Governance and Management Review (GMR)
Volume 9, No. 2, 2025

Rebuilding Trust in Fragile Democracies: Algorithmic Accountability and Citizen Perceptions of AI in Sri Lankan Digital Public Services

[1]Yohan Perera

[1]Senior Lecturer, Department of Human Resources Management, University of Colombo, Sri Lanka yohan@hrm.cmb.ac.lk

Abstract

This study investigates how algorithmic accountability and citizen perceptions shape trust in AI-driven public services within evolving post-crisis governance context in Sri Lanka. Based on 45 semi-structured interviews with government officials, technocrats, and citizens, the research identifies five interrelated challenges: institutional distrust, algorithmic opacity, digital exclusion, cultural disconnect, and the demand for human-mediated accountability. These themes reveal that citizen trust in digital governance is not solely a function of system transparency or efficiency but is profoundly influenced by historical marginalization, socio-cultural conditions, and procedural fairness. Hence, this study advances the concept of Trust-Centric AI Design, a framework grounded in institutional trust theory, which integrates explainability, grievance redressal, and participatory co-design. The findings contribute to regional debates on inclusive e-governance and emphasize the limitations of Global North-derived models in fragile democracies. As Sri Lanka expands its digital transformation agenda, this research offers a context-sensitive roadmap for ethically resilient algorithmic governance.

Keywords:
Algorithmic Accountability, Digital Governance, Institutional Trust, Participatory Design, Public Service Delivery, Sri Lanka

1. Introduction

The evolving landscape of public sector governance in Sri Lanka is undergoing an unprecedented shift following the political and institutional transformations of 2024 (Centre for Policy Alternatives, 2024). Amid widespread public discontent resulting from the 2022 governance crisis and the broader economic disruptions after 2019, the new administration has initiated a swift digital transformation agenda aimed at restoring confidence, improving administrative efficiency, and redefining the state–citizen interface. Digital platforms such as Aswesuma (welfare delivery) and the e-Land Registration System (eLR) exemplify attempts to embed artificial intelligence (AI) into core state functions. Yet, the coexistence of institutional fragility and rapid automation raises pressing questions concerning algorithmic accountability, transparency, and public trust in digital public services (CPA, 2024; Hettiarachchi & Harinda, 2023).

AI adoption in governance is frequently accompanied by techno-optimistic narratives that present automation as a corrective mechanism for bureaucratic inefficiencies while understating

its socio-political implications (Yuan & Chen, 2025; Hanna, 2024). In Sri Lanka, these narratives intersect with pre-existing structural challenges such as compromised institutional legitimacy, entrenched clientelist networks, and pronounced digital disparities between urban and rural populations (Karunasena & Deng, 2011; Erangi & Stecenko, 2023). With only 28% of citizens expressing confidence in government systems (CPA, 2024), AI-driven decisions are being introduced into an environment marked by distrust and skepticism. Insufficiently transparent algorithmic processes such as welfare eligibility filtering or automated tax assessments offer users neither accessible explanations nor meaningful appeal mechanisms (Aysolmaz et al., 2023). Consequently, the Sri Lankan context provides a distinctive empirical setting for examining how algorithmic governance interacts with political legitimacy in post-crisis democracies and whether AI technologies serve to mitigate or amplify existing governance asymmetries (Biju & Gayathri, 2025; Dunleavy & Margetts, 2023; Zhou et al., 2025).

A central concern addressed in this study is the shortage of empirical research on citizen perceptions of algorithmic decision-making systems in South Asian administrative contexts. While substantial scholarship exists on AI ethics, fairness, and explainability within Global North bureaucracies, contexts typically characterized by stable institutions and higher baseline trust (Alhosani & Alhashmi, 2024; Elbehairy et al., 2025; Ferreira & Santos, 2025). These frameworks are not automatically transferable to post-crisis states like Sri Lanka. Here, political instability, socioeconomic vulnerability, and accelerating digital reforms produce a governance environment in which algorithmic outputs are interpreted not merely as technical determinations but as politically consequential decisions embedded in lived experiences of fairness, dignity, and inclusion (Tangi et al., 2025; Sadik-Zada et al., 2022).

Three key scholarly gaps motivate this inquiry. First, algorithmic accountability frameworks in current scholarship remain predominantly Global North-centric (Aysolmaz et al., 2023; Biju & Gayathri, 2025), assuming bureaucratic consistency, accessible legal recourse, and widespread digital literacy (Bright et al., 2024; Elbehairy et al., 2025; Hettiarachchi & Harinda, 2023). In Sri Lanka, state institutions continue to grapple with legitimacy challenges, intensifying the need to contextualize accountability frameworks using indigenous administrative traditions such as Gham Sabha participatory practices (Erangi & Stecenko, 2023). Second, despite the growing role of AI in welfare distribution, taxation, land administration, and service delivery, empirical insights into how digitally marginalized populations perceive and experience algorithmic decisions remain limited (Aysolmaz et al., 2023; Bhopal et al., 2025; Hjaltalin & Sigurdarson, 2024). Existing studies often overrepresent urban and digitally literate segments and overlook the influence of socioeconomic status, geography, and local infrastructure on trust in state technologies (Biju & Gayathri, 2025; Latupeirissa et al., 2024; Singh et al., 2025). Third, there is limited engagement with how AI-mediated governance intersects with post-crisis institutional trust dynamics and historical memory (Bhopal et al., 2025; Yuan & Chen, 2025).

Assumptions that algorithmic transparency alone increases trust underestimate the reality that trust in fragile democracies is historically conditioned and socially negotiated (Levi & Stoker, 2000). In Sri Lanka, citizens often distrust fully automated systems, viewing them as unaccountable and indifferent to nuance. By contrast, confidence tends to be higher when algorithmic decisions remain embedded within hybrid models involving human mediation such as through grievance officers or community-level adjudication (Hettiarachchi & Harinda, 2023). This challenges purely technical framings of trust-building and emphasizes the relational,

communicative, and participatory elements through which legitimacy is constructed (Anshari et al., 2024; Schmager et al., 2024). Institutions such as the South Asian Network on Public Administration (SANPA) advocate context-sensitive governance approaches, yet a notable theoretical disjunction persists between these frameworks and current AI adoption trajectories in the region (Sadik-Zada et al., 2022; Tangi et al., 2025), limiting their influence on digital policymaking (Singh et al., 2025).

