



TRUST CENTRIC FRAMEWORK FOR ARTIFICIAL INTELLIGENCE INTEGRATION IN PUBLIC ADMINISTRATION

Shaza Khalid

University Islam International Indonesia

*) Correspondence regarding this article should be addressed to: E-mail:
shaza.khalid@uiii.ac.id

Abstract: Artificial Intelligence (AI) and digital transformation are changing the global public governance landscape in today's fast-changing world. This paper addresses the important nexus of technological progress and institutional trust in the context of the Presidential Decree, which has been proclaimed as the Year of Digitalization and Artificial Intelligence. While AI can introduce an unprecedented level of efficiency and data-driven decision-making in public administration, it is also associated with the "black box" issue, where algorithmic processes raise citizen skepticism and resistance within public administration. This study adopts a qualitative and conceptual approach, aiming to address the question: How can public institutions build and maintain trust during the application of AI? Based on the concepts of Trust Management and Knowledge Management, the study posits a "Trust-Centric Framework" for digital government. The four pillars of this framework are Transparency, Accountability, Competence and Benevolence. To ensure the sustainable digital transformation of the civil service in Kazakhstan, it is necessary to focus not only on technical infrastructure but also on socio-technical management of trust. The study concludes that the implementation of AI is not merely a technological change but a fundamental transformation of the organization, which must be accompanied by new digital skills for civil servants and extensive ethical-legal frameworks. This study brings together international good practice and the geopolitical situation in Kazakhstan to provide policy recommendations for administrators.

Key Words: Public Administration, Digital Transformation, Kazakhstan, Artificial Intelligence, Trust Management.

Abstrak: Kecerdasan Buatan (AI) dan transformasi digital sedang mengubah lanskap pemerintahan publik global di dunia yang cepat berubah saat ini. Penelitian ini membahas hubungan penting antara kemajuan teknologi dan kepercayaan institusional dalam konteks Keputusan Presiden, yang telah diumumkan sebagai Tahun

Digitalisasi dan Kecerdasan Buatan. Sementara AI dapat memperkenalkan tingkat efisiensi yang belum pernah terjadi sebelumnya dan pengambilan keputusan berbasis data dalam administrasi publik, AI juga terkait dengan masalah "kotak hitam", di mana proses algoritmik menimbulkan skeptisisme dan resistensi warga dalam administrasi publik. Studi ini mengadopsi pendekatan kualitatif dan konseptual, dengan tujuan untuk menjawab pertanyaan: Bagaimana institusi publik dapat membangun dan mempertahankan kepercayaan selama penerapan AI? Berdasarkan konsep Manajemen Kepercayaan dan Manajemen Pengetahuan, studi ini mengusulkan "Kerangka Berbasis Kepercayaan" untuk pemerintahan digital. Empat pilar dari kerangka ini adalah Transparansi, Akuntabilitas, Kompetensi, dan Kebajikan. Untuk memastikan transformasi digital yang berkelanjutan dari layanan sipil di Kazakhstan, perlu untuk fokus tidak hanya pada infrastruktur teknis tetapi juga pada manajemen sosial-teknis kepercayaan. Studi ini menyimpulkan bahwa penerapan AI bukan sekadar perubahan teknologi tetapi transformasi mendasar dari organisasi, yang harus disertai dengan keterampilan digital baru bagi pegawai negeri dan kerangka etika-hukum yang luas. Studi ini menggabungkan praktik baik internasional dan situasi geopolitik di Kazakhstan untuk memberikan rekomendasi kebijakan bagi para administrator.

Kata kunci: Administrasi Publik; Transformasi Digital; Kazakhstan; Kecerdasan Buatan; Manajemen Kepercayaan.

INTRODUCTION

All over the world, public governance is at a pivotal point of structural transformation, driven by the Fourth Industrial Revolution. At the core of this transformation is artificial intelligence (AI), a collection of technologies such as machine learning, natural language processing, and predictive analytics, which will shape the interaction between the State and its citizens and the delivery of public services. This technological revolution is not only a matter of government administration in the Republic of Kazakhstan but also a matter of national priority. This is evidenced in the Presidential Decree titled “Year of Digitalization and Artificial Intelligence, 2026” which is the main catalyst for the application of high-tech solutions at all tiers of the public service.

It is necessary to clarify the situation in respect of the Decree. The reasoning in this paper assumes that the 2026 Presidential Decree is not yet a fully-fledged regulatory instrument with settled implementing rules, but the empirical anchor and policy premise of the analysis, a declared national policy priority. This is because several of the specific mechanisms discussed later in this paper (such as department-level AI Ethics Oversight Boards or mandatory Explainability Audits) are not proposed as existing administrative procedure but rather as a forward-looking policy scenario in line with the priorities set by the Decree. This distinction is retained throughout the paper, and should be interpreted as a boundary condition on the practical recommendations made in the paper.

However, there is a major socio-technical paradox in the fast uptake of AI in public administration. The sophistication of algorithmic systems is rapidly accelerating, and the human factor (trust) is the most difficult part to get right. The implementation of AI in public management is not a purely technical issue, as argued by Wirtz and Müller (2019). Even the most advanced digital system will not be used by the civil servants who must use it and by the citizens who are the end users of public services if they do not trust it.

Trust in the context of the modernization of Kazakhstan's civil service is multifaceted. It encompasses the trust relationship within the organization among different levels of government, the trust of civil servants in algorithmic support, and the trust of citizens in the fairness, transparency, and accountability of decision-making. As Paliszkievicz (2011) says, trust is a key ingredient in proper knowledge management, and in the digitalized world, this trust is the oil that makes the data flow and helps in implementing innovation. The growing significance of Artificial Intelligence in the public sector is reflected in recent advances in digital governance. International evaluations indicate a growing trend of using AI-enabled systems by governments to enhance service provision, administrative efficiency, and evidence-based decision making. But the implementation of these technologies is not only a technical matter, but also one of public and institutional trust.

