社会技術研究開発事業
研究開発実施終了報告書

科学技術の倫理的・法制度的・社会的課題（ELSI）への
包括的実践研究開発プログラム

研究開発プロジェクト

「人工知能の開発・利用をめぐる自律性および関係性の理論分析と社会実装」

研究開発期間
（２０２１(令和３)年１０月１日 ～ ２０２５(令和７)年３月３１日）

宇佐美　誠
（京都大学 大学院地球環境学堂 教授）

目次

Ⅰ サマリー 2
Ⅱ 本編 4
1. プロジェクトの目標と背景 4
1-1. プロジェクトの達成目標 4
1-2. 研究開発の対象・テーマ、その意義と背景 4
（1）研究開発の対象 4
（2）創出したアウトプット 5
（3）本研究開発プログラムにおける位置付け 5
（4）本プロジェクトの意義・必要性 5
1-3. プロジェクトが考察する根源的問い 6
1-4. プロジェクトの先に見据える将来展望 6
2. 研究開発の実施内容 6
2-1. 実施項目 6
2-2. 実施内容 7
2-3. 研究開発マネジメント 12
3. 研究開発の成果 14
3-1. プロジェクト目標における研究開発成果 14
3-2. プログラム目標における成果・示唆 16
3-3. 成果の活用・展開状況や今後の展望 17
3-4. 成果の他分野への適応可能性・知見の一般化可能性 17
4. 研究開発実施体制 18
4-1. 研究開発実施体制図 18
4-2. 研究開発実施者 19
4-3. 研究開発の協力者 20
5. 研究開発成果の発表・発信状況、アウトリーチ活動など 21
5-1. 社会に向けた情報発信状況、アウトリーチ活動など 21
5-2. 論文発表 23
5-3. 口頭発表（国際学会発表及び主要な国内学会発表） 26
5-4. 新聞/TV 報道・投稿、受賞など 28
5-5. 特許出願 28

Ⅰ サマリー

プロジェクト概要

本プロジェクトは、人工知能（AI）と人・社会の調和という終極理念の下、次の 2 点を具体的達成目標とした。

① AIの倫理的課題の理論分析：AI・ビッグデータ分析の技術開発や企業の開発者との共同探索的会合を通じて、開発・利用上の倫理的課題を把握し、人間・AI の自律性および人間－AIの関係性という観点から構造化した。次に、AI倫理学の国際学界での先端的研究成果を活用するとともに、日本の状況の特徴に留意しつつ、新たな規範理論を構築し、理論分析を実施した。これらの作業を通じて、AIの倫理的課題について精緻な規範的分析を行うことにより、本プロジェクトの理論的目標を達成した。

② 技術開発者の行動指針の策定：倫理的課題の理論分析の成果を基盤に、ステークホルダーとの連携の下、AI の開発・利用の現実やそれを取り巻く日本社会の特徴を踏まえつつ、開発者がとるべき行動や、開発・利用を行う企業が設けるべき制度に関して、行動指針（CoC）を公表した。これにより、責任ある研究・イノベーション（RRI）の制度化に向けた本プロジェクトの実践的目標を達成した。

〈参画・協力機関〉

京都大学、九州大学、共立女子大学、慶應義塾大学、東京大学、日本大学、名古屋大学、南山大学、武蔵野大学、明治学院大学
アイルランド・ゴールウェー国立大学、カリフォルニア大学バークレー校、台湾中正大学、ニュー・スクール、フリードリヒ・アレクサンダー大学エアランゲン＝ニュルンベルク、ミュンヘン工科大学

〈キーワード〉

AI、ビッグデータ、ロボット、倫理、行動指針

結果・成果の概要

本プロジェクトの結果・成果は、次の 3 点に要約できる。

（1）自律性をめぐる AI の倫理的課題の理論分析

人間を取り巻く環境を構成している AI が、個人の自律性に与えつつある影響を発見・予見した上で、技術開発や企業の開発者との共同探索的会合を通じて、開発・利用上の倫理的課題を把握した。次に、AI 倫理学の国際学界での先端的研究成果を活用するとともに、日本の状況の特徴に留意しつつ理論分析を行い、その成果を書籍や学術論文を通じて公表した。

（2）関係性をめぐる AI の倫理的課題の理論分析

AI が、個人間の関係性に与えつつある影響についても、技術開発や開発者との共同探索的会合を通じて、開発・利用上の倫理的課題を捕捉した。次に、国際学界の先端的研究成果を活用し日本の状況にも留意しつつ理論分析を進め、その成果を書籍や論文により公表した。

（3）行動指針の策定

2 種類の理論分析の結果に基づいて、民間企業による採用が推奨される新たな行動指針を策定し、公表した。

問い合わせ先

京都大学大学院地球環境学堂地球環境政策論分野
https://www.envpolicy.ges.kyoto-u.ac.jp/

Summary

As artificial intelligence (AI) has been rapidly developed, humanity has met an increasing number of philosophical and moral challenges including new forms of threat against individual autonomy, the transformation of intimate relationships between persons, and risks from autonomous machines. This research project, carried out by a multidisciplinary team consisting of legal, moral, and political philosophers as well as data scientists and engineers, uncovered or foresaw ethical challenges raised in the contexts of AI development and deployment. The team then analyzed these challenges from the perspectives of individual and AI autonomy and of human affinity, utilizing recent findings in the literature of AI ethics and considering the context of Japanese society. Finally, it proposed the code of conduct for AI engineers to actualize the principle of responsible research and development.

〈Joint R&D Organizations〉

Kyoto University, Keio University, Kyoritsu Women's University, Kyushu University, Meiji Gakuin University, Musashino University, Nagoya University, Nanzan University, Nihon University, University of Tokyo
Friedrich-Alexander University of Erlangen-Nuremberg, National Chung Cheng University, Technical University of Munich, The New School, University of California Berkeley, University of Galway

〈Key words〉

artificial intelligence, big data, code of conduct, ethics, robot

Ⅱ 本編

1. プロジェクトの目標と背景

1-1. プロジェクトの達成目標

本プロジェクトは、人工知能（AI）と人・社会の調和という終極理念の下、次の 2 点を具体的達成目標とした。

① AI の倫理的課題の理論分析：AI・ビッグデータ分析の技術開発や企業の開発者との共同探索的会合を通じて、開発・利用上の倫理的課題を把握し、人間・AI の自律性および人間－AI の関係性という観点から構造化する。次に、AI 倫理学の国際学界での先端的研究成果を活用するとともに、日本の状況の特徴に留意しつつ、新たな規範理論を構築し、理論分析を実施する。これらの作業を通じて、AI の倫理的課題について精緻な規範的分析を行うことが、本プロジェクトの理論的目標である。

② 技術開発者の行動指針の策定：倫理的課題の理論分析の成果を基盤に、ステークホルダーとの連携の下、AI の開発・利用の現実やそれを取り巻く日本社会の特徴を踏まえつつ、開発者がとるべき行動や、開発・利用を行う企業が設けるべき制度に関して、行動指針（CoC）を策定し公表する。これは、責任ある研究・イノベーション（RRI）の普及・定着に向けた本プロジェクトの実践的目標となる。

カテゴリ	☑ ELSI への具体的な対応方策の創出 ¨ 共創の仕組みや方法論の開発 ¨ トランスサイエンス問題の事例分析とアーカイブに基づく将来への提言 ¨ その他（新型コロナウイルス感染症など新興感染症に関する研究開発）

1-2. 研究開発の対象・テーマ、その意義と背景

（1）研究開発の対象

人間には、日常生活から人生の目標にいたるまで自ら選択し実行するという自律的側面と、他者と親密性をもって様々な相互行為をなすという関係的側面がある。自らの判断の自律性を急速に高めつつ、多様な場面で人間と相互作用を行う AI は、人間の 2 つの側面をどのように変容させつつあるか。そして、この変容の行く先を見据えて、人間と社会の新たなあり方をいかに構想できるか。このような基本的問題関心に基づき、本提案は、人間・AI の自律性と人間－AI の関係性をテーマとした。

これらのテーマの下、AI 技術の進歩にしたがって問題が緩和されずむしろ深刻化してゆく技術相反型問題に焦点をあわせた上で、国際的な AI 倫理学の成果を活用するとともに、海外で未開拓な研究領域でも日本社会の特徴を踏まえた分析を行うことを意図して、具体的な研究対象を次の通り設定した。

①行動誘導型情報提供

AI を活用したターゲティング広告や生活習慣・交通経路等の助言は、自らの過去と大きく異なる意思決定を、本人が認識しえない仕方で困難にする点で、個人の自律性を脅かしうる。こうした自律性への新たな脅威をいかに抑制しうるかを探究する。

②自律型致死兵器

自律型 AI のうち自律型致死兵器は、相手国兵士への殺傷力を高め、戦争の意味を変質させるが、日本では倫理学的検討がようやく緒についた段階にあることを踏まえて、本格的研究を実施する。

③職場協働型 AI

オフィスや作業現場で労働者と共に労務を行う AI・ロボットが発達するほど、労働者の存在意義が低下しうることを踏まえて、労働者と AI・ロボットの関係性を考察する。

③介護ロボット

高齢者介護で活用されるロボットの発達につれて、高齢者と介護者の関係が変質してゆく可能性を見据えつつ、利用者と AI の関係性を分析する。

(2) 創出したアウトプット

本プロジェクトが創出したアウトプットは、理論的なものと実践的なものに大別される。理論的アウトプットは 3 点ある。①研究開発の代表者・実施者が個別的に、研究活動の中間成果物を発表した。具体的には、国内外の学術誌での論文公刊や、国際会議・国内学会大会等での個別報告などを行った。②国際水準を踏まえた理論分析を実施した成果として、海外の研究開発協力者を含む AI 倫理学の第一線の研究者が参集して、代表者・実施者とともに基調講演・研究報告を行う国際シンポジウムを開催した。その成果は、会議録または欧米の出版社から公刊する英語論文集として記録に残す。③倫理的課題の理論分析の最終成果物を邦語書籍として公刊する。

実践的アウトプットとしては、④開発者がとるべき行動や、開発・利用を行う企業が設けるべき制度に関する CoC を策定し、公表する。

(3) 本研究開発プログラムにおける位置付け

AI の発展・普及は、モノのインターネット（IoT）の進展と相まって、新たな経済成長を促進し、人間の福利を増大させる大きな潜在力をもつ。他方、個人の自律性・プライバシーが脅かされ、個人間の真正な関係・連帯が掘り崩されるリスクもある。それゆえ、AI から得られる便益を人類が享受しつつ、各種のリスクを予見し防止することを意図した規範的考察が、不可欠となる。こうした基本認識の下、本プロジェクトは、AI からのリスクの予防に向けた規範的理論分析を行い、またその分析に基づき CoC を策定した。これは、AI・ロボットと人や社会との調和を終極目標とした上で、自律性と関係性をめぐって生起しつつある諸問題を発見・予見し、RRI の普及・定着への貢献を企図するものである。このような研究開発の実施を通じて、本プロジェクトは、本研究開発プログラムに大きく貢献した。

