

Reconstructing Ethical Principles and Sociotechnical Infrastructure for the Defense of Democracy in the Age of AGI - Disclosure of an Awarded Research Plan -

Yusaku Fujii^{1, a, *}

¹Graduate School of Science and Technology, Gunma University, 1-5-1 Tenjin-cho, Kiryu, Gunma, Japan

* Corresponding author

^a fujii@gunma-u.ac.jp

Keywords: AGI, human dignity, democracy, privacy, VRAIO, AI output governance

Abstract. This article discloses the objectives, methods, and budget plan of the research project “Reconstructing Ethical Principles and Sociotechnical Infrastructure for the Defense of Democracy in the Age of AGI” (Grant No. 26K21972), which was awarded funding under the FY2026 Grants-in-Aid for Scientific Research, Challenging Research (Exploratory). It is an expanded version of a Japanese explanatory article published in the Japanese Journal of Social Safety and Privacy (JpnJSSP), prepared with a view to publication in English in the Journal of Technology and Social Science (JTSS), and supplemented with concise summaries of five related papers. The project seeks ways to secure “human dignity,” a precondition for democracy, and “privacy protection,” one of democracy’s necessary conditions, in the age of artificial general intelligence (AGI). Task 1 addresses the ethical redefinition of human dignity against the background of the loss of the economic value of human labor, concentration of wealth, and widening inequality caused by the convergence of AGI and robotics. It draws on ancient Rome’s “bread and circuses” as a form of civic privilege in a labor-free society and on contemporary universal basic income (UBI) experiments. Task 2 proposes institutional design guidelines for a technical and institutional foundation that protects privacy by extending VRAIO (Verifiable Record of AI Output) to AI systems as a whole, particularly as a safeguard against misuse and abuse by insiders. Human dignity provides the ethical basis for privacy protection, while VRAIO provides a technical and institutional basis for protecting human dignity. The project will run for three years, from FY2026 to FY2028, with total direct costs of JPY 5 million. This article is intended to serve as a common reference point through which project participants can share the research content and budget plan and assess specific budget proposals.

1. Introduction

It remains uncertain when and in what form artificial general intelligence (AGI) will emerge. However, if artificial systems with intellectual capabilities equal to or greater than those of humans come to replace a broad range of human cognitive and physical labor, the consequences will extend far beyond industrial structure and employment. They may transform the modern assumption that human beings are the agents who create wealth, the concept of dignity historically linked to labor and social contribution, the motivation for political participation, and even the image of citizenship on which democracy rests.

At the same time, AI is increasingly being integrated into public-space surveillance, information recommendation, administrative decision-making, healthcare, education, transportation, security, and everyday sensing. The social impact of AI is realized not only through what it computes internally, but also through what information, judgment, recommendation, or control command it ultimately outputs, to whom, and under what conditions. AI ethics must therefore be connected to institutions that allow

society to verify the rules under which each output was produced and the procedures through which it was provided, rather than remaining at the level of abstract declarations of principle.

This article is based on the Japanese research-project article published in JpnJSSP [1], but it is not intended as a simple republication or literal translation. It has been substantially expanded with a view to publication as an international review article. It systematically reviews two foundational papers and subsequent developments, situates the funded research within a coherent line of scholarship, and enables research collaborators, those consulted on budget planning, reviewers, and readers to understand the project's background, objectives, methods, budget, and anticipated outcomes from this article alone.

The project has two central tasks. The first is to explore ethical principles capable of establishing “human dignity without labor contribution” as a precondition for democracy in the age of AGI. The second is to explore a sociotechnical infrastructure based on VRAIO and GLO that protects privacy and makes AI outputs verifiably governable as a necessary condition for safeguarding democracy. The first task asks what foundation is required to recognize human beings—the subjects of democracy—as citizens with dignity. The second asks what institutional and technical conditions are required to prevent those citizens from being dominated by AI-based surveillance, sorting, manipulation, or arbitrary outputs. The project begins by clarifying these two indispensable conditions as far as possible and examining how they may be connected.

2. Overview of the Research Project

Table 1. Overview of the research project.

