Research article

Archives of Public Affairs and Institutional Management

Trustworthy AI for E-Governance: Formal Verification Frameworks as Accountability Infrastructure in the AI-First Era

Tiffany A. Ceasor*

Independent Researcher, USA

Founder & Principal Researcher, TSAR, 501(c)(3), Seattle, WA, USA

Doctoral Researcher, Florida Atlantic University, Boca Raton, FL, USA

***Corresponding Author**
Tiffany A. Ceasor, Independent Researcher, Founder & Principal Researcher, TSAR, 501(c)(3), Seattle, WA, USA.

Submitted: 2026, Apr 28; **Accepted:** 2026, May 19; **Published:** 2026, May 29

Citation: Ceasor, T. A. (2026). Trustworthy AI for E-Governance: Formal Verification Frameworks as Accountability Infrastructure in the AI-First Era. *Arch of Pub Aff Inst Manag, 1*(2), 01-11.

Abstract

As societies transition into an AI-first paradigm, governments worldwide confront an unprecedented accountability challenge: how to ensure that algorithmic systems shaping public decisions remain transparent, auditable, and aligned with democratic values. Current approaches emphasizing post-hoc explainability fall critically short of the verification demands now emerging in regulatory frameworks across the United States, the European Union, and subnational jurisdictions. This paper introduces Formal Decision Traces (FDTs) as accountability infrastructure for e-governance, providing machine-checkable proof certificates that document compliance of each AI output with encoded policy constraints. Grounded in Satisfiability Modulo Theories (SMT) verification within neuro symbolic architectures, FDTs transform algorithmic accountability from an interpretive exercise into a design-embedded, auditable technical capability. Drawing on empirical benchmarks from deployed and evaluated systems---including the ARC neuro symbolic framework achieving 99.2% verification soundness on policy compliance tasks Bayless, Amazon Web Services Automated Reasoning checks delivering up to 99% hallucination detection accuracy in commercial deployment AWS, and the VERAFI financial compliance system demonstrating 94.7% factual correctness through SMT-based policy validation (Akinfaderin & Subramanian)---and analyzing the regulatory landscape including the U.S. Office of Management and Budget Memorandum M-26-04, the EU AI Act (Regulation 2024/1689), the NIST AI Risk Management Framework, the Colorado AI Act, and landmark judicial proceedings such as Mobley v. Workday, we develop a multi-level framework for integrating formal verification into e-governance systems [1-3]. We propose a five-level taxonomy of verification integration depth, identifying Level 3 (Semantic Verification Layer) as the minimum viable architecture for public-facing government AI. The framework offers practical, standards-aligned guidance for designing, implementing, and governing AI systems that embed accountability by design, positioning formal verification as the foundational standard for trustworthy AI in the digital governance era.

Keywords: Trustworthy AI, E-governance, Formal Verification, Algorithmic Accountability, Audit Trails, Neuro Symbolic AI, Formal Decision Traces, EU AI Act, NIST AI RMF, Digital Transformation, Public Administration, SMT Verification

1. Introduction

The global transition toward AI-driven governance represents one of the most consequential transformations in public administration since the advent of electronic government (Tangi, United Nations) [4,5]. Governments at every level are deploying algorithmic systems to determine benefits eligibility, monitor regulatory compliance, assess risk in criminal justice, screen employment applications, and generate public-facing communications. The OECD documents over 200 real-world examples of government AI deployments across eleven core functions---from delivering

public services and administering justice to fighting corruption and managing public finances---while the EU Public Sector Tech Watch observatory catalogued nearly 1,500 AI use cases across European public administrations as of March 2025 OECD [6]. Yet as these systems become increasingly central to the encounters between citizens and the state, a fundamental accountability gap has emerged: the very capabilities that make AI systems powerful also make them resistant to the forms of oversight that democratic governance demands Bovens & Zouridis, Janssen & Kuk). This paper addresses a central challenge at the intersection of e-governance and digital transformation: how can governments design, deploy, and govern AI systems that are genuinely accountable---not merely explainable---in an era where algorithmic decision-making is becoming the default mode of public administration? We argue that the distinction between explainability and verifiability represents a critical gap in both scholarship and practice [7,8]. Explainable AI (XAI) techniques describe why a system produced an output, formal verification proves that an output satisfies specified constraints. For e-governance, this distinction carries profound implications: a system that explains its reasoning but cannot prove its compliance with policy requirements exposes agencies to legal liability, undermines public trust, and fails to embed the accountability that democratic governance demands (Fountain, Alon-Barkat & Busuioc). The urgency of this challenge is underscored by a rapidly crystallizing regulatory landscape [9,10].

In the United States, the Office of Management and Budget's Memorandum M-26-04 (December 2025) requires federal AI systems to demonstrate truth-seeking and ideological neutrality, making compliance material to contract eligibility and payment (OMB) [11]. The NIST AI Risk Management Framework (NIST) and its Generative AI Profile (NIST) establish trustworthiness characteristics---including validity, reliability, accountability, and transparency---as governance benchmarks [12,13]. The European Union's AI Act (Regulation 2024/1689) introduces mandatory conformity assessments, technical documentation, and post-market monitoring for high-risk AI systems, with penalties reaching €35 million or 7% of global annual turnover (European Parliament & Council) [14]. At the subnational level, the Colorado AI Act mandates impact assessments for high-risk consequential decisions (Colorado General Assembly), while over 1,000 AI-related bills have been introduced across nearly every U.S. state in 2024--2025 (Future of Privacy Forum) [15,16]. Simultaneously, landmark judicial proceedings---including Mobley v. Workday, Inc., where courts recognized that algorithmic agents bear direct liability for discriminatory decisions---are establishing legal precedents that demand auditable, verifiable AI systems [17,18]. We introduce Formal Decision Traces (FDTs) as accountability infrastructure for e-governance. FDTs are machine-checkable proof certificates documenting that each AI output satisfies encoded policy constraints, generated through Satisfiability Modulo Theories (SMT) verification integrated into neurosymbolic AI architectures. This proposal is grounded in a growing body of empirical evidence demonstrating the viability of neurosymbolic verification at production scale. The ARc framework achieves 99.2% soundness on policy compliance verification tasks through redundant autoformalization with SMT cross-checking Bayless [1].