(4) 本プロジェクトの意義・必要性

AI がもたらすリスクについて、わが国では法学者が精力的に研究を進めてきた。最大の論点は、自動運転車が交通事故を起こした場合の法的責任の所在である。この論点は自動運転車の導入期には極めて重要であるが、技術革新と普及に伴って事故数は急激に減少すると予測されているから、技術発展に伴って緩和されてゆくいわば技術親和型問題に属する。それに対して、本研究は、技術発展に応じて深刻化してゆくいわば技術相反型問題を検討することにより、AI が人間の自律性・関係性に与える永続的影響という問題圏を探査する。AI の影響に関する法学的研究の急速な進展と対照的に、日本での倫理学的研究はようやく緒についた段階にある。だが、倫理学的研究は、法制度設計の基本方針を提供しうるのみならず、企

業・開発者・利用者等のステークホルダーに行動の道徳的理由を提示するという重要な実践的機能をもつ。

以上の諸点を踏まえて、自律性・関係性をめぐる多様な技術相反型問題を研究対象とし、倫理学的理論分析を行った上で、その社会実装をめざす本プロジェクトには、大きな意義および必要性がある。

1-3. プロジェクトが考察する根源的問い

人間には自律的と関係的という両面がある。人類史を回顧すれば、狩猟社会では、小集団での日常的協力が生存に不可欠であり、農業社会に入ると、灌漑等のためには大規模な組織的協力が必要となったが、原則的に村落共同体での協力が中心だった。これらの時代には、個人の自律性はいまだ弱く、関係性が強かった。続く工業社会で、自律性が次第に強化され、情報社会にいたって、自律性はいっそう昂進した。だが、人間は本質的に他者との関わりや繋がりを求める存在であり、家族や地域社会の紐帯の弱化が懸念される今日、新たな関係性の模索が日本社会の随所で続いている。こうした人類史上の現段階にある日本社会において、AI は、個人の自律性と人間相互の関係性に対していかなる影響をもたらしうるか。そして、発見・予見される影響を踏まえて、どのような個人のあり方、人間関係、社会制度を構想し形成してゆくべきか。本プロジェクトは、〈我々がいかなる科学技術を手にしたいか〉にとどまらず、〈どのような人間でありたいか〉を根源的に考察する。

1-4. プロジェクトの先に見据える将来展望

本プロジェクトは、理論的アウトプットの国際的発信を行うものであり、その成果が国境を越え AI 倫理学等の研究者によって広く認知されると想定している。また、実践的アウトプットは、技術開発者などのステークホルダーによって活用されると考えている。特に、CoC は、技術開発者が、推論過程がブラックボックス化した自律的機械に関してもなお一定の倫理的責任を負うことを自覚し、それに基づき研究開発を進めるよう求める。これは、RRI を普及し定着させてゆく一里塚となる。その先に展望されるのは、個人の新たな自律性・関係性に基づき、責任ある研究者・技術開発者によって支えられた〈倫理的に強靭な社会〉である。

2. 研究開発の実施内容

2-1. 実施項目

■項目 1：AI・ビッグデータ分析の倫理的論点の把握
■項目 2：理論分析の枠組みの構築
■項目 3：理論分析の実施
■項目 4：分析結果の公表とその社会実装に向けた発信

2-2. 実施内容

■項目 1：AI・ビッグデータ分析の倫理的論点の把握

項目 1 は、開発グループを中心としつつ自律性グループ・関係性グループも参画して実施された。倫理的論点を把握する手法は、外部者との共同探索、市民意識調査、実施体制内部での技術開発に大別できる。その具体的な内容・方法・結果は、以下の 5 点に要約される。

（1）技術開発・システム開発の専門家や AI 関連諸領域の研究者との共同探索的会合

開発グループに属する工学系研究者である中西崇文・大越匡・伊藤昌毅は、民間企業での技術開発・システム開発の専門家との緊密な連携関係をそれぞれ保持することから、同グループを中心的媒介として、専門家との共同探索的会合を開催した。佐々木由樹氏（株式会社リンクアンドコミュニケーション）との会合の結果、行動誘導型情報提供システム（1-2.（1）[1]の研究対象①）をめぐる倫理的課題の一部を同定できた。また、松田雄馬氏（株式会社オンギガンツ代表取締役）と会合を通じて、人間の生物的情報処理システムに対する AI の含意について理解を深化させることができた。

民間企業の開発者との共同探索的会合に続き、AI 関連諸領域の多様な研究者を招いて、各領域の現状や先端的研究に関する情報提供を受けた上で、倫理的課題の共同探索を重ねていった。職場協働型 AI（1-2.（1）[1]の研究対象③）に関連して、山本陽大氏（労働政策研究・研修機構）が、ドイツにおける職場での AI 普及をめぐる最新の議論状況と政策動向を紹介した後、日本の労働市場で予見される AI 普及の影響とそれへの制度的対応について熟議した。また、AI の倫理的課題を同定し検討する際、一般市民がいかなる予期・意見・姿勢等をもっているかは、1 つの重要な考慮事由となりうることから、本プロジェクトでは意識調査の実施を計画していた（後出（2））。そのため、AI に関する意識調査の専門家から自らの最新研究を紹介してもらった上で、調査方法、結果の解釈、今後の研究課題、理論分析との関連づけなどについて討論を重ねた。専門家の人選は、関係性グループが中心となって行った。共同探索的会合では、一方井祐子氏（金沢大学）が、AI 利用に対する市民の倫理的態度の定量分析について、また馬奈木俊介氏（九州大学）は、リハビリテーション・ロボットに対する利用者の信頼の定量解析について、それぞれ報告を行った。さらに、人間－AI の関係性をめぐる倫理的課題に直截につながる論点として、AI による人間への働きかけのタイミングがある。そこで、早川正祐氏（東京大学）が、問いかけ・働きかけのタイミングが対話さらには共同的決定の促進にどのように関わるかを哲学的に論じ、その知見が人間－AI 関係にいかなる含意をもつかについて共同探索を試みた。

（2）市民意識の定量分析

AI の倫理的課題は、社会実践の外部に位置する外的視点から理論家によって一方的に設定されるべきでなく、実践の内部に定位した内的視点を十分に考慮に入れつつ学術的に同定される必要がある。すなわち、一般市民がいかなる予期・意見・姿勢等をもっているかは、倫理的課題の同定に際して 1 つの重要な考慮事由となる。こうした基本認識の下、関係性グループの荒井弘毅が、自律性グループの松元雅和とともに、自律型致死兵器（1-2.（1）[1]の研究対象②）に関する市民意識調査を実施した。荒井はまた、代表者兼自律性グループリーダーの宇佐美誠とともに、介護ロボット（研究対象④）の市民意識調査も行った。いずれもインターネット調査であり、パイロット調査の後に大規模な本調査を実施した。

（3）生活を取り巻く IT・AI サービスにおける自律性

開発グループの大越は、生活を取り巻く情報技術（IT）や AI のサービスにおける自律性に関する倫理的論点を把握するため、最も多くの人々に普及している情報機器としてのスマートフォンにおける、ソーシャル・ネットワーク・サービス（SNS）をはじめたサービスの依存的利用について研究を実施した。人々の生活に深く浸透し、単純に利用を禁止する方策が

現実には採用できない特殊な状況を設定した上で、特定のアプリケーションの依存的利用を抑制する手法として、「衝動発生プロセス」の「状況修正」戦略に着眼した、ユーザビリティ低下を伴って依存的利用の低減をめざす新手法を複数考案し、実証・評価を行った。その結果として、従来の手法と比較してより大きな効果が確認でき、この効果は、依存的利用の低減と自律的利用の回復・維持の実現に向けた役割を果たすと考えられる。

（4）交通における自律性

開発グループの伊藤は、交通領域における AI の活用が自律性に対して提起する倫理的論点を把握するべく、研究を実施した。交通には、個人の身体や個人が制御する装置によって自律的意思に基づく移動が実現されるという側面と、社会インフラとして多くの人が空間や設備を共有し一定のルールに従って移動するという側面がある。それゆえ、IT・AI を活用した研究開発においても、システムとしての交通体系全般をより効率的に実現しようという方向性と、個人の意思をより高度に実現しようという方向性とが併存して、研究が進んでいる。この構図は、本研究開発課題での自律性という一般的テーマが交通という具体的な文脈において社会実装されている事態として捉えることができる。だが、時間や空間を共有して移動する以上、個人の自律的意思だけを追い求めることは不可能であり、社会的な調停の仕組みが不可欠となる。それゆえ、AI には、個人の自律的な交通意志決定と、社会的・全体的な交通制御とを橋渡しする役割が期待されている。だが、交通制御は時に個人の自律的行動を制約することもある。そこで、どのように AI を埋め込むべきかを、実証実験とデータ解析を通じて探究した。

（5）AI の説明可能性を通じた人間－AI 間協働

人間と AI が真の意味で協働しうる環境を構築してゆくためには、AI の説明可能性の進展が不可欠である。そこで、開発グループリーダーの中西は、かねてより技術的側面の研究を進め、独自の説明可能 AI （XAI）手法である AIME を開発し、すでに特許も取得している。本プロジェクトでは、この AIME を用いて目的変数（AI の出力）と説明変数の貢献度を算出することにより、AI がなぜそのような結果を導いたのかを説明することが可能となった。これにより、AI のもつバイアスや学習の不備などを発見することができた。こうした作業を通じて、人間－AI 間協働を可能とする環境の構築に向けた論点が明確化した。

■項目 2：理論分析の枠組みの構築

項目 2 は、自律性グループ・関係性グループによって実施された。いずれのグループでも、現代社会における AI 全般の考察と特定場面に焦点をあわせた検討とを、同時並行的かつ相互参照的に進行させた。その内容・方法・結果は、次の 3 点に要約される。

（1）人間の自律の揺らぎと自律的 AI による殺傷への倫理的評価

自律性グループにおいては、宇佐美が、自律概念の哲学的分析の最新知見を収集することから始め、次に従来の自律の構造を再構成することにより、AI による行動誘導型情報提供が利用者の自律を揺るがせているという具体的場面を扱いうる枠組みを構築した。成原慧は、各 AI が他の AI やさまざまなモノ、さらには人とネットワークを通じて接続・連携する時代における自律の揺らぎを示すとともに、自律の再建を構想するための枠組みを形成した。松尾陽は、権力と自由の相克という文脈において自律を分析する枠組みを考察した。松元は、軍事部門における AI の利活用を検討するため、正戦論を中心とする戦争倫理学の知見に依拠することで、分析枠組みを構築していった。具体的には、区別原理（民間人保護原理）を前提としつつ、機械が人間を殺傷しうることの倫理的是非を、尊厳や責任の観点から検討する枠組みを確立した。