Item	Description
Funding scheme	JSPS KAKENHI, Challenging Research (Exploratory)
Grant number	26K21972
Project title	Reconstructing Ethical Principles and Sociotechnical Infrastructure for the Defense of Democracy in the Age of AGI
Project period	FY2026–FY2028
Research structure	Single principal investigator
Direct costs	JPY 5 million (full amount requested and awarded)
Two pillars	Ethical redefinition of human dignity / AI output governance centered on VRAIO
Integration objective	Translate ethical principles into rules and make their application verifiable at the AI output stage

This project is not merely a summary of previous research. In accordance with the logic of the grant application, it is organized around two tasks. The first is to explore the establishment of human dignity as a precondition for democracy in the age of AGI. The second is to explore a sociotechnical system based on VRAIO and GLO as a necessary condition for protecting privacy and governing AI outputs in a verifiable manner. Without the former, human beings may lose the basis on which they remain the subjects of democracy. Without the latter, even if dignity and democracy are affirmed in principle, society cannot verify whether specific AI outputs violate them. Among the vast range of issues involved, this exploratory project focuses on two conditions regarded as indispensable and seeks to make them concrete wherever possible.

3. Task 1: Loss of the Human Role as the Source of Wealth and the Reconstruction of Dignity

3.1 The Historical Link among Labor, Wealth, and Dignity

In modern society, labor has been associated not only with earning an income but also with social contribution, recognition, self-realization, responsibility, and civic standing. Under this structure, the ability to work and contribute to society through work has constituted an important basis of human dignity. If AGI and robotics broadly replace human activity in production, management, research, education, healthcare, public administration, and other domains, human beings may cease to be the principal agents of wealth creation.

What may be lost in such a transition is not employment alone. The sense of being needed by society, the pathway through which individuals participate in a community by exercising their abilities, and the motivation to assume responsibility for political judgment may all be weakened. Even if universal basic income (UBI) is introduced as a form of livelihood security, income protection does not automatically secure a social role, self-determination, recognition, or political agency.

3.2 Insights from “Bread and Circuses”

Fujii [2] reread ancient Rome’s “bread and circuses” from a perspective different from the conventional interpretation that treats them solely as symbols of civic decay or political appeasement. The central question is why grain and entertainment continued to be provided over a long period to large numbers of Roman citizens who could hardly be said to contribute directly to society through production, military service, public administration, or other functions. Explaining the policy merely as a means of preventing unrest may fail to capture the significance of the fact that these people continued to be treated as Roman citizens and remained entitled to livelihood support and public consideration.

This reinterpretation is also consistent with major findings in ancient history. Veyne [3] comprehensively analyzed public benefaction (euergetism) in the ancient Mediterranean world and showed that it cannot be reduced to instrumental motives such as buying popular support or concealing exploitation. Rather, such benefaction expressed the obligations and privileges of urban notables and emperors toward the civic community. Thus, the popular view of “bread and circuses” as merely a device of appeasement and depoliticization is also historically one-sided.

Fujii’s argument identifies the possibility that, behind this practice, being a Roman citizen itself carried a status and value independent of economic usefulness or labor contribution—in other words, that a certain degree of civic dignity had been established. This is not an attempt to idealize ancient Rome. Rather, it seeks to draw a lesson for the age of AGI from the fact that even people who were not socially useful could be regarded as entitled to livelihood protection and social consideration simply because they were citizens. Even if AGI produces most wealth and human labor contribution becomes unnecessary, could a principle be established under which human beings are treated not as surplus population or dependents, but as citizens possessing dignity in themselves? The first objective of this project is precisely to explore this possibility and clarify the ethical and institutional foundations of human dignity independent of labor capacity and economic usefulness.

4. Task 2: Privacy Protection as a Necessary Condition and AI Output Governance through VRAIO and GLO

4.1 Why Govern AI “Outputs” Rather than Its “Interior”?

AI governance has emphasized training data, model architecture, safety constraints, explainability, and alignment [4]. Yet complete disclosure of the internal workings of complex AI systems is difficult both technically and from the perspectives of intellectual property and security. Even when internal processes can be explained to some extent, retrospective verification and attribution of responsibility remain impossible unless society records which outputs were actually provided and under which rules they were authorized.

At the institutional level, existing AI regulations, including the EU AI Act and GDPR and the U.S. AI Risk Management Framework, primarily rely on documentation requirements such as fundamental rights impact assessments (FRIAs) and data protection impact assessments (DPIAs), together with ex post audits and sanctions. However, these frameworks are not sufficient to address misuse and abuse by system owners and operators—that is, insiders—which is more difficult to control than attacks by external adversaries. This project focuses precisely on such insider threats. VRAIO differs qualitatively from existing regulatory approaches because it provides preventive control at the output boundary through real-time rule checking and tamper-resistant records, complementing ex post audits and sanctions.