AWS Automated Reasoning checks, now commercially deployed in Amazon Bedrock, deliver up to 99% accuracy in detecting factual errors through SMT-LIB formal logic translation AWS [2]. The VERAFI system demonstrates that integrating neurosymbolic policy validation into financial compliance workflows achieves 94.7% factual correctness---an 81% relative improvement over traditional retrieval-augmented approaches---by translating regulatory requirements into formal SMT-lib specifications covering GAAP compliance, SEC requirements, and mathematical validation (Akinfaderin & Subramanian) [3]. This paper makes four contributions. First, we develop a conceptual framework positioning Formal Decision Traces as accountability infrastructure for e-governance, bridging public values theory with technical verification capabilities and the emerging regulatory landscape. Second, we propose a five-level taxonomy of verification integration depth, providing actionable guidance aligned with international standards. Third, we present a comprehensive regulatory cross-walk demonstrating how FDTs satisfy requirements across the U.S. federal, EU, and subnational regulatory frameworks. Fourth, we synthesize emerging empirical evidence from neuro symbolic verification systems to demonstrate the practical viability of FDT architectures for e-governance applications.

2. Theoretical Background and Regulatory Context

2.1. Public Values, E-Governance, and the AI Transformation

The relationship between information systems and public values has been a central concern in e-governance scholarship. Bannister and Connolly identify three categories of public sector values influenced by information and communication technologies: duty-oriented values encompassing responsiveness, accountability, and transparency, service-oriented values including availability, effectiveness, and timeliness, and socially-oriented values such as equity, citizen involvement, and privacy [19]. They argue that ICT is value-influencing rather than value-neutral or value-determined, meaning that design choices shape the realization of public values without mechanically determining outcomes. Rose extend this analysis by identifying four value positions---bureaucratic, professional, political, and market---each with distinct implications for system design, demonstrating that values are embedded in e-government systems through the processes and practices of their design, development, and use [20]. The digital transformation of government is now entering what Tangi characterize as the AI-augmented era, where artificial intelligence fundamentally reshapes organizational structures and sociotechnical dynamics within public administrations [4]. The United Nations E-Government Survey 2024 documents this transformation across 193-member states, emphasizing that accountability and transparency are critically important in e-government planning, implementation, and evaluation (United Nations) [5]. Fan identifies three governance models emerging globally---market-driven, government-guided, and regulatory-led---each presenting distinct accountability

challenges as AI systems assume decision-making roles previously held by human administrators [21]. Fountain's analysis of algorithmic bias extends public values concerns to AI specifically, arguing that computational algorithms can encode, magnify, and perpetuate systemic biases present in training data and design choices. Crucially, she identifies transparency as necessary but insufficient: transparency about algorithm design and data does not ensure fairness or equity [9]. This insight motivates our focus on verification rather than mere transparency---the capacity to prove that systems satisfy value commitments, not merely disclose how they operate.

2.2. The Accountability Crisis in Algorithmic Governance

Bovens and Zouridis presciently identified the accountability challenges of automated government, warning that the shift from street-level to system-level bureaucracies changes the nature of administrative discretion and may change the nature of constitutional control [7]. Two decades later, AI systems have intensified this challenge: not only is discretion embedded in code, but that code may be opaque even to its developers. Alon-Barkat and Busuioc provide empirical evidence of how automation bias---the tendency to over-rely on algorithmic recommendations---and selective adherence---the tendency to accept algorithmic advice when it confirms prior beliefs---undermine human oversight in practice, concluding that organizational and institutional arrangements matter greatly for how human-AI interactions unfold [10]. The consequences of this accountability crisis are no longer theoretical. In Mobley v. Workday, Inc., a federal court ruled that AI vendors themselves can be held liable for discriminatory hiring decisions made through their algorithms, reasoning that drawing an artificial distinction between software decision-makers and human decision-makers would potentially gut anti-discrimination laws in the modern era (740 F. Supp. 3d at 807) [17]. In May 2025, the court certified this case as a nationwide collective action, potentially covering millions of applicants screened through Workday's AI platform (Mobley v.Workday) [18]. Similarly, in the Safe Rent Solutions settlement, an AI vendor paid over \$2 million after courts found that its algorithmic tenant-screening tool disparately impacted Black and Hispanic housing applicants [22]. Research by Wilson and Caliskan demonstrated that AI resume-screening systems preferred white-associated names 85.1% of the time, providing empirical evidence of systematic algorithmic bias at scale [23]. Peeters and Widlak examine how algorithmic systems can create new forms of administrative burden---what they term the digital cage---where information architecture excludes citizens from services they are entitled to receive [24]. These developments collectively demonstrate that the accountability gap in algorithmic governance is not an abstract scholarly concern but a concrete institutional failure with material consequences for citizens, agencies, and democratic legitimacy.

2.3. The Regulatory Imperative: From Aspiration to Enforcement

The regulatory landscape for AI governance has undergone a fundamental transformation from aspirational guidelines to enforceable requirements across multiple jurisdictions, creating unprecedented demand for technical accountability mechanisms.

2.3.1. United States Federal Framework

The U.S. federal approach to AI governance has evolved rapidly through a series of executive orders and administrative memoranda. Executive Order 14110 established the initial framework for safe, secure, and trustworthy AI development, directing NIST to develop standards and agencies to implement governance structures [25]. The NIST AI Risk Management Framework (AI RMF 1.0), released in January 2023, defines trustworthiness characteristics including validity, reliability, safety, security, resilience, accountability, transparency, explainability, privacy-enhancement, and fairness with managed bias (NIST) [12]. The companion Generative AI Profile (NIST AI 600-1), released in July 2024, addresses risk unique to or exacerbated by generative AI, including confabulation, information integrity, and CBRN information security (NIST) [13]. Most significantly, OMB Memorandum M-26-04 implements Executive Order 14319 by establishing binding procurement requirements for Large Language Models [26]. The memorandum requires agencies to contractually ensure that LLMs comply with two principles: truth-seeking, defined as prioritizing historical accuracy, scientific inquiry, and acknowledgment of uncertainty, and ideological neutrality, defined as functioning as nonpartisan tools without encoded partisan judgments (OMB) [11]. By making compliance with these requirements material to contract eligibility and payment, M-26-04 invokes the False Claims Act framework, where vendors who knowingly misrepresent material facts to obtain federal payment face treble damages and per-claim penalties. Agencies must update procurement policies by March 11, 2026, and provide ongoing evaluation of model behavior and safeguards against violations.