（2）環境構成物および代替的奉仕者としての AI と人間との関係性

関係性グループでは、AI が人間との関係でもつ 2 つの側面に着目しつつ、理論分析の枠組みを構築した。グループリーダーの稲葉振一郎は、AI が人間にとって個別的道具であるよりもむしろ環境の構成物となりつつあるという一般的趨勢に着目する枠組みを探究した。神崎宣次は、人間を取り巻く環境としての AI の持続可能性という概念に着目し、従来のように電力確保のみに視野を限定するのでなく、アラインメントを含む広い視野でこの概念を分析する枠組みを構成した。他方、介護を典型例とするいくつかの具体的場面では、AI 実装のロボットが、濃密な関係性が生成しやすいサービスの提供者にとって代わりうる。荒井は、こうした代替の場面において AI への信頼の構造を解明するための枠組みを定立した。

（3）市場分析および司法論への応用

以上 2 つの実施内容に加えて、自律性・関係性の応用的な枠組み構築も進めた。荒井は、自律性・関係性の理論分析の一環として、デジタル市場における競争環境の変化に着目し、競争法の事前規制と事後規制の経済的影響を比較する理論的枠組みを構築し、市場競争への影響を評価するための規制コストの概念を明確化した。他方、宇佐美は、諸外国で進展している司法での AI 利用を取り上げ、市民の自律性の毀損という観点からこの利用を峻拒する最新学説に注目した上で、AI による裁定と個人の自律性の関係を分析するための枠組みを構成した。

■項目 3：理論分析の実施

項目 3 は、項目 2 の結果を踏まえつつ、自律性グループ・関係性グループによって実施された。その内容・方法・結果は、項目 2 に対応して 3 点に要約される。

（1）自律の再建と自律的 AI の誤射責任

自律性グループでは、宇佐美が、AI の行動誘導型情報提供が利用者の自律を揺るがせているという事態について分析を実施した。これにより、従来の自律観を所与とすれば、AI による自律侵害として現状を評価することは困難である一方で、自律観を再編成するならば、自律空洞化という評価が可能となり、その対策の方向性を構想できるという結論を得た。成原は、AI 社会における個人の自律の揺らぎという一般的状況を踏まえて、自律を再建するためには、これを社会的文脈から切り離された個人の抽象的な権能や属性として捉えることは困難であると論じた上で、自律を他者との関係性や具体的な文脈に即して築き直すという見通しを示した。松尾陽は、権力と自由の相克という文脈で自律の分析を進めた。自律型致死兵器の倫理的是非については、松元が、兵器が引き起こしうる誤射に対して誰がどのように責任を問えるのかを焦点化して検討した。まず、行為主体に責任を問いうる条件として、制御条件・認識条件・非難条件を設定し、自律型致死兵器がこうした条件を充たす責任主体になりうるか、主体になれないとすれば代わりに製造者や使用者に責任が問えるかを理論的に精査した。また、自律型致死兵器の投入が人間の尊厳にとってどのような正負の影響をもたらすかを、尊厳の概念分析を交えつつ検討した。

（2）AI の安全保障への影響と AI への信頼

関係性グループでは、稲葉が、環境の構成物としての AI という基本認識の下、AI と在来型技術の比較や、AI が埋め込まれた社会制度と従来の社会制度との比較を実施した。神崎は、AI の持続可能性という枠組みに基づく理論分析の 1 つとして、人間の安全保障に対して AI が持ちうる影響を考察し、さらに AI 研究者が学会等を通じて果たしうる役割について検討した。荒井は、人間にとって代わる AI 実装のロボットに対する信頼の規定要因について分析を実施した。

（3）市場分析・司法論への応用

応用的な分析としては、荒井が、デジタル市場における競争環境の変化を分析するために、競争法の事前規制と事後規制の経済的影響について株価イベント分析を用い、各イベントが市場に与える影響を定量的に測定し、企業価値の変動を通じて規制の経済的コストを評価した。宇佐美は、精緻化された自律概念を活用して、司法での AI と市民の自律の関係について法哲学的分析を進めた。

■項目 4：分析結果の公表とその社会実装に向けた発信

項目 4 は、代表者・実施者全員によって行われた。

（1）理論分析結果の個別的公表

自律性グループにおいては、宇佐美が、自律性の観点から司法での AI 利用を理論的に分析した結果を、論文集所収論文 1 編および国際会議報告 2 件として公表した。そのなかで、生成 AI の活用とその長短にも論及している。また、これに密接に関連する法的シンギュラリティというテーマについて、国内学会大会報告および論文公刊を行っている。その他、自律性の観点から顔認証技術の倫理学的評価を行った国際会議報告 1 件や、国内学会大会における AI と法全般に関する総括討論 1 件および AI と司法に関する総括討論 1 件もある。松元は、自律型致死兵器の責任問題を分析した英語論文において、責任主体が不在となり、行為に対する責任が帰責できないときに責任ギャップが生じうることを論証し、ギャップを埋めるための責任の制度化という可能性を示唆した。また、尊厳に基づく自律型致死兵器反対論には、複数の相異なる尊厳概念への注目が見られ、そのいくつかは反対論として有望であるが、他のものは理論内在的な困難を抱えていることを明らかにした。松尾は、AI を用いた権力による監視と権力に対する監視や、個人のプライバシーについて、論文を公表している。成原は、個人の自律と AI の自律に関する英語論文を上梓した他、自律の揺らぎと再建に関する国内学会報告を実施した。また、アーキテクチャー・プライバシー・生成 AI 等について、多くの邦語論文を公刊するとともに口頭発表を行ってきた。

関係性グループでは、稲葉が、環境としての AI を理解する手掛かりを農業技術や土地制度をめぐる議論に求めた論考を公表した。生成 AI を生身の人間がどのように経験するかについて現象学的に考察した近刊論文もある。荒井は松元とともに、自律型致死兵器に関する市民意識を分析する英語論文を公刊した。また、宇佐美とともに、介護ロボットの市民意識を解明した英語論文を公表する予定である。さらに、荒井は、人工知能を含むデジタル市場の競争環境の最適化と規制の合理化に貢献するべく英語論文も公刊している。神崎は、AI と持続可能性に関するサーヴェイを人工知能学会誌上で公表した。

（2）理論分析結果の社会実装に向けた個別的発信

成原は、総務省「自治体における AI の利用に関するワーキンググループ」および宮崎県延岡市の生成 AI 検討委員会の委員を務めている。本研究プロジェクトでの分析結果を踏まえ、これらの会議において、AI の導入自体を自己目的化することなく、また従来は人間の職員が行ってきたことを単に AI に担わせるのではなく、AI を導入することを通じて業務のプロセス全体を見直し、職員と AI が協調・連携しつつ、住民のニーズに応えていくことが求められるとの提言を行った。また、総務省の情報通信法学研究会などで「生成 AI の時代におけるメディアの役割」について報告を行い、報道機関が AI 開発事業者に AI の学習に適した記事データベースを有償で提供し、AI 開発事業者が報道機関向けの生成 AI サービスを提供するなど、両者の連携を模索することにより、報道機関の経営基盤を強化するとともに、生成 AI の持続的で健全な発展にも貢献するエコシステムが構築されることが期待されるとの見通しを示した。

宇佐美は、司法過程での AI 利用に関する訪日学生団向け英語講演や、AI などの新技術を

取り巻く倫理的論点を解説する社会人向け講演を行ってきた。また、所属する日本学術会議「リスク社会と法」が公表を予定している意思の表出（報告）の作成に参画し、AI がもたらすリスクに関する部分を担当する。松尾は、『朝日新聞』上の連載論説文で、デジタル社会における権力と自由について論じている。さらに、神崎は、AI と持続可能性に関するサーヴェイを一般市民向けに調整した上で、新書として出版する予定である。これらは、理論分析の結果の社会的な普及・実装に向けた直接的または間接的な発信として位置づけられる。

（3）技術開発の個別的公表とその社会実装に向けた個別的発信

開発グループにおいては、社会実装に直結する多様な技術開発の個別的公表が行われてきた。中西は、AIME の応用面を強化するべく、多角的アプローチを通じて、説明可能性の向上や実用化のための技術展開を進めてきた。第 1 に、主成分分析（PCA）を活用して AIME の説明能力を拡張するための手法を提案した。既存の AIME では扱いきれない高次元データや複雑な変数間の相互作用に対して、次元分解と拡張を組み合わせることで、モデルの挙動をより直観的に解釈できる仕組みを示している。これにより、応用分野を拡大し、より複雑な課題へ AIME を適用可能とする基盤を築いた。第 2 に、特徴選択に焦点を当て、グローバルな特徴重要度を活用したシナジー効果のある前進・後退型の削除手法を提案した。AI モデルの性能と説明可能性を両立させる上で不要な特徴量を除去しながら、重要な特徴量を適切に残す必要があるなかで、AIME が算出する重要度を用いて効率的にモデルを洗練し、モデルの可搬性や推論速度の向上にも寄与する可能性を示した。第 3 に、PCaLDI という新たな枠組みを提案し、主成分負荷量を基盤とする特徴重要度を用いて類似度や距離計算を行う方法を開発した。従来のユークリッド距離やコサイン類似度などに比べて、特徴の相対的重要度を反映しながら似ているか異なるかを判断できるため、多様な場面で説明力と解釈性を高める仕組みを提供している。これは、AIME がめざす説明可能性をさらに他分野のメトリクス設計へ拡張するものであり、各種分析タスクの透明性向上に寄与する。これらの研究は、AIME の説明可能性を中心に据えながら、高次元データの扱いや特徴選択という多様な領域と連携することで、実用化への道筋を具体化している。こうした成果は、今後の社会実装において活用の幅を広げ、信頼性と透明性を両立する AI システムの普及に寄与するだろう。

大越は、生活を取り巻く IT・AI サービスにおける自律性に関する研究として考案した衝動発生プロセスの「状況修正」戦略にもとづく新手法を論文として公表した。この新手法の内容およびシステムは公開されており、社会実装を大きく促進するものである。また、伊藤は、AI 交通信号の開発やそこから得られた知見を招待講演等で公表するとともに、公共交通での情報技術のインパクトについて一般向けセミナーで論じた。このように、交通における AI のインパクトについて、一般市民向けや、交通関連の行政官・技術者等の専門家向けに情報発信を行っている。

（4）理論分析・技術開発の結果の集合的公表

本プロジェクトで行われた理論分析および技術開発の結果の集合的な公表方法は、2 つに大別される。第 1 に、代表者と実施者全員が寄稿して、研究成果の要点を平易に解説した一般市民向けの書籍を公刊する予定であり、2025 年 1 月時点で準備の最終段階にある。本書は、自律性と関係性の 2 部で構成されているが、2 つの視点を架橋する論考も含んでいる。自律性および関係性の観点から、AI を取り巻く国内外の最新状況を伝えるのにとどまらず、読者が、個人の自律性や個人相互の関係性に AI が与える衝撃について自ら予見し評価した上で、近未来社会を主体的に展望するように促すことを意図している。この出版企画は、理論分析の結果の公表であるとともに、その社会実装に向けての普及活動でもある。

第 2 に、代表者と一部の実施者、海外の協力者らが研究報告を行い、集中的に討論する国際会議「AI の開発・利用上の倫理的・法制度的・社会的課題に関する国際ワークショップ」を 2025 年 2 月に開催した。報告者は、スチュアート・ラッセル（カリフォルニア大学バークレー校）、クリストフ・リュトゲ（ミュンヘン工科大学）、ツゥイピン・リー（国立中正大学）、