The VRAIO framework proposed by Fujii [5] places the governance mechanism at the social boundary of an AI system—that is, on its output side. A Recorder preserves, in a tamper-resistant form, the output request, applicable rules, authorization conditions, candidate output, actual output provided, time, and relevant parties, thereby enabling verification by citizens or independent auditors. This approach does not require full disclosure of the AI's internal processes; instead, it focuses on the outputs that act upon society and the procedures through which they are released.

4.2 Prototype for Public-Space Surveillance

The original VRAIO model was presented for public spaces monitored by AI-connected cameras [5]. Images from public spaces can be useful for crime prevention, accident detection, and the search for missing persons, but unrestricted external disclosure can cause serious privacy violations. When the police, a municipality, or another entity requests an image or analytical result, an egress firewall checks the output request against the authorization conditions, while the Recorder logs the entire procedure. False output requests are deterred by combining substantial penalties, rewards for reporting violations, and compensation for victims, thereby eliminating the expected benefit of noncompliance. In the grant application, the third-party Recorder was described in simplified form as both reviewing and recording outputs. In this article, however, mechanical rule checking is assigned to the egress firewall, while the Recorder is dedicated to maintaining a tamper-resistant record of the entire procedure.

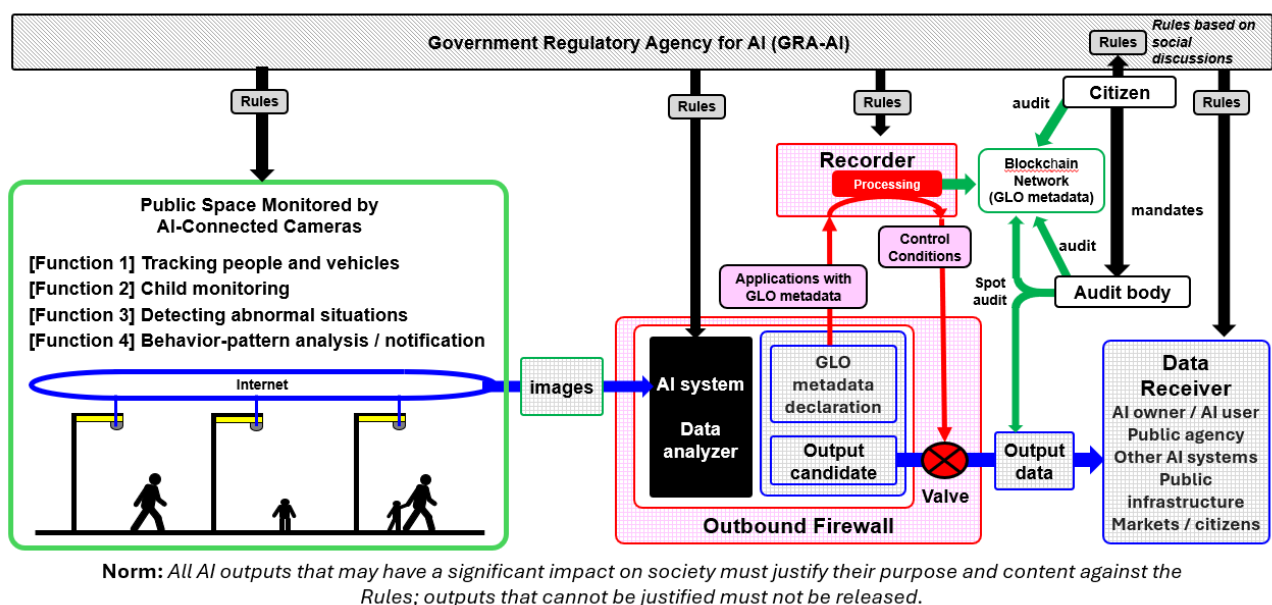


Fig. 1. Basic architecture of an AI-connected-camera public-space monitoring system implementing VRAIO (redrawn from Ref. [5]).

The significance of this architecture lies in avoiding the false choice between banning surveillance technology altogether and accepting it unconditionally. Useful functions can be retained while external outputs are made subject to rules governing purpose, authority, subject, scope, and retention period, and compliance can be verified retrospectively. VRAIO is not a single privacy-protection technology; it is a sociotechnical system linking rule formation, machine-executable checking, recording, auditing, and mechanisms for objection.

