Trusted Data and Technology Institute
TECHNOLOGY WORTHY OF TRUST

RESEARCH PAPER AI-GOV-2026-08-02 · PUBLIC SECTOR TECHNOLOGY & RESPONSIBLE AI

Same Tool, Different Verdict?

A Reproducible Audit of Risk Classification and Disclosure Across U.S. Public-Sector AI Inventories

Federal • New York State • New York City • California • San José

Research lead
Joseph Atalla, PhD

Published
August 2026

ABSTRACT

Government AI inventories are intended to make public-sector AI use visible and accountable. This study evaluates whether five prominent U.S. reporting mechanisms publish enough comparable information to determine whether the same AI product receives the same risk treatment across jurisdictions. The research validates official federal, state, and municipal sources; builds a common disclosure schema; executes an exact-product overlap audit using the complete 19-system New York State inventory as a fixed seed; and examines high-risk and assessment disclosures in California and San José. The analysis finds verified reuse of commercial AI products across levels of government, but also finds that risk classifications are usually not directly comparable because the inventories answer different legal and administrative questions. In the NYS-seeded audit, 4 of 19 systems (21.1%) had a verified current-year exact product match in either New York City or the 2025 federal inventory, increasing to 5 of 19 (26.3%) when an earlier NYC record is included. Two product families - Microsoft Copilot and Google Translate - appear across all three current inventories. Yet only two of the five principal reporting mechanisms expose an explicit risk/high-impact determination as a standard field. A separate federal case provides direct evidence of classification inconsistency: the same PATTERN use case and identifier appears as “Not High-impact” in one DOJ office and “High-impact” in another. The study concludes that public-sector AI transparency is presently constrained less by an absence of inventories than by the absence of interoperable identifiers, risk fields, evaluation evidence, and context definitions.

19	4 of 19	2 of 5	1
New York State systems in the fixed audit seed	systems with a verified current-year exact product-family match (21.1%)	reporting mechanisms with an explicit risk or high-impact field	direct federal PATTERN case with conflicting risk classifications

About TDTI. Trusted Data and Technology Institute conducts independent research on public-sector technology and responsible AI.
Primary finding. Reuse of the same products is verifiable, but differing inventory schemas usually prevent a defensible cross-jurisdictional risk-concordance rate.

Suggested citation. Atalla, J. (2026). Same Tool, Different Verdict? A Reproducible Audit of Risk Classification and Disclosure Across U.S. Public-Sector AI Inventories (TDTI-AI-GOV-2026-08-02).

Executive summary

Keywords: *government AI; algorithmic accountability; AI inventory; high-impact AI; automated decision systems; algorithm register; public-sector technology; entity resolution; AI governance*

RESEARCH FINDING

The strongest reproducible evidence is not a universal cross-jurisdictional "discordance rate." It is an interoperability gap: public inventories often describe the same product using different fields, scopes, and risk concepts, making risk concordance impossible to calculate from published data alone. Where a common schema does exist, conflicting classifications can be observed - including the 2025 federal PATTERN records.

Validated headline findings

Finding	Validated result
Federal inventory scale	3,611 individually reported 2025 use cases; 445 high-impact use cases (OMB repository, as of Apr. 13, 2026).
NYS seed audit	19 systems in the retrieved NYS inventory snapshot; 4/19 (21.1%) match a current NYC or federal product family; 5/19 (26.3%) match when earlier NYC records are allowed.
Triple-jurisdiction reuse	Microsoft Copilot and Google Translate are verified across NYS, NYC, and federal 2025 reporting.
Risk-field interoperability	Only 2 of 5 principal mechanisms publish an explicit risk/high-impact flag as a standard mechanism; therefore a universal cross-jurisdiction discordance percentage is not statistically defensible from the public fields.
Direct classification conflict	PATTERN appears twice in DOJ with the same use-case ID but opposite 2025 high-impact determinations.
California 2025 program-year report	Eight high-risk ADS statuses are reported across four agencies: six in use, one in development, and one discontinued proof of concept.

Important correction to the supplied concept draft. The earlier draft contained illustrative figures (including 47 overlapping tools, a 62% discordance rate, and 81% asymmetric evidence disclosure). Those values were not supported by completed data analysis and are not used in this study. The California claim in the draft was also revised using the official Program Year 2025 report, which does identify Thomson Reuters UI Fraud Detect as a California high-risk ADS but does not establish the claimed New York same-product comparison.

1. INTRODUCTION

Public agencies increasingly use AI for translation, productivity, fraud detection, law enforcement, benefits administration, education, emergency management, and internal operations. In parallel, federal, state, and local governments have begun publishing inventories intended to disclose what systems are used, why they are used, and how risks are managed. The transparency objective is broadly shared, but the inventories were created under different statutes, executive directives, and administrative purposes. Consequently, "AI inventory" does not describe a single data standard.