The 2026 Presidential Decree clearly emphasizes the need for comprehensive work in the field of Trust Management. However, there is an urgent need for research on the introduction of specific models of trust management in the

administrative and cultural environment of Central Asian states. Many of the currently available models of AI integration are designed for Western or East Asian bureaucratic settings and not adequately account for the unique characteristics of bureaucratic culture and historical development in transitional areas such as Kazakhstan. Additionally, the "black box" strategy, which is widely adopted in many AI applications, whereby the decision-making process is not easily understood by the general public, is contrary to the Listening State model promoted by the national leadership.

The Research Gap: The Need for a Central Asian Trust Model

This isn't only a geographical difference, it's a structural difference. Western trust-in-AI models, for instance, GDPR-inflected transparency regimes and the risk-tiered approach of the EU AI Act (Regulation (EU) 2024/1689); are built on a tradition of adversarial, rights-based citizen–state relations, in which institutional trust is earned through contestability and oversight mechanisms that sit outside, and can check, the executive. In contrast, East Asian techno-corporatist models are likely to embed AI governance in already highly capable and high-trust bureaucracies that have high continuity in administration. The trust architecture inherited in Kazakhstan; as in most post-Soviet Central Asian states is also different: an historically centralized and vertically hierarchical bureaucratic culture, where trust is more likely to originate from the top, and where citizen oversight mechanisms are quite recent and institutionally underdeveloped. Therefore, the Western model of strong external contestability and the East Asian model of strong pre-existing bureaucratic trust do not easily correspond to the transitional administrative culture, where both are (simultaneously) in a developmental state. This mix – decree-mandated, top-down digitalization moving faster than institutionalized citizen oversight is the research gap driving this study.

While there is an increasing literature on artificial intelligence governance, the current studies have mainly addressed the technical, regulatory and operational aspects of AI adoption. Trust-centered approaches in transitional administrative systems have received little attention, especially in Central Asia. This leaves a lack

of understanding of how institutional trust can be systematically incorporated into AI-enabled public administration. This study draws on the perspectives of Trust Management, Knowledge Management, and Socio-Technical Systems. Such approaches can serve as a basis for understanding the interplay between technological innovation, organizational behaviour and institutional trust in the process of digital transformation in public administration.

This paper aims to fill these gaps by introducing a "Trust Architecture"—a conceptual and strategic approach to the ethical and effective use of AI in public administration. The research is not just about the 'how' of digitalisation, but also the more complicated 'why' of trust. This paper examines global best practices and local administrative needs to offer guidance for a people-centred digital transformation that is consistent with openness, accountability and quality service values. The paper contributes to the literature not only on the policy level, but also theoretically, as it extends the Organizational Trust Theory (Mayer, Davis, & Schoorman, 1995) and Trust Management theory (Paliszkiewicz, 2011), which were developed for a private sector and Western/Northern European organizational context, into a transitional, decree-driven public-sector context, and explicitly links them with the Socio-Technical Systems Theory (Trist, 1981) via a new construct, the Trust-Knowledge Feedback Loop, proposed in this paper. In this regard, the following research question is posed: How can public institutions develop and sustain trust in the process of the implementation of Artificial Intelligence in public administration?

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

Trust Management and Organizational Trust Theory

Researches on trust in organizational contexts have progressed from a psychological curiosity to a central pillar of strategic management. Trust is generally defined as the willingness of one party to be vulnerable to the actions of another party on the basis of the expectation that the other will act in a way that is beneficial to the trustor, even when the trustor cannot control or monitor the other party's actions (Mayer, Davis, & Schoorman, 1995). This Integrative Model of

Organizational Trust is the gold standard in the explanation of interpersonal interaction within bureaucracies. It focuses on three key perceived attributes of the trustee, namely, ability (competence), benevolence (loyalty to the user), and integrity (conformance to a code of ethics). In the field of public administration, they become even more important. After a government begins to utilize AI in decision making, the trustee is not a human partner anymore, but an algorithmic system. Trust, as Paliszkiewicz (2011) states, is a complex phenomenon and a "strategic asset" for the organization. Her research on trust management indicates that trust is not only a social enabler but it is also a prerequisite for sharing of knowledge. With low institutional trust, communication is limited, and the intelligence of the organization (human or machine) is compromised.

To ensure optimal performance of an AI system, it is essential that high-quality data is provided and that employees are motivated to engage with the technology. Paliszkiewicz (2012) states that the "trust-knowledge" relationship is the basis for management in the digital world. In low-trust organizations, digital transformation can be seen as a risk to employee autonomy or job security. This can lead to the "knowledge hiding" of critical tacit knowledge, which could undermine even the most advanced AI applications. The adoption of AI requires trust, and the use of AI influences future trust levels. Wirtz and Müller (2019) state that AI is often used to make high-stakes decisions in public management, such as the distribution of social benefits or the monitoring of public safety. An algorithm is seen as an instrument of surveillance or punishment, trust between the state and the people is lost.

Socio-Technical Systems Theory

This study is also rooted in Socio-Technical Systems (STS) Theory (Trist, 1981), which views organizational performance as being dependent on the optimization of both the technical and the social subsystems and not just the technology. STS theory is applied to the introduction of AI in the civil service in Kazakhstan, shifting the research question away from the purely technological question of whether an algorithm works, towards an interdependency perspective

in which the efficiency of the algorithm and the trust of the institution are mutually formed. An AI system, technically excellent but socially unaccepted will work badly in a socio-technical sense even if it works technically in a narrow (engineering) sense: this is the theory behind why in this paper the variable that is of interest is not only technical performance, but also trust.

Algorithmic decision making and knowledge hiding

A relationship only briefly mentioned in previous drafts of this line of research and developed here is that of how distrust of AI can lead to "knowledge hiding" (Paliszkiewicz, 2012) and, ultimately, to worse algorithmic performance. Digital transformation may be seen by civil servants in low-trust organizations as a threat to their autonomy, discretion, or even job security. Three mechanisms seem to link this perception to a tangible knowledge flow issue. First, civil servants with undeniable suspicions that an algorithmic system is monitoring or evaluating their own performance have an incentive to withhold the informal, tacit details, exceptions, caveats and situational judgments that would be shared with a human colleague or supervisor. Second, if workers feel that more thorough and accurate data input will hasten the automation of their job, they may engage in "defensive under-reporting"; putting in only the essential data that will make the model work, and not the complete story that would benefit the model. Third, low trust decreases the chances that civil servants will identify edge-cases where the algorithm is not good, and that they will report them, as it takes some level of interaction and goodwill to do so, and that is undermined by low trust.

