

The Trust Deficit in Digital Government: A Conceptual Model of AI-Driven Public Services

Dumitru Gabriel Cristian¹; Toma Alexandru²;
Deak Elena Roxana³; Neguțoiu Milian⁴

^{1,2,3,4} Bucharest University of Economic Studies, Bucharest, Romania

Corresponding Author: Deak Elena Roxana*

Publication Date: 2026/07/20

Abstract: This paper provides a comprehensive conceptual model examining the intersection of artificial intelligence (AI) and citizen trust within public administration. As governments transition from traditional e-government to algorithmic governance, implementing automated systems such as AI-driven chatbots, automated benefit allocation and predictive administrative routing create a significant psychological and systemic trust deficit. This research conceptualizes a three-dimensional framework to analyze and mitigate this deficit with a particular focus on algorithmic transparency, decisional equity, and data security. Hence, the proposed model addresses the shift from human-mediated bureaucracy to black-box algorithmic execution, therefore offering public managers an ethical blueprint for designing citizen-centric digital services that align with emerging regulatory mandates like the European Union AI Act.

Keywords: Algorithmic Governance, Public Administration, Citizen Trust, Artificial Intelligence, Ethical AI, Digital Transformation.

How to Cite: Dumitru Gabriel Cristian; Toma Alexandru; Deak Elena Roxana; Neguțoiu Milian (2026) The Trust Deficit In Digital Government: A Conceptual Model of AI-Driven Public Services. *International Journal of Innovative Science and Research Technology*, 11(7), 381-386. <https://doi.org/10.38124/ijisrt/26jul390>

I. INTRODUCTION

The relentless acceleration of digital transformation has pushed modern public administrations past the boundaries of traditional e-government. Modern evaluation frameworks traditionally measure electronic government success by technical uptime or processing speed. These standard metrics fail when software scripts begin making discretionary choices that alter citizens' lives. The interface between the state and the individual requires deep normative alignment. Our model addresses this shifting focus from purely functional attributes to protective, relational safeguards. Public sector automation demands structural accountability that public organizations can map and counteract the opacities that estrange the public by anchoring administrative systems to transparency, equity and data security. Moreover, the expansion of data systems fundamentally alters how state agencies manage public resource distribution [1]. Early administrative efforts focused mainly on converting paper files into downloadable forms [2]. Contemporary public management moves beyond these simple transitions. Institutions now implement algorithmic governance to oversee daily administrative routines [3]. Software applications process tax liabilities, evaluate social benefits, and handle civic queries without direct human mediation [4].

These tools optimize internal processing speeds. Yet, this mechanical acceleration causes severe institutional friction.

Scholars look closely at system reliability to measure citizen satisfaction with e-government tools. Traditional evaluation approaches examine web platform availability, technical uptime, and online submission speeds. These common measures miss a critical element. They ignore how citizens react when automated routines make discretionary choices that impact human lives. A computer script operates without a visible chain of personal responsibility. When a software application denies a medical subsidy, people demand answers. The traditional foundation of administrative fairness breaks down when a program fails to provide a plain-language explanation.

This research addresses this gap by creating a relational conceptual framework. We bypass functional operational metrics to investigate the normative bond connecting the public to automated state systems. Our model identifies three specific pillars: structural transparency, decisional equity, and data sovereignty. We show how public organizations can preserve administrative legitimacy during technological transitions [5]. The following sections review the literature, outline our framework, discuss policy execution and propose

future field studies. to ground this investigation with systemic analytical rigor, this paper seeks to answer a central unresolved research question:

RQ: How do the intersecting qualitative dimensions of algorithmic transparency, decisional equity, and data sovereignty relationally constrain the systemic equilibrium of citizen trust in automated public administrations?

The answer to this question allows the inquiry to move beyond loose, descriptive goals to establish verifiable constraint model for digital governance.

II. LITERATURE REVIEW

➤ *Institutional Trust and Automated Changes*

Several public policy studies suggest that institutional trust serves as the primary currency for civic compliance [6]. When citizens believe an agency operates with competence, benevolence, and integrity, they cooperate with state mandates. Standard administrative models, like the classic Weberian framework, maintain those clear legal rules and visible human responsibility build this institutional confidence.

However, recent investigations warn that automated systems disrupt this traditional relational link [7]. Scholars emphasize that algorithms act as inscrutable boxes that separate the state from the population [8]. The problem with current technology acceptance models is their narrow focus on user convenience. They skip the moral obligations of public enforcement. This study addresses that gap by collecting distinct building blocks from ethical governance literature [9]. We choose these specific sources because they prioritize public accountability over mere corporate efficiency.