Accordingly, this study examines how Sri Lankan citizens and public officials perceive, negotiate, and interpret algorithmic decision-making in public services, and how these interactions shape institutional trust within AI-mediated governance. Positioning Sri Lanka as the focal case, this research advances a localized, post-Westphalian approach to AI governance attentive to political fragility, grassroots accountability, and the affective dimensions of trust. Rather than merely evaluating the technical performance of AI systems, it interrogates the normative frameworks through which digital infrastructures exercise authority (Kopponen et al., 2024; Levi & Stoker, 2000). The analysis is grounded in Institutional Trust Theory and informed by thematic constructs including algorithmic bias, procedural fairness, and legitimacy (Braun & Clarke, 2006; Ferreira & Santos, 2025; Karunasena & Deng, 2011).

Ultimately, this study proposes a Trust-Centric AI Design Framework aligned with Sri Lanka's National Digital Transformation Strategy, emphasizing explainable AI, meaningful grievance redress mechanisms, and participatory audit practices as essential components for restoring democratic legitimacy in AI-mediated public service systems (Alahakoon & Jehan, 2020; Karunasena & Deng, 2011; Dunleavy & Margetts, 2023). Conceptually, this study advances a Sri Lanka-rooted logic of algorithmic accountability that moves beyond Western techno-bureaucratic approaches and foregrounds post-crisis state psychology, localized participatory governance traditions, and culturally embedded trust mechanisms.

2. Literature Review

2.1 Digital Governance and AI in Public Sector Transformations

Public sector transformation through digitalization has been a central policy goal for both developed and developing nations (Anshari et al., 2024; Hanna, 2024). The move toward algorithmically mediated public services is described as part of the "third wave of digital-era governance," where artificial intelligence (AI), machine learning, and data analytics are reshaping institutional logics and administrative architectures (Dunleavy & Margetts, 2023; Ruvalcaba-Gomez, 2023). Digital governance initiatives, especially in developing countries, are increasingly adopting AI-driven tools for service personalization, fraud detection, and efficiency enhancement (Anshari et al., 2024; Tangi et al., 2025; Valle-Cruz & García-Contreras, 2023). Yet, while these transitions aim to modernize the state apparatus, scholars argue that technological sophistication alone is insufficient in contexts marked by institutional fragility, political volatility, and weak civic engagement (Alahakoon & Jehan, 2020; Ferreira & Santos, 2025).

Sri Lanka, like many South Asian states, is undergoing a critical period of digital reform in the aftermath of successive political and economic crises (Hettiarachchi et al., 2023). Erangi and Stecenko (2023) argue that while digital transformation in Sri Lanka has gained momentum through projects, there is minimal emphasis on institutional robustness or participatory design,

thereby raising concerns over sustainability and citizen trust. This is exacerbated by the lack of localized frameworks to guide ethical and accountable AI adoption.

2.2 Algorithmic Accountability: Beyond Technical Fixes

Algorithmic accountability refers to the capacity of AI systems to be explainable, transparent, and subject to public scrutiny (Aysolmaz et al., 2023). In public governance, this becomes critical as decisions made by algorithmic agents affect resource allocation, benefits eligibility, and even civil liberties (Biju & Gayathri, 2025; Yuan & Chen, 2025). Technical solutions such as explainable AI (XAI), audit trails, and decision logs are often posited as mechanisms for accountability (Zhou et al., 2025). However, Biju and Gayathri (2025) contend that these tools often fall short in real-world governance contexts, particularly where bureaucracies lack evaluative mechanisms and citizens have limited digital literacy. One of the key issues with algorithmic accountability in the Global South is that it often gets conflated with technical transparency while ignoring the socio-political structures within which these technologies operate (Schmager et al., 2024; Yigitcanlar et al., 2024).

Sundberg and Holmström (2024) critique this by suggesting that the co-production of domain knowledge and machine learning design is essential to ensure contextual relevance and ethical responsiveness. Without such integration, AI systems may reinforce existing power hierarchies and exacerbate digital injustices (Bhopal et al., 2025; Bright et al., 2024). In Sri Lanka, institutional trust is precarious, and the opacity of algorithmic decision-making has intensified citizen skepticism (Alahakoon & Jehan, 2020). Moreover, Hettiarachchi and Harinda (2023) demonstrate that even among digitally literate populations, AI interventions in welfare or taxation are viewed with suspicion due to the lack of grievance redressal and participatory design. As Sadik-Zada et al. (2022) demonstrate, digital platforms in fragile states are particularly vulnerable to corruption and misuse unless accountability is embedded not only in technology but also in institutional cultures.

2.3 Public Trust and Citizen Perceptions of AI

Trust is not merely a by-product of technological deployment; it is a complex political and affective process (Decuypere & Van De Vijver, 2025). Levi and Stoker's (2000) foundational work on institutional trust emphasizes that citizens evaluate government legitimacy based not only on performance but also on fairness, openness, and perceived accountability. In AI-mediated governance, trust becomes even more contentious as algorithmic processes are often non-transparent and unfamiliar to most citizens (Latupeirissa et al., 2024; Schmager et al., 2024). Empirical studies from the Global North show that algorithmic systems can erode perceptions of fairness and procedural justice, especially when citizens are unaware of how decisions are made (Decuypere & Van De Vijver, 2025). Meanwhile, Aysolmaz et al. (2023), in a large-scale study, found that the lack of human oversight and appeal mechanisms in algorithmic decisions significantly decreased public confidence in government systems.