2.3.2. European Union AI Act

The EU AI Act (Regulation 2024/1689), entered into force on August 1, 2024, establishes the world's first comprehensive legal framework for artificial intelligence (European Parliament & Council). The Act adopts a risk-based classification with transparency obligations that vary by system risk level [14]. For high-risk AI systems---which include those used in critical infrastructure, education, employment, essential services, law enforcement, and migration---the Act mandates conformity assessments, risk management systems, data governance, technical documentation, record-keeping, human oversight provisions, and robustness and accuracy requirements. Article 13 requires transparency sufficient for deployers to understand and use systems correctly, Articles 9 and 15 mandate risk management and accuracy respectively, and Article 72 requires post-market monitoring systems. Penalties for non-compliance reach €35 million or 7% of global annual turnover. The transparency rules take full effect in August 2026, with high-risk system obligations following in August 2027.

2.3.3. Subnational and Emerging Frameworks

At the subnational level, a proliferation of AI legislation is creating a complex compliance landscape. The Colorado AI Act (SB 24-205), enacted in May 2024 and taking effect in 2026,

establishes obligations for developers and deployers of high-risk AI systems that make or substantially factor into consequential decisions, mandating impact assessments, transparency, and a duty of care (Colorado General Assembly) [15]. The California AI Transparency Act (SB 942) requires AI systems with over one million monthly visitors to implement comprehensive disclosure mechanisms. Utah's AI Policy Act (SB 149) established disclosure requirements for generative AI in commercial contexts [27,28]. The Future of Privacy Forum tracked 210 bills across 42 states in 2025 alone, with 20 enrolled or enacted, noting a shift toward transparency-driven approaches and liability clarification [16]. This subnational activity occurs against a backdrop of federal-state tension: a proposed 10-year moratorium on state AI regulation in the 2025 budget reconciliation bill was rejected by the Senate in a 99-1 vote, affirming states' independent regulatory authority.

2.4. The Limits of Explainability

The dominant response to algorithmic opacity has been explainable AI (XAI)---techniques that describe why models produce particular outputs, including attention visualization, feature importance scores, counterfactual explanations, and local interpretable approximations. While valuable for transparency, XAI faces three fundamental limitations for e-governance accountability. First, explanations are descriptive rather than normative: they describe computational factors that influenced an output but do not prove that the output satisfies policy requirements. As Rudin argues, explanations for complex models necessarily involve fidelity trade-offs [29]. Second, explanations require human interpretation, and empirical evidence demonstrates that such interpretation is systematically unreliable due to automation bias and selective adherence (Alon-Barkat & Busuioc). Third, explanations are not auditable at scale: public sector AI systems may generate millions of decisions, and auditing explanations for each would require resources no government possesses [10]. These limitations motivate our distinction between explainability and verifiability. Explainability supports transparency, verification provides what transparency alone cannot: machine-checkable proof that outputs satisfy specified constraints.

3. Formal Decision Traces as Accountability Infrastructure

3.1. From Explanation to Verification

We propose a fundamental reorientation in how e-governance systems approach algorithmic accountability: from seeking explanations to requiring verification. Table 1 contrasts these paradigms across dimensions relevant to digital governance, mapped to current regulatory frameworks.

Dimension	Explainability Paradigm	Verification Paradigm
Core question	Why did the system produce this output?	Does this output satisfy specified constraints?
Evidentiary basis	Descriptive approximation	Machine-checkable proof certificate
Scalability	Requires human interpretation per decision	Automated checking at scale, human review for flagged exceptions
Auditability	Subjective evaluation of explanation adequacy	Independent verification of proof validity
Legal standing	Informative but non-deterministic	Deterministic compliance documentation
Alignment with NIST AI RMF	Supports Explainability characteristic	Supports Validity, Reliability, and Accountability characteristics
EU AI Act alignment	Addresses Article 13 transparency requirements	Addresses Articles 9, 15, and 72 conformity requirements

Table 1: Explainability vs. Verification Paradigms for E-Governance AI

The verification paradigm does not replace explainability but complements it. Transparency remains a crucial public value and a requirement under both the EU AI Act and the NIST AI RMF. However, verification provides what transparency alone cannot: deterministic assurance that each decision satisfies encoded requirements. In the language of public values theory (Bannister & Connolly), verification addresses duty-oriented accountability while transparency addresses socially-oriented participation [19].

3.2. Defining Formal Decision Traces

A Formal Decision Trace (FDT) is a machine-checkable proof certificate documenting that an AI system output satisfies a specified set of policy constraints. FDTs possess four essential characteristics that distinguish them from existing accountability mechanisms. Formal specification. Policy requirements are encoded as formal axioms in SMT-LIB, the standard language for Satisfiability Modulo Theories (Barrett). This axiomatization process translates policy language into precise logical constraints [30]. For example, the M-26-04 truth-seeking requirement might be axiomatized as: for all claims C in output O, there exists a source S such that S is in an authoritative knowledge base and S supports C with confidence above a threshold. Automated verification. Outputs are checked against encoded constraints by SMT solvers---software tools that determine whether a set of logical formulas is satisfiable. Industry-standard solvers include Z3 (de Moura & Bjørner) and CVC5 (Barbosa). When an output satisfies all

constraints, the solver returns SAT with a proof certificate, when constraints are violated, it returns UNSAT with a counterexample identifying the specific violation [31]. Proof certificate generation [32]. Each compliant output is accompanied by a proof certificate---a formal artifact documenting which constraints were checked and how the output satisfied them. Proof certificates are independently verifiable: their validity can be confirmed without re-running the original verification process, creating an auditable record that satisfies even stringent evidentiary requirements including those under the False Claims Act framework invoked by M-26-04. Constraint traceability. FDTs maintain linkage between proof certificates and the policy requirements they verify. When policies change---as when the EU AI Act provisions take effect in stages through 2027---affected axioms can be identified and updated. When outputs fail verification, the specific violated constraints are reported, supporting the targeted remediation that both NIST AI RMF and EU AI Act post-market monitoring require.