➤ *Theoretical Assertions on Algorithmic Bias*

Many researchers argue that algorithms are never neutral because they mimic the social inequalities present in historical training records [10; 11]. Some theorists claim that software objectivity is a myth [12]. Conversely, tech-centric reports note that automation removes human favoritism from public queues.

Our study challenges the view that speed equals progress. We hypothesize that unreviewable automated choices undermine the social contract. To ground this view, we examine how algorithmic design impacts citizen compliance. If an individual cannot contest an automated decision, institutional trust drops. This dynamic shows why opinion congruence between the public and administrators matters [13; 14]. Therefore, we organize our conceptual dimensions to show how public organizations can restore transparent communication.

III. MATERIALS AND METHODS

This study utilizes a qualitative conceptual synthesis to build a normative framework for public managers. We restate our primary aim, which is to define the specific

dimensions required to protect citizen trust during automation. This methodological framework is explicitly calibrated to resolve our primary research question (RQ). The qualitative synthesis establishes a direct operational bridge between the normative variables found in literature and the practical risk profiles faced by modern municipal managers.

An empirical dataset is not used here. Instead, we use a conceptual design method to analyze the structural characteristics of public administrative software.

Our methodological approach reviews existing public sector frameworks to extract core ethical criteria. We limit our scope to high-risk administrative decisions, such as benefit allocation and risk-scoring audits. This focus matches the regulatory boundaries established by international frameworks like the European Union AI Act [15]. The analysis maps these legal constraints against psychological trust variables. This procedure allows future researchers to replicate our model using quantitative survey methods within specific public organizations.

IV. RESULTS

The conceptual synthesis reveals that citizen trust in automated public services is not a static reaction to technical software performance. Instead, it is an emergent property governed by a three-dimensional normative matrix. The analysis shows that when public institutions replace human administrative discretion with automated scripts, public confidence follows a weakest-link dynamic. Technical optimization fails to secure cooperation if any single protective safeguard experiences a structural deficit. Figure 1 illustrates our modelling and indicates that these clear pillars must act concurrently within a unified public sector ecosystem. If an administrative platform offers excellent data encryption but hides its underlying rule-based processing logic, public rejection occurs immediately. True procedural transparency means translating abstract statistical thresholds into reviewable, plain-language narratives. Equity demands continuous active intervention rather than static initial calibration. Automated scripts run on historical databases, meaning they constantly risk producing past socioeconomic discrimination against marginalized groups. Decisional equity requires public institutions to deploy real-time outcome auditing. These internal protocols track whether automated routing disproportionately flags vulnerable demographics for administrative penalties or social service rejections. Finally, data safety cannot depend on simple consent checkboxes. Our model shows that state data sovereignty requires an absolute legal shield. It demands strict barriers prohibiting any commercial monetization of civic registries. Together, these components turn a closed technical black box into an open democratic channel.

To move beyond loose definitions, we formally establish the *Metric-integrated institutional trust stability model* (MIITS). This operational framework models overall civic trust (T) as a non-linear function dependent on three distinct qualitative variables: Algorithmic transparency index

(x), Decisional equity score (y), and Data security & sovereignty compliance (z).

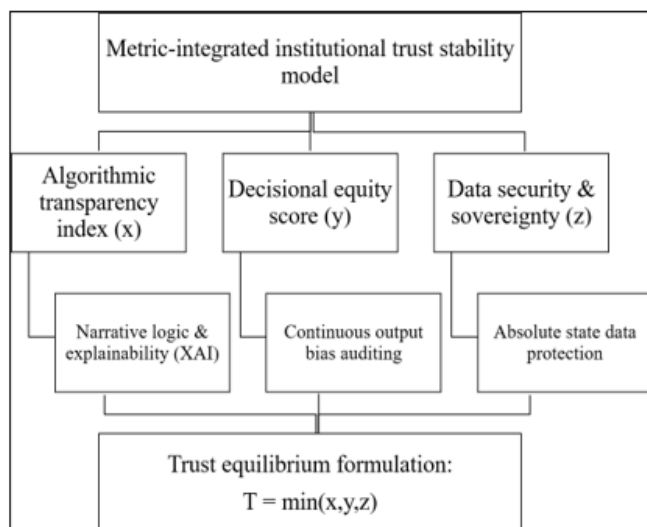


Fig 1 The Conceptual Framework of Trust Equilibrium.