While these findings are insightful, their applicability to South Asian contexts is limited due to different institutional histories, technological access, and social structures (Elbehairy et al., 2025; Sadik-Zada et al., 2022). In Sri Lanka, trust in digital governance is shaped by a blend of historical marginalization, ethnic fragmentation, and rural-urban divides (Karunasena & Deng, 2011; Erangi & Stecenko, 2023). Meanwhile, Bhopal et al. (2025) emphasize that digital innovations during the post-COVID-19 era have intensified inequalities, especially as

marginalized populations remain excluded from platform-based services due to infrastructure gaps and socio-cultural barriers. Therefore, algorithmic governance must be assessed not only on efficiency metrics but also on its capacity to foster inclusive and context-sensitive trust architectures (Singh et al., 2025; Valle-Cruz & García-Contreras, 2023).

2.4 Citizen-Centric and Ethical AI in Governance

Recent literature calls for a paradigm shift toward citizen-centric AI that prioritizes values of inclusion, responsiveness, and moral accountability (Kopponen et al., 2024; Schmager et al., 2024). Alhosani and Alhashmi (2024) suggest that ethical AI governance requires a broader vision than compliance checklists or data protection laws as it demands embeddedness in participatory policy-making and adaptive institutions. Furthermore, Singh et al. (2025) propose the reimagining of democracy itself in the digital era, where citizens are co-creators of algorithmic frameworks, rather than passive data subjects. One promising direction is the integration of local administrative practices into AI design (Latupeirissa et al., 2024; Yuan & Chen, 2025). In the Sri Lankan context, participatory traditions such as the *Gham Sabhaa* (village-level governance) offer culturally resonant mechanisms for citizen involvement in oversight and accountability (Erangi & Stecenko, 2023).

Meanwhile, Kopponen et al. (2024) and Yigitcanlar et al. (2024) explore how digital twins and personalized AI tools can be adapted to local governance, though caution that without normative grounding, such innovations risk elite capture or technocratic overreach. Moreover, Bright et al. (2024) demonstrate that even in technologically advanced settings like the UK, frontline public officials often use generative AI informally without institutional safeguards, underscoring the need for comprehensive policy architectures. In fragile states like Sri Lanka, this unregulated use can be dangerous, reinforcing systemic biases and procedural opacity (Hettiarachchi & Harinda, 2023; Ruvalcaba-Gomez, 2023).

2.5 Gaps in the South Asian Governance Literature

Despite the increasing deployment of AI in governance across South Asia, region-specific theoretical and empirical frameworks are rare (Ahmad et al., 2023; Anshari et al., 2024). SANPA has called for the development of indigenized governance models, yet few studies explore how algorithmic systems intersect with local administrative logics or trust regimes (Chen et al., 2025; Erangi & Stecenko, 2023). This disconnect is especially troubling in light of the normative assumptions embedded in imported AI tools, many of which are trained on data and institutional logics from the Global North (Elbehairy et al., 2025). As Yuan and Chen (2025) argue, AI accountability frameworks must reflect contextual variables such as fragility, informality, and historical mistrust. In this regard, Sri Lanka offers a fertile ground for advancing the Global South's theoretical contributions to AI governance (Aloshani & Alhashmi, 2024). Its socio-political volatility, combined with an ambitious digital transformation agenda, presents both a cautionary tale and a source of innovation for equitable, accountable governance (Ferreira & Santos, 2025).

3. Methodology

3.1 Research Design

This study employs a qualitative exploratory research design to investigate how algorithmic accountability is perceived, interpreted, and experienced by both implementers and users of AI-

driven public services in Sri Lanka. Given the limited empirical research on this topic in South Asian contexts, particularly in fragile democracies, a qualitative approach is appropriate to explore complex and context-sensitive dimensions such as trust, fairness, institutional legitimacy, and citizen perception (Levi & Stoker, 2000; Braun & Clarke, 2006). The research does not aim for generalizability, but rather to build rich contextual insights and develop conceptual understanding that can inform theory and policy.

3.2 Sampling and Participants

The study adopts purposive stratified sampling to ensure representation from both institutional stakeholders and diverse citizen groups. A total of 45 semi-structured interviews were conducted. 10 public officials working in AI-integrated public service delivery (taxation, welfare, land registry). 5 technocrats and developers from government IT institutions (e.g., ICTA, Lanka Government Network). 30 citizens equally divided between rural and urban residents, and further stratified across age, income level, and digital access. Fieldwork was conducted in Colombo, Kandy, and Batticaloa to ensure geographical diversity, ethnic inclusion, and exposure to different digital infrastructure levels.

3.3 Data Collection

Interviews were conducted between March and May 2025 using a semi-structured format. Interviews were held in Sinhala or English, based on the participant's preference, and later translated and transcribed verbatim. Each session lasted between 45 to 60 minutes. Interviews with government officials and technocrats focused on implementation challenges, ethical safeguards, and transparency mechanisms. Citizen interviews explored themes of procedural fairness, explainability, trust, and perceived risks. All interviews were conducted with informed consent, with assurances of anonymity and confidentiality in line with standard ethical research protocols. Ethics approval was obtained from the researcher's home institution prior to data collection.

3.4 Data Analysis

Data was analyzed using thematic analysis (Braun & Clarke, 2006), allowing for systematic identification of patterns across stakeholder narratives. The coding process was both deductive (based on predefined categories such as bias, trust, and transparency) and inductive (emerging themes from the data). MAXQDA (2024) software was used to organize and visualize code frequencies, co-occurrences, and interrelations. Rigour was maintained through peer debriefing and member checking with select participants.

3.5 Interview Protocol

The semi-structured interview protocol was designed to ensure thematic depth and flexibility while maintaining consistency across participant categories. Two customized guides were developed: one for government officials and technocrats, and another for citizens. For officials, questions explored the integration of AI in service delivery, public communication around algorithmic decisions, institutional accountability mechanisms, and perceptions of public trust and ethical risk. For citizens, the interviews examined lived experiences with digital public services (e.g., MyGov, Aswesuma), perceptions of fairness and transparency, levels of trust in algorithmic systems, and expectations for procedural justice. This dual-path approach enabled

the study to capture multi-stakeholder perspectives on algorithmic governance in Sri Lanka's evolving public administration landscape.