ピーター・アサロ（ニュー・スクール）、ジョン・ダナハー（アイルランド・ゴールウェー国立大学）、荒井、松本、成原、宇佐美である。その記録を英語論文集として公刊することをめざしている。

（5）社会実装に向けた集合的発信

理論分析・技術開発の結果の社会実装に向けた集合的な発信として、CoC の策定と公表を行う。まず、種々の国際機関や各国の公的機関・民間団体が策定した CoC およびそれに類する文書について、包括的収集と比較検討を実施した。その結果、多くの CoC とその類似文書は、各企業で開発される AI 技術が備えるべき実体的側面に力点をおく一方、AI 技術を倫理的観点から評価するプロセスに関わる手続的側面を等閑視しがちであることが判明した。このような従来の策定状況を踏まえて、本プロジェクトでは、民間企業が、自社の開発による AI 技術やそれを活用したシステムを倫理的観点から評価する制度が備えるべき特徴を規定する、顕著な独創性を備えた CoC を策定する。
邦語・英語の両版で公表される本プロジェクトの CoC は、国内外でこれまで策定されてきた CoC への代替物でなく保管物となることを意図している。2025 年 1 月時点で完成に向けた最終段階にあり、同年春に公表する予定である。

2-3. 研究開発マネジメント

■プロジェクト・マネジメントの実施方針と工夫

研究開発の実施体制は、工学系研究者からなる開発グループと、人文社会科学系研究者の自律性グループ・関係性グループにより構成される。開発グループは、技術開発を通じて AI 開発・利用上の倫理的課題を同定し、自律性グループ・関係性グループに提示した。また、企業の技術開発・システム開発の専門家との共同探索的会合を通じて、全グループが倫理的課題を特定した。次に、自律性グループ・関係性グループが、倫理的課題の理論分析を行い、その結果を開発グループや、さらには技術者へとフィードバックした。このように、ステークホルダーの参画を得つつ課題の同定を行った上で、工学系研究者と人社系研究者の間で研究成果の螺旋的な共有・発展が起こるための工夫が、本プロジェクトの一大特徴である。

グループ間で実施者の専門分野が大きく異なることから、グループ間の有機的連携・協働をいかに促進するかが、研究開発期間の当初から重要な課題だった。この課題を達成するためのマネジメント上の工夫として、自律性グループリーダーの宇佐美と関係性グループリーダーの稲葉が、開発グループにも所属し、継続的に情報共有を行った。また、自律性グループと関係性グループの連携・協働を図るため、宇佐美と稲葉が緊密に連絡・調整を重ねた。さらに、グループ会合よりもむしろ全体会合をおもに開催することにより、研究成果の統合を強力に推進した。

■課題や障壁と、その対応方策

本プロジェクトの開始時点で想定された障壁は、AI に関する ELSI のうち法制度的課題や社会的課題から切断して、倫理的課題に特化して研究を進めてしまうリスクだった。そこで、法制度的課題との連結については、3 つの対応策を講じた。第 1 に、成原は情報法を専門とし、また法哲学を専攻する宇佐美や松尾も、AI に精通した法学者とのネットワークをもつから、これらの知識やネットワークを活用して、法制度的課題との連結を意識しつつ研究を進めた。第 2 に、アイルランドの法科大学院で教鞭をとり、法制度的課題にも通暁している研

究開発協力者ジョン・ダナハーから、研究助言を得た。第 3 に、RISTEX の研究開発プログラムでの法学系プロジェクトとの意見交換・情報交換を積極的に行った。他方、AI の社会的課題のなかでは、職場での AI・ロボットの普及が失業率や所得格差に与える経済的影響が分析されてきた。経済的影響に関しては、宇佐美がすでに邦語論文・英語論文を公刊し、実施者の稲葉・中西も考察しており、さらにダナハーが著書で詳細に分析していたため、共同討議を通じて倫理的課題と社会的課題の関連づけを行うよう努めた。

研究開発の開始時点では明確に想定されていなかった最大の新たな論点は、2022 年 11 月に OpenAI 社が ChatGPT を一般公開したことを機に、日本で急速に耳目を集めた生成 AI について、倫理的課題の特定や理論分析、その成果の社会実装をどのように進めてゆくかであった。生成 AI の萌芽は、つとに 1960 年代に開発されたイライザに見出され、後に 2014 年には、敵対的生成ネットワーク（GAN）が現れていた。だが、日本において大規模言語モデルの生成 AI が社会的関心を集めるにいたったのは、2022 年秋以降だと言ってよい。この新たな事態に対して、本プロジェクトがどのように対応するかが大きな課題となった。宇佐美・稲葉・成原は、相異なった領域における生成 AI の活用について検討や提案を行っていたため（2-2.の項目 4（1）・(2)）、情報交換・意見交換を行うと同時に、工学系研究者の実施者からの示唆も受けて、この新たな論点も視野に収めつつ理論分析を進めることができた。その成果は、公刊が間近な邦語書籍にも反映される予定である（1-2.（2））。

■プロジェクトを通じた ELSI/RRI 人材の育成

ELSI/RRI 人材の育成計画の起点となる基本認識は、人社系の院生・若手研究者にとって、ELSI/RRI のスキルの体得は決して容易ではないということである。体得を妨げる最大の障壁は、専門分化にある。自然科学よりも遅れて 19 世紀後半に各分野が成立した社会科学においては、その後急速に分化が進行した。こうした歴史的趨勢は今日の欧米諸国では大きく変容し、各分野でさらなる分化が進行する一方で、学際的な研究・教育が拡大し定着している。他方、わが国では、学際的な学部・研究科の増加から競争的外部資金での学際領域の新設にいたるまで、学問的越境・融合の努力が行われてきたにもかかわらず、人社系の各分野で中枢的位置を占める研究者の間では、従前の専門分化が堅持されているように見受けられる。しかるに、ELSI/RRI は、特定の科学技術を前提として初めて成立するから、これに関わるスキルを体得するためには、何よりもまず当該科学技術の基礎知識を習得しなければならない。また、倫理的課題・法的課題・社会的課題はしばしば複雑に関連しあっているため、その 1 つを深く理解するには他との関連性を正確に把握する必要がある。

別の障壁は、予見の必要性にある。AI を典型例とする科学技術の加速度的発展を所与として、それが提起する課題に的確に取り組むためには、未来の技術や社会を学術的根拠のある仕方で予想し、そこで生じるだろう課題を予見する必要がある。ところが、上述の厳格な専門分化の下、社会科学の各分野では、過去の事象の記述または因果的説明が中心となり、予想や予見は 2 次的なものと位置づけられてきた。また、人文学のうち特に法哲学・政治哲学・倫理学を見るならば、日本では従来、過去から現在までの知的遺産の解釈に重きが置かれ、欧米における歴史上の思想や当代の理論の紹介・検討に多大な精力が傾注されてきた。近年には、種々の今日的問題への関心が、欧米の理論家の目を通す形で日本でも生じているものの、来たる未来を予想し、新たな課題を予見する研究は、いまだごく僅かにとどまる。

これらの障壁を克服する最初の一歩は、宇佐美が指導生に長年説いてきた〈基軸ある学際性〉にあると考える。基軸ある学際性は、まず自らの分野を選び、専門的に深く学んだ後、研究テーマにあわせて 隣接する諸分野さらには距離ある諸分野を探訪し、知識を拡張しつつ相互に関連づけてゆく。このような形態の学際性こそが、ELSI/RRI の基盤となる科学技術の理解や予見に基づく考察を可能とするだろう。

その上で、ELSI/RRI を扱いうる研究者へと陶冶されるためには、少なくとも 3 種類の能

力が必要だと思われる。第 1 は〈見抜く力〉である。経済学者ケネス・アロウがかつて述べたように、専門技術的論点と深遠な論点つまり哲学的論点は別々に存在するのでなく、連関しさらには融合している。それゆえ、技術的問題のなかに潜む哲学的・倫理学的問題を看取する能力が求められる。第 2 は〈繋がる力〉である。独力では理解しがたい他分野での専門的知識を獲得するためには、当該分野の専門家との連携・協働が不可欠となる。無論、連携・協働が相手にとっても有益であるためには、相手に対して提供できる自らの分野での専門知を習得していることが前提となる。第 3 は〈変わる力〉である。科学技術が日進月歩で発展し、そこから生じる課題も急速に変化し多様化してゆく以上、研究生活の当初の分野・主題・方法に安住するのではなく、常に新たな分野・主題・方法を貪欲に学習して、自らが発展的に変容してゆかねばならない。

これらの能力を研究人材に獲得させるという宇佐美の育成計画の下、本プロジェクトで雇用された服部久美恵研究員は、自律性グループと関係性グループの双方の活動に従事して、AI の開発・利用上の技術的論点に連関した倫理的課題を探査するとともに、多様な分野の研究者との討議にも参加した。また、企業等の技術開発者・システム開発者との共同探索的会合にも同席した。服部研究員は、2021 年 12 月から 2024 年 3 月まで京都大学に勤務した。

3. 研究開発の成果

3-1. プロジェクト目標における研究開発成果

（1）諸成果の特徴と目標の達成状況

前述の通り、本プロジェクトで代表者・実施者によって個別的に発表された諸成果は、理論分析／技術開発、成果の公表／実装化に向けた発信という 2 次元で大別できる（2-2.の項目 4（1）～（3））。自律性グループ・関係性グループが実施した理論分析の成果の多くは、国際的な研究動向に棹さし、日本社会という文脈を超えた普遍性を備えている一方で、市民意識調査のように日本的文脈を反映した知見も含んでいる。技術開発の成果も、AI の説明可能性の向上のような普遍的結果と、特定の都市における AI 交通信号の試用結果などの特殊的結果を含んでいる。同様に、実装化に向けて個別に行われた発信もまた、普遍性を備えたものと日本社会に定位したものを含んでいる。代表者・実施者の全員が集合的に行う成果の発表は（(4)）、AI をめぐる最新状況の知識を一般市民に提供し、近未来社会を展望するよう促す邦語書籍の出版と、海外の第一級の研究者たちとともに研究報告・討論を行う国際会議の開催およびその成果の公刊からなる。ここにもまた、日本社会の特徴への着目と普遍的論点の探究の両方が含まれている。このように、本プロジェクトの諸成果の一特徴は、普遍的レベルと特殊日本的レベルをともに含む重層性にある。

本プロジェクトの理論的目標は、①企業の開発者との共同探索的会合などを通じて、開発・利用上の倫理的課題を把握し、人間・AI の自律性および人間－AI の関係性という観点から構造化した後、②AI 倫理学の国際学界での先端的研究成果を活用するとともに、日本の状況の特徴に留意しつつ、新たな規範理論を構築し、理論分析を実施することを通じて、AI の倫理的課題に関する精緻な規範的分析を行うことだった（1-1.）。この目標は、実施内容の項目 1～3 によって（2-2.）、十全に達成された。

