

社会技術研究開発事業
令和5年度研究開発実施報告書

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

「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」

宇佐美 誠

（京都大学 大学院地球環境学堂 教授）

目次

1. 研究開発プロジェクト名.....	2
2. 研究開発実施の具体的内容.....	2
2 - 1. プロジェクトの達成目標.....	2
2 - 2. 実施内容・結果.....	2
2 - 3. 会議等の活動.....	8
3. 研究開発成果の活用・展開に向けた状況.....	10
4. 研究開発実施体制.....	10
5. 研究開発実施者.....	11
6. 研究開発成果の発表・発信状況、アウトリーチ活動など.....	13
6 - 1. シンポジウム等.....	13
6 - 2. 社会に向けた情報発信状況、アウトリーチ活動など.....	13
6 - 3. 論文発表.....	14
6 - 4. 口頭発表（国際学会発表及び主要な国内学会発表）.....	15
6 - 5. 新聞／TV報道・投稿、受賞等.....	15
6 - 6. 知財出願（出願件数のみ公開）.....	15

1. 研究開発プロジェクト名

「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」
 Autonomy and Affinity in AI Development and Use

2. 研究開発実施の具体的内容

2-1. 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する。

①AI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）の制度化に向けた本プロジェクトの実践的目標となる。

2-2. 実施内容・結果

（1）スケジュール

研究実施項目	2021年度 (6ヵ月)	2022年度 (12ヵ月)	2023年度 (12ヵ月)	2024年度 (12ヵ月)
1. AI・ビッグデータ分析の倫理的論点の把握				
・開発から生じる課題の探索（開発G）		←————→		
・利用から生じる課題の探索（すべてのG）		←————→		
・論点の抽出と予備分析（自律性G・関係性G）		←————→		
2. 理論分析の枠組みの構築				
・学問的蓄積の共有化（自律性G・関係性G）		←————→		
・先端的研究の把握（自律性G・関係性G）		←————→		
・枠組みの設定と彫琢（自律性G・関係性G）			←————→	
3. 理論分析の実施				
・自律性の観点からの分析（自律性G）			←————→	

・関係性の観点からの分析（関係性G）			←			→
・利用者の信頼の調査（自律性G・関係性G）			←			→
4. 分析結果の公表とその社会実装に向けた発信						
・理論分析結果の発表（自律性G・関係性G）			←			→
・社会実装に向けた発信（すべてのG）				←		→
・国際シンポジウム（すべてのG）					←	→
・邦語論文集の公刊（自律性G・関係性G）					←	→
・行動規範の策定と提案（すべてのG）					←	→

（２）各実施内容

実施項目：項目2（理論分析の枠組みの構築）

実施体制：自律性グループ（以下、自律性G）・関係性グループ（以下、関係性G）

実施内容：先端的研究の把握

前年度に引き続き、自律性G・関係性Gは2023年度前期に、体系的・網羅的な文献調査や国内外の学会大会・ワークショップ等での情報収集などを通じて、倫理的課題の理論分析に関連する諸分野である倫理学・法哲学・政治哲学において、先端的研究の把握を進めた。また、その結果を研究会合での発表・討論・情報交換により共有した。

具体的な実施内容には以下が含まれる。自律性Gでは、研究開発協力者リュトゲ教授らが主催する責任あるAI国際会議（ミュンヘン）において、世界各地から参集した多数の研究者と情報交換・意見交換を行い、最新の研究状況について情報収集を行った。また、アメリカに赴き、情報法の研究者・実務家との意見交換を通じて、AIによる個人の意思決定に対する操作のリスクや個人の自律を保護するための規制の最新動向について調査を実施した。さらに、本プロジェクトの研究開発課題に密接に関連することが新たに認識された司法過程におけるAI使用について、法の支配の稀薄理論に依拠するとした上で、これを市民の自律性と関連付ける最新学説を把握し、国際的学界に貢献することを企図して、この論点を射程に入れることとした。

自律性Gと関係性Gの協働による研究活動として、自律型致死兵器（lethal autonomous weapons systems (LAWS)）に関する倫理的研究の最新の学説状況を網羅的に把握し、LAWSに対する市民意識を調査する意義を特定した。

実施体制：自律性G・関係性G

実施内容：枠組みの設定と彫琢

項目2における先端的研究の把握を踏まえて、自律性G・関係性Gの各人が、理論分析のための枠組みを設定した上で、グループ単位および全グループ合同の研究会合を通じて枠組みを彫琢した。具体的には、自律性Gでは、自律感覚という内的側面と選択肢群の確保や強制・教化の不在という外的側面から自律性を再構成した理論枠組みを設定した。また、AI時代において個人の自