4. Findings

4.1 Overview

This section presents the thematic findings from 45 semi-structured interviews across three respondent groups: government officials, technocrats, and citizens. The data reveal five dominant themes: (1) Institutional Distrust and Democratic Deficits; (2) Algorithmic Opacity and User Disempowerment; (3) Digital Inequality and Rural Exclusion; (4) Preference for Human-Mediated Accountability; and (5) Local Traditions and Participatory Governance as Design Anchors. In order to provide a structured overview of the core themes that emerged from the qualitative data, Table 1 presents a MAXQDA-based coding summary. It outlines the five dominant codes, their frequency across the 45 interviews, associated sub-codes, illustrative quotations, and key thematic insights. This visual summary offers readers a foundation for interpreting the detailed findings that follow in subsequent subsections.

Table 1: MAXQDA-Based Thematic Coding Summary from 45 Interviews

Code (Theme)	Frequency (n=45)	Example Sub-Codes	Participant Quote	Thematic Insight
Institutional Distrust	37	Low trust in government; political bias	"People don't believe the government cares, no matter how modern it is."	Trust is relational, shaped by past political failures
Algorithmic Opacity	33	No explanation; black-box decisions	"The website showed 'ineligible'. No explanation. No one to ask."	Lack of explainability undermines procedural fairness
Digital Inequality	29	Limited access; low digital literacy	"We go to a shop that does internet. We don't know what happens inside."	Digital exclusion mirrors social and geographic divides
Human Accountability	35	Need for human interaction; hybrid models	"A machine doesn't know my struggles. A government officer can listen."	Citizens prefer systems where humans mediate algorithmic decisions

Participatory Governance	31	*Gham Sabhaa*; citizen audits	"*Gham Sabhaa* is not about apps. It's about respect, voice, and listening."	Cultural traditions can anchor ethical, inclusive AI system design

This table displays the frequency, sub-codes, and sample narratives associated with the five central themes identified through qualitative thematic analysis in MAXQDA
Source: Author (2025)

4.2 Institutional Distrust and Democratic Deficits

Across all respondent categories, a recurring narrative was the pervasive *distrust in state institutions*. Public officials acknowledged this legitimacy deficit, attributing it to years of patronage politics, corruption scandals, and economic mismanagement. One senior welfare official stated:

"After 2022, we are dealing with a trust vacuum. People don't believe the government cares, no matter how modern the system is."

Citizen respondents, especially from rural areas expressed deep skepticism toward digital platforms, often linking algorithmic decisions to opaque political agendas. A 36-year-old farmer from Batticaloa remarked:

"They say AI decides, but we know someone behind the curtain pulls the strings. It's not technology. It's politics."

These findings resonate with Levi and Stoker's (2000) argument that political trust is relational and historical, not merely procedural. The introduction of AI into this fragile trust environment did little to enhance legitimacy and, in some cases, deepened the perception of state irresponsiveness.

4.3 Algorithmic Opacity and User Disempowerment

Both citizen and official narratives pointed to a shared lack of *understanding and explainability* in AI-driven decisions. Most citizens who interacted with platforms such as *Aswesuma* or *e-Land Registration System (eLR)* did not know that algorithmic processes were involved. Even among those aware, there was little knowledge of how decisions were made or how to challenge them.

A citizen from Kandy recounted:

"I applied for the benefit. The website showed 'ineligible'. No explanation. No contact person. Nothing."

Technocrats admitted the limited implementation of explainability features due to "budget, speed, and political pressure." As Biju and Gayathri (2025) as well as Decuypere and Van De Vijver (2025) argue, the absence of procedural fairness mechanisms in algorithmic systems leads to user disempowerment and erodes legitimacy. This aligns with Sundberg and Holmström (2024) who caution that absent domain-specific knowledge integration, AI becomes a black-box instrument of exclusion. Several officials emphasized the tension between political urgency and ethical system design:

"We are pushed to deliver fast. But meaningful transparency takes time-interface design, multilingual support, grievance options. These are luxuries right now."

4.4 Digital Inequality and Rural Exclusion

The *digital divide* emerged as a significant theme, especially between rural and urban respondents. While urban users demonstrated some level of comfort navigating digital interfaces, rural citizens, particularly women and elders found the systems confusing and alienating. Digital exclusion was both infrastructural (connectivity, device access) and cognitive (digital literacy, linguistic barriers).

One rural participant noted:

"We usually go to a small shop in the village where someone helps us with online services that they 'do the internet' for us. But we don't really understand what they do on the computer. We just wait, and they tell us the result. We don't see how the decision is made or why..."

This reflects findings by Bhopal et al. (2025) as well as Karunasena and Deng (2011), who highlight how techno-solutionism can reinforce stratified access to state resources. Several citizens expressed fear that their inability to "use the system" would render them permanently excluded from welfare entitlements.

Technocrats acknowledged this gap, but also revealed institutional lethargy:

"We've raised it before how rural people are left behind. But digital equity is not a political priority."

The *absence of inclusive design* is a direct violation of ethical AI principles articulated by Yuan and Chen (2025) as well as Anshari et al. (2024), making digital transformation not just a technical issue, but a democratic one.

4.5 Preference for Human-Mediated Accountability

One of the most compelling findings is the *overwhelming citizen preference for hybrid models* of AI governance that retain some form of human mediation. Interviewees repeatedly emphasized the need for a "person to talk to" when facing algorithmic outcomes. Many distrusted purely automated decisions, regardless of whether they were technically correct. A 42-year-old urban mother in Colombo responded as follows.

"A machine doesn't know my struggles. A government officer can listen, maybe understand. AI cannot explain suffering."

This perception cuts against popular techno-optimistic policy discourse and validates concerns raised by Schmager et al. (2024) and Hettiarachchi and Harinda (2023). Their research suggests that algorithmic decision-making, without a pathway for appeal or dialogic interaction, leads to what they call "citizen alienation through automation." Even among officials, there was a growing consensus that *citizen trust cannot be designed through code alone*. As one Gov.lk administrator explained:

"We cannot expect people to believe in a system that cannot explain itself or apologize."

These findings reinforce Levi and Stoker's (2000) theory that trust is relational, requiring affective reciprocity and accountability. Respondents indicated strong support for mechanisms like community grievance officers, local ombudspersons, or public explainability dashboards, none of which are currently embedded in Sri Lanka's digital governance architecture.

