戦略的創造研究推進事業
（社会技術研究開発）
令和元年度研究開発実施報告書

「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、エージェンシー、トラストの文化を超えた実現方法」

中川 裕志

（国立研究開発法人理化学研究所・革新知能統合研究センター、チームリーダ）

目次

１．研究開発プロジェクト名

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

２．研究開発実施の具体的内容

２－１．研究開発目標

- 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)
  - 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する。
  - 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。
- AIシステムのトラスト
  - AIを含むシステムのトラストの在り方を調査、分析し、社会に受け入れられるトラストの在り方、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。
- AIを含む社会のガバナンス
  - 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。

２－２．実施内容・結果

（１）スケジュール

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～ 2022.12)
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（主に英国側）			← →	
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				← →

（２）各実施内容

（目標）AI倫理指針についての世界的な発展方向の調査・分析および具体化

実施項目①-1：　AI倫理規範の背景調査

実施内容：

2017年のFLI Asilomar 23原則からOECD閣僚理事会で承認された Recommendation of the Council on OECD Legal Instruments Artificial Intelligenceまでに提案、公開されたAI倫理指針における主要論点を分析する。

変更点としては、これらの指針に加えて、総務省　AIネットワーク社会推進委員会　AI利活用ガイドライン、Guidance for Regulation of Artificial Intelligence Applications: USA Whitehouse. MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIESも調査、分析の対象に加えた。

実施項目①-2：パーソナルAIエージェントの設計

実施内容：

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。

変更点としては、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とである。

佐倉先生

（目標）AIの文化、社会的な側面を分析し提言する。

実施項目②-1：文化、社会の歴史的背景調査、分析、提言

グループの役割の説明：

人間-AI共生エコシステムにおける重要概念であるプライバシー（privacy)、エージェント（agency)、トラスト（trust)、セキュリティ／リリーフ（securoity/relief）などの内実が日英でどのように異なるかを、文化的諸領域において比較検討し、健全なエコシステムの形成に寄与する方法を考察するための準備作業として、比較検討する領域と項目の選定をおこなう。

実施項目②-2：美的感覚の調査とAIによる学習

グループの役割の説明：

②−1で抽出・選定した諸概念が日英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、2019年度は研究対象とする文学作品や芸術作品の選定、注目すべき項目、ファッション界の実態の予備調査などをおこなう。

大屋先生

（目標）AIの社会における法的側面を明確化する

実施項目③-1：政策、法制度、経済性に関する基礎調査

グループの役割の説明：

主要なAI倫理指針や個人情報の利活用に関連する提言・ガイドライン類においてトラストの概念がどのように用いられているか、どのような意味を持つものと想定されているかに関する調査を行なう。

特に変更はなく進行している。ただし、特にガバナンスとの関連については実施項目④と強く関係するため合同での研究会開催を計画したが、新型コロナウイル

ス問題により当面延期となった。

実施項目③-2：文化、社会の歴史的背景調査、分析、提言
グループの役割の説明：
より一般的な信用としてのトラストについて、社会理論においてどのように扱われてきたかについて調査する。イングランドにおいてトラスト（信託）制度が必要となった制度的条件について調査する。
特に変更はなく進行している。イングランド法制度に関する基礎的な調査は終了した。

実施項目③-3：⑦既存のガバナンス枠組の網羅的調査、分析、提言
グループの役割の説明：
イングランド法における法制度としてのトラスト（信託）と、それを含む一般的概念としての信認（fiduciary）、その重要な要素としてのaccountability（説明責任・答責性）について、先行研究を調査する。
特に変更はなく進行している。基礎調査をもとに、システムを利用する根拠となる信頼性をcredibilityとやや中立的に位置付け、その実現手法を利用者・被用者の知識水準の差異によって分類する基本的な枠組を提案した。

成原先生
（目標）AIのガバナンスのモデル化を行う
実施項目④-1：政策、法制度、経済性に関する基礎調査
グループの役割の説明：
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）のについて調査・検討を行う。

実施項目④-2：既存のガバナンス枠組の網羅的調査、分析、提言
グループの役割の説明：
AIエージェントのガバナンスの枠組の構成要素となる法規範、倫理規範、アーキテクチャの間の役割分担と相互作用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枠組みを構想・提示する。

