2-2. 実施内容」の「実施項目⑤」を参照。

(2) 活用・展開：この項目に関しては、中川が多数の論文、講演し、また総務省・情報通信政策研究会AI分科会(2020年1月27日開催)でコメントとして発信を行ってきた。橋田浩一氏のムーショットプロジェクトにおいて、ここで検討してきたアイデアを利活用した実装に活かす方向などを検討してもらっている。

(3) その他：

パーソナルAIエージェントの一種である自律的アバターにかかわる諸問題は2022年度に明確になってきた。とくに本人死後の自律的アバターの在り方や法的位置づけは引き続き

き法学系研究者と議論を続けていきたい。

実施項目⑥ 美的感覚の調査と AI による学習

(1) 内容：「2-2. 実施内容」の「実施項目⑥」を参照。

(2) 活用・展開：この項目に関しては、日英比較については佐倉が、それ以外については水上が担当した。ファッション研究と技術哲学の接点については論文として発表し、ソーシャルな AI の行為者性についての研究は複数の学会にて発表した他、論文化を進めている。今後はファッション研究に従事する研究者やソーシャルな AI やロボットの開発に携わる工学研究者や企業の担当者との情報交換が重要であり、これについては引き続き議論を進めていきたい。

実施項目⑦ 既存のガバナンス枠組の網羅的調査、分析、提言

(1) 内容：「2-2. 実施内容」の「実施項目⑦」を参照。

(2) 活用・展開：AI エージェントのガバナンスのあり方について、関連する論文や記事を多数執筆するとともに、法学系の学会、人工知能学会、IEEE、政府機関、企業などで多数の講演・報告を行うことにより、研究成果の活用・展開を図ってきた。また、成原は NTT データの AI アドバイザリーボードの構成員として、同社の AI ガバナンスの構築・運用について助言を行ってきた。

実施項目⑧ 利害関係者からの直接、間接の情報収集（英国側、日本側）

(1) 内容：「2-2. 実施内容」の「実施項目⑧」を参照。

(2) 活用・展開：実施項目④と同様である。

実施項目⑨ 利害関係者との議論による国際的ガバナンス枠組の設計（主に英国側）

(1) 内容：「2-2. 実施内容」の「実施項目⑨」を参照。

(2) 活用・展開：英国側プロジェクトが主催し、日本側も共催した、全世界から約 40 人の研究者、実務者を集めて、UNESCO AI 倫理指針に基づき、医療チャットシステムを具体的対象にして、3 回にわたっておこなったシンポジウムを行った。ステークホルダーを網羅すること、各国で準備すべき方策など多岐にわたる意見が集約された。このシンポジウムには UNESCO の担当者も出席していた。集約された意見は各国の指導的研究者、UNESCO の当事者にも伝わったと思われる。

実施項目⑩ 日英双方の共同作業による国際的ガバナンス枠組の設計

(1) 内容：「2-2. 実施内容」の「実施項目⑩」を参照。

(2) 活用・展開：「2-2. 実施内容」の「実施項目⑩」を参照。

実施項目⑪ 国際的ガバナンス枠組に関する会議開催

(1) 内容：「2-2. 実施内容」の「実施項目⑪」を参照。

(2) 活用・展開：理論的な調査研究に加えて、2020年度に英国側との定期的オンライン会議で COVID-19 対策の現状と課題を議論し、日本の状況、とりわけ COCOA の問題点を明らかにした。この結果を論文、理化学研究所の一般向け公開セミナーでも発信してきた。また、英国のボードミーティング、公開国際シンポジウムでも発信して、海外へのアウトリーチも行った。

(3) その他：代表者：中川は JST-RISTEX ELSI 領域のアドバイザーとして IT 技術を用いたパンデミック対策のプロジェクトのアドバイザーを行っているが、我々に研究成果を必要に応じて伝えている。

実施項目⑫ 実施項目⑪の会議に結果の出版などによるアウトリーチ

(1) 内容：「2-2. 実施内容」の「実施項目⑫」を参照。

(2) 活用・展開：6-1 節～6-4 節で記載した発信を行ってきた。研究成果は今後も継続的に発信する予定であるが、年度末の終了後の公開のワークショップを日英共同、および日本側独自の形で開催する計画を立てている。

3-3. 今後の成果の活用・展開に向けた状況

AI 倫理はすでに政策、場合によっては法制度として実現される時代に入った。一方で、まだ明確になっていなかった AI 倫理の問題も生じてきている。例えば、最近急速に発展した Foundation Model に係わる問題、例えば(1) 画像生成 AI である midjourney や Stable Diffusion などが生成した画像の著作権の問題、(2) 自然言語生成 AI である ChatGPT の著作権ないし作成された文章の文責者や影響範囲、著作権誤用、さらには悪用の問題が顕在化してくるであろう。また、自律的アバターの著作権など未検討の問題も出てきている。AI 倫理に携わった経験をこれらの問題にも応用していく可能性を検討する。

分担した各チームのリーダーとは密接に連絡をとっているが、さらに研究開発実施者、協力者とも共同研究などを通じて研究を展開させていきたい。寺田麻佑氏が代表者である SECOM 財団の助成研究が採択され、開始している。橋田氏のムーンショット研究に中川が協力を開始した。英国側の James Wright 氏とは引き続き研究協力を続け、同氏の日本滞在と共同研究を計画中である。

本研究で行ったパーソナル AI エージェントの研究の成果は、引き続き自律性を持つソフトウェアのアバターとして展開できる部分が多いので、この方向の研究を継続していく予定である。死者の SNS データの扱いは、本研究実施中に明らかになってきたテーマであるが、本人死後のパーソナル AI エージェントの扱い方についての研究を開始しているため、上記のような他の研究と共同することによって、本研究を継続していく。

人々が技術に対して持つトラストの現実的な状況に関する調査、トラスト概念の分析をもとにした制度的検討については JST-CRDS の戦略プロポーザル「デジタル社会における新たなトラスト形成」と結び付く点があり、中川・大屋が継続的に協力している。

内閣府・ムーンショット目標 1「2050 年までに、人が身体、脳、空間、時間の制約から解放された社会を実現」において採択されたプロジェクト「アバターを安全かつ信頼して利用できる社会の実現」には大屋が協力しており、他者のアバターに対するトラストの根拠とそれを反映した社会的制度について分析を進める予定となっている。また、同目標 1 において採択されたプロジェクト「身体的能力と知覚能力の拡張による身体の制約からの解放」には成原が参加しており、本研究で得られた AI ガバナンスに関する知見を活用して、サイバネティックアバターのガバナンスのあり方について研究している。

4. 領域目標達成への貢献

プロジェクトの領域目標達成への貢献

- ・ 情報技術がもたらす正負の側面については、文化依存性、人々の心理と感性と密接に関係していることがある程度知られているが、これを AI と文化のグループが精密に調査、分析して領域全体に情報提供することができる。
- ・ 法制度や企業活動においても情報技術がもたらす正負の側面があり、さらに多様な価値観、倫理意識があることは認識されているが、個別事象だけではなく、むしろトラスト、ガバナンスという大局的価値体系について調査、分析したうえで将来の方向を提示することによって、領域の目標に沿った成果を発信する。
- ・ 社会と技術の望ましい共進化の具体例としてパーソナル AI エージェントの概念設計と技術的可能性を示す。さらに、社会で受け入れられるための法制度、社会制度の改善を促す提言を、上記のトラスト、ガバナンスの成果をもとに行う。

初年度から最終年度までの研究開発の経過および成果により、領域目標達成へどのように貢献したか

- ・ AI と文化に関する目標は主として佐倉が中心に行った。理論的基礎については書籍、論文などで発信し、複数回の公開シンポジウムにより領域内外の方々にも情報提供を行った。とくに ELSI 領域の方には発表、参加をいただくことができた。
- ・ 法制度およびガバナンスに関しては主に大屋、成原が多数の書籍、論文、講演を行って、その体系、現状、あるべき姿を本研究の分野以外にも十分に発信できた。また、中川は産業界における AI ガバナンスに関して、EU の AI 白書、AI 規制法案の紹介を通じて日本企業が留意すべき点を発信した。ガバナンスという観点からは、中川、大屋が経産省の Society5.0 における新たなガバナンスモデル検討会、アジャイル・ガバナンス実装 WG に委員としての参加を通して、AI を活用したアジャイル・ガバナンスの在り方に関して本プロジェクトの成果を利用して貢献した。
- ・ パーソナル AI エージェントに関しては、主として中川が概念設計、法的位置づけを調査を行った。メタバースなどのインターネット空間においてパーソナル AI エージェントと対応する概念である自律的アバターに関してトラストの形成と保持をめぐる諸問題、本人死後の自律的アバターの在り方を検討し、論文、講演などを通して発信することによって当初の目標達成に貢献した。

社会技術研究開発
「人と情報のエコシステム」研究開発領域
「PATH-AI: 人間-AI エコシステムにおけるプライバシー、
エージェンシー、トラストの文化を超えた実現方法」
研究開発プロジェクト 実施終了報告書