4.6 Local Traditions and Participatory Governance as Design Anchors

Amid critiques of technocratic overreach, many respondents proposed grounding AI development in local governance logics, especially those linked to participatory governance traditions, such as the *Gham Sabhaa* system. These traditions, rooted in collective decision-making, consultation, and face-to-face interaction, were cited as culturally resonant alternatives to imported algorithmic models. A retired senior officer of Sri Lanka Administrative Service (SLAS) responded as follows.

"Gham Sabhaa is not about apps. It's about respect, voice, and listening. Maybe AI should learn from that."

This sentiment aligns with emerging Global South scholarship calling for *indigenized algorithmic accountability* (Erangi & Stecenko, 2023; Singh et al., 2025). Technocrats also acknowledged that international models often overlook contextual realities:

"We copy Estonia or Singapore, but our democracy is different. Our people ask different questions. That matters."

The findings also align with Yigitcanlar et al. (2024) and Kopponen et al. (2024), who argue that AI governance must not only be ethical but *contextually meaningful*, drawing from local administrative heritage rather than abstract universalism. Embedding community audits, village consultations, and multilingual grievance systems were among the suggestions raised by respondents as ways to reconcile *technological modernity with democratic legitimacy.*

4.7 Reframing Trust: A Cornerstone for Democratic AI Governance

These findings converge on a singular insight of **a**lgorithmic governance in fragile democracies must prioritize political trust, not just technical accuracy. In Sri Lankan post-crisis context, trust is shaped by histories of exclusion, institutional opacity, and socio-cultural asymmetries. The mere automation of state functions are certainly efficient, but it cannot substitute for the relational scaffolding that underpins democratic legitimacy. Citizens reject "dark AI" that offers no voice, no redress, and no explanation. What they desire instead is a *trust-centric AI design* in which that includes human mediation, participatory governance principles, local language accessibility, and culturally grounded accountability mechanisms. Hence, these insights reinforce the need to reject Global North techno-solutions that fail to engage with Global South realities (Elbehairy et al., 2025; Yuan & Chen, 2025).

These five themes offer a multidimensional view of how AI-driven public service delivery is understood on the ground, not as neutral technology, but as a *contested political terrain*, where power, identity, and legitimacy are actively negotiated. In order to enhance transparency in the thematic structure and demonstrate empirical grounding, table 2 below summarizes the five core themes identified through qualitative thematic analysis. Each theme is accompanied by

representative codes derived from the coding process and illustrated through anonymized participant quotations. These codes not only capture semantic meanings but also reflect the affective and political layers underlying citizens' and officials' experiences with algorithmic governance. The triangulation of perspectives across citizen users, public officials, and technocrats reinforces the robustness of the analysis and illustrates how themes intersect with institutional trust, procedural fairness, digital inclusion, and cultural legitimacy.

Table 2: Thematic Summary of Key Findings on Algorithmic Accountability and Trust in Sri Lanka's Digital Public Services

Themes were developed through inductive and deductive coding of 45 semi-structured interviews with government officials, technocrats, and citizens.

Theme	Representative Codes / Concepts	Sample Participant Quote
1. Institutional Distrust and Democratic Deficits	Low trust in government; perceived political bias; post-crisis legitimacy collapse	"People don't believe the government cares, no matter how modern the system is."
2. Algorithmic Opacity and User Disempowerment	Black-box systems; no explanations; lack of appeal mechanisms; dark AI'	"The website showed 'ineligible'. No explanation. No contact person. Nothing."
3. Digital Inequality and Rural Exclusion	Rural digital illiteracy; lack of infrastructure; urban-rural divide; third-party access	"We go to a shop that does internet. We don't know what happens inside."
4. Preference for Human-Mediated Accountability	Desire for human contact; mistrust in automated-only systems; demand for hybrid models	"A machine doesn't know my struggles. A government officer can listen."
5. Local Traditions and Participatory Governance	*Gham Sabhaa* ideals; local consultations; need for culturally embedded AI design	"*Gham Sabhaa* is not about apps. It's about respect, voice, and listening."

Source: Author (2025)

5. Discussion

This study advances Institutional Trust Theory by demonstrating how trust in AI is shaped by post-crisis state legitimacy, cultural memory, and grassroots administrative practices unique to South Asian contexts. In doing so, it contributes to a Global South theory of algorithmic legitimacy. Analyzing Sri Lankan post-2024 governance through the lens of institutional trust and citizen perception, the findings nuance dominant Global North discourses on AI in public administration. While much of the existing literature assumes that algorithmic transparency automatically enhances public trust (Decuypere & Van De Vijver, 2025; Yuan & Chen, 2025), this study reveals that trust is contextually constructed, often shaped more by political memory and perceived legitimacy than by system design. Echoing Levi and Stoker's (2000), trust in

public services is relational and historically situated, not merely a product of visible algorithms. Participants often described algorithmic services as alienating, particularly when human oversight was absent or when decisions lacked clear justification. These findings challenge the sufficiency of Explainable AI (XAI) frameworks (Biju & Gayathri, 2025; Sundberg & Holmström, 2024), emphasizing the need for redressal mechanisms and procedural fairness alongside transparency. They affirm Schmager et al.'s (2024) contention that algorithmic trust is co-produced through cultural familiarity and communication, not simply technical auditability.

The findings also resonate with critical scholarship that warns of the unintended harms of AI in post-crisis states. Bright et al. (2024) and Ruvalcaba-Gomez (2023) highlight that digital reforms without institutional safeguards can exacerbate distrust. In line with Sadik-Zada et al. (2022), this study shows how Sri Lankan officials often deploy AI symbolically to signal modernization, despite limited preparedness or oversight. Even efficiency-driven systems like *Aswesuma* risk legitimacy loss when citizens perceive them as opaque or unresponsive. Ferreira and Santos (2025) underline that digital transformation must be accompanied by inclusive governance if it fosters legitimacy which is a theme strongly supported here. Digital exclusion emerged as a critical tension. Bhopal et al. (2025) as well as Karunasena and Deng (2011) argue that digital divides mirror and deepen structural inequalities. Consistent with this, rural Sri Lankan participants relied heavily on unregulated intermediaries, exposing gaps in accessibility, data transparency, and procedural clarity. Citizens frequently expressed disempowerment when decisions were made without consent or explanation. These realities confirm Anshari et al.'s (2024) call to align digital public services with the Sustainable Development Goals (SDGs), particularly equity and inclusiveness.