実施項目④-3：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
グループの役割の説明：
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。

（３）成果

目標）AI倫理指針についての世界的な発展方向の調査・分析および具体化

実施項目①-1：　AI倫理規範の背景調査

成果：

以下の重要なAI倫理指針を調査し分析し、AI倫理の方向性を抽出できた。

① FLI: Asilomar AI Principles (23原則) (2017)
② IEEE Ethically Aligned Design(EAD), version 2(2017/12)
③ Partnership on AI (2016~
④ 総務省　AIネットワーク社会推進委員会　AI開発ガイドライン(OECDに提案(2017))
⑤ 内閣府　人間中心のＡＩ社会原則(2019/3/29) AI ready な社会の在り方 G20に提案
⑥ IEEE Ethically Aligned Design(EAD), first edition (2019/3)
⑦ EU: High Level Expert Group: Ethics Guidelines for Trustworthy AI (2019/4/8)
⑧ Recommendation of the Council on　OECD Legal　Instruments　Artificial Intelligence OECD　閣僚理事会承認　(2019/5/22)
⑨ 総務省　AIネットワーク社会推進委員会　AI利活用ガイドライン(2019)
⑩ Beijing AI Principle (2019/5/25)
⑪ Guidance for Regulation of Artificial Intelligence Applications: USA Whitehouse. MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES (Draft 2019/4/24)

分析結果を以下の表にまとめることができた。名宛人は、
AIの基礎技術の開発者を　基、実用に供するAIシステム開発者を　実、AIシステムの末端利用者を　利、AI事業者を　業、政策立案者を　政、と略記した。

	AIの脅威と制御	法的位置づけ	安全性	プライバシー	AIエージェント	悪用、誤用	予見性	透明性、説明可能性	アカウンタビリティ	トラスト	公平性非差別、バイアス	文化的多様性の許容	教育	政策環境	遵法性	軍事利用	名宛人
Asilomar AI Principles	○			○												○	実政
人工知能学会・倫理指針	△	○	○	○		○											基実
総務省AI開発ガイドライン	○		○	○				○	○		○						業実
Partnership on AI			○	○	○			○	○		○		○	○			業実
IEEE EAD ver2	○	○	○	○	○	○	○	○	○	○	○	○	○	○		○	基実業政
IEEE EAD 1e		○	○	○	○	○	○	○	○	○	○	○	○	○			基実業政
人間中心AI社会原則			○	○		○		○	○	○	○	○	○	○			実業
Trustworthy AI			○	○	○	○	○	○	○	○	○	○	○	○		○	基実業政
OECD Recommendation			○	○		○	△	○	○	○	○	○		○			実業政
総務省AI利活用ガイドライン			○	○		○	△	○	○	○	○	○					業利

Beijing Principle	○		○	○				○	○	○		○	○	○			実業
Whitehouse Guidance	×		○	○		○	○	○	○	○	○	○		○			業

表中、予見性における△はリスクという単語のみで表現されている場合である。Whitehouse GuidanceのAIの脅威と制御の項目に×がついているのは、人間の脅威になる強いAIはこの指針では視野に入っていないと言い切っているからである。

この表から読み取れる解釈を以下に述べる。

・　　AI脅威論の退潮

AI倫理指針が注目されるようになった契機の一つともいえるAI脅威論に関しては、AI制御（AIを人間の制御できるように設計する）という項目が対応する。これは初期のAsilomar AI Principleでは取りあげられ、「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」「発達したAIは地球生命の歴史に重大な変化を及ぼすかもしれないため、相応の配慮と資源を用意して計画、管理しなければならない」「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」とまで書かれている。以降、IEEE EAD version2　までAI制御について倫理指針の一部に記載されたものの、それ以降は全く取り上げられなくなっている。つまり、人間の脅威になるようなAI、AGI 、超知能は、当面あるいは遠い将来まで実現しそうにないことがAI研究者たちの間でコンセンサスを得たことによると考えられる。