水上 拓也	ミズカミ タクヤ	理化学研究所	革新知能統合研究センター	ポスドク研究員
宇津呂 武仁	ウツロ タケヒト	筑波大学	システム情報系知能機能工学域	教授
高橋 達二	タカハシ タツジ	東京電機大学	理工学科	准教授
堀 浩一	ホリ コウイチ	東京大学	工学系研究科	教授
江間 有紗	エマ アリサ	東京大学	政策ビジョン研究センター	准教授
橋田 浩一	ハシダ コウイチ	東京大学	情報理工学系研究科	教授
武田 英明	タケダ ヒデアキ	国立情報学研究所	情報学プリンシプル研究系	教授
折田 明子	オリタ アキコ	関東学院大学	人間共生学部	教授
寺田 麻佑	テラダ マユ	一橋大学	ソーシャル・データサイエ ンス教育研究推進センター	教授
加藤 綾子	カトウ アヤコ	東洋大学	経済学部	准教授
吉田光男	ヨシダ ミツオ	筑波大学	ビジネスサイエンス系	准教授
吹金原 榛耶	フキハラ シンヤ	東京電機大学	理工学科	修士課程大学院生
村瀬 達也	ムラセ タツヤ	東京電機大学	理工学科	修士課程大学院生
西村友海	ニシムラ トモウ ミ	九州大学	法学研究院	准教授

(2) AI 文化グループ（リーダー氏名：佐倉統）

役割：AI の文化依存性ととりわけ日本の文化への依存性に関しては、日本人の AI に対するメンタリティを明らかにする

氏名	フリガナ	所属機関	所属部署	役職（身分）
佐倉 統	サクラ オサム	東京大学	大学院 情報学環	教授
福住 伸一	フクズミ シンイチ	理化学研究所	革新知能統合研究センター	研究員
猪口 智広	イノクチ トモヒロ	東京大学	大学院 学際情報学府	博士課程大学院生
Wang Yuhui	ワン ユーファイ	東京大学	大学院 学際情報学府	修士課程大学院生

(3) AI 法制度グループ（リーダー氏名：大屋雄裕）

役割：AI と人間とのインタフェースで重要になるトラストと法律の関係を検討する。

氏名	フリガナ	所属機関	所属部署	役職（身分）
大屋 雄裕	オオヤ タケヒロ	慶応義塾大学	法学部	教授

(4) AI ガバナンスグループ（リーダー氏名：成原慧）

役割：AI あるいは AI で利用されるデータに関する統治機構にあり方について検討する。

氏名	フリガナ	所属機関	所属部署	役職（身分）
成原 慧	ナリハラ サトシ	九州大学	法学研究院	准教授
松瀬 萌々香	マツセモモカ	九州大学	大学院法学府	修士課程大学院生

(5) 英国側のカウンターパート（リーダー氏名：David Leslie）

氏名	フリガナ	所属機関	所属部署	役職（身分）
David Leslie	デビッドレスリー	The Alan Turing Institute	Public policy programme	Ethics Fellow
Florian Otsmann	フロリアンオスマン	The Alan Turing Institute	Public policy programme	Policy Fellow
James Wright	ジェイムズライト	The Alan Turing Institute	Public policy programme	Researcher
Charles Raab	チャールスラーブ	The University of Edinburgh	School of Social and Political Science	Professor
Fumi Kitagawa	フミキタガワ	The University of Edinburgh	Entrepreneurship & Innovation	Lecturer

英国側グループとの関係は以下の通りである。David Leslie が率いる AI 倫理と文化をテーマとしたグループ、および AI 倫理に造詣が深い Florian Otsmann のグループ、AI と法律、社会学に造詣が深い Charles Raab、Fumi Kitagawa のグループの 4 グループの間で意見交換、議論を行う。このような成果の発表などを行う国際セミナーなどのアウトリーチ活動を行う。

5-3. 研究開発の協力者

氏名	フリガナ	所属	役職（身分）	協力内容
工藤 郁子	クドウ イクコ	大阪大学	招へい教員	AI に対する法制度の見地から研究

社会技術研究開発
「人と情報のエコシステム」研究開発領域
「PATH-AI: 人間-AI エコシステムにおけるプライバシー、
エージェンシー、トラストの文化を超えた実現方法」
研究開発プロジェクト 実施終了報告書

藤田 卓仙	フジタ タカノリ	慶應義塾大学	特任准教授	医療法制度における AI の在り方の研究
-------	----------	--------	-------	----------------------

機関名	部 署	協力内容
なし		

6. 研究開発成果の発表・発信状況、アウトリーチ活動など

6-1. 社会に向けた情報発信状況、アウトリーチ活動など

6-1-1. プロジェクトで主催したイベント（シンポジウム・ワークショップなど）

主催した公開シンポジウム

年月日	名称	主催者	場所	参加人数	概要
2020/ 1/21	AIP 科学技術と 社会チーム公開 シンポジウム (1)	理化学研究所 AIP センター 科学技術と社 会チーム(佐 倉、福住、中川)	理化学 研究所 AIP セ ンター (日本 橋)	84	AI と身体性 その1 (詳細は以下に記述)
2020/ 6/18	(一財)消費科学 センター「消費 者大学講座」	大屋雄裕	消費科 学セン ター会 議室		「IT 時代に必要なルー ルとは」なにかについて 講演を行った(録画公開 あり)
2020/ 7/2	ビッグデータ・AI と人権	成原慧:公益 財団法人世界 人権問題研究 センター人権 大学講座	オンラ イン	100	ビッグデータ・AI の利活用 に伴う差別やプライバシー 侵害など人権問題について 概説した。
2020/ 7/7	Society5.0 にお ける自由と規制	大屋雄裕:慶 應義塾大学	オンラ イン	100	来るべき Society5.0 におけ る人々が持つべき自由と課 されるべき規制の在り方
2021/ 3/19	AIP 科学技術と 社会チーム公開 シンポジウム (2)	理化学研究所 AIP センター 科学技術と社 会チーム(佐 倉、福住、中川)			AI と身体性 その2 (詳細 は以下に記述)
2022/ 1/10	第四次産業革命 における情報通 信融合法制フォ ーラム	成原慧	オンラ イン	20	日韓の研究者が第四次 産業革命の伸展位向け た法的課題について議 論した。

2022/1 1/18	理研-JST 合同 AIP シンポジウ ム.	理化学研究 所・JST の共同 主催	オンラ イン	500	共同主催者の成果報告 を行い、本プロジェクト からは中川が AI 倫理と EU の AI 倫理から AI 規 制法への展開について 講演した。
2022/2 /8	人間-AI エコシ ステム・プロジ ェクト&空飛ぶ クルマ プロジ ェクト合同公開 シンポジウム	中川裕志、小 島立	オンラ イン	156	JST-RISTEX ELSI 領域： 「空飛ぶクルマ」の社会実 装において克服すべき ELSI の総合的研究（代表：小島 立）との合同公開シンポジ ウム（詳細は以下に記述）
2022/3 /23	AIP 科学技術と 社会チーム公開 シンポジウム (3)	理化学研究所 AIP センター 科学技術と社 会チーム(佐 倉、福住、中川)	オンラ イン	64	ステークホルダ分析を通じ た自動運転バスの社会受容 性 (詳細は以下に記述)
2022/ 10/18	AIP 科学技術と 社会チーム公開 シンポジウム (4)	理化学研究所 AIP センター 科学技術と社 会チーム(佐 倉、福住、中川)	オンラ イン	46	COVID-19 関連アプリの受 容性に関する文化の違いか らの比較 (詳細は以下に記述)

◆会議名：AIP 科学技術と社会チーム公開シンポジウム(1)

テーマ：AI と身体性 その1

期間：2020年1月21日 15:00-18:00

開催場所：理化学研究所 AIP センター（日本橋）

主催等の機関名：主催：理化学研究所 AIP センター科学技術と社会チーム

趣旨：

AI 技術のと社会の関係を検討する際に、社会の文化的特性が重要な論点であることは明らかだ。理化学研究所革新知能統合研究センターの佐倉チームでもその点を主たる研究対象として活動を進めてきた。しかし、この「文化」というのがなかなかの難物で、調べれば調べるほど、まるで砂漠の蜃気楼のように遠ざかっていってしまう。

そこで本シンポジウムでは、「文化」はいったん背景に退かせ、文化差を検討するための別のアプローチとして「身体性」に注目することにした。「身体」は、文化や歴史や人間関係など、さまざまな要素が交錯するフィールドでありメディアであるからだ。AI と身

体性の関係自体が AI 技術の今後を考える上で重要なテーマであるだけでなく、その社会的文化的側面を考察するための、良いゲイトウェイとしても機能するはずである。

当日は、人間拡張技術を駆使して身体の境界を突破しようとしている稲見昌彦氏、身体性の生物的特性を認知発達科学的に研究している明和政子氏、能楽師として身体性の文化差を実践し俯瞰的に分析もする安田登氏から話題提供をいただき、討論した。

●プログラム

15:00-15:10 開会あいさつ

このシンポジウムの目的 佐倉統 (東京大/理研 AIP)

15:10-15:50 基調講演 1

デジタルサイボーグ 稲見昌彦 (東京大)

