# Bureaucrat-Expert Collaboration in LLM Adoption: An Institutional Logic Perspective on China's Public Sector

**Abstract**

This study investigates how collaboration between bureaucrats and experts shapes the development and application of Large Language Models (LLMs) in the Chinese public sector. The existing research has increasingly focused on how artificial intelligence (AI) drives transformations in public services, while our study investigates the less-explored but equally important aspect of how the AI's role and functionality are shaped by the dynamic interactions and negotiations among cross-sector stakeholders. Drawing on institutional logic theory, this study examines the interplay of state and professional logics in shaping AI applications, using a case study of LLM adoption by a Chinese local government. We identified key challenges—model selection, data security, censorship integration—and strategies to navigate these issues. Our findings highlight how bureaucratic priorities for safety and stability contrast with expert-driven innovation, creating tensions that require negotiation and compromise.

*Keywords:* Large Language Models, AI Governance, Institutional Logic, Cross-Sector Collaboration, Public Administration

## Introduction

Artificial intelligence (AI) has become a cornerstone of technological innovation, driving transformation across numerous sectors, including public administration (Legner et al. 2017; Misuraca, van Noordt & Boukli, 2020; Mikalef, Fjørtoft & Torvatn, 2019; Hjaltalin & Sigurdarson, 2024). Governments worldwide are leveraging AI for tasks ranging from data analysis and resource optimization to citizen engagement and decision-making support (Wirtz, Weyerer & Geyer, 2019; Dwivedi et al. 2021; Nasseef et al. 2022). This growing adoption has sparked extensive debates about the promises and perils of AI in public governance (Mikalef et al. 2023).

Existing literature on AI adoption in the public sector often focuseseon factors influencing whether AI is adopted, such as political will, technological infrastructure, and budgetary constraints (De Vries, Tummers & Bekkers, 2018; Dwivedi et al. 2021; Wang, Zhang & Zhao, 2022; Madan & Ashok, 2023). While these studies provide valuable insights into the prerequisites for AI adoption, they frequently overlook the critical collaborative processes that shape the design and deployment of AI applications. This oversight is particularly significant given that AI systems in public administration are not static technological tools but evolving socio-technical systems shaped by the interactions between government actors, technical experts, and embedded institutional frameworks. Understanding how these interactions influence the design and functionality of AI systems is essential for ensuring their alignment with public value and service delivery effectiveness.

Moreover, much of the existing literature on AI in public administration focus on traditional AI techniques, such as classical machine learning and predictive analytics, which primarily aim to automate routine tasks and enhance operational efficiency (Vogl et al., 2020; Dhungel et al., 2021). However, the emergence of advanced AI models, particularly large language models (LLMs) like those developed by OpenAI, have introduced new possibilities for complex, context-sensitive applications. These models, with their ability to process and generate nuanced textual data, hold particular promise for enhancing decision-making processes in governance. Yet, the unique challenges and opportunities presented by LLMs in public sector contexts remain underexplored. For instance, how can public institutions effectively integrate LLMs while addressing issues such as transparency, accountability, and equity? How do collaborative dynamics between government actors and AI experts influence the design of these cutting-edge systems? These questions are particularly pressing given the heightened expectations surrounding LLMs compared to earlier AI technologies.

In addition to these technological considerations, the literature often portrays AI adoption as a unidirectional process, where technology reshapes bureaucratic processes (Davenport & Ronanki 2018; Androutsopoulou et al. 2019; Kuziemski & Misuraca, 2020; Shareef et al. 2021; Ranerup & Henriksen, 2022; Nasseef et al. 2022; Schiff, Schiff & Pierson, 2022; Ahn & Chen, 2022). This perspective neglects the socio-technical nature of this relationship (Rinta-Kahila et al., 2022; Sanina, Balashov & Rubtcova, 2023)—how institutional logics, governance frameworks, and stakeholders interactions actively shape AI functionalities. Public administration, with its unique ethical, legal, and operational imperatives, requires a more nuanced understanding of how AI technologies are co-designed by governments and technical experts to meet specific public sector needs.

This study addresses these gaps by examining how government bureaucracies and expert collaborators shape the functionality of AI applications in the public sector, with a particular focus on LLMs. Using a case study of a Chinese local government's adoption of AI technologies, this research applies the theoretical lens of institutional logic to explore the interaction types between government officials and technical experts. Our findings reveal several challenges and tensions that emerged during the development of LLM services for the Chinese public sector. Key issues include fundamental model selection, content security, user privacy, and functional versatility, each reflecting broader concerns around cost, ethics, and performance. These challenges illustrate the interplay between different institutional logics, as university researchers prioritized innovation and long-term digitalization while government officials emphasized safety, cost-efficiency, and adherence to political directives. Through strategies such as open dialogue, perspective-taking, and technical adaptations, the university team sought to bridge multiple perspectives with their government counterparts, ultimately persuading officials to adopt LLM-based solutions over traditional approaches like knowledge graphs. These efforts highlight the intricate negotiation processes necessary to align advanced AI capabilities with institutional priorities and public service requirements.

Additionally, our study demonstrates how collaborative strategies address both technical and institutional concerns. Privacy concerns were addressed through robust anonymization and encryption measures, fostering greater willingness among officials to share data essential for effective model training. The decision to use open-source LLMs reflected a pragmatic compromise between performance, cost, and accessibility, while debates over incorporating sentiment analysis for prioritizing requests underscored ongoing ethical and practical dilemmas. Together, these findings illuminate the dynamic socio-technical negotiations involved in adapting cutting-edge AI to the complex demands of public administration.

This study contributes to the growing body of literature on AI in public administration by addressing critical gaps in understanding the socio-technical and institutional dynamics underlying the development of LLM-based services in government. Existing research on traditional AI adoption frameworks (De Vries, Tummers & Bekkers, 2018; Dwivedi et al., 2021; Madan & Ashok, 2023) often neglects the collaborative processes that shape AI system design and functionality (Rinta-Kahila et al., 2022; Sanina et al., 2023). Our findings extend this discussion by examining how the interplay between government actors and technical experts, operating under divergent institutional logics, influences the co-design of LLM applications in the public sector. This perspective highlights the socio-technical negotiations required to align innovative AI capabilities with the operational and ethical imperatives of public administration (Thornton et al., 2012; Reay & Hinings, 2009).

By situating our analysis within the context of China's public sector, this study also expands the institutional logic framework, traditionally applied in democratic settings, to an authoritarian governance context (Zhou et al., 2017; Vogel et al., 2022). We demonstrate how the state's unique incentive structures and relational dynamics shape the resolution of challenges related to cost, security, privacy, and functional performance in LLM adoption. Furthermore, we identify and theorize multiple collaborative strategies—such as addressing cognitive dissonance, implementing censorship, and negotiating data-sharing protocols—as interrelated approaches for overcoming these challenges (Emerson, 2012; Bryson et al., 2006). These insights offer a nuanced understanding of cross-sector collaborations and institutional logics in the design of generative AI-based public services, providing actionable implications for both scholars and practitioners in governance and technology adoption (Boxenbaum & Jonsson, 2017; Greenwood et al., 2011).

## Literature Review

Artificial Intelligence (AI) technologies are increasingly integrated into public services, including machine learning (Pan et al., 2017), computer vision (Sun et al., 2022), reinforcement learning (Kwak et al., 2021), robotics and automation (Abdi et al., 2018). Despite scholarly attention to these AI technologies, large language models (LLMs) have experienced recent breakthroughs such as the Transformer architecture (Vaswani et al., 2017), the Generative Pre-trained Transformer (GPT) models (Brown et al., 2020), LLMs may revolutionize government services by offering advancements beyond the capabilities of traditional chatbots, but this crucial domain remains underexplored.

