

研究プロジェクト

AGI時代における民主主義防衛のための 倫理原則再構築と社会技術インフラ整備（第一報）

藤井 雄作^{1,*}¹群馬大学*fujii@gunma-u.ac.jp

要旨：本稿は、令和8（2026）年度科学研究費助成事業・挑戦的研究（萌芽）として採択された研究課題「AGI時代における民主主義防衛のための倫理原則再構築と社会技術インフラ整備」（課題番号26K21972）の研究計画を開示するものである。本研究は、人工汎用知能（Artificial General Intelligence：AGI）の進展によって、人間の労働、尊厳、プライバシーおよび民主主義の基盤が大きく変容し得るとの問題意識に立つ。第一の課題は、AGI時代にも民主主義が成立するための前提条件として、「労働貢献なき人間の尊厳」をいかに確立し得るかを探究することである。第二の課題は、民主主義を維持するための必要条件の一つとして、プライバシーを保護しつつAI出力を統治する社会技術制度を探究することであり、VRAIO（Verifiable Record of AI Output）とGLO（Guide to the Expression of Legitimacy of Output）をその中核に置く。これら二つの課題を達成しても、AGI時代の民主主義防衛にとって十分であるとは限らない。本研究は、それでもなお避けて通れない前提条件と必要条件を明らかにし、実装可能な基礎的枠組みへ進めることを目指す。研究期間は令和8～10年度の3年間、研究代表者単独で実施し、直接経費は申請額どおり総額500万円が措置された。

キーワード：AGI、人間の尊厳、民主主義、プライバシー、VRAIO、AI出力統治

1. はじめに

人工汎用知能（Artificial General Intelligence：AGI）がいつ、どのような形で実現するかは確定していない。しかし、人間と同等またはそれ以上の知的能力を持つ人工システムが実現した場合、その影響は個別の産業や職業にとどまらず、人間の生き方、社会参加、政治的意思決定、情報環境および民主主義の制度基盤にまで及ぶ可能性がある。

特に、AGIが人間の知的・身体的労働を広範に代替する社会では、人間は富の源泉としての役割を失い得る。その帰結として、富はAGIやロボットの所有者に集中し、格差の拡大が社会の結束と民主主義の基盤を脅かし得る。生活保障のためにユニバーサル・ベーシック・インカム（Universal Basic Income：UBI）等が導入されたとしても、それだけで、人間が民主主義の担い手として尊重され続けることが保証されるわけではない。そこで本研究は、労働や経済的有用性に依存しない人間の尊厳を確立することを、AGI時代にも民主主義が成立するための前提条件として位置付ける[1]。

同時に、AIシステムが公共空間の監視、情報推薦、行政判断、教育、医療、交通その他の社会機能に組み込まれると、プライバシー侵害、恣意的な選別、説明不能な判断、世論形成への介入等が大規模かつ継続的に生じ得る。したがって、民主主義を維持するためには、プライバシーを保護しつつ、社会へ出されるAI出力を検証可能な形で統治することが必要である。本研究は、これを必要条件の一つとして位置付け、VRAIOとGLOを中核とする社会技術制度を探究する[2-7]。

本研究計画は、令和8（2026）年度科学研究費助成事業・挑戦的研究（萌芽）として採択された。課題番号は26K21972、研究期間は令和8～10年度、研究代表者単独で実施する。直接経費は申請額どおり満額の500万円である。本稿は、研究成果の報告ではなく、採択された研究計画の背景、目的および実施構想を社会に開示することを目的とする。

2. 研究の目的

2. 1. AGI時代における人間の尊厳の倫理的再定義

本研究の第一の目的は、AGIが人間の労働を広範に代替し、人間が富の源泉としての役割を失い得る社会において、人間の尊厳をどのような原理によって確立できるかを検討することである。これは、AGI時代にも民主主義が成立するための前提条件に関する探究である。近代社会では、労働は所得獲得の手段であるだけでな

く、社会的貢献、承認、自己実現および尊厳と深く結び付けられてきた。この結び付きが失われるとき、尊厳を労働能力や経済的有用性によって測る考え方そのものを再検討する必要がある。