Notably, the study underscores the relevance of localized, participatory AI ethics. Participants advocated for trust-building mechanisms rooted in local traditions, *Gham Sabhaa* models, community mediation, and human interfaces. These perspectives validate Erangi and Stecenko's (2023) argument that AI must be embedded within culturally intelligible governance models. Singh et al. (2025) similarly urge governments to prioritize democratic digital cultures over imported techno-solutions. The preference for hybrid governance, where algorithmic decisions are mediated by human accountability, aligns closely with Yigitcanlar et al.'s (2024) call for "AI with a human face." Finally, this study reconceptualizes accountability as both a top-down and bottom-up process. While formal frameworks favor audits and legal enforcement (Alhosani & Alhashmi, 2024), citizens demand interpersonal accessibility, grievance channels, and interpretability. This bottom-up accountability echoes Schmager et al.'s (2024) emphasis on social trust as a communicative and participatory construct. As one participant noted, *"We don't need dashboards. We need someone to talk to."* In essence, trust in algorithmic public services cannot be programmed and it must be cultivated through relational, ethical, and culturally grounded systems of engagement.

In order to visually consolidate the core insights of this study, the following conceptual diagram illustrates the interrelated themes that emerged from the field data. Each theme represents a distinct but interconnected challenge in algorithmic governance by ranging from institutional distrust to digital exclusion and collectively informs a central proposition: that public sector AI systems in fragile democracies must be grounded in *Trust-Centric Design* principles. Rather than viewing trust as a technical byproduct of algorithmic transparency, this framework repositions trust as the central democratic condition necessary for the success of digital governance. The

diagram reflects how socio-political, technological, and cultural dynamics converge to either erode or cultivate citizen trust in AI-mediated public services.

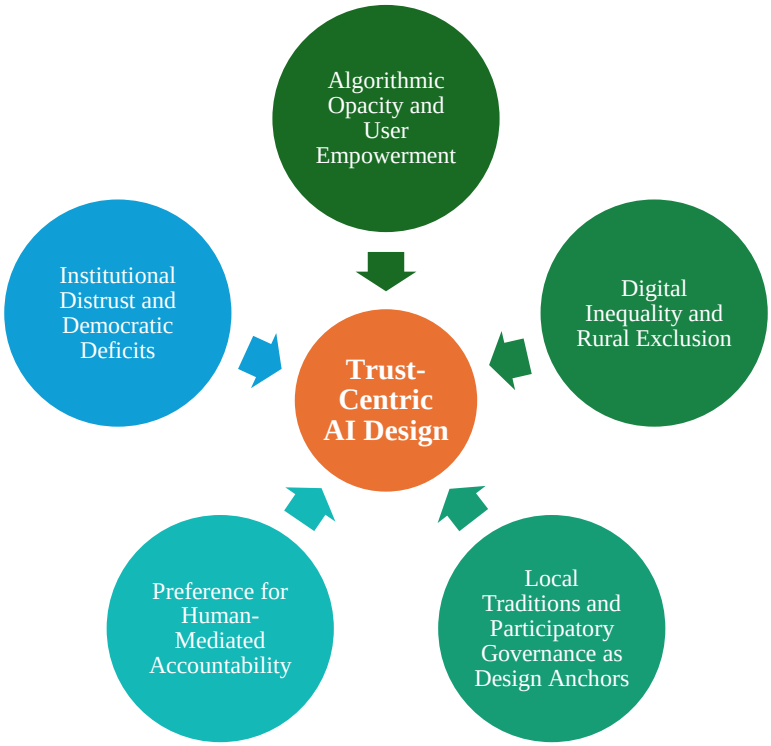

Figure 1: Conceptual Framework — Trust-Centric AI Design in Fragile Democracies

The diagram synthesizes five interrelated themes derived from qualitative fieldwork in Sri Lanka: (1) Institutional Distrust, (2) Algorithmic Opacity, (3) Digital Inequality, (4) Human Accountability Preferences, and (5) Local Participatory Traditions.

6. Conclusion and Policy Implications

This study has illuminated the complicated realities of algorithmic governance in Sri Lanka, where AI adoption intersects with legacies of institutional distrust, social fragmentation, and digital inequality. Following the 2024 political realignment, the state accelerated AI-based service reforms across welfare, taxation, and administration. Yet, as the findings show, these efforts unfold in a governance environment historically shaped by opacity and citizen marginalization. The promise of modernization through AI risks being undermined when algorithmic systems reproduce rather than resolving long-standing governance deficits (Chen et al., 2025). Theoretically, this study contributes to the growing discourse on algorithmic accountability by foregrounding the limitations of Global North-centric governance models. Elbehairy et al. (2025) and Yuan and Chen (2025) emphasize the embedded assumptions in many

AI governance frameworks such as legal enforceability and trust in state institutions which rarely hold in fragile democracies like Sri Lanka. As this study reveals, untransparent decision-making, weak redress mechanisms, and loss of citizen agency are not exceptions but recurring patterns in digital public service platforms in Sri Lanka.

In line with Levi & Stoker's (2000) notion of trust as socially constructed, this study repositions algorithmic legitimacy as a relational outcome by emerging from dialogue, historical memory, and institutional responsiveness. Citizens do not encounter AI as a neutral tool but as an extension of the legitimacy of state, often filtered through past grievances (Hjaltalin & Sigurdarson, 2024). This aligns with Decuypere and Van De Vijver (2025), who argue that procedural fairness, not just performance, shapes public trust in digital systems. In Sri Lanka, citizens demanded not only transparency but also human mediation, accountability, and cultural recognition (Hettiarachchi & Harinda, 2023). These insights call for a move beyond infrastructure-centric reforms. Rather than assuming that digitization itself will rebuild trust, this study proposes a shift toward a participatory, ethical, and context-sensitive design of AI in public governance. Grounded in the lived experiences of citizens and frontline bureaucrats, the *Trust-Centric AI Design* framework introduced here offers a policy blueprint centered on four pillars: (1) participatory accountability, (2) contextual transparency, (3) procedural justice, and (4) human oversight. This framework directly responds to citizen demands for accessible, responsive, and empathetic systems of digital governance.