(Source: authors own contribution)

Because trust behaves as an interdependent chain rather than a sum of isolated features, it is governed by a strict limiting constraint. We formulate this dynamic using a standard minimum-value threshold:

$$T = \min(x, y, z),$$

Where each parameter is mapped qualitatively across a boundary scale of [0,1], where 0 represents total institutional failure and 1 denotes complete normative compliance. This mathematical minimum function provides a direct, verifiable answer to our central research question. It demonstrates that the qualitative variables do not influence trust independently, but rather constrain it through a strict limiting relationship.

This formula establishes that the overall stability of digital public services is explicitly capped by the weakest performing safeguard. A public agency cannot compensate for systemic discrimination or data leaks by simply increasing the speed or clarity of its interface.

To establish systemic rigor for public sector deployment, this model isolates the specific administrative failures that occur when individual pillars are compromised. Trust does not scale linearly with isolated infrastructure upgrades. Rather, it is governed by specific structural vulnerabilities. First, the Transparency deficit occurs when x approaches 0. When a public agency maximizes operational processing speed and data security but fails to provide explainable outputs, the citizen experiences an asymmetric distribution of state power. The individual cannot form a coherent administrative defense or appeal an unfavorable act, leaving them alienated by an inscrutable digital bureaucracy.

Next, the Equity deficit occurs when y approaches 0. An explainable system that continually reproduces biased historical distribution patterns simply makes institutional

discrimination more visible. It fails to restore normative trust because it codifies societal inequalities under a false pretense of technological objectivity. Finally, the Security deficit occurs when z approaches 0. Trust collapses completely if citizens suspect that their private data feeds unauthorized state surveillance architectures or aggregates into commercial profiling networks. This specific failure triggers defensive civic behavior, low data compliance, and structural resistance to digital public interfaces.

The MIITS framework provides an exact blueprint to map these qualitative attributes against real-world administrative risk profiles. Public sector applications operate under different levels of algorithmic complexity and human consequence. Therefore, the required performance baseline for each index shifts depending on the administrative context (Table 1).

Our structural mapping reveals that the combination of these three dimensions creates a stabilizing feedback loop. Algorithmic transparency validates the existence of decisional equity by allowing external researchers and legal entities to inspect the variables driving state choices. Simultaneously, strict data sovereignty guarantees that the extensive collection of personal parameters required for high-precision public automation remains isolated from commercial profiling.

Table 1 Qualitative Variable Requirements Across Administrative Environments

Public service context	Primary variable	Qualitative mapping requirement	Empirical indicator
Automated tax auditing	Transparency index (x)	High explainability ($x > 0.8$)	Verifiable narrative justification for tax adjustments
Social welfare allocation	Equity score (y)	Zero demographic shift ($y \rightarrow 1.0$)	Parity in benefit routing across distinct income tiers
Municipal registry access	Security & sovereignty (z)	Full data shield ($z = 1.0$)	Complete separation of civic data from external platforms

(Source: authors own contribution)

The structural utility of the Metric-Integrated Institutional Trust Stability Model (MIITS) extends beyond identifying individual failures. It maps the dynamic feedback loops that occur when these parameters interact inside a live administrative environment. Our modeling shows that algorithmic transparency (x) acts as an indispensable verification channel for decisional equity (y). Without legible text outputs and clear narrative justifications, citizens cannot verify whether an automated systems operates without demographic or socioeconomic drift. Consequently, a high transparency index creates an external audit trail. This

visibility allows independent oversight bodies and civil society to validate the system's actual equity score, transforming public trust from a blind leap of faith into an evidence-informed institutional consensus.

Conversely, a strong data sovereignty shield (z) prevents the behavioural distortions that typically degrade public cooperation during digital transitions. When public sector platforms fail to isolate individual files from commercial profiling or third-party data-mining networks, citizens exhibit defensive compliance strategies. These defensive behaviours include inputs of incomplete datasets, evasion of digital interfaces, and systematic resistance to administrative mandates. The MIITS framework eliminates the underlying anxiety of state surveillance. This baseline security stabilizes the trust equilibrium (T) ensuring that high performance in clarity and fairness translates directly into wider civic compliance and lower systemic friction.