・法的位置づけと遵法性

AIに人格権を与えるか否かという法的位置づけは、AIには人格権を与える根拠が当面はないというコンセンサスが形成され、人間中心AI社会原則以降は取りあげられなくなった。一方、現状のAIあるいは現在に機械学習技術を用いて開発されるAIは現行法を遵守するべきであるという主張はTrustworthy AI以降、しばしば取り上げられている。これは一見、当然すぎるようにみえるが、その一方でAIの能力や機械学習の技術が未知だった時代に作られた法律は、技術に適合しなくなってきており、むしろ法律を変えるべきだという主張もある。

・予見性

IEEE EAD version2以降の多くのAI倫理指針で予見性の必要性が指摘されている。予見性とは主として悪い状態の発生に係わることが多い。したがって、より具体的

にリスクという用語で表現されている場合が多い。リスクの予見ないし予知は機械学習における研究、開発のテーマであるが、AIシステム構築の費用を増大させ、開発企業にとってはありがたい概念ではない。にもかかわらず、多くのAI倫理指針で言及されるようになったことは、予見性がAIシステムの実用において重要な要件になってきたこと、および機械学習の評価技術に進展が要因であろう。

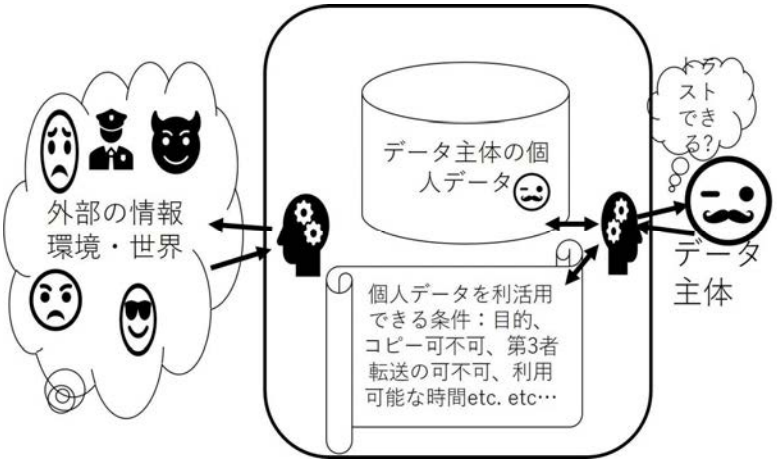
・透明性、説明可能性、アカウンタビリティ、トラスト
これらの諸概念は初期のAsilomar AI Principle、人工知能学会倫理指針では取り上げられなかった。しかし、AIの実用化の進展を鑑みて、AI関係者はその重要さにすぐに気付き、それ以後のAI倫理指針では継続的に取り上げられている。直観的にその必要性が理解しやすい透明性、説明可能性、およびアカウンタビリティはほぼ同時期に取り上げられるようになった。ただし、これらの実装は機械学習の技術によっても困難なものであることが理解され始めたため、社会学的な信用の近い概念であるトラストを終着点に置くのは少し時間遅れがでたと考えられる。

実施項目①-2：パーソナルAIエージェントの設計
成果：

エージェントは代理人の意味なので、AIエージェントは、個人、集団、組織などの代理を勤めるプログラムにAIの知的機能が組み込まれたものである。このようなAIエージェントは今やいたるところに存在している。ITプラットフォームの個人に対する知的なインタフェースはITプラットフォームのAIエージェントと見なすこともできる。

さて、重要なエージェントは個人の代理をするパーソナルAIエージェントである。個人データの生成者すなわちデータ主体が持っている種々の個人データ、たとえば購買履歴、移動履歴、医療ないし健康状態履歴、などを外部の事業者がサービスと引き換えに使いたいとデータ主体の個人にアクセスしてくることは頻繁に起こる。プライバシーを守るためには、それらの全ての慎重に対応しなければならないが、これは容易なことではない。たとえば、ITプラットフォームにどこまで自分の個人情報を与えるかは判断に迷うこともしばしばである。そこで、この判断を支援、ないし代理してくれるのがパーソナルAIエージェントである。

データ主体のパーソナルAIエージェントは、データ主体のそれまで個人データを集積して保持、管理していることに加えて、それらを外部からの要請があったとき、どのように使用許諾したかの利用許可履歴から学習した個人データ利活用条件のデータベースも同時に保持している。パーソナルAIエージェントの構造を図に示す。