The result is a vicious cycle: distrust blocks the sharing of good tacit knowledge into the system, which decreases the practical performance of the system and its ability to handle edge cases, leading to more distrust. According to Wirtz and Müller (2019), the use of AI to make important decisions regarding the public, for instance, in the allocation of social benefits or the monitoring of public safety, is particularly significant in this regard: The trust of citizens in the state and the trust of the management in the staff can hardly be restored once they perceive AI

as an instrument of surveillance or punishment. This mechanism is later formally stated in this paper as the "Trust-Knowledge Feedback Loop".

Contemporary Developments: Generative AI, Explainable AI, and Global AI Regulation

Three streams of scholarship have developed since the foundational literature on trust management and socio-technical literature discussed above was written that directly apply to AI-enabled public administration and are integrated here. *First*, generative AI has shifted from the experimental to directly being deployed in government service delivery. Consistent with this paper, Neumann, Guirguis, and Steiner (2024) offer a comparative case study on the adoption of AI technologies in public organizations that demonstrates that institutional and governance context, not just technical capability; account for differences in the adoption outcomes. The OECD's (2025) review of AI in public service design and delivery captures a fast-growing trend of the use of generative AI within government systems globally, not only for its operational promise but also for the trust risks that generative systems bring to the design and delivery of public services in addition to the risks of the predictive and rule-based AI processes covered in previous public-administration scholarship.

Secondly, the research agenda of explainable AI (XAI) has gone from a purely technical to a policy instrument. Explainability is politically popular, and is often under-specified, as Nannini, Balayn, and Smith (2023) critically review the specification of explainability requirements across EU, US, and UK AI policy documents; they find that, much of the time, explainability is only mentioned in policy documents, but not clearly defined in terms of target, threshold, type, rendering it a symbolic rather than substantive accountability mechanism. Nannini, Alonso-Moral, Catalá, Lama and Barro (2024) then explore the operationalization of explainability in the EU's mandatory regulatory framework in particular. *Thirdly*, the EU Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council, published in the Official Journal of the European Union on 12 July 2024, is the first binding, risk-tiered legal framework for AI

internationally, and offers an external regulatory benchmark against which the accountability and transparency pillars of the proposed Trust Architecture can be compared. While Kazakhstan does not have the obligations under the EU AI Act, and copies of its risk-tiering logic are not possible, the AI Act's general approach to determining obligations (higher risk equals higher obligations) serves as an external benchmark for the graduated oversight under Pillar 2 (Accountability).

Synthesis: Toward a Trust-Centric Conceptual Framework

This literature indicates that a trust-centric AI framework for Kazakhstan's administrative context should include three elements: firstly, trust is a two-way dynamic relationship between technical systems, civil servants, and citizens (Mayer et al., 1995; Paliszkiewicz, 2011) and not a precondition; secondly, such a framework should be based on joint optimization of social and technical subsystems (Trist, 1981), rather than merely on technical performance; and thirdly, it should take into account the operational lessons learnt from recent generative-AI, XAI, and regulatory research (Neumann et al., 2024; Nannini et al., 2023, 2024; Regulation (EU) 2024/1689), making its transparency and accountability requirements specific enough to be audited. The three requirements give shape to the four-pillar Trust Architecture.

METHODS

Research Design: Socio-Technical Approach

This study uses qualitative conceptual research design, based on Socio-Technical Systems (STS) Theory. This study moves beyond the purely technological perspective of AI adoption in the Kazakhstan civil service to an interdependency perspective, exploring the relationship between algorithmic efficiency and trust in government, as proposed by STS theory (Trist, 1981). The conceptual framework of this paper is needed because of the dynamic evolution of the 2026 Presidential Decree. The lack of empirical findings regarding the long-term consequences of these new AI policies in the civil service of Kazakhstan, makes conceptual approach a wise and scientifically sound approach to foresight and avoid potential pitfalls in the digital transformation process (Snyder, 2019).

Data Collection and Scoping: The Digital Kazakhstan Context

The analysis in particular includes:

- State Program Digital Kazakhstan: To identify the current infrastructure.
- Presidential Decree on AI 2026: To grasp specific strategic directives.
- OECD Digital Government Studies: To indicate Kazakhstan's progress.
- The research was conducted using the snowballing approach, beginning with the basic theories of Trust Management (Paliszkiewicz, 2011) and Mayer et al. (1995). The sources were analyzed for relevance to the concept of transitioning economies, as well as to ensure that western models of transitioning economies were not simply applied to the central Asian context without cultural adaptation.

Approach to the literature and narratives

A methodological framing note. An initial version of this study reported a literature search which was conducted according to PRISMA (Moher et al., 2009). In retrospect, the description was excessive: this paper was not a systematic review in the PRISMA sense, so we do not report a PRISMA flow diagram (with numbers identified, screened and excluded). It is better thought of as a narrative synthesis (Snyder, 2019) of the trust-management, socio-technical-systems, and AI-governance literatures, which is guided by but not formally aligned with systematic-review transparency norms.

The narrative synthesis is based on foundational and contemporary literature found through iterative and snowball searching using the keywords "trust management", "institutional trust", "artificial intelligence governance", "public administration", "socio-technical systems", "explainable AI", and "Central Asia" / "Kazakhstan" in Scopus, Web of Science and Google Scholar. Literature was not only searched for during the publication period 2018–2025, but also literature covering the foundational theory (e.g., Mayer et al., 1995; Trist, 1981) and generative AI, XAI-as-policy, and the EU AI Act (Section 2.4). Sources were filtered for relevance to trust-centric governance of AI and specifically to the

applicability of the conceptual models to non-Western and transitional governance contexts, avoiding uncritical transposition of Western models to the Central Asian context. If a future revision of this conceptual paper were to be expanded to become a full systematic review, a PRISMA flow diagram with screening counts would be provided at that stage, but is not part of this conceptual paper.

