Service Sphere: Citizen Service and Complaint Management System

Wazeer Ali[1], Sana Riaz, Ali Raza[1], Muhammad Rizwan[1], Jawad Hussain[2], Sabeen Fatima[1]
[1]Department of Computer Science Quaid-e-Awam University of Engineering, Sciences & Technology, Nawabshah, Sindh, Pakistan
[2]Department of Information Technology, Shaheed Benazir Bhutto University, Shaheed Benazir Abad, Sindh, Pakistan
*Correspondence: wazeeralipanhwar786@gmail.com, sanariaz272@gmail.com, aliafzal0314@gmail.com, khokharr438@gmail.com, jawadq049@gmail.com, sabeenfatima5858@gmail.com

Citation | Ali. W, Riaz. S, Raza. A, Rizwan. M, Hussain. J, Fatima. S, "Service Sphere: Citizen Service and Complaint Management System", IJIST, Vol. 7 Issue. 10 pp 33-43, November 2025
DOI | https://doi.org/10.33411/IJIST/1709
Received | October 29, 2025 **Revised |** November 20, 2025 **Accepted |** November 22, 2025
Published | November 26, 2025.

Globally, particularly in developing countries, registering and tracking citizen complaints remains a significant challenge due to complex government procedures, slow response times, and limited follow-up mechanisms. To address this issue, this study proposes an integrated citizen service and complaint management system that enables users to draft, submit, and monitor complaints through a unified platform. The system leverages fine-tuned large language models (LLMs), including GPT and LLaMA, along with a custom backend that manages routing, microservices, and business logic. The AI models automatically classify complaints and forward them to the appropriate government departments. The platform supports both English and Urdu languages, reducing manual processing and improving accessibility. Experimental results demonstrate that the system achieves 92% classification accuracy, reduces average response time by 40%, and automates 68% of complaint cases. These findings indicate that the proposed system can significantly enhance efficiency, transparency, and citizen satisfaction in public service complaint management.

Keywords: Artificial Intelligence; Automation; Complaint Management System; Human–AI Collaboration; Large Language Models; MERN Stack; Multilingual System; Natural Language Processing; Public Service Delivery

Introduction:

Public service complaint management remains a major challenge worldwide, particularly in developing countries where administrative procedures are often slow, complex, and inefficient [1][2][3][4]. Citizens frequently encounter difficulties in submitting complaints, tracking progress, and receiving timely responses. Many complaints are delayed or ignored due to manual processing, communication gaps, and limited transparency. These challenges highlight the need for intelligent automated systems that can improve efficiency, responsiveness, and accessibility in government service delivery.

In the 21st century, the world has become more technologically advanced, and artificial intelligence has played an important role in improving digital systems. Artificial intelligence introduced modern technologies like Natural Language Processing, automation, and an intelligent chat system, which helps organizations handle user requests more effectively. Large Language Models (LLMs), such as GPT and LLaMA, can understand human language, support multiple languages, and accurately identify user intent. These technologies provide strong support to build smart and automated complaint management platforms.

This study presents the Citizen Service Automation and Complaint Relief System, which allows people to register and track complaints in both English and Urdu through a single platform. The automated system sends the complaint to the relevant department and updates users in a timely manner. This reduces delay and confusion. Centralized dashboards help departments manage cases easily and also improve transparency.

Key Objectives of the Study:

To develop a simple web-based system for registering and tracking citizen complaints.
To help citizens write complaints easily using natural language.
To automatically classify complaints and send them to the correct government department.
To reduce delays and manual work in the complaint handling process.
To support both the English and Urdu languages for better accessibility.
To improve transparency and follow-up in government complaint services.

Literature Review:

During the past ten years, research has continually demonstrated that public administration is one of the areas that can benefit most from AI due to its potential to enhance efficiency, responsiveness, and citizen participation. The original overview of Mehr emphasizes the fact that AI can conduct citizen services and administrative workflows more effectively, pointing out chatbot applications and process automation in municipal and national services as examples [1]. Furthermore, Battina considers AI to be a game changer in the area of citizen services and makes the case that automation and data-driven decision-making can greatly enhance the delivery of public-sector services [5]. Comprehensive studies of AI in the public sector consistently find that the main applications are service delivery, administrative support, and decision-making, but large-scale, integrated systems are still quite uncommon [2].