Finally, our model puts emphasis on a critical warning for public sector modernisation strategies> technical optimization cannot patch normative vulnerabilities. Many administrative bodies attempt to recover from a public relations crisis or data leak by investing heavily in intuitive user interfaces or faster automated chatbots. In our model, such investments merely increase processing velocity without affecting the limiting constraint dictated by the minimum function. Because overall trust is bound by the weakest link ($\min(x,y,z)$), operational resources must be directed systematically toward resolving whichever specific pillar experiences a structural deficit. Thus, this balanced mitigation strategy ensures that digital tools serve to reinforce democratic rule-of-law principles rather than obscuring arbitrary administrative power behind sophisticated code.

This protective alignment directly shifts the citizen's perception of the state from an aggressive, data-harvesting surveillance apparatus to a predictable, rule-of-law institution. This way, public managers can move beyond superficial public relations strategies and achieve accountability by design.

V. DISCUSSION

Our conceptual framework demonstrates that software speed cannot replace institutional fairness. Traditional research focuses heavily on infrastructure efficiency. In contrast, our findings highlight a deeper need for relational accountability between citizens and digital state interfaces. This unexpected emphasis suggests that technical optimization might actually decrease trust if it reduces human contact.

Our model compares closely with recent ethical governance studies. It expands previous work by treating transparency as a legal right rather than a technical feature. However, a clear limitation of this study is its lack of empirical data. We propose a conceptual structure, but we do not measure real-world citizen responses in a specific municipal office.

Extending this analysis to the operational layer reveals that the MIITS framework provides an analytical tool to evaluate public administration modernization. Many digital transformation strategies fail because they treat public sector automation as a direct mirror of private sector e-commerce optimization. While private platforms maximize convenience and transaction speed to retain users, public agencies operate under a different constitutional mandate. State software choices mediate legal rights, social security entitlements, and civil statuses. Therefore, applying the MIITS framework forces public managers to shift their core performance indicators from market-driven metrics like user clicks or rapid processing times to democratic markers like procedural explanation and demographic parity.

Furthermore, this framework supports directly the international regulatory compliance mainly under the evolving requirements of the European Union AI Act. The regulation introduces strict boundaries for high-risk automated tools deployed in public management demanding mandatory fundamental rights impact assessments, human oversight loops, and clear transparency protocols. The MIITS model shows how technical teams can implement accountability by design by treating the three pillars as a single limiting function ($T = \min(x,y,z)$). For instance, during the initial procurement phase, public departments can use this model to stress-test third-party applications. If a vendor provides a predictive auditing system with high statistical accuracy but an uninterpretable data layer ($x \rightarrow 0$), the model warns managers that deploying the tool will trigger a severe transparency deficit and erode institutional legitimacy.

Finally, the non-linear weakest-link dynamic identified in this study redefines how public funds should be distributed within government IT budgets. In a typical municipal budget configuration, the vast majority of financial resources are directed toward front-end design, database migration, and cybersecurity infrastructure (z). Meanwhile, post-deployment fairness audits and clear narrative translation modules (x,y) are treated as secondary concerns. The MIITS model proves that this lopsided allocation creates a false sense of security. Because overall public confidence is explicitly bound by the lowest-performing safeguard, building an ironclad digital perimeter around a fundamentally biased or unexplainable automated decision mechanism yields no net gain in civic trust. Hence, organisational leadership must use this framework to realign budgets, systematically sending resources to diagnose and fix whichever qualitative pillar is actively suppressing the trust equilibrium.

This research contributes to the body of knowledge by offering a new understanding of automated public systems. It provides administrators with a direct implementation checklist. We recommend that future researchers test these three pillars through field surveys. Gathering feedback from users interacting with live municipal platforms will validate these specific criteria.

VI. CONCLUSIONS

Resolving the central research question reveals the clear structural boundaries of civic confidence. Technical efficiency cannot serve as a functional substitute for administrative accountability. This paper builds a tridimensional model to address the growing trust deficit in digital public services. We show that converting human bureaucracy into automated code alters the social contract. Efficiency does not guarantee institutional integrity. When automated choices lack clarity, citizens withdraw their trust from public bodies. Our framework connects these issues back to the core principles of administrative law. Notably, governments can protect their institutional legitimacy provided that they manage transparency, equity, and data safety as absolute requirements.

Digital tools must make the state more accessible so they should not hide public choices behind an unreadable layer or code. Replacing a human clerk with an automated script updates the basic social contract between a state and its population. Efficiency cannot substitute for institutional fairness. When software outputs lack clear human explanations, individuals pull away from public offices. Hence, the framework developed in this study provides public administrators with a practical strategy. It offers a structured method to evaluate automated services before wide-scale citizen deployment.