Thematic synthesis and coding: methodological frameworks

Thematic synthesis was carried out using a three-step qualitative coding process (Braun & Clarke, 2006) to develop the "Trust Architecture" in Section 4.

- Stage 1 (Open Coding): Raw data points identified with regard to the failure of AI (e.g., bias, opacity).
- Stage 2 (Axial Coding): Categorizing these points into Trust Barriers (e.g., "black box" problem) and Trust Drivers (e.g., transparency and competency).
- Stage 3 (Selective Coding): Putting these drivers together in a strategic model.

Particular attention was paid to the "Trust-Knowledge Nexus" as described by Paliszkievicz and Madra-Sawicka (2016) in which trust is the primary method of knowledge transfer in digital environments. This coding process will ensure that the final framework will not be a list of ideas, but a framework where each "pillar" of the architecture directly responds to a barrier identified.

Selection Criteria for the Comparative Cases

The research design includes case studies on Estonia and Poland. The cases have been selected based on the 'Most Different Systems Design' (MDSD) (Przeworski & Teune, 1970) logic. The first is a decentralized/digital-native model and the second is a centralized/transitioning model with a historical administrative legacy with Kazakhstan. This comparison is built on secondary and documentary

data, and is not based on new country-level primary or quantitative data, but instead aims to identify transferable principles and constraints of design; and does not aim to test causal claims about the impact of any given policy instrument. This is a conscious limitation of scope in the conceptual paper presented here, which is explicitly returned to in the Conclusions. Three significant measures are examined:

- Legislative Transparency: Legal frameworks to justify AI decisions.
- Public Trust: Eurobarometer historical data and domestic statistics.
- Public Administration AI Competencies: Specialization training level and quality.

These variables are compared and portable best practices are identified that can be transferred to the Kazakhstani "Hearing State" (Listening State) model.

Operationalizing Key Variables

An integral component is operationalising trust in the digital public service. The four qualitative variables that were employed in the coding of the Estonian and Polish case studies are summarized under the umbrella term "Trust in AI" based on the results of research by Paliszkiewicz (2011) and Wirtz (2019): These variables are used to create a Measurement Matrix for comparison of international models and the administrative state of Kazakhstan.

Table 1. Operationalization of Trust Variables

Variable	Operational Definition
System Competence	Technical accuracy and reliability
Procedural Fairness	Explainability and reason-giving
Privacy Integrity	Protection of citizen data
Benevolent AI	Public-interest orientation

Source: Author's compilation based on Paliszkiewicz (2011) and Wirtz & Müller (2019)

These variables are used to create a Measurement Matrix for comparison of international models and the administrative state of Kazakhstan.

RESULTS AND DISCUSSION

“Trust Architecture” for AI Adoption

The key finding of this study is the "Trust Architecture" (TA) – a socio-technical strategy for bridging the gap between technological possibilities and institutional acceptance in the Kazakhstani civil service. The TA serves as a detailed response to the 2026 Presidential Decree, outlining a strategic pathway to transition from experimental AI toward institutionalized AI. The architecture is based on an international comparative study of Estonia and Poland and a thematic coding of the trust management literature. It posits that trust in AI isn't a fixed condition, but instead a dynamic balance supported by four interdependent pillars: algorithmic transparency, institutional accountability, professional competence, and public benevolence.

Table 2. Four Pillars of the Trust Architecture

Pillar	Purpose	Expected Outcome
Transparency	Explainability	Public understanding
Accountability	Oversight	Institutional legitimacy
Competence	Digital skills	Effective implementation
Benevolence	Citizen-centricity	Public trust

Source: Developed by the authors based on thematic synthesis of Paliszkiewicz (2011), Mayer et al. (1995), and Wirtz & Müller (2019).

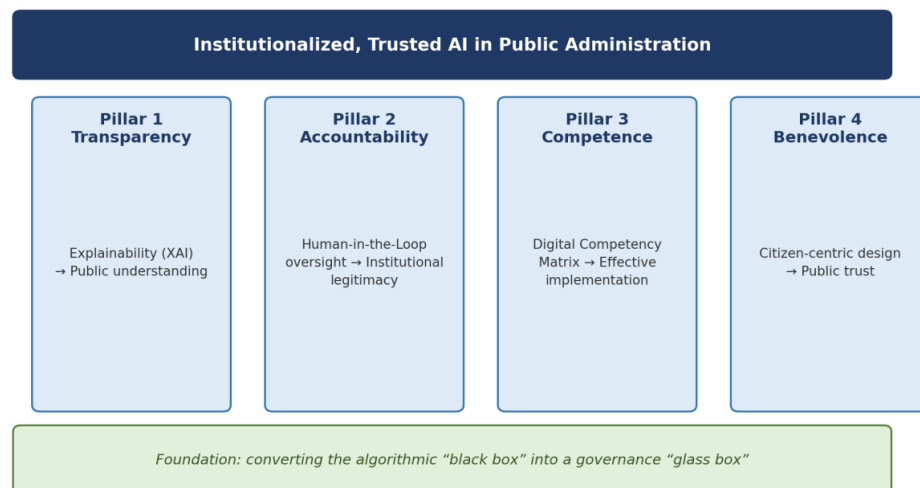


Fig. 1. The four pillars of the Trust Architecture are those supporting the institutionalised trusted AI in public administration.

Every pillar tackles a concrete trust barrier that the methodology has identified, and helps to turn the "black box" of AI into a glass box of governance. The Trust Architecture (TA) goes beyond "technocratic determinism" in approaches to implementing AI. It is clear that the use of AI doesn't eliminate the need for trust in humans, in fact, it enhances it. Also, Pillar 1 (Transparency) supports Paliszkievicz (2017), which will be discussed later in the paper. Transparency shows AI's technical ability and institutional integrity. If a government decides to use an algorithm to make a decision on which citizens it cannot be transparent. Then the "black box" phenomenon is not only a technical issue but a trust violation, because the government essentially breaches the social contract of transparency. The TA framework tackles this by introducing Explainable AI (XAI).