5. Development of VRAIO Research: From Individual Applications to Society-Scale Trust Infrastructure

5.1 Application to Autonomous Driving

Autonomous vehicles continuously collect information about their surroundings, pedestrians, occupants, location, and driving history and communicate with external systems. Fujii [6] applied VRAIO to autonomous driving and presented a framework for governing the external disclosure of images, recognition results, location data, and decision information generated by vehicles. This work extended VRAIO from fixed public-space surveillance to autonomous systems that acquire and process large volumes of personal information while moving.

The autonomous-driving application showed that AI output governance concerns not only privacy but also social acceptance, accountability after accidents, requests from law-enforcement agencies, data sharing among service providers, and citizen oversight. The governance question is not merely whether the AI can drive accurately, but also what the vehicle outputs to society and who is permitted to use those outputs.

5.2 Application to Continuous Smartphone Sensing

Using audio, acceleration, location, ambient-environment, and other signals, smartphones can serve as social-safety sensors for detecting violence, accidents, sudden illness, disasters, and other emergencies. At the same time, continuous sensing carries the risk of turning into permanent surveillance by governments or corporations. Building on the proposal for a smartphone-based emergency-detection network [10], Fujii [7] outlined a path for connecting continuous smartphone sensing to democratic homeland security by making on-device processing the default and governing only externally transmitted alerts, features, location information, and evidentiary data through VRAIO.

This application goes beyond the view that privacy protection consists only in refraining from data collection. It clarifies an approach in which sensing needed for public safety is performed while the information transmitted externally is limited, recorded, and audited. The key to reconciling the utility of sensing with individual freedom is to distinguish on-device processing from external output and to govern the output boundary.

5.3 Society-Scale Extension to AI Platforms

Generative AI and large-scale AI platforms continuously provide text, images, recommendations, search results, evaluations, and decision support to vast numbers of users. Fujii [8] positioned VRAIO as a “safety valve” for society-scale AI platforms and proposed governing each platform output by linking it to applicable rules, intended purpose, target population, risk category, and authorization conditions.

Here, VRAIO expands from a mechanism attached to a particular camera, vehicle, or smartphone into an infrastructure supporting society’s entire information environment. The crucial point is to institutionalize rule declaration, machine-verifiable checking, tamper-resistant records, independent auditing, and citizen objections rather than relying solely on the voluntary goodwill of AI companies.

5.4 Designing the Conditions for Trustworthiness Rather than Demanding “Trust”: VRAIO and GLO

Based on the Extended Kelvin Principle, Fujii [9] extended the idea that what cannot be measured cannot be properly controlled to the governance of AI outputs. Legitimate trust cannot be created merely by asking society to trust AI. What can be designed are the objective conditions under which compliance with rules can be verified—namely, trustworthiness—and a trust infrastructure through which those conditions can be repeatedly confirmed. In this framework, VRAIO provides a sociotechnical infrastructure that verifiably links the output request, applicable rules, authorization decision, actual output, and record. GLO, by contrast, provides guidance for expressing in a common form the rules, grounds, constraints, scope of application, uncertainty, and degree of legitimacy associated with each candidate output. Whereas VRAIO is a mechanism for verification and recording, GLO functions as a common language for describing, checking, and auditing legitimacy. GLO was not included in the original KAKENHI application; it was formulated by Fujii [9] in the subsequent development of the research.

Through this generalization, VRAIO and GLO can be understood as trust infrastructure supporting the legitimacy, legality, safety, fairness, and accountability of AI outputs, rather than merely as a record-keeping method for privacy protection. Nevertheless, building such a mechanism does not automatically make AI ethical or safeguard democracy. Human society must continue deciding what counts as legitimate, which rules should take priority, and which decisions must remain reserved for humans. This project therefore explores VRAIO and GLO as one necessary condition for securing privacy and democratic control.

The background to this project also includes sociotechnical research aimed at reconciling public technologies with individual freedom. In work on public-use powered air-purifying respirators (PAPRs), the author has examined their feasibility as an alternative to lockdowns [11], social requirements [12], respiratory-state estimation [13], estimation of respiratory flow and mask leakage using a single differential-pressure sensor [14], and time-managed use as a means of reconciling infection control with personal freedom [15]. These studies share with the present project an approach that treats technical performance, operation, public acceptance, freedom, and social governance as an integrated whole.