Technological processing moves fast. In contrast, institutional compliance and administrative habits change slowly. This specific structural friction is where civic confidence disappears. Provided they treat transparency, equity, and data security as absolute legal requirements rather than optional software updates, governments can preserve their institutional legitimacy. Future research must now test these conceptual pillars through empirical field studies. Quantitative surveys and comparative case studies within specific agencies will show whether these specific criteria truly improve how the public views automated state power. Digital tools must serve to open up the state. They should never conceal administrative choices behind unreadable lines of code.

REFERENCES

- [1]. Lina Dencik, Anne Kaun; Datafication and the Welfare State. *Global Perspectives* 11 May 2020; 1 (1): 12912. <https://doi.org/10.1525/gp.2020.12912>.
- [2]. Song, Yifan, Takashi Natori, and Xintao Yu. 2024. "Tracing the Evolution of E-Government: A Visual Bibliometric Analysis from 2000 to 2023" *Administrative Sciences* 14, no. 7: 133. <https://doi.org/10.3390/admsci14070133>.
- [3]. J. I. Criado, R. Sandoval-Almazan, and Gil-Garcia, J Ramon, "Artificial intelligence and public administration: Understanding actors, governance, and policy from micro, meso, and macro perspectives," *Public Policy and Administration*, vol. 40, no. 2, pp. 173-184, March 2025. <https://doi.org/10.1177/09520767241272921>
- [4]. Kerstin Hötte, Angelos Theodorakopoulos, Pantelis Koutroumpis, Automation and taxation, *Oxford Economic Papers*, Volume 76, Issue 4, October 2024, Pages 945–969, <https://doi.org/10.1093/oep/gpae006>.
- [5]. Viale Pereira, G., Ciesielska, M., Rizun, N. *et al.* Exploring Adverse Effects from Technological Change in the Context of Smart City Governance. *J Knowl Econ* 17, 5476–5507 (2026). <https://doi.org/10.1007/s13132-025-02875-7>.
- [6]. Murtin, Fabrice, David Pipke, Gianluca Grimalda, Julien Bonnet, Lara Fleischer, Santiago Gonzalez, Vincent Siegerink, et al. "What Explains Trust in Institutions? Multi-Disciplinary Evidence from an International Survey." *European Journal of Political Research*, 2026, 1–25. <https://doi.org/10.1017/S1475676526100875>.
- [7]. Kups, Sarah. "4: Towards evidence-informed policy interventions to increase trust in public institutions". *A Research Agenda for Political Trust*. Cheltenham, UK: Edward Elgar Publishing, 2026. <https://doi.org/10.4337/9781035317486.00009>.
- [8]. C. Castelluccia and D. Le Métayer, "Understanding algorithmic decision-making: Opportunities and challenges," *European Parliamentary Research Service (EPRS)*, Scientific Foresight Unit (STOA), Brussels, Report PE 624.261, March 2019. <https://doi.org/10.2861/536131>
- [9]. S. J. Mikhaylov, M. Esteve, and A. Campion, "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration," *Philos. Trans. R. Soc. A*, vol. 376, no. 2128, pp. 1-18, September 2018. <https://doi.org/10.1098/rsta.2017.0357>
- [10]. Smuha, Nathalie A. "Algorithmic Regulation." Chapter. In *Algorithmic Rule By Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law*, 26–94. Cambridge: Cambridge University Press, 2024. <https://doi.org/10.1017/9781009427500.005>
- [11]. E. Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, 2nd ed. New York: PublicAffairs, 2014, pp. 112-145. Available at: <https://archive.org/details/tosaveeverything0000morozov>
- [12]. O. H. Gandy, "The panoptic sort: A Political Economy of Personal Information," Oxford: Westview Press, June 1993. Retrieved from: <https://truthcloud.net/truthcloud/networks/the-panoptic-sort.pdf>
- [13]. J. Hillo, M. Vento, and T. Erkkilä, "Algorithmic Governance: Experimental Evidence on Citizens' and Public Administrators' Legitimacy Perceptions of Automated Decision-Making," *Public Administration*, vol. 38, September 2025. <https://doi.org/10.1111/padm.70028>
- [14]. Rizk, Aya, Lindgren, Ida. Automated decision-making in public administration: Changing the decision space between public officials and citizens, *Government Information Quarterly*, Volume 42, Issue 3, 2025, 102061, ISSN 0740-624X, <https://doi.org/10.1016/j.giq.2025.102061>

- [15]. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>