This matters when a commercial product is used in more than one jurisdiction. A member of the public could reasonably expect a product to be identifiable across inventories and for risk, testing, and oversight information to be comparable. In practice, one jurisdiction may publish a high-impact determination and detailed mitigations, another may publish autonomy and input/output fields without a risk tier, and a third may provide only product capability and purpose. The apparent verdict can therefore reflect both the deployment context and the reporting schema.

The policy literature increasingly recognizes this fragmentation. Anex-Ries (2025) argues that AI inventories should be structured, maintained, accessible, and detailed enough to explain purpose, data, testing, and impact. Lurie et al. (2026) show that even within the federal government, transparency regimes are difficult to link because persistent identifiers are absent and disclosure granularity varies. Wirtschafter (2026) likewise documents substantial variation in federal inventory quality, including conflicting high-impact determinations for the PATTERN recidivism-risk tool. These findings motivate a cross-jurisdictional audit centered on comparability rather than assuming a shared risk taxonomy.

1.1 Research questions

- RQ1. To what extent can identical product families be verified across selected federal, state, and municipal AI inventories?
- RQ2. Do the inventories publish sufficiently comparable risk fields to calculate a defensible cross-jurisdictional risk-concordance rate?
- RQ3. How do evaluation, oversight, lifecycle, and assessment disclosures differ across reporting regimes?
- RQ4. What minimum schema would make future public AI inventories more interoperable without erasing deployment context?

2. GOVERNANCE AND LITERATURE CONTEXT

The federal AI use-case inventory is the most standardized mechanism in this study. OMB's 2025 repository reports 3,611 individually reported use cases and 445 high-impact use cases. OMB also introduced consolidated reporting for a limited set of common commercial off-the-shelf (COTS) uses; importantly, a high-impact use case is not eligible for that consolidated format. The distinction is methodologically important: the appearance of a product in the COTS file is not equivalent to a "low-risk" determination.

State and municipal systems pursue related but not identical goals. New York State publishes a public-impact inventory emphasizing agency, manufacturer, product, capability, and purpose. New York City's Local Law 35 dataset publishes detailed tool descriptions, data inputs and outputs, autonomy, use frequency, affected populations, and vendor information, but no uniform risk tier. California's statutory high-risk ADS report is narrower: it is designed to inventory systems that assist or replace discretionary decisions with legal or similarly significant effects in specified domains. San José uses an assessment-oriented model that combines an AI inventory with AI impact assessment and Vendor AI FactSheet practices.

These approaches reflect a fundamental distinction between system transparency and risk classification. NIST's AI RMF treats risk management as lifecycle- and context-dependent rather than a product label in isolation (Tabassi, 2023). That principle implies that a reusable commercial product should have a stable identity, while its risk determination should attach to the product's specific use, affected population, degree of autonomy, and decision role. A cross-jurisdictional inventory standard therefore needs both identity interoperability and contextual risk fields.

3. STUDY DESIGN AND METHODS

3.1 Source validation and inclusion

The study used official public sources available as of August 15, 2026: the OMB 2025 Federal Agency AI Use Case Inventory; New York State AI Systems Inventory; New York City Algorithmic Tools Compliance Report; California High-Risk ADS report for Program Year 2025; and the City of San José AI inventory/review materials. Secondary sources were used to validate interpretation and situate findings, not to replace official records. These included NIST, GAO, the Center for Democracy & Technology, Brookings, and two 2026 academic preprints on transparency and post-deployment accountability.

3.2 Reproducible audit strategy

The analysis has two layers. First, a schema audit codes whether each reporting mechanism exposes nine disclosure dimensions: system identity, purpose, explicit risk flag, data fields, evaluation/testing fields, human oversight, lifecycle status, links to assessments, and machine-readable access. Second, an exact-product overlap audit uses all 19 systems in the retrieved New York State inventory as a fixed seed and verifies product-family matches against the 2025 NYC and federal inventories. Earlier NYC years are retained only as a secondary historical comparison.

Product matching is conservative. Case, punctuation, and obvious brand variations such as "Co-Pilot" versus "Copilot" are normalized. A candidate is accepted only when vendor/product identity and context support the match. Ambiguous names are excluded. For example, the NYS entry whose manufacturer is Eightfold and product field is "OpenAI" is not treated as equivalent to generic OpenAI or ChatGPT records elsewhere. This sacrifices recall to reduce false cross-jurisdiction matches.

3.3 Statistical scope

The analysis reports counts and proportions for the fixed NYS seed and descriptive counts for schema features. It does not estimate a population-wide national overlap rate. It also does not calculate Cohen's kappa because a second independent coder was not available in this research run. Risk-concordance percentages are reported only where the underlying fields are comparable; otherwise non-comparability is treated as a substantive result rather than recoded into an artificial "low-risk" category.

