Conversational AI in Public Service Delivery: Trust, Accessibility, Accountability, and the Communication Imperative in Citizen–Institution Interaction

Vanja Stojković*

Deputy Director, National Employment Service of Republic Serbia

Abstract: Public institutions are among the most consequential contexts for the deployment of conversational artificial intelligence. When AI systems mediate communication between citizens and government employment agencies, healthcare providers, social welfare offices, and administrative bodies, the stakes of miscalibrated trust, inadequate interpretability, and poor interaction efficiency extend beyond individual user experience to include access to rights, benefits, and services. This paper examines the specific challenges and opportunities of deploying conversational AI in public service communication contexts and develops the Public Service Conversational AI (PSCAI) framework as both a conceptual and an implementation-oriented model. The revised framework links citizen-facing dialogue interfaces to authoritative knowledge bases, case-management systems, human escalation pathways, audit logs, and continuous feedback mechanisms. It further provides practical guidance for applying the framework in real government service environments such as employment services, welfare benefit guidance, healthcare administration, and municipal service portals. We argue that public sector conversational AI requires a fundamentally different design philosophy from commercial AI: one centered on institutional accountability, citizen dignity, legal accuracy, equitable access, and appealable human oversight rather than pure efficiency optimization.

Keywords: Public service AI, Conversational AI, Citizen–institution communication, Trust calibration, Accountability, Accessibility, AI governance, Human-machine communication, Public administration, Employment services, Public service implementation, AI service architecture.

1. INTRODUCTION

In 2023, an unemployed citizen in a European country applied for jobseeker's allowance through an AI-powered conversational interface deployed by the national employment service. The system, trained on policy documents and procedural manuals, provided detailed guidance on eligibility criteria, documentation requirements, and application timelines. The citizen, following the AI's instructions precisely, submitted a set of documents that turned out to be incomplete under a recent regulatory change that the system's training data did not yet reflect. The application was rejected. The rejection triggered a financial crisis. The AI had communicated with confidence and fluency; the citizen had trusted it appropriately given all available signals; the consequences were severe.

This scenario—composite but empirically grounded in patterns documented across AI deployments in public administration (Saxena *et al*., 2021; Molina & Sundt, 2023)—illustrates the distinctive stakes of conversational AI in public service communication. The public sector presents the most consequential and the most demanding context for AI-mediated communication: consequential because the outputs of government–citizen interactions directly affect access employment; demanding because it combines the to rights, benefits, healthcare, housing, and complexity of natural language interaction with the precision requirements of legal and regulatory compliance, the diversity of the citizen population, and the accountability expectations of democratic governance.

Yet public sector AI deployment has accelerated rapidly, often driven by efficiency imperatives that do not fully account for the distinctive communication challenges of citizen-facing institutions. Many deployments have been undertaken with commercial AI communication tools developed primarily for customer service contexts, without adequate adaptation to the trust, accountability, and equity requirements of public service. The consequences have included documented failures of AI systems in welfare benefit administration, employment guidance, and health information services that have produced tangible harms to vulnerable citizens (Eubanks, 2018; Saxena *et al*., 2021; Molina & Sundt, 2023).

This paper addresses the design and governance principles required for responsible conversational AI deployment in public service communication. Its core contribution is the PSCAI framework, which connects public-administration values to concrete AI communication system components: service intake, verified knowledge retrieval, response generation, legal accuracy checking, human escalation, audit logging, and citizen feedback integration. Rather than treating conversational AI as a generic chatbot layer, the framework positions it as a governed communication

*Address correspondence to this author at the Deputy Director, National Employment Service of Republic Serbia;
E-mail: vanjastojkovic988@gmail.com

Received on 20-04-2026 Accepted on 29-05-2026 Published on 05-07-2026

E-ISSN: 2979-1081/26

infrastructure embedded within real administrative workflows.

The paper is organized as follows. Section 2 reviews the distinctive features of citizen-institution communication. Section 3 surveys empirical evidence and adds practical examples from government service environments. Section 4 presents the PSCAI framework, including a conceptual implementation model. Section 5 addresses governance requirements, Section 6 provides evaluation metrics and a governance checklist, and Section 7 concludes.