LLMs surpass traditional rule-based chatbots in handling complex queries and generating natural, context-aware responses. Traditional chatbots rely on rule-based or limited machine learning frameworks, using predefined rule sets or decision trees to parse user inputs and provide corresponding responses (Zhen, Zhao & Stylianou, 2020). While effective for structured queries, these systems often falter when faced with complex or unanticipated user inquiries. In contrast, LLMs employ deep learning and natural language processing techniques to understand and generate human language. By training on extensive text datasets, LLMs develop a broad understanding of language and robust generative capabilities, enabling them to address diverse user needs and produce more natural, human-like responses (Rahwan et al., 2019). Moreover, LLMs can be fine-tuned for specific tasks, allowing them to engage in highly context-aware and nuanced dialogues, further enhancing their practical utility (Dong et al., 2023).

Governments across the globe are integrating large language models (LLMs) and generative AI into public service systems, aiming to enhance efficiency and innovation. In the United States, the General Services Administration (2023) has introduced policies to facilitate access to generative AI technologies, supporting their integration into public service delivery. Similarly, China's State Council (2024) emphasizes the transformative potential of emerging technologies, including big data, blockchain, and artificial intelligence, advocating for a shift from manual processes to human-computer interaction and from experiential judgment to data-driven decision-making. On a more localized level, the Anhui Provincial Government (2023) has unveiled plans to advance LLM adoption through extensive training programs, public awareness campaigns, and strategic research facilitated by government procurement. Additionally, Anhui aims to construct intelligent computing centers, including specialized government platforms, to support these technological advancements. These initiatives highlight a global trend toward leveraging LLMs to revolutionize governance and public administration.

In the realm of public services, research has extensively compared large language models (LLMs) to traditional chatbots, focusing on their differences in versatility, privacy, efficiency, and cost. Among these, LLMs such as GPT-3 stand out for their exceptional versatility in addressing diverse public service applications. Leveraging their ability to comprehend and generate natural language, LLMs can handle a wide array of tasks, from automating responses to public inquiries to drafting complex policy documents. Rahwan et al. (2019) underscore LLMs' adaptability to various contexts and tasks—an advantage far beyond the limited, predefined responses of traditional chatbots. For instance, when equipped with relevant data and minimal prompting, pre-trained LLMs have demonstrated the ability to perform highly specialized functions, such as detecting inconsistencies in legal frameworks (Nay et al., 2024) or resolving disputes in online settings (Westermann et al., 2023). These capabilities highlight LLMs' potential to revolutionize public service delivery, setting them apart as a transformative tool in governance and administration.

Also, privacy preservation remains a critical concern in the use of LLMs for public service, highlighting both risks and emerging solutions compared to traditional chatbots. Peris et al. (2021) highlight the significant privacy risks associated with LLMs, including their tendency to memorize and potentially disclose sensitive information. This characteristic raises concerns when deploying LLMs in contexts where confidentiality is paramount. In contrast, traditional chatbots, with their limited functionality and narrower scope, inherently reduce the likelihood of such breaches, as they process and store less complex data. However, recent advancements in privacy-preserving technologies, such as federated learning and differential privacy, are beginning to address these challenges. These techniques allow LLMs to perform their functions without directly accessing or storing sensitive information, creating a pathway to harness their powerful capabilities in public service while minimizing privacy risks (Adnan et al., 2020). By integrating these safeguards, governments and organizations can confidently explore LLMs as tools for innovation in public administration.

In addition, efficiency is a key advantage of LLMs in public service, encompassing both the speed and contextual accuracy of their responses compared to traditional chatbots. Kumar et al. (2023) demonstrate that LLMs can produce high-quality responses more rapidly, streamlining interactions in public service contexts. This efficiency extends beyond speed, as LLMs excel in generating relevant, contextually appropriate responses—a critical feature for handling the diverse and complex inquiries often encountered in public administration. Zhang et al. (2023) emphasize the importance of contextual understanding in public service applications, where the ability to process nuanced language and provide precise answers can drastically improve user satisfaction and operational effectiveness. By integrating LLMs into public service systems, governments and organizations can achieve the dual benefit of faster service delivery and enhanced response quality, positioning LLMs as a transformative tool for modern governance.

LLMs incur high initial costs, their scalability and multitasking capabilities position them as a cost-effective solution for public service in the long term. For instance, GPT-4's training cost over $100 million (Meyer, 2024). Treviso et al. (2023) further detail the considerable expenses associated with data acquisition, processing time, storage, and energy requirements for training LLMs. Deploying pre-trained LLMs in public services also incurs ongoing costs, such as maintaining computational resources, conducting training, and supporting operations. In contrast, traditional chatbots demand less computational power and are generally more cost-effective for narrow, task-specific applications. However, the scalability and multitasking capabilities of LLMs can offset these initial expenses by reducing redundancy and improving operational efficiency over time. Mani et al. (2023) argue that this long-term cost-effectiveness makes LLMs a viable and transformative investment for governments aiming to modernize public service delivery.

The comparative analysis underscores both the advantages and challenges of deploying LLMs in public services, presenting a compelling case for their transformative potential despite certain trade-offs. Technological advancements, such as privacy-preserving techniques, and strategic planning, like optimized resource allocation, can effectively mitigate privacy concerns, ensuring secure implementation. Moreover, while the initial deployment of LLMs entails substantial costs, their long-term efficiencies and scalability significantly offset these expenses, offering economic profitability that traditional chatbots cannot achieve. These comparisons demonstrate that LLMs hold considerable promise for application in public services, providing governments with a powerful tool to enhance operational effectiveness and citizen engagement. By carefully balancing incentives and concerns, the deployment of LLMs emerges as a promising and innovative direction for modern governance.

| Dimension | Large Language Models (LLMs) | Chatbots |
|---|---|---|
| Algorithm | Deep learning using Transformer neural networks | Rule-based, retrieval-based, or simple machine learning models |
| Versatility | High | Low |
| Privacy | More privacy concerns | Fewer privacy concerns |
| Efficiency | Efficient for simple and complex tasks | Less efficient, insufficient for complex tasks |
| Cost per request | Substantial for deployment, low for long-term adjustment | Low for deployment but substantial for long-term adjustment |

Table 1. Comparison between LLMs and Chatbots

While existing research has focused on enhancing the technical performance and cost-efficiency of LLMs, their practical application in public sector contexts remains underexplored. Existing studies often evaluate the social application of LLM within controlled laboratory environments, relying on technical metrics such as F1-scores, precision, and recall as benchmarks for success (Costello et al., 2024; Luo et al., 2024; OpenAI, 2023). While these studies provide valuable insights into the technical potential of LLMs, they do not fully capture the complexities and unpredictable challenges of real-world implementation, particularly in the public sector. The sanitized conditions of laboratory settings may overlook critical factors such as diverse user needs, policy constraints, and integration with legacy systems. Consequently, the tangible impact of embedding advanced LLMs into public sector mechanisms remains insufficiently understood, highlighting an urgent need for empirical research in real-world public service contexts.

*AI in Government Globally*

The adoption of AI in public sectors has accelerated in recent years, encompassing a diverse range of transformative applications. These advancements include AI-driven process automation systems, virtual agents, cognitive robotics, and autonomous systems designed to streamline operations and enhance efficiency (Ahn & Chen, 2020; Wirtz et al., 2019). Beyond these foundational technologies, AI adoption also extends to recommendation systems and intelligent digital assistants that improve decision-making and service delivery. Moreover, AI is increasingly applied in critical sectors such as public safety, healthcare, education, and urban planning, where its potential to analyze data, predict outcomes, and optimize resources is transforming traditional approaches(Wirtz et al., 2019; Dwivedi et al. 2021; Nasseef et al. 2022). This multifaceted integration underscores the growing reliance on AI as a key driver of public sector innovation, paving the way for more efficient, accessible, and effective governance.

Governments worldwide are leveraging AI to enhance governance while grappling with its ethical challenges and implications for accountability. Governments are increasingly adopting AI to bolster administrative capacity, streamline decision-making, and enhance citizen-government interactions (Medaglia et al., 2023; Medaglia & Tangi, 2022). AI's transformative potential offers significant benefits to public governance, from automating routine tasks to providing data-driven insights for policy decisions. However, as a double-edged sword, AI also raises pressing ethical concerns. These include fears of human replacement in critical roles, the opacity of AI-driven decision-making, and questions about accountability for decisions made by autonomous systems (Boyd & Wilson, 2017). Given these complexities, government agencies must approach AI adoption strategically, balancing its advantages against potential risks. This requires implementing robust regulatory frameworks, fostering transparency in AI systems, and ensuring that ethical considerations are integral to their deployment.