その手掛かりとして、本研究は古代ローマの「パンとサーカス」に注目する[1]。通常、「パンとサーカス」は、市民を物質的給付と娯楽によって懐柔した政治の象徴として否定的に理解される。しかし、本研究が問うのは、社会の富の生産に直接寄与しない多数のローマ市民に、なぜパンとサーカスが与え続けられたのか、という点である。その背景には、「ローマ市民であること」自体に固有の地位と価値、すなわち市民としての尊厳が認められていた可能性がある。この通常とは異なる解釈に、労働貢献を失うAGI時代においても、人間を民主主義の主体として尊重し続けるためのヒントが存在するかもしれない。本研究は、この可能性を起点として、「労働貢献なき人間の尊厳」を支える倫理原則を探究する。

表1. 研究プロジェクトの概要

研究種目	科学研究費助成事業・挑戦的研究（萌芽）
課題番号	26K21972
研究課題名	AGI時代における民主主義防衛のための倫理原則再構築と社会技術インフラ整備
研究期間	令和8（2026）年度～令和10（2028）年度
研究体制	研究代表者単独
直接経費	総額500万円（申請額どおり満額採択）
中心課題	人間の尊厳の倫理的再定義／AIシステムからプライバシーを守る社会技術制度

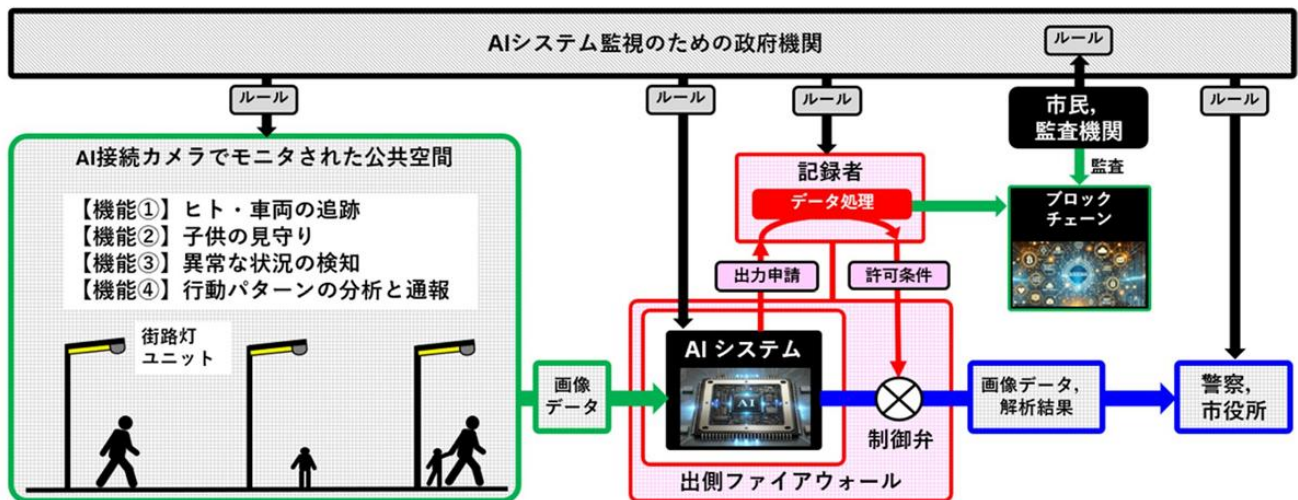


図1 「AI出力の検証可能な記録（VRAIO）」を実装した「AI接続カメラでモニタされた公共空間」

2. 2. AIシステムからプライバシーを守る社会技術制度

第二の目的は、民主主義を維持するための必要条件として、AIシステムからプライバシーを守りつつ、外部に提供されるAI出力を検証可能にする社会技術制度を構築することである。その中核となるのが、著者が提案してきたVRAIO（Verifiable Record of AI Output）と、出力の正当性を表現するためのGLO（Guide to the Expression of Legitimacy of Output）である[2-7]。

VRAIOでは、AIシステムの外部境界に出側ファイアウォールを設け、外部提供を求める出力申請、社会的議論に基づいて定められたルール、許可条件、候補出力および実際に提供された出力に関する記録を、記録者（Recorder）が検証可能な形で保存する。GLOは、候補出力がどのような根拠、ルールおよび正当性の下で提示されるのかを表現する共通言語としてVRAIOを補完する。なお、GLOは科研費申請書の記載には含まれず、その後の研究[7]で定式化した枠組みである。これにより、AIの学習データや内部モデルの全面公開を前提とせず、

