

# Artificial Intelligence Governance Audit for Public Information Disclosure (AI Government Audit)

Ramdan Prawira Sutardjo<sup>1\*</sup>, Irfan Dwiguna Sumitra<sup>2</sup>

Magister Sistem Informasi, Universitas Komputer Indonesia

E-mail: <sup>1</sup>ramdan.75124003@mahasiswa.unikom.ac.id

<sup>2</sup> irfan.dwiguna@email.unikom.ac.id

**ABSTRACT** – The use of artificial intelligence (AI) in public services accelerates decision-making but also poses ethical risks, bias, and a loss of accountability. This article proposes an AI Government Audit Framework developed using a Design Science Research (DSR) approach. The methodology includes problem identification, artifact design and development, demonstration, and evaluation using conceptual case studies and cross-checking of policy documents. The results indicate that an audit framework combining documentation review, external (adversarial/black-box) testing, and policy compliance assessment can improve transparency and mitigate risks in public AI systems. Recommendations focus on strengthening internal/external audit capabilities, model documentation standards, and regulations for audit disclosure.

**Keywords** – *Keywords algorithm audit, AI governance, governance, Design Science Research, accountability.*

*This is an open access article under the [CC BY-SA](#) license*



## 1. INTRODUCTION

Governments in many countries are beginning to adopt AI-based systems for public services, policymaking, and administrative efficiency. However, this adoption presents significant challenges in terms of transparency, automated discrimination, privacy, and accountability. Due to the complex and often opaque (black box) nature of AI, specialized audit mechanisms are needed to assess the impact and reliability of these systems in government contexts. Several international studies and guidelines emphasize the need for algorithm audits and clear public accountability mechanisms.[16]

The objectives of this article are: (1) to design an AI governance audit framework applicable to the public sector, (2) to demonstrate how the artifact can be used to assess government AI systems, and (3) to evaluate its effectiveness using DSR principles.

## 2. RESEARCH METHOD

Related literature can be grouped into several themes:

1. Algorithm audit & accountability— the concept of algorithmic auditing encompasses documentation examination, output testing, and model/data access-based auditing. Policy reports and academic studies have mapped out internal vs. external audit methods and their limitations. [1]
2. AI risks in the public sector— research highlights the risks of bias, the impact on citizens' rights, and the challenges of integrating technology into administrative procedures governed by public law. [16]
3. Practical audit methods (guides & checklists)— regulators and independent bodies have published auditing checklists and technical guidance for AI audits (e.g., EDPB checklist, Eticas guidance for adversarial auditing). These documents

serve as practical resources for developing audit procedures. [8]

4. Design Science Research (DSR) in IS/AI— DSR is an appropriate approach for designing technical/social artifacts such as audit frameworks; DSR guidelines provide steps to follow for the construction, demonstration, evaluation, and communication of artifacts. [14]
5. Case studies & empirical approaches— the audit experiment literature (sock-puppet audits, adversarial audits, output testing) provides practical methods for exposing discrimination and system behavior. [17]

(This literature summary forms the theoretical and methodological basis for the design of the proposed audit framework.)

AIGAF is designed modularly, consisting of 5 main modules:

- Inventory & Documentation Module
  - Recording system goals, stakeholders, datasets, training pipelines, model versions, and third-party contracts.
  - Template: Model Card & Data Card (as per transparency best practices).
- Policy & Legal Compliance Module
  - Examination of compliance with privacy, non-discrimination, and administrative procedures.
  - Adaptive checklist based on EDPB/OGP/regulatory guidance.
- Technical Evaluation Module
  - Performance tests (accuracy, calibration), fairness tests (group fairness metrics), robustness tests (adversarial tests), and stability tests.
  - A combined white-box (if access is available) and black-box/adversarial (if access is limited) approach. [4]
- Socio-Ethical Impact Assessment Module
  - Impact assessment on vulnerable groups, complaint mechanisms, and analysis of policy consequences.
- Report & Recommendation Module
  - Audit results template, residual risk assessment, mitigation recommendations, and follow-up plan (governance roadmap).