律性を可能にする諸条件について、プライバシーおよびデータ保護にもとづく分析枠組みを設定した。

関係性Gでは、職場でのAI技術の利用に関して、令和4年度に研究組織外の専門家の講演を通じて得られたドイツの労働4.0をめぐる最新情報も踏まえつつ、AI利活用が労働環境や厚生に及ぼす質的影響に焦点をあわせて分析枠組みの設定を行った。また、介護者－被介護者間の関係性の構図においてAI実装ロボットの導入がもたらしうる被介護者の感情変容を分析するための枠組みを発展させた。さらに、AI導入による関係性の変化を考察する具体的場面を、職場・自宅・介護施設等に限定せず都市空間にまで拡張する理論的可能性についても、検討を進めた。

これらに加えて、自律性Gと関係性Gを架橋する横断的作業として、個人レベルでのナビゲーションやモビリティなどに属する新規技術と、そのユーザーの組み合わせを単位としたエージェント分析を進めた。

実施項目：項目3（理論分析の実施）

実施体制：自律性G

実施内容：自律性の観点からの分析

自律性Gでは、前年度後半に引き続き、各メンバーが項目2の研究開発活動の結果に基づいて、AIの利用によって個人の自律性がいかなる意味でどのように影響を受けうるかに関して、近未来の予見を含めた理論分析を行った。また、自律的AIの利用が人間の自己理解や社会实践にどのような影響を及ぼしうるかについても、予見を含めた理論分析を進めた。

具体的には、自律の内的側面と外的側面の両面から行動誘導型アルゴリズムがもたらす問題状況の分析を進めた。また、司法過程でのAI使用が市民の自律性を損なうという法哲学上の最新学説を取り上げ、その当否を精査した。さらに、コンテンツモデレーションに用いられるAIを含めたアルゴリズムの中立性について、米国の裁判例や学説を題材としつつ検討を行った。その他、LAWSをめぐる政治哲学的諸論点に関連して、市民の意識の理論的な位置づけについて分析を行った。すなわち、政治哲学理論において市民は主として啓蒙対象と位置づけられるべきか、あるいはむしろ市民意識を踏まえた理論構築がめざされるべきかについて、検討を行った。その際には、一方では政治的・社会的状況が市民意識に影響を与える点を、他方では技術の進展が意識を規定する点を考慮した。

実施体制：関係性G

実施内容：関係性の観点からの分析

関係性Gでは、前年度後半に引き続き、各メンバーが項目2の結果に基づいて、日本の職場の特徴に留意するとともに社会性ロボット等に関する欧米の研究成果も活用しつつ、AI利活用によって個人間の関係性がいかなる意味でどのように変容しうるかに関して、近未来の予見を含めた理論分析を進めた。具体的には、社会性ロボットのなかでも、超高齢化社会である日本においてとりわけ社会的重要性が高まっている介護ロボットに焦点をあわせ、ロボットの導

入が介護者－被介護者の感情的交流を含む関係性にいかなる影響を及ぼしうるかについて、分析を行った。また、本プロジェクト開始後にその重要性を改めて認識するにいたった視角として、国際的なAI倫理学での一大論点であるアライメント問題に着目した上で、個人間関係の変容や人間－AI関係の創発について、理論分析に着手している。

実施体制：自律性G・関係性G

実施内容：利用者の信頼の調査

自律性Gでは、すでに実施したLAWSの仮想的利用に関する市民意識調査の結果について、解釈・考察をいっそう深化させた。他方、関係性Gでは、介護ロボットに焦点をあわせて、ロボットへの信頼に関する市民意識調査を実施した。先行研究の包括的レビューに基づいて、介護の場面における個人間の関係性にAI実装ロボットが参入することは、被介護者の信頼を規定する各要因には変化をもたらさない一方で、各要因の作用力には影響を及ぼしうると想定した上で、一般市民の意識調査を実施し、予期した結果をおおよそ得られた。