The AI Special Issue “Governing Trustworthy AI Outputs in a Sensor-Dense Society: Privacy, Auditability and Responsible Deployment” [16] will also serve as a forum for international discussion and dissemination of findings related to this project.

6. Relationship between the Two Tasks: A Precondition and a Necessary Condition

The establishment of human dignity explored in Task 1 is a precondition for human beings to remain the subjects of democracy in the age of AGI. If, when people lose labor capacity or economic usefulness, they also lose the basis on which they are treated as citizens with dignity, democracy loses its subjects. Privacy protection and AI output governance through VRAIO and GLO, explored in Task 2, are necessary conditions for maintaining democracy within an actual AI society. Even if human dignity is established as a principle, it lacks substance unless citizens can be protected from AI-based surveillance, sorting, manipulation, and arbitrary outputs.

The two tasks are not simply merged into a single system. The first investigates the ethical and political basis for why and in what sense human beings should be treated as beings with dignity. The second investigates a sociotechnical system that protects those human beings from AI systems by safeguarding privacy, expressing the legitimacy of AI outputs through GLO, and enabling rule checking, recording, and auditing through VRAIO. At their point of connection, the project examines how principles concerning dignity, freedom, self-determination, political participation, and privacy can be translated into rules applicable to AI outputs and how application of those rules can be verified.

Meeting the precondition and the necessary condition would not, however, be sufficient to guarantee democracy in the age of AGI. Many issues would remain, including concentration of power, distribution of wealth, education, political participation, security, international competition, and control of AGI itself. The two tasks addressed here are not a universal framework capable of solving all of these problems. Nevertheless, sustaining democracy would be extremely difficult in a society where human dignity is not established and privacy and AI output governance are not secured. The project therefore does not pretend to be complete; instead, it directly addresses two important but unresolved conditions and seeks to advance them step by step through both theory and sociotechnical design.

7. Research Objectives and Methods

7.1 Research Objectives

The first objective is to explore the possibility of establishing human dignity as a precondition for democracy in a society where AGI broadly replaces human labor. Taking as a clue the possibility that the continued provision of “bread and circuses” in ancient Rome rested on civic dignity independent of labor contribution, the project examines the ethical and institutional foundations of dignity that do not depend on labor capacity, earning capacity, or economic usefulness. The second objective is to design, as a necessary condition for protecting democracy, a sociotechnical system that safeguards privacy against AI systems, expresses the legitimacy of AI outputs through GLO, and enables rule checking, recording, and auditing through VRAIO. Together, the two objectives seek to establish foundations indispensable to the defense of democracy.

7.2 Examination of Ethical Principles

Following the methods set out in the grant application, the project combines thought experiments (theoretical reasoning), reviews of historical and ethical literature (including Roman history and UBI research), case studies (including Finland’s UBI experiment), and discussion at international conferences. The specific approaches are: (1) internalization of values through culture and narrative, such as constructing social narratives that affirm dignity; (2) establishment of practices, such as legal norms that respect dignity; (3) guarantees of social participation, such as opportunities for dialogue among diverse stakeholders; and (4) AI-adapted ethics education, such as value-education curricula for an AI society. These approaches will be combined in the design of social institutions, and the ethical basis of human dignity independent of labor will be modeled through a three-layer structure of cultural narrative, social institutions, and education systems. The ultimate aim is to present a new dignity model supporting democracy as an internationally shareable guideline.

The thought experiments will consider the following settings: (1) a society in which human labor has been almost entirely replaced; (2) a society in which livelihoods are secured by UBI or similar measures but social roles are lost; (3) a society in which AGI makes policy judgments more accurately than humans; (4) a society in which citizens excessively delegate decision-making to AGI; and (5) a society in which AI intervenes in preference formation through the information environment. For each setting, the project will examine changes in dignity, self-determination, political participation, objection, education, culture, care, and responsibility toward the community.

7.3 Institutional Design Based on VRAIO

Using public-space surveillance, generative AI and social media, administrative AI, continuous sensing, and autonomous driving as illustrative domains, the project will examine who defines the rules, which outputs should be recorded, what information should be disclosed to citizens and auditors,

how the independence of the Recorder should be secured, and how objection and remedy should be institutionalized. It will also examine how GLO can express the rules, grounds, constraints, scope of application, uncertainty, validity period, and other elements constituting the legitimacy of a candidate output and how this information can be connected to VRAIO-based recording and verification. Evaluation metrics will include legality, reduction of privacy violations, auditability, accountability, rule-compliance rate, responsiveness to objections, and institutional operating costs.