外部出力が定められたルールに適合していたか、また、どの程度の正当性を持つものとして提供されたかを事後に検証・監査できる。

VRAIOが特に対象とするのは、外部からの攻撃者に比べて対策が困難な、システムの所有者・運用者すなわち内部者による悪用・誤用である。EUのAI ActおよびGDPR、米国のAI Risk Management Frameworkといった既存のAI規制が、基本的人権影響評価（FRIA）やデータ保護影響評価（DPIA）等の文書化義務と事後的な監査・制裁を主眼とするのに対し、VRAIOは、出力境界におけるリアルタイムのルール照合と改ざん困難な記録によって内部脅威を事前に統制する点に特徴がある。虚偽の出力申請に対しては、巨額の罰則、通報に対する報奨金および被害者への補償制度を組み合わせる。なお、科研費申請書では、第三者機関としてのRecorderが出力の審査と記録の双方を担う構成として簡潔に記述しているが、本稿では、ルール照合を出側ファイアウォールが担い、Recorderは記録に専念する構成として記述する。

図1は、AI接続カメラによってモニタされた公共空間を対象としたVRAIOの基本構成を示す。公共空間から得られた画像データはAIシステムで処理されるが、画像データや解析結果を警察、市役所等へ提供する際には、出側ファイアウォール内で出力申請と許可条件が照合される。記録者はデータ処理の記録を保持し、市民・監査機関はブロックチェーン等を利用して監査する。この構成は、監視機能の有用性とプライバシー保護を両立させるための一例である[2]。

2. 3. 二つの課題の統合

第一課題は、AGI時代にも民主主義が成立し得るための前提条件として、人間の尊厳の確立を探究するものである。第二課題は、その民主主義をAIシステムの下で現実維持するための必要条件として、プライバシー保護とVRAIO+GLOによるAI出力統治を探究するものである。人間の尊厳が確立されなければ、人間が民主主義の主体であり続ける根拠そのものが失われ得る。一方、人間の尊厳を掲げるだけでは、AIシステムによる具体的なプライバシー侵害や不当な出力を防ぐことはできない。

したがって、本研究は、第一課題を第二課題のルール形成を方向付ける倫理的基盤として位置付け、人間の尊厳、プライバシー、自己決定、政治参加および説明責任に関する原則を、AIシステムへ適用可能なルールへ翻訳し、その適用過程をVRAIO+GLOによって検証可能にする枠組みを検討する。ただし、これら二つの課題が達成されても、AGI時代の民主主義防衛の十分条件が満たされるとは限らない。経済制度、教育、権力分配、安全保障その他の課題も残る。本研究は、民主主義防衛を完成させると主張するのではなく、避けて通れない前提条件と必要条件を、可能なところから具体化していく挑戦である。

3. 研究方法

3. 1. 倫理原則の検討

AGIが人間の労働、所得、社会的役割および政治参加に与える影響について、文献研究と思考実験を用いて検討する。具体的には、①人間の労働がほぼ全面的に代替された社会、②UBI等によって生活は保障されるが、人間が富の生産主体としての役割を失った社会、③AGIが人間より高い精度で政策判断を行う社会、④市民が意思決定をAGIへ過度に委任する社会等を想定する。また、古代ローマにおいて、生産への直接的寄与を失った市民にも「市民であること」自体に価値が認められていた可能性を検討し、AGI時代における人間の尊厳の根拠を探究する。

具体的なアプローチとして、(1)文化・物語を通じた価値の内面化（尊厳を肯定する社会ナラティブの構築）、(2)習慣の定着（尊厳を尊重する法規範）、(3)社会参加の保障（多様なステークホルダーの対話機会）、(4)AI適応型倫理教育を組み合わせた社会制度設計を検討し、労働に依拠しない人間の尊厳の倫理的根拠を、文化的ナラティブ・社会制度・教育システムの三層構造としてモデル化する。検討にあたっては、フィンランドのUBI実験等のケーススタディを参照し、国際会議での議論を通じて考察を深める。最終的には、民主主義を支える新たな尊厳モデルを、国際的に共有可能な指針として提示することを目指す。