Pillar 1: Transparency and Explainability of Algorithms (XAI)

One of the key takeaways from this research is that transparency is a prerequisite to trust. The obligation of the government to inform citizens on the grounds of the decision-making processes is a legal and social requirement within the framework of the "Hearing State" (Listening State) in Kazakhstan. The results

show that for AI to be included in the public administration system, it must become Explainable AI (XAI). Data governance security is an integral part of the why of AI decision-making, which can enhance the perceived integrity as proposed by Janssen et al. (2020) and help to solve the problem of arbitrary or biased automation, which undermines the citizen's right to fair treatment, as written in the Constitution of Kazakhstan.

The "Black Box" problem of deep learning and other advanced AI systems is one of the primary issues that have been identified in recent researches. Public administration is based on the concept that government actions must be justifiable and open to review, a concept known as administrative reason-giving. Many AI systems, are based on non-linear processes which are often opaque even to the designers themselves, as Janssen et al. (2020) argue. This Explainable AI (XAI) issue results in a crisis of accountability.

The lack of AI transparency is legally and ethically questionable, especially in a Hearing State (Listening State) like Kazakhstan, where the government is attempting to become more responsive to the needs of the people. An institutional problem exists with the integrity (as per Mayer's model) if the civil servant cannot explain to the citizen why they were denied a business permit by the AI system. This research indicates that transparency is the link between this gap. Transparency is not only related to code visibility, but also to meaningful human oversight and interpretability (Dwivedi et al., 2021). The literature states that for AI to be accepted in the public sector, it should be a 'glass box' rather than a 'black box' – the entire process of AI decision-making should be transparent.

Institutional Accountability and Oversight (Pillar 2)

The second part of the architecture is the accountability gap produced by the automated systems. The research shows that the protocol that maximizes the trust is the "Human-in-the-Loop" (HITL) protocol. The Trust Architecture proposes a multi-level supervision system for the public service in Kazakhstan:

Technical Level: Bias-detection protocols to check algorithmic processes for potential errors or discrimination.

Administrative Level: The introduction of an AI Ethics Commissioner within every government department to address complaints from citizens regarding AI decision making.

Legal Level: Complex legal framework to establish roles of AI developers and government bodies.

The system is developed on the basis of international comparative studies and it is an effective model of Ombudsman functioning in some European countries, adapted to the administrative system in Kazakhstan. Jobin et al. (2019) found that the most frequently cited ethical principle in international guidance on AI is also the most challenging to put into practice: accountability. The TA addresses this by requiring that for every AI system implemented under the 2026 Decree, a Responsible Human Officer must be appointed who is legally and ethically liable for the system's operation. This avoids an accountability vacuum where the software takes the blame and not the system. The other aspect of the literature is related to AI ethics and the law. With the growing reliance of governments on machines to make decisions, administrative accountability is a key question. A global survey of AI ethics guidelines (Jobin, Ienca and Vayena, 2019) revealed that transparency, justice and non-maleficence were the most frequently mentioned principles. However, the gap between the higher moral theories and the actual realization of the same in the routine work of public administration is very large.

Kazakhstan's legislation should remain current with the intricacies of AI, including bias and privacy concerns. According to the “Hearing State” (Listening State) model, any governmental decision can be challenged. Research indicates that a regulatory sandbox framework, in which AI systems are tried out in a legal safe space prior to deployment on a larger scale, can stimulate innovation and public safety. It has been asserted in this review that the ‘Trust Architecture’ proposed in this paper should be accompanied by citizen rights-respecting and technology-friendly legal frameworks.

In addition to institutional accountability, protecting citizen information is also critical to maintaining trust. An important issue in this regard is the security-trust dilemma. In the unique geopolitical context of Central Asia, data sovereignty and security play a crucial role. Gathering this data can be an overreach of state surveillance, and AI needs lots of data. The Trust Architecture aims to solve this problem by promoting a privacy-first approach to AI. Data sovereignty is highlighted, so in Estonia, the government can not only monitor what citizens are doing with AI, but can also defend their rights. The more the government is a guardian of data and not just a user of data, the more we can trust it. This means that in Kazakhstan, the Roadmap 2026-2030 should focus on building safe local cloud platforms to ensure that data remains within Kazakhstan's legal jurisdiction, strengthening the 'integrity' pillar of the proposed model.

Pillar 3: Civil Service Capabilities and "Trust-Knowledge"

The third pillar is on trust within the civil service. This study reveals that the resistance to AI is often caused by “information asymmetry”, which is an insufficient understanding of the functioning of the technology that public servants use. The Trust Architecture includes a Digital Competency Matrix, similar to Paliszkievicz (2017) on trust and knowledge management. This matrix isn't limited to just technical skills. It requires civil servants to have the ability to recognise when an AI is producing incorrect or biased outcomes, referred to as algorithmic criticality. The results show that Kazakhstan needs to launch a massive retraining program for its staff in the context of the "Year of Digitalization". The TA hypothesis suggested that the professional trust of a civil servant increases when they have confidence in the use of AI. This ultimately leads to knowledge sharing and raises the probability that the workforce will innovate from technology other than simply obeying bureaucratic orders. The crisis of confidence of the civil service, which is digitalized without the professional competence pillar, will have a significant impact on the national strategy. This discussion points to the fact that the “Year of Digitalization” must be a year of human investment and not a year of hardware investment.

Also, the use of Artificial Intelligence in Public Administration does not only rely on the technical strength of the algorithms, it hinges upon on the digital readiness of the public service. Mergel, Edelmann and Haug (2019) state that digital transformation is a cultural and organizational transformation that calls for competency development. AI adds to digital literacy, encompassing what's called data fluency and algorithmic awareness as well.

Trust in AI often comes as a consequence of understanding. AI-generated insights lead to higher trust among civil servants. Conversely, low digital competence contributes to technostress (the sense of being overloaded by technological advances). Paliszkiewicz (2020) focuses on the role of leadership in the formation of such competence. There is a need to promote a culture of lifelong learning in the public sector, where AI is perceived as an augmenting tool than a replacement tool. This is an important attitude for the Kazakhstan civil service in the period of its functioning under the Decree of 2026. If employees do not feel in control of the AI, then the whole digitalization process is in doubt.