7.4 Integrated Evaluation

Ultimately, the project will identify which elements of the human-dignity principles examined in Task 1 can be made concrete as rules for AI output governance and will evaluate whether they can be consistently connected to the expression of legitimacy through GLO, output requests, authorization conditions, actual outputs, VRAIO records, and independent auditing. This connection does not mean reducing human dignity as a whole to machine-executable rules. As exploratory research, the project will present foundational guidelines and an evaluation framework that can lead to future empirical studies, standardization, legal institutionalization, international collaboration, and further work on the remaining issues.

8. Three-Year Research Plan and Budget

Table 2. Three-year research plan and budget (direct costs).

Fiscal year	Main research activities	Direct costs
FY2026	Organize issues concerning human dignity in the age of AGI; establish scenarios; systematize existing VRAIO research; and clarify target domains and the basic architecture.	JPY 1.3 million
FY2027	Systematize candidate ethical principles; design institutions for AI output governance; and examine case-specific procedures for rules, records, audits, and objections.	JPY 1.4 million
FY2028	Integrate ethical principles and sociotechnical infrastructure; present an evaluation framework, common specifications, institutional design guidelines, and future research issues.	JPY 2.3 million

The total direct cost is JPY 5 million. The initial allocation consists of JPY 1.2 million for travel, JPY 0.3 million for personnel and honoraria, and JPY 3.5 million for other expenses. Funds will be used for workshops, exchanges with domestic and international researchers, conference presentations, collection of materials, publication, AI-related services and computing environments, surveys, and institutional design. This allocation is the publicly disclosed baseline plan at the time of the award and may be reasonably modified in light of research progress, external conditions, and the flexibility permitted under the KAKENHI system.

The significance of using this article as a common reference point for research planning and budget execution does not lie in mechanically matching every expenditure to a particular phrase in the text. Rather, it lies in making it possible to explain how each expenditure, research collaboration, or outsourcing arrangement contributes to a research objective, annual plan, and anticipated outcome, thereby sharing the project's purposes among the principal investigator, collaborators, the host institution, and society.

9. Anticipated Outcomes and Academic Significance

First, the project does not treat ancient Rome's "bread and circuses" only as a negative metaphor. It reexamines why livelihood support and entertainment continued to be provided even to citizens who did not directly contribute to society and uses the possible existence of "dignity inherent in being a citizen" as a clue for exploring the establishment of human dignity in the age of AGI. This is an attempt to identify a precondition under which human beings, as the subjects of democracy, can be recognized as citizens with dignity independently of labor contribution or economic usefulness.

Second, the project makes privacy protection against AI systems and AI output governance concrete as necessary conditions for safeguarding democracy. GLO expresses the legitimacy of outputs in a common format, while VRAIO makes it possible to verify and audit the outputs provided externally, the applicable rules, authorization decisions, and records. Society can thereby confirm not only that ethical principles have been declared, but also how they were applied to individual AI outputs.

Third, the project clarifies the relationship between Task 1 as a precondition and Task 2 as a necessary condition. Without human dignity, democracy loses its subjects; without privacy protection and AI output governance, those subjects cannot be protected from AI-based surveillance and manipulation. The academic significance of the project lies not in offering a universal solution, but in extracting two particularly fundamental conditions from a vast and unfinished problem and beginning their concrete exploration on the basis of prior research.

Fourth, publication of the awarded research plan, budget, and evaluation policy at the beginning of the project will make it possible to compare subsequent outcomes with the original plan. This provides a baseline for increasing research transparency and for clearly identifying the reasons for changes, achievements, unfulfilled objectives, and newly emerging issues.

10. Research Considerations

Because the project concerns AGI, whose timing and form of realization remain uncertain, future projections involve inherent uncertainty. Rather than asserting a single future scenario, the project will explicitly state multiple scenarios and assumptions and examine which principles or institutions are required under which conditions.

There are also value conflicts among human dignity, democracy, freedom, and safety that cannot be reduced to a single optimum. VRAIO-based recording and auditing may themselves be misused for excessive surveillance or control. The institutional core must therefore include limitation of the recorded content, prohibition of secondary use, retention periods, access rights, independence of auditors, citizens' right to object, and auditability of the Recorder itself.