実施項目：項目4（分析結果の公表とその社会実装に向けた発信）

実施体制：自律性G・関係性G

実施内容：理論分析結果の発表

項目3の中間成果を、国内学会大会や国際会議などでの研究報告で発表し、専門家からフィードバックを得て、考察をいっそう発展させた。おもな発表として、自律性Gと関係性Gの協働により、自律性兵器に関する市民意識調査を実施し、その考察を深化させた結果を国際学術誌に投稿し掲載されたことが挙げられる。また、国際技術哲学会の国際会議（東京）や、研究開発協力者リュトゲ教授らが主催する責任あるAI国際会議（ミュンヘン）において、理論分析結果を発表した。

その他、関連する中間成果の発表には、下記のものが含まれる。まず、AIの開発・利活用が今後いっそう進展し、さらには人口超知能（artificial superintelligence (ASI)）が将来的に開発されると、これらは人類全体を脅かしうる存亡リスクとなりうるという国際学界での予見を踏まえて、存亡リスクの認識論的・価値論的・政策論的検討を行う英語論文を公刊した。また、電子情報通信学会での招聘講演や豊田工業大学・次世代文明センターのシンポジウムにおける招待講演なども行った。

実施体制：すべてのグループ（以下、すべてのG）

実施内容：社会実装に向けた発信

項目4における理論分析結果の発表も踏まえつつ、社会実装に向けた発信の仕方について、自律性G・関係性G・開発グループ（以下、開発G）による共同熟議を進めた。これに関連して、理論分析結果の一部を一般市民に向けて平易に解説し、社会実装に向けた意識の喚起や考察の促進を意図して、邦語論文集を公刊する予定としており、出版社も内定している。

（3）成果

実施項目：項目2（理論分析の枠組みの構築）

成 果：自律性Gでは、コンテンツモデレーションに用いられるAIを含めたアルゴリズムの中立性について、米国の裁判例・学説を題材に検討を行った結果、この中立性を判定するための客観的な基準を定立することは困難であり、アルゴリズムが一定の価値判断を元に設計されたものであることを前提として、アルゴリズムの設計者および利用者の役割と責任のあり方を検討する必要があるとの知見を得ている。

項目2は元々、おもに自律性G・関係性Gにより実施される計画であったが、本項目に該当する活動が開発Gでも推進されたため、いくつかの重要な結果を得ている。まず、個人およびAIの自律性や人間-AIの関係性を考察する際に技術的に重要となる説明可能なAI（explainable AI（XAI））について、独自の手法であるapproximate inverse model explanations（AIME）を開発し、その実装も進めた。表データ・画像データについては、特徴重要度の確定を通じて、AIがなぜ当該推定結果を導出したかについて検証を行った。本手法は、機械学習モデルについて人間が理解可能な単純な関数からなる近似逆作用を構築することにより、導出過程の説明を実現するものである。

開発Gにおける別種の成果としては、個人およびAIの自律性の理論分析に関連して、アミューズメントパークをはじめとした娯楽施設における人の主観的満足度と客観的表情・仕草の関係を解明した知見が挙げられる。具体的には、顔表情映像データから主観的満足度を推定するAI技術を開発するべく、データ収集実験およびデータ解析を実施した。データ収集実験から、感情をさほど表出しない日本人について、より明確に顔表情や身振り手振りをしてもらう必要があるという実験設計上の課題が浮かび上がった。また、データ解析からは、映像からの顔表情（facial action coding system（FACS））の検出技術をいっそう向上させる必要性などの課題が明らかとなった。

実施項目：項目3（理論分析の実施）

成 果：自律性Gでは、司法過程でのAI使用と市民の自律性の関係について、前者が後者を毀損するという法哲学上の学説とは反対に、AI使用は自律性を損なわないと論じる論文を公刊した。また、生成AIを利用して「対談」を創作することを通じて、創作者である個人の自律の揺らぎと個人と生成AIとの共創的な関係の構築の可能性を示した。あわせて、生成AIの出力により、個人の意思決定が誘導されたり操作されたりすることで、個人の自律が脅かされる可能性があるとの知見を新たに得た。

LAWsの是非をめぐる戦争倫理的検討において、注目すべき成果が上がっている。具体的には、人間の尊厳に関する哲学的研究に基づき、価値構想と地位構想を区別した上で、意味の重層性を細分化して五つの指標を設定した。次に、この指標に基づき、尊厳に基づく既存のLAWs

反対論を分類し整理した。最後に、以上の分類を踏まえて実施した社会調査の結果を提示し、その含意を分析した。こうした検討を通じて、尊厳に訴えかける議論を構築する際には、人間と機械の存在論的区別に注意を払うべきことが明らかとなった。このような成果に基づく英語論文を執筆し、国際学術誌に投稿中である。