15:50-16:30 基調講演 2

能における身体性 安田登 (能楽師)

16:30-16:40 休憩

16:40-17:20 基調講演 3

ヒトの脳と心の創発・発達 明和政子 (京都大)

17:20-18:00 パネル・ディスカッション

指定討論者：松田雄馬 (iQβ)

パネリスト：稲見昌彦、安田登、明和政子、福住伸一 (理研 AIP)

司会：佐倉統

◆会議名：AIP 科学技術と社会チーム公開シンポジウム(2)

テーマ：AI と身体性 その2

期間：2021年3月19日 13:30-17:15

開催場所：zoom ウェビナー

主催等の機関名：主催：理化学研究所 AIP センター科学技術と社会チーム

趣旨：社会の中における AI について多方面から考えるシンポジウム、昨年引き続き身体性を中心テーマとして幅広い分野の論者からの講演とパネルディスカッションです。2020 年はコロナ禍のためオンラインでの仕事や生活の比重が以前より格段に増えました。デジタル世界におけるコミュニケーションのありかたをさまざまな観点から検討し、人と人工物とのより調和的な関係を追求することが求められています。今年は感性工学、ヒューマンインターフェイス、障害と開発経済を専門とされる方に講演をいただき、オンラインの世界における人の身体性を AI などの自律的人工物がどう変えていくのかをみなさんと一緒に討論した。当日は手話による通訳も行った。

●プログラム

- 13:30-13:40 開催趣旨説明 佐倉統 (東京大学/理研 AIP)
- 13:40-14:25 大澤博隆 (筑波大学) : COVID-19 下でのヒューマンエージェントイン
タラクションの貢献
- 14:25-15:10 大倉典子 (芝浦工大) : 「かわいい」という感性価値と身体性～触感を軸として～
- 15:10-15:20 休憩 10分
- 15:20-16:05 森壮也 (アジア経済研) : ろう者という身体性と AI
- 16:05-16:15 コメント 中川裕志 (理研 AIP)
- 16:15-17:15 パネルディスカッション 講演者+中川裕志 (理研 AIP) +福住伸一
(理研 AIP) +荒井ひろみ (理研 AIP) 司会: 佐倉統 (東京大学/理研 AIP)

◆会議名: 人間-AIエコシステム・プロジェクト&空飛ぶクルマ プロジェクト合同公開
シンポジウム: JJST-RISTEX ELSI領域: 「空飛ぶクルマ」の社会実装において克服すべ
きELSIの総合的研究 (代表: 小島立) との合同公開シンポジウム

◇ 日時: 2022年2月8日 13時~16時 オンライン (zoom webinar)

登録者156名、実質参加者138名

プログラム

- ・ 医療AIの社会調査報告 (中川)
- ・ 日本文化とAI/ロボット (佐倉)
- ・ 空飛ぶクルマに関する社会的合意形成セッション (小島立・九州大学)
- ・ パネルディスカッション

◇ パネリスト: 山本薫 (九州大学)、有吉亮 (横浜国立大学)、平山賢太郎 (九州大学)、寺田麻佑 (ICU)、大屋雄裕 (慶応大学) 成原慧 (九州大学)

◆会議名: AIP科学技術と社会チーム公開シンポジウム (3)

テーマ: 自動運転バスの社会受容性

概要:

理研AIPセンター「科学技術と社会チーム」では、技術の社会的形成と技術の社会受容性について、「社会・文化的側面」と「人間的側面」から取り組んでいる。本チームではこれまで、「科学技術と社会」、「AIと身体性、AIの身体性、AIと社会の身体性」、「AIと身体性: その2」というテーマでシンポジウムを開催してきた。今回は対象をより具体的にしぼり、自動運転バスの社会受容性について実例を踏まえつつ人文社会科学とユーザユーザエクスペリエンスの視点から議論する。まず、BOLDLY(株)の改

發氏から自動運転バスの実証実験を御紹介いただき、次いで東京大学の唐沢先生から自動運転に関する一般社会の態度とその社会受容性との関連をご講演いただく。その後理研AIPセンターから福住が製品やシステムを「使う」際のステークホルダへの影響をモデル化した「利用時品質」を紹介し、最後に小樽商科大学の平沢先生から実証実験を通じたステークホルダ分析と利用時品質モデルの適用についてお話ししていただく。講演後、技術の社会受容性などについて議論を行う。

◇ 日時：2022年3月23日

形式：ウェビナー． 参加者 6 4 名

プログラム：

14:00 趣旨説明 佐倉統（理研AIP／東京大学）

14:10 「自動運転バス実用化の先行事例」 改發社（BOLDLY(株)）

14:30 「自動運転をめぐる態度構造と社会受容」 唐沢かおり（東京大学）

15:10 「利用時品質モデル」 福住伸一（理研AIP）

15:30 休憩

15:40 「自動運転バスを対象としたステークホルダ分析」 平沢尚毅（小樽商科大学）

16:20 指定討論 中川裕志（理研AIP）、吉田直可（法律事務所愛宕山）

16:40 ディスカッション

17:30 終了

◆会議名：AIP 科学技術と社会チーム公開シンポジウム(4)

テーマ：COVID-19 関連アプリの受容性に関する文化の違いからの比較期間：2022 年
10 月 18 日 14:00-17:00

形式：ウェビナー

主催等の機関名：主催：理化学研究所 AIP センター科学技術と社会チーム

趣旨：

理研 AIP 科学技術と社会チームでは、技術が社会で受け入れられるための要件（技術の社会受容性）についての研究を、社会的側面と人間的側面から取り組んでいる。今回は COVID-19 関連アプリについて、それらがどのように受け入れられ使われているのか、あるいは使われていないのかを社会的側面から比較し、将来への示唆を検討する場としてシンポジウムを企画した。アジア数か国と欧州を対象にした実態調査結果を紹介し、その違いを踏まえて、新たな感染症が発生したときに今回の教訓をどのように活かすのか議論した。

I 総論

問題点の所在、日本の現状を含む：中川裕志

II 各論 日本以外の各国状況：

シンガポール：Hong Chew TOH（東大）

台湾：緒方健（千葉大）

韓国：ホ・ドゥイム（東大）

Ⅲ パネル討論

司会：佐倉統

コメンテーター：福住（AIP）、神里（千葉大）

パネリスト：中川、緒方、ホ、Toh

英国側プロジェクト主導の会議

年月日	名称	場所	概要
2021/6/14	第 1 回国際オー プンワークショップ	オンライ ン (zoom)	AI ガバナンスに関する報告と議 論(詳細は 2 章⑪に記載)
2022/3/15	第 2 回国際オー プンワークショップ	オンライ ン (zoom)	医療 AI の調査報告(詳細は 2 章⑪ に記載)
2022/7/15	PATH-AI ワーク ショップ第 1 回	オンライ ン (zoom)	UNESCO の AI 倫理指針をベース にして、医療チャットアプリにつ いてステークホルダーの在り方 について議論。参加者は、日米欧ア ジア、アフリカから計 3 2 名(日本 からの参加者 4 人)
2022/8/15	PATH-AI ワーク ショップ第 2 回	オンライ ン (zoom)	UNESCO の AI 倫理指針をベース にして、医療チャットアプリにつ いて PIA (影響評価) について議 論。参加者は、日米欧アジア、ア フリカから計 3 5 名(日本からの参 加者 4 人)
2022/9/23	PATH-AI ワーク ショップ第 3 回	オンライ ン (zoom)	UNESCO の AI 倫理指針をベース にして、医療チャットアプリにつ いて readiness(実現準備状況)につ いて議論。参加者は、日米欧アジ ア、アフリカから計 3 0 名(日本か らの参加者 2 人)

日本および英国のプロジェクトの日英共同研究内部で行った会議一覧

なお、各回の参加者は10名程度。

年月日	名称	場所	概要
2020/4/2	英国側の共同研究グループとの定期的会議	オンライン (zoom)	本年度の共同研究の進め方の打ち合わせ
2020/5/15	同上	同上	COVID-16 パンデミックにおける日英政府の行動分析
2020/5/21	同上	同上	中川、成原、寺田の発表
2020/6/18	同上	同上	COVID-19 の接触通知アプリの実態の報告と共有
2020/8/11	同上	同上	英国側からのパンデミック対策の社会的影響の報告
2020/10/27	同上	同上	日英両国の Web サイト構築および論文発表経過報告
2021/3/18	同上	同上	利害関係者からの直接、間接の情報収集 4月以降に予定する公開ワークショップでの成原による発表の下打ち合わせ
2021/6/24	同上	同上	第1回国際ワークショップの振り返り
2021/7/6	同上	同上	医療AIの調査結果の議論
2021/10/26	同上	同上	研究進捗状況打ち合わせ
2022/1/17	同上	同上	第2回国際ワークショップの計画

本プロジェクトの日本側のみで行ったワークショップ。プロジェクト外部の研究者にも参加いただいた。参加者数は毎回10名～15名。

年月日	名称	場所	概要
2020/11/30	日本側のグループ会議	同上	発表者：藤田卓仙、平山健太郎、タイトル：「医療データと独禁法に関する現状」
2021/1/6	同上	同上	発表者：藤嶋陽子、タイトル：「デザインにおける AI」、加藤綾子、タイトル：「パ