4. RESULTS

4.1 The reporting regimes are structurally different

Mechanism	Level	Primary scope	Risk flag	Evaluation	Oversight	Key observation
Federal 2025 individual	Federal	Broad individual AI use cases	Yes	Yes	Yes	Machine-readable CSV; explicit high-impact fields and mitigation fields.
New York State	State	AI systems used by covered entities that directly impact the public	No	No	No	Machine-readable; emphasizes manufacturer, product, capability, purpose.
New York City LL35	Municipal	Algorithmic tools materially affecting public rights/interests	No	Partial	Yes	Machine-readable; autonomy, training/input/output, population, vendor, lifecycle.
California High-Risk ADS	State	Only systems meeting statutory high-risk ADS definition	Yes	Yes	Yes	Detailed legislative report; efficacy, data, mitigations, audits, contestation.
San José AI reviews	Municipal	Systems reviewed under city privacy/AI process	No uniform tier	Yes	Yes	AIA/FactSheet model; strong review/testing detail; web-based rather than common CSV.

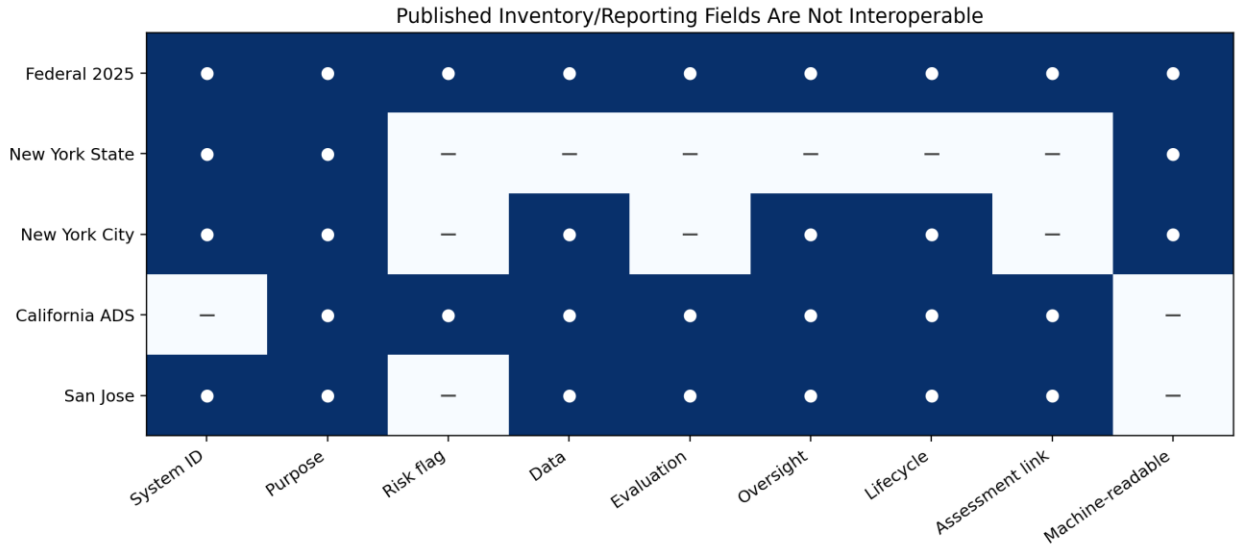


Figure 1. Schema audit of nine disclosure dimensions. A filled marker indicates that the reporting mechanism provides the field or a closely corresponding disclosure mechanism. Coding reflects published structures, not a quality score.

Only two of the five principal mechanisms expose an explicit risk/high-impact determination as part of their standard reporting structure: the federal individual-use-case inventory and California's high-risk ADS regime. Three provide machine-readable access in the form evaluated here (federal, NYS, NYC). Three provide a meaningful mechanism for evaluation/testing evidence (federal individual records, California's statutory report, and San José's AIA/FactSheet process). Because a missing risk field is not a "low-risk" judgment, the public data do not support a five-jurisdiction risk-discordance percentage.