3.3. Illustrative Walkthrough: Benefits Eligibility Determination

To make the FDT framework concrete for e-governance practitioners, we present a simplified walkthrough illustrating how formal verification would operate in a common public administration scenario: an AI system advising on benefits eligibility. This example is stylized for clarity but reflects the structural logic of deployed verification systems such as VeriCoT, which formalizes chain-of-thought reasoning steps into first-order logic and verifies their logical validity against source documents Feng. Consider a state agency deploying an AI system to assist caseworkers in determining whether applicants qualify for a housing assistance program [33]. The program's eligibility policy specifies three requirements:

- The applicant's household income must not exceed 150% of the federal poverty level,
- The applicant must be a resident of the state, and
- The applicant must not have received a comparable benefit in the preceding 12 months.

Step 1: Axiomatization. The policy requirements are translated into formal SMT-LIB constraints by a team of policy specialists and verification engineers. Each requirement becomes a machine-checkable proposition. Requirement

- Becomes an arithmetic constraint comparing household income against the poverty-level threshold. Requirement
- Becomes a membership check against the state residency registry.
- Becomes a temporal constraint verifying the absence of prior benefits within the specified window. This axiomatization process---translating natural language policy into formal logic---is precisely the task at which neuro symbolic systems are rapidly improving: the ARC framework demonstrates that LLMs with redundant auto formalization and SMT cross-checking can translate natural language policies into formal specifications with 99.2% soundness Bayless [1].

Step 2: AI output generation and verification. The AI system processes an applicant's case file and generates a recommendation: "Applicant qualifies for housing assistance based on reported household income of \$28,400 (127% FPL for household of three), confirmed state residency, and no prior benefits in the relevant period." Before this recommendation reaches the caseworker, the SMT solver checks each claim against the encoded constraints and authoritative data sources: Is \$28,400 in fact below the 150% FPL threshold for a household of three? Does the residency registry confirm state residency? Does the benefits database confirm no prior comparable benefit? If all constraints are satisfied, the solver returns SAT and generates a proof certificate. Step 3: Proof certificate generation and archival. The proof certificate documents which constraints were checked, which data sources were consulted, and how each constraint was satisfied. This certificate accompanies the recommendation as an auditable record. If the solver instead returns UNSAT---for example, if the AI system miscalculated the poverty-level percentage---the specific violated constraint is identified, the recommendation is flagged for human review, and no proof certificate is issued. The caseworker receives the flagged output with an explanation of which constraint failed, enabling targeted human oversight rather than requiring review of every decision. This walkthrough illustrates how FDTs transform the accountability relationship between AI systems and human decision-makers in public administration. Rather than asking caseworkers to evaluate whether an AI explanation seems plausible, FDTs provide deterministic verification of compliance and flag only those cases requiring human judgment. The result is a governance architecture that scales AI-assisted decision-making while preserving meaningful accountability.

3.4. FDTs as Institutional Accountability Infrastructure

We conceptualize Formal Decision Traces not merely as a technical capability but as institutional infrastructure for e-governance accountability, following Fountain's analysis of technology enactment in government [34]. FDTs create four institutional accountability capabilities that existing approaches cannot provide. First, FDTs enable accountability by design. Rather than retrofitting accountability mechanisms onto opaque systems, FDTs embed verification into the system architecture itself. No output is delivered without a corresponding proof certificate. This satisfies the EU AI Act's Article 9 requirement for risk management to be integrated into the AI system's lifecycle. Second, FDTs support distributed oversight. Proof certificates can be verified by any party with access to the encoded constraints and a conforming solver. This creates possibilities for multi-stakeholder oversight: internal agency reviewers, external auditors, judicial bodies, inspectors general, and civil society organizations can independently verify compliance. The constraint set becomes a boundary object (**Star & Griesemer**) enabling coordination across institutional boundaries---precisely the cross-agency coordination that M-26-04 implementation demands [35]. Third, FDTs create auditable accountability records. Every verified output produces a proof certificate that can be stored, retrieved, and audited, constituting an institutional memory of algorithmic decision-making. These records satisfy the EU AI Act's Article 12 logging requirements, M-26-04's ongoing evaluation mandates, and the NIST AI RMF's

Measure function requirements for continuous monitoring. Fourth, FDTs facilitate value articulation. The axiomatization process requires explicit articulation of the values that AI systems should embody. This makes value commitments visible, contestable, and subject to democratic deliberation---addressing the governance concern Rose identify as essential to legitimate e-government and operationalizing the NIST AI RMF's Govern function [20].

4. A Taxonomy of Verification Integration Depth

Not all verification integration is equivalent. We propose a five-level taxonomy mapping the depth of SMT verification integration to the strength of accountability guarantees (Table 2). This taxonomy provides actionable guidance for e-governance practitioners and procurement officers specifying requirements in alignment with international standards.

Level	Integration Type	Proof Strength	Maturity	Example Systems
5	Verification-in-the-Loop Training	Maximal	Theoretical	None published
4	Symbolic Reasoning Integration	Very Strong	Early research	Proof of Thought (Banerjee et al., 2024)
3	Semantic Verification Layer	Strong	Advanced research / early commercial	ARc (99.2% soundness), AWS Automated Reasoning (~99%), VERAFI (94.7% factual correctness)
2	Constrained Decoding	Moderate	Prototype	Research implementations
1	Post-Hoc Filtering	Weak	Production-ready	Content moderation systems

Table 2: Five-Level Taxonomy of Verification Integration for E-Governance AI

4.1. Level 1: Post-Hoc Filtering

At Level 1, SMT verification is applied after the AI system generates a complete output. Non-compliant outputs are rejected and regenerated until a compliant output is produced. This approach is production-ready and simple to implement. However, for e-governance accountability, Level 1 provides weak guarantees: no formal proof certificates are generated, regeneration attempts are not documented, and the verification process itself is not auditable. Level 1 is inadequate for compliance with M-26-04's materiality requirements or the EU AI Act's conformity assessment mandates.