			パーソナル AI エージェントの利用」
2021/3/9	同上	同上	公開ワークショップでの成原慧による発表の下打ち合わせ
2021/6/14	同上	同上	発表者：稲谷龍彦. タイトル：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」
2021/7/26	同上	同上	発表者：福住伸一. タイトル：製品・システム・サービスの利用時品質モデルをベースにした AI 問診の社会受容性及び自動運転バスへの適用」
2021/8/25	同上	同上	発表者：西村友海. タイトル：AI システム開発契約」
2021/10/21	同上	同上	発表者：松尾剛行. タイトル：中国個人情報保護法-国際比較と政府の情報利用の観点から」
2021/12/7	同上	同上	発表者：折田明子. タイトル「死後のソーシャルメディアを巡るジレンマ」
2022/4/21	同上	同上	発表者：相澤彰子. タイトル「自然言語処理技術の最新動向と AI」
2022/9/7	同上	同上	発表者：寺田麻佑. タイトル「公法と AI」

◆主催者ではなかったが、本プロジェクトの成果をアウトリーチした社会活動

- ・ 全脳アーキテクチャシンポジウム（2021年9月10日）において、「人間は動物を必要とするが、AIは人間を必要とするか？」と題した講演を行なった。（大屋雄裕）
- ・ 日本弁護士連合会による第29回司法シンポジウム「民事裁判手続のIT化とこれからの司法」（2021年10月30日）第4部パネルディスカッション「裁判手続のIT化のこれから・市民にとって利用しやすい裁判とは」に登壇し、AIの司法に対する影響についても発言した。（大屋雄裕）
- ・ JST・NSFによるWorkshop on: Exploring new research agendas connecting Artificial Intelligence、Robotics、Humanities and Social science [Engagement Initiative 2021]に参加し、アメリカ側参加者とともに関連する研究シーズに関する討議を行なった（2022年2月2日・8日）（大屋雄裕）。
- ・ AI倫理に関して富士通にて社内向けオンラインセミナー（2020/9/15）、関連会社を含む一般向けセミナー富士通ソフトウェア・テクニカルカンファレンス2020へのビデオセミナー資料作成に講師として出演した。（中川裕志）

- ・ 公正取引委員会「デジタル市場における競争政策に関する研究会」に(委員)座長代理として参加し、報告書「アルゴリズム/AIと競争政策」
https://www.jftc.go.jp/houdou/pressrelease/2021/mar/210331_digital.html の作成、公開に寄与した。(中川裕志)
- ・ 総務省主催「AIネットワーク社会フォーラム」(2021年3月1日開催)パネルディスカッション「ニューノーマルにおけるAIへの期待」のパネリストとして参加し、本課題のAI倫理とガバナンスについて講演した。(中川裕志)
- ・ 総務省:AIネットワーク社会推進会議に幹事として参加し、本課題のAI倫理とガバナンスの研究成果についてアウトリーチを図った。(中川裕志:2018年～現在)
- ・ 人間中心のAI社会原則会議に構成員として参加し、本課題のうちAIエージェントの法的位置づけについて発表を行った(2021年3月24日)。(中川裕志)

6-1-2. 書籍、DVD など論文以外に発行したもの

- (1) 中川裕志：機械学習プロフェッショナルシリーズ『機械学習工学』【ISBN】9784065285862 1章1節 pp.1-9,7章 pp.199-241, 講談社 2022年7月22日刊行.
- (2) 中川裕志：データサイエンス入門シリーズ『教養としてのデータサイエンス』【ISBN】978-4-06-523809-7 3.1節 データ・AIを扱う上での留意事項 P.175-204 .講談社サイエンティフィック. 2021年5月25日刊行.
- (3) 中川裕志：AI倫理とプライバシー, 医学のあゆみ, 274巻9号(特集 AIが切り拓く未来の医療), 医歯薬出版株式会社, 2020年8月29日刊行.
- (4) 土屋俊・中島秀之・中川裕志・橋田浩一・松原仁 : AI事典 初版復刻版, 543ページ, 近代科学社, 2020年3月31日刊行, 【ISBN】978-4-7649-0528-3 C0550
- (5) 中川裕志: AI研究の歴史的経緯, 人工知能と人間・社会の第1章第1節, pp.2-15, 勁草書房, 2020年2月20日刊行.
- (6) AI研究の歴史的経緯, 中川裕志, 勁草書房, 人工知能と人間・社会の第1章第1節, pp.2-15, 2020年2月20日刊行.
- (7) 佐倉統: 科学とはなにか——新しい科学論, いま必要な三つの視点, 講談社, 2020年12月17日刊行.
- (8) Charles Keller 著, 夏目大訳, 佐倉統監修: ダーウィン『種の起源』を漫画で読む, いそっぷ社, 2020年05月刊行.
- (9) 佐倉統「生物は『悪条件だと進化できる』のは本当か 寄生の仕組みから考える」『講談社ブルーバックス』ウェブサイト, 2020年刊行. (<https://gendai.ismedia.jp/articles/-/71191>).
- (10) 福住伸一, 笠松慶子「製品開発のためのHCD実践-ユーザの心を動かすモノづくり」, 近代科学社, 2021年4月刊行.

- (11) 福住伸一, 平沢尚毅, 小林大二: 「ユーザビリティのための産業共通様式と人間中心設計プロセス -国際標準の全貌とその使い方」日本規格協会, 2021年4月刊行.
- (12) 五十嵐美樹・佐倉統・梶太一: 鼎談 サイエンスコミュニケーターの今, 現代化学 (617) pp.22-27 2022年8月1日刊行.
- (13) 佐倉統: 科学は誰のものなんだろうか? 現代化学 (610) pp.39-42, 2022年1月1日刊行.
- (14) 大屋雄裕: 「民主政は可能か?: 合理性と「個人」の再設計」那須耕介・橋本努(編) 『ナッジ!?: 自由でおせっかいなリバタリアン・パターナリズム』勁草書房, pp. 100-124, 2020年5月刊行.
- (15) 宍戸常寿・大屋雄裕・小塚荘一郎・佐藤一郎(編) 『AIと社会と法: パラダイムシフトは起きるか?』有斐閣, 2020年8月刊行.
- (16) 大屋雄裕 「AIにおける可謬性と可傷性」宇佐美誠(編) 『AIで変わる法と社会: 近未来を深く考えるために』岩波書店, pp. 45-62, 2020年9月刊行.
- (17) 大屋雄裕: プロファイリング・理由・人格, 勁草書房, 人工知能と人間・社会(第3編第1章), pp. 260-296, 2020年2月20日刊行.
- (18) 大屋雄裕 「パンデミック・監視・自己決定」林良嗣・森田紘圭(編) 『感染症とソーシャルディスタンス: COVID-19による都市・交通・コミュニティの変容と設計』第6章, 明石書店, pp. 112-131, 2022年12月刊行.
- (19) 大屋雄裕 「現代社会のリスクと個人の自由を考える」『中央公論』2022年9月号, 中央公論新社, pp. 194-199, 2022年8月刊行.
- (20) 成原慧: 「それでもアーキテクチャは自由への脅威なのか?」那須耕介・橋本努(編) 『ナッジ!?: 自由でおせっかいなリバタリアン・パターナリズム』勁草書房, pp. 75-99 2020年5月刊行.
- (21) 成原慧・大屋雄裕ほか: AIで変わる法と社会, 174ページ, 岩波書店, 2020年9月17日刊行.
- (22) 成原慧 「個人の自律とAIの自律」宇佐美誠(編) 『AIで変わる法と社会: 近未来を深く考えるために』岩波書店, pp. 26-44, 2020年9月刊行.
- (23) プリマヴェラ・デ・フィリッピ, アーロン・ライト(著), 片桐直人(編訳), 成原慧ほか(訳): ブロックチェーンと法, 316ページ, 弘文堂, 2020年11月刊行.
- (24) 成原慧ほか: 人工知能と人間・社会, 3章, 勁草書房, 2020年2月20日刊行. ISBN, 978-4-326-10280-8.
- (25) 宮下萌, 明戸隆浩, 石川優実, 川口泰司, 上瀧浩子, 曹慶鎬, 唐澤貴洋, 佐藤佳弘, 金尚均, 成原慧, 佐藤暁子: 『テクノロジーと差別: ネットヘイトから「AIによる差別」まで』【ISBN】978-4-75-926801-0, 10章 「『AIによる差別』にいかに向き合うか」, 解放出版社, 2022年2月刊行.
- (26) AI ネットワーク社会におけるアーキテクチャと法のデザイン, 成原慧, 勁草書房, 人工知

能と人間・社会の第3章第1節, pp.92-119, 2020年2月20日刊行.

- (27) 成原慧「プライバシー：プライバシー1.0, 2.0, 3.0, そしてその先のプライバシー」駒村圭吾(編)『Liberty2.0—自由論のバージョン・アップはありうるのか?』, 弘文堂, 2023年2月刊行.
- (28) Satoshi NARIHARA, “Personal Autonomy and Machine Autonomy in Human-Robot Interaction,” in Woodrow Barfield, Yueh-Hsuan Weng & Ugo Pagallo (eds.), Cambridge Handbook on Law, Policy, and Regulations for Human-Robot Interaction (Cambridge University Press, 2023, forthcoming).