なお、新規に開始した誕生以前、死後の個人データをパーソナルAIエージェントで扱う方法は、今年度はアイデアの想起であり、具体的な検討は次年度に行う予定である。

（目標）AIの文化、社会的な側面を分析し提言する。
実施項目②-1：文化、社会の歴史的背景調査、分析、提言
成果：
人間-AI共生エコシステムにおける重要概念であるプライバシー（privacy)、エージェント（agency)、トラスト（trust)、セキュリティ／リリーフ（securoity/relief）などの内実が日英でどのように異なるかを、文化的諸領域において比較検討し、健全なエコシステムの形成に寄与する方途を考察するための準備作業に着手した。日英間で比較検討する領域と項目の選定を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イギリス側との共同討議は行えなかった。

実施項目②-2：美的感覚の調査とAIによる学習
成果：
②−1で抽出・選定した諸概念が日英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、2019年度は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。これらを対象に、AIやロボットの描かれ方や作品の受容が日英でどのように異なるかを比較分析する予定である。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。

AIの社会における法的側面を明確化する
実施項目③-1：政策、法制度、経済性に関する基礎調査
成果：
主要なAI倫理指針や個人情報の利活用に関連する提言・ガイドライン類におけるトラストの概念について基本的な調査を行ない、あまり明確な分析がされていないこと、その一方で重要性は強く意識されていること、利用者側の概念であるtrustとシステム側の概念であるtrustworthinessのあいだにある差異ないし乖離が意識されつつも十分には分析されていないこと、などを知見として得ている。

実施項目③-2：文化、社会の歴史的背景調査、分析、提言
成果：
イングランド法上のトラストが成立した背景については、田中英夫『英米法総論〔上下〕』（東京大学出版会、1980）などの基礎文献による確認を終えている。具体的には、伝統的なイングランド法（狭義のcommon law）においては制限権利能力者（女性・未成年者）などが土地を保有すること、また土地所有権に条件を付すことが認められていなかったため（封建的義務の代償としての性質を持つため、と一般的には説明されている）、未成年の相続人しか存在しないような場合に、被相続人が保有する土地所有権を形式上は友人など信頼すべき人物（受託者）に対して無条件で譲渡し、相続人が成人に達したのちに返還することを信じて託すという行動が発生した。しかし、受託者がこの信頼に背いて行動した場合（被相続人への返還を拒んだ場合）、上述の通りcommon law上は受託者の所有権が無条件に成立しているため、裁判などを通じた救済を受けることができない。このため、国王に対する訴願によりある意味で超法規的に救済を得ようと相続人が考え、教会の首長である国王の宗教的権威を活用することにより「他者の信頼に応える」という倫理的な行為が実質的に強制されることになった。これがcommon lawと呼ばれる法体系（広義）のもう一つの構成要素である「衡平法」（equity）と呼ばれるものの歴史的基礎であったと考えることができる。
ここからは、イングランド法におけるトラストが。相続人・被相続人の法的無能力と受託者の形式的には無限定な所有権という能力的な極度の非対称性を背景にしていること、そのような状況で実質的に正義にかなった解決を保障するために、「信認関係」（fiduciary）と呼ばれる非対称的な法的関係の存在を想定していること、などを読み取ることができる。AIに対するトラストについても、このような状況の有無や程度によって分析の精度を高めることができると想定される。

実施項目③-3：⑦既存のガバナンス枠組の網羅的調査、分析、提言
成果：
基礎調査をもとに、システムを利用する根拠となる信頼性をcredibilityとやや中立的に位置付け、その実現手法を利用者・被用者の知識水準の差異によって分類する基本的な枠組を提案した。具体的には以下のようになる。
まず、利用者の側に十分に高い知識・能力水準がある場合に、自分の代わりに被用

