"Minwon_Itda": An AI-Based Personalized Public Service Guidance System Using Hybrid Retrieval and Retrieval-Augmented Generation

Jisoo Han[1], Yeseo Kim[2], Hyuckjin Lee[3], BogyeongJeon[4], Seungjae Lee[5]*

Department of Computer Engineering, Sun Moon University, Korea

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1309000203

Received: 24 September 2026; Accepted: 29 September 2026; Published: 20 October 2026

ABSTRACT

Accessing public administrative services often requires citizens to navigate complex procedures, interpret specialized terminology, and identify eligibility requirements and supporting documents. Conventional e-government portals primarily rely on keyword-based search and predefined categories, which can make service discovery difficult for users who are unfamiliar with administrative terminology. This paper presents Minwon Itda, an AI-based personalized public service guidance system designed to facilitate the discovery and completion of administrative procedures through natural-language interaction. The proposed system integrates a two-stage service classification mechanism combining rule-based keyword mapping with a large language model (LLM), together with a hybrid retrieval pipeline that combines semantic vector search and keyword matching. Retrieved information is used to generate context-grounded responses through retrieval-augmented generation (RAG). Beyond conversational question answering, the system provides structured, step-by-step procedural guidance, personalized document checklists, administrative terminology explanations, and multilingual interaction. A session-management mechanism maintains users' application progress, while caching and fallback mechanisms support responsive and resilient service delivery. The system was implemented using a web-based three-tier architecture with a FastAPI backend, a Next.js frontend, MariaDB, and Redis. The reported evaluation achieved a public-service classification accuracy of 98%. In latency measurements conducted over 100 requests, the average response time was approximately 0.4 s for keyword-matched queries and 1.8 s when LLM augmentation was invoked, while cached responses were delivered in less than 0.1 s on average. These results demonstrate the technical feasibility of integrating hybrid retrieval, LLM-based response generation, and structured procedural guidance into a unified public-service application. The proposed system provides an implementation framework for improving access to administrative information, although further user studies and broader evaluations are required to establish its effectiveness across diverse populations and service scenarios.

Keywords: E-Government, Public Service Chatbot, Personalized Service Guidance, Large Language Models, Retrieval-Augmented Generation, Hybrid Information Retrieval

INTRODUCTION

Digital government initiatives have expanded the availability of public administrative information through online portals and web-based service platforms. These systems enable citizens to access administrative procedures, eligibility requirements, and application documents without necessarily visiting government offices. However, the availability of information does not always translate into effective access to public services. Citizens may still encounter difficulties identifying the appropriate service, understanding administrative terminology, determining their eligibility, and completing the required application procedures. Research on e-government chatbots has identified information discovery, service accessibility, and the integration of conversational systems into administrative processes as important issues in the development of digital public services [1], [2].

A major limitation of conventional administrative information systems is their dependence on predefined categories and keyword-based search. Such mechanisms generally require users to possess at least some knowledge of the terminology used by the relevant administrative agency. When users describe their needs in everyday language rather than official service names, the retrieved results may not adequately reflect their intentions. These difficulties can be particularly relevant to older adults, foreign residents, and other users who are unfamiliar with administrative language or digital service procedures. Although conversational interfaces provide an alternative means of accessing information, their practical value depends on the reliability of the responses and the extent to which the interaction supports users' actual tasks [3], [4].

Recent advances in large language models (LLMs) have enabled conversational systems to interpret natural-language queries and generate contextually appropriate responses. Nevertheless, directly relying on an LLM to provide administrative guidance introduces important challenges. Public-service information is domain-specific and may change over time, while an unsupported or inaccurate response can lead users to misunderstand eligibility conditions or application requirements. Retrieval-augmented generation (RAG) addresses part of this problem by incorporating information retrieved from an external knowledge source into the generation process [6], [7]. However, the reliability of a RAG-based system also depends on the quality of retrieval and the degree to which generated responses remain consistent with the retrieved evidence.

Information retrieval research provides complementary approaches for addressing the challenge of matching user queries with relevant documents. Dense retrieval uses semantic representations to identify passages that are conceptually related to a query, whereas lexical retrieval can identify explicit terms appearing in the query and the source documents [8]. Benchmarking studies have demonstrated that retrieval approaches exhibit different performance characteristics across datasets and tasks, motivating the evaluation of retrieval methods in their intended application domains [9]. Combining semantic and lexical retrieval is therefore a practical design choice for administrative queries, which may contain both informal descriptions and specific legal or procedural terms.