Multiple studies present prototype systems for handling complaints and their demonstrations. Among these studies, Shedge et al. developed a prototype of an AI-based complaint management system that utilized Natural Language Processing and sentiment analysis. They found that the proposed system could do an excellent job in prioritizing and organizing complaints, but their method was based on lexicon or rule-based techniques and did not utilize modern Large Language Models (LLMs) or even basic evaluation metrics [6]. Another research also confirmed that such automatic sentiment detection can be useful for decision-makers to recognize and classify citizens' issues with respect to their importance and thus allocate resources accordingly, but the analysis usually lacks a strong understanding of the situation or testing of its expansion [7][8].

The landscape of conversational agents and automated text processing has been significantly transformed by recent advances in foundation models, particularly large language models (LLMs) such as GPT and LLaMA. Technical reports and publications about the models showcase the LLM's abilities in terms of understanding context, handling different languages, and adapting to new situations with very few examples—characteristics that surpass the drawbacks of the previous rule-based chatbots [7][9]. The use of these models allows a better understanding of users' intents, maintaining the context of the conversation even after different sessions, and giving more human-like interaction—all these features are very important for systems dealing with citizens since they often face issues such as ambiguity, use of local languages, and input in several languages at the same time.

In Carvalho's view of Government Augmented Intelligence, AI is a human-supportive and enhancement tool and not the other way around; the study accentuates the cooperative human-AI workflows for CitRM (Citizen Relationship Management) at the municipal level and the requirement of expert supervision in public contexts [10]. This combined viewpoint is also supported in other places: although automation is capable of performing simple and everyday cases, complex or sensitive complaints still need the input of humans and have their corresponding mechanisms of accountability [2][11].

Public trust is a prerequisite for the implementation of AI in the government sector. Aoki and his colleagues point out that public acceptance of AI-driven chatbots relies heavily on the factors mentioned above - transparency, understanding regarding the role of AI, and the nature of the information solicited [11]. Other research done in different geographical areas also refers to ethical and policy issues: Jadoon et al., for instance, see the need for ethical frameworks and governance policies in Pakistan to regulate AI usage in social governance [12]. The deployment of systems with human intervention options for monitoring, tracking, and auditing will then be the only option available for consumers who might use capable systems but would not trust them due to the lack of the mentioned features.

Numerous studies centered on Pakistan and the surrounding area investigate the role of ICT and AI in increasing the efficiency of the governance and justice system, showing both potential and structural limitations. Shafiq et al. propose that the application of AI/ICT can help eliminate the backlog of cases in Pakistan courts; however, they also refer to the emergence of institutional bottlenecks and the requirement for policy readiness [3]. Baig and Mughal scrutinize the compatibility of automation in local government governance, pointing out that the adoption of the new technology by the organization and the provision of regulatory support must go hand in hand with the technical feasibility [4]. Afzal & Khan; Jadoon et al. have recently underscored the importance of user-friendly attributes such as the perceived usefulness, the psychological profile of the adopters, and the ethical issues surrounding the use of AI in the public sector that will determine the uptake of the AI systems by the public [9][12].

Advanced sentiment analysis and evaluation techniques are important for sorting complaints and measuring satisfaction. [7] introduce the concepts of fine-grained sentiment analysis and fuzzy linguistic hedges, which will be of great help when attempting to assess citizens' feelings and the urgency of the matters in unstructured text. Similarly, [8] demonstrate that using sentiment as a guiding principle in management can lead to community governance benefits, and the same approach can be applied to public complaint triaging.

Material and Methods:

The Citizens Service Automation and Complaints Relief System was developed through a systematic multi-phase methodology, which aimed at public-sector settings, technical correctness, human-friendliness, and applicability. Progress in each phase contributed to refining system requirements, architectural components, AI capabilities,

process automation, and evaluation. Below are the seven phases of the methodology that have been elaborated.