Pillar 4 – Public Benevolence and Citizen Centric Design

Public Benevolence is the fourth and possibly most important pillar of the Trust Architecture. Benevolence in the context of trust management is the perception that the trustee (AI and government) cares about the welfare of the trustor (citizen), than its own (bureaucratic efficiency or surveillance). Comparative analysis results indicate that technology adoption is unsuccessful if the use of AI is seen as a means of bureaucratic distancing — using machines as a way to avoid direct human contact or to deny citizens' rights.

The Trust Architecture encourages human-centric AI design, in line with the Hearing State (Listening State) approach. This implies that in the Kazakhstan civil service, AI solutions should be assessed not only on technical ability, but also on the social impact. The findings show that when AI is used to make complex bureaucratic processes more convenient for the citizen, such as in the field of tax administration with AI support for automated tax filing, or in education with AI personalization of learning paths, the level of trust is highest. As Bannister and

Connolly (2014) explain, trust in e-government is easily shattered, and it must be a benevolent intent which the public can see. The focus on benevolence also guarantees that the 2026 Presidential Decree will result in a more responsive and empathetic government that is not just “smarter.”

The assessment of AI integration's performance should not be limited to technical aspects such as efficiency or cost reduction, in order to fit into the benevolence pillar. Infact, a 'Benevolence Metric' or 'Public Trust Score' should be instituted. This would include regular citizen surveys, through e-Gov platforms, to assess if the AI tools have indeed made the public interaction simpler and more transparent. If an AI tool makes things more efficient, but trust in the citizen decreases, it needs to be recalibrated.

The "Trust-Knowledge" Feedback Loop

An important finding of this research is the discovery of a cyclical link between trust and information quality, the so-called Trust-Knowledge Feedback Loop. This study reveals that trust is the most important mediator for the dissemination of high-quality information in digital systems as previously mentioned by Paliszkiewicz, Koohang and Nord (2021). The operation of this loop is as follows:

Trust as Input: Transparency and accountability promote citizens and public servants to input more accurate and complete data.

Knowledge Generation: Quality data leads to more accurate AI predictions and better administrative decisions.

Trust as Output: These decisions become more effective and accurate as institutional trust grows, leading to greater digitalization.

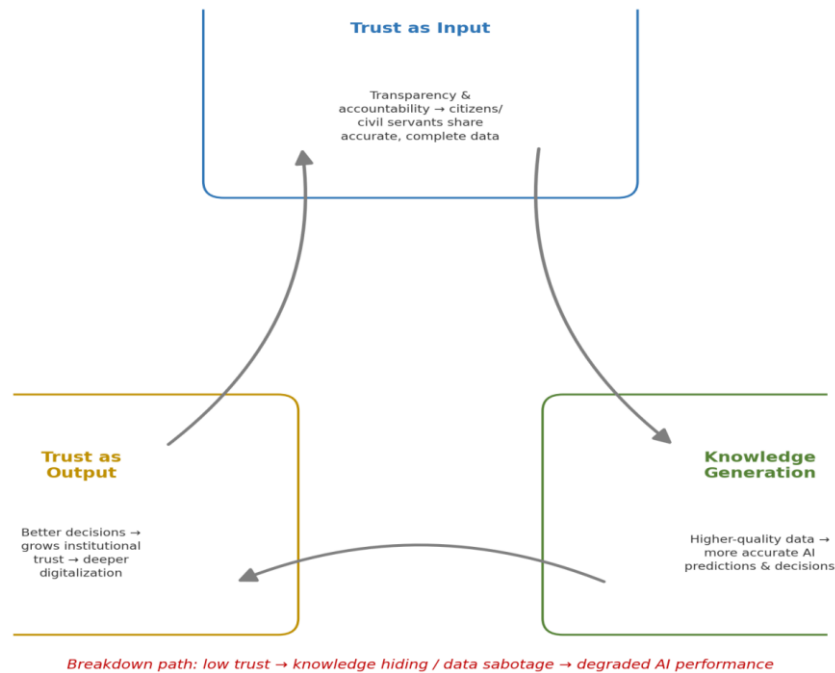


Fig. 2. *The Trust-Knowledge Feedback Loop, with the breakdown path due to low levels of institutional trust.*

But in case, if the trust is not present, users can engage in data sabotage, or knowledge hiding, where they may purposefully enter inaccurate or incomplete data to protect their privacy or autonomy. This is significant for Kazakhstan, where the transition to a data-driven government calls for a high degree of information altruism. The Trust Architecture addresses this by adding data sovereignty – the control of data and how it is utilized by AI as an integral component of the feedback loop.

Comparative Lessons for Kazakhstan

In addition to learning from international experience, the proposed Trust Architecture places Kazakhstan as a potential regional leader in the field of ethical AI governance. Kazakhstan is not merely adapting foreign models but can also transfer foreign experiences to its administrative environment and provide a more lucrative model for other transition economies. This study is unique in that it places the issue of AI in the context of "Hearing State". The Kazakhstan model of AI governance is different from the western one, which prioritizes individual rights

over collective efficiency, as the Kazakhstan model is geared towards the benefits of AI, also protecting the rights of users. The case study analysis of Estonia and Poland shows that trust-by-design is not a universal fit and needs to be culturally adapted.

In Central Asia, there is a hierarchical trust in institutions. The Trust Architecture overcomes this challenge by ensuring the Multi-Level Oversight Mechanism (Pillar 2) is not an external bureaucratic burden but is integrated into institutional frameworks. This is a wise digitalization, in which AI is employed to improve state responsiveness without undermining the traditional conceptions of order and stability. According to the United Nations (2024) E-Government Survey, Kazakhstan has already made significant progress in the digital transformation of its infrastructure, but to achieve full AI-powered governance, the country must advance beyond the phase of data availability to data trust. The results of this research suggest that if the public sees AI as a tool for benevolent governance (Pillar 4), Kazakhstan can overcome many of the innovation-resistance issues associated with more decentralized governments.