2. THE DISTINCTIVE COMMUNICATION CONTEXT OF PUBLIC SERVICES

2.1. Power Asymmetry and Vulnerability

The most distinctive feature of citizen–institution communication is the fundamental power asymmetry between the parties. Government agencies hold decision-making authority over access to benefits, services, legal status, and in some cases liberty. Citizens approaching public institutions are often in positions of economic vulnerability, emotional distress, or informational disadvantage. This power asymmetry has direct implications for trust calibration in AI-mediated communication: citizens interacting with government AI systems may feel compelled to trust AI outputs even when they have doubts, because challenging institutional authority carries perceived costs that do not apply in commercial contexts.

The vulnerability dimension of public service communication has been extensively documented in research on administrative burden (Herd & Moynihan, 2018) and digital exclusion (Ragnedda & Ruiu, 2017). Administrative burden—the costs of time, effort, stress, and documentation that citizens bear when accessing public services—disproportionately affects those in greatest need: the elderly, the poorly educated, non-native speakers, and those experiencing mental health challenges. Conversational AI systems have the potential to substantially reduce administrative burden by providing accessible, patient, multilingual guidance—but poorly designed systems can also amplify burden by adding confusing technological intermediaries between citizens and the services they need.

2.2. Legal and Regulatory Precision Requirements

Public service communication occurs within a framework of legal rights and procedural obligations that demands precision not typically required in commercial communication contexts. A customer service AI that gives slightly imprecise product information risks a customer inconvenience; a benefits administration AI that gives imprecise eligibility guidance can deprive a citizen of income, housing, or healthcare. This precision requirement creates a fundamental tension with the probabilistic, fluency-optimized nature of LLM output, which can generate plausible-sounding but legally inaccurate guidance with high confidence and no warning.

Research on AI deployment in welfare administration (Eubanks, 2018; Saxena *et al*., 2021) and employment services (Molina & Sundt, 2023) documents systematic failures in which AI systems provided citizens with incorrect procedural guidance, failed to update when regulations changed, and could not be held accountable for the consequences of their errors in ways that human case workers could. These failures suggest that public sector AI deployment requires fundamentally different accuracy and accountability standards than commercial deployment.

2.3. Equity and Non-Discrimination Obligations

Public sector institutions operate under legal and constitutional non-discrimination obligations that have direct implications for AI system design. AI systems trained on historical data from public institutions risk encoding and amplifying existing patterns of discriminatory treatment—favoring citizens who have historically been better served by the institution, penalizing those who have faced barriers, and replicating the implicit assumptions and values embedded in historical administrative practice (Eubanks, 2018; Noble, 2018; Wachter *et al*.,. 2021). The deployment of AI in public service communication therefore requires not only general bias mitigation but specific auditing against the legally protected characteristics and equity obligations of the public institution.

2.4. The Public Trust Dimension

Citizens' trust in government institutions is a foundational public good with effects on democratic participation, social cohesion, and institutional legitimacy (Citrin & Stoker, 2018). The deployment of AI systems in citizen-facing public services is not merely an operational decision but a trust governance decision: it signals to citizens how the institution values the quality of its communicative relationship with them. Research on public trust in government AI (Saxena *et al*., 2021; Wirtz *et al*., 2019) finds that perceived procedural fairness, transparency, and dignity of treatment are stronger predictors of public sector AI trust than perceived efficiency or accuracy. This finding has important design implications: efficiency optimization at the expense of communicative dignity or procedural transparency may produce short-term

Table 1:

Dimension	Traditional Public Service	Rule-Based Chatbot	LLM-Based Conversational AI
Communicative Flexibility	High (human judgment)	Low (scripted trees)	High (natural language understanding)
Availability	Office hours only	24/7	24/7, multilingual
Personalization	Case worker discretion	Very limited	Context-aware, adaptive
Accountability	Clear chain	Rule audit trail	Complex; requires explicit design
Trust Calibration	Relationship-based	Minimal trust management	Requires explicit engineering
Equity Risks	Human bias	Encoding of rule bias	Amplification of training data bias
Error Redress	Supervisor / appeal	Defined escalation path	Must be explicitly designed

operational gains while eroding the public trust foundations on which institutional legitimacy depends.

2.5. Comparative Overview

The table above illustrates how LLM-based conversational AI differs from both traditional public service delivery and earlier rule-based chatbots across dimensions critical to the public sector context. Each distinguishing advantage of LLM-based systems is accompanied by a specific governance challenge that the PSCAI framework addresses.