6-1-3. ウェブメディア開設・運営

- (1) 本プロジェクトのHPの作成 2020/11/1
英国側：<http://path-ai.org/>
日本側：<https://sites.google.com/site/nakagawa3/path-ai>
- (2) 本プロジェクトのビデオアーカイブによる成果発表 2020/12/2
1. 代表者：中川の所属先の理研 AIP の open seminar として研究成果を発表公開した.
 2. 日本語：<https://aip.riken.jp/video/aip-open-seminar-4-japanese-part/>
 3. 英語：<https://aip.riken.jp/video/aip-open-seminar-4-english-part/>
- (3) RIKEN AIP researchers talked at the Symposium on Machine Learning and Fairness, <https://aip.riken.jp/news/mlfsymposium200218/>, 2020年2月18日
- (4) 大屋「Society5.0における自由と規制」(2020年7月7日)
1. 慶應丸の内シティキャンパスの動画サイト「クロシング」において動画が公開されている。

6-1-4. 学会以外のシンポジウムなどでの招へい講演 など

- (1) 中川裕志：理研-JST 合同 AIP シンポジウム, 2022年11月18日 (オンライン) 内容：AI倫理とEUのAI倫理からAI規制法への展開.
- (2) Hiroshi Nakagawa: EJEAC conference 2022 “Innovation and Action for Managing Urgent Future Local and Global Issues and Domains in Japan and Europe”. 2022/11.21. Title: AI and Society – ELSI About Autonomous Avatars.
- (3) Hiroshi Nakagawa: Delegation of Digital Heritage to Personal AI Agents. Invited talk (Japanese Embassy in Finland and gHAIR (Global Human-Centric AI-Transformation Research). March 3, 2021 Yle Uutiset .
- (4) 中川裕志, 丸山宏, 江間有沙, 神寫敏夫：機械学習と公平性に関するシンポジウム, 一橋講堂, 2020年1月9日 18:00-20:00, 参加者数350 内容：機械学習の利用が公平性に与える影響の問題への対処法を社会一般の方々共有.

- (5) 中川裕志, 山川宏, 岡ノ谷 一夫, 長谷敏司: 生物進化の終焉とシンギュラリティ後の世界・セミナー 東京大学・駒場キャンパス 21 KOMCEE West 地下講義室
開催日: 2020年3月19日 14:30-17:00 参加者数 160(Web参加) 内容: 高度AIとその知能爆発をリスク要因として捉えるよりも, 大局的なリスクを乗り越えるために活用するという視点を探る.
- (6) 中川裕志: AIの利活用を巡る課題, NICTサーバーセキュリティシンポジウム
2020, 招待講演, 品川フロントビル会議室 品川フロントビル B1F, 2020年2月12日.
- (7) 中川裕志: AI倫理の動向とプライバシーデータを意識した情報エコシステム, インフォメーションバンクコンソーシアムシンポジウム 第7回「シン・情報銀行」"これから目指すべき本当の情報銀行の姿", 招待講演, 日本科学未来館 7階未来館ホール, 2020年2月5日.
- (8) 中川裕志: AI倫理とエージェントの概念設計, JEITA 超スマート社会とデータ流通専門委員会 招待講演, 2020年1月17日, 2.場所;【ハロー貸会議室飯田橋駅前】9階 room B.
- (9) 佐倉統: AIと安心感, 学術会議学術フォーラム「安心感への多面的アプローチ」, 2022年11月5日, (オンライン).
- (10) O. Sakura: Education for an "AI Ready Society", The 3rd Japanese-German-French AI symposium -AI for Planetary Challenges in the Anthropocene-, 28 Oct, 2022, 日本科学未来館.
- (11) 佐倉統: JAAS+Elsevier 円卓会議「研究への信頼 Confidence in Research」2022年9月5日.
- (12) 佐倉統: 科学技術は誰のもの? —歴史を振り返って考える—, 成形プラスチック歯車研究専門委員会 第147回研究会・特別講演, 2022年2月25日.
- (13) 佐倉統: トークセッション「総合知が築く『食・農・環境』の未来」, 東京農業大学HUB構想第2回シンポジウム, 2021年12月16日.
- (14) 佐倉統: 科学知と生活知の間(はざま)にどう立ち向かうか, NEDO2021年度「産業技術総合研修」, 2021年12月2日.
- (15) 佐倉統: 科学技術は誰のもの? —歴史を振り返って考える—, 「移動の価値とモビリティの未来(3)」2020年10月20日 名古屋大学未来社会創造機構モビリティ社会研究所.
- (16) 佐倉統: 社会の側から科学技術を考える, 東北大学工学教育院トップリーダー特別講義, 2021年5月10日, 東北大学工学教育院.
- (17) 福住伸一, 笠松慶子: 「CPE セミナー」ABWに関する研究のご紹介, 2022年11月22日, オンライン.

- (18) 福住伸一：「HCD ビジネスシンポジウム 2022」～アーキテクチャが描く人間中心のデジタル社会～，デジタル技術と人の関係を考察するための国際規格，2022年2月4日，(オンライン開催)。
- (19) 福住伸一：ET&IoT 展テーマ別セミナー：AI/デジタル社会に必須の人間中心の取組み，【AI/デジタル技術】と【人】の関係を考察するための国際規格，2021年11月17日，パシフィコ横浜。
- (20) 佐倉統：社会の中の技術を考えるために，機械学習と公平性に関するシンポジウム，招待基調講演，2020年1月9日，一橋講堂。
- (21) 佐倉統：人類と地球の未来，第3回「科学者が見通す46億年の地球」，招待講演，2020年1月20日，毎日メディアカフェ。
- (22) 大屋雄裕：一財)消費科学センター「消費者大学講座」 開催日:2020年6月18日，内容「IT時代に必要なルールとは」。
- (23) 大屋雄裕:AIと基本的人権，青森自治研 特別分科会「AIと自治体」，自治労，2020年10月(ビデオ収録による開催)。
- (24) 大屋雄裕:『悪のAI論』とは何かーコメント・悪と人格の可能性について，先端技術と立法課題研究会，2020年10月30日(オンライン開催)。
- (25) 大屋雄裕:人格なき統治における社会科学，科学技術未来戦略ワークショップ「Society 5.0 実現に向けた計算社会科学」JST-CSDS，2020年7月9日(オンライン開催)。
- (26) 大屋雄裕:連帯の二つの基礎ー個別性と一体性，日本教育学会第79回大会・課題研究「技術革新とエンハンスメントの時代における教育学の課題」2020年8月26日(要旨公表による代替開催)。
- (27) 大屋雄裕:自律・代理・代表ー権限と責任の分配に関する制度，東京法哲学研究会，2020年7月23日(オンライン開催)。
- (28) 大屋雄裕：信用・信頼・信託ー責任と説明に関する概念整理，2020年度 人工知能学会全国大会（第34回）オーガナイズドセッション「人工知能におけるプライバシー，公平性，説明責任，透明性への学際的アプローチ」招待講演，2020年6月11日（オンライン開催）。
- (29) 大屋雄裕：AIの普及・進展とその課題，自治大学校地域づくりセミナー，2020年1月16日，総務省自治大学校。
- (30) 大屋雄裕：個人信用スコアの社会・経済的意義，GLOCOM 研究ワークショップ，2020年1月24日，国際大学グローバルコミュニケーションセンター
- (31) 大屋雄裕：プロファイリングと個人信用スコア，クレジットマネジメント研究会，2020年1月28日，早稲田大学国際会議場。
- (32) 大屋雄裕「Society5.0における自由と規制」丸の内政経懇話会，2021年4月15日。
- (33) 大屋雄裕「AIと規制:法との関係で」自由民主党 人工知能未来社会経済戦略本部，2021年5月12日。

- (34)大屋雄裕「監視社会における自由と安全」最高裁判所 司法研修所基盤研究会, 2021年7月12日.
- (35)大屋雄裕「法制度とトラスト」JST-CRDS 俯瞰セミナー「トラスト研究俯瞰」2021年8月18日.
- (36)大屋雄裕「AI と司法と法の未来」日本弁護士会 司法シンポジウム運営委員会勉強会, 2021年6月9日.
- (37)大屋雄裕「我らの社会とその基礎: 人格・可傷性・コミュニケーション」日本学会議シンポジウム「誰もが夢を追求できるアバター共生社会の実現を目指して」2022年7月26日.
- (38)大屋雄裕「メタバースと着地問題」総務省 WEB3 時代に向けたメタバース等の利活用に関する研究会, 2022年8月1日.
- (39)大屋雄裕「行政のデジタル化と地方自治の未来」地方財務協会 地方行政研究会, 2022年11月15日.
- (40)大屋雄裕「意思・契約・個人: 変容と消費者法の未来」イノベーションガバナンスと法理論研究会, 2023年1月7日.
- (41)成原慧: AI 時代の差別と公平性, オンラインシンポジウム「AI と差別」, 招待講演, オンライン, 2020年3月25日.
- (42)成原慧: 情報法の視点からみた情報ガバナンスと文理融合教育の課題, シンポジウム「情報ガバナンスと文理融合教育の課題」, 2020年1月24日, 九州大学伊都キャンパス.
- (43)成原慧「採用選考におけるデータ・AI の活用と法的問題」一般社団法人公正採用人権啓発推進センター第12回総会記念講演, 2023年1月27日.
- (44)成原慧: ビッグデータ・AI と人権, 公益財団法人世界人権問題研究センター人権大学講座, 2020年7月2日.