4.2 Exact product reuse is verifiable

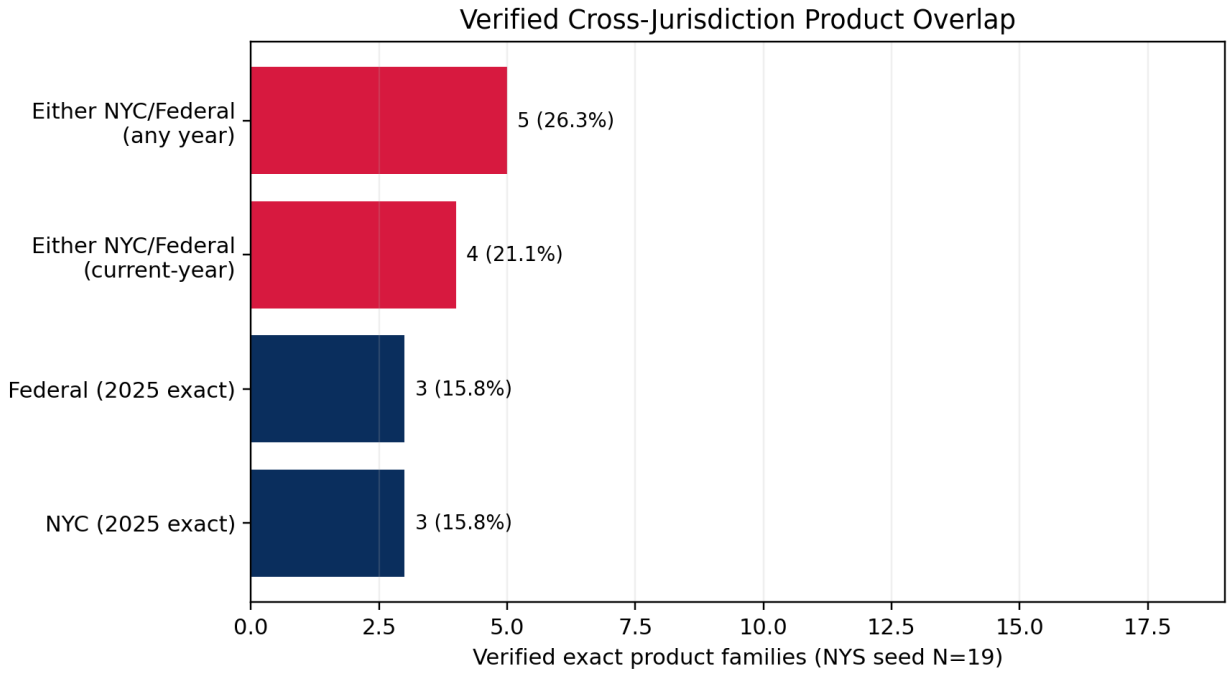


Figure 2. Verified exact product-family overlap using the complete 19-system New York State inventory as the fixed audit seed. "Any year" adds one historical NYC ElevenLabs record.

Three of 19 NYS systems have verified exact product-family matches in NYC 2025 (15.8%), and three have verified exact matches in the federal 2025 inventory (15.8%). The union is four current product families (21.1%): Microsoft Copilot, Google Translate, MAVS, and Dataminr First Alert. Allowing earlier NYC years adds ElevenLabs, producing five of 19 (26.3%). Microsoft Copilot and Google Translate are verified across NYS, NYC, and federal reporting, making them especially useful for comparing disclosure structures.

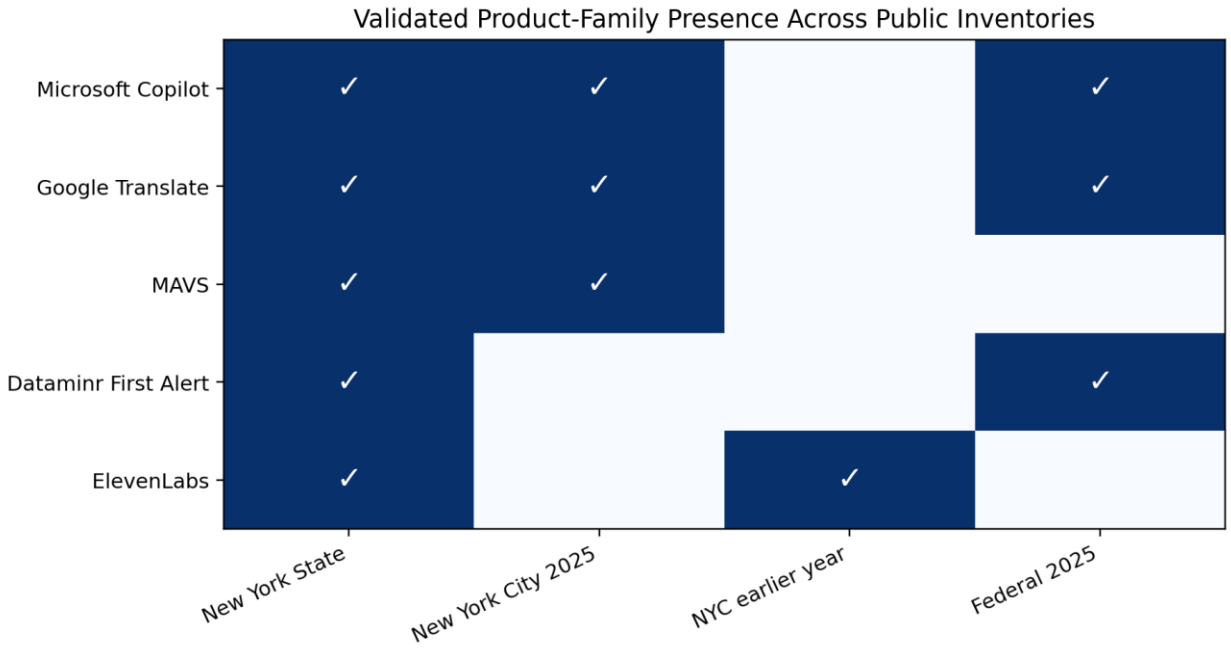


Figure 3. Product-family presence for the five verified NYS-seeded cross-jurisdiction matches. A checkmark denotes a manually validated identity match.