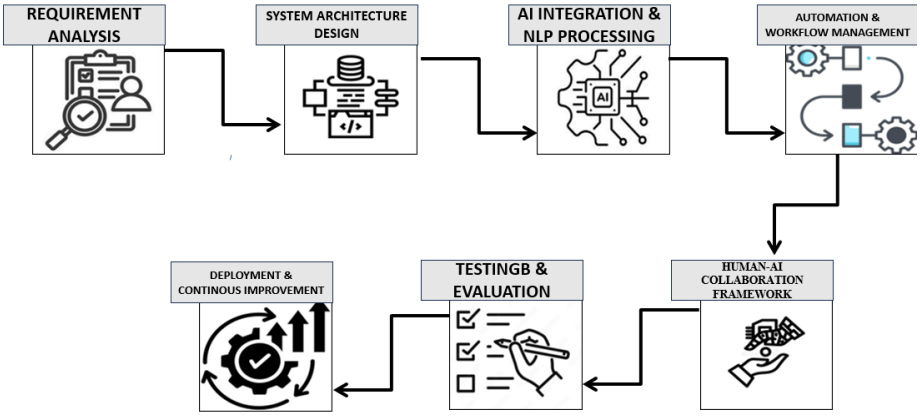


Figure 1. Phases of Methodology

Phases of Methodology:

Phase 1: Requirement Analysis:

An initial requirement analysis was conducted to identify the challenges citizens face when submitting complaints to government departments. Common complaint types and user experiences with existing procedures were analyzed. Three primary user roles were identified: citizens, officers, and administrators.

Initially, user requirements were analyzed, and the main features—such as complaint registration, tracking, notifications, and automatic forwarding—were finalized. We also evaluated different AI models and selected GPT/LLaMA because these models perform well for both English and Urdu.

Phase 2: System Architecture Design:

After defining the requirements, the system architecture was designed for the structure of our system. We chose the MERN stack because of its single-page application architecture and its ability to work with different databases.

We fine-tuned existing AI models such as GPT/LLaMA to reduce development time and improve system accuracy. We set up a custom backend to communicate with the AI models using RESTful APIs.

Phase 3: AI Integration and NLP Processing:

Natural Language Processing tasks were performed using fine-tuned Large Language Models (LLMs), including GPT and LLaMA. These models were integrated into the system via RESTful APIs to analyze complaint text, extract key information, and determine the appropriate department. The NLP pipeline performs language detection, text normalization, sentiment analysis, and intent classification.

Phase 4: Automation and Workflow Management:

To automate the core operational workflows, a dedicated AI Agent module was specially designed. The module does the automatic logging of complaints in the database, classification, and forwards them to the right department according to the established rules. Moreover, it determines the time limits for solutions, sends out reminders for follow-ups, and increases the priority of the cases that are still unresolved after the expected time. The integration of notification services (either emails or SMS) keeps the public aware throughout the entire process of dealing with complaints, which means that there is transparency and people are involved.

Phase 5: Human–AI Collaboration Framework:

To ensure oversight and accountability, a Human-AI collaboration framework was designed to be part of the platform. The supervisor dashboard gives department officials the

power to look at, control, and bring to a conclusion complaints. The system enables seamless transfer from AI processing to human review whenever manual verification is required or when AI confidence levels are insufficient. Feedback from the officers and the public was built in as a continuous process of improving the model's performance and the system's trustworthiness.

Phase 6: Testing and Evaluation:

To guarantee that the system is stable and its performance is optimal, multiple testing cycles were carried out. These consisted of unit testing of separate modules, integration testing for communication and data flow across layers, and user acceptance testing (UAT) with actual users. The system was assessed with the help of key performance indicators, such as response accuracy (how well the AI understood the user's intent), resolution efficiency (average time taken to resolve complaints), user satisfaction scores, and automation success rate (percentage of cases managed without human intervention). In addition, security and privacy testing were carried out to check compliance with digital governance standards.

Phase 7: Deployment and Continuous Improvement:

The cloud-based server became the location for the system prototype deployment for the purpose of real-time testing and monitoring. Constantly, user interactions, logs, and officer feedback were being evaluated to detect performance issues and to spot areas that can be improved. Over time, prompts, workflows, and routing logic were adjusted in order to gain accuracy and cut down the operational delays. The method also takes into account the possibility of future growth, for example, the connecting of more sophisticated localized LLMs and the enhancement of domain-specific AI capabilities.