これらの想定を通じて、労働能力や経済的有用性とは独立に、人間を民主主義の主体として尊重するための条件を整理する。ここで目指すのは、人間の尊厳について最終的・唯一の定義を与えるのではなく、AGI時代にも民主主義を成立させるために不可欠な前提条件として、どのような倫理原則を確立すべきかを明らかにすることである。

3. 2. VRAIOに基づく社会技術制度の設計

AIシステムから外部へ出される情報を統治対象として、出力申請、許可条件、ルール照合、GLOによる正当性の表現、記録、監査および異議申立ての手续を設計する。対象例として、AI接続カメラ、生成AIおよびSNS、行政・公共サービスに用いられるAI等を想定する[2-7]。

各事例について、①誰がルールを定めるか、②どの出力を記録対象とするか、③GLOによってどのような正当性情報を表現するか、④どの情報を市民および監査機関へ開示するか、⑤記録者の独立性をどのように確保するか、⑥制度の有効性をどのような指標で評価するかを検討する。必要に応じて、出力の適法性、プライバシー侵害の抑制、監査可能性、説明責任等を評価指標として用いる。

3. 3. 制度設計指針と評価枠組みの提示

第一課題で探究する人間の尊厳の倫理原則と、第二課題で設計するVRAIO+GLOに基づく社会技術制度を接続し、AGI・AIシステムを民主的統制の下に置くための制度設計指針を提示する。本研究は萌芽的研究であるため、大規模な社会実装や民主主義防衛の完成を直ちに目標とするものではない。人間の尊厳という前提条件と、プライバシー保護およびAI出力統治という必要条件について、今後の実証研究、標準化および制度化につながる基礎的枠組みを整備する。また、公共的な技術を個人の自由と両立させる社会技術システムの構築という観点では、著者が進めてきた市民用PAPR研究[8-12]も参照し、技術性能だけでなく、運用、受容性、自由および社会的統治を一体として扱う。

4. 研究計画

本研究は3年間で実施する。申請書に示したとおり、本研究は統合モデルの理論的検討にフォーカスし、個別の実装モデルについては別途の大型プロジェクトとの連携を予定する。研究の進行に応じて、第一課題である人間の尊厳の探究と、第二課題であるプライバシー保護およびVRAIO+GLOに基づく社会技術制度の設計を往復させ、最終年度に両者の接続を試みる。ただし、この統合はAGI時代の民主主義防衛を完成させるものではなく、今後の研究と制度形成へ向けた基礎を整えるものである。

表 2. 年度別研究計画および直接経費

年度	主な研究内容	直接経費
令和8年度	AGI時代の人間の尊厳に関する論点整理、思考実験の設定、VRAIOの基本構造と適用対象の整理	130万円
令和9年度	倫理原則候補の体系化、AI出力統治の制度設計、事例別のルール・記録・監査手続の検討	140万円
令和10年度	倫理原則と社会技術インフラの統合、評価枠組みの整理、制度設計指針および今後の研究課題の提示	230万円

直接経費の総額は500万円であり、旅費120万円、人件費・謝金30万円、その他350万円を計画している。研究会、学会発表、国内外の研究者との意見交換、資料収集、論文公表および制度設計に必要な調査等に使用する。

5. 期待される成果と学術的意義

第一に、AGI時代にも民主主義が成立し得るための前提条件として、人間の尊厳を、労働能力や経済的有用性だけに依存しない形で再検討する。古代ローマにおいて、社会の富の生産に直接寄与しない市民にもパンとサーカスが与え続けられた背景に、市民としての尊厳が確立していた可能性を見だし、そこから「労働貢献なき人間の尊厳」を支える倫理原則の候補を提示することが期待される。

第二に、民主主義を維持するための必要条件として、プライバシーを保護しつつ、AIシステムの具体的な出力を検証・記録・監査する制度へ接続する。VRAIO+GLOを用いることにより、AI内部の全面公開が困難な場合でも、外部へ提供された出力、その正当性の表現および適用された手続を検証する道を示すことができる。