The examples of Estonia and Poland show that institutional trust is essential for effective AI governance, as well as technological innovation. Estonia exemplifies the trust-by-design concept, which is based on transparency, data privacy, and citizen control, to boost public trust in digital government (Velsberg, Westergren and Jonsson, 2020). Likewise, Poland shows that institutional trust and knowledge management are crucial for successful digital transformation in transition economies (Paliszkiewicz, 2011). Based on these experiences, it can be said that technologies like AI, blockchain and cloud computing can be transferred but whether it is effective or not depends on the culture and institutional context of each country.

What Transfers, and What Requires Adaptation

The Estonian and Polish cases demonstrate that institutional trust is as important to effective AI governance as technological capability, but the comparative reading above also cautions against treating either case as a direct

template. Estonia exemplifies 'trust-by-design,' grounded in transparency, data privacy, and citizen control of digital identity (Velsberg, Westergren, & Jonsson, 2020); Poland illustrates the role of institutional trust and knowledge management in the digital transformation of a transitioning, historically centralized economy (Paliszkiewicz, 2011). The table below sets out, at the level of design principle rather than quantitative measurement, which elements this study assesses as more directly transferable to Kazakhstan's administrative culture and which are likely to require substantial local adaptation.

Table 3. Illustrative Transferability of Estonian and Polish Design Principles to Kazakhstan

Design Principle (Source)	Assessed Transferability	Rationale / Required Adaptation
Citizen-controlled digital identity and data access logs (Estonia)	Moderate technically transferable, institutionally demanding	Requires the independent oversight and legal contestability infrastructure that underpins Estonian data sovereignty; in Kazakhstan this pillar would need to be built alongside, not after, the technical rollout.
Decentralized, citizen-facing transparency portals (Estonia)	Low-to-moderate in the short term	Assumes a decentralized, disclosure-oriented administrative culture; in a more centralized system this principle is more realistically adapted as centrally coordinated but publicly auditable disclosure, rather than a fully decentralized model.
Sector-based institutional trust-building through incremental digitalization (Poland)	High	Poland's transitioning, historically centralized administrative legacy is structurally closer to Kazakhstan's; incremental, sector-by-

		sector trust-building (health, tax, then justice) is a directly relevant sequencing lesson, reflected in the phased Roadmap.
Knowledge-management-centred change management for digital transformation (Poland)	High	Directly applicable: Poland's experience reinforces this paper's emphasis on the competence pillar and the Trust-Knowledge Feedback Loop as preconditions for, not consequences of, successful digitalization.

Taken together, this comparison suggests that Kazakhstan is better positioned to adapt Poland's incremental, competence-led sequencing than to transplant Estonia's decentralized transparency infrastructure wholesale — though elements of Estonian practice, particularly around citizen-controlled data logs, remain a longer-term aspiration once independent oversight capacity is more fully institutionalized. As with the rest of this comparative analysis, this assessment is offered as an illustrative, literature-informed judgment rather than a statistically validated transferability score, consistent with the scope limitation noted in the Methods section.

Kazakhstan's Strategic Implementation Roadmap (2026-2030)

This research presents a three-stage roadmap for the Republic of Kazakhstan to address the needs of the 2026 Presidential Decree by combining global best practices with the proposed TA framework.

Phase I (Years 1-2): Establish legal guardrails and an AI Ethics Charter. Assert on "explainability modules" from all vendors.

Phase II (Years 2-3): Launch National Digital Fluency Certification and pilot projects in municipal management to build early public trust.

Phase III (Years 4-5): Full deployment in healthcare and justice, with Multi-Level Oversight mechanisms in place and functioning.

Recommendations for the Kazakhstan Civil Service Policy

The “Trust Architecture” presented in this study is not only a conceptual framework, it is also a practical instrument to realize the Presidential Decree of 2026. In the context of the four pillars of transparency, accountability, competence, and benevolence, the following practical recommendations are made to the policymakers and top managers of the Republic of Kazakhstan:

Institutionalization of Explainability Audits: The Academy of Public Administration and the Ministry of Digital Development should ensure that all AI systems that provide services to citizens are subject to Explainability Audits to ensure transparency. Before introducing any algorithm, whether for social welfare or taxation, it should be audited so that the grounds for its decisions can be explained to the public in natural language (Kazakh and Russian). This is to avoid a black box approach which will lead to a loss of institutional trust.

AI Ethics Oversight Boards (AEOB): In order to achieve the principle of accountability, each public institution should have its own AEOB. These boards should not consist of technical experts alone, but should also include lawyers, ethicists and representatives from civil society. The boards will act as a Human-in-the-Loop (HITL) mechanism to address complex cases where citizens are concerned about AI-driven decisions. This is in conjunction with the Hearing State (Listening State), in which the machine is not the ultimate and uncontroversial source of authority.

Digital Trust Bootcamp: Civil Service should move beyond IT literacy to reach the pillar of competence. A national certification program is proposed, entitled Trust Management in the Digital Era, which Paliszkievicz (2020) investigated the relationship between trust and leadership, and which should help managers to identify and manage technostress in their team. Public servants must

be equipped to be algorithmic managers than passive managers who simply take the recommendations made by the machine.

Geopolitical aspects: Kazakhstan as a digital leader

Finally, there are geopolitical consequences of the Trust Architecture. If Kazakhstan manages to implement AI on the basis of trust, it can become a Central Asian Digital hub. The 2026 Presidential Decree is not only a domestic project, but also a confidence-building initiative for investors and partners from around the world, as it demonstrates that Kazakhstan is in the process of building a modern, reliable, and high-tech government. The “Trust-Centric Framework” proposed in this research can be used as a model for other countries in the region (such as Uzbekistan and Kyrgyzstan) with similar administrative backgrounds. Kazakhstan's emphasis on public benevolence and data sovereignty can pave the way for the development of a Regional AI Ethics Standard tailored to the Eurasian context. This would make the country a leader in what is called academic diplomacy.