者に行為させる類型が「代理」（agency）であり、被用者の行為の適切性を利用者が自ら判断することが十分に可能なので、透明性transparencyが機能することになろう。
これに対し、被用者の側に高い知識・能力水準がある場合が「信認関係」であり、利用者による判断や自己決定が十分に機能すると期待できないので、何らかの外的な統制を被用者に対して加える必要がある。このうち、依頼される行為が定型的・反復的であって当該分野についての十分な能力や行為の適切性を有していることを事前に検証できる場合が「権威」（authority）であり、同業者団体や国家による能力証明（認証・許可・免許など）によって被用者の行為を基礎付けることができる。弁護士や医師など、古典的な専門家をこの典型として想定することができよう。
最後に、利用者の期待が多様であったり、状況の変化が激しいなどの理由で定型性・反復性が十分に成立しない場合が狭義の「信託」（trust）であり、事前の検証も成り立たないため事後の正当化・検証作業が必要になると予想される。このように事後的な正当化を支えるのが「説明可能性・答責性」（accoutability）であり、正当理由なく利用者の期待に反する結果が生じた場合に責任を負うシステムだと想定することができる。
以上のような分析から、AIシステムに対するトラストを問題にする場合であってもAIと利用者の知識・能力水準の格差や利用形態の定型性によって異なる統制手法を採用すべきであるという見解が示唆される。

（目標）AIのガバナンスのモデル化を行う
実施項目④-1：政策、法制度、経済性に関する基礎調査
成果：
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になるとの見通しを得た。

実施項目④-2：既存のガバナンス枠組の網羅的調査、分析、提言
成果：
日英を中心に新型コロナウィ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。

実施項目④-3：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
成果：
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、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の間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。

（４）当該年度の成果の総括・次年度に向けた課題

○今年度の研究開発を総括し、以下の点について簡潔に記載してください。

・AI倫理、トラスト、ガバナンスのグループの活動は当初の目標を十分に達したと考えられる。AIと文化に関しては、概ね当初の予定通り進んだが、英国との共同研究と調査が新型コロナ感染症の蔓延により実施できなかった。

・まだ、具体的に大きな成果はでていないが、新型コロナ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、AI技術、倫理的視点、法制度などを総合的に組み合わせた研究が必要であることが分かってきた。次年度以降、このプロジェクトの理論的成果の応用局面として社会への貢献という意味に含めて進めていくべきと考えている。

・　新型コロナ感染症の蔓延により、英国との間での実際の移動を伴う研究はしばらくの間はできそうもない。そこで、本プロジェクトでは、オンライン会議（主としてZoomを利用）で代用する方針として、新年度からオンライン会議での集中的な議論を行っている。オンライン会議はすでに3回行い、十分な効果を得られており、今後も継続する予定である。

２－３．会議等の活動

○実施体制内での主なミーティング・ワークショップ等の開催状況について記入してください。

年月日	名称	場所	概要
2020 年 1 月 21日	AIと文化を考える公開シンポジウム《AIと身体性、AIの身体性、AIと社会の身体性》	理化学研究所革新知能統合研究センター	AIをめぐる文化差を分析する際に重要なトピックのひとつである身体性について、人体拡張研究、能楽、発達認知科学の専門家を招いて討論した。
2020年2月5日	富士通　AI教育コンテンツ開発ミーティング	富士通・汐留オフィス	中川が富士通が作成するAI倫理に基づくAI倫理教育コンテンツ内容のコンサルティングを行った
2020 年 3 月 16日	公正取引委員会とのAI倫理打ち合わせ	千代田区霞が関：公正取引委員会	AI倫理と個人情報の扱いに関する公正取引委員会からのヒアリングに中川が対応した。
2020 年 2 月 27日	AI型社会的ジレンマの共同研究	新型コロナ対応のため	AIの社会応用に伴う倫理的問題の類型化について、検討した。

	打ち合わせ	Slack にて行なう	
2020 年 1 月 13日	AI法制度グループ内打ち合わせ	商事法務研究会会議室	研究遂行の方法、グループ間討議の可能性について検討した。
2020 年 2 月 10日	デジタル規制改革ヒアリング	行政改革推進本部	内閣官房行政改革推進本部のヒアリングに大屋が対応した。

３．研究開発成果の活用・展開に向けた状況

AI倫理に関しては、機械学習の公平性シンポジウムなどの公開シンポジウムを行い、新聞でも報道されるなど大きな反響をえることができた。また、書籍に関しては、AI倫理、プロファイリング、法的アーキテクチュアの各々のついて刊行し、さらに論文もAI倫理と社会、生物学的観点から発表した。これらは、AI倫理などの分野に興味を持つ必ずしも専門家でない方々も射程に入れており、社会へのアウトリーチとしては大きな影響を持つも思われる。