The adoption of AI in the public sector is accompanied by unique legal, ethical, technical, and societal challenges, distinct from those of earlier IT innovations. Research indicates that many obstacles to AI adoption mirror those encountered during the introduction of earlier information technology (IT) innovations, as both can introduce uncertainty and disrupt established governance processes (Selten & Klievink, 2024; Meijer, 2015). However, AI differs from traditional IT innovations in its capacity for autonomous behavior, often encapsulated by the "black box" phenomenon, which obscures the inner workings of AI systems and complicates monitoring and regulation (Radanliev & De Roure, 2023; Burrell, 2016). This opacity brings a tension between exploiting AI's capabilities and exploring its implications, amplifying concerns related to transparency, accountability, data security, and ethical governance (Grimmelikhuijsen

& Meijer, 2022; Selten & Klievink, 2024). Addressing these unique challenges requires a nuanced approach that balances innovation with safeguards to maintain trust and accountability in public administration.

Additionally, the political context is curcial in shaping the adoption and governance of AI, as countries align technological strategies with national priorities and global competitiveness. Both developing and developed countries recognize AI's potential to enhance administrative efficiency and secure a competitive edge in the global economy. For instance, China and the United States have made significant investments in AI-driven governance, leveraging its capabilities to improve public services while simultaneously solidifying their leadership in the global AI race (Wirtz et al., 2019). Similarly, Europe has carved out a strategic approach by emphasizing the use of data from enterprises and government entities to fuel AI innovation, positioning itself as a key player on the global stage (Groth & Straube, 2021). These examples illustrate how the intersection of technological advancement and political priorities shapes AI adoption, reflecting broader national goals and international aspirations.

Political structures and cultural values significantly influence the determinants and approaches to AI adoption, shaping distinct pathways across different regions. For example, research shows that individualism culture is positively associated with the development of open government data initiatives, which is closely associated with how AI systems are adopted (Zhang et al., 2023). Similarly, public expectations and trust in data-driven technologies in England are conditional, with a strong emphasis on transparency, inclusiveness, and accessibility (Wong et al., 2023). In China, prior studies highlight that vertical administrative pressure, driven by directives from higher levels of government, and horizontal competition pressure among local governments are key factors influencing technological innovation (Wang et al., 2022). Conversely, in Europe, the integration of AI within government sectors is shaped by a complex interplay of environmental, organizational, and contextual factors, underscoring the importance of high-quality data as well as interconnected institutional dynamics (Mikalef et al., 2023). From a cultural perspective, the policy cycle further illustrates how deeply embedded societal values impact AI governance. In the United States, the Protestant ethic emphasizes individual responsibility and decentralized innovation, influencing technology policies to prioritize autonomy and market-driven advancements. In contrast, China's Confucian ethics, central to its development strategy, advocate for centralized authoritarian guidance, balancing the push for AI innovation with the imperative to maintain social stability (Hine & Floridi, 2024). These variations demonstrate how political and cultural contexts create divergent strategies for integrating AI into governance, reflecting broader societal goals and priorities.

*Institutional Logic Perspective*

Cross-sector collaboration has been long discussed as an effective way of improving the quality of public service (Bryson et al. 2006; Emerson 2012; Quélin et al. 2017). Different stakeholders own different resources and knowledge, and the cooperation between multiple sectors enables resource integration and functional extension (Gary 1985; Gray and Purdy 2018; George et al. 2024). Cross-sector collaboration has also recently been recoginzed by the existing studies about technology adoption in the public sector (Wirtz et al. 2019; Mikhaylov et al. 2018; Charles et al. 2022; Williams et al. 2023). In a nutshell, the lack of government in the knowledge of computer science provides an incentive for effective cross-sector collaboration.

The cross-sector collaboration has the potential to integrate multiple advantages but also leads to new problems of coordination and conflicts between different sectors. Different stakeholders have different values (Lounsbury 2007), priorities and risk management strategies (Klijn & Teisman 2003). For instance, public organizations bear responsibility towards their service recipients as well as the broader public community, whereas private entities primarily answer to their shareholders (Nutt 2006). The involvement of multiple stakeholders can also lead to the failure of technology adoption projects in the public sector (Gauld 2007). In a similar vein, the ignorance of coordination difficulty between multiple stakeholders in overall planning renders ineffective budget and time control (Anthopoulos et al. 2016). Also, effective and fair negotiation was undermined by the unequal power relationship between different stakeholders (Curtis 2008).

Institutional logic provides a helpful framework for us to understand the sources of conflicts between the stakeholders from different sectors (Baxter et al 2023). The perspective of institutional logic facilitates our comprehension of organizations by examining the interrelations between individuals, organizations, and society within the confines of an institutional environment (Friedland 1991). These logics not only underscore the interconnectedness that promotes collective goals and solidarity across an organizational landscape but also highlight the inherent diversity, as manifested through the presence of simultaneous, competing paradigms (Reay and Hinings 2009). The binding of different logic also deeply affects the quality of the cooperation process and the eventual outcome of the project.

Some adaption is required when contextualizing? this theory to analyze the cross-sector collaboration in China. Despite a few exceptions (Zhou et al. 2017; Zhu et al. 2021), the institutional logic theory has predominantly focused on the institutional logic of democratic governments within developed nations (Thornton et al. 2012; Ma et al. 2022). As Vogel et al. (2022) suggest, the framework of governmental logic is tailored to democratic systems and does not universally apply to authoritarian regimes. China is an authoritarian state, and the logic of government is political incentive rather than the incentive of bottom-up accountability. To note, it does not mean the Chinese government is not responsive, as suggested by the previous studies (Chen et al. 2016), but rather indicates that state actors in authoritarian states may share different sources of legitimacy with their counterparts in democracies. Second, given the primacy of inter-personal relationships/connections? (or guanxi) in Chinese society (Yan 1996), the communication and trust-building process may also vary from the situation in Western society and the importance of relational leadership should be further emphasized (Kinder et al. 2021).

In this study, we focus on the collaboration between the government and experts in developing an innovative pattern for resolving the public's complaints by using generative AI, and more specifically, the Large Language Model. Depart from cross-sector collaboration largely focuses on public-private sector collaborations (Petrescu 2019), here we examine the interplay between the government and the expert in the public service upgradation process. As an exploratory study, we do not aim to illustrate all the potential challenges, but to explore how the different logics of the government and experts interact and jointly shape the resolution of prominent challenges. Several strategies were also indicated by the previous studies as effective ways of facilitating cross-sector collaborations, such as the cultivation of Facilitative leadership (Day et al. 2014), social capital (Smith et al. 2004), shared objectives (Boyne and Chen 2007) as well as knowledge gathering and sharing (Chen and Lee 2021). While these strategies have been identified as powerful catalysts for cross-sector collaboration, their implementation is not universally straightforward or feasible in all contexts. This paper delves into the specific challenges and nuances of applying these approaches within the Chinese context, highlighting the complexities of building effective cross-sector collaborations in developing generative AI-based service projects in the public sector.

The development of Language Model (LLM)-based services in the public sector is influenced by the coexistence of state and professional logic. These dual logics often generate both synergies and tensions, influencing the strategic choices of government agencies (Thornton et al., 2012; Scott, 2008). Unlike private organizations, government bodies do not merely respond to institutional logic; they are pivotal in shaping this logic, wielding both regulatory powers and the responsibility of public service (Waardenburg et al., 2020; Greenwood et al., 2011).

State logic typically prioritizes governance, regulatory compliance, and public accountability, aligning closely with bureaucratic principles that emphasize control and standardization (Meyer & Rowan, 1977). On the other hand, professional logic in the context of LLM services champions innovation, technical excellence, and adaptability, often driven by advancements in artificial intelligence and computational linguistics (Pache & Santos, 2013). The tension between maintaining regulatory standards and fostering innovation is a central challenge in this domain (Di Domenico et al., 2010; Gray & Purdy, 2018; Ma et al., 2022).