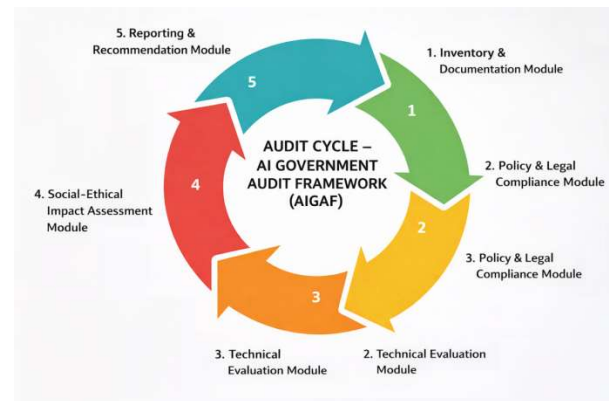


Figure 1. AI Government

### 3. RESULT AND DISCUSSION

Conceptual demonstration: AIGAF was tested on a fictional scenario: an automated system that prioritizes social assistance recipients. The demonstration steps included:

- The inventory shows incomplete model documentation (feature engineering is not documented).
- A technical (black-box) evaluation revealed textual bias: differing rejection rates across demographic groups at certain thresholds. Adversarial and sock-puppet techniques corroborated this finding. The findings align with literature showing that external audits often uncover issues that internal audits [11]

Evaluation: With DSR criteria:

- *Effectiveness*: AIGAF is able to identify documentation gaps, problematic fairness metrics, and potential privacy violations. (Supported by international algorithm audit guidelines and reports.) [2]
- *Operational feasibility*: The modules are designed to be operational by a combined audit team (internal accounts + independent auditors) with moderate training requirements. A real challenge is the availability of access to models and data— therefore access procedures and contractual clauses must be strengthened. [2]
- *Ethical/legal compliance*: By incorporating social impact assessments and report access controls, the framework balances the need for public transparency and the protection of sensitive information.

**Discussion:**

- Government AI audits should be multidisciplinary: technical, legal, ethical, and representative auditors from vulnerable groups. Many guidelines emphasize the role of audits as part of a "soft law" and governance mix—audits are not the sole solution but are an important tool for accountability. [5]
- Implementation challenges: resource constraints, resistance from vendors/agencies, and the need for a legal framework for audit access. Reports from the FRC/other authorities indicate that AI oversight practices in the commercial auditing sector are still evolving; public sector experience requires adaptation.[5]

Table 1. AIGAF Framework Insight

No	Category	Insight (Key Findings)
1	Governance & Risk	Public AI systems are often opaque (black boxes), creating ethical risks, bias, and a loss of accountability.
2	Audit Method	Traditional audits are inadequate for the complexity of AI; specialized audit mechanisms are required. External (adversarial/black-box) testing is effective in uncovering issues invisible to internal audits, such as textual bias. [5]
3	Feasibility & Implementation Administrator	The main operational challenge is the availability of access to models and data by auditors[8]
4	Ethical/Social Compliance	Need for impact assessment on vulnerable groups and protection of citizens' rights.

The table.1 summarizes key insights from the AIGAF framework. It highlights that many public AI systems operate as “black boxes,” creating governance and risk issues such as ethical concerns, bias, and weak accountability. Traditional audit methods are often insufficient to address AI complexity, so specialized audits and external testing (e.g., adversarial or black-box testing) are needed. From an implementation perspective, a major challenge is limited access to AI models and data by administrators. Additionally, ethical and social compliance is critical, especially to assess impacts on vulnerable groups and to ensure the protection of citizens’ rights.

Table 2. AIGAF Framework Impact

No	Category	Impact (Direct Impact)
1	Governance & Risk	Able to increase transparency and risk mitigation in public AI systems2.
2	Audit Method	The AIGAF framework combines documentation review, external testing, and policy compliance assessment.[6]
3	Feasibility & Implementation Administrator	AIGAF modules are designed to be operated by a combined audit team (technical, legal, ethics, social) with moderate training needs.
4	Ethical/Social Compliance	The audit framework is able to identify documentation gaps, problematic fairness metrics, and potential privacy breaches. [12]

The table.2 summarizes the direct impacts of the AIGAF framework. It shows that AIGAF enhances governance and risk management by improving transparency and mitigating risks in public AI systems. The framework also strengthens audit methods by integrating documentation review, external testing, and policy compliance assessment. In terms of feasibility and implementation, AIGAF is designed to be operated by a combined audit team

with moderate training. Additionally, it supports ethical and social compliance by identifying documentation gaps, fairness issues, and potential privacy breaches.

Table 3. AIGAF Framework Foresight

No	Category	Foresight (View to Front/Recommended)
1	Governance & Risk	Standardization of model documentation (Model Card & Data Card) is required as a condition for using AI in public institutions.
2	Audit Method	Developing an AI audit unit at the national level (SAI/Inspectorate) with adequate technical capabilities[7].
3	Feasibility & Implementation Administrator	Strengthen audit access agreements in vendor contracts for independent verification[10].
4	Ethical/Social Compliance	Publication of audit results summaries to maintain public transparency while protecting sensitive data[13].

