Responsible AI Governance: Transparency and Readiness in Federal AI Systems

Oai Tran[a], *Farhana* Faruqe[b]

[a]School of Data Science,University of Virginia, Charlottesville, Virginia, USA, dzn7nf@virginia.edu

[b]Faculty at School of Data Science,University of Virginia, Charlottesville, Virginia, USA, xrh6cm@virginia.edu

Publishing history: Submitted*: 31 January 2026, Revised*: , Accepted*: , Published*:

Abstract. Federal agencies are required to disclose their use cases for artificial intelligence (AI) through centralized inventories to support transparency and responsible digital government. While these inventories improve visibility, they are rarely used to assess governance readiness, institutional capacity, or oversight maturity, limiting their value for accountability, particularly in high-impact systems. This paper evaluates transparency in federal AI deployments across public service and digital government functions by constructing empirical indicators of system impact, development stage, disclosure practices, and core governance controls, including impact assessments, monitoring, independent evaluation, and appeals. Focusing on rights- and safety-impacting systems, we assess whether governance practices scale with deployment maturity. We found that high-impact AI systems are already operational, while governance controls often lag behind. Disclosure quality varies systematically with institutional capacity: agencies with greater technical staffing report more complete information, whereas vendor-dependent systems exhibit higher opacity. These findings reveal a structural governance gap in federal AI adoption.

Keywords. AI governance, Responsible AI, Transparency and oversight, Transparency and oversight

1. Introduction

Recent federal mandates require agencies to disclose their AI use cases as part of a broader push toward responsible and transparent AI adoption. Artificial intelligence has transitioned from experimental pilot projects to core infrastructure within digital government. Federal agencies increasingly deploy AI systems to support border security, benefits administration, workforce planning, fraud detection, document analysis, and operational decision-making[Executive Order 14110].

However, federal agencies are still in the early stages of disclosing how AI systems are designed, deployed, and governed. The five ethical principles, beneficence, non-maleficence, autonomy, justice, and explicability, provide a foundation (Floridi and Cowls, 2019), but federal agencies continue to struggle with translating these ideals into procurement, risk assessment, documentation, and day-to-day oversight. The inventories remain largely descriptive and vary substantially across agencies in depth, specificity, and quality. Agencies differ not only in what they disclose, but also in the clarity with which they communicate data sources, model design, risks, and governance practices.

Transparency tools such as accountable algorithms (Kroll et al., 2017), model documentation (Mitchell et al., 2019), and dataset disclosures (Gebru et al., 2021)offer mechanisms for evaluating system behavior. Many automated systems cannot be adequately governed through existing legal and administrative processes (Edwards and Veale, 2017). Administrative law scholars argue that algorithmic systems challenge traditional accountability structures by obscuring responsibility and reducing contestability (Veale and Brass, 2019). Also, abstraction errors, the removal of systems from their real institutional contexts, produce governance failures. Further demonstrate that internal audits are insufficient, calling for ecosystem-level third-party oversight to evaluate public-sector AI (Raji et al., 2022). Furthermore, the U.S. Government Accountability Office (GAO) repeatedly identifies gaps in agencies' AI inventories, documentation practices, risk assessments, and oversight plans, revealing a mismatch between policy expectations and institutional readiness.

This study uses data from the Federal CIO AI Use Case Inventory to examine two questions: (1)How transparent are U.S. federal agencies about the design, data, and risks of the AI systems they deploy, and (2) what agency characteristics predict higher transparency? To contextualize and validate patterns observed in the CIO inventory, we integrate complementary data from FedScope, GAO reports, and OMB guidance as proxies for institutional capacity, external oversight, and governance expectations. Initial evidence reveals substantial variation in disclosure depth, governance maturity, and documentation quality across agencies.

2. Data and Methodology

2.1. Data Sources

The primary dataset is the Federal CIO AI Use Case Inventory, which records AI systems reported by federal agencies under OMB guidance. We use this inventory to measure governance readiness across transparency, risk, and oversight dimensions (Table 1).

Tab. 1 – Data sources and analytical roles.

Layer	Dataset	Role
Core	CIO AI Inventory 2024	Establish current and How governable AI is.
Capacity	FedScope	Staffing capacity proxy.
Oversight	GAO/OMB	External accountability pressure.

