PRESS RELEASE

群馬大学 GUNMA UNIVERSITY

2026 年 6 月 18 日

報道関係者 各位

AI の「出力」を統治するための原理と基盤を提示した論文が、学術誌『AI』に掲載されました。

― 拡張ケルビン原理に基づき、AI 出力規制の必要性と、その実装基盤 VRAIO を提案 ―

群馬大学大学院理工学府知能機械創製部門の藤井雄作教授が提唱する AI 統治の枠組みに関する論文が、この度、学術誌『AI』に掲載されました。本論文は、AI が社会に及ぼす影響を「出力（アウトプット）」の段階で検証可能な形に統治するための原理と基盤を、計量標準（メートル法に代表される計測の制度設計）との構造的な対応関係から導いたものです。

社会に深く浸透する AI は、推薦・判断・警告・行政支援・医療助言など、多様な「出力」を通じて人々や社会に作用します。しかし、こうした出力が「どの規則に基づき、誰が、どのように検証・記録したのか」を事後に確認できる仕組みは、社会基盤の規模では十分に整備されていません。本論文は、この欠落こそが AI に対する正当な社会的信頼を妨げていると指摘し、「信頼は宣言ではなく基盤によって生まれる」という観点から、AI 出力統治のあるべき制度設計を提示します。

論文の概要

背景：2020 年以降に各国で導入された接触確認アプリ（日本の COCOA 等）は、技術的にプライバシー保護的な設計がなされていても、社会的な効果を十分に発揮できませんでした。著者はこの経験を、「社会的計測（social measurement）は技術だけでは成立しない」ことを示す象徴的事例と位置づけます。人は、測られることに対して意思を持つため、システムへの信頼がなければデータは集まらず、計測は成立しないからです。

原理（拡張ケルビン原理）：「測定できなければ理解できない、理解できなければ制御できない」というケルビンの命題に対し、社会的計測の前提として次の連鎖を前置します。すなわち「信頼基盤なくして社会的信頼なし、社会的信頼なくして正当な社会的計測なし」。本論文はこれを拡張ケルビン原理として定式化します。

枠組み（GLO／VRAIO）：計量分野において GUM（計測の不確かさの表現ガイド）と校正インフラが計測値への社会的信頼を支えているのと同様に、AI 出力統治には、出力の正当性を表現する共通言語 GLO（Guide to the Expression of Legitimacy of Output）と、それを実装する基盤 VRAIO（Verifiable Record of AI Output：検証可能な AI 出力記録）が必要であると論じます。GUM と校正インフラの関係が、GLO と VRAIO の関係に構造的に対応する――これが本論文の核心です。

結論：VRAIO は、メタデータ宣言・規則照合・改ざん耐性のある記録・独立監査を統合します。封印された決定論的な検証器（Rule-Judgment AI）により出力の正当性を再現可能な形で検証し、計算上の虚偽は再計算で、事実の虚偽は権威ある記録との照合で検知します。重い制裁と抜き打ち監査を組み合わせることで、虚偽申告を「割に合わない」選択とし、全件監視によらずに統治を成立させる点に特徴があります。

関連する最新の研究成果

本論文は、藤井教授が提唱する AI 出力統治手法 VRAIO を軸とする一連の研究の中核をなすものです。関連して、以下の論文も最近、相次いで採択・出版されました。

- **プラットフォーム AI の統治（YouTube を題材）：** 論文[2]は、YouTube 等の推薦・広告を主たる題材に、過激化に至る経路、エンゲージメント最大化のループ、未成年者を狙った広告など、個別の出力にとどまらない**累積的・連鎖的な出力パターン**の統治にも VRAIO を適用できることを示しています。GDPR・EU AI 法・デジタルサービス法（DSA）が宣言する義務

を、出力レベルでのリアルタイムな実効性へと翻訳する「接続基盤」としての役割も論じています。

- **高感度な公共安全応用への展望：** 論文[3]（およびその基となった論文[4]）は、スマートフォン網を社会的な安全基盤として活用し、重大犯罪やテロの予兆を早期に検知・通報する構想を、**プライバシー保護を前提として**検討したものです。常時盗聴のような重大なプライバシー侵害のリスクを正面から論じたうえで、本論文[1]が示す出力統治の基盤が確立されてはじめて、こうした高感度な応用も検証可能で民主的な統制の下で現実味を帯びる――という、提案的な位置づけにあります。

これら[1]～[4]は、いずれも藤井教授が提案する AI 出力統治手法 VRAIO に関する一連の研究です。

掲載先

雑誌名：*AI*（MDPI）

公開日：2026 年 6 月 14 日（オンライン公開済み）

題　名：No Trust Without Trust Infrastructure: The Extended Kelvin Principle and Its Application to AI Output Governance

著者名：Yusaku Fujii（藤井 雄作）

掲載巻号：Vol. 7, No. 6, Article 218 (2026)

ＵＲＬ：https://www.mdpi.com/2673-2688/7/6/218

ＤＯＩ：10.3390/ai7060218

【関連論文】

- [2] 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
- [3] Y. Fujii, “Governing Continuous Smartphone Sensing: A Privacy-Preserving Path Toward Democratic Homeland Security”, ***J. Homel. Secur. Emerg. Manag.***, 2026. https://doi.org/10.1515/jhsem-2026-0014

・ [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://www.mdpi.com/2076-3417/16/2/1032

【本件に関するお問合せ先】

〈研究に関すること〉

群馬大学　大学院理工学府知能機械創生部門　教授　藤井　雄作（フジイ　ユウサク）

E-MAIL：fujii@gunma-u.ac.jp

〈取材についてのお問合せ〉

群馬大学　理工学部庶務係（広報）

TEL：0277-30-1895

E-MAIL：rikou-pr@ml.gunma-u.ac.jp