自律性Gと関係性Gの協働によるLAWSの市民意識調査についても、注目すべき成果が出ている。倫理的問題について市民の意見表明が比較的容易だと思われる軍事作戦およびその結果を記述したシナリオを作成し、市民に提示して回答を収集した。その結果、第1に、LAWS投入への否定的姿勢を予期する仮説は否定された。第2に、戦闘員よりも民間人の死傷者数がLAWS投入への肯定的姿勢に影響を及ぼすとする仮説は、強く有意に確認された。第3に、LAWSよりも遠隔兵器に対して市民は投入に積極的だとする仮説は否定され、両者間の差異はほぼ皆無だった。第4に、自国防衛に際してはLAWS投入に対して肯定的だとする仮説は、一定程度まで確認された。第5に、男性や若年者は投入に肯定的だとする仮説については、男性に関しては強く有意に確認された一方で、年齢による傾向はほぼ皆無だった。

関係性Gにおける他の成果としては、AIアライメント問題における倫理性という論点、特に人間の尊厳について分析を行ない、アライメントの先にある価値はいかなるものかを検討したことが挙げられる。このように人間の尊厳という一大理念が、自律性Gと関係性Gに共通する論点として浮上してきている。

開発Gにおいては、とりわけXAIに関して注目すべき成果が上がっている。例えば、近年大きく発展を遂げている生成AIでも、入力したプロンプトがいかに関係性Gに解釈され、言語・画像等が生成しているかはしばしば不明である。そのため、生成AIを含む今後のAIでは、精度等による評価のみならず、説明可能性による適切性の評価もまた重要となる。このようなXAI技術を発展させることが、個人の自律性の確保や人間-AI間の健全な関係のために重要であることを示した。

（4）当該年度の成果の総括・次年度に向けた課題

・本研究開発プロジェクトの達成目標に対する現在の進捗状況については、全般的に順調に実施してきたものと評価している。加えて、項目4（分析結果の公表とその社会実装に向けた発信）については、中間成果の論文公刊・国際会議報告等の面で当初の予定よりも進展している。

・各実施項目で得られた結果や成果を俯瞰・統合した結果、本プロジェクトの中心的目標を達成するためには、当初に予定していた各項目での研究開発作業を複数の方向で拡張する必要があることが明らかとなった。その理由の一つは、新興技術に対する理論的応答の必要性であり、例えば生成AIが固有または典型的に提起する新たな問題群を視野に入れるべきだと判断している。作業を拡張する別の理由は、問題構造の分析を通じて、当初はその重要性を強く意識していなかった論点を詳細に解明する必要性が新たに認識されたことにある。例えば、重要な技術的課題である

XAIや、AI倫理学上の主要概念であるアラインメントは、自律性・関係性の双方に関連するため、これらへの取組みを本格的に行うべきだと考えるにいたった。

・本プロジェクトの最終年度にあたる次年度には、当初から予定している研究成果の発表および行動規範の策定を十全に実施する予定である。その際、上記のような研究開発作業の新たな拡張部分をどのように組み込むかが一課題となる。研究組織内で重ねてきた討議の結果も踏まえつつ、拡張部分の組み込み方についてさらなる検討を行う予定である。

2 - 3. 会議等の活動

年月日	名称	場所	概要
2023年5月13日	サイトビジット・定例研究会合	Zoom	研究開発代表者・実施者が順次、2022年度の研究活動・成果および2023年度の研究計画について報告を行った上で、アドバイザーから質問・助言等を受けた。
2023年6月24日	定例研究会合	Zoom	研究実施者による以下の研究報告が行われた。 ・荒井弘毅「介護の場面における関係性：人とロボットやAIの親近性の要素」 ・稲葉振一郎「RISTEX-ELSI「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」2023年度研究計画」
2023年8月25日	定例研究会合	Zoom	最終成果物の一部をなす邦語論文集の公刊に向けて、打合せを行った。
2023年9月30日	外部研究者招聘講演会兼定例研究会合	Zoom	外部研究者による講演と研究開発実施者による研究報告を、以下のとおり行った。 ・伊吹友秀「ロボットやAIに看護業務を委託することの可能性と倫理的課題について：ロボット倫理と看護倫理の交わる場所」 ・伊藤昌毅「交通行動変容と未来の交通」
2023年12月16日	定例研究会合	Zoom	研究実施者による以下の研究報告が行われた。 ・松尾陽「AIと弁護士制度」
2024年2月	定例研究会合	Zoom	研究実施者による以下の研究報告