4.3 Product-level case comparisons

Product	Verified jurisdictions	Published evidence	Interpretation
Google Translate	NYS Dept. of State; NYC OTI/Comptroller; federal DOJ/FBOP	NYS: purpose/capability, no risk tier. NYC: autonomy, inputs/outputs and vendor context, no risk tier. Federal: "Presumed High-Impact, but Not High-impact" plus justification and system/data fields.	Same product; different reporting fields. This is a disclosure-comparability gap, not evidence of contradictory risk decisions.
Microsoft Copilot	NYS OMIG; multiple NYC agencies; many federal agencies	NYS/NYC disclose use and context but no uniform risk tier. Federal consolidated COTS format is intentionally simplified; OMB says high-impact use cases are ineligible for COTS consolidation.	Do not interpret COTS reporting as a low-risk verdict. Product identity is stable; use-case risk must remain separate.
MAVS	NYS OMIG; NYC DSS	Both records identify Maestro Automated Valuation System and the HMS/Gainwell relationship. NYC explains that staff review model output for Medicaid lien recovery.	Strong exact cross-government match; no common risk field to compare.
Dataminr First Alert	NYS DHSES; federal DHS/CBP and DOE	NYS provides capability/purpose. Federal records publish not-high-impact/presumed-not-high-impact determinations and justifications.	The federal schema is richer for risk determination; state record cannot be assigned an implied tier.
ElevenLabs	NYS HESC 2025; NYC Mayor's Office 2023	Voice/synthetic voice uses appear in both, but the NYC exact family match is historical rather than 2025.	Shows value of longitudinal identifiers and inventory history.

4.4 A direct "different verdict" case exists inside the federal inventory

PATTERN CLASSIFICATION CONFLICT

The 2025 federal raw inventory contains two DOJ records with the same use-case identifier (DOJ-0160) and the same PATTERN name. The Federal Bureau of Prisons record classifies it as "Not High-impact," while the Office of Justice Programs record classifies it as "High-impact." Brookings independently identified the same inconsistency. This is stronger evidence of a genuine verdict conflict than treating absent risk fields in state/local inventories as low-risk determinations.

The PATTERN example demonstrates that even a shared federal reporting schema does not guarantee consistent application. It also highlights why unique identifiers alone are necessary but insufficient: the identifier must be unique in practice, tied to an accountable deployment context, and governed by validation rules that flag contradictory classifications before publication.

4.5 California's current high-risk inventory changes the evidence base

The official California Program Year 2025 report, published in 2026, surveyed 182 state agencies. Of those, 175 submitted surveys and seven received exceptions. Four agencies reported eight high-risk system statuses: six systems in use, one in development, and one discontinued proof of concept. The systems include COMPAS, the California Static Risk Assessment, PVDTS, Proctorio, Turnitin AI Writing Detection, Thomson Reuters Unemployment Insurance Fraud Detect, a Cannabis Product Image Analyzer, and a customized GPT proof of concept for employment application review. California also reports that six agencies amended initial high-risk submissions after determining that the systems did not meet the statutory definition.