3. EMPIRICAL EVIDENCE ON AI-MEDIATED PUBLIC SERVICE COMMUNICATION

3.1. Documented Benefits

Empirical studies of AI deployment in public service contexts document several significant benefits when systems are appropriately designed and governed. Research on AI-assisted employment services (Molina & Sundt, 2023; Desiere & Struyven, 2021) finds that well-designed conversational AI systems can substantially reduce first-contact resolution times, extend service availability to non-business hours, and provide more consistent initial guidance than human case workers under high workload conditions. In healthcare contexts, AI-mediated patient communication has been associated with improved appointment adherence, more effective pre-consultation preparation, and reduced administrative burden for patients with literacy challenges (Cai *et al*., 2023).

Accessibility benefits are among the most compelling. AI systems capable of real-time language translation and register adaptation can serve citizen populations that have historically been underserved by monolingual, technical public service communication. Studies of AI-mediated immigration services (Saxena *et al*., 2021) and healthcare communication (Cai *et al*., 2023) document significant access improvements for non-native speaker populations when AI systems are designed with linguistic accessibility as a primary design target.

3.2. Documented Failures and Risks

Against these benefits, the empirical literature documents a series of significant failures and risks that must inform PSCAI design. In automated benefits administration, AI systems have produced both false denials—incorrectly identifying citizens as ineligible—and false approvals—directing citizens through application processes that will ultimately be rejected, wasting time and raising expectations (Eubanks, 2018; Saxena *et al*., 2021). These errors are particularly costly when systems do not provide citizens with clear explanations of decisions or accessible pathways for review and appeal.

Procedural knowledge obsolescence is a documented failure mode specific to the public sector context. LLM-based systems with static training data cannot be kept current with the rapid pace of regulatory change in complex welfare and employment systems. Studies document multiple instances in which citizens acted on AI-provided guidance that reflected superseded regulations, with consequential harms to benefit access (Molina & Sundt, 2023). This failure mode requires continuous knowledge update mechanisms and explicit uncertainty flagging when system confidence in regulatory currency is low.

Equity impacts of AI public service systems have been documented across multiple dimensions. Research by Eubanks (2018) and Noble (2018) demonstrates that AI systems trained on historical administrative data from systems characterized by racial, socioeconomic, and gender inequities reproduce and in some cases amplify those inequities in their outputs. Employment AI systems have been shown to rate candidates from disadvantaged backgrounds less favorably, direct users toward services less appropriate to their needs, and apply eligibility criteria inconsistently across demographic groups.

3.3. Practical Application Examples in Government Service Environments

The practical relevance of PSCAI can be illustrated through common government service environments in which conversational AI is already technically feasible but institutionally sensitive. In a public employment service, a citizen-facing AI assistant may explain registration procedures, prepare a document checklist, translate eligibility language into plain speech, and route complex cases to a human counsellor. However, it should not make a final determination of benefit eligibility or sanction status unless the decision is fully traceable, legally verified, and subject to human review.

In social welfare administration, a PSCAI-compliant system may help citizens identify which benefits are potentially relevant, pre-fill forms using user-confirmed information, and warn when a case involves recent regulatory changes. The system should maintain a clear boundary between guidance and decision-making: it may guide a citizen toward the appropriate application path, but an accountable public official must remain responsible for consequential outcomes such as denial, suspension, or recovery of benefits.

In healthcare administration, conversational AI can support appointment scheduling, rights information, pre-visit preparation, and navigation of insurance or reimbursement procedures. Clinical advice, emergency triage, or contested entitlement questions must be escalated to qualified human staff. In municipal services, similar tools can support citizens applying for permits, certificates, tax information, or local-service complaints, provided that each answer is linked to current official rules and an auditable interaction record.

4. THE PSCAI FRAMEWORK

The Public Service Conversational AI (PSCAI) framework adapts the general principles of responsible AI communication design to the distinctive requirements of citizen-facing public service contexts. It is organized around six design pillars that address the specific accountability, equity, dignity, and accessibility obligations of public sector institutions.

4.1. Technical Architecture and Conceptual Model of the PSCAI Framework

For applied public sector deployment, the PSCAI framework should be understood not only as a set of ethical principles but also as a communication-system architecture. A PSCAI-compliant system begins with a citizen channel, such as a web portal, mobile application, kiosk, voice interface, or messaging channel. The intake layer identifies the citizen intent, accessibility needs, language requirements, and service domain. The conversational AI layer then generates a response, but only after grounding the answer in an authoritative and versioned knowledge layer containing current laws, administrative procedures, service catalogues, and institutional rules. The output is routed through accountability controls: legal accuracy checks, uncertainty flags, audit logging, and human escalation triggers.