一方、実際に聴衆を集めての講演会では各グループから多数の多様な発表を行い、本プロジェクトの成果を多数の人々と共有できたと考えられる。

４．研究開発実施体制

（１）AI倫理グループ
①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）
②実施項目

実施項目①-1：　AI倫理規範の背景調査

実施項目①-2：パーソナルAIエージェントの設計

（２）AI文化グループ（佐倉統）
①佐倉統（東京大学大学院情報学環、教授）
②実施項目

実施項目②-1：文化、社会の歴史的背景調査、分析、提言
実施項目②-2：美的感覚の調査とAIによる学習

（３）AI法制度グループ
①大屋雄裕（慶應義塾大学法学部、教授）
②実施項目

実施項目③-1：政策、法制度、経済性に関する基礎調査
実施項目③-2：文化、社会の歴史的背景調査、分析、提言

実施項目③-3：⑦既存のガバナンス枠組の網羅的調査、分析、提言

（4）AIガバナンスグループ

①成原慧（九州大学法学研究院 、准教授）

②実施項目

実施項目④-1：政策、法制度、経済性に関する基礎調査
実施項目④-2：既存のガバナンス枠組の網羅的調査、分析、提言
実施項目④-3：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

５．研究開発実施者

AI倫理グループ

氏名	フリガナ	所属機関	所属部署	役職（身分）
中川裕志	ナカガワヒロシ	理化学研究所	革新知能統合研究センター	チームリーダ
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	特任講師
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	准教授
吉田光男	ヨシダ　ミツオ	豊橋技術科学大学	大学院工学研究科	助教
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	特任講師
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

AI文化グループ

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	大学院生
水上　拓也	ミズカミタクヤ	東京大学	大学院 学際情報学府	大学院生
Wang Yuhui	ワン　ユーフイ	東京大学	大学院 学際情報学府	大学院生
藤嶋　陽子	フジシマヨウコ	ZOZO研究所		リサーチサイエンティスト

AI法制度グループ

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
工藤　郁子	クドウ　イクコ	東京大学・未来ビジョン研究センター		客員研究員
藤田　卓仙	フジタ　タカノリ	慶応義塾大学		特任講師

AIガバナンスグループ

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
小島　立	コジマ　リュウ	九州大学		准教授
平山　賢太郎	ヒラヤマケンタロウ	九州大学		准教授

赤坂　幸一	アカサカ コウイチ	九州大学		准教授
冨川　雅満	トミカワ マサミツ	九州大学		准教授
新屋敷　恵美子	アラヤシキ エミコ	九州大学		准教授

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．シンポジウム等

年月日	名称	場所	参加人数	概要
2020年1月9日	機械学習と公平性に関するシンポジウム	一橋講堂	350	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
2020年3月19日	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	160(Web参加)	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

６－２．社会に向けた情報発信状況、アウトリーチ活動など

（１）書籍・冊子等出版物、DVD等

・ 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

・プロファイリング・理由・人格、大屋雄裕、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

･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

（２）ウェブメディアの開設・運営

・RIKEN AIP researchers talked at the Symposium on Machine Learning and Fairness、https://aip.riken.jp/news/mlfsymposium200218/、2020年2月18日

（３）学会（6-4.参照）以外のシンポジウム等への招聘講演実施等

・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

・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

・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

・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

・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

・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

・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

・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

・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

・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

６－３．論文発表

（１）査読付き（ 2 件）

●国内誌（ 2 件）

・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), I-1,23 、招待論文、2020 年

・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
(https://gendai.ismedia.jp/articles/-/71191)

●国際誌（ 0 件）

・

（２）査読なし（ 0 件）

・

６－４．口頭発表（国際学会発表及び主要な国内学会発表）

（１）招待講演（国内会議 0 件、国際会議 0 件）

・

（２）口頭発表（国内会議 0 件、国際会議 0 件）

・

（３）ポスター発表（国内会議 0 件、国際会議 0 件）

・

６－５．新聞／ＴＶ報道・投稿、受賞等

（１）新聞報道・投稿（ 2 件）

・毎日新聞、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

・毎日新聞、１月22日、毎日メディアカフェの報告記事

（２）受賞（ 0 件）

・

（３）その他（ 0 件）

・

６－６．知財出願

（１）国内出願（ 0 件）

・

（２）海外出願（ 0 件）

・