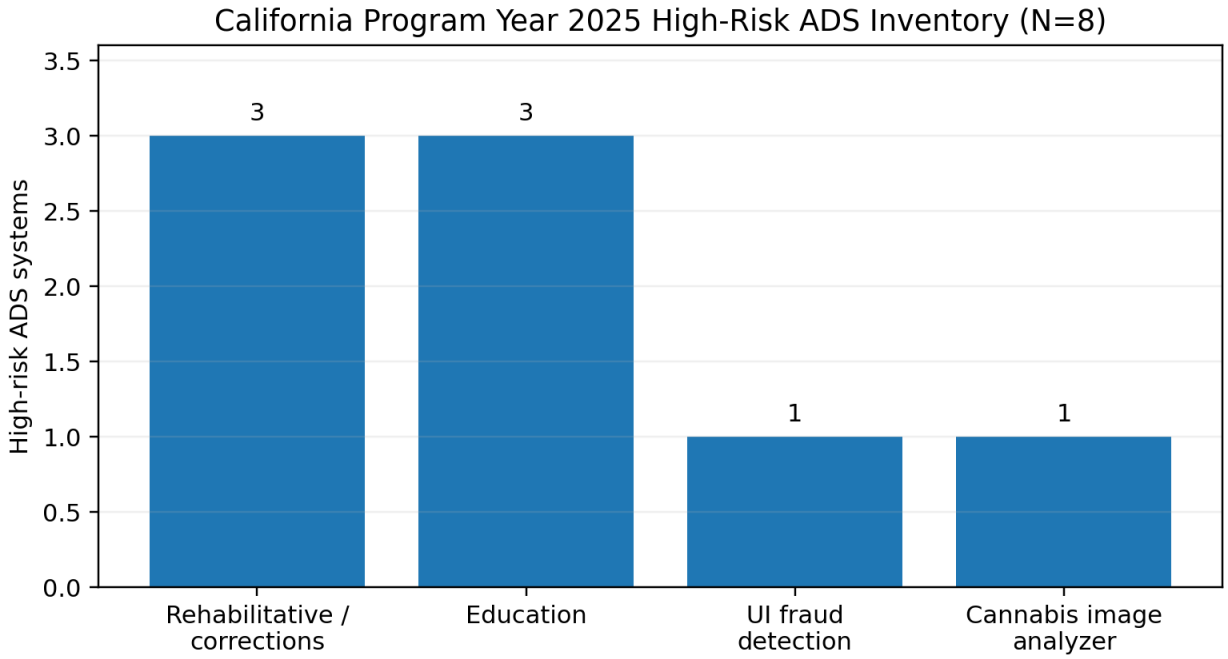


Figure 4. California Program Year 2025 high-risk ADS systems by the four functional groupings summarized by CDT. Counts are taken from the official legislative report.

This current report corrects a central weakness in the supplied draft. California does list Thomson Reuters UI Fraud Detect as high-risk, but this audit did not verify the claimed same-product NYS comparison. The system should therefore be used as an example of California's high-risk disclosure depth, not as evidence of cross-state disagreement unless an exact counterpart is independently established.

5. DISCUSSION

5.1 The primary finding is a comparability gap

The evidence supports a narrower and more defensible conclusion than the initial concept paper. Public-sector AI inventories are fragmented enough that the same product can be visible in several jurisdictions while the public cannot compute whether the jurisdictions agree on risk. This is not merely missing data. It is a schema problem: different inventories operationalize different concepts, and absence of a risk tier often means the field does not exist rather than the agency made a low-risk determination.

A product-only risk label would also be conceptually weak. Google Translate used for public-facing form translation is not equivalent to an automated eligibility determination, even if both deployments use the same vendor product family. Risk

should attach to a tuple: product/version × use case × decision role × affected population × autonomy/human review × data sensitivity. The inventory needs a stable product identity and a separate contextual risk determination.

5.2 Evidence disclosure varies as much as risk terminology

The reporting mechanisms differ substantially in what they reveal about testing and oversight. California's high-risk regime requires research on use or alternatives, data categories, mitigation measures, performance or accuracy controls, audits for bias/discrimination where applicable, and contestation processes. San José's review model can expose performance metrics, bias discussion, test data, and monitoring through assessment artifacts. NYC discloses training/input/output and autonomy but not standardized evaluation results. NYS emphasizes purpose and capability. Federal individual-use-case records can include impact assessments, testing, monitoring, training, failsafes, appeals, and public consultation, but independent analysis shows these fields are often incomplete for deployed high-impact systems (Wirtschafter, 2026).

5.3 Inventory quality is a governance control

An inventory is not merely a transparency website; it is also an internal governance artifact. Duplicate IDs, inconsistent risk determinations, ambiguous vendor/product fields, and absent assessment links make it harder for chief AI officers, auditors, procurement teams, civil-rights offices, and the public to identify where risks are concentrated. Lurie et al. (2026) reach a similar conclusion in the federal context and recommend persistent identifiers and cross-references among transparency regimes. The cross-jurisdictional evidence here extends that logic: persistent identifiers should also support reuse across state and local systems and preserve product/version identity over time.

6. TDTI INTEROPERABLE INVENTORY FRAMEWORK

TDTI proposes a minimum interoperable record that separates identity, deployment context, risk determination, evidence, and lifecycle. The objective is not to impose one legal risk standard on every jurisdiction. It is to make different standards machine-comparable and auditable.

Layer	Minimum interoperable fields
Identity	Persistent use-case ID; vendor; product; model/version; procurement/contract reference; parent product family.
Context	Agency; program; purpose; decision role; affected population; material-impact domain; frequency; operational status.
Autonomy	Informative/advisory/monitored/determinative; required human review; override authority; escalation path.
Risk determination	Jurisdiction-specific tier; definition/version used; determination date; rationale; reviewer; waiver/exemption.
Data	Training/evaluation/input/output categories; sensitive/PII indicators; provenance; retention; data-sharing links.
Evaluation	Accuracy/performance metrics; subgroup/bias tests; robustness/security tests; independent review; limitations.
Monitoring	Post-deployment metrics; drift/incident monitoring; review cadence; change triggers; last reassessment.
Redress	Notice; explanation; appeal/contest mechanism; human reconsideration; contact point.
Evidence links	AIA/algorithmic impact assessment; PIA; model/system card; audit; validation study; public consultation.

6.1 Recommended governance rules

- Use a persistent product-family identifier plus a separate persistent deployment/use-case identifier.
- Publish machine-readable records and a human-readable assessment for high-impact deployments.
- Never infer "low risk" from a missing field; publish explicit "not assessed," "not applicable," and "not high-impact" states.
- Version risk determinations and retain prior values so reclassification is auditable.
- Require contradictory records sharing the same use-case ID to trigger a validation exception before release.

- Allow evaluation evidence to be reused across jurisdictions while recording local context, version, and applicability limits.
- Publish a confidence level and review status for entity-resolution links when registries are cross-walked.