Results and Discussion:

Service Sphere is implemented using a modular architecture combining the MERN stack with an AI-driven NLP layer powered by pre-trained Large Language Models (LLMs). The entire system was developed as a functional prototype capable of autonomous complaint handling, multilingual text interpretation, workflow automation, and human—AI collaboration. The following subsections discuss the implementation of each component.

System Architecture:

Our System follows a layered architecture consisting of four main parts:
Frontend Interaction Layer (React.js)
Backend Service Layer (Node.js + Express)
AI Integration & Workflow Automation Layer
Database Layer

Frontend Implementation:

The frontend was developed using React.js for a better user experience. Overall, this system is designed around these three portals:

Citizen Portal:

Complaint submission form
AI-powered chatbot
Complaint status tracking
Users can select either English or Urdu

Officer Dashboard:

Assigned complaint list
Status update panel
Support notes management and verification mechanisms
Escalation handling

Admin Dashboard:

Department-level statistics
Case handling performance metrics for officers

Automation success rate
Complaint analytics

Backend Implementation:

The Implementation of the backend is using Node.js and Express, which create Robust API endpoints and Business logic.

Controller Layer:

The controller layer manages all requests coming from the frontend and links them to backend features, and it also ensures smooth communication between the user interface and system logic. It includes routes for:

Authentication routes
Complaint routes
Officer and admin routes
AI service controllers

Service Layer:

Manages the main system logic for:
Complaint classification
Department routing
User role validation
Workflow automation
AI communication

Middleware Layer:

Includes:
Authentication
Request validation
Logging and auditing

AI Integration and NLP Pipeline:

The core component of the system is an AI-based pipeline that processes citizen text and converts it into structured complaint data.

LLM Communication:

The model was guided using custom prompts to extract:
Complaint category
Relevant government department
Sentiment polarity
Urgency level
Structured summary

Multilingual Processing:

The system first detects the languages of the input using a fine-tuned model, then applies normalization rules to handle:
mixed-language sentences,
informal expressions,
spelling variations.
This ensures more accurate classification.

Context Management:

A separate component stores conversation records in the database, enabling the AI to continue interactions smoothly when users return, and also helps maintain context and improve overall clarity.

Workflow Automation Module:

The workflow module manages complaint automation and controls key management tasks, including:
Logging new complaints

Directing complaints to the responsible department
Monitoring deadlines
Forwarding unresolved complaints for further action

The system follows an event-based approach to manage complaint assignment and updates. During every interaction by a user or officer, the backend validates the complaint status and records all updates in the database, which enables users to see up-to-date information through their dashboards. If complaints remain unanswered for more than five days, then automatically identified and escalated. This approach ensures consistent progress without the need for manual monitoring.

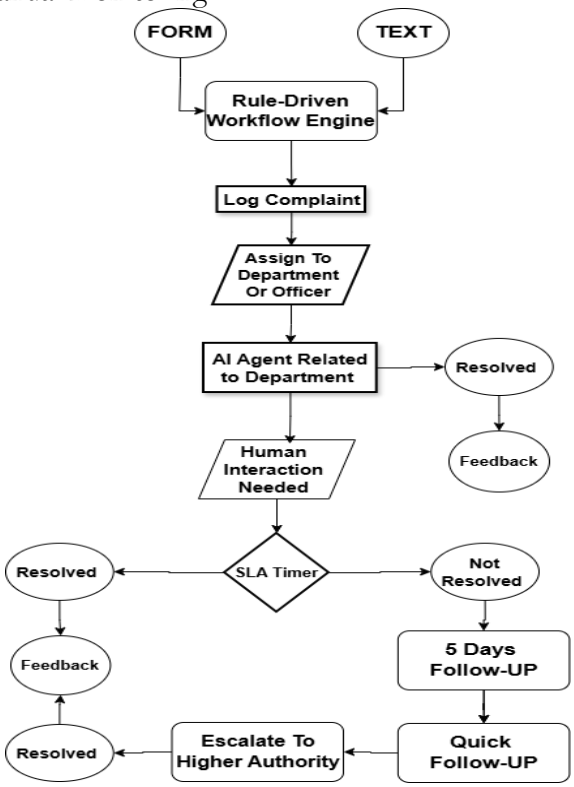


Figure 2. Workflow Automation Module:

Explanation:

In Figure 2 illustrates the workflow automation module of the proposed system. Complaints can be submitted either through structured forms or free-text input, after which the rule-driven workflow engine logs the complaint and assigns it to the appropriate department or officer. An AI agent associated with the department automatically processes routine cases. If the complaint requires human intervention, it is forwarded to the responsible officer for manual handling. The system continuously monitors resolution progress using an SLA timer. If the issue is resolved, feedback is collected from the citizen. Otherwise, unresolved cases trigger follow-up actions, including automated reminders or escalation to higher authorities after a defined period (e.g., five days). This workflow ensures timely resolution, accountability, and minimal manual oversight.

Results and Performance Evaluation:

To evaluate the performance of the proposed citizen service automation and complaint management system, experiments were conducted using clearly defined evaluation metrics and a controlled dataset. The evaluation focused on the effectiveness of the LLM-based classification model, system response behavior, and workflow efficiency. The following subsections present details of the dataset, ground-truth preparation, evaluation metrics, and performance outcomes.

Evaluation Dataset and Ground Truth:

A test dataset consisting of 200 citizen complaints was created for evaluation. The complaints were written in both English and Urdu and covered common public service issues such as identity services, municipal problems, tax queries, and utility-related complaints. Each complaint was manually labeled with its category and the relevant government department. Annotations were performed by two independent evaluators, and disagreements were resolved through consensus. This labeled dataset served as the ground truth for evaluating the outputs produced by the LLM.

Classification Accuracy:

Classification accuracy is used to evaluate how well the large language models assign complaints to the appropriate departments. A complaint is considered correctly classified when the model's output matches the ground-truth category.

Calculation Method:

$$\text{Classification Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Complaints}} \times 100$$

Understanding Results:

Out of 200 evaluated complaints, 184 were correctly classified, resulting in an overall accuracy of 92%. This demonstrates that the proposed system can accurately identify complaint categories and perform comparably to manual classification methods.

Confusion Matrix:

Total complaints evaluated: 200
Complaint categories:
Municipal Services
Utility Services
Identity Services
Tax & Revenue
Correctly classified complaints: 184
Misclassified complaints: 16

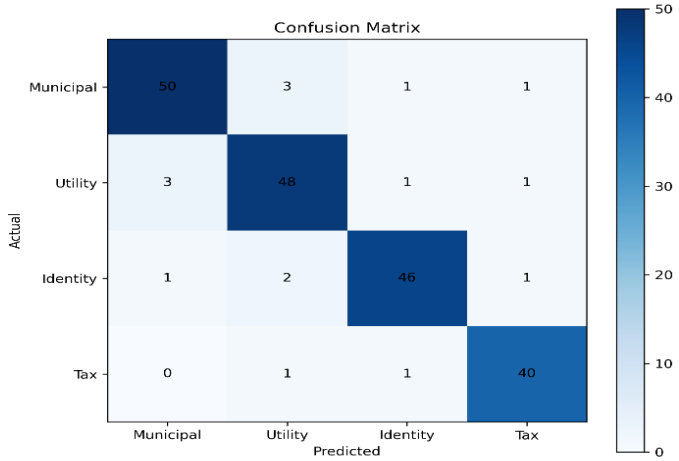


Figure 3. Confusion Matrix for workflow classification.

Explanation:

Figure 3 presents the confusion matrix of the LLM-based complaint classification model used in the Service Sphere system. The rows represent the actual complaint categories, while the columns indicate the categories predicted by the model. Most values are concentrated along the main diagonal, demonstrating that the the large language models correctly classified the majority of complaints, resulting in an overall accuracy of 92%.

Minor misclassifications occur primarily between semantically similar categories, particularly municipal and utility services, indicating overlapping characteristics in real-world complaint descriptions. Identity Services and Tax & Revenue categories exhibit higher

precision, suggesting clearer linguistic patterns that are easier for the model to distinguish. Overall, the confusion matrix confirms that the LLM provides reliable classification performance across diverse complaint categories.