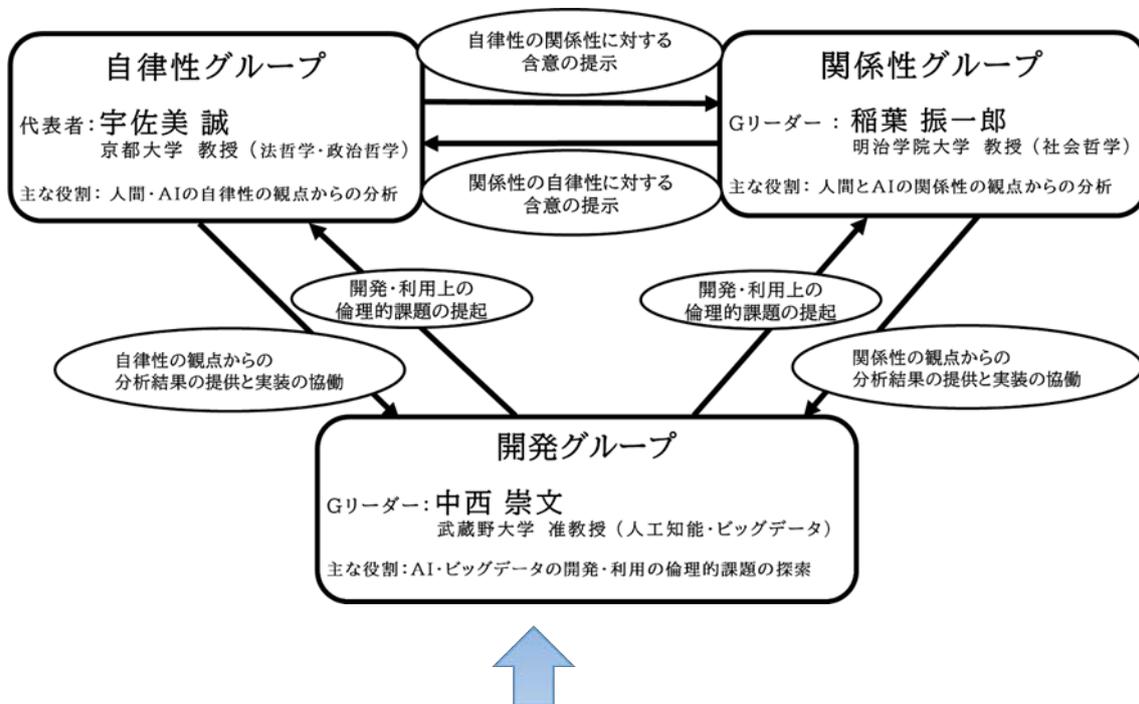
10日			が行われた。 ・大越匡「ユビキタス・コンピューティング研究と自律性を含む倫理的課題の最近の動向」 ・宇佐美誠「法的シンギュラリティと法の支配：市民の合意および自律の観点から」
2024年3月30日	定例研究会合	Zoom	研究実施者による以下の研究報告が行われた。 ・神崎宣次「人間とAIの関係性という観点からのアライメント問題」

3. 研究開発成果の活用・展開に向けた状況

該当なし

4. 研究開発実施体制

自律性G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ドローンを研究対象とする。関係性Gは、人間-AIの関係性をテーマとし、職場協働型AI・ロボットと介護・会話ロボットを対象とする。開発Gは、実施者による研究開発や、技術者等とのトランス・ディシプリナリーな共同探索を通じて、上記の研究対象にかかる開発・利用上の倫理的課題を同定し、他の2グループに提示する。自律性G・関係性Gは理論分析を実施して、その結果を開発Gさらには連携先の技術者へとフィードバックする。このように螺旋的に発展する学際的協働は、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)

5. 研究開発実施者

自律性グループ（リーダー氏名：宇佐美 誠）

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	国立大学法人京都 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	国立大学法人東海 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	国立大学法人九州 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	国立大学法人京都 大学	大学院地球環 境学堂	研究員

関係性グループ（リーダー氏名：稲葉 振一郎）

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	学校法人明治学院	明治学院大学 社会学部	教授
荒井 弘毅	アライ コ ウキ	学校法人共立女子 学園	共立女子大学 ビジネス学部	教授
神崎 宣次	カンザキ ノブツグ	学校法人南山学園	南山大学 国 際教養学部	教授
服部 久美恵	ハットリ クミエ	国立大学法人京都 大学	大学院地球環 境学堂	研究員