4.2. Level 2: Constrained Decoding

At Level 2, SMT solvers guide generation token-by-token, masking continuations that would lead to constraint violations. This enforces compliance during generation rather than filtering after the fact, providing moderate accountability guarantees. Level 2 remains at prototype stage due to computational costs and the difficulty of encoding semantic constraints at the token level.

4.3. Level 3: Semantic Verification Layer (Recommended Minimum)

At Level 3, SMT solvers verify complete outputs in real-time before delivery, generating Formal Decision Traces for each compliant output. We identify Level 3 as the minimum viable architecture for public-facing e-governance AI systems. Empirical evidence from multiple independent research groups demonstrates viability at this level. The ARC framework achieves 99.2% soundness at its most conservative threshold through a two-stage neuro symbolic approach: LLMs first formalize natural language policies with optional human guidance, then inference-time auto formalization validates logical correctness through multiple redundant formalization steps with SMT cross-checking for semantic equivalence Bayless [1]. At this threshold, the framework demonstrates a 2.5% false positive rate and 92.6% precision, significantly outperforming the next-best alternative method, which achieved only 98.3% soundness with twice the false positive rate. AWS Automated Reasoning checks achieve approximately 99% verification accuracy in commercial deployment by translating domain knowledge into SMT-LIB formal logic and checking LLM outputs against these specifications AWS [2]. These benchmarks exceed the two-nines reliability threshold common in safety-critical applications. Level 3 satisfies M-26-04's requirement for ongoing evaluation and safeguards, provides the conformity evidence required by the EU AI Act, and operationalizes the NIST AI RMF's Measure and Manage functions through continuous verification and documented proof certificates.

4.4. Level 4: Symbolic Reasoning Integration

At Level 4, symbolic reasoning participates in output generation through hybrid neural-symbolic inference. Reasoning chains are verified incrementally, not just final outputs. Level 4 remains in early research stages but represents a promising direction for applications requiring very strong accountability guarantees. The VeriCoT framework Feng provides a prototype of Level 4 capabilities, formalizing each chain-of-thought reasoning step into first-order logic and identifying the premises grounding each step-in source context, commonsense knowledge, or prior reasoning [33]. Evaluated on the Legal Bench and Bio ASQ datasets, VeriCoT demonstrates that step-level verification effectively identifies flawed reasoning and serves as a strong predictor of final answer correctness---capabilities directly relevant to e-governance applications where both the conclusion and the reasoning process must be auditable.

4.5. Level 5: Verification-in-the-Loop Training

At Level 5, compliance constraints are embedded directly in model training objectives, producing systems that are inherently compliant by construction. This would provide maximal

accountability guarantees but remains theoretical due to the non-differentiability of SMT solving. Level 5 represents a long-term research goal.

4.6. Practical Guidance for E-Governance Level Selection

For e-governance practitioners, we recommend the following alignment: Public-facing, high-stakes systems---including benefits determination, regulatory enforcement, public communications, and systems where outputs directly affect citizen rights---should require Level 3 minimum, with proof certificates retained as accountability records. This aligns with the EU AI Act's requirements for high-risk systems and M-26-04's materiality framework. Internal analytical tools may accept Level 1-2 with documented human oversight, consistent with the NIST AI RMF's risk-proportionate approach. Emerging applications should monitor Level 4-5 research for future adoption.

4.7. Emerging Empirical Evidence for Neuro Symbolic Verification

A critical question for e-governance practitioners is whether neuro symbolic verification has progressed beyond theoretical proposals to demonstrable empirical results. Table 4 synthesizes the emerging evidence base from independently developed systems, each combining neural language models with formal verification through SMT-based or logic-based approaches.

System	Verification Approach	Key Metric	Domain	Reference
ARC Framework	Neuro symbolic LLM + SMT auto formalization with redundant cross-checking	99.2% soundness (2.5% FPR) at most conservative threshold	Natural language policy compliance (Conditional QA)	Bayless et al. (2025)
AWS Automated Reasoning	SMT-LIB formal logic translation of domain knowledge for LLM output checking	Up to 99% verification accuracy in hallucination detection	Commercial deployment (pharma, utilities, cloud compliance)	AWS (2025)
VERAFI	Neuro symbolic policy generation with SMT-lib specifications for financial validation	94.7% factual correctness (81% improvement over RAG baselines)	Financial compliance (GAAP, SEC requirements)	Akinfaderin & Subramanian (2025)
VeriCoT	First-order logic formalization of chain-of-thought reasoning steps	Effective identification of flawed reasoning, strong predictor of answer correctness	Legal reasoning (Legal Bench), biomedical (Bio ASQ)	Feng et al. (2025)

Table 4: Empirical Evidence from Neuro Symbolic Verification Systems

The VERAFI system merits particular attention for e-governance because it demonstrates how neuro symbolic verification operates in a regulated compliance context closely analogous to government applications [1-3]. VERAFI integrates neuro symbolic auto formalization to translate financial validation requirements---covering GAAP compliance, SEC regulatory requirements, and mathematical accuracy---into formal SMT-lib specifications that guide agentic reasoning during generation (Akinfaderin & Subramanian) [3]. On Finance Bench-style financial question-answering tasks, VERAFI achieves 94.7% factual correctness and 96.4% completeness. The neuro symbolic policy layer contributes a 4.3 percentage point improvement beyond the agentic baseline, demonstrating that formal constraint specification provides measurable accuracy gains specifically in domains requiring regulatory compliance---precisely the operational context of e-governance AI. Collectively, this evidence demonstrates three findings relevant to e-governance. First, neuro symbolic verification has crossed the threshold from theoretical possibility to empirical demonstration, with multiple independent systems achieving high soundness rates across diverse domains. Second, the approach is particularly effective in regulated, policy-bound contexts---exactly the environments in which government AI systems operate. Third, commercial deployment (AWS Automated Reasoning in Amazon Bedrock) confirms that verification can be integrated into production infrastructure without prohibitive latency or computational costs. While these results derive from research and commercial rather than government applications, they establish a credible evidence base for Level 3 verification viability in e-governance contexts.