6-2. 論文発表

6-2-1. 査読付き (24 件)

- (1) 水上拓哉: なぜファッション研究において技術哲学が重要なのか, 『vanitas』, vol.8 採録決定
- (2) 中川裕志: サイバネティックアバターのトラスト, レクチャーシリーズ: 「AI と社会と人間〜ぶつかる・なじむ・とけこむ〜」〔第1回〕, 人工知能学会誌, 38巻1号 pp. 65-72, 2023年1月). https://doi.org/10.11517/jjsai.38.1_65
- (3) 中川裕志: AI エージェント, サイバネティック・アバターにおけるトラストの構築と問題点『情報通信政策研究』第6巻第1号 pp. IA-45~IA-60. 2022年11月 https://www.soumu.go.jp/main_content/000851753.pdf

- (4) Hiroshi Nakagawa, Akiko Orita. Using deceased people' s personal data. AI & Soc . 2022. <https://doi.org/10.1007/s00146-022-01549-1>.
<https://link.springer.com/article/10.1007/s00146-022-01549-1#citeas>
2023年2月12日時点でのアクセス数は1100
- (5) 中川裕志 : デジタルアーキテクチャデザイン : 6. デジタル社会における AI ガバナンス -倫理と法制度- (Digital Architecture Design : AI Governance in Digital Society - Ethics and Legal System). 情報処理, 62(6), pp. e34-e39 (2021-05-15). [info:doi/10.20729/00210983](https://doi.org/10.20729/00210983)
- (6) 中川裕志 , AI 倫理指針の動向とパーソナル AI エージェント, 総務省 学術雑誌『情報通信政策研究』第3巻第2号 (Journal of Information and Communications Policy Vol.3 No.2), pp. I1~I23 , 2020年
- (7) 湯沢友之, 佐倉統 : 科学コミュニケーション養成講座での訓練は聴衆特性への配慮にどのような影響を与えているか? 科学技術コミュニケーション 31 pp.1-17 2022年9月.
- (8) Sakura O: Robot and Ukiyo-e: Implications to Cultural Varieties in Human-Robot Relationships, AI & Society,
<https://link.springer.com/article/10.1007/s00146-021-01243-8>, 2021年
- (9) Qinyu E, Sakura O: Who Is Responsible for Causing and Solving the Problem? Responsibility Attribution of Medical Disputes in Chinese Print Media, International Journal of Health Services, 52(4), 2020年7月16日
- (10) 佐倉統 : 挑発に応答はあったか? : 挑発としての展示, その後とこれから, 展示学, Vol.59, pp.23-30, 2020年3月15日.
- (11) 福住伸一, 平沢尚毅, 改發 壮 : 新たな利用時品質モデルの考え方—自動運転バスの運用を事例として—, 情報処理デジタルプラクティス特集「超スマート社会実現に向けた情報技術活用のプラクティス」, Vo.63.No.5. 2022年4月15日.
<https://www.ipsj.or.jp/dp/contents/publication/50/S1302-S02.html>
- (12) 水上拓哉 : ソーシャルロボットの倫理における仮想的行為者性概念の可能性と限界, 『新進研究者 Research Notes』Vol.3, pp.28-36. 2020年
- (13) 水上拓哉 : 技術的人工物の「発話」の道徳的身分について——言語行為論の枠組みを用いた検討, 『哲学の門』Vol.2. pp.162-175. 2020年
- (14) 大屋雄裕 : Society5.0と人格なき統治 (特集 : With/Afterコロナ時代におけるICTの役割と利活用), 情報通信政策研究 5(1), 総務省情報通信政策研究所, pp.I-1 - I-14, 2021年. https://doi.org/10.24798/jicp.5.1_1
- (15) 大屋雄裕 : AIによる危機, AIをめぐる危機, 公共政策研究(21), 日本公共政策学会, ,

pp. 102-110, 2021年12月.

- (16) 大屋雄裕:「計算可能性と自由な行為」地方自治 892, ぎょうせい, pp. 2-15, 2022年3月.
- (17) Takehiro OHYA, "Surveillance Society" in: Mortimer Sellers and Stephan Kirste (eds.), *Encyclopedia of the Philosophy of Law and Social Philosophy*, Springer, 2020, https://doi.org/10.1007/978-94-007-6730-0_137-2
- (18) 大屋雄裕「メタバースの可能性と限界 (特集:メタバースがやってくる!)」『法学セミナー』817号(2023年2月号), 日本評論社, pp. 20-25, 2023年1月.
- (19) Satoshi NARIHARA, "AI Networking and its Governance: Autonomy, Complexity and Meta-Design" in A-B-C (Artificial Intelligence-Big Data-Cyber Security) Law Review Vol.1 No.1, pp.156-164, 2020
- (20) 成原慧:個人情報保護法制の官民一元化に向けた検討状況と課題, 九州大学法政研究87巻3号(村上裕章教授 退職記念論文集), pp.779-813, 2020年12月
- (21) 成原慧「インド太平洋地域におけるディスインフォメーションの流通とその対策—米国政府とプラットフォーム事業者による対策に着目して—」笹川平和財団国際情報ネットワーク分析 IINA, 2022年12月5日.
https://www.spf.org/iina/articles/narihara_01.html
- (22) Shinyashiki Emiko: Changing Circumstances in Contracts of Employment, Contract Law, and Employment Legislation in English Law, *Journal of Law and Politics* (Kyushu University), Nov. 88, pp.190-124, 2022.
- (23) 新屋敷恵美子:イギリスにおける代位責任(Vicarious Liability)法理の近時の展開(2):就業関係をめぐるリスクの多様化と管理監督機能の曖昧化の観点から, 法政研究88(1), pp.324-266, 2021年7月.
- (24) Kentaro HIRAYAMA: Interaction between Information Law and Competition Law in Japan, *Academia Letters*, Vol. 2307, pp.1-5, July 2021.

6—2—2. 査読なし (28 件)

- (1) 中川裕志: 本人死後のサイバネティック・アバターに関する考察. 日本ロボット学会誌. 41 巻 1 号 p. 9-13 . 2023年1月. <https://doi.org/10.7210/jrsj.41.9>
- (2) 中川裕志:最近公開されたAI倫理の動向と展望, 法とコンピュータ No.38, pp.101-108, 2020年7月(招待)
- (3) 中川裕志:AI 倫理指針における課題(Problems Appeared on AI Ethics Guidelines), レクチャーシリーズ「人工知能の今」[第11回], 人工知能学会誌.35(6), P.845-854. 2020年11月(招待)
- (4) 佐倉統:自動運転を社会論の側から考える, 学術の動向 25(5), pp. 67-69, 2020年

- (5) 佐倉統:「新しい知見」が意味するところは専門家と社会とで異なる, 学術の動向 Vol. 25, No. 12, pp. 22-25, 2020年12月.
- (6) 佐倉統:自動運転を社会の側から考える, 学術の動向 Vol, 25, No. 5, pp. 67-69, 2020年5月.
- (7) 佐倉統:科学技術は暴走しているのか? 世界思想, Vol. 47, pp. 27-31, 2020年4月.
- (8) 福住伸一・吉武良治・平沢尚毅:『責任の境界』に向けた人間中心のガイドライン, J-Stage, 2020年6月23日.
- (9) 大屋雄裕「AIと基本的人権〔特集:ウィズ・コロナ時代の自治研活動〕」『月刊自治研』2021年1月号(63巻736号), 自治労サービス, 2021年1月.
- (10) 大屋雄裕「パンデミックと超監視社会の可能性」『国際問題』698号 pp. 47-52. (2021年1・2月号), (公財)日本国際問題研究所, pp. 23-31, 2021年1月.
- (11) 大屋雄裕「「成熟した市民社会」とその敵〔特集:いま, 社会のあり方を考える〕」『法学セミナー』785号(2020年6月号), 日本評論社, pp. 31-37, 2020年6月.
- (12) 大屋雄裕「現代科学技術への「法」のアプローチ:事後・事前的規制から同時協働へ」『ビジネス法務』2020年9月号, 中央経済社, pp. 65-68, 2020年7月.
- (13) 大屋雄裕「人格なき統治における社会科学」科学技術未来戦略ワークショップ報告書『Society5.0実現に向けた計算社会科学』科学技術振興機構 研究開発戦略センター, pp. 36-43, 2020年11月.
- (14) 大屋雄裕「危機における個人と集団〔時論〕」『ジュリスト』1552号(2020年12月), 有斐閣, pp. 83-88, 2020年11月.
- (15) 大屋雄裕「AIとルール:マルチステークホルダー・プロセスの意味するもの〔特集:倫理と社会的責任〕」『労働の科学』2020年11月号, 大原記念労働科学研究所, pp. 22-25, 2020年11月.
- (16) 大屋雄裕(シンポジウム記録)「特集 裁判手続のIT化のこれから・市民にとって利用しやすい裁判とは」(上)『判例時報』2511号, 判例時報社, 2022年4月, pp. 126-135, (下)同2512号, pp. 101-109, 2022年5月.
- (17) 大屋雄裕(シンポジウム記録)「技術革新とエンハンスメントの時代における教育学の課題:「個別最適化された学び」は公教育に何をもたらすか」『教育学研究』89巻1号, 日本教育学会, pp. 76-83, 2022年3月.
- (18) 成原慧:媒介者責任の再検討:プロバイダ責任制限法改正および関連する取組の意義と課題, 法学セミナー66(12), pp.45-51, 2021年12月.
- (19) 成原慧:情報法ナビゲーション(第10回)Society5.0は近代の夢を見るか?, 法学セミナー66(1), pp. 53-60, 招待論文, 2021年1月.