Precision, Recall, F1-Score (Derived from Figure 3):

Table 1. Classification Performance Metrics

Complaint Category	Precision (%)	Recall (%)	F1-Score (%)
Municipal Services	92.6	90.9	91.7
Utility Services	88.9	90.6	89.7
Identity Services	93.9	92.0	92.9
Tax & Revenue	93.0	95.2	94.1
Macro Average	**92.1**	**92.2**	**92.1**

Explanation:

Table 1 summarizes the classification performance of the LLM-based model across different complaint categories using precision, recall, and F1-score metrics derived from the confusion matrix in Figure 3. The results indicate consistently high performance, with macro-average values of approximately 92% across all evaluation metrics.

Identity Services and Tax & Revenue categories achieved the highest scores, suggesting that complaints in these domains contain more distinct linguistic features that the model can identify effectively. Slightly lower performance for Municipal and Utility Services reflects occasional confusion between closely related service areas. Overall, the results demonstrate that the LLM provides accurate and balanced classification performance suitable for real-world complaint management systems.

Average Response Time Reduction:

Average response time refers to the time between complaint registration and the system's initial response:

Complaint Registration:

Response time was measured from when the complaint was submitted until the system gave its first reply or placed it into a category, and then we compared this time with manual processing, where complaints were checked by human officers in a test setting.

Calculation Method:

Let:

T_{manual} be the average response time under manual handling

$T_{AI\ is}$ the average response time under the AI-assisted system

$$\text{Response Time Reduction (\%)} = \frac{T_{manual} - T_{AI}}{T_{manual}} \times 100$$

Understanding Results:

We tested complaints, and we found that our AI-based system reduced average response time by about 40% compared to manual handling. This shows that our system gives faster replies and processes complaints more quickly.

Automation Rate:

Automation rate shows the number of complaints handled completely by an AI system without human help, and a complaint is counted as automated if AI resolves it, sends it to the relevant department, and no manual checking or escalation is required.

Calculation Method:

$$\text{Automation Rate} = \frac{\text{Number of AI-Resolved Complaints}}{\text{Total Number of Complaints}} \times 100$$

Understanding Results: A total of 150 complaints were tested. Among these, 102 complaints were resolved automatically, while the remaining cases required intervention by a human officer because they were difficult, sensitive, or unclear because they were difficult, sensitive, or unclear. This shows an automation rate of 68%, which indicates that a significant portion of routine complaints can be managed automatically by the system.

Evaluation Validity and Limitations:

Although the results are promising, they were obtained under controlled datasets and simulated operational conditions. When deployed in real government environments, system performance may vary due to differences in complaint types, operational constraints, and user behavior. Future work will focus on long-term testing using real government data to further improve system reliability and effectiveness.

Usability and User Acceptance:

To evaluate system usability, a small pilot test was conducted with several citizens and government officers who were asked how easy it was to submit complaints, understand system messages, and use the system overall. Most users said that writing complaints in simple language and using English or Urdu made the process much easier, and officers reported that cases could be reviewed more clearly and had less manual work.

Large-scale testing was not conducted in this study, but the early feedback was positive. This shows that users found the system helpful and easy to use, and it can be tested further for real government use.

Comparison with Other Studies:

Compared with earlier complaint management systems, such as [6], Service Sphere demonstrates improved performance. Earlier systems mainly used basic NLP methods and did not clearly report standard evaluation results. In contrast, Service Sphere evaluates its performance using clear measures such as precision, recall, F1-score, and a confusion matrix.

The system achieved a classification accuracy of 92%, which is comparable to or higher than other AI-based public service systems discussed by [2], where detailed performance results were not clearly reported. Unlike those studies, Service Sphere shows real improvements, including a 40% reduction in average response time.

[10] discussed the importance of human and AI working together in citizen service systems, but did not provide numerical results. Service Sphere follows this human–AI collaboration approach and also provides clear performance numbers. In addition, it supports both English and Urdu and uses automated workflows. These results show that Service Sphere is more practical and effective than many existing complaint management systems.

Conclusion:

This study demonstrates that an AI-based complaint management system can significantly improve public service delivery and citizen satisfaction. Service Sphere accurately classifies complaints, routes them to the appropriate departments, and provides timely updates to users.

