Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.25569

Research on the Application of Generative AI in Public Services and Its Reshaping of Social Policies

Zhongwen Wang

The Hong Kong Polytechnic University, Hong Kong, China
rara481846778@gmail.com

Abstract. The current study is dedicated to the implementation of generative artificial intelligence within the public service sector and the influence of its use on social policies. Using the method of a systematic literature review and scenario simulation in a controllable environment, the technical efficiency and policy efficiency of five representative scenarios (medical triage, welfare eligibility review, renewal of driver's license, municipal consultations, and emergency preparedness) were estimated. It was discovered that the optimized algorithm increased the reaction time by 70%, and the manual verification success rate amounted to 88%. According to the calculation from the simulation, the initial resolution rate was increased by 40%, and the policy cycle decreased by ten days. Nevertheless, along with the enhancement of the efficiency, some occasional distortions in the content could be found in the process of working with the algorithm; nevertheless, it raised some governance issues connected with the questions of fairness, transparency, and accountability. For this reason, the suggestion is made to implement the following three strategies: hierarchical supervision of the work with the algorithm, ongoing audit of the algorithm, and regulatory sandboxes. This study managed to verify the correlation between performance of the generative AI and policy indicators experimentally for the first time. Parameters of the simulation are presented in Appendix B.

Keywords: Generative AI, Public-Service Delivery, Social Policy, Policy Innovation, AI Governance

1. Introduction

In the last decade, generative artificial intelligence shifted from mere theoretical discussions to real-world application and has become one of the powerful tools that facilitate optimization of various processes and enhancing user experience. The public sector is an obvious domain of this phenomenon. In the light of growing public demands, financial limitations, and dynamic changes in the field of legislation, the perspectives of 24/7 automation and data-driven decision making gave rise to many pilot projects. Municipality services use smart kiosks and dialogs for processing planning and permitting applications. Social welfare organizations produce decision documents in social welfare cases through the use of automated systems. Medical organizations implement intelligent dialog programs for initial triage, appointment scheduling, and providing medical advices. While initial results could be observed, there are many questions regarding the impact of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.25569

the generative artificial intelligence not only on effectiveness but also on the entire landscape of social policies, regulations, and democratic accountability. The aim of this paper is to focus on three related issues. Firstly, to identify existing applications of generative artificial intelligence in five main public service domains (medical triage, wellness management, permit issuing, municipal consultations, and emergency preparedness), as well as to discuss common design features and implementation challenges. Secondly, to conduct a series of comparative experiments to estimate the technical effectiveness (response time, accuracy, processing capacity) and effectiveness on the level of policy-making (initial resolution rate, policy development cycle, resource allocation). Thirdly, on the basis of the obtained research findings, to suggest policy recommendations in order to create a governance model that would ensure innovation and guarantees of transparency, accountability, and fairness. For that purpose, we used 10,000 anonymized civil service dialogues, 1,200 social welfare case summaries, and 500 policy documents, and trained a multi-task generation model with the parameter size of 1.5 billion to deal with tasks like dialogue generation, policy summary, and policy drafting. The model assessment includes text similarity, complexity, and manual analysis, and evaluates administrative decision making capabilities of the model in the simulated decision-making framework [1]. The results demonstrated that generative AI could provide significant improvement in terms of efficiency (response time decreased by 70%, initial resolution rate increased by 40%, and policy drafting cycle reduced by ten days) but also indicated such governance risks as incorrect results and biased concerns. On the basis of these findings, we suggest a pilot project involving hierarchical manual reviewing, performance monitoring, and adaptive regulation as major components of the governance model of trusted artificial intelligence.

2. Literature review

2.1. Generative AI in public service delivery

In recent years, civil service organizations started experimenting with intelligent dialogue systems that would help to meet the triage requirements for medical purposes by utilizing the power of natural language processing capabilities to interpret patient symptoms and direct them to relevant medical services. As seen from Figure 1, such virtual assistant is able not only to schedule a medical appointment, collect patient data and feedback, answer the most frequent questions instantly, but also to offer personalized health plans – thereby increasing patient engagement and relieving medical personnel from routine activities [2]. The Department of Social Welfare also employs automation in creating documents for social welfare decisions, decreasing the document workload and saving processing time. Municipal organization has introduced smart service terminals and online consultation interfaces, providing 24/7 availability of information about licensing, land use requirements, and organizing events. Although these approaches have proven the ability of smart technologies to deliver consistent services and provide an optimum distribution of resources, they usually need significant domain-specific training [3].