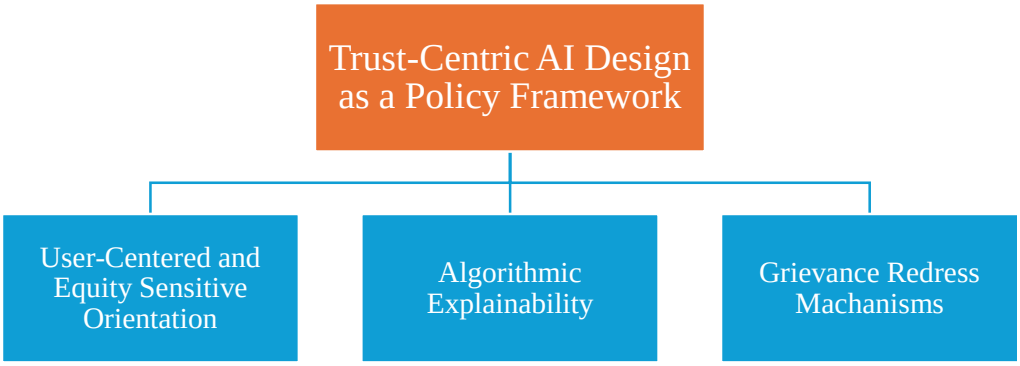

Figure 2: Trust-Centric AI Design as a Policy Framework

This framework synthesizes three essential design pillars, user-centered equity, algorithmic explainability, and grievance redress mechanisms, as guiding principles for AI integration in public service delivery in fragile democratic contexts.

Participants in this study repeatedly referenced Sri Lanka's indigenous *Gham Sabhaa* traditions as offering practical, not merely nostalgic, tools for ethical governance. These localized accountability mechanisms, rooted in consultation and collective redress, resonate with the work of Erangi & Stecenko (2023) and Singh et al. (2025), who advocate for the integration of cultural governance logics into AI policy. However, as noted in the findings, Sri Lanka currently lacks institutional pathways to translate such indigenous models into algorithmic design (Hettiarachchi & Harinda, 2023). This absence represents a critical policy gap, particularly as the state intensifies its commitment to digitization. To operationalize these insights, the following policy implications are proposed.

First, AI systems must be developed through *inclusive co-design processes,* involving local communities, civil society actors, and frontline service workers. This is essential for institutional legitimacy in post-crisis states (Schmager et al., 2024; Karunasena & Deng, 2011). Second, *algorithmic explainability* must go beyond technical disclosure. Visual dashboards, multilingual guidance, and community training can help ensure that non-expert users understand how decisions are made (Alhosani & Alhashmi, 2024; Aysolmaz et al., 2023; Kopponen et al., 2024). Third, every AI-driven platform must incorporate *grievance redress systems* with clear procedures, timelines, and offline access, as emphasized by Sadik-Zada et al. (2022). Without these mechanisms, perceived injustices will intensify distrust. Lastly, policy development must be situated within *existing participatory governance structures,* drawing on models such as community audits or ombudsperson oversight to mediate between algorithmic logic and democratic accountability.

In conclusion, AI adoption in the public sector of Sri Lanka must elevate trust from a secondary outcome to a foundational design principle. This study has demonstrated that in fragile democracies, technological efficiency alone is insufficient to rebuild public confidence or foster democratic legitimacy. Instead, algorithmic systems must be developed within frameworks that acknowledge historical distrust, institutional asymmetries, and cultural expectations of state-citizen interaction. By embedding ethical safeguards, participatory grievance mechanisms, and culturally resonant governance traditions into digital infrastructures is not a technical option which is a democratic necessity. The proposed Trust-Centric AI Design framework offers one such pathway by emphasizing procedural justice, contextual transparency, and human accountability. While grounded in the Sri Lankan experience, these insights contribute to a broader rethinking of algorithmic governance in postcolonial and Global South contexts, where public trust must be actively cultivated rather than assumed. As governments increasingly integrate AI into public administration, the real challenge lies not in coding systems that work, but in designing systems that are trusted, inclusive, and fair.

References

Ahmad, N., Du, S., Ahmed, F., Amin, N. U., & Yi, X. (2023). Healthcare professionals satisfaction and AI-based clinical decision support system in public sector hospitals during health crises: a cross-sectional study. *Information Technology and Management*. https://doi.org/10.1007/s10799-023-00407-w

Alahakoon, M. U. I., & Jehan, S. N. (2020). Efficiency of Public Service Delivery—A Post-ICT Deployment analysis. *Economies*, *8*(4), 97. https://doi.org/10.3390/economies8040097

Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00111-w

Anshari, M., Hamdan, M., Ahmad, N., & Ali, E. (2024). Public service delivery, artificial intelligence and the sustainable development goals: trends, evidence and complexities. *Journal of Science and Technology Policy Management*. https://doi.org/10.1108/jstpm-07-2023-0123

Aysolmaz, B., Müller, R., & Meacham, D. (2023). The public perceptions of algorithmic decision-making systems: Results from a large-scale survey. *Telematics and Informatics*, *79*, 101954. https://doi.org/10.1016/j.tele.2023.101954

Bhopal, R., Nazma, N., & Devi, R. (2025). Post-COVID-19: leveraging digital payment services for a sustainable future through bibliometric analysis. *IIMT Journal of Management*. https://doi.org/10.1108/iimtjm-07-2024-0077

Biju P.R., & Gayathri O. (2025). Algorithmic solutions, subjectivity and decision errors: a study of AI accountability. *Digital Policy, Regulation and Governance* 24 June 2025; 27 (5): 523–552. https://doi.org/10.1108/DPRG-05-2024-0090