The architecture shown in Figure **1** emphasizes that public service conversational AI should not operate as a stand-alone chatbot. It must be integrated with

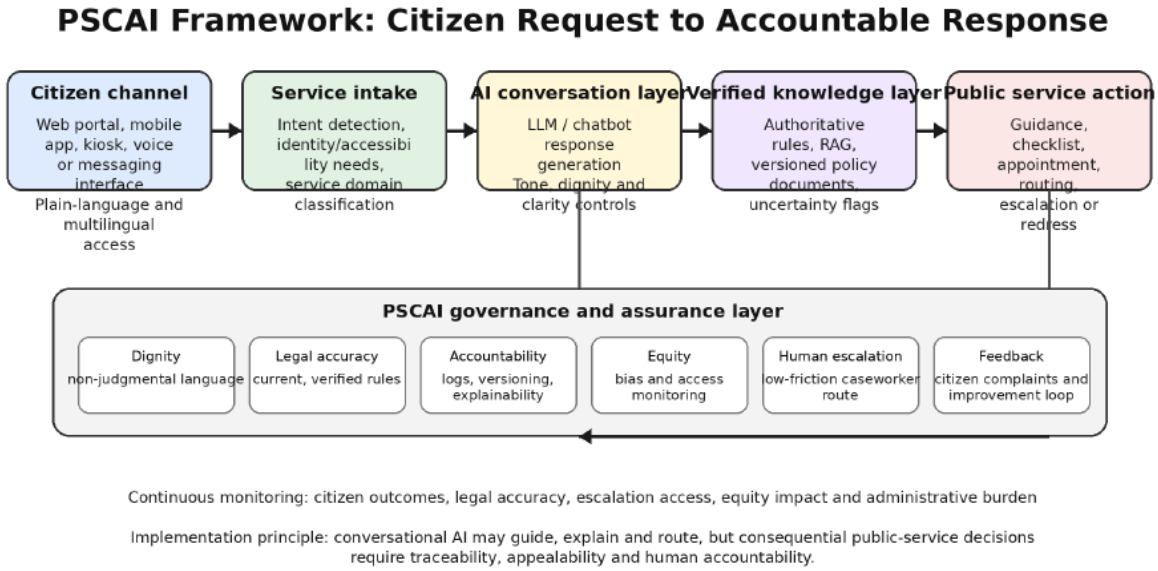

Figure 1: Conceptual implementation model of the PSCAI framework for citizen-facing conversational AI in public service delivery.

case-management systems, legally verified knowledge repositories, identity and accessibility services where appropriate, escalation queues, and institutional monitoring dashboards. This integration is essential because many public service interactions are not merely informational: they may influence access to legal rights, employment support, healthcare, social assistance, or administrative remedies.

4.2. Pillar 1: Dignity-Centered Communication Design

Dignity-centered communication design establishes that every interaction with a public service AI must honor the citizenship status and human dignity of the person it serves, regardless of their reason for contact, their administrative history, or the outcome of their request. Concretely, this requires: (a) non-judgmental language design that avoids stigmatizing framing in responses to sensitive requests (benefit applications, debt queries, unemployment claims); (b) communicative patience mechanisms that allow users to ask the same question multiple times, in different formulations, and at different points in the interaction without triggering dismissive or impatient response patterns; (c) respectful refusal design that delivers negative decisions, limitations, and redirections in ways that acknowledge the human significance of the denial without sacrificing accuracy or honesty.

Dignity-centered design is not merely an ethical aspiration—it has practical implications for trust and system effectiveness. Research consistently shows that citizens who experience public service interactions as undignified—whether from human officials or AI systems—are less likely to provide accurate information, more likely to abandon applications, and more likely to develop long-term avoidance of the services they need (Herd & Moynihan, 2018; Wirtz *et al*., 2019).

4.3. Pillar 2: Legal Accuracy and Knowledge Currency

The PSCAI framework requires that public service AI systems be engineered with explicit mechanisms for legal accuracy assurance and knowledge currency maintenance. This includes: (a) structured knowledge update protocols that integrate regulatory changes into system knowledge on a defined schedule with mandatory accuracy verification before deployment; (b) temporal uncertainty flagging that explicitly communicates to citizens when provided guidance may be affected by recent regulatory changes and directs them to verification resources; (c) confident-scope definition that clearly delineates the domains within which the system can provide legally reliable guidance and those requiring human case worker consultation.