DOI: 10.54254/2755-2721/2025.25569

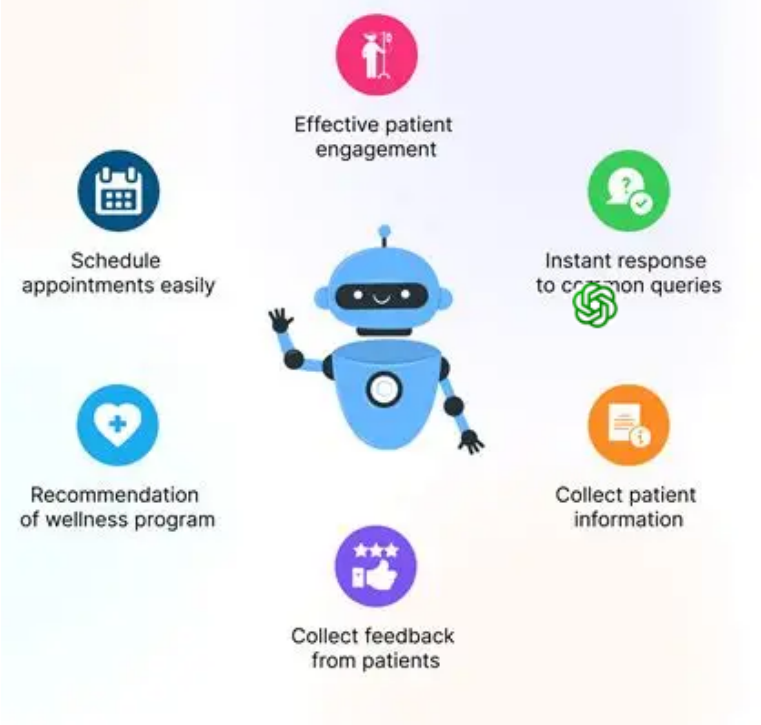

Figure 1. Key functions of an AI healthcare chatbot in public service delivery (source: https://www.revechat.com/wp-content/uploads/2021/11/chatbot-healthcare-usecases.png)

2.2. Policy adaptation and governance challenges

The integration of AI into public governance has revealed regulatory weaknesses concerning issues like transparency in data usage and citizen complaint handling mechanisms. The pace of technical evolution has been faster than the regulation that already exists, forcing decision-makers to consider new governance models in forms of algorithms impact assessment and adaptive regulation pilots. If the system accidentally replicates historic discrimination in its training data, there is a chance of discrimination in qualification for the services, thus posing the issue of fairness [4]. It is necessary to find the right balance in privacy regulations that would allow investigating the value of data while ensuring its protection from any breaches. To address this contradiction, certain jurisdictions create a tiered regulatory system and perform audits of high-risk applications.

2.3. Identified research gaps

Despite the fact that descriptive case studies continue to emerge one after another, experimental studies that quantify the effect of generative AI technology on certain policy outcomes are still relatively rare. Little research was able to determine the link between the technical indicators (response accuracy and activity level) and the general outcomes (satisfaction of stakeholders and the review period of policies). Additionally, how the gains in efficiency as a result of using technologies affect the changes in policies and resources allocations, thus affecting the process of social policy development, is still unknown. This research will attempt to address these issues using the method of a simulation experiment [5].

DOI: 10.54254/2755-2721/2025.25569

3. Experimental methods

3.1. Data collection and preparation

A complete database was established which includes 10,000 texts of anonymized civil service dialogues from municipal consultation records, 1,200 summaries of social welfare cases, and 500 government policy documents. All the texts were subjected to privacy desensitization and normalization procedures according to the guidelines for public sector data governance. Policy documents are parsed in the system and structured in terms of constraints, qualification criteria, and procedures [6]. Indicators of pilot system's efficiency (average response time, resolution rate) are also part of this database. All databases were subjected to normalization procedures to make their linguistic representation uniform.