Faced with institutional pressures rooted in single or multiple logics, the actors still own the autonomy to select different strategies to respond. Oliver (1991) identifies five strategic approaches that organizations may adopt when confronted with external pressures: acquiesce, compromise, avoid, defy, and manipulate. The "acquiesce" strategy involves full compliance with external demands or norms, reflecting a passive acceptance. In contrast, "compromise|" denotes a negotiation process where the organization seeks a middle ground to partially satisfy both its interests and external expectations. The "avoid" strategy is characterized by efforts to circumvent the necessity of conforming, such as concealing non-compliance or buffering the organization from external scrutiny. The strategy "Defy" represents an open challenge to external pressures, where the organization actively resists or rejects imposed demands. Lastly, the "manipulate" strategy entails attempts to influence or control external institutions or stakeholders to align them with the organization's objectives.

Research on decoupling strategies suggests that government agencies might symbolically adopt innovative practices to satisfy professional logic while substantively adhering to state logic to maintain legitimacy and compliance (Boxenbaum & Jonsson, 2017). However, an emerging strategy is the integration of these competing logics. This approach not only enhances the legitimacy of government initiatives but also ensures that innovations are robust and compliant with established norms (Lounsbury, 2007).

Professionalization within government entities that manage LLM-based services is also critical. This involves adopting strategic planning, independent evaluations, and rigorous standards to align with both state and professional logic (Hwang & Powell, 2009). Such professionalization is essential to navigate the complexities of implementing technically advanced services that meet both administrative mandates and evolving technological standards.

**Research Design**

As one of the most advanced algorithms, LLM still has not been widely adopted in the government sector. Thus, we selected a single-case approach to provide detailed insights into a specific instance of collaboration between bureaucrats and experts in LLM adoption in public service. This case is suitable for a single-case study because it represents a pioneering effort to deploy LLMs for public service within a unique institutional and governance context, combining the practical challenges of authoritarian governance with the technical complexities of AI adoption. Single-case studies are commonly used in public administration research to explore rare, noteworthy, or particularly illustrative phenomena. This case is uniquely positioned to contribute because it captures both the political and technological nuances of LLM adoption in a moderately developed region, offering insights that extend beyond high-profile cases in developed nations. This case, a partnership between a Chinese local government and experts offers an opportunity to explore how the complexity of differing institutional priorities influences the adoption of LLMs.

*Case Description*

The case focuses on the development of an LLM-based public service platform aimed at improving the classification and response to citizen complaints . The project involved multiple phases of collaboration between government officials and technical experts, highlighting distinct institutional priorities and negotiation strategies. The case is particularly relevant for understanding the challenges and strategies of integrating LLMs into public administration within an authoritarian governance context in the Global South. This collaboration is contextualized in a small city J in a province of moderate development in China. J is located in the north of China with a gross domestic product (GDP) per capita (2023) of around 13,300 US dollars and a population of 4.67 million.

In our case, the development of the LLM application is driven by a multi-stakeholder collaboration involving officials, a university professor, experts, and government technicians with varying interests and responsibilities. The project is initiated by officials who define strategic objectives, allocate resources, and provide endorsement. Then, a university professor introduces its students to this task and ensures coherence in project objectives and resource allocation. As experts in this field, the professor's graduate students translate government requirements into actionable solutions, including requirement analysis, feasibility assessment, LLM development, and validation. Finally, government technicians oversee the technical implementation and ensure alignment with the government's demand. Upon project completion, they assess security, risks, and quality. Overall, this multi-stakeholder collaboration reflects diverse institutional interests that collectively shapes the adoption of LLMs in public service, which can be understood through the perspective of institutional logic.

This study employed a combination of semi-structured interviews, participant observation, and document analysis. First, 26 semi-structured interviews were conducted with government officials and technical team members to explore their perspectives on the evolving collaboration. These interviews focused on several key aspects of LLM application development and deployment, such as safety, compliance, and efficiency, as well as technical concerns, including model performance, hallucination mitigation, and censorship integration.

Second, participant observation was employed to document real-time interactions during meetings, technical development sessions, and feedback discussions. This approach provided valuable insights into decision-making processes, the negotiation of competing priorities, and the power dynamics inherent in the collaboration. Observations were recorded as detailed field notes, which were periodically reviewed to identify patterns and themes. This method complemented the interviews by capturing the subtleties of stakeholder behavior and organizational culture that might not be explicitly articulated.

Third, document analysis further enriched the study by providing contextual and triangulated data. Key documents, including policy guidelines, project progress reports, preliminary test results, and meeting minutes, were analyzed to understand the institutional framework within which the LLM was being developed. Document analysis also offered insights into the technical specifications of the implemented models and recorded stakeholder feedback on the iterative adjustments made during the project. This method ensured that the study was grounded in both formal documentation and participants' narratives, enhancing the reliability and validity of the findings.

Ethical considerations in this study centered on stakeholder engagement and data collection. Government bureaucrats and university scholars provided informed consent after being briefed on the study's objectives, scope, and methodology. Participants were explicitly informed of their right to withdraw at any stage without repercussion, ensuring voluntary participation. Researchers addressed potential power imbalances by encouraging open communication and valuing participants' perspectives throughout the research. Moreover, confidentiality was prioritized by anonymizing and securely storing sensitive data, such as complaint records and internal documents. Respectful and constructive interactions were emphasized to align with stakeholder expectations and institutional norms. Reflexivity was a guiding principle; the research team critically examined their own roles and potential influences on the findings. Finally, participants were regularly consulted and invited to review preliminary results, fostering mutual accountability and bolstering the study's reliability.

*Collaboration Process*

The collaboration process consists of four phases:

1. Planning (January–March 2023)

The collaboration began with the government seeking technical support to address an overwhelming influx of citizen complaints related to post-pandemic urban challenges, such as infrastructure inadequacies and consumer grievances. The initial goal was to automate and scale the classification of complaints while ensuring compatibility with limited computational resources.

2. Model Training I: Knowledge Graph-Based Model (March–June 2023)

The technical team proposed a knowledge graph-based classification model capable of efficiently categorizing complaints in real time. This solution demonstrated satisfied performance in high-demand scenarios and met government requirements for low-resource environments.

3. Model Training II: LLM-Based System Development (June–November 2023)

As the government's needs evolved, the focus shifted to creating a large-scale, robust LLM-based customer service system with enhanced generalization and user interaction capabilities. The solution integrated the original classification model with a large language model (LLM) to improve response accuracy and user experience.

4. Deployment and Evaluation (November 2023–March 2024)

The government evaluated the LLM's performance, emphasizing data security, user privacy, and content control. In response, the technical team implemented privacy-preserving measures, refined the model to ensure legal and ethical compliance, and optimized it for efficient deployment in resource-constrained environments.