CIO AI Use Case Inventory (2024):

- 1,757 AI use cases reported across the federal government
- 62 standardized fields capturing system characteristics, governance practices, and information
- 37 federal departments and agencies represented

Capacity (FedScope – OPM):

- Federal workforce employment data published by the U.S. Office of Personnel Management
- Provides agency-level staffing counts by occupational series and functional area

Oversight GAO Reports and OMB Guidance:

- Public audit and evaluation reports produced by the U.S. Government Accountability Office and Federal policy guidance and memoranda issued by the Office of Management and Budget
- Capture external oversight, accountability pressure, and identified governance gaps related to federal AI systems

2.2. Method

We evaluate the current state of AI governance in U.S. federal agencies. Rather than treating the federal AI inventory as a catalog of technical systems, we treat it as an institutional artifact that reveals how agencies articulate, structure, and operationalize governance capacity. The analysis, therefore, proceeds in two complementary stages: (1) descriptive inventory analysis as foundational and (2) governance readiness diagnostics as diagnostic.

Stage 1 provides a descriptive characterization of federal AI deployment using the Federal CIO AI Use Case Inventory. The objective of this stage is to establish an institutional context by documenting where AI systems exist, how they are distributed across agencies and policy domains, and how concentrated adoption is across government. This does not evaluate governance quality, system performance, or compliance.

Stage 2 evaluates whether deployed AI systems are accompanied by documented governance mechanisms sufficient for oversight, accountability, and evaluability. Rather than assessing technical model performance, this phase operationalizes governance readiness using disclosure-based indicators derived from the Federal CIO AI Use Case Inventory.

Tab. 2 – Stage 2 Methods overview

Finding	Analytical Focus	Methods
Finding 1	Federal AI inventories are uneven, inconsistent, and limited for oversight	Disclosure completeness audit: field-level fill rates, use-case completeness scores, cross-agency and cross-topic comparisons
Finding 2	High-impact system lack corresponding governance capacity	Risk–governance gap analysis: comparison of governance control prevalence between high-impact and non-high-impact systems
Finding 3	Agencies with more capacity/governance maturity disclose more	Construction of a Data and Engineering Readiness Index and descriptive correlation with governance outcomes

3. Results

The empirical findings draw on data from the Federal CIO reports show that AI deployment across the federal government has outpaced institutional governance capacity, particularly for high-impact systems. AI adoption is unevenly distributed across agencies, with a small subset accounting for a disproportionate share of use cases, many of which are classified as rights- or safety-impacting.

3.1. Stage 1 - Current Stage of Federal AI Deployment

We establish a baseline of federal AI deployment using the 2024 Federal CIO AI Use Case Inventory ($N = 1{,}757$). AI adoption is highly uneven across agencies, with a small number accounting for over half of all reported use cases (Figure 1). The top five agencies: HHS ($n = 271$), VA ($n = 229$), DHS ($n = 183$), DOI ($n = 180$), and USAID ($n = 137$), they represent approximately 57% of deployments, highlighting substantial cross-agency differences in scale and governance burden.

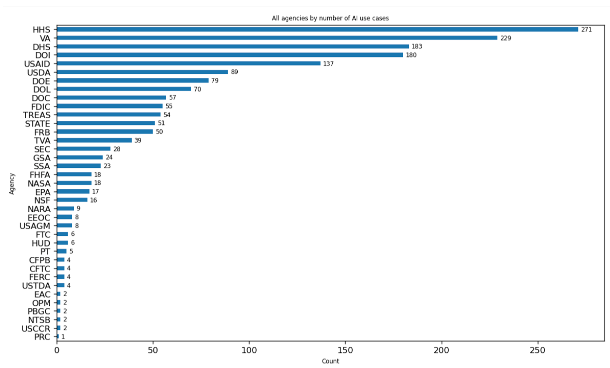

Fig. 1 – Agencies AI use cases

Further more looking at High impact cases, these departments also top of high-impacts area (fig 2).

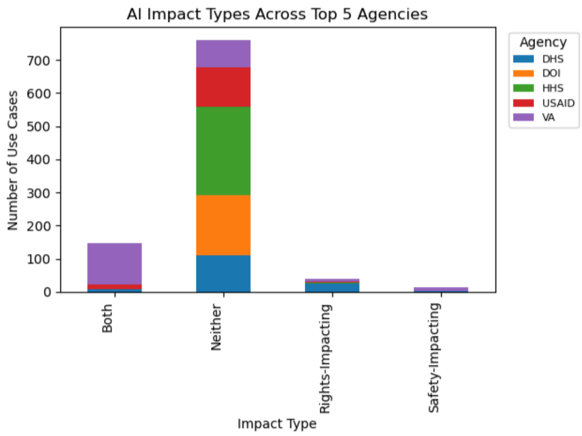

Fig. 2 – High Impact use cases by agencies - top 5 agencies

To assess whether federal AI adoption is broadly distributed or concentrated among a small subset of agencies, we compute two standard concentration measures: the Herfindahl–Hirschman Index (HHI) and the Gini coefficient.