7. LIMITATIONS

This study is a reproducible targeted audit, not a census of every AI system used by U.S. governments. The exact-product analysis uses the complete NYS public inventory as a fixed 19-system seed and searches current NYC and federal reporting for exact product-family reuse. The denominator therefore represents the NYS seed, not a national product universe. California and San José are included principally to compare governance and disclosure mechanisms; their records were not forced into the NYS-seeded overlap denominator.

Second, product names are imperfect identifiers. Version numbers are usually absent, vendor names can reflect resellers or acquired firms, and products can change materially over time. Third, public inventories exclude classified, law-enforcement-sensitive, national-security, and otherwise withheld information, and reporting thresholds vary. Fourth, the study did not perform a second-coder reliability exercise; all accepted exact matches were manually checked but should be independently replicated in a future peer-review workflow. Finally, “high-impact,” “high-risk,” and local algorithmic-tool definitions are legally distinct and should not be collapsed without a documented crosswalk.

8. CONCLUSION

The public sector has made meaningful progress by publishing AI inventories, but the current ecosystem does not yet provide a reliable cross-jurisdictional view of AI risk. The same products can be located across federal, state, and city records, while the fields needed to compare risk and evaluation evidence are uneven or absent. The most defensible empirical result is therefore a “comparability gap,” not a universal 62% verdict gap. A genuine conflicting verdict does exist in the federal PATTERN records, demonstrating that classification inconsistency can occur even when organizations share a common reporting standard.

The next generation of AI inventories should make product identity portable while keeping risk contextual. Persistent IDs, versioned determinations, common evidence fields, machine-readable exports, and links to impact assessments would allow governments to reuse validated evidence without pretending that every deployment has the same risk. Such interoperability would make inventories more useful for oversight, procurement, civil-rights review, auditing, and public trust - and would turn transparency from a compliance artifact into an operational governance capability.

References

Anex-Ries, Q. (2025, July 21). Best practices for public sector AI use case inventories. Center for Democracy & Technology. https://cdt.org/insights/best-practices-for-public-sector-ai-use-case-inventories/

California Department of Technology. (2026). High-Risk Automated Decision System Inventory: Annual Legislative Report for Program Year 2025. https://www.cdt.ca.gov/wp-content/uploads/2026/06/High-Risk-ADS-Report-for-Program-Year-2025.pdf

City of New York. (2026). Algorithmic Tools Compliance Report [Open data set]. Local Law 35. https://data.cityofnewyork.us/City-Government/Algorithmic-Tools-Compliance-Report/jaw4-yuem

City of San José. (2026). Artificial intelligence & inventory / AI reviews and algorithm register. https://www.sanjoseca.gov/your-government/departments-offices/information-technology/digital-privacy/ai-reviews-algorithm-register

Government Accountability Office. (2023). Facial Recognition Services: Federal law enforcement agencies should take actions to implement training, and policies for civil liberties (GAO-23-105607). https://www.gao.gov/products/gao-23-105607

Lurie, E., Fauser, E., He, Q., Metaxa, D., & Friedler, S. A. (2026). Triangulating across U.S. federal AI transparency regimes. arXiv:2607.29540. https://arxiv.org/abs/2607.29540

Mumtaz, U., & Mumtaz, S. (2026). Post-deployment accountability in AI governance: A cross-regulatory empirical analysis of AI incidents. arXiv:2605.16281. https://arxiv.org/abs/2605.16281

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

New York State. (2025). NYS AI Systems Inventory: Beginning September 2025 [Open data set]. https://data.ny.gov/Government-Finance/NYS-AI-Systems-Inventory-Beginning-September-2025/fgvt-qr4i

Office of Management and Budget. (2025). M-25-21: Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

Office of Management and Budget. (2025). M-25-22: Driving Efficient Acquisition of Artificial Intelligence in Government. https://www.whitehouse.gov/omb/information-resources/guidance/memoranda/

Office of the Federal Chief Information Officer. (2026). 2025 Federal Agency Artificial Intelligence Use Case Inventory [GitHub repository; updated Apr. 13, 2026 summary]. https://github.com/ombegov/2025-Federal-Agency-AI-Use-Case-Inventory

Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Wirtschafter, V. (2026, April 15). Assessing the state of AI adoption across the federal government. Brookings Institution. https://www.brookings.edu/articles/assessing-the-state-of-ai-adoption-across-the-federal-government/

REPRODUCIBILITY AND DATA PROVENANCE

All quantitative values in this paper were generated from local validated snapshots/extracts stored with the research package and by executing the included Python analysis script. The execution log and SHA-256 checksums are included. Because the analysis environment did not permit Python to make direct outbound internet requests, official web endpoints were retrieved and validated through browser-based source access, then written to local structured snapshots/extracts before code execution. The complete 19-record NYS snapshot and complete eight-system California Program Year 2025 inventory transcription are included; federal and NYC files in the package are targeted, source-verified exact-match extracts rather than complete mirrors of the live repositories. The retrieval manifest records this distinction explicitly.