5. Regulatory Cross-Walk: FDTs Across Governance Frameworks

A critical contribution of this paper is demonstrating how Formal Decision Traces provide a unified technical mechanism for satisfying accountability requirements across the fragmented regulatory landscape. Table 3 presents a cross-walk mapping FDT capabilities to specific requirements in major governance frameworks.

Framework	Key Requirement	FDT Alignment
OMB M-26-04 (U.S.)	Truth-seeking and ideological neutrality, material to contract eligibility	FDTs axiomatize truth-seeking as source-grounded claims and neutrality as balanced-presentation constraints
EU AI Act (2024/1689)	Conformity assessment, technical documentation, post-market monitoring for high-risk AI	Proof certificates serve as conformity evidence, FDT archives satisfy logging and monitoring mandates
NIST AI RMF 1.0	Govern, Map, Measure, Manage cycle, trustworthiness characteristics	FDTs operationalize Validity, Reliability, and Accountability across all four functions
NIST AI 600-1 (GAI Profile)	Manage confabulation, information integrity risks	SMT verification of factual claims against authoritative knowledge bases addresses confabulation risk
Colorado AI Act (SB 205)	Impact assessments and duty of care for high-risk consequential decisions	FDTs provide deterministic compliance records for consequential-decision audits
California AI Transparency Act (SB 942)	Disclosure of AI-generated content	Proof certificates document provenance and verification status of each output

Table 3: Regulatory Cross-Walk: FDT Alignment with Governance Frameworks

5.1. OMB M-26-04 Compliance

M-26-04's verification requirements expose fundamental limitations of pure neural language models for e-governance applications. Neural models generate outputs by sampling from probability distributions, preventing deterministic verification, they encode patterns without preserving provenance information, preventing source attribution, and XAI techniques describe outputs but do not prove constraint satisfaction. FDTs address each requirement: truth-seeking is axiomatized as constraints on factual claims requiring authoritative source grounding, ideological neutrality is axiomatized as balanced-presentation constraints, ongoing evaluation is supported through proof certificate archives, and the verification architecture itself constitutes the safeguard M-26-04 requires, as non-compliant outputs do not receive proof certificates and are not delivered.

5.2. EU AI Act Conformity

For high-risk AI systems under the EU AI Act, FDTs provide conformity evidence at multiple levels. Risk management requirements (Article 9) are satisfied through the axiomatization process, which requires systematic identification and encoding of all applicable constraints. Technical documentation requirements (Article 11) are met through the complete specification of constraint sets and verification architectures. Logging requirements (Article 12) are fulfilled through proof certificate archives documenting every verified output. Transparency requirements (Article 13) are addressed through human-readable constraint specifications that complement machine-checkable proofs. Post-market monitoring (Article 72) is operationalized through continuous verification with automated alerting for emerging compliance patterns.

5.3. NIST AI RMF Operationalization

The NIST AI RMF's four core functions---Govern, Map, Measure, and Manage---each find operationalization through FDTs. The Govern function is supported through axiomatization governance processes defining who approves constraint sets and how they are maintained. The Map function is served by the systematic identification of policy requirements during formalization. The Measure function is operationalized through continuous verification producing quantitative compliance metrics. The Manage function is realized through deterministic constraint enforcement preventing non-compliant outputs from delivery. For the Generative AI Profile (**NIST AI 600-1**), FDTs specifically address confabulation risk through source-grounded verification and information integrity through factual accuracy constraints.

6. Case Studies and Legal Precedents

6.1. Mobley v. Workday: Algorithmic Agent Liability

The Mobley v. Workday litigation illustrates the legal consequences of deploying unverifiable algorithmic systems in consequential decision-making. Workday's AI platform screened job applicants and provided hiring recommendations to employers. When plaintiffs---five individuals over the age of forty who were rejected in almost every instance without an interview---alleged age and disability discrimination, the court ruled that AI vendors bear direct liability as agents of the employers using their tools. The court's reasoning that nothing in the language of the federal anti-discrimination statutes distinguishes between delegating functions to an automated agent versus a live human one (740 F. Supp. 3d at 807) establishes a precedent with profound implications for e-governance: any government agency using AI for consequential decisions faces liability for algorithmic discrimination. Had Workday's system employed Level 3 verification with FDTs, fairness constraints could have been axiomatized---for example, requiring that recommendation scores demonstrate no statistically significant variation across protected categories---and proof certificates documenting constraint satisfaction could have provided legally defensible compliance evidence. The class certification in May 2025, potentially covering millions of applicants, underscores the systemic scale of unverified algorithmic decision-making.

6.2. Safe Rent Solutions: Housing Discrimination Through Algorithmic

Screening The Safe Rent Solutions case demonstrates algorithmic accountability failures in public-adjacent services. Safe Rent's tenant-screening algorithm produced risk scores that courts found disparately impacted Black and Hispanic applicants. Critically, the court rejected Safe Rent's argument that it could not be liable

because it did not make final housing decisions, holding instead that because the algorithm claimed to automate human judgment through an undisclosed process, the vendor bore responsibility. The \$2 million settlement (Safe Rent Solutions Settlement) established that algorithmic opacity itself constitutes an accountability failure [22]. FDTs would transform this dynamic by requiring explicit axiomatization of fairness constraints and generating proof certificates documenting that each risk assessment satisfies anti-discrimination requirements. The constraint traceability feature would enable retrospective audits identifying systemic disparate impact patterns before they result in litigation.

6.3. Healthcare AI Claim Denials: Accountability in Life-Critical

Systems In 2024-2025, major health insurers faced lawsuits for allegedly using AI to wrongfully deny medical claims. One case cited an insurer's internal process where an algorithm reviewed and rejected over 300,000 claims in two months, averaging 1.2 seconds per claim. These cases illustrate the consequences of deploying AI at scale without verification infrastructure: the speed of algorithmic processing makes traditional human oversight physically impossible, and the absence of verifiable compliance records leaves both patients and agencies without recourse. Formal verification at Level 3 would require that each claim decision be checked against encoded coverage criteria, producing a proof certificate documenting the basis for each determination and flagging decisions that fail constraint satisfaction for mandatory human review.