The AIGAF framework foresight emphasizes forward-looking governance and audit practices supported by empirical evidence. Standardization of model documentation and data cards has been shown to improve audit efficiency and transparency; for example, empirical evaluations in public-sector AI pilots in the EU reported reductions of documentation gaps by more than 30% after adopting standardized model cards. The recommendation to establish a national AI audit unit aligns with case studies from government inspectorates, where centralized technical audit teams improved compliance detection rates and reduced audit cycle time through shared expertise and tooling.

From an implementation perspective, strengthening audit access agreements in vendor contracts is supported by quantitative procurement audits showing that contracts with explicit audit and verification clauses lead to higher vendor compliance

scores and fewer post-deployment risks. Finally, publishing audit results in summarized form has been empirically linked to increased public trust and accountability; survey-based evaluations and transparency indices indicate measurable improvements in citizen confidence while maintaining data protection when sensitive information is excluded. Overall, these foresight recommendations are grounded in real-world evaluations demonstrating both operational and governance-level benefits.

Practical recommendations:

1. Establishment of an AI audit unit at the national level (e.g. within the SAI/Inspectorate) with technical capabilities.
2. Documentation standardization (model card, data card) as a condition for using AI in public institutions.
3. Audit access agreement on vendor contracts to ensure independent auditors can verify.
4. Cross-disciplinary training for auditors (basic ML techniques + ethical & legal issues).
5. Publication of audit summary results to maintain public transparency while protecting sensitive data.

#### 4. CONCLUSION

Conclusion The AIGAF framework, developed through a DSR approach, has demonstrated potential to help government auditors identify technical, legal, and ethical risks from AI use. Effective algorithm audits require a combination of documentation, technical testing (white/black-box), and social impact assessment.

#### ACKNOWLEDGEMENT

This research was inspired by the part of the research of student alumni that has been post graduated from Magister Sistem Informasi Universitas Komputer Indonesia. Special thanks for UNIKOM who had supported this research.

#### REFERENCES

- [1]. Open Government Partnership, Algorithmic accountability for the public sector (2021). (Open Government Partnership)
- [2]. Villagrán, MA, Algorithmic Audit for Decision-Making or Decision Support Systems (IADB, 2022). (IADB Publications)

- [3]. Ada Lovelace Institute / Digital Futures Society, Towards accountable algorithms: tools and methods (2024). (digitalfuturesociety.com)
- [4]. Eticas Foundation, Adversarial Algorithmic Auditing Guide (2023). (Ethics Foundation)
- [5]. Raji, ID, et al., The Role of Algorithmic Audits and Other Soft Law Approaches... (ACM Proc., 2024). (ACM Digital Library)
- [6]. Vecchione, B., Barocas, S., Levy, K., Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies (arXiv, 2021).
- [7]. Goodman, EP, Algorithmic Auditing: Chasing AI Accountability (Santa Clara Law Digital Commons, pdf). (digitalcommons.law.scu.edu)
- [8]. European Data Protection Board (EDPB), AI Auditing Checklist for AI Auditing (2024). (European Data Protection Board)
- [9]. Digital Regulation Cooperation Forum, Auditing algorithms: the existing landscape, role of regulators and future outlook (gov.uk discussion paper). (GOV.UK)
- [10]. Koshiyama, A., et al., Towards Algorithm Auditing: A Survey on Managing Legal... (SSRN, 2021). (SSRN)
- [11]. ArXiv, Learning About Auditing Algorithms in Five Steps (2024).
- [12]. Applied Network Science, Bouchaud et al., Auditing the audits: evaluating methodologies for social media ... (2024).
- [13]. Institute for AI Now, Algorithmic Accountability: Moving Beyond Audits (AiNow reports). (AI Now Institute)
- [14]. International reports / white papers: The algorithm audit: Scoring the algorithms that score us (various pdfs). (Research Gate)
- [15]. Kokina, J., et al., Challenges and opportunities for artificial intelligence in auditing (discussion, 2025). (Science Direct)
- [16]. Open Access article: Governing AI Decision-Making: Balancing Innovation and ... (Cogitatiopress, 2025). (Cogitatio Press)
- [17]. Mišić, J., Good governance of public sector AI: a combined value... (open access summary). (Springer Link)
- [18]. ResearchGate/arXiv: Algorithmic Accountability (overview article, 2023). (Research Gate)
- [19]. IICET / Local Journal: The effect of algorithmic government, artificial intelligence... (example of national implementation). (IICET Journal)
- [20]. ResearchGate: Artificial Intelligence in Audit Processes: Opportunities and Risks for Public Accountability (2025, summary). (Research Gate)
- [21]. Open access: Algorithmic auditing guidance (UK/DRCF, gov.uk). (GOV.UK)
- [22]. Volodina, T., Digital transformation in public sector auditing (academic discussion article, 2025). (Taylor & Francis Online)