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bright, J., Enock, F., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector: evidence from a survey of UK public sector professionals. *Digital Government Research and Practice*. https://doi.org/10.1145/3700140

Centre for Policy Alternatives (CPA) (2024). *Statement on the Online Safety Act No. 09 of 2024.* Centre for Policy Alternatives. https://www.cpalanka.org/statement-on-the-online-safety-act-no-09-of-2024/

Chen, X., Ge, E., Xu, X., & Zhou, Q. (2025). Does digitalization of government activities improve business environment? The influence of public service standardization. *Economic Analysis and Policy*, *87*, 533–560. https://doi.org/10.1016/j.eap.2025.05.034

Decuypere, A., & Van De Vijver, A. (2025). AI: Friend or foe of fairness perceptions of the tax administration? A survey experiment on citizens' procedural fairness perceptions. *Government Information Quarterly*, *42*(1), 102002. https://doi.org/10.1016/j.giq.2024.102002

Dunleavy, P., & Margetts, H. (2023). Data science, artificial intelligence and the third wave of digital era governance. *Public Policy and Administration*, *40(2),* https://doi.org/10.1177/09520767231198737

Elbehairy, F., Mora, L., & Soe, R. (2025). Who are you? Examining the multifaceted innovation roles of municipal governments in AI governance. *Technovation*, *147*, 103303. https://doi.org/10.1016/j.technovation.2025.103303

Erangi, D. H. N., & Stecenko, I. (2023). Digital Transformation in Sri Lanka: Evaluating Progress and Addressing Challenges. *Green Blue and Digital Economy Journal*, 5(2), 1–11. https://doi.org/10.30525/2661-5169/2024-2-1

Ferreira, A., & Santos, C. (2025). Digital transformation in public sector. In *IGI Global eBooks* (pp. 265–288). https://doi.org/10.4018/979-8-3693-9251-5.ch008

Hanna, N. K. (2024). *Driving digital transformation in developing nations*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-54573-3

Hettiarachchi, N. & Harinda L., Public Awareness, Usage of E-government Services in Sri Lanka (2023). JETIR August 2023, Volume 10, Issue 8, Available at SSRN: https://ssrn.com/abstract=4552475

Hjaltalin, I. T., & Sigurdarson, H. T. (2024). The strategic use of AI in the public sector: A public values analysis of national AI strategies. *Government Information Quarterly*, *41*(1), 101914. https://doi.org/10.1016/j.giq.2024.101914

Karunasena, K., & Deng, H. (2011). Critical factors for evaluating the public value of e-government in Sri Lanka. *Government Information Quarterly*, *29*(1), 76–84. https://doi.org/10.1016/j.giq.2011.04.005

Kopponen, A., Hahto, A., Villman, T., Kettunen, P., Mikkonen, T., & Rossi, M. (2024). Personalised public services powered by AI: the citizen digital twin approach. In *Edward Elgar Publishing eBooks* (pp. 170–186). https://doi.org/10.4337/9781802207347.00020

Levi, M., & Stoker, L. (2000). *Political Trust and Trustworthiness.* Annual Review of Political Science, 3, 475–507. https://doi.org/10.1146/annurev.polisci.3.1.475

Latupeirissa, J. J. P., Dewi, N. L. Y., Prayana, I. K. R., Srikandi, M. B., Ramadiansyah, S. A., & Pramana, I. B. G. A. Y. (2024). Transforming Public Service Delivery: A Comprehensive Review of Digitization Initiatives. *Sustainability*, *16*(7), 2818. https://doi.org/10.3390/su16072818

Ruvalcaba-Gomez, E. A. (2023). Systematic and axiological capacities in artificial intelligence applied in the public sector. *Public Policy and Administration*. https://doi.org/10.1177/09520767231170321

Sadik-Zada, E. R., Gatto, A., & Niftiyev, I. (2022). E-government and petty corruption in public sector service delivery. *Technology Analysis and Strategic Management*, *36*(12), 3987–4003. https://doi.org/10.1080/09537325.2022.2067037

Schmager, S., Grøder, C. H., Parmiggiani, E., Pappas, I., & Vassilakopoulou, P. (2024). Exploring citizens' stances on AI in public services: A social contract perspective. *Data & Policy*, *6*. https://doi.org/10.1017/dap.2024.13

Singh, B., Jain, A., Wongmahesak, K., & Chandra, S. (2025). Reimagining democracy in the digital era. In *IGI Global eBooks* (pp. 1–18). https://doi.org/10.4018/979-8-3693-8749-8.ch001

Sundberg, L., & Holmström, J. (2024). Fusing domain knowledge with machine learning: A public sector perspective. *The Journal of Strategic Information Systems*, *33*(3), 101848. https://doi.org/10.1016/j.jsis.2024.101848

Tangi, L., Müller, A. P. R., & Janssen, M. (2025). AI-augmented government transformation: Organisational transformation and the sociotechnical implications of artificial intelligence in public administrations. *Government Information Quarterly*, *42*(3), 102055. https://doi.org/10.1016/j.giq.2025.102055

Valle-Cruz, D., & García-Contreras, R. (2023). Towards AI-driven transformation and smart data management: Emerging technological change in the public sector value chain. *Public Policy and Administration*. https://doi.org/10.1177/09520767231188401

Yigitcanlar, T., David, A., Li, W., Fookes, C., Bibri, S. E., & Ye, X. (2024). Unlocking Artificial Intelligence Adoption in Local Governments: Best Practice Lessons from Real-World Implementations. *Smart Cities*, *7*(4), 1576–1625. https://doi.org/10.3390/smartcities7040064

Yuan, Q., & Chen, T. (2025). Holding AI-Based Systems Accountable in the Public Sector: A Systematic review. *Public Performance & Management Review*, 1–34. https://doi.org/10.1080/15309576.2025.2469784

Zhou, Z., Liu, D., Chen, Z., & Pancho, M. (2025). Research on government adoption of generative artificial intelligence and ambidextrous innovation drivers. *International Review of Economics & Finance*, 103953. https://doi.org/10.1016/j.iref.2025.103953