（2）アウトプットとインパクト

本プロジェクトのアウトプットは、実施内容の項目 4 に記載の通り、（1）理論分析結果の個別的公表、（2）理論分析結果の社会実装に向けた個別的発信、（3）技術開発の個別的公表とその社会実装に向けた個別的発信、（4）理論分析・技術開発の結果の集合的公表、（5）社

会実装に向けた集合的発信に類別できる（2-2.）。各アウトプットの学術的・社会的インパクトは、以下のように要約できる。（1）の学術的インパクトは、国際会議・国内学会大会における代表者・実施者の報告等に対するフロアからの活発かつ建設的な反応によって例証されている。また、特筆するべき学術的評価として、荒井のデジタル市場に関する英語論文が、2025 Antitrust Writing Awards の Academic Articles における Digital 部門にノミネートされたことを挙げたい。（2）に関しては、成原が、国・自治体の両レベルにおいて政策過程への重要なインプットを行っている。また、宇佐美は、日本学術会議の所属分科会による意思の表出への参画を通じて、社会実装に向けた発信を行う予定である。（3）については、中西・大越・伊藤がそれぞれ AI 技術の社会実装に向けて重要な貢献をなし、成果の周知・応用が進みつつある。（4）・（5）のインパクトは今後の確認を待つ他ないが、特に（5）として、民間企業における自社の AI 技術の倫理的評価制度が備えるべき手続的特徴を規定する独創的な CoC を、日英両語で公表するので、国内外で大きなインパクトが生じることを期待している。

今後期待できる社会実装上のインパクトとしては、（3）の一部をなす AIME を特記しておきたい。AIME は、AI が出力した結果を説明する際、目的変数（AI の推定結果や判断）に対する説明変数の貢献度を可視化し、どの要因がどの程度影響を及ぼしているのかを明らかにするメカニズムをもつ。このメカニズムによって、ユーザーは AI の意思決定プロセスを理解しやすくなり、必要に応じて偏りや学習不備の検証も行えるため、AI の出力を単に鵜呑みにするのではなく、結果の真偽や妥当性を自律的に判断・検証することが可能となる。こうした説明可能性をもつ AI を社会実装してゆく意義は大きい。医療・公共交通・教育など、人命や社会基盤に関わる分野では、AI と人間との協働が不可欠であり、その協働が安全かつ適正に行われるためには、AI の判断根拠を開示する仕組みが必要となる。AIME のような XAI 技術を活用することで、AI の判断を人間が検証しながら意思決定に組み込むプロセスを確立し、社会的合意形成や責任所在の明確化などを可能にできる。これにより、AI を導入する業務やサービスの信頼性を高め、多様な場面で AI 技術を活用していくことが期待される。したがって、AIME をはじめとする XAI 技術の普及と発展は、大きな社会的インパクトをもつだろう。今後は、研究段階で得られた知見を基に、社会実装に向けた具体的な工程表を策定して、実際に運用する場面での課題を同定し解決してゆく必要がある。こうした取り組みによって、人間と AI の協働をより高次元で実現し、社会の様々な領域における価値創出と安全・安心の確保を同時に実現することが可能となるだろう。

（3）研究開発プロセスからの知見

本プロジェクトの実施体制がもつ 1 つの特徴は、顕著な学際性にある。代表者・実施者の専門分野は、人社系では法哲学（宇佐美・松尾）・政治哲学（松元）・倫理学（神崎）・社会倫理学（稲葉）・経済学（荒井）など、また工学系ではデータサイエンス（中西）・情報通信（大越）・交通工学（伊藤）など、多岐にわたっている。こうした極めて学際的な布陣の下で、各人が専門知や研究実績を持ち寄り、相異なった視点から分野横断的討議を重ね、そして専門融合的な共同研究を実施してきた。その結果、研究開発期間の開始時点で予定していた研究対象について分析の著しい深化を実現できたのみならず、当初は予期していなかった新たな論点も視野に入れつつ研究を進めることができた。例えば、自律的 AI の人間に対する作用で厳しく問われる人間の尊厳、個人の関係性のみならず自律性を扱う際にも重要となる消費者の感情、AI 技術の生産者を理解する際に避けられない企業の行動原理などについて、認識または再認識をすることができた。また、AI 開発を評価する上で、アラインメントという倫理的要請や、説明可能性という技術的要請にも焦点を当てることができた。さらに、研究開発期間の途中でにわかに社会的注目を集めるにいたった生成 AI について、迅速な理論的対応をなしえたのも（2-3.）、技術開発の最前線にいる工学系研究者と AI 実装状況に通暁する人社系研究者が集まっている本実施体制のゆえだと考えられる。要言すれば、学際的協働が実り豊かな学術的創発を生みだし、その社会実装に向けた提案をもたらすという点が、研究開発

プロセスから得られた知見である。

3-2. プログラム目標における成果・示唆

（1）プログラム目標の達成への貢献

本プログラムの目標は、新興科学技術の ELSI を発見・予見しながら、科学技術と人・社会が調和する RRI の営みの普及・定着に資する実践的協業モデルを創出することだとされている。本プロジェクトでは、行動誘導型情報提供・自律型致死兵器・職場協働型 AI・介護ロボットという具体的研究対象にそくして（1-2.（1）[1]）、企業の技術者や AI 関連諸分野の研究者との共同探索的会合および実施者による技術開発を通じて個々の倫理的課題を発見・予見した（2-2.の項目 1）。その上で、課題の理論分析を実施し（項目 2・3）、その結果の社会実装に向けて、RRI の普及・定着に貢献するために、民間企業向けの CoC を公表する（項目 4（5））。これらを通じて、本プロジェクトは本プログラムの目標に対して重要な貢献を行った。

（2）根源的問いの探究結果

本プロジェクトは、人間が自律的側面と関係的側面をあわせもつという人間観の下、関係性が支配的だった狩猟社会・農業社会から、自律性が次第に強まる工業社会を経て、情報社会にいたった日本社会において、新たな関係性の模索が続いているという基本認識に立脚している。その上で、AI が個人の自律性と人間相互の関係性に対していかなる影響をもたらしうるか、また発見・予見される影響を踏まえて、どのような個人のあり方、人間関係、社会制度を構想し形成してゆくべきかを探究した。こうした探究のなかで繰り返し現れたのは、〈どのような人間でありたいか〉という根源的問いである（1-3.）。

上記の根源的問いに対する本プロジェクトの暫定的解答は、人間が備える自律性および関係性を再構成することによって、AI を不可欠の要素とするエコシステムにおける人間の繁栄（human flourishing）を追求することだと要約できる。こうした自律性・関係性の再構成に向けて、本プロジェクトの理論分析の成果は資するところが大きい。また、技術開発の実装や CoC の策定・公表は、理論分析と緊密に連結した社会実践の発展に役立つと期待できる。

（3）日本の文脈の考慮

欧米社会と対比しつつ日本社会を特徴づける人口に膾炙した言説には、文化的宿命論とでも呼ぶべき陥穽がある。文化的宿命論は、ヨーロッパ諸国と北米を一括した上で、それとは極めて異質である独特の伝統的文化をもつものとして日本社会を位置づけた後、日本の後進性を論難するか、あるいは反対にその伝統的文化を称揚する。ここには、①欧米諸国の間に存在する多様性の看過という他者認識的誤謬、②明治の文明開化に端を発し第 2 次世界大戦後にいっそう強まった西洋に範を求める日本社会の顕著な傾向の忘却という自己理解的誤謬、③日本社会を含むあらゆる社会が時代により変容してゆく動態性の否定という静態主義的誤謬が見出される。

本プロジェクトでは、文化的宿命論に陥ることなく、日本社会に伏在する普遍性と特殊性、そして両者がもつ動態性を冷静に観察し考察するように努めた。その結果、例えば、自律型致死兵器・介護ロボットに関する意識調査の結果について、社会の如何を問わず観察される普遍的傾向と、欧米での調査結果との比較を通じて浮き彫りになる日本での結果の偏差とを正確に把握することができた。また、理論分析・技術開発の成果に含まれる普遍的含意を自覚できた。

グローバルなものとローカルなものとの乖離・衝突は、民間企業の内部でも生じうる。経

済のグローバル化が進展する今日、多国籍企業は元より一国内の企業も、輸出先の法的規制を守り、サプライチェーンが位置する国々での規制や社会規範に服することを要求される。他方、企業が本社・事業所を置く社会の文化的環境を無視して経営を行うことは不可能である。このようなグローバルとローカルの乖離・衝突の可能性を考慮に入れ、本プロジェクトで公表する CoC では、両者のハーモナイゼーションを図る予定である（2-2.の項目 4（5））。

3-3. 成果の活用・展開状況や今後の展望

本プロジェクト全体の成果やアウトプットのなかでも、特に CoC は、広範な社会的活用が期待される。CoC を国内外に広く周知するため、邦語版・英語版をともに策定した上で、ユーザー数が 1 億人を超えるソフトウェア開発のプラットフォーム GitHub 上で公開する。また、恒常的に閲覧可能となるよう、宇佐美が勤務先で主催する研究室の公式サイトに掲載する（https://www.envpolicy.ges.kyoto-u.ac.jp/）。さらに、本プロジェクトに協力していただいた企業の開発者と共有するのにとどまらず、各実施者による他方面への拡散も計画している。

3-4. 成果の他分野への適応可能性・知見の一般化可能性

（1）スマートフォンの依存的利用に関する知見

本プロジェクトの成果・知見のなかでも、開発グループによる技術開発の成果は、他分野への適用可能性や一般化可能性が顕著である。第 1 に、人々に広く普及しているコンピューティングデバイスは、本プロジェクトで大越が研究対象としたスマートフォンに限定されない。スマートウォッチ、ヘッドマウントディスプレイなどの新しいモバイルやウェアラブルデバイスなどが登場している。これらはいずれも、普及とともに多くのアプリケーション・サービスを伴って、日常生活から切り離せないデバイスへと変化してゆく。こういった端末上での依存的利用についても、スマートフォンに関する研究結果は適用可能である。第 2 に、大越は SNS での依存的利用を解明したが、その知見は、SNS 以外の多くのアプリケーション・サービスに対しても、人間と IT・AI サービスの対話における依存的利用の低減という意味で、一般化でき適用できる。今日、多種多様なアプリケーション・サービスが次々に登場しているから、本研究が適用可能である範囲の検証は、1 つの重要な研究課題となるだろう。