第三に、前提条件としての人間の尊厳と、必要条件としてのプライバシー保護・AI出力統治を接続することで、倫理原則をAIシステムへ適用可能なルールへ翻訳し、その適用を検証可能にする基礎的枠組みを提示する。ただし、この枠組みだけでAGI時代の民主主義防衛が完結するわけではない。本研究の学術的意義は、完全な解決を装うことなく、避けて通れない二つの条件を明示し、具体的な研究対象として前進させる点にある。

これまでの研究では、VRAIOを公共空間監視[2]、自動運転[3]、スマートフォンによる緊急検知[4]、および継続的センシング[5]へ展開し、さらに、社会規模のAIプラットフォーム出力を対象とする安全弁としてのVRAIO[6]、拡張ケルビン原理に基づく信頼インフラとAI出力統治[7]へと発展させてきた。また、市民用PAPRについても、ロックダウン代替手段としての実現可能性[8]、社会的要件[9]、呼吸状態推定[10]、単一差圧センサによる呼吸流量・漏れ推定[11]、時間管理型利用による感染制御と自由の両立[12]を検討してきた。これらは、倫理、プライバシー、検証可能性、社会受容性および個人の自由を、個別技術の外側にある社会技術システムとして扱う一連の研究である。

本研究は、これらの到達点を単に総括するものではない。AGIによって人間の労働、尊厳および民主的意思決定の前提自体が変容し得る状況を対象として、第一に、人間の尊厳の確立という前提条件を探究し、第二に、プライバシー保護とVRAIO+GLOによるAI出力統治という必要条件を具体化する。そして両者を可能な範囲で接

続する。この二つを達成しても、AGI時代の民主主義防衛にとって十分ではない可能性を認めた上で、なお、そのための基礎を一步步築くことを目指す。あわせて、著者が編集するAI誌特集号「Governing Trustworthy AI Outputs in a Sensor-Dense Society: Privacy, Auditability and Responsible Deployment」[13]を国際的議論の場として活用し、本研究で得られる構想を広く検証する。

6. 研究上の留意点

本研究は、現時点で実現形態が確定していないAGIを対象とするため、将来予測には本質的な不確実性がある。また、人間の尊厳や民主主義に関する価値判断には複数の立場があり、第一課題から唯一の倫理原則が導かれるとは限らない。さらに、前提条件としての人間の尊厳と、必要条件としてのプライバシー保護・AI出力統治を確立しても、民主主義防衛の十分条件が満たされるとは限らない。したがって、本研究では特定の未来像や完全な解決を断定せず、複数のシナリオを比較し、前提と限界を明示した上で議論する。

VRAIO+GLOについても、記録、正当性表現および監査の仕組み自体が過度な監視や統制に利用される危険がある。このため、記録対象の限定、目的外利用の禁止、監査主体の独立性、市民による異議申立ておよび制度自体の検証可能性を重視する。

7. おわりに

AGI時代における民主主義の防衛は、AGIを技術的に制御するだけで達成できるものではない。第一に、富の源泉としての役割を失った後も、人間が民主主義の主体として尊重されるための尊厳を確立しなければならない。これは民主主義成立の前提条件である。第二に、AIシステムからプライバシーを守り、社会へ出されるAI出力を検証可能に統治しなければならない。これは民主主義維持の必要条件の一つである。

本研究は、第一課題として「労働貢献なき人間の尊厳」の根拠を探究し、第二課題としてVRAIO+GLOを中核とするプライバシー保護・AI出力統治の社会技術制度を探究し、両者の接続を試みる。これらを達成しても、AGI時代の民主主義防衛が完成するとは限らない。それでも、避けて通れない前提条件と必要条件を明らかにし、可能なところから具体化していくことには意義がある。本稿は採択された研究計画を開示する第一報であり、今後、研究の進展に応じて理論的検討、制度設計、適用事例および残された課題を順次報告する予定である。

謝辞

本研究は、日本学術振興会科学研究費助成事業・挑戦的研究（萌芽）「AGI時代における民主主義防衛のための倫理原則再構築と社会技術インフラ整備」（課題番号26K21972、令和8～10年度）の助成を受けて実施する。