CONCLUSIONS

The challenges of introducing Artificial Intelligence (AI) to the public service of the Republic of Kazakhstan were addressed in this study, which developed a "Trust Architecture". This study is not only concerned with the technical part of digital transformation, but also examines the socio-technical aspects of institutional trust in the light of the Presidential Decree on the Year of Digitalization and Artificial Intelligence (AI) issued in 2026. The central contribution of this study is the four-part framework of algorithmic transparency, institutional accountability, professional competence, and public benevolence for AI to be effective in public services. This study demonstrates that trust is the oil of digital governance and can be learned from good practices in other countries and adapted to the needs of the “Hearing State” (Listening State) of Kazakhstan. The rapid incorporation of AI, without explanation and human control, could create an accountability gap," thereby breaching the social pact between government and citizens. The results indicate that public institutions can build and maintain public

trust in AI by incorporating transparency, accountability, professional competence, and public benevolence throughout the AI governance process.

The novelty of this work lies in the scientific aspect of creating the “Trust Architecture” model on the basis of the theory of trust management and the requirements of the legislation for a transitioning economy. This is a regional AI governance solution for Central Asian countries and a different approach from the western-centric models. Moreover, the study proposed a new concept, the "Trust-Knowledge Feedback Loop," to explain the relationship between institutional trust and data quality in the context of AI-powered decision-making.

Moreover, this study contributes to the holistic knowledge by integrating the concept of the Hearing State and the context of the Fourth Industrial Revolution. It defines algorithmic integrity as one of the key components of the bureaucratic code of ethics, that is, the link between the old rules of public administration and the new rules of the digital age. In this paper, civil servants and policymakers are provided with a roadmap of three stages to help them transition from digital experimentation to a digital revolution with trusted AI, with recommendations that can be applied to ongoing reforms in Kazakhstan.

The Trust Architecture seeks to mitigate the risks of technostress and algorithmic bias within public sector of Kazakhstan and to make the administration of the 2026 Presidential Decree more efficient, transparent and benevolent. The study underscores the fact that the Year of Digitalization should be considered a Year of Human Investment: the aim is not to replace the bureaucrat with a machine, but to enable a digital civil servant, who can more effectively serve the public interest.

While this study provides a strong Trust Architecture to facilitate the integration of AI, there are some limitations to consider. First, the proposed model is a conceptual project, which is based on the literature review and the international standards, but not on the new primary data of the civil service of Kazakhstan. The coding of the sources was thorough, but there might be cultural or bureaucratic

barriers to implementing the TA approach that were not addressed in the general literature on "trust management".

In the journey towards full-scale AI integration, several avenues for further research lie ahead for Kazakhstan. The next steps for research should be to test the Trust Architecture empirically in a longitudinal citizen survey and to do a field experiment in a concrete government institution, e.g. the Ministry of Healthcare or the Ministry of Social Protection. In addition, additional research on "legal liability of AI" in the judiciary system of Kazakhstan is required to strengthen the institutional responsibility pillar.

To sum up, this research concludes that trust is the main currency in the Digital Age. The Trust Architecture is both a compass and a shield, guiding the Republic of Kazakhstan on its ambitious path toward AI-based governance while ensuring that the technological promises of the future are realized in an ethical, reliable, and human-centered way.

Bibliography

- Bannister, F., & Connolly, R. (2014). ICT, trust and government: A model of the trust of citizens in e-government. *Government Information Quarterly*, 31(1), 73–86.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Digital Kazakhstan State Program. (2017, Updated 2024). On the approval of the State Program "Digital Kazakhstan". Government of the Republic of Kazakhstan. <https://adilet.zan.kz/rus/docs/P1700000827>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Gertsen, G., & et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994.
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, OJ L, 2024/1689, 12 July 2024.
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance, trust and aspects of AI-based decision-making. *Government Information Quarterly*, 37(3), 101447.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.
- Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation in the public sector. *Government Information Quarterly*, 36(4), 101385.

- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097.
- Nannini, L., Balayn, A., & Smith, A. L. (2023). Explainability in AI policies: A critical review of communications, reports, regulations, and standards in the EU, US, and UK. *arXiv preprint arXiv:2304.11218*.
- Nannini, L., Alonso-Moral, J. M., Catalá, A., Lama, M., & Barro, S. (2024). Operationalizing explainable artificial intelligence in the European Union regulatory ecosystem. *IEEE Intelligent Systems*, 39(4), 37–48.
- Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, 26(1), 114–141.
- OECD. (2025). *AI in public service design and delivery: Governing with artificial intelligence*. OECD Publishing.
- Paliszkiewicz, J. (2011). Trust management: Literature review. *Management*, 6(4), 315–331.
- Paliszkiewicz, J. (2012). *Orientation on trust and organizational performance*. Warsaw University of Life Sciences Press.
- Paliszkiewicz, J., & Madra-Sawicka, M. (2016). Trust and social media: A literature review. *Management*, 11(2), 127–140.
- Paliszkiewicz, J. (2017). *Trust and knowledge management in organizations*. Warsaw University of Life Sciences Press.
- Paliszkiewicz, J. (2020). Leadership, trust, and knowledge management in the digital era. *Journal of Computer Information Systems*, 60(2), 112–121.

- Paliszkiewicz, J., Koohang, A., & Nord, J. H. (2021). Digital transformation, trust, and knowledge management: Perspectives from the public and private sectors. Routledge.
- Presidential Decree of the Republic of Kazakhstan. (2026). On declaring the Year of Digitalization and Artificial Intelligence. Official Portal of the President of the Republic of Kazakhstan (Akorda).
- Przeworski, A., & Teune, H. (1970). The logic of comparative social inquiry. Wiley-Interscience.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339.
- Trist, E. L. (1981). The evolution of socio-technical systems: A conceptual framework and an action research program. Occasional Paper No. 2. Ontario Quality of Working Life Centre.
- United Nations. (2024). E-government survey 2024: Accelerating digital transformation in public administration. UN Department of Economic and Social Affairs.
- Velsberg, O., Westergren, U. H., & Jonsson, K. (2020). Exploring smartness in public sector innovation: A socio-technical perspective. *Government Information Quarterly*, 37(3), 101469.
- Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. *Public Management Review*, 21(7), 1076–1100.