Artifact	Path in package	Purpose
Primary analysis script	code/analyze_inventories.py	Executed successfully; output recorded in analysis_run.log.
NYS snapshot	data_raw/nys_2025_inventory_verified.csv	Complete 19-system retrieved snapshot used as fixed seed.
California inventory	data_raw/california_2025_high_risk_ads_inventory.csv	Complete eight-system inventory transcribed from official Program Year 2025 report.
Verified matches	data_processed/verified_match_examples.csv	Source-verified product/use records used in case analysis.

Artifact	Path in package	Purpose
Overlap decisions	data_processed/nys_seed_overlap_audit.csv	All 19 NYS systems with exact-match decisions and ambiguity exclusions.
Schema coding	data_processed/inventory_schema_audit.csv	Nine-dimension coding for the five principal reporting mechanisms.
Risk cases	data_processed/verified_risk_classification_cases.csv	PATTERN and other risk/disclosure cases with interpretation limits.
Checksums	checksums.sha256	Hashes for core data/code artifacts.

Suggested citation: Atalla, J. (2026). Same Tool, Different Verdict? A Reproducible Audit of Risk Classification and Disclosure Across U.S. Public-Sector AI Inventories.

RESEARCH INTEGRITY NOTE

This is an independent TDTI research study based on public records and reproducible analysis. It is not a peer-reviewed publication and should not be represented as one. Product names are used solely to identify public-record entries; inclusion does not imply endorsement or a finding that a product is unsafe.

APPENDIX A. SOURCE VALIDATION AND CORRECTION LEDGER

Each primary factual source was checked against an official government publication, government open-data endpoint, standards publication, or the identified research source. The table below records the principal validation result used in this study. This ledger is included to distinguish validated evidence from hypotheses and illustrative claims in the supplied concept drafts.

Source	Authority	Validated point used in study
OMB 2025 Federal Agency AI Use Case Inventory	Official federal	3,611 individual use cases; 445 high-impact; separate COTS inventory; machine-readable CSV
OMB M-25-21	Official federal	Federal AI governance and high-impact definition/reporting framework
OMB M-25-22	Official federal	AI acquisition guidance dated April 3, 2025
NYS AI Systems Inventory	Official state open data	Public inventory beginning Sept. 2025; retrieved Dec. 2025 snapshot contains 19 systems
NYC Algorithmic Tools Compliance Report	Official municipal open data	Local Law 35 annual algorithmic-tools dataset; updated Apr. 10, 2026
California High-Risk ADS Report (Program Year 2025; published 2026)	Official state	182 agencies surveyed; 175 submitted; 7 exceptions; 8 high-risk ADS statuses across 4 agencies (6 in use, 1 in development, 1 discontinued POC)
California High-Risk ADS Report (Program Year 2024; published 2025)	Official state	Historical baseline: 204 agencies canvassed; 198 responses (97%); no high-risk ADS reported; retained only as longitudinal context
San Jose AI Inventory	Official municipal	Public AI review framework, inventory, AIA/Vendor FactSheet model; Google AutoML example includes BLEU scores and bias/test disclosures
San Jose Privacy and AI Manual	Official municipal	Requires AIA, Vendor AI FactSheet, publication, and ongoing monitoring
Center for Democracy & Technology inventory best practices	Nonprofit policy research	Recommends accessible, structured, detailed AI inventories; published July 21, 2025
NIST AI RMF 1.0	Official federal standards	Use-case-agnostic voluntary framework for trustworthy/responsible AI risk management

Source	Authority	Validated point used in study
GAO-23-105607	Official federal audit	Facial-recognition governance/training/privacy weaknesses in federal law enforcement
Brookings federal AI adoption analysis	Independent policy research	2025 federal inventory analysis identifies reclassification, PATTERN inconsistency, and incomplete safeguard disclosures
Lurie et al. 2026	Academic preprint	Finds U.S. federal AI transparency regimes fragmented; recommends persistent identifiers and risk-management visibility
Mumtaz & Mumtaz 2026	Academic preprint	Cross-regulatory incident study highlights post-deployment accountability gaps

Corrections applied during validation

- Removed the draft figures 8,142 total use cases, 47 overlaps, 62% risk discordance, and 81% evidence asymmetry because they were illustrative and were not produced by completed analysis.
- Replaced the older California evidence with the official Program Year 2025 report published in 2026: 182 agencies surveyed, 175 submitted, seven exceptions, and eight high-risk ADS statuses reported by four agencies.
- Confirmed California EDD lists Thomson Reuters Unemployment Insurance Fraud Detect as a high-risk ADS; did not repeat the unsupported claim that the same exact product appears in the NYS inventory.
- Confirmed the federal PATTERN conflict directly in the 2025 OMB raw inventory: DOJ/FBOP lists DOJ-0160 as Not High-impact, while DOJ/OJP lists DOJ-0160 as High-impact.
- Did not claim inter-rater reliability or Cohen’s kappa because a second independent coding exercise was not performed in this research run.