These requirements have significant engineering implications. Unlike commercial AI deployment where knowledge currency is a quality-of-service issue, in public service contexts knowledge obsolescence is a safety-critical failure mode with potential legal liability. PSCAI systems should be designed with knowledge update mechanisms at least as robust as those applied to safety-critical software systems.

4.4. Pillar 3: Transparent Accountability Architecture

Every consequential output of a public service AI must be traceable to a specific system version, knowledge state, and decision rule, and must be attributable to a defined institutional accountability chain. This requires: (a) immutable interaction logging that preserves a complete, tamper-evident record of citizen–AI interactions for audit, review, and redress purposes; (b) decision explainability that can render any AI output into a form accessible to the citizen affected and to oversight bodies; (c) defined human accountability that establishes which institutional role is responsible for the AI system's outputs at each point in its operation, including during both design and deployment phases.

Transparent accountability architecture also requires that citizens be clearly informed of the AI-mediated nature of their interaction at the outset and throughout. Research on public trust in government AI finds that citizens who are not clearly informed of AI mediation and later discover it report sharply reduced institutional trust, while those who are clearly informed and receive a high-quality interaction report trust levels comparable to human-mediated service (Wirtz *et al*., 2019).

4.5. Pillar 4: Equity-Aware Design and Monitoring

PSCAI systems must be designed with explicit equity awareness and subjected to ongoing equity monitoring against the non-discrimination obligations of the deploying institution. Design requirements include: (a) bias auditing against legally protected characteristics before deployment and at defined intervals thereafter; (b) differential impact analysis that examines whether AI-mediated service produces systematically different outcomes for different demographic groups; (c) accessibility-first design that treats linguistic accessibility, plain language, and accommodations for cognitive and communicative challenges as primary requirements rather than optional additions.

Equity monitoring must be institutionalized as an ongoing governance function, not a one-time pre-deployment audit. The dynamic nature of both LLM

behavior and the citizen populations being served means that equity impacts can emerge over time even in systems that passed initial audits. Institutions deploying PSCAI systems should establish dedicated equity monitoring functions with authority to suspend system operations when monitoring identifies significant discriminatory impacts.

4.6. Pillar 5: Accessible Human Escalation

No public service AI system should be designed as a terminal communication channel. Every citizen interaction must include a clearly accessible, low-friction pathway to human case worker escalation, and this pathway must be available without preconditions, fees, or navigational complexity. Human escalation is not a failure mode to be minimized—it is a right of every citizen and a design requirement of every public service AI system.

Research on service escalation behavior (Herd & Moynihan, 2018) shows that barriers to human escalation—including complex navigation, long wait times, or social stigma associated with requesting human assistance—disproportionately disadvantage the citizens most in need of human judgment. Accessible human escalation design must therefore account for the full range of barriers that different citizen populations face, including digital literacy, language, disability, and time constraints.

4.7. Pillar 6: Continuous Citizen Feedback Integration

Public service AI systems should be designed with institutionalized mechanisms for continuous citizen feedback integration. Unlike commercial systems where user feedback is primarily a quality improvement signal, in public service contexts citizen feedback is also a democratic accountability mechanism: it provides the institution with evidence about whether its AI-mediated communication is serving the citizenship it is obligated to serve. Feedback mechanisms should be accessible, multilingual, designed to reach marginalized users who are least likely to voluntarily engage with standard feedback channels, and should be reviewed by accountable human officials at defined intervals.

4.8. Practical Implementation Guidance for Public Institutions

A practical PSCAI deployment should proceed through a staged implementation model rather than immediate full-scale automation. First, the institution should define a narrow service domain and classify citizen inquiries into low-risk informational guidance, medium-risk procedural support, and high-risk consequential cases requiring human review. Second, all AI responses should be grounded in an authoritative, version-controlled knowledge base, preferably using retrieval-augmented generation combined with rule-based legal constraints. Third, every interaction should include visible escalation options and a clear statement of the system's scope limitations. Fourth, the deployment should begin as a pilot with monitored performance across legal accuracy, escalation access, equity impact, citizen dignity, and administrative burden before expansion to additional services.