開発グループ（リーダー氏名：中西 崇文）

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	学校法人武蔵野大 学	データサイエ ンス学部	准教授
大越 匡	オオコシ タダシ	学校法人慶應義塾	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	国立大学法人東京 大学	大学院情報理 工学系研究科	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

該当なし

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、フリーペーパー、DVD

- ・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、181-206頁、2023年9月
- ・Makoto Usami, “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, pp. 71-91. September 2023.
- ・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
- ・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

(2) ウェブメディアの開設・運営

該当なし

(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

- ・Makoto Usami, “AI Judges: Potentials and Pitfalls,” Lecture for the students of Northwestern Polytechnical University in China, Kyoto University, 12 July 2023. (招聘講演)
- ・Makoto Usami, “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. (招聘講演)
- ・Makoto Usami, “Incorporating Distributive Justice into Existing Risk Studies,” Special Seminar on Evaluating Catastrophic Risks, Kyoto University, Kyoto and online, 26 March 2024. (招聘講演)
- ・Nobutsugu Kanzaki, “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 (SPT2023), 国立オリンピック記念青少年総合センター、2023年6月8日
- ・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

- ・ 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（招聘講演）
- ・ 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（招聘講演）

6-3. 論文発表

(1) 査読付き (4 件)

●国内誌 (0 件)

該当なし

●国際誌 (4 件)

- ・ Koki Arai (2024), “Law and Economics of Language Model Development: Empirical Examination of Corporate Strategies and Vaporware Claims,” *Asian Journal of Law and Economics*, doi.org/10.1515/ajle-2023-0118
- ・ Arai, Koki and Masakazu Matsumoto (2023), “Public Perceptions of Autonomous Lethal Weapons Systems,” *AI and Ethics*, online first <https://link.springer.com/article/10.1007/s43681-023-00282-9>
- ・ Matsumoto, Masakazu (2023), “The Right and Justice of Subsistence Wars as Necessity: A Grotian Account,” *International Politics*, online first <https://link.springer.com/article/10.1057/s41311-023-00540-x>
- ・ T. Nakanishi (2023), “Approximate Inverse Model Explanations (AIME): Unveiling Local and Global Insights in Machine Learning Models,” in *IEEE Access*, vol. 11, pp. 101020-101044. <https://doi.org/10.1109/ACCESS.2023.3314336>

(2) 査読なし (5 件)

- ・ 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
- ・ 成原慧「ChatGPTと法」『法律時報』95巻5号1-3頁、2023年
- ・ 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
- ・ Satoshi Narihara (2023), “Personal Autonomy and Machine Autonomy in Human-Robot Interaction,” ALSA 2023 Law and Technology Conference, 20231202
- ・ 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.istage.jst.go.jp/article/psaj/60/0/60_600104/_pdf/-char/ja

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 1 件）

- ・ Makoto Usami, “Address,” Impulse Session: Global Perspectives on AI Ethics, The Responsible AI Forum (TRAI F) 2023, Technical University of Munich, Munich, 13 September 2023.

(2) 口頭発表（国内会議 0 件、国際会議 4 件）

- ・ Makoto Usami, “The Use of Artificial Intelligence in the Judiciary: Possibilities, Principles, and Pitfalls,” The Responsible AI Forum (TRAI F) 2023, Technical University of Munich, Munich, 13 September 2023.
- ・ Shin-ichiro Inaba, “The Contact Paradox and Its Analogues—or How to Deal with Existential Risks.” 23rd Biennial Conference of the Society for Philosophy and Technology (SPT2023), 国立オリンピック記念青少年総合センター、2023年6月10日
- ・ T. Nakanishi, “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, 2024 pp. 191-196. <https://doi.org/10.1109/ICSC59802.2024.00036>
- ・ T. Nakanishi, “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>

(3) ポスター発表（国内会議 0 件、国際会議 0 件）

該当なし

6-5. 新聞／TV報道・投稿、受賞等

(1) 新聞報道・投稿（0 件）

- ・ 該当なし

(2) 受賞（0 件）

- ・ 該当なし

(3) その他（0 件）

- ・ 該当なし

6-6. 知財出願（出願件数のみ公開）

(1) 国内出願（0 件）

- ・ 該当なし

(2) 海外出願（0 件）

- ・ 該当なし