When the English version of this article is published in JTSS as a review article, it will explicitly state in the Abstract and Introduction that it is an expanded version based on the JpnJSSP article [1] and will cite the original Japanese paper. Although both journals are published by the same nonprofit organization, the origin of the article and the added content will be disclosed transparently to readers. The JTSS version will clearly distinguish itself from simple duplicate publication by adding value as a review of prior research and an international presentation of the research program.

11. Conclusion

Defending democracy in the age of AGI cannot be achieved solely through technical control of AGI. First, principles must be established under which human beings continue to be treated as citizens with dignity even when they are no longer the principal agents of wealth creation and lose their direct usefulness to society. This is a precondition for democracy. The fact that "bread and circuses" continued to be provided to large numbers of citizens who were not directly useful to ancient Roman society, together with the possibility that civic dignity lay behind this practice, provides an important clue for addressing this task.

Second, those citizens with dignity must be protected from privacy violations, surveillance, sorting, manipulation, and arbitrary outputs by AI systems. This is a necessary condition for maintaining democracy in practice. The project explores a sociotechnical system in which GLO expresses the legitimacy of AI outputs and VRAIO makes the output request, rule checking, authorization decision, actual output, record, and audit verifiable. Whereas Task 1 asks for the foundation that supports human beings as the subjects of democracy, Task 2 asks for the mechanisms that protect those subjects within an AI society.

Even achievement of these two tasks would not complete the defense of democracy in the age of AGI. Many major issues would remain, including concentration of wealth and power, education, political participation, security, international institutions, and control of AGI itself. This project does not claim that all of these issues can be solved at once. Nevertheless, the establishment of human dignity and AI output governance through VRAIO and GLO, including privacy protection, are conditions that cannot be avoided. As a publicly disclosed baseline at the beginning of this three-year project, this article expresses a commitment to explore these two conditions as deeply as possible and to carry the resulting insights forward to the next set of challenges.

Acknowledgements

This work is supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Challenging Research (Exploratory), “Reconstructing Ethical Principles and Sociotechnical Infrastructure for the Defense of Democracy in the Age of AGI” (Grant No. 26K21972, FY2026–FY2028).

Disclosure Regarding the Prior Japanese Version

This article is based on a Japanese research-project article published in the Japanese Journal of Social Safety and Privacy [1]. The JTSS version is a substantially expanded article that adds a review of prior research, the developmental history of VRAIO, integration of the two lines of research, the role of the budget plan, and an explanation of the project as an international research program.

Conflict of Interest

The author is the Editor-in-Chief of both JTSS and JpnJSSP. For submission to JTSS, the author will not participate in reviewer selection, the peer-review process, or the editorial decision. The process will be managed by an independent handling editor.

References

- [1] Y. Fujii, “Reconstructing Ethical Principles and Sociotechnical Infrastructure for the Defense of Democracy in the Age of AGI: First Report,” *Japanese Journal of Social Safety and Privacy*, Vol. 10, No. 1, pp. 1–6, 2026. (in Japanese).
- [2] 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>
- [3] P. Veyne, *Le Pain et le Cirque: Sociologie Historique d’un Pluralisme Politique*, 1976; English translation: *Bread and Circuses: Historical Sociology and Political Pluralism*, trans. B. Pearce, Allen Lane, London, 1990.

- [4] Y. Bengio, G. Hinton, A. Yao, D. Song, P. Abbeel, T. Darrell, et al., “Managing Extreme AI Risks amid Rapid Progress,” *Science*, Vol. 384, No. 6698, pp. 842–845, 2024.
<https://doi.org/10.1126/science.adn0117>
- [5] 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>
- [6] 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>
- [7] 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>
- [8] 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>
- [9] 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>
- [10] 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>
- [11] 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>
- [12] 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>
- [13] 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>
- [14] 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>
- [15] 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>

[16] *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

Author Biography

Yusaku Fujii is a Professor at the Graduate School of Science and Technology, Gunma University, Japan, and holds a Ph.D. degree. His research interests include AI governance, the defense of democracy, social safety engineering, privacy protection, and precision measurement. He is President of the NPO(Non-Profit Organization) e-JIKEI Network Promotion Institute and Editor-in-Chief of the Journal of Technology and Social Science, the Journal of Mechanical and Electrical Intelligent System, and the Japanese Journal of Social Safety and Privacy.