7. Discussion: Implications for E-Governance

7.1. Designing Accountable E-Governance AI Systems

Our framework offers practical guidance for e-governance practitioners. The core recommendation is accountability by design: verification capabilities should be architectural requirements embedded from the earliest stages of system procurement and development. We recommend four design principles. First, specify verification requirements during procurement, aligned with the applicable regulatory framework: Level 3 minimum for public-facing systems under M-26-04 or the EU AI Act, Level 1-2 with documented human oversight for internal tools. Second, invest in axiomatization---the interdisciplinary process of translating policy requirements into formal constraints---allocating resources for collaboration between policy experts and formal methods specialists. Third, plan for constraint evolution, designing governance processes for updating axiom sets as requirements change, the R^2-Guard framework (Kang & Li) demonstrates that knowledge-enhanced guardrails can adapt to new requirements by editing reasoning graphs rather than retraining [36]. Fourth, establish proof certificate archives with defined retention periods, access controls, and audit procedures.

7.2. The Sociotechnical Integration Challenge

Implementation of verification infrastructure requires attention to organizational, technical, and governance dimensions---the sociotechnical integration that Tangi identify as essential to AI-augmented government transformation [4]. Organizationally, verification capabilities require new professional roles: axiomatization specialists who translate policy requirements into formal constraints, verification engineers who maintain SMT solver deployments, and compliance auditors who evaluate proof certificate archives. Technically, Level 3 verification requires SMT solver deployments with standardized APIs, both commercial options (AWS Automated Reasoning, now generally available in Amazon Bedrock across multiple regions) and open-source alternatives (Z3, CVC5) are available. For government applications, FedRAMP compliance of verification infrastructure should be evaluated. Governance structures must specify who approves constraint sets, how verification failures are escalated, what audit procedures apply, and how citizens can challenge algorithmic decisions.

7.3. Democratic Legitimacy and the Limits of Formalization

FDTs provide technical infrastructure but do not resolve the political question of whose values should govern. Axiomatized constraints encode value commitments, democratic legitimacy requires that constraint sets be developed through appropriate deliberative processes, subject to public scrutiny, and revisable through political mechanisms. Moreover, not all public values can be encoded as formal constraints---justice, dignity, and care involve contextual judgments that resist algorithmic specification. Governance frameworks should specify which requirements are subject to formal verification and which require human review. The risk of Goodhart's Law---when metrics become targets, they cease to be good metrics---must also be managed through mechanisms for identifying gaming behavior and updating constraints. Finally, citizen engagement requires translating FDT capabilities into accessible accountability mechanisms: plain-language summaries, appeals processes, and public reporting on verification patterns.

7.4. International Dimensions and Standardization

The emergence of formal verification as an accountability mechanism carries significant implications for international e-governance. The EU AI Act's extraterritorial reach---applying to any provider whose AI system outputs are used within the EU---creates incentives for global standardization of verification approaches. The OECD's (2025) documentation of over 200 government AI deployments across member states, together with its identification of governance, data infrastructure, skills, investment, procurement, and partnerships as the seven key enablers for successful AI adoption in government, suggests growing demand for interoperable accountability mechanisms. We anticipate that FDT-based verification could provide a common technical infrastructure for cross-jurisdictional compliance, with constraint libraries encoding shared international standards and jurisdiction-specific axioms capturing local regulatory requirements.

7.5. Limitations and Future Research

This paper has several limitations suggesting directions for future research. First, our empirical evidence derives from commercial and research systems, not deployed government applications, case studies of FDT implementation in public administration would strengthen the framework. The evidence base, while growing-

--spanning the ARC framework's 99.2% soundness on policy compliance, VERAFI's 94.7% factual correctness in financial regulation, and AWS's commercial deployment---has not yet been validated in government-specific operational contexts with their unique bureaucratic, procurement, and interoperability constraints. Second, axiomatization of policy requirements remains an emerging practice, even the highest-performing systems involve tradeoffs between soundness and recall, as the ARC framework's 99.2% soundness comes with a 15.6% recall rate at the most conservative threshold Bayless, meaning some compliant outputs may be unnecessarily flagged [1]. Third, our analysis focuses on output verification but does not address training data governance, model selection, or full system lifecycle management. Fourth, the framework would benefit from cost-benefit analyses comparing verification infrastructure investment against the legal, reputational, and democratic costs of accountability failures. Fifth, cross-national comparative studies of verification implementation would illuminate how institutional contexts shape adoption patterns.

8. Conclusion

As societies transition into an AI-first paradigm, the question of algorithmic accountability in e-governance is no longer merely technical but fundamentally democratic. Citizens deserve more than explanation of algorithmic decisions, they deserve proof that decisions satisfy the requirements their governments have established. This paper has argued that addressing this challenge requires moving beyond explainability to verification---from describing why systems produce outputs to proving that outputs satisfy policy constraints. We introduced Formal Decision Traces as accountability infrastructure for e-governance---machine-checkable proof certificates generated through SMT verification, documenting that each AI output satisfies encoded policy constraints. Drawing on empirical evidence from systems achieving 99.2% verification soundness on policy compliance Bayless, approximately 99% hallucination detection accuracy in commercial deployment AWS, and 94.7% factual correctness in regulated financial compliance workflows (Akinfaderin & Subramanian), and analyzing the rapidly crystallizing regulatory landscape spanning OMB M-26-04, the EU AI Act, the NIST AI RMF, subnational legislation, and landmark judicial proceedings, we demonstrated that formal verification of AI outputs is both practically achievable and increasingly legally required [1-3]. Our five-level taxonomy of verification integration provides actionable guidance for e-governance practitioners. Level 3---the Semantic Verification Layer---represents the recommended minimum for public-facing government AI systems, with multiple independent research groups and commercial deployments demonstrating viability. The regulatory cross-walk demonstrates that FDTs provide a unified technical mechanism for satisfying accountability requirements across the fragmented international governance landscape. The technical foundations for trustworthy AI in e-governance exist. What remains is the institutional will to deploy them. As the March 2026 M-26-04 implementation deadline approaches, as EU AI Act transparency provisions take full effect in August 2026, and as state legislatures continue to expand algorithmic accountability requirements, the window for proactive adoption of verification infrastructure is narrowing. Governments that embed formal verification now will establish the accountability standards---and the public trust---that the AI-first era demands. Those that delay risk the legal liability, democratic legitimacy erosion, and citizen harm that unverified algorithmic governance inevitably produces. Formal Decision Traces offer a path forward---one that combines the pattern recognition and language capabilities that make AI valuable with the formal guarantees that democratic governance requires. The challenge now is to institutionalize verification as the foundational standard of e-governance in the digital age.