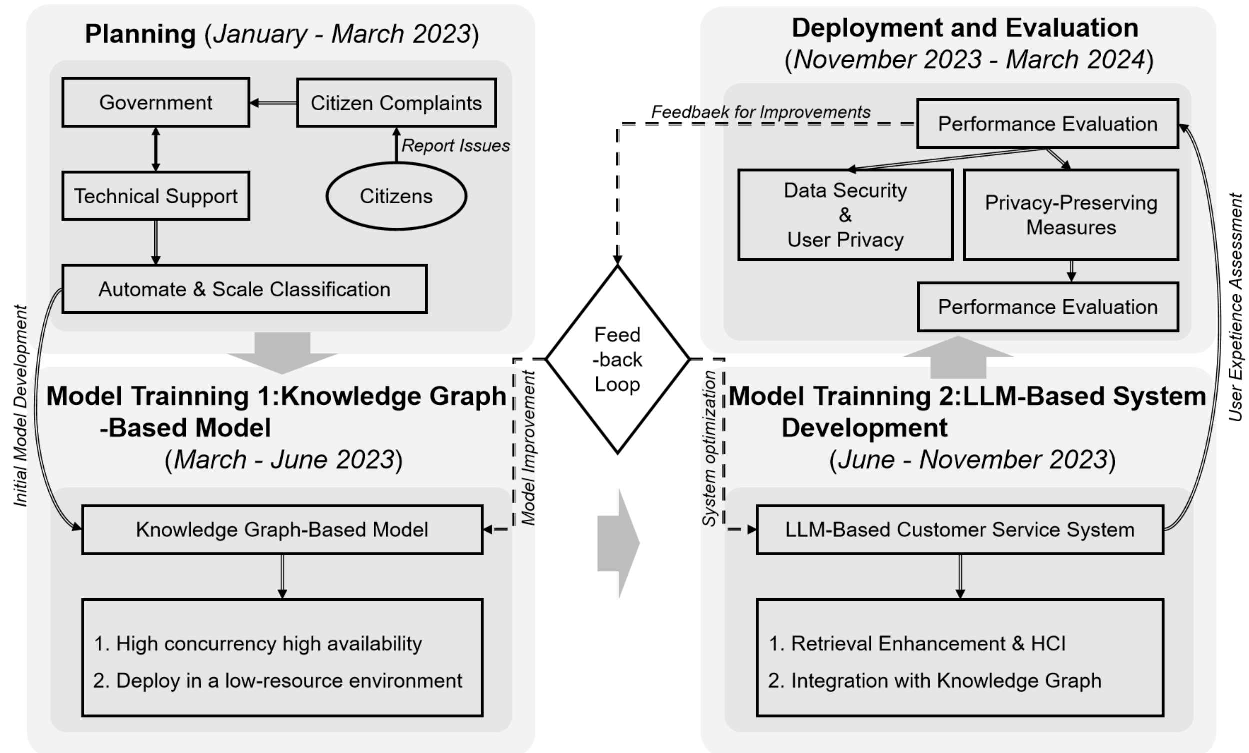


Figure 1 The four stages of collaboration process

## Findings

Through interviews with the project team from University A and the government officials they collaborated with, our findings indicate several challenges in the different stages of developing LLM services in government: fundamental model selection, generated content security, user privacy, and functional versatility. These challenges reflect concerns about cost, ethics, and functional performance of the LLM project, revealing both sharedand divergent perspectives due to different institutional logic. Our findings also demonstrate that multiple strategies can be adopted by the project team to cope with the tensions between them and government officials. It is important to note that these strategies should not be interpreted as separate prescriptions for individual challenges but rather as intercorrelated approaches, as we will show later.

| Issues | State Logic | Professional Logic | Strategy | Outcome |
|---|---|---|---|---|
| Technique Selection (Knowledge Graph vs. LLM) | Prioritize mature, stable, secure technology (Knowledge Graph); Lower initial cost, lower security risk | - Prioritize long-term, versatile technology (LLM)<br>- Higher performance, better conversationality, future adaptability (multi-modal)<br><br>- Potential long-term cost savings (replace human staff)<br><br>- Align with central gov't directives on AI adoption<br><br>- Potential for recognition and awards | Challenge<br>- Emphasize long-term benefits<br><br>- Highlight limitations of Knowledge Graph<br><br>- Leverage central government directives | Government adopts LLM |
| Censorship Module | - Prioritize political safety, service survival<br><br>- Avoid generating politically incorrect content | - Concern about service inclusiveness, data diversity<br><br>- Recognize potential negative impact on model training | Acquiescence<br><br>- Acknowledge political risk<br><br>- Prioritize government's concerns | Censorship module added |
| Data Sharing | Prioritize data security, minimize data leakage risk<br><br>- Initially provided only 3,000 cases | - Emphasize data necessity for model performance<br><br>- Demonstrate performance improvement with increased data | Challenge<br><br>- Show negative evaluation results with limited data<br><br>- Introduce data anonymization and | Government shares more data with privacy protection measures implemented |

| | | | | |
|---|---|---|---|---|
| | | | privacy protection measures | |
| Open Source vs. Commercial Models | - Prioritize cost-effectiveness<br><br>- Avoid using banned services (e.g., OpenAI) in authoritarian states | - Consider performance, cost, data security, accessibility<br><br>- Open Source: Lower performance, but free, secure, and accessible<br><br>- Commercial: Higher performance, but costly, potential data security risks, and accessibility issues | Acquiesce<br><br>- Educate: Present pros and cons of both options | Both agree on Open Source LLM |
| Emotion-Based Request Prioritization | - Concern about moral hazard (users exaggerating urgency<br><br>- Note that emotion is also influenced by personalized tones of the request senders | - Emphasize efficient resource allocation<br><br>- Enable empathetic responses by addressing emotional states<br><br>- Technically feasible, minimal cost | Strategic Suspension (Defer)<br><br>- Agree that it needs further research<br><br>- Focus on project timeline<br><br>- Plan for future research | Further research needed, no delay on project; Future collaborative ranking experiments planned |

Table 2 State and professional logic and responding strategies

*Model Selection With Multiple Expectations*

The lack of technical knowledge about LLMs left the government confused about appropriate techniques and specific formats for government services based on LLMs. Initially, the government planned to develop a project to automatically classify complaints into different categories. For example, the model could act as a telegraph recipient, correctly sorting and sending the public's requests to the appropriate government departments. For instance, the model should be capable of sending a request like "Pothole on Main Street needs repair" to "Municipal Engineering Administration."

The government formulated an initial plan using WeChat small programming (微信小程序) as a front-end interface and a knowledge graph as back-end technology . They gathered examples of knowledge graph-based WeChat chatbots from big cities such as Beijing and Shanghai, aiming to develop a similar service. Although government officials are aware of recent technological breakthroughs in LLMs, they are reluctant to adopt LLMs in their service due to concerns about data security and model hallucination. The government officials has explicitly expressed that theyprioritize safety in public  services.

The university team adopted a “challenge” strategy regarding the fundamental technology for the project. They proposed using LLMs based on considerations of the government’s long-term digitalization strategy and their own research needs. The university team agreed that the knowledge graph approach is more mature, stable, and transparent than LLMs, with lower computational and latency costs when handling user inquiries, but they were also aware of its weaknesses, particularly its lack of conversational capacity and versatility.Since the service will eventually deal with complaints from the public, responsiveness is crucial for appeasing worries and increasing public satisfaction.

Regarding service development costs, although the knowledge graph approach is cheaper than the LLM approach, it cannot fully replace human staff working in the hotline system responsible for responding to public requests. Therefore, the university team recognized the potential of LLMs to replace human staff as request recipients, with the capacity to be easily upgraded with multi-modal data processing capabilities.

They recognized and acknowledged the government's concerns regarding the high costs associated with training models using the LLM approach. By openly sharing and discussing these concerns, the university team demonstrated their understanding and empathy towards the government’s position. However, the university team highlighted the tradeoff between local and aggregate costs, noting that while the knowledge graph requires less upfront cost, it cannot replace human receptionists. They also referenced documents issued by central and provincial governments promoting AI to reinforce public services in persuading local officials. Since LLMs represent significant AI breakthroughs, their adoption aligns with government directives to improve service quality through AI.

The university team emphasized that adopting LLMs could bring additional benefits for the local government, such as evidence of their efforts to comply with upper-level authority directives and potential awards for leading in public service delivery improvements compared to their peers. In summary, the increased performance, long-term cost reduction, and potential demonstration of commitment to implementing top-down directives convinced the government to eventually agree with the university team’s proposal to use LLMs.

*Incoporating Digital Censorship in The Service AI*

The university team and the government initially had different understandings of model security. The university team primarily focused on mitigating the model's hallucinations and addressed this issue by adding an information retrieval module before generating content. This module acts as a reference book for the LLM to consult when queried by users, significantly improving the model's classification accuracy compared to versions without this module. Conversely, the government was concerned not only with content accuracy but also with political safety. They questioned whether the content generated by the model would align with official ideology, especially on topics related to China's political system and leaders, and the guidelines issued by the central propaganda department.

The university team adopted an acquiescence strategy toward the government's request to add a 'censorship' module. To avoid generating politically incorrect answers, the university prepared a dictionary of politically sensitive words. If a user's prompt includes any of these sensitive words, the model will not generate any related content. For example, if a user asks, "How do you evaluate China's political system?" the presence of the sensitive term "political system" will prevent the model from generating a response, instead returning a message such as, "Please try another topic."