- (20) 成原慧「AI とコンプライアンス —デジタル時代の企業と法—」『情報処理』63 巻 9 号, 情報処理学会, pp. e8-e11, 2022 年 9 月.
- (21) 新屋敷恵美子: イギリスにおける近時の代位責任 (vicarious liability) 法理の展開 (1): 就労をめぐるリスクの多様化と管理監督機能の曖昧化の観点から, 法政研究 87 巻 4 号, 2021 年 3 月.
- (22) 成原慧: 情報法ナビゲーション(第 1 回) プラットフォームはなぜ情報法の問題になるのか, 法学セミナー65(4), pp. 54-61, 招待論文, 2020 年 4 月.
- (23) 成原慧: 情報法ナビゲーション(第 7 回) データの世紀におけるプライバシー, 法学セミナー65(10), pp. 65-72, 招待論文, 2020 年 10 月.
- (24) 平山賢太郎: 情報法と競争法の相互作用: プラットフォームビジネスに対する規制の視点 (特集 情報法というフロンティア), 法学教室 (479), pp. 34-37, 招待論文, 2020 年 8 月.
- (25) 平山賢太郎: 楽天株式会社から申請があった確約計画を公取委が認定した事例 (楽天トラベル事件), 新・判例解説 Watch, 招待論文, 2020 年 7 月.
- (26) 平山賢太郎: 特定デジタルプラットフォーム取引透明化法 デジタルプラットフォームをめぐる取引の「透明化」 (特集 2020 年通常国会 成立・注目法案の影響度), Business law journal 13(6), pp. 34-39, 招待論文, 2020 年 6 月.
- (27) 平山賢太郎: 情報法ナビゲーション(第 2 回) プラットフォームと独禁法(競争法), 法学セミナー 65(5), pp. 60-65, 招待論文, 2020 年 5 月.
- (28) 新屋敷恵美子: 働き方の変化と労働法規制の意義と限界: イギリスにおける労務提供契約の不確定化に起因する諸問題を素材として, 民商法雑誌 156 巻 1 号, pp. 4-31, 招待論文, 2020 年 4 月.

6-3. 口頭発表 (国際学会発表及び主要な国内学会発表)

6-3-1. 招待講演 (国内会議 11 件, 国際会議 4 件)

- (1) 中川裕志(理化学研究所): AIとトラスト.俯瞰セミナーシリーズ.トラスト研究俯瞰.CRDS 国立研究開発法人科学技術振興機構 研究開発戦略センター, 2021年8月6日.
- (2) 中川裕志: AI 倫理とガバナンス:世界動向と産業界の取り組み.人工知能学会チュートリアル 3A4-TS-2 : 2020年6月11日.
- (3) 中川裕志:AIと倫理と社会状況. 第1回インダストリアルAIシンポジウム SIAI2020. 人工知能学会, 2020年12月16日.
- (4) 大屋雄裕: AIによる危機, AIに対する危機, 日本公共政策学会 (企画委員会セッション: グローバル・リスクと公共政策), 2021年6月5日.
- (5) 大屋雄裕: 連帯の二つの基礎: 個別性と一体性, 日本教育学会 (課題研究III・技術

革新とエンハンスメントの時代における教育学の課題), 2021年8月27日.

- (6) Shin-ichi Fukuzumi : Quality-in-Use model -considering bias between direct users and indirect users-, Bots Against Bias2021, <https://aspirelab.io/boab2021/>, Aug.9. 2021.
- (7) 佐倉統 : 科学技術は誰のもの? ——歴史的に振り返って考える, 2020年12月2日, .
- (8) 佐倉統 : 社会と文化の側から科学技術を考える, 第30回日本神経回路学会全国大会, 2020年7月29日.
- (9) Satoshi Narihara: Future of Online Speech and Content Moderation: From a Japanese Perspective, Symposium: Human Rights in the Digital Sphere, Nov. 18, 2021.
- (10) 成原慧 (九州大学) : デジタル社会におけるアルゴリズム/アーキテクチャと法, GLOCOM六本木会議オンライン#32, 2021年12月17日.
- (11) 成原慧 : AI・デジタル時代のプライバシー・個人情報保護, 第二東京弁護士会 情報公開・個人情報保護委員会定例研修会, 2022年1月21日.
- (12) Satoshi Narihara : Does Trimming of a Photograph Using AI in Retweets Infringe Moral Rights of the Author of the Photograph?, Future science & technology and legal policy joint International Conference, Feb. 25, 2022.
- (13) 成原慧 : 情報法からみたプラットフォームビジネスをめぐる問題, 九州法学会シンポジウム「プラットフォームビジネスと法」, 2022年6月19日.
- (14) 成原慧 : アメリカにおける個人情報保護と監視社会(論)の展開, 比較法学会2022年度総会シンポ, 2022年6月5日.
- (15) Satoshi Narihara : Human Autonomy and Interaction with Intelligent Robots: Toward Symbiosis between Persons and Robots, The 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022) Ethics Forum, Oct. 25, 2022.

6—3—2. 口頭発表 (国内会議 28 件, 国際会議 8 件)

- (1) 中川裕志 : AI 応用システムの社会的受容性の調査と分析. 情報処理学会 EIP 研究会 99, 2023年2月16日 (オンライン) .
- (2) 中川裕志 : 本人死後のサイバネティック・アバターに関する考察 . 情報処理学会 EIP 研究会 97(4), 2022年9月1日 (オンライン) .
- (3) 中川裕志 : AI エージェントとサイバネティック・アバターのトラスト . 情報処理学会 EIP 研究会 96(9), 2022年6月9日 (オンライン) .
- (4) 中川裕志 : 死後の個人データの公開された利用法に関する社会調査. 情報処理学会 EIP 研究会 95(15), 2022年2月8日 (オンライン) .
- (5) Hiroshi Nakagawa: Trust among data subject, AI agent and cybernetic avatar, and the service provider. 6th France-Japan Cybersecurity Workshop WG5

- “Technologies on Sanitization, Generalization and Data Mining for privacy preserving”. April 21, 2022. <https://project.inria.fr/FranceJapanICST/save-the-date-6th-franco-japanese-cybersecurity-workshop-online-april-20-22/>
- (6) Hiroshi Nakagawa:Trust relationships. Symposium “Personal AI Agent for Management and Utilization of Personal Data” at HICSS-55. Jan 4, 2022. <http://www.sict.i.u-tokyo.ac.jp/members/hasida/pai/>
- (7) 中川裕志:医療系 AI アプリに関する社会調査報告. 情報処理学会 EIP 研究会 94(2), 2021 年 11 月 8 日 (オンライン) .
- (8) 中川 裕志: 医療系 AI アプリに関する社会調査の記述意見. 情報処理学会 EIP 研究会 94(3), 2021 年 11 月 8 日 (オンライン) .
- (9) 菊池 浩明, 荒井 ひろみ, 井口 誠, 小栗 秀暢, 黒政 敦史, 千田 浩司, 中川 裕志, 中村 優一, 西山 賢志郎, 野島 良, 波多野 卓磨, 濱田 浩気, 古川 諒, 馬 瑞強, 前田 若菜, 村上 隆夫:PWS Cup 2021 - 糖尿病罹患リスクを予測するヘルスケアデータの匿名化コンテスト . 情報処理学会 CSS-2021. 4B1-1: 2021 年 10 月 29 日.
- (10) 中川 裕志:死後の個人データのビジネス化 . 情報処理学会EIP研究会 93(11), 2021 年9月9日 (オンライン) .
- (11) 中川裕志: A I の基本と競争政策: 公正取引委員会 AI 講習会 (デジタルアドバイザーとして講演) 2021 年 11 月 12 日.
- (12) 吹金原榛耶, 村瀬達也, 高橋達二, 中川裕志: 個人の購買行動を支援するパーソナル AI エージェントの試案(Experimental idea of personal AI agent to support individual purchasing behavior). 人工知能学会大会 4H2-GS-11c-02 , 2021 年 6 月 11 日.
- (13) 中川裕志: AI エージェントの法的位置づけの言説の経緯 , 情報処理学会 EIP 研究会 92(14), 2021 年 6 月 8 日 (オンライン) .
- (14) 中川裕志: 「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」を読む , 情報処理学会 EIP 研究会 92(3), 2021 年 6 月 7 日 (オンライン) .
- (15) 中川裕志: パーソナル AI エージェントの構造. サービス学会第 9 回国内大会 C000076. 2021 年 3 月 9 日.
- (16) Hiroshi Nakagawa: Twist of Contact Notification App for COVID-19. French Japanese Cybersecurity Intermediate Workshop 2021. Feb. 25, 2021(online).
- (17) 加藤綾子, 中川裕志: パーソナル AI エージェントの社会制度的位置づけ, EIP 研究会 91(25), 2020 年 11 月 26 日 (オンライン).
- (18) 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, EIP 研究会 91(26), 2020 年 11 月 26 日 (オンライン)