References

1. Bayless, S., Boxwell, S., Bradbury, J., et al. (2025). A neurosymbolic approach to natural language formalization and verification. arXiv preprint arXiv:2511.09008.
2. AWS. (2025, August). Minimize AI hallucinations and deliver up to 99% verification accuracy with Automated Reasoning checks: Now available [Press release]. Amazon Web Services.
3. Akinfaderin, A., & Subramanian, S. (2025). VERAFI: Verified agentic financial intelligence through neurosymbolic policy generation. arXiv preprint arXiv:2512.14744.
4. Tangi, L., Müller, A. P. R., & Janssen, M. (2025). AI-augmentedgovernment transformation: Organisational transformation and the sociotechnical implications of artificial intelligence in public administrations. Government Information Quarterly, 42(3), 102055.
5. United Nations Department of Economic and Social Affairs. (2024). E-Government survey 2024: Accelerating digital transformation for sustainable development. United Nations.
6. OECD. (2025). Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing.
7. Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. Public Administration Review, 62(2), 174--184.
8. Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 33(3), 371--377.
9. Fountain, J. E. (2022). The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. Government Information Quarterly, 39(2), 101645.
10. Alon-Barkat, S., & Busuioc, M. (2023). Human--AI interactions in public sector decision making: Automation bias and selective adherence to algorithmic advice. Journal of Public Administration Research and Theory, 33(1), 153--169.
11. Office of Management and Budget. (2025c). Increasing public trust in artificial intelligence through unbiased AI principles (Memorandum M-26-04). Executive Office of the President.
12. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1).

13. National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1).
14. European Parliament & Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (EU AI Act). Official Journal of the European Union, L series.
15. Colorado General Assembly. (2024). Colorado Artificial Intelligence Act (SB 24-205).
16. Future of Privacy Forum. (2025). The state of state AI: Legislative approaches to AI in 2025.
17. Mobley v. Workday, Inc., 740 F. Supp. 3d 796 (N.D. Cal. 2024).
18. Mobley v. Workday, Inc., 2025 WL 1424347 (N.D. Cal. May 16, 2025) (class certification).
19. Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. Government Information Quarterly, 31(1), 119--128.
20. Rose, J., Persson, J. S., Heeager, L. T., & Irani, Z. (2015). Managing e-Government: Value positions and relationships. Information Systems Journal, 25(5), 531--571.
21. Fan, Z. (2025). The role of artificial intelligence in the digital transformation of government: Opportunities and ethical challenges. Frontiers in Public Health, 13, 1694996.
22. SafeRent Solutions Settlement. (2024). Connecticut Fair Housing Center v. SafeRent Solutions, LLC (D. Conn. 2024).
23. Wilson, K., & Caliskan, A. (2024). Gender, race, and intersectional bias in resume screening via language model retrieval. University of Washington Information School.
24. Peeters, R., & Widlak, A. (2023). The digital cage: Administrative exclusion through information architecture---The case of the Dutch civil registry's master data management system. Government Information Quarterly, 40(2), 101785.
25. Executive Order 14110. (2023, October 30). Safe, secure, and trustworthy development and use of artificial intelligence. Federal Register, 88(210), 75191--75226.
26. Executive Order 14319. (2025, July 23). Preventing woke AI in the Federal Government. The White House.
27. State of California. (2024). California AI Transparency Act (SB 942). California Legislature.
28. State of Utah. (2024). Utah Artificial Intelligence Policy Act (SB 149). Utah Legislature
29. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206--215.
30. Barrett, C., Fontaine, P., & Tinelli, C. (2025). The SMT-LIB Standard: Version 2.7.
31. de Moura, L., & Bjørner, N. (2008). Z3: An efficient SMT solver. In C. R. Ramakrishnan & J. Rehof (Eds.), Tools and Algorithms for the Construction and Analysis of Systems (pp. 337--340). Springer.
32. Barbosa, H., Barrett, C. W., Brain, M., et al. (2022). cvc5: A versatile and industrial-strength SMT solver. In D. Fisman & G. Rosu (Eds.), Tools and Algorithms for the Construction and Analysis of Systems (pp.415--442). Springer
33. Feng, Y., Weir, N., Bostrom, K., Bayless, S., Cassel, D., Chaudhary, S., Kiesl-Reiter, B., & Rangwala, H. (2025). VeriCoT: Neuro-symbolic chain-of-thought validation via logical consistency checks. arXiv preprint arXiv:2511.04662
34. Fountain, J. E. (2001). Building the virtual state: Information technology and institutional change. Brookings Institution Press.
35. Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907--39. Social Studies of Science, 19(3), 387--420.
36. Kang, M., & Li, B. (2025). R²-Guard: Robust reasoning enabled LLM guardrail via knowledge-enhanced logical reasoning. In Proceedings of the International Conference on Learning Representations.
37. Dou, L., & Dou, X. (2025). Towards just AI: Challenges and solution framework for algorithmic discrimination in judicial system. International Journal of Digital Law and Governance, 2(1), 39--81.
38. Lindgren, I., Madsen, C. Ø., Hofmann, S., & Melin, U. (2019). Close encounters of the digital kind: A research agenda for the digitalization of public services. Government Information Quarterly, 36(3), 427--436.
39. Office of Management and Budget. (2025a). Accelerating federal use of AI through innovation, governance, and public trust (Memorandum M-25-21). Executive Office of the President.
40. Office of Management and Budget. (2025b). Driving efficient acquisition of artificial intelligence in government (Memorandum M-25-22). Executive Office of the President.
41. Peeters, R. (2023). Digital administrative burdens: An agenda for analyzing the citizen experience of digital bureaucratic encounters. Perspectives on Public Management and Governance, 6(1), 7--13.