（2）説明可能 AI

中西が開発した AIME は、多様な分野で適用・応用され、AI と人間の協働を円滑かつ安全に進める不可欠なツールになってゆくと予想される。AIME の一大特徴は、連続か非連続かを問わず、また線形か非線形かという関数形に依存せず、あらゆる AI モデルへ適用可能だという汎用性にある。それにとどまらず、データ分布が偏っている場合や、データセットが多様な特徴量を含む場合にも、アルゴリズム上の柔軟性を活かして説明を導出できる。つまり、AI のモデル構造やデータの分布特性にとらわれずに説明可能性を実現するという汎用性を備えている。それゆえ、まず医療・金融など、高い信頼性が求められる領域では、AIME の活用により、専門家は判断プロセスを追跡しながら、AI モデルのバイアスを検証し、精度向上やリスク軽減策の検討を行うことができる。また、教育現場や自治体業務でも、導入コストや運用上の負荷を考慮しつつ、システムの透明性やユーザーに対する説明責任を確保できる。さらに、マーケティングや交通インフラなど、異なる分野で異なるニーズがある場面でも、有効に機能すると期待される。今後は、社会実装の実績を積み重ねることが重要となる。複数の産業セクターや自治体との連携を通じて AIME を試験導入し、ユーザーからのフィー

ドバックを反映しながら改良を重ねることが求められる。その過程で、用途に応じて柔軟にカスタマイズできるインタフェースや、より大規模なデータに対して効率的に説明を生成できるアルゴリズムを整備していく必要がある。また、社会的・倫理的側面への配慮として、利用者が AI の判断根拠を十分理解し、責任の所在が曖昧になるのを防ぐため、ガイドラインや運用指針の策定を技術開発と併行して進めることが望ましい。こうした取り組みを通じて、AIME への信頼性が高まり、人間の専門知識・経験と AI の高速処理・分析能力を組み合わせることで、より高度な問題解決や新たな価値創造につながる可能性が開かれる。この研究成果を基盤として、社会の多様な課題に対して最適な解決策を探求し、AI 技術を効果的かつ安全に活用できる仕組みを整えることが、今後は重要である。

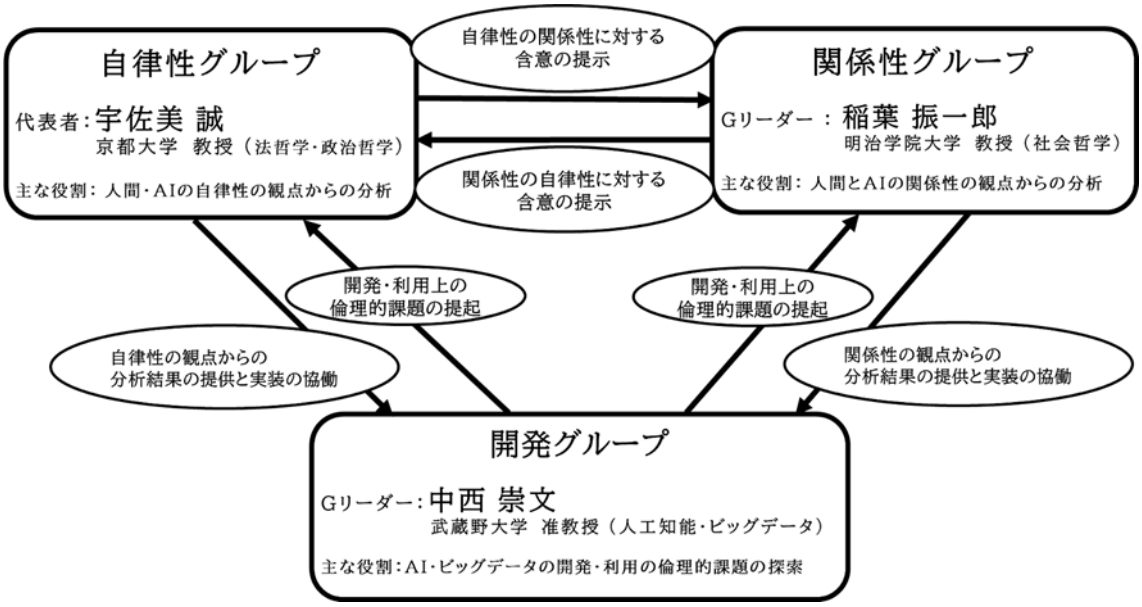
（3）AI の持続可能性の理論分析

理論分析の成果のなかでは、神崎による AI と持続可能性に関する分析が、ウェルビーイングやスマートシティなどと密接な関連をもつ。そのため、これらのテーマに取り組む他のプロジェクトや地方自治体によって活用されると期待できる。

4. 研究開発実施体制

4-1. 研究開発実施体制図

自律性グループは、人間・AIの自律性をテーマとし、行動誘導型情報提供と自律型致死兵器を研究対象とした。関係性グループは、人間－AIの関係性をテーマとし、職場協働型AIと介護ロボットを対象とした。開発グループは、技術開発を通じてAI開発・利用上の倫理的課題を同定し、自律性グループ・関係性グループに提示した。また、企業の技術開発・システム開発の専門家との共同探索的会合を通じて、全グループが倫理的課題を特定した。続いて、自律性グループ・関係性グループが、倫理的課題の理論分析を行い、その結果を開発グループや、さらには技術者へとフィードバックした。このように螺旋的に発展する学際的協働が、本研究開発実施体制の大きな特徴である。



協力者

スチュアート・ラッセル（AIの研究開発とリスク対処法に関する助言）
ヴィンセント・C・ミューラー（AI倫理学の最先端の情報提供）
クリストフ・リュトゲ（AI倫理の市場での実装に関する助言）
ピーター・アサロ（自律的兵器の倫理学的研究に関する研究協力）
ジョン・ダナハー（AIの労働者への倫理的影響に関する研究協力）
鹿島　久嗣（統計的機械学習の開発・応用で生じる倫理的問題に関する情報提供）
久木田　水生（社会的ロボットとの相互作用の倫理学的研究に関する研究協力）

4-2. 研究開発実施者

（1）自律性グループ（リーダー氏名：宇佐美 誠）
役割：人間・AI の自律性の観点からの分析

氏名	フリガナ	所属機関	所属部署	役職（身分）
宇佐美 誠	ウサミ マコト	京都大学	大学院地球環境学堂	教授
松元 雅和	マツモト マサカズ	日本大学	法学部	教授
松尾 陽	マツオ ヨウ	東海国立大学機構	名古屋大学大学院法学研究科	教授
成原 慧	ナリハラ サトシ	九州大学	大学院法学研究院	准教授
服部 久美恵	ハットリ クミエ	京都大学	大学院地球環境学堂	研究員

（2）関係性グループ（リーダー氏名：稲葉 振一郎）
役割：人間と AI の関係性の観点からの分析

氏名	フリガナ	所属機関	所属部署	役職（身分）
稲葉 振一郎	イナバ シンイチロウ	明治学院大学	社会学部	教授
荒井 弘毅	アライ コウキ	共立女子学園	ビジネス学部	教授
神崎 宣次	カンザキ ノブツグ	南山大学	国際教養学部	教授
服部 久美恵	ハットリ クミエ	京都大学	大学院地球環境学堂	研究員

（3）開発グループ（リーダー氏名：中西 崇文）

役割：開発・利用上の倫理的課題の提起

氏名	フリガナ	所属機関	所属部署	役職（身分）
中西 崇文	ナカニシ タカフミ	武蔵野大学	データサイエンス学部	准教授
大越 匡	オオコシ タダシ	慶應義塾大学	環境情報学部	准教授
伊藤 昌毅	イトウ マサキ	東京大学	大学院情報理工学系研究科	准教授
宇佐美 誠	ウサミ マコト	京都大学	大学院地球環境学堂	教授
稲葉 振一郎	イナバ シンイチロウ	明治学院大学	社会学部	教授

4-3. 研究開発の協力者

氏　名	フリガナ	所　属	役職（身分）	協力内容
スチュアート・ラッセル	スチュアート・ラッセル	カリフォルニア大学バークレー校	教授、マイケル・H・スミス＝ロトフィ・A・ザデ工学主任	AI の研究開発とリスク対処法に関する助言
クリストフ・リュトゲ	クリストフ　リュトゲ	ミュンヘン工科大学	教授、AI 倫理学研究所長	AI 倫理の市場での実装に関する助言
ヴィンセント・C・ミューラー	ヴィンセント C ミューラー	フリードリヒ・アレクサンダー大学エアランゲン＝ニュルンベルク	Av・フンボルト教授、哲学・AI 研究所長	AI 倫理学の最先端の情報提供
ツゥイピン・リー	ツゥイピン・リー	国立中正大学	教授	公務員の AI への信頼に関する研究協力
ピーター・アサロ	ピーター　アサロ	ニュー・スクール	メディア研究准教授	自律型致死兵器に関する研究協力
ジョン・ダナハー	ジョン　ダナハー	アイルランド・ゴールウェー国立大学	准教授	AI 倫理学に関する研究協力
鹿島 久嗣	カシマ ヒサシ	京都大学	教授	統計的機械学習の開発・応用で生じる倫理的問題に関する情報提供
久木田 水生	クキタ ミナオ	名古屋大学	准教授	社会的ロボットとの相互作用の倫理学的研究に関する研究協力
馬奈木 俊介	マナギ シュンスケ	九州大学	教授	リハビリテーション・

				ロボットに対する利用者の信頼に関する情報提供
一方井 祐子	イッカタイ ユウコ	金沢大学	准教授	AI 利用に対する市民の倫理的態度に関する情報提供
山本 陽大	ヤマモト ヨウタ	労働政策研究・研修機構	主任研究員	ドイツの職場での AI 普及と労働政策に関する情報提供
松田 雄馬	マツダ ユウマ	株式会社オンギガンツ	代表取締役	人間・AI の生物的情報処理システムに関する情報提供
佐々木 由樹	ササキ ユキ	株式会社リンクアンドコミュニケーション	最高公衆衛生責任者・事業開発マネージャー	AI 健康アプリに関する情報提供

5. 研究開発成果の発表・発信状況、アウトリーチ活動など

5-1. 社会に向けた情報発信状況、アウトリーチ活動など

5-1-1. 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）

年月日	名　称	場　所	概要・反響など	参加人数
2023 年 3 月 12 日	令和 4 年度 JST - RISTEX プロジェクト間連携シンポジウム「気候変動をめぐる市民参加・情報技術・倫理」	滋賀県立県民交流センター ピアザ淡海	政策のための科学プログラムの馬場 PJ と RInCA プログラムの宇佐美 PJ は、ビッグデータ分析の技術開発を実装しようとする側と、技術開発・利活用に係わる倫理的課題を探る側という相補的な関係をもつ。こうした相補性に基づき、両 PJ の構成メンバーが中間的研究成果を発表し共有する公開シンポジウムを開催した。その結果、今後の両 PJ における技術開発や社会実装の課題や指針について重要な示唆を	

			得た。	
2024 年 10 月 9 日	Beyond 2050 プロローグ第 6 節 ～京都大学が描く未来の社会像Ⅱ	京都大学吉田キャンパス 国際科学イノベーション棟	2050 年よりも先の社会を展望しつつ、気候変動・災害・生命倫理など、自然と人為の間の多様な課題について話題提供と意見交換を行う公開シンポジウムを開催した。宇佐美 PJ の宇佐美が講演者の 1 人となり、出口 PJ の出口康夫がファシリテーターを務めた。	

5-1-2. 書籍、DVD など論文以外に発行したもの

(1) Usami, Makoto, “Digitization, Unemployment, and Distributive Justice,” in Christoph Lütge, Matthias Uhl, Alexander Kriebitz, and Raphael Max (eds.), *Business Ethics and Digitization*, Springer, September 2022, pp. 65-87.