- (19) 中川 裕志: 接触通知アプリの捻じれ , 情報処理学会 EIP 研究会 91(28), 2020 年 11 月 26 日 (オンライン).
- (20) Satoshi Narihara (九州大学) : AI Governance in Japan: Lessons from Experiences in the COVID-19 era, Workshop-AI Governance in the UK and Japan in the COVID-19 era, Jun 2, 2021
- (21) 佐倉統: ロボット, 浮世絵, 類人猿——AI への文化的アプローチ—— , 「高齢者と対話ロボットのコミュニケーションに関する量的・質的調査研究(代表: 高木美也子)」2020 年度第 2 回講演, 2020 年 10 月 1 日.
- (22) 佐倉統: 科学技術を社会の中で考える, 日本学術会議公開シンポジウム「新知見の扱いとその活用」, 日本学術会議総合工学委員会原子力安全分科会, 2020 年 9 月 10 日.
- (23) Shin'ichi Fukuzumi, Thomas Geis, Jonathan V Earthy: What is the difference between usability in ISO 25000 and ISO 9241-11? -The term “usability” as part of the Product Quality model used in the SQuARE series in contrast to the term usability in ISO 9241-11-, IWESQ2022, Dec. 6, 2022.
- (24) 福住伸一, 小林大二, 平沢尚毅: サービスデザインにおける RIAS の活用とステークホルダ分析, HI 学会 UX とサービスデザイン研究会, 2022 年 11 月 17 日, オンライン.
- (25) 福住伸一: 人間工学に関連した研究倫理の課題, 日本人間工学会北海道支部大会, 2022 年 11 月 26 日, 小樽商科大学.
- (26) 福住伸一: 没入環境に関する人間工学的標準化の動向, 日本バーチャルリアリティ学会全国大会, 札幌市立大学, 2022 年 9 月.
- (27) 福住伸一, 笠松慶子, 澤田享, 大林史明: ABW の導入効果指標化, 日本人間工学会全国大会, 尾道市, 2022 年 7 月.
- (28) 福住伸一: HCI 関連人間工学標準化の最新動向, 日本人間工学会全国大会, 尾道市, 2022 年 7 月.
- (29) Fukuzumi, S. and Wada, N.: Quality in Use -Case Study for Evaluation-, Human-Computer Interaction. Theory, Methods and Tools, Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24-29, 2021, Proceedings, Part I, Editors: Kurosu, Masaaki (Ed.) pp. 343-350, 2021.
- (30) 福住伸一・佐倉統・神野真理子・稲垣香澄・澤虹之介・野田夏子・中川裕志: ロボットイメージと社会受容性, 第 34 回人工知能学会全国大会, 2020 年 06 月 10 日.
- (31) Shin'ichi Fukuzumi, Hiroshi Nakagawa, Osamu Sakura: User Experience in smart city -UX and safety management for self-driving car, 'Smart Cities at Play Workshop Lived Experiences, Emerging Forms of Playfulness, and Problems of Participation' CHI' 20 workshop, Hawaii, April 25th to 30th, 2020.

- (32) Fukuzumi, S., Wada, N. and Hirasawa, N.: Quality in Use -Issues and proposal-, IWESQ2020, <http://ceur-ws.org/Vol-2800/paper-07.pdf>, 2020.
- (33) 水上拓哉: ソーシャルロボットの倫理のための概念工学——道徳的行為者性の虚構性をめぐって, 人工知能学会全国大会, 2020年6月
- (34) 水上拓哉: デジタルクローンの行為者性とその設計の倫理, 科学技術社会論学会, 2021年12月.
- (35) 水上拓哉: 対話システムの倫理における設計者の道徳的責任について, 人工知能学会全国大会, 2022年6月.
- (36) 寺田麻佑, 成原慧, 松尾剛行, : オンラインプラットフォーム上におけるプライバシー情報収集規制の日本・EU・米国・中国における規制の最新動向と比較, 情報ネットワーク法学会第22回研究大会, 2022年12月4日.

6-3-3. ポスター発表 (国内会議 0 件、国際会議 0 件)

6-4. 新聞/TV 報道・投稿、受賞など

6-4-1. 新聞/TV 報道・投稿

- (1) 中川裕志: 「AIの悪用、可能性は常に」、毎日新聞 2020年8月12日朝刊 オピニオン、戦争と科学
- (2) 中川裕志: 国営放送 (YLE) のニュース Yle Uutiset <https://yle.fi/uutiset/3-11818326> 2020年3月11日. 在フィンランド日本大使館で作成したこの記事に英語版(仮訳)は https://drive.google.com/file/d/1Yg5e_pmph9p4QQFhtyq_3tdEZfgGutbv/view
- (3) 中川裕志: 毎日新聞、2020年2月6日、機械学習と公平性に関するシンポジウムの報告記事
- (4) 中川裕志: 毎日新聞、2020年1月22日、毎日メディアカフェの報告記事
- (5) 水上拓哉: 「ソーシャルロボットの社会浸透を「哲学」する」, 『THE ROBOT』, 3月号, 日刊工業新聞社, 2022年3月
- (6) 水上拓哉: 「『ロボットいじめ』はプロセス、ロボット『倫理』研究者が語る社会浸透への課題」, ニューススイッチ, 2022年3月, URL (<https://newswitch.jp/p/31171>)
- (7) 佐倉統: 「生活を豊かに」の視点で、毎日新聞 9 2021年11月10日
- (8) 佐倉統: 生活者の役割・関与重み増す、日本経済新聞 2021年8月13日
- (9) 佐倉統: 感染症と人類——いがみ合うか励まし合うか、朝日新聞 13 2020年4月4日
- (10) 佐倉統: 今、求められる総合知とは、Science Window 15(2), 3-5
- (11) 大屋雄裕 「危機下の民主主義: 次の危険への備え平時に〔経済教室〕」日本経済新聞

2020年6月8日朝刊

- (12)大屋雄裕「社会のあり方を問い直す機会に」日本経済新聞 2020年5月10日 朝刊
- (13) 成原慧:「家賃滞納予測AIがプチ炎上、「公平性」巡る問題は技術で解決できるのか、日経BP 日経クロステック、2021年8月17日
- (14)成原慧:トランプ氏のアカウント凍結、ツイッター社の責任とは、朝日新聞デジタル、2021年2月23日
- (15)成原慧:「トランプのアカウント凍結」と「メルケルのツイッター批判」、多くの人が誤解していること、現代ビジネス、2021年1月18日
- (16)成原慧: AI はなぜ“差別”するのか?: AI 時代の“人間”の生き方~シンポジウム「AI と差別」成原慧さん基調講演、日刊サイゾー、2020年7月15日
- (17)成原慧:感染症対策のための規制、ナッジ、データそして民主主義、SYNODOS、2020年4月28日

6-4-2. 受賞

- (1) 編者 北川 源四郎、竹村 彰通、著者:内田 誠一、川崎 能典、孝忠 大輔、佐久間 淳、椎名 洋、中川 裕志、樋口 知之、丸山 宏、日本統計学会 第15回日本統計学会出版賞;「教養としてのデータサイエンス(データサイエンス入門シリーズ)」.2022年5月13日
- (2) 中川裕志:令和2年度 情報通信月間推進協議会 情報通信功績賞 :2020年5月29日
- (3) 福住伸一:経済産業省産業標準化事業表彰-経済産業大臣賞、2021年10月20日
- (4) 鱗原 晴彦、仲谷 尚郁、福住 伸一、山口 恒久、吉武 良治:HCD 研究発表会優秀講演賞「自動運転社会における HAI (Human AI Interaction) の考察」、2020年6月6日

6-4-3. その他

- (1) 以下の報告書の内容作成の審議に参加し、修文などを行った。公正取引委員会:デジタル市場における競争政策に関する研究会 報告書「アルゴリズム/AIと競争政策」が公開され、
https://www.jftc.go.jp/houdou/pressrelease/2021/mar/210331_digital.html
さらに公正取引委員会デジタルスペシャリストアドバイザーとして、この報告書の解説講演会などを2021年、2022年の2回行った。

6-5. 特許出願

6-5-1. 国内出願 (0 件)

6-5-2. 海外出願 (0 件)