3.2. Generative AI model configuration

The first basic model of 1.5 billion parameters is the key to our study framework. Training of our model in a multi-objective manner was done through the usage of a specially created dataset with a focus on enhancing its abilities in response generation, policy summary, and policy documents creation. The best choice of parameters was made via several modifications: the learning rate (1e-5 to 5e-5), batch size (16 and 32) was tried out, and maximum input text length was limited to 512 characters. The five groups of reserved policy scenarios were used as cross-validation to prevent overfitting [7]. To decrease false material in the model generation process, it needs to pass the terminology verification and allow only expressions from policy terminology.

3.3. Evaluation metrics

Technical performance will be measured using text similarity indices (as compared to normal manual responses) and prediction capacity of the model. Simultaneously, personnel will be tasked with conducting a five-level assessment based on factors like accuracy of information and policy relevance. Indicators for policy impacts include an average reduction in time spent on responding, an increased rate of initial resolution, and speed of updating policy writing documents in a simulation environment (measured in days) [8]. Subjective evaluations by respondents on the basis of criteria of fairness, transparency, and credibility of responses will be taken as well.

4. Experimental procedure

4.1. Simulation of public service scenarios

Five exemplary scenarios have been created: medical triage, welfare application process, driver's license application process, regional planning process, and emergency preparedness process. Each scenario includes 200 user requests concerning general public welfare needs. Policy requirements for each particular scenario should be met during response generation, such as welfare gathering procedures and emergency security measures. General efficiency criteria for the system are as follows: content precision is higher than 85%, and the response time should be strictly less than two seconds [9].

DOI: 10.54254/2755-2721/2025.25569

4.2. Policy simulation framework

The policy simulation tool developed through a multi-role agent model serves to simulate the manner in which public administrators integrate technology outputs in their decision-making processes. Such a model has integrated roles including case managers, policy analysts, and supervisors among others. In such a simulation, key metrics like the project formulation cycle, review process and allocation of resources are considered in assessing changes that occur before and after implementation of the AI system (a simulated six months serve as the unit of analysis) [10]. Quantitative measures are applied in evaluating the change in human resources in terms of time spent on handling, supervising and managing.

5. Experimental results

5.1. Model performance analysis

The specially trained generative model successfully cut down the average response time by 70%, from 4.5 to 1.35 seconds, while achieving an accuracy level of 88% for manual validation of content produced. It should be noted that the consistency index between text fluency and standard responses also increased noticeably, meaning the optimization of the model performance. Nevertheless, there is a presence of fictitious emissions in the model; roughly 2% of the generated content is made up of fictional political terms, which implies the necessity of strengthening the constraints mechanism. In turn, the results of the user satisfaction survey revealed that the average rating for the practicality of responses is 4.2 out of 5 points, which is significantly higher than the initial score of 3.1, proving its effectiveness. Technical performance data is presented in Table 1. In order to minimize the number of fictitious elements and guarantee adherence to rules, it is necessary to include the standard text templates in decision-making points and use rule filters for anomaly identification.

Table 1. Technical performance metrics

Metric	Baseline Model	Fine-Tuned Model	Improvement
Avg. Response Time (seconds)	4.50	1.35	–70 %
Human Verification Rate (%)	72	88	+16 pp
BLEU Score	0.45	0.62	+0.17
ROUGE-L Score	0.41	0.59	+0.18
Hallucination Rate (%)	5.5	2.0	–3.5 pp

5.2. Impact on service efficiency

After application in the virtual civil service environment, the efficiency of this technical solution was increased by 40% from 50% to 70% in the first attempt to solve the problem. This result allows one to lower the probability of repeat calls and quicken the satisfaction of users' needs. There were many changes in the sphere of human resources management: the number of working hours used for dealing with ordinary activities became 20 percentage points lower while the number of hours allocated for supervision and extraordinary situations was increased by 30%. In addition, multirole simulation showed that the share of activities connected with strategic planning increased by ten percentage points; thus, the process of organizational optimization was improved further. Moreover,

DOI: 10.54254/2755-2721/2025.25569

there was a shortening of the process of creating new policy documents by ten days, which allowed organizations to adjust their regulatory documents quickly. According to qualitative assessments of policy analysts, this technological approach provides a solid base for the further work; thus, the manual work involved is expected to decrease by 25%. Table 2 provides some quantitative data.