Tab. 3 – Institutional concentration of federal AI use cases across agencies.

Metric	Value
Herfindahl–Hirschman Index (HHI)	0.0806
Gini Coefficient	0.6518

An HHI of 0.0806 indicates moderate concentration in federal AI adoption, while a Gini coefficient of 0.6518 reveals substantial inequality across agencies. These measures show that AI deployment is concentrated among a small subset of agencies, creating uneven oversight burdens and elevating governance risks where AI portfolios are largest.

3.2. Stage 2

As a result, governance gaps or capacity constraints within a small number of high-volume agencies may have outsized effects on overall federal AI accountability, particularly by amplifying weaknesses in oversight and monitoring.

1. Federal AI inventories are uneven, inconsistent, limited for oversight

This pattern reflects high-level disclosure without corresponding documentation for validation or oversight. As shown in Table 4, agencies frequently report AI use but provide limited information on impacts or evaluability. Figure 3 further confirms this pattern at the agency level, where high-level disclosures are common while validation and documentation lag behind.

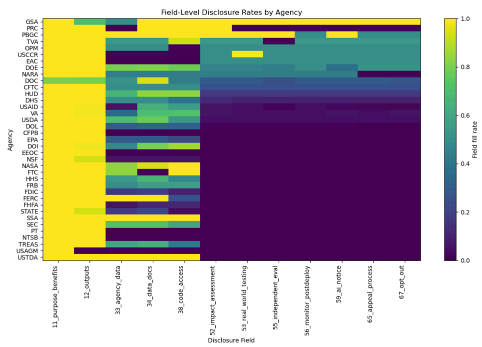

Fig. 3 – Disclosure rates by agency

Tab. 4 – Field-Level Disclosure Rates in the Federal AI Inventory

Field-Level	Fill Rate
Purpose Benefits	0.9937
Outputs	0.9767
Data docs	0.6198
Code access	0.5470
Agency data	0.4491
Impact assessment	0.0894
Real world testing	0.0848
Monitor postdeploy	0.0848
Opt out	0.0820
Appeal process	0.0814
AI notice	0.0797
Independent eval	0.0677

2. High-impact system lack corresponding governance capacity.

Tab. 5 – High-Impact Field-Level in the Federal AI Inventory

Governance Control	High-Impact (%)
Impact assessment (52 == Yes)	3.4924
Independent eval (55 starts with Yes)	2.5029
Monitoring exists (56 >= 1)	6.2282
Appeals available (65 == Yes)	4.4237
Notice provided (59 not None/N/A)	6.7520
Real-world testing (53 >= Benchmark)	5.3551

Over 93–97% of high-impact AI systems lack each core governance safeguard. It showing that the governance capacity is lagging behind deployment and High-impact classification is not yet coupled to mandatory governance actions, risk labeling does not translate into actions.

3. Agencies with more capcity/governance maturity disclose more.

Further exploring the high-impacts, we identified that High-impact AI systems are frequently deployed under vendor arrangements that constrain transparency and oversight.

Tab. 6 – High-Impact governance On Vendor in the Federal AI Inventory

Impact Group	No Code Access (%)	No Data Ownership	Weak Data Docs
High-Impact	20.8964	1.5134	50.6403

Table 6 shows that high-impact AI systems frequently rely on vendor-provided infrastructure, with limited code access and weak data documentation constraining agencies' ability to evaluate and govern deployed systems. A substantial share of high-impact systems operate under conditions of vendor opacity: approximately one-fifth report no access to source code, and over half exhibit weak or missing data documentation. Although lack of formal data ownership is less prevalent, these results indicate that agencies frequently lack the technical artifacts necessary for independent evaluation, auditing, or modification of deployed systems.

Tab. 7 – High-Impact governance on Public in the Federal AI Inventory

Public Info	No(%)	Yes	Missing/NaN
High-Impact	44.5868	11.3504	44.0628

Table 7 shows that public disclosure for high-impact AI systems is rare, with most systems reporting no public notice or providing no information to the public. Public disclosure is limited and highly inconsistent. Fewer than 12

In short, the results describe institutional readiness and evaluability gaps, not technical failure, ethical misconduct, or demonstrated harm.

4. Discussion

This study shows that federal AI inventories can be used to assess governance readiness when interpreted as institutional disclosures rather than technical documentation. By constructing disclosure-based indicators, we identify a systematic gap between AI deployment maturity and governance capacity, particularly for high-impact systems.