Built using modern web technologies and Large Language Models such as GPT and LLaMA, the system achieved 92% classification accuracy, reduced response time by 40%, and automated 68% of complaint cases. These results indicate that the proposed approach can reduce manual workload, enhance operational efficiency, and improve transparency in government services.

Furthermore, the system incorporates human oversight for complex cases, ensuring accountability and reliability. Overall, Service Sphere provides a scalable and practical solution for modern digital governance.

References:

[1] Hila Mehr, "Artificial Intelligence for Citizen Services and Government," *Harvard Ash Cent. Technol. Democr. Fellow*, 2017, [Online]. Available: https://ash.harvard.edu/wp-content/uploads/2024/02/artificial_intelligence_for_citizen_services.pdf

[2] G. M. Colin van Noordt, "Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union," *Gov. Inf. Q.*, vol. 39, no. 3, p. 101714, 2022, [Online]. Available: http://sciencedirect.com/science/article/abs/pii/S0740624X22000478

[3] H. M. S. S. Javairia Shafiq, "Use of ICTs and Artificial Intelligence to Overcome Judicial Trial Delays in Pakistani Courts," *Pakistan Lang. Humanit. Rev.*, vol. 6, no. 2, 2022, [Online]. Available: https://ojs.plhr.org.pk/journal/article/view/131

[4] A. D. Samuel-Okon and O. Abejide, "Bridging the Digital Divide: Exploring the Role of Artificial Intelligence and Automation in Enhancing Connectivity in Developing Nations," *SSRN Electron. J.*, May 2024, doi: 10.2139/SSRN.4839447.

[5] "Research on Artificial Intelligence For Citizen Services and Government." Accessed: Feb. 10, 2026. [Online]. Available: https://www.researchgate.net/publication/357033081_Research_On_Artificial_Intelligence_For_Citizen_Services_And_Government

[6] P. Wadkar, A. Raorane, S. Bushra, and shreyas shedge, "AI-Driven Complaint Management System," *SSRN Electron. J.*, May 2021, doi: 10.2139/ssrn.3866078.

[7] Asad Khattak, Waqas Tariq Paracha, Muhammad Zubair Asghar, "Fine-Grained Sentiment Analysis for Measuring Customer Satisfaction Using an Extended Set of Fuzzy Linguistic Hedges," *Int. J. Comput. Intell. Syst.*, vol. 13, pp. 744–756, 2020, [Online]. Available: https://link.springer.com/article/10.2991/ijcis.d.200513.001

[8] Z. Y. Xudong Zhang, "Community Governance Based on Sentiment Analysis: Towards Sustainable Management and Development," *Sustainability*, vol. 15, no. 3, p. 2684, 2023, [Online]. Available: https://www.mdpi.com/2071-1050/15/3/2684

[9] Basit Afzal, Dr. Muhammad Imran Khan, "AI Adaptability and Customer Satisfaction in Telecom Complaint Handling: The Mediating Role of Customer Psychographics and Perceived Usefulness," *Soc. Sci. Rev. Arch.*, vol. 3, no. 3, pp. 1278–1291, 2025, [Online]. Available: https://www.researchgate.net/publication/394834111_AI_Adaptability_and_Customer_Satisfaction_in_Telecom_Complaint_Handling_The_Mediating_Role_of_Customer_Psychographics_and_Perceived_Usefulness

[10] Ana Marta Goucha de Carvalho, "Government Augmented Intelligence the Use of AI to Improve Citizen Relationship Management," *Nov. Inf. Manag. Sch. Inst. Super. Estatística e Gestão Informação*, 2022.

[11] Naomi Aoki, "An experimental study of public trust in AI chatbots in the public sector," *Gov. Inf. Q.*, vol. 37, no. 4, p. 101490, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0740624X1930406X

[12] D. I. N. Tayyaba Jadoon, "Leveraging Artificial Intelligence for Social Governance in Pakistan: Toward an Ethical and Policy-Oriented Framework," *Acad. Int. J. Soc. Sci.*, vol. 4, no. 3, 2025, [Online]. Available: https://academia.edu.pk/index.php/Journals/article/view/541?articlesBySimilarityPage=1



Copyright © by authors and 50Sea. This work is licensed under the Creative Commons Attribution 4.0 International License.