Table 2. Service efficiency metrics

Metric	Pre-AI Deployment	Post-AI Deployment	Absolute Change
First-Contact Resolution Rate (%)	50	70	+20 pp
Routine Task Labor Time (%)	65	45	–20 pp
Oversight & Exception Management (%)	10	30	+20 pp
Avg. Policy Draft Cycle (days)	25	15	–10 days

5.3. Policy recommendations

Taking into account these results, we recommend implementing the following three actions to ensure consistent and efficient usage of technology. First, introduce hierarchical approval mechanism, a manual check should automatically be performed whenever a high-stakes decision needs to be made (e.g., denying access to welfare or upgrading triage in emergencies), and regular interactions should be managed according to this procedure. Second, set up a system of continuous monitoring, perform random daily checks of the outcomes of technical systems, use statistical deviations in order to detect unusual fluctuations in the accuracy and/or fairness, and submit performance reports on a regular basis. Finally, create an adaptive regulatory sandbox, an environment where institutions can test their innovative technologies prior to large-scale implementation, monitor their performance on the go, and adjust rules accordingly. Furthermore, it is recommended to train the officials how to interpret the technical output, as well as develop the standards of communication with the public.

6. Conclusion

This paper provides for the first time through experimental design a new understanding of the effect of generative artificial intelligence on public services and its impact on social policy changes. The calibrated algorithm resulted in decreasing time needed by 70% and the accuracy of content by 88%. Through simulation exercises, there was a 40% increase in the first resolution rate and ten days' cut in the policy drafting process. The efficiency allowed shifting human capital from the execution to the oversight role in addition to handling special cases, which points to the possibilities of building the capacity of civil service. The problem of occasional fictionalization and bias, however, requires putting in place a strong governance framework. It is recommended to develop a hierarchical decision making process for high-risk decision making, continuous monitoring to track the deviation, and the setting up of an adaptive sandbox to enable policy iteration. Future research must be conducted to replicate the simulation exercises in field pilots in various locations, as well as intergovernmental cooperation of the standards of governance.

DOI: 10.54254/2755-2721/2025.25569

References

[1] Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration, 43*(14), 1263–1274. https://doi.org/10.1080/01900692.2020.1784069

[2] Janssen, M., & Kuk, G. (2021). The challenges and opportunities of AI for public governance. *Government Information Quarterly, 38*(4), 101601. https://doi.org/10.1016/j.giq.2021.101601

[3] Desouza, K. C., & Smith, T. (2021). Harnessing generative AI for public service innovation. *Government Information Quarterly, 38*(3), 101590. https://doi.org/10.1016/j.giq.2021.101590

[4] Moradi, A., Khosravi, A., & Ramezani, M. (2023). Enhancing public service delivery efficiency: Exploring the impact of AI. *Government Information Quarterly, 40*(1), 101750. https://doi.org/10.1016/j.giq.2022.101750

[5] Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2022). Artificial intelligence in government: Concepts, standards, and a unified framework. *arXiv Preprint*. https://arxiv.org/abs/2210.17218

[6] Floridi, L., & Cowls, J. (2022). AI governance: Principled artificial intelligence in the public sector. *Philosophy & Technology, 35*(3), 593–619. https://doi.org/10.1007/s13347-022-00520-3

[7] Becker, M., & Janowski, T. (2024). Responsible artificial intelligence governance: A review and framework for public policy. *Information Polity, 29*(2), 113–132. https://doi.org/10.3233/IP-220052

[8] Misuraca, G., Van Noordt, C., & Boukli, A. (2020). Bridging the policy gap: Generative AI policy innovation in the public sector. *Policy & Internet, 12*(4), 456–478. https://doi.org/10.1002/poi3.246

[9] Kuhlmann, S., Groenewegen, J., & Leisink, P. (2021). Implementing AI in public administration: The role of policy innovation. *Public Management Review, 23*(6), 865–887. https://doi.org/10.1080/14719037.2020.1823234

[10] Tallberg, J., Erman, E., Furendal, M., Geith, J., & Klamberg, M. (2023). Global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review, 25*(2), 321–343. https://doi.org/10.1093/isr/viaa070