参考文献

- 1) Y. Fujii, “Lessons from the Roman Empire: ‘Bread and Circuses’ as a Model for Democracy in the AGI Age,” *AI & Society*, Vol. 41, pp. 467-468, 2026. <https://doi.org/10.1007/s00146-025-02449-w>
- 2) Y. Fujii, “Verifiable Record of AI Output for Privacy Protection: Public Space Watched by AI-Connected Cameras as a Target Example,” *AI & Society*, Vol. 40, pp. 3697-3706, 2025. <https://doi.org/10.1007/s00146-024-02122-8>
- 3) Y. Fujii, “Governing AI Output in Autonomous Driving: Scalable Privacy Infrastructure for Societal Acceptance,” *Future Transportation*, Vol. 5, No. 3, 116, 2025. <https://doi.org/10.3390/futuretransp5030116>
- 4) Y. Fujii, “Smartphone-Based Sensing Network for Emergency Detection: A Privacy-Preserving Framework for Trustworthy Digital Governance,” *Applied Sciences*, Vol. 16, No. 2, 1032, 2026. <https://doi.org/10.3390/app16021032>
- 5) Y. Fujii, “Governing Continuous Smartphone Sensing: A Privacy-Preserving Path Toward Democratic Homeland Security,” *Journal of Homeland Security and Emergency Management*, Opinion, 2026. <https://doi.org/10.1515/jhsem-2026-0014>
- 6) Y. Fujii, “VRAIO as a Safety Valve: Governing AI Platform Outputs at Societal Scale,” *AI and Ethics*, Vol. 6, 340, 2026. <https://doi.org/10.1007/s43681-026-01206-z>
- 7) Y. Fujii, “No Trust Without Trust Infrastructure: The Extended Kelvin Principle and Its Application to AI Output Governance,” *AI*, Vol. 7, No. 6, 218, 2026. <https://doi.org/10.3390/ai7060218>

- 8) Y. Fujii, “An Engineering Alternative to Lockdown During COVID-19 and Other Airborne Infectious Disease Pandemics: Feasibility Study,” *JMIR Biomedical Engineering*, Vol. 9, e54666, 2024. <https://doi.org/10.2196/54666>
- 9) Y. Fujii, “Examination of the Requirements for Powered Air-Purifying Respirator (PAPR) Utilization as an Alternative to Lockdown,” *Scientific Reports*, Vol. 15, 1217, 2025. <https://doi.org/10.1038/s41598-024-82348-0>
- 10) Y. Fujii, A. Takita, S. Hashimoto and K. Amagai, “Estimation of Respiratory States Based on a Measurement Model of Airflow Characteristics in PAPR Using Differential Pressure and PWM Control Signals: In the Development of a Public-Oriented PAPR as an Alternative to Lockdown Measures,” *Sensors*, Vol. 25, No. 9, 2939, 2025. <https://doi.org/10.3390/s25092939>
- 11) Y. Fujii, “The Real-Time Estimation of Respiratory Flow and Mask Leakage in a PAPR Using a Single Differential-Pressure Sensor and Microcontroller-Based Smartphone Interface in the Development of a Public-Oriented Powered Air-Purifying Respirator as an Alternative to Lockdown Measures,” *Sensors*, Vol. 25, No. 17, 5340, 2025. <https://doi.org/10.3390/s25175340>
- 12) Y. Fujii, “Time-Managed PAPR Use Enables a Balanced Approach to Infection Control and Personal Freedom,” *Scientific Reports*, Vol. 16, 2878, 2026. <https://doi.org/10.1038/s41598-025-32625-3>
- 13) AI, Special Issue, “Governing Trustworthy AI Outputs in a Sensor-Dense Society: Privacy, Auditability and Responsible Deployment,” edited by Yusaku Fujii, MDPI, submission open. https://www.mdpi.com/journal/ai/special_issues/74CM45UW2P



藤井雄作 FUJII, Yusaku
群馬大学 大学院理工学府

1989年3月東京大学工学部船舶工学科卒業，1991年3月東京大学大学院工学系研究科修士課程修了．2001年東京大学より博士（工学）の学位修得．1991年4月川崎製鉄株式会社入社，工業技術院計量研究所，産業技術総合研究所 を経て，現在群馬大学理工学府教授．2004年よりNPO法人e自警ネットワーク研究会理事長．AIガバナンス，民主主義の防衛，社会安全工学，防犯カメラシステム，精密計測，光波干渉計などが専門．