(2) 宇佐美誠「AI裁判官：技術・機能・原理」田中成明・足立幸男編『政治における法と政策：公共政策学と法哲学の対話に向けて』勁草書房、2023年9月、181-206頁。

(3) Usami, Makoto, “Responding to Existential Risks: Grounds, Targets, and Strategies,” in Yukio Adachi and Makoto Usami (eds.), *Governance for a Sustainable Future: The State of the Art in Japan*, Springer, September 2023, pp. 71-91.

(4) 稲葉振一郎「アンドロイドは水耕農場の夢を見るか？」梶谷懐・岸政彦編『所有とは何か：ヒト・社会・資本主義の根源』中央公論新社、2023年6月。

(5) 成原慧「「AIによる差別」にいかに向き合うか」宮下萌編『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』解放出版社、2022年2月。

(6) 成原慧「プラットフォーム事業者によるアーキテクチャのデザインとモデレーションに伴う民事責任」千葉恵美子編『デジタル化社会の進展と法のデザイン 』商事法務、2023年10月。

(7) 成原慧「フェイク・ニュースを規制すべきか？」瀧川裕英編『もっと問いかける法哲学』法律文化社、2024年5月。

(8) 松元雅和「平和・安全保障」山岡龍一・大澤津編著『現実と向き合う政治理論』放送大学教育振興会、2022年、124-138頁。

(9) 松元雅和「パトリオティックな配慮とその限界：国際社会における消極的／積極的義務」有賀誠・田上孝一・松元雅和編『普遍主義の可能性／不可能性：分断の時代をサバイブするために』法政大学出版局、2024年2月、85-109頁。

(10) 松元雅和「自由の制限はいつどのように許容されうるか：チルドレス諸原則の分析」岩崎正洋編『コロナ化した世界：COVID-19は政治を変えたのか』勁草書房、2024年3月、19-37頁。

5-1-3. ウェブメディア開設・運営

該当なし

5-1-4. 学会以外のシンポジウムなどでの招へい講演　など

（1） Usami, Makoto, “AI Judges: Potentials and Pitfalls,” lecture for the students of Northwestern Polytechnical University, Kyoto University, 12 July 2023.

（2） Usami, Makoto, “Do Algorithms Dream of Robes? A Moral Appraisal of AI Judges,” lecture for the Students of Beijing University of Posts and Telecommunications, Kyoto University, 3 August 2023.

（3） Usami, Makoto, “Existential Risk: Relevance of Justice,” lecture for the students of the Chinese University of Hong Kong, Kyoto University, Kyoto, 3 June 2024.

（4） 神崎宣次「AIと人権について考えてみませんか」令和4年度天白生涯学習センター公開講座 人権問題特別講演会、名古屋市、2023年3月4日。

（5） Kanzaki, Nobutsugu, “Autonomous Vehicle Ethics as City Ethics: On What Shared Value We Should Base Our Discussion,” 23rd Biennial Conference of the Society for Philosophy and Technology (SPT 2023), National Olympics Memorial Youth Center, 8 June 2023.

（6） 神崎宣次「人工知能と倫理の関係: アラインメントとサステイナビリティ」南山学会シンポジウム、南山大学、2023年10月25日。

（7） 神崎宣次「生成AI時代の「人間」と「社会」のゆくえを問う」豊田工業大学次世代文明センターシンポジウム「生成AI時代の「人間」と「社会」のゆくえを問う」、豊田工業大学、2024年3月6日。

（8） 神崎宣次 「アライメントと尊厳: 「AIが」尊厳を侵害するかもしれない状況はあるか」ワークショップ「AI・ロボット社会的受容と尊厳」、慶応義塾大学、2024年3月20日。

5-2. 論文発表

5-2-1. 査読付き（ 31 件）

（1） Arai, K. and Y. Hayakawa, “Three E-commerce Case Studies in the Context of Japanese Competition Enforcement: Comparative Considerations with the US Experience.” *World Competition* 46(1): 99-120, 2023.

（2） Arai, Koki and Masakazu Matsumoto, “Public Perceptions of Autonomous Lethal Weapons Systems,” *AI and Ethics*, online first, 2023. https://link.springer.com/article/10.1007/s43681-023-00282-9

（3） Arai, Koki, “Law and Economics of Language Model Development: Empirical Examination of Corporate Strategies and Vaporware Claims,” *Asian Journal of Law and Economics*, 2024. doi.org/10.1515/ajle-2023-0118

（4） Arai, Koki, “Comparative Impacts of Ex-Ante and Ex-Post Regulation in Digital Markets: An Event Study Analysis of Stock Price Reactions,” *Asian Journal of Law and Economics*, online first, 2024. https://doi.org/10.1515/ajle-2024-2002

（5） 吉添衛・服部宏充・江間有沙・大澤博隆・神崎宣次・久木田水生・小川祐樹「グループディスカッションにおける可視化情報提示に基づく気づき支援」『ヒューマンインタフェース学会論文誌』23 巻 4 号 501-512 頁、2021 年。 https://doi.org/10.11184/his.23.4_501

（6） 大谷卓史・大澤博隆・壁谷彰慶・川口嘉奈子・川口由起子・神崎宣次・久木田水生・杉

本俊介「良い倫理的意思決定のための倫理綱領―研究・イノベーションと倫理―」『電子情報通信学会誌』106 巻 3 号 1-6 頁、2023 年。

(7) Matsumoto, Masakazu, "The Right and Justice of Subsistence Wars as Necessity: A Grotian Account," *International Politics*, online first, 2023. https://link.springer.com/article/10.1057/s41311-023-00540-x

(8) Matsumoto, Masakazu and Koki Arai, "What's Wrong with 'Death by Algorithm'? Classifying Dignity-based Objections to LAWS," *AI and Society*, online first, 2024.

(9) Matsumoto, Masakazu, "A Threshold Account of Last Resort in the Ethics of War," *International Relations*, online first, 2024.

(10) 松尾陽「アーキテクチャ論から新型コロナ禍の対応を考える」『法哲学年報 2020』80-92 頁、2021 年。

(11) 松尾陽「コロナ禍における多様な規制手法とその制御のあり方：ポスト／ウィズ・コロナの法哲学」『学術の動向』2022 年 3 月号 40-46 頁。

(12) 松尾陽「プライバシーの論じ方：「放っておいてもらう権利」の登場背景とその受容の苦難」『α-SYNODOS』 304 号、オンライン、2022 年。

(13) 松尾陽「監視と自由：権力による監視と権力に対する監視」『法律時報』95 巻 1 号 78-82 頁、2023 年。

(14) Ikegami, A. and T. Nakanishi, "Interpretable Predictive Results in Classification of Waka Poets," *Proceedings of the 12th IIAI International Congress on Advanced Applied Informatics* 2022: 436-442, 2022. https://doi.org/10.1109/IIAIAAI55812.2022.00092

(15) Nitta, T., S. Hagimoto, A. Yanase, R. Okada, V. Sornlertlamvanich, and T. Nakanishi, "Realization for Finger Character Recognition Method by Similarity Measure of Finger Features," *International Journal of Smart Computing and Artificial Intelligence* 6(1), 2022. https://doi.org/10.52731/ijscai.v6.i1.684

(16) Ikegami, A., R. Okada, and T. Nakanishi, "The Discovery of Historical Transition in Aesthetic Notions Through Changes in Co-occurrence Words Mainly Used in Waka Poetry in Three Major Poetry Anthologies," in R. Lee (ed.), Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. SNPD 2021. *Studies in Computational Intelligence*, 1012, Springer, 2022. https://doi.org/10.1007/978-3-030-92317-4_12

(17) Nakanishi, T., "Approximate Inverse Model Explanations (AIME): Unveiling Local and Global Insights in Machine Learning Models," IEEE Access, 11: 101020-101044, 2023. https://doi.org/10.1109/ACCESS.2023.3314336

(18) Nakanishi, T., "PCAIME: Principal Component Analysis-Enhanced Approximate Inverse Model Explanations Through Dimensional Decomposition and Expansion," IEEE Access, 12: 121093-121113, 2024. https://doi.org/10.1109/ACCESS.2024.3450299

(19) Nakanishi, T., P. Chophuk, and K. Chinnasarn, "Evolving Feature Selection: Synergistic Backward and Forward Deletion Method Utilizing Global Feature Importance," IEEE Access, 12: 88696-88714, 2024. https://doi.org/10.1109/ACCESS.2024.3418499

(20) Sano, K., K. Ojima, T.Nakamura, R. Okada, T. Nakanishi, "Self-mental Care Management System by Emotion Estimation Method for Heart Rate Variability," *International Journal of Smart Computing and Artificial Intelligence* 8(1), 2024. https://doi.org/10.52731/ijscai.v8.i1.760

(21) Nakanishi, T., "PCaLDI: Explainable Similarity and Distance Metrics Using Principal Component Analysis Loadings for Feature Importance," IEEE Access, 12: 52623-52640, 2024. https://doi.org/10.1109/ACCESS.2024.3387547

(22) Nakanishi, T., A. Minematsu, R. Okada, O. Hasegawa, V. Sornlertlamvanich, "A New Global Sign Language Recognition System Utilizing the Editable Mediator: Integration with Local Hand Shape Recognition," *Information Modelling and Knowledge Bases* XXXV: 227-238, 2024. https://doi.org/10.3233/faia231158
(23) Nakanishi, T., "AIME Image Insights: Transforming Image Recognition with Approximate Inverse Model Explanations," 2024 16th IIAI International Congress on Advanced Applied Informatics, 535-540, 2024. https://doi.org/10.1109/IIAI-AAI63651.2024.00102
(24) Hayashi, K., and T. Nakanishi, "Consumer Price Index Dynamics Unveiled: Approximate Inverse Model Explanations (AIME) of News Text Data and Consumer Price Index Data," 2024 16th IIAI International Congress on Advanced Applied Informatics, 444-449, 2024. https://doi.org/10.1109/IIAI-AAI63651.2024.00087
(25) Ishii, Y., A. Sugiyama, K. Fukushima, R. Okada, and T. Nakanishi, "A Learning Enhancement System for Learner's Community," in R. Lee (ed.), Software Engineering and Management: Theory and Application, Studies in Computational Intelligence, vol. 1137, Springer, 15, 2024. https://doi.org/10.1007/978-3-031-55174-1_15.
(26) 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、2022 年。
(27) 成原慧「メタバースのアーキテクチャと法:世界創造のプラットフォームとそのガバナンス」『Nextcom』52 巻 24-32 頁、2022 年。
(28) 成原慧「差別と選別 : プライバシーとデータ保護の間」有斐閣 ONLINE、2023 年。
(29) 羽柴彩月・大越匡・中澤仁「ユーザビリティの低下による不快感を用いた SNS 利用に対するセルフコントロール手法」『情報処理学会論文誌』64巻4号、2023年。
(30) 中嶋桃香・浜中智己・佐々木航・大越匡・中澤仁「顔表情解析による自然な笑いと作り笑いの識別」『情報処理学会 IoT 行動変容学研究グループ第 3 回研究会（BTI-3）論文集』、 2023 年。
(31) 浜田惇矢・浜中智己・佐々木航・大越匡・中澤仁「Vacuum Colorner: 複合現実を用いた掃除における行動変容を促すアプリケーションの提案」『情報処理学会 IoT 行動変容学研究グループ第 3 回研究会（BTI-3）論文集』、2023 年。