Several limitations apply. The CIO AI Inventory is self-reported and reflects variation in reporting practices and institutional capacity. The data do not include outcome-level or subgroup impacts, preventing assessment of model performance, fairness, or real-world harm. High levels of missingness in governance fields further limit comparability, while vendor dependence constrains agencies' ability to disclose or evaluate certain systems.

5. Conclusion and future work

This study assesses transparency and governance readiness in federal AI systems using the Federal CIO AI Use Case Inventory. By constructing disclosure-based indicators of impact classification, system maturity, and governance controls, we show that AI deployment across the federal government has outpaced the institutional capacity required for consistent oversight, particularly for rights- and safety-impacting systems.

High-impact AI systems are already deployed across multiple agencies, yet key governance mechanisms; including impact assessments, post-deployment monitoring, independent evaluation, public notice, and appeal processes are frequently absent or incompletely documented. These gaps persist even among systems in advanced lifecycle stages, indicating that deployment maturity does not reliably translate into governance maturity. Governance readiness varies systematically with institutional capacity: agencies with greater technical staffing disclose more complete information, while vendor-dependent systems exhibit structural opacity that limits evaluability.

This analysis is limited by its reliance on self-reported inventory data and the absence of outcome-level or demographic impact measures. Future work should link inventory disclosures to audits, evaluations, and operational data, examine governance trends longitudinally, and further develop standardized evaluation framework to support consistent, capacity-aware oversight of AI in digital government.

Data/Software Access Statement

github - codes and data

https://github.com/oaitran1/Responsible-AI-Governance-Transparency-and-Readiness-in-Federal-AI-Systems

CIO: https://github.com/ombegov/2024-Federal-AI-Use-Case-Inventory

GAO AI accountability: https://www.gao.gov/products/gao-21-519sp

FedScope: https://www.fedscope.opm.gov

Contributor Statement

Author 1 led the study design, data collection, data processing, analysis, and manuscript preparation. Aurthor 2 provided guidance, and review.

Use of AI

During the preparation of this work, the author(s) used Grammarly in order to check grammar. After using this tool/service, the author(s) reviewed, edited, made the content their own and validated the outcome as needed, and take(s) full responsibility for the content of the publication.

Conflict Of Interest (COI)

There is no conflict of interest.

References

Edwards, L., & Veale, M. (2017). Slave to the algorithm? why a 'right to an explanation' is probably not the remedy you are looking for. DOI: https://doi.org/10.31228/osf.io/97upg.

Faruqe, F., Watkins, R., & Medsker, L. (2021). Competency model approach to AI literacy: Research-based path from initial framework to model [Version Number: 1]. DOI: https://doi.org/10.48550/ARXIV.2108.05809.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for ai in society. *Harvard Data Science Review*. DOI: https://doi.org/10.1162/99608f92.8cd550d1.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for datasets. https://arxiv.org/abs/1803.09010

Kasneci, E., et al. (2023). ChatGPT and other LLMs in education. *Computers and Education: Artificial Intelligence*, *4*, 100111. DOI: https://doi.org/10.1016/j.lindif.2023.102274.

Kroll, J., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*, 633–705.

Long, D., & Magerko, B. (2020). What is AI literacy? competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, 1–16. DOI: https://doi.org/10.1145/3313831.3376727.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. DOI: https://doi.org/10.1145/3287560.3287596.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1–21. DOI: https://doi.org/10.1177/2053951716679679.

OECD. (2023). AI in education: Policy perspectives for equitable adoption. https://www.oecd.org/education/ai-in-education.htm

Raji, I., Kumar, I., Horowitz, A., & Selbst, A. (2022). The fallacy of ai functionality. DOI: https://doi.org/10.48550/arXiv.2206.09511.

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527–563. DOI: https://doi.org/10.1007/s40593-021-00270-2.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *CoRR*, *abs/2002.04087*. https://arxiv.org/abs/2002.04087

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. DOI: https://doi.org/10.1177/0002764213498851.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. DOI: https://doi.org/10.1016/j.respol.2013.05.008.

Veale, M., & Brass, I. (2019). Administration by algorithm? public management meets public sector machine learning.

Wirtz, B., & Müller, W. (2018). An integrated artificial intelligence framework for public management. *Public Management Review*, *21*, 1–25. DOI: https://doi.org/10.1080/14719037.2018.1549268.

Woolf, B. (2010, January). *Building intelligent interactive tutors, student-centered strategies for revolutionizing e-learning*.