The university team recognized the drawbacks of blocking inputs containing sensitive words. In the short term, this censorship decreases the service inclusiveness by excluding certain requests. Since the censorship is dictionary-based, some reasonable requests containing sensitive words might not receive responses. This issue is particularly pronounced in China, where the public often references central government documents and speeches by top leaders when petitioning local governments (O'Brien & Li, 1999). In the long term, blocking parts of questions reduces input diversity, which could jeopardize the quality of training data and potentially affect the model's capacity in future generations.

Despite these concerns, the university team eventually agreed to install the censorship module, considering the government's perspective on the service's survival. Government officials persuaded the university team by highlighting the 'existential risk' of generating improper content. They stressed that the service could be suspended or even terminated if the model produced politically incorrect information, which could also jeopardize the careers of the officials involved in the project. The university team agreed and prepared a comprehensive list of sensitive words. In summary, the perspective-taking approach was effectively used by the government to convince the university team to add a censorship module to the responsive service.

*Securing Data Sharing for Effective LLM Training*

The university team encountered another issue in the process of model training, specifically the government's initial reluctance to share large amounts of data. Although the necessary training data was fully owned by the single department the university collaborated with, the government only provided 3,000 cases of request data for the university team to build the model. However, the university team found it difficult to build an LLM with good classification accuracy due to the lack of data diversity. When the university team communicated with government officials to request more data, the government expressed concerns about data security, particularly regarding potential data leaks.

The university team adopted a challenge strategy by addressing the government's concerns by emphasizing the goal of better model performance and the adoption of data and privacy protection measures. They shared the negative evaluation results of the model trained only on 3,000 cases, which showed only 50% classification accuracy. They also clearly demonstrated the gradual improvement in model performance with more data fed into the model. Based on this solid evidence, the government realized that data scale is crucial for improving model performance.

The government's willingness to share data further improved when the university team introduced additional measures to remove personal private information before feeding data into the model. For example, they showed how user information was anonymized before and after the data cleaning process. The university team also informed the government about privacy protection measures implemented when the model was queried. The portal website employs secure access controls and anti-scraping techniques to safeguard data access.

*Open Source vs. Commercial Models for Government Services*

The fundamental model for the service is an open-source LLM, based on a consensus between the government and the university team. Using open-source or commercial models for building government responsive services comes with distinct advantages and disadvantages. One key difference is performance. Compared to open-source models, commercial LLMs generally exhibit more powerful capabilities in text generation, reasoning, and other aspects, according to comprehensive evaluations. However, commercial LLM providers charge users for access to their most advanced models, significantly raising costs if used as the foundation for government services. Additionally, the use of LLMs poses risks related to data leakage, which can threaten the privacy of the general public's request data.

Another challenge with commercial LLMs, such as those provided by OpenAI, is that they are banned in some authoritarian states, making it difficult, if not impossible, for governments in these countries to develop their services using OpenAI's API. Even if the government could access OpenAI's API using VPN techniques, it remains highly unstable since OpenAI continuously monitors and bans API requests from prohibited countries, even those disguised via VPN. The university team thoroughly introduced the advantages and disadvantages of both open-source and commercial LLMs to the government. After careful consideration, both sides agreed to use an open-source LLM.

*Unsettled Debate: Ranking Requests Based on Emotion*

The university team and government officials also had a disparity over the issue of prioritizing requests based on the emotions expressed within them. Delivering public services to those in urgent need is an effective way of administrative resource allocation, and urgency can be reflected through the emotional appeals within the text of requests. Additionally, incorporating emotion enables the model to provide responses that address both the specific requests and the emotional states, such as anger and anxiety, expressed by the users. The university team argued that performing sentiment analysis on text is technically feasible and would not generate substantial costs by adding a function to display the sentiment scores of citizens' requests.

However, government officials expressed concerns about the potential negative effects of incorporating sentiment analysis for ranking requests. They pointed out that using sentiment scores to determine priority could lead to a moral hazard where users might exaggerate the urgency of their requests. Additionally, they noted that the emotion in requests is not solely dependent on the issues reported but is also influenced by the tones of the request senders, which is highly personalized. Although the government acknowledged the potential benefits of using emotion to rank requests, they insisted it might not be suitable for implementation at the initial stage.

The university team and government officials adopted a strategic suspension approach regarding the issue of emotional analysis. Both sides agreed that emotional analysis for priority ranking requires more research and deliberation. More importantly, there are other ethical problems that both sides do not yet have agreed answers to, which is similar to the debate about AI's role in public administration decision making. However, the both sides also concurred that these unresolved issues should not delay the project schedule, which was a shared consensus. Furthermore, the university team plans to conduct more scientific research on this aspect by inviting government officials to rank the requests and then training a machine learning model to predict priority ranking, an approach that has been embraced by the government.

**Discussion**

This study highlights the evolving interplay between bureaucrats and experts in the design and deployment of LLM based AI application in Chinese public sector. The collaborative development process revealed both synergies and tensions stemming from institutional logic differences. While LLMs offer tremendous versatility and efficiency in automating public service functions, the study found significant challenges in areas such as model security, data sharing, and alignment with government priorities.

By leveraging the theoretical lens of institutional logic, our findings address key gaps in the literature. Specifically, while existing studies have predominantly focused on the technical and infrastructural prerequisites for AI adoption (De Vries, Tummers & Bekkers, 2018; Dwivedi et al., 2021), this research highlights the collaborative processes that shape the socio-technical design of LLM applications. The study underscores the role of state and professional logics in influencing decision-making, thereby advancing academic discussions on cross-sector collaboration (Thornton et al., 2012; Reay & Hinings, 2009).

The findings also challenge the traditional notion of AI adoption as a unidirectional process, wherein technology reshapes bureaucratic functions (Davenport & Ronanki, 2018; Androutsopoulou et al., 2019). Instead, this study reveals a bidirectional relationship, where institutional priorities and collaborative negotiations actively shape AI functionality. For instance, the government's preference for data security and political stability required compromises in the technical design, such as integrating censorship modules, which reflect broader socio-political imperatives. By situating this discussion in the context of LLM adoption, the study provides a nuanced understanding of how socio-technical negotiations underpin public sector innovation.

In addition, this study reveals the intricate interplay of institutional logics in shaping the adoption of LLMs in the public sector. Our analysis demonstrates that state logic, characterized by a focus on safety, compliance, and political stability, often contrasts with professional logic, which emphasizes innovation, technical excellence, and adaptability. This tension manifested in various stages of the project, such as the negotiation over model selection, where the government's cautious approach to data security conflicted with the experts' push for leveraging cutting-edge LLM technologies. Ultimately, these negotiations resulted in pragmatic compromises, such as the adoption of open-source LLMs and the integration of privacy-preserving measures. These findings contribute to the broader literature by illustrating how competing institutional logics can be reconciled through collaborative strategies (Lounsbury, 2007; Pache & Santos, 2013; George et al., 2024), including open dialogue, perspective-taking, and evidence-based decision-making.

Our case also suggests that the adoption of LLMs in public services raises important social and ethical considerations. First, the automation of administrative functions may reduce demand for certain roles, such as human request handlers, while creating opportunities for higher-skilled positions. The inclusion of censorship and sentiment analysis highlights ethical concerns about fairness and bias in algorithmic decision-making. Sensitive word blocking may exclude legitimate requests, reducing the inclusiveness of public services. In addition, the reliance on digital infrastructure for LLM services risks marginalizing populations without adequate access to technology or digital literacy. Ensuring that AI decisions are explainable and accountable remains a significant challenge. The study emphasizes the importance of regulatory frameworks to monitor and address these risks.

This research underscores the need for robust policies and governance models to support LLM integration. The governments may need to create guidelines to address ethical concerns, such as censorship and bias, while fostering innovation. Transparent standards for AI in public administration are critical to maintaining public trust. Effective integration requires standardized protocols across departments and agencies. Open-source LLMs offer flexibility but necessitate collaboration to achieve consistent implementation. The interplay between state and professional logics demonstrates the importance of balancing regulatory compliance with technical excellence. The decoupling and integration strategies identified here offer valuable insights into navigating these tensions.