5. GOVERNANCE AND POLICY RECOMMENDATIONS

The PSCAI framework has governance implications at multiple levels of public administration. At the institutional level, public sector organizations deploying conversational AI must establish dedicated AI governance functions with cross-disciplinary composition, including legal, equity, technology, service design, data protection, and frontline case-work expertise. These functions should approve the service scope, maintain the official knowledge base, define

Table 2: Practical Application of the PSCAI Framework in Real Public Service Environments

Service environment	Typical AI-supported interaction	Required safeguard	Human accountability point
Employment service	Jobseeker registration guidance, document checklist, appointment routing, training-program information	Current labour-market and benefit rules; uncertainty flag for recent regulation changes	Employment counsellor or case officer verifies eligibility and sanctions
Social welfare office	Benefit navigation, form preparation, missing-document alerts, plain-language explanation of rights	No final denial by AI; appeal and review information must be displayed	Social worker or benefits officer remains responsible for entitlement decisions
Healthcare administration	Appointment scheduling, pre-visit information, insurance/reimbursement procedure guidance	Clinical and emergency questions escalated immediately; medical scope clearly limited	Qualified health professional or administrative supervisor reviews high-risk cases
Municipal service portal	Permits, certificates, local tax inquiries, complaint intake, service status updates	All answers linked to official municipal rules and logged for audit	Designated municipal officer signs off on final administrative action

escalation thresholds, review citizen complaints, and suspend the system if legal accuracy, equity, or dignity metrics fall below acceptable thresholds.

At the regulatory level, national and supranational frameworks—including the EU AI Act, which classifies certain public sector AI applications as high-risk—establish minimum requirements for transparency, accuracy, and human oversight in AI systems deployed by public authorities. PSCAI framework compliance should be designed to meet or exceed these requirements, not merely to satisfy minimum compliance thresholds.

At the professional level, public administrators and service designers require specific training in AI communication design and governance, including the distinctive requirements of public sector contexts. Professional development frameworks for public servants should incorporate AI literacy as a core competency, with particular attention to the accountability and equity dimensions that distinguish public sector AI deployment from commercial contexts.

6. EVALUATION METRICS AND GOVERNANCE CHECKLIST

The following evaluation metrics and governance checklist are designed specifically for public sector conversational AI contexts and reflect the six pillars of the PSCAI framework.

6.1. Core Evaluation Metrics

- First Contact Resolution Rate (FCRR): proportion of citizen inquiries resolved accurately in a single AI-mediated interaction, benchmarked against human case worker FCRR and against legal accuracy verification.
- Legal Accuracy Rate (LAR): proportion of AI-provided guidance that is legally accurate at the time of provision, assessed through expert review sampling and regulatory currency audits.
- Equity Differential Index (EDI): ratio of service quality and accuracy metrics across protected demographic groups, with threshold alerts when differential exceeds pre-defined equity bounds.
- Escalation Access Rate (EAR): proportion of citizens who successfully access human escalation when they request it, with time-to-escalation and barrier documentation.
- Citizen Dignity Score (CDS): composite measure of perceived respectfulness, patience, and non-judgment in AI interaction, assessed through validated survey instrument.
- Administrative Burden Change (ABC): net change in time, effort, and stress costs borne by citizens accessing services, compared to pre-AI deployment baseline.

6.2. Governance Checklist

1. Has a citizen impact assessment been completed, covering equity, accessibility, and dignity implications?
2. Is there a defined institutional accountability chain for AI outputs, including named role holders?
3. Are knowledge update protocols in place, with mandatory verification before updates go live?
4. Is there an immutable interaction log with defined retention and access policies?
5. Has bias auditing been completed against all legally protected characteristics?
6. Is human escalation accessible without barriers from every point in the AI interaction?
7. Are citizens clearly informed of AI mediation at the outset of every interaction?
8. Is there a citizen feedback mechanism designed to reach marginalized populations?
9. Is there a defined process for citizen redress when AI-mediated guidance causes harm?
10. Has the system been reviewed for compliance with applicable AI regulation (*e.g.*, EU AI Act)?
11. Is there a scheduled equity monitoring program with authority to suspend operations?
12. Does system documentation include scope limitations, known failure modes, and knowledge currency status?