5-2-2. 査読なし（ 15 件）

(1) 宇佐美誠「法的シンギュラリティ」『批判的合理主義研究』15 巻 2 号 28-37 頁、2024 年。
(2) 稲葉振一郎「巨大事故、グローバル災害と人類の未来」『明治学院大学社会学・社会福祉学研究』160 号 107-126 頁、2023 年。
(3) 稲葉振一郎「「コンタクト・パラドックス」とその同類たち」『明治学院大学社会学・社会福祉学研究』161 号 1-31 頁、2023 年。
(4) 神崎宣次「AI と持続可能性」『人工知能』39 巻 6 号 957-966 頁、2024 年。https://doi.org/10.11517/jjsai.39.6_957
(5) 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 巻 12 号 45-51 頁、2021 年。
(6) 成原慧「「インターネットとガバナンス」をめぐる国際的な動向と議論の展開」『法とコンピュータ』41 号 45-52 頁、2023 年。
(7) 成原慧「ChatGPTと法」『法律時報』95巻5号1-3頁、2023年。
(8) 成原慧「民事裁判手続におけるアーキテクチャの機能とその法的意義・統制」『法律時報』95巻11号56-62頁、2023年。

(9) 成原慧「民事判決情報データベース化・オープンデータ化の意義と課題：司法における AI の活用に向けて」『法の支配』215 号 99-110 頁、2024 年。
(10) 成原慧・実積寿也・小熊美紀「規制手法の多様化と法哲学の課題」『情報法制研究』16 巻 34-48 頁、2024 年。
(11) 松尾陽「規制手法の多様化と法哲学の課題」『法律時報』1174号52-57頁、2022年。
(12) 松元雅和「『反攻の象徴』としてのドローンと戦争倫理」『自由思想』166号20-24頁、2022年。
(13) 松元雅和「ウクライナ戦争と平和主義のゆくえ」『世界』964号201-210頁、2022年。
(14) 松元雅和「平和研究としての政治哲学：『理想』を再定義する」『平和研究』60号1-24頁、2023年。https://www.jstage.jst.go.jp/article/psaj/60/0/60_600104/_pdf/-char/ja
(15) 松元雅和「戦争終結の正義」『法と哲学』10号75-94頁、2024年。

5-3. 口頭発表（国際学会発表及び主要な国内学会発表）

5-3-1. 招待講演（国内会議 9 件、国際会議 5 件）

(1) Usami, Makoto, “Address,” Impulse Session: Global Perspectives on AI Ethics, The Responsible AI Forum (TRAIF) 2023, Technical University of Munich, Munich, 13 September 2023.
(2) Usami, Makoto, “Incorporating Distributive Justice into Existing Risk Studies,” Special Seminar on Evaluating Catastrophic Risks, Kyoto University, Kyoto and online, 26 March 2024.
(3) 伊藤昌毅「公共交通オープンデータ第 2 幕：「静的データは出来た、次はリアルタイム」と決めつける前に考えること」令和 3 年度新モビリティサービス検討会、福井県、2021 年。
(4) 伊藤昌毅「行動変容と交通インフラの動的制御によるスマートな都市交通基盤を目指して」電子情報通信学会第 26 回ヒューマンプローブ研究会（HPB）、オンライン、2022 年。
(5) 神崎宣次「倫理の観点からみた EdTech の ELSI 論点」シンポジウム「学習データ利活用 EdTech（エドテック）の ELSI（倫理的・法的・社会的課題）」、オンライン、2022 年。
(6) 成原慧「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、オンライン、2021 年。
(7) 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、オンライン、2022 年。
(8) Narihara, Satoshi “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, hybrid, 2022.
(9) Narihara, Satoshi “Artificial Intelligence, Mis/Disinformation, and Trustworthy Information Ecosystems,” 台湾法務部司法官學院 2024 國際研討會「航向未來：法律如何面對 AI 新浪潮之挑戰」、台湾、2024 年。
(10) Narihara, Satoshi “Personal Autonomy in the Age of Social Robots,” IAS-FRIS Symposium on Social Robots and Ethical Design, 2024.
(11) 成原慧「生成 AI の時代におけるメディアの役割」総務省情報通信法学研究会メディア法分科会（令和 6 年度第 1 回）、2024 年。
(12) 成原慧「AI 社会における自律の揺らぎと再建」2024 年度日本法哲学会学術大会、2024 年。
(13) 松尾陽「監視と自由：権力による監視と権力に対する監視」第 13 回基礎法学総合シンポジウム「『危機は法を破る』のか？　危機管理における人権制約と権力統制の問題」、オンライン、2022 年。
(14) 松元雅和「国家、戦争、平和：カント平和論を側面から探る」日本カント協会第 49 回大会、2024 年。

5-3-2. 口頭発表（国内会議 5 件、国際会議 9 件）

（1） Usami, Makoto, “The Use of Artificial Intelligence in the Judiciary: Possibilities, Principles, and Pitfalls,” The Responsible AI Forum (TRAIF) 2023, Technical University of Munich, Munich, 13 September 2023.

（2） Usami, Makoto, “Robot Judges and the Rule of Law,” 11th Annual International Conference on Business, Law & Economics, Athens Institute for Education and Research, Athens and online, 7 May 2024.

（3） Usami, Makoto, “Facial Recognition and Human Rights: An Autonomy-based Approach,” International Conference on Human Rights (ICHRS-2024), online, 12 September 2024.

（4） Arai, Koki “Data Distribution and Competition Law Issues,” 4th International Workshop on Cross-disciplinary Data Exchange and Collaboration (CDEC2021), online, 2021. DOI: 10.1109/BigData52589.2021.9671329

（5） 寺園結基・伊藤昌毅・大口敬「時刻表データを用いた公共交通網評価手法の提案」『第64 回土木計画学研究発表会』オンライン、2021 年。

（6） 伊藤昌毅「新しい情報通信インフラが実現する行動変容とモビリティ「ITが変えるモビリティとその課題」」情報処理学会 連続セミナー2022「その先へ　情報技術が貢献できること」第5回、2022年9月2日。

（7） 白畑健・伊藤昌毅・大口敬「信号切替り損失時間の解析的評価に基づく信号現示設計」第66回土木計画学研究発表会・講演集、琉球大学、2022年11月12日。

（8） 大塚理恵子・伊藤昌毅・太田恒平・瀬崎薫「複数の交通ビッグデータを組み合わせた地方都市における通勤者の交通利用状況分析」第66回土木計画学研究発表会・講演集、琉球大学、2022年11月13日。

（9） Inaba, Shin-ichiro, “The Contact Paradox and Its Analogues—or How to Deal with Existential Risks,” 23rd Biennial Conference of the Society for Philosophy and Technology (SPT 2023), National Olympics Memorial Youth Center, 10 June 2023.

（10） 神崎宣次「AI 倫理の基礎的な論点」国際シンポジウム「AI（人工知能）の生命倫理」、一橋講堂、2023 年 3 月 5 日

（11） Nakanishi, T., “An Inquirer-Responder Architecture Using LLMs: Emulating Virtual Hearing Q&A in Education,” Proceedings of 2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), International Workshop on ChatGPT for Business Applications, pp. 526–533, 2023. https://doi.org/10.1109/WI-IAT59888.2023.00088

（12） Nakanishi, T., “Detection of Latent Gender Biases in Data and Models Using the Approximate Generalized Inverse Method,” 2024 IEEE 18th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, pp. 191-196, 2024. https://doi.org/10.1109/ICSC59802.2024.00036

（13） Nakata, H. and T. Nakanishi, “Music Impression Extraction Method by Chord Impressions and Its Application to Music Media Retrieval,” Proceedings of 22nd IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2021-Fall), pp. 68-73, 2021.

（14） Narihara, Satoshi, “Personal Autonomy and Machine Autonomy in Human-Robot Interaction,” ALSA 2023 Law and Technology Conference, 2023.

5-3-3. ポスター発表（国内会議 2 件、国際会議 1 件）

（1） Hamanaka, Satoki, Kazunori Sakamoto, Yuki Sasaki, Shinicihiro Mizuno, Yasunori Kawasaki, Wataru Sasaki, Jin Nakazawa, and Tadashi Okoshi, “Adaptive

Push Notification for Behavioral Change in Lifelogging Services," Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services (MOBISYS '24). Association for Computing Machinery, New York, NY, USA, 746–747, 2024. https://doi.org/10.1145/3643832.3661448

(2) 山中翔太・浜中智己・佐々木航・大越匡・中澤仁「スマートフォンのセンサデータに基づくユーザコンテキスト認識型音楽推薦システム」AIoT 行動変容学会第 8 回研究会論文集、岡山市、2024 年 12 月。

(3) 服部久美恵「AI 司法の法哲学的問題：アルゴリズムは法の権威を主張するか」2022 年度「科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践」研究開発プログラム全体会議、東京、2022 年 12 月 18 日。

5-4. 新聞/TV 報道・投稿、受賞など

5-4-1. 新聞/TV 報道・投稿

(1) 松尾陽「憲法季評：デジタル社会の「恵沢」　自由と権力、再考のとき」『朝日新聞』2021 年 7 月 8 日朝刊

(2) 松尾陽「憲法季評：ワクチンの奨励政策 緩やかな規制、生かすには」『朝日新聞』2021 年 10 月 14 日朝刊

(3) 松尾陽「憲法季評：ワクチンの奨励政策　緩やかな規制、生かすには」『朝日新聞』2021 年 10 月 14 日朝刊

(4) 松尾陽「憲法季評：弁護士の電源使用、裁判長が制止　今問う「法の支配」の理念」『朝日新聞』2022 年 1 月 13 日朝刊

(5) 松尾陽朝刊「憲法季評：日本が守るべきものは何か 生命と尊厳と、重い問い 」『朝日新聞』2022 年 4 月 14 日朝刊

(6) 松尾陽「憲法季評：権力構造の「健診」が必要」『朝日新聞』2022 年 8 月 12 日朝刊

(7) 松尾陽「憲法季評：外国人受け入れ拡大から考える　問題は、日本の人権保障」『朝日新聞』2022 年 11 月 10 日朝刊

(8) 松尾陽「憲法季評：自殺やメンタルヘルス対策　ケアと憲法理念の緊張関係」『朝日新聞』2023 年 2 月 9 日朝刊

5-4-2. 受賞

該当なし

5-4-3. その他

該当なし

5-5. 特許出願

5-5-1. 国内出願（０件）

該当なし

5-5-2. 海外出願（０件）

該当なし