Despite providing valuable insights into bureaucrat-expert collaboration on LLM-based public services in China, several limitations are noteworthy. First, the research is focused on a single case study within a specific local government setting in China, limiting the extent to which the findings can be extrapolated to other contexts, particularly in democracies or countries with different institutional logics. The unique political and cultural environment of China, including its emphasis on centralized authority and ideological alignment, may not reflect the dynamics in more decentralized or pluralistic governance systems. Second, the study relies heavily on qualitative data from interviews, which, while rich in detail, may be subject to biases from participants' perceptions and interpretations. This approach also limits the ability to draw causal inferences or establish robust patterns across broader samples. Third, the exploratory nature of the study, while appropriate for uncovering initial insights, does not allow for a comprehensive analysis of all potential challenges or solutions in adopting LLMs in the public sector. Finally, the rapid pace of technological development in AI means that some findings may become outdated as new capabilities and applications of LLMs emerge. These limitations suggest the need for further research across diverse governance contexts and with mixed-method approaches to strengthen the external validity and applicability of the findings.

The integration of LLMs into public services requires more research endeavors. A key direction involves exploring the long-term impacts of LLM adoption on public service delivery, particularly in transformative applications like real-time citizen engagement and policy drafting. Additionally, unresolved ethical issues demand sustained attention, especially concerns about inclusivity, fairness, and the implications of censorship modules in AI-driven services. Addressing these issues is crucial to ensuring that LLM implementations remain equitable and socially responsible. In addition, it is also imperative to examine the the generalization of our findings; future studies should assess how insights from this case study can inform LLM applications in diverse governance contexts, including those of democracies with differing institutional logics. Furthermore, innovative applications of LLMs, such as emotional analysis, multi-modal data integration, and enhanced citizen feedback mechanisms, present promising avenues for enhancing public sector responsiveness and efficiency. While challenges persist, the integration of LLMs offers transformative opportunities. Collaborative strategies that reconcile competing institutional logics, address ethical considerations, and prioritize scalable solutions are essential for unlocking the full potential of AI in governance.

**Reference:**

Abbas, N., Følstad, A., & Bjørkli, C. A. (2022, November). *Chatbots as part of digital government service provision–a user perspective*. In *International Workshop on Chatbot Research and Design* (pp. 66–82). Cham: Springer International Publishing.

Abdi, J., Al-Hindawi, A., Ng, T., & Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open, 8*(2), e018815.

Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific Reports, 12*(1), 1953.

Ahn, M. J., & Chen, Y. C. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly, 39*(2), 101664.

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly, 36*(2), 358–367.

Anthopoulos, L., Reddick, C. G., Giannakidou, I., & Mavridis, N. (2016). Why e‐government projects fail? An analysis of the Healthcare.gov website. *Government Information Quarterly, 33*(1), 161–173.
Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology and Community Health, 60*(10), 854.
Baxter, D., Dacre, N., Dong, H., & Ceylan, S. (2023). Institutional challenges in agile adoption: Evidence from a public sector IT project. *Government Information Quarterly, 40*(4), 101858.
Boxenbaum, E., & Jonsson, S. (2017). Isomorphism, diffusion and decoupling: Concept evolution and theoretical challenges. In *The Sage handbook of organizational institutionalism* (2nd ed., pp. 77–101). Sage.
Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? *Policy Quarterly, 13*(4).
Boyne, G. A., & Chen, A. A. (2007). Performance targets and public service improvement. *Journal of Public Administration Research and Theory, 17*(3), 455–477.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., … & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.
Bryson, J. M., Crosby, B. C., & Stone, M. M. (2006). The design and implementation of cross‐sector collaborations: Propositions from the literature. *Public Administration Review, 66*, 44–55.
Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1). https://doi.org/10.1177/2053951715622512
Charles, V., Rana, N. P., & Carter, L. (2022). Artificial intelligence for data-driven decision-making and governance in public affairs. *Government Information Quarterly, 39*(4), 101742.
Chen, J., Pan, J., & Xu, Y. (2016). Sources of authoritarian responsiveness: A field experiment in China. *American Journal of Political Science, 60*(2), 383–400.
Chen, Y. C., & Lee, J. (2021). Collaborative data networks for public service: Governance, management, and performance. In *Digital Government and Public Management* (pp. 40–58). Routledge.
Cornish, F., Breton, N., Moreno-Tabarez, U., Delgado, J., Rua, M., de-Graft Aikins, A., & Hodgetts, D. (2023). Participatory action research. *Nature Reviews Methods Primers, 3*(1), 34.
Curtis, T. (2008). Finding that grit makes a pearl: A critical re‐reading of research into social enterprise. *International Journal of Entrepreneurial Behavior & Research, 14*(5), 276–290.
Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly*, *25*(1), 63-82.
De Vries, H., Tummers, L., & Bekkers, V. (2018). The diffusion and adoption of public sector innovations: A meta-synthesis of the literature. *Perspectives on Public Management and Governance, 1*(3), 159–176.
Dhungel, A. K., Wessel, D., Zoubir, M., & Heine, M. (2021). Too bureaucratic to flexibly learn about AI? The human-centered development of a MOOC on artificial intelligence in and for public administration. In *Proceedings of Mensch und Computer 2021* (pp. 563–567).
Di Domenico, M., Haugh, H., & Tracey, P. (2010). Social bricolage: Theorizing social value creation in social enterprises. *Entrepreneurship Theory and Practice, 34*(4), 681–703.

Dong, X. L., Moon, S., Xu, Y. E., Malik, K., & Yu, Z. (2023, August). Towards next-generation intelligent assistants leveraging LLM techniques. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 5792–5793).

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., … & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994.

Emerson, K., Nabatchi, T., & Balogh, S. (2012). An integrative framework for collaborative governance. *Journal of Public Administration Research and Theory, 22*(1), 1–29.

Friedland, R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In *The new institutionalism in organizational analysis* (pp. 232–263).

Gauld, R. (2007). Public sector information system project failures: Lessons from a New Zealand hospital organization. *Government Information Quarterly, 24*(1), 102–114.

Gray, B. (1985). Conditions facilitating interorganizational collaboration. *Human Relations, 38*(10), 911–936.

Gray, B., & Purdy, J. (2018). *Collaborating for our future: Multistakeholder partnerships for solving complex problems*. Oxford University Press.

Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional complexity and organizational responses. *Academy of Management Annals, 5*(1), 317–371.

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance, 5*(3), 232–242.

Groth, O. L. A. F., & Straube, T. (2021). Analysis of current global AI developments with a focus on Europe. Konrad-Adenauer-Stiftung.

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

Hine, E., & Floridi, L. (2024). Artificial intelligence with American values and Chinese characteristics: A comparative analysis of American and Chinese governmental AI policies. *AI & Society, 39*(1), 257–278.

Hwang, H., & Powell, W. W. (2009). The rationalization of charity: The influences of professionalism in the nonprofit sector. *Administrative Science Quarterly, 54*(2), 268–298.

Kinder, T., Stenvall, J., Six, F., & Memon, A. (2021). Relational leadership in collaborative governance ecosystems. *Public Management Review, 23*(11), 1612–1639.

Klijn, E. H., & Teisman, G. R. (2003). Institutional and strategic barriers to public–private partnership: An analysis of Dutch cases. *Public Money & Management, 23*(3), 137–146.

Kumar, V., Srivastava, P., Dwivedi, A., Budhiraja, I., Ghosh, D., Goyal, V., & Arora, R. (2023, December). Large‑language‑models (LLM)‑based AI chatbots: Architecture, in‑depth analysis and their performance evaluation. In *International Conference on Recent Trends in Image Processing and Pattern Recognition* (pp. 237–249). Cham: Springer Nature Switzerland.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision‑making in democratic settings. *Telecommunications Policy, 44*(6), 101976.

Kwak, G. H., Ling, L., & Hui, P. (2021). Deep reinforcement learning approaches for global public health strategies for COVID-19 pandemic. *PloS One, 16*(5), e0251550.

Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., … & Ahlemann, F. (2017). Digitalization: Opportunity and challenge for the business and information systems engineering community. *Business & Information Systems Engineering, 59*, 301–308.
Lounsbury, M. (2007). A tale of two cities: Competing logics and practice variation in the professionalizing of mutual funds. *Academy of Management Journal, 50*(2), 289–307.
Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., … & Love, B. C. (2024). Large language models surpass human experts in predicting neuroscience results. *Nature Human Behaviour, 1–11*.
Ma, M., Kang, Y., & Feng, Y. (2022). Can cross‐sector support help social enterprises in legitimacy building? The mixed effects in Hong Kong. *Journal of Public and Nonprofit Affairs, 8*(3), 375–398.
Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly, 40*(1), 101774.
Mani, S. K., Zhou, Y., Hsieh, K., Segarra, S., Eberl, T., Azulai, E., … & Kandula, S. (2023, November). Enhancing network management using code generated by large language models. In *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks* (pp. 196–204).
Meyer, D. (2024). The cost of training AI could soon become too much to bear. *Fortune.* https://fortune.com/2024/04/04/ai-training-costs-how-much-is-too-much-openai-gpt-anthropic-microsoft/
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340–363.
Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019). Artificial Intelligence in the public sector: A study of challenges and opportunities for Norwegian municipalities. In *Digital Transformation for a Sustainable Society in the 21st Century: 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway, September 18–20, 2019, Proceedings 18* (pp. 267–277). Springer International Publishing.
Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., … & Niehaves, B. (2023). Examining how AI capabilities can foster organizational performance in public organizations. *Government Information Quarterly, 40*(2), 101797.
Misuraca, G., van Noordt, C., & Boukli, A. (2020, September). The use of AI in public services: Results from a preliminary mapping across the EU. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 90–99).
Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross‐sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 20170357.
Medaglia, R., & Tangi, L. (2022, October). The adoption of Artificial Intelligence in the public sector in Europe: Drivers, features, and impacts. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance* (pp. 10–18).
Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial intelligence in government: Taking stock and moving forward. *Social Science Computer Review, 41*(1), 123–140.
Meijer, A. (2015). E-governance innovation: Barriers and strategies. *Government Information Quarterly, 32*(2), 198–206.
Nasseef, O. A., Baabdullah, A. M., Alalwan, A. A., Lal, B., & Dwivedi, Y. K. (2022). Artificial intelligence-based public healthcare systems: G2G knowledge-based exchange to enhance the decision-making process. *Government Information Quarterly, 39*(4), 101618.

Nay, J. J., Karamardian, D., Lawsky, S. B., Tao, W., Bhat, M., Jain, R., … & Kasai, J. (2024). Large language models as tax attorneys: A case study in legal capabilities emergence. *Philosophical Transactions of the Royal Society A, 382*(2270), 20230159.
Nutt, P. C. (2006). Comparing public and private sector decision-making practices. *Journal of Public Administration Research and Theory, 16*(2), 289–318.
Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review, 16*(1), 145–179.
OpenAI. (2023). *GPT-4 Technical Report*. https://cdn.openai.com/papers/gpt-4.pdf
Pache, A. C., & Santos, F. (2013). Inside the hybrid organization: Selective coupling as a response to competing institutional logics. *Academy of Management Journal, 56*(4), 972–1001.
Pan, I., Nolan, L. B., Brown, R. R., Khan, R., van der Boor, P., Harris, D. G., & Ghani, R. (2017). Machine learning for social services: A study of prenatal case management in Illinois. *American Journal of Public Health, 107*(6), 938–944.
Peris, C., Dupuy, C., Majmudar, J., Parikh, R., Smaili, S., Zemel, R., & Gupta, R. (2023, February). Privacy in the time of language models. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 1291–1292).
Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., … & Jennings, N. R. (2019). Machine behaviour. *Nature, 568*(7753), 477–486.
Ranerup, A., & Henriksen, H. Z. (2022). Digital discretion: Unpacking human and technological agency in automated decision making in Sweden's social services. *Social Science Computer Review, 40*(2), 445–461.
Radanliev, P., & De Roure, D. (2023). Review of the state of the art in autonomous artificial intelligence. *AI and Ethics, 3*(2), 497–504.
Reay, T., & Hinings, C. R. (2009). Managing the rivalry of competing institutional logics. *Organization Studies, 30*(6), 629–652.
Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems, 31*(3), 313–338.
Sanina, A., Balashov, A., & Rubtcova, M. (2023). The socio-economic efficiency of digital government transformation. *International Journal of Public Administration, 46*(1), 85–96.
Schiff, D. S., Schiff, K. J., & Pierson, P. (2022). Assessing public value failure in government adoption of artificial intelligence. *Public Administration, 100*(3), 653–673.
Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly, 41*(1), 101885.
Shareef, M. A., Kumar, V., Dwivedi, Y. K., Kumar, U., Akram, M. S., & Raman, R. (2021). A new health care system enabled by machine intelligence: Elderly people's trust or losing self control. *Technological Forecasting and Social Change, 162*, 120334.
Smith, G., Maloney, W., & Stoker, G. (2004). Building social capital in city politics: Scope and limitations at the inter‐organisational level. *Political Studies, 52*(3), 508–530.
Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., … & John, A. (2022). A review of Earth artificial intelligence. *Computers & Geosciences, 159*, 105034.
The Anhui Government. (2023). 安徽省通用人工智能创新发展三年行动计划（2023—2025年）. https://kjt.ah.gov.cn/kjzx/tzgg/121837851.html
The Chinese State Council. (2024). 国务院关于进一步优化政务服务提升行政效能推动"高效办成一件事"的指导意见，国发〔2024〕3号.
https://www.gov.cn/zhengce/content/202401/content_6926255.htm

The U.S. General Services Administration. (2023). *Security Policy for Generative Artificial Intelligence (AI) Large Language Models (LLMs)*. https://www.gsa.gov/directives-library/security-policy-for-generative-artificial-intelligence-ai-large-language-models-llms
Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure and process*. Oxford University Press.
Treviso, M., Lee, J. U., Ji, T., Aken, B. V., Cao, Q., Ciosici, M. R., … & Schwartz, R. (2023). Efficient methods for natural language processing: A survey. *Transactions of the Association for Computational Linguistics, 11*, 826–860.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.
Vogel, R., Göbel, M., Grewe‐Salfeld, M., Herbert, B., Matsuo, Y., & Weber, C. (2022). Cross‐sector partnerships: Mapping the field and advancing an institutional approach. *International Journal of Management Reviews, 24*(3), 394–414.
Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review, 80*(6), 946–961.
Wang, Y., Zhang, N., & Zhao, X. (2022). Understanding the determinants in the different government AI adoption stages: Evidence of local government chatbots in China. *Social Science Computer Review, 40*(2), 534–554.
Waardenburg, M., Groenleer, M., de Jong, J., & Keijser, B. (2020). Paradoxes of collaborative governance: Investigating the real-life dynamics of multi-agency collaborations using a quasi-experimental action-research approach. *Public Management Review, 22*(3), 386–407.
Westermann, H., Savelka, J., & Benyekhlef, K. (2023). LLMediator: GPT-4 assisted online dispute resolution. *arXiv preprint arXiv:2307.16732*.
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.
Williams, K., Berman, G., & Michalska, S. (2023). Investigating hybridity in artificial intelligence research. *Big Data & Society, 10*(2), 20539517231180577.
Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. *Big Data & Society, 10*(2), 20539517231184892.
Yan, Y. (1996). The culture of guanxi in a North China village. *The China Journal, (35)*, 1–25.
Zhang, X., Khan, F., Wang, X., & Tang, C. (2023). Exploring the impact of national culture on the development of open government data: A cross-cultural analysis. *Big Data & Society, 10*(2), 20539517231206809.
Zheng, Y., Zhao, K., & Stylianou, A. (2020). An examination of the use of chatbots in government services: The case of 311. *Government Information Quarterly, 37*(3), 101487.
Zhu, Z., Zhao, R., & Tao, C. (2021). Chinese NPOs in service contracting at the community level: Challenges and strategies. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 32*(4), 780–794.