7. DISCUSSION AND LIMITATIONS

The PSCAI framework contributes to the literature on AI in public administration by providing an integrated design and governance structure that addresses the distinctive communication requirements of citizen-facing public services. Its six pillars—dignity-centered communication, legal accuracy, transparent accountability, equity-aware design, accessible human escalation, and citizen feedback integration—reflect the convergence of public administration theory, HCI research, and AI ethics scholarship.

The framework has important limitations. It is primarily designed for text-based conversational AI systems in direct citizen-facing applications. Extension

to voice-based, multimodal, or decision-support AI systems used by public servants rather than directly by citizens would require significant adaptation. The framework is also primarily oriented to democratic governance contexts with rule-of-law institutions; its applicability in contexts with weaker institutional accountability infrastructure requires separate analysis.

The tension between efficiency imperatives and the extensive design and governance requirements of the PSCAI framework deserves explicit acknowledgment. Implementing the framework's requirements imposes significant resource costs on deploying institutions, and there is a real risk that institutions operating under resource constraints will treat compliance as a burden to be minimized rather than a design philosophy to be internalized. Addressing this tension requires policy frameworks that provide adequate resources for responsible AI deployment in the public sector, rather than assuming that efficiency gains will self-finance governance requirements.

8. CONCLUSION

The deployment of conversational AI in public service communication is not merely a technology adoption decision—it is a fundamental choice about the communicative relationship between democratic institutions and the citizens they serve. When that relationship is mediated by AI, the design of the AI system becomes a political as well as a technical act, with implications for citizen rights, institutional legitimacy, and democratic accountability.

The PSCAI framework proposed in this paper argues that responsible public sector AI communication requires a design philosophy centered on dignity, accuracy, accountability, equity, accessibility, and practical implementability. The framework is therefore not only a normative checklist but also an operational model for connecting conversational interfaces to official knowledge sources, audit trails, human escalation pathways, and continuous citizen feedback.

As public sector AI deployment accelerates across employment services, healthcare, welfare administration, and regulatory compliance, the research and practitioner communities must engage urgently with the governance challenges this paper has identified. The citizens most likely to be served—and most likely to be harmed—by public service AI are those who have the least power to demand better. Designing well for them is not optional; it is the fundamental purpose of public service.

REFERENCES

Cai, C. J., Winter, S., Steier, D., Rabelo, L., & Terry, M. (2023). Hello AI: Uncovering the onboarding needs of medical clinicians for human-AI collaborative decision-making. ACM CHI 2023.

Citrin, J., & Stoker, L. (2018). Political trust in a cynical age. Annual Review of Political Science, 21, 49-70. https://doi.org/10.1146/annurev-polisci-050316-092550

Desiere, S., & Struyven, L. (2021). Using artificial intelligence to classify jobseekers: The accuracy-fairness trade-off. Journal of Social Policy, 50(2), 367-385. https://doi.org/10.1017/S0047279420000203

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

Herd, P., & Moynihan, D. P. (2018). Administrative Burden: Policymaking by Other Means. Russell Sage Foundation. https://doi.org/10.7758/9781610448789

Kasirzadeh, A., & Gabriel, I. (2023). In conversation with artificial intelligence: Aligning language models with human values. Philosophy & Technology, 36(2). https://doi.org/10.1007/s13347-023-00606-x

Liao, Q. V., & Vaughan, J. W. (2023). AI transparency in the age of LLMs: A human-centered research roadmap. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8036d03b

Molina, M., & Sundt, M. (2023). Algorithmic accountability in public employment services: A comparative study of AI-mediated job placement systems in Europe. Government Information Quarterly, 40(2), 101805.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Ragnedda, M., & Ruiu, M. L. (2017). Social capital and the three levels of digital divide. In Theorizing Digital Divides (pp. 21-34). Routledge. https://doi.org/10.4324/9781315455334-3

Saxena, D., Badillo-Urquiola, K., Wisniewski, P., & Guha, S. (2021). A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. Proceedings of the ACM on Human-Computer Interaction (CSCW), 5. https://doi.org/10.1145/3476089

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law & Security Review, 41, 105567. https://doi.org/10.1016/j.clsr.2021.105567

Weidinger, L., *et al.* (2022). Taxonomy of risks posed by language models. Proceedings of FAccT 2022. https://doi.org/10.1145/3531146.3533088

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. International Journal of Public Administration, 42(7), 596-615. https://doi.org/10.1080/01900692.2018.1498103

https://doi.org/10.31875/2979-1081.2026.02.10