This paper presents Minwon Itda, an AI-based personalized public service guidance system that integrates conversational interaction, hybrid retrieval, LLM-based response generation, and structured procedural navigation. The system is designed not merely to return relevant administrative information but also to organize that information into actionable guidance. Its principal components include a two-stage classification mechanism combining rule-based keyword mapping and an LLM, a hybrid search pipeline combining vector similarity and keyword matching, and a structured guidance interface that presents procedural steps and required-document checklists. The system additionally incorporates multilingual interaction, administrative terminology explanations, session-based progress tracking, caching, and fallback mechanisms.

The main contributions of this study are as follows:

1. Hybrid intent classification and information retrieval: A two-stage classification mechanism and a hybrid retrieval pipeline are integrated to identify relevant public services from natural-language queries.
2. Evidence-grounded administrative guidance: RAG-based response generation is combined with structured administrative data to support domain-specific responses.
3. Procedural support beyond question answering: Retrieved information is organized into step-by-step instructions and document checklists, with session-based tracking of application progress.
4. End-to-end system implementation and evaluation: A web-based system incorporating multilingual interaction, database management, caching, and fault-tolerance mechanisms is implemented and evaluated in terms of classification accuracy, response latency, and operational behavior.

The remainder of this paper is organized as follows. Section II reviews related work on e-government chatbots, conversational interfaces, and retrieval-augmented information systems. Section III describes the proposed system architecture and its principal components. Section IV presents the implementation and experimental results. Finally, Section V concludes the paper and discusses directions for future research.

Related Work

Conversational Systems for E-Government Services

Conversational agents have been investigated as an alternative interface for accessing digital information and services. Unlike conventional graphical interfaces, which often require users to navigate menus and predefined categories, chatbots allow users to express information needs through natural language. General research on chatbot technology has examined the evolution of conversational systems, their underlying architectures, and their applications across different domains [4]. From a human–computer interaction perspective, conversational interfaces also introduce opportunities and challenges concerning usability, user expectations, and the design of interactions between people and automated systems [5].

In the public sector, chatbots have been adopted to support access to government information and administrative services. Cortés-Cediel et al. [1] examined e-government chatbots through a sociotechnical framework and discussed challenges in their implementation and evaluation. Their work highlights the importance of considering not only information access but also the broader administrative and citizen-participation context. Similarly, Larsen and Følstad [2] investigated the effects of chatbots on public service provision through interviews with citizens and service providers. Their findings indicate that chatbots can contribute to access and efficiency while also illustrating that their effects depend on how the technology is incorporated into public-service processes.

Trust is another important consideration in the deployment of conversational systems for government services. Aoki [3] experimentally examined initial public trust in AI chatbots in the public sector and found that trust varied according to the subject of the inquiry and the purposes communicated by the government. These findings suggest that the design of a public-service chatbot should account for the consequences of administrative information and the expectations of its users.

Despite these developments, providing a conversational answer is not equivalent to helping a citizen complete an administrative task. Users may need to identify eligibility conditions, gather supporting documents, follow a sequence of procedural steps, and monitor their progress. *Minwon Itda* addresses this implementation requirement by combining conversational information access with structured procedural instructions and personalized checklists. In this respect, the system extends the interaction from information retrieval toward task-oriented administrative guidance.

Retrieval-Augmented Generation and Domain-Specific Question Answering

Large language models can interpret natural-language requests and generate fluent responses, but their outputs may contain unsupported or inaccurate information. Retrieval-augmented generation provides a framework in which a generative model uses information retrieved from an external knowledge source to support its responses. Lewis et al. [6] introduced a retrieval-augmented generation architecture for knowledge-intensive natural-language processing tasks, combining a parametric language model with a non-parametric retrieved knowledge source. Subsequent research has surveyed the development of RAG architectures and identified retrieval quality, knowledge integration, and generation reliability as important components of these systems [7].

For public administrative applications, grounding responses in a structured and maintainable information source is particularly relevant because service descriptions, eligibility requirements, and procedural instructions are domain-specific. However, RAG does not independently guarantee factual correctness. Errors may arise when relevant documents are not retrieved, when retrieved information is incomplete, or when the generated response is inconsistent with the source material [7]. Accordingly, a practical administrative guidance system should combine retrieval with domain-specific data organization and an output structure that makes procedural requirements explicit.

The proposed system applies this principle by retrieving administrative information from its service database and using the retrieved content to support LLM-generated guidance. In addition, the system separates the

presentation of procedural steps and required documents from the conversational response. This design allows users to inspect the relevant application requirements in a structured format rather than relying exclusively on a free-form generated answer.

Hybrid Retrieval and Re-Ranking

Information retrieval methods can be broadly classified into lexical and semantic approaches. Lexical retrieval identifies documents through explicit term matching, whereas dense retrieval represents queries and passages as vectors to identify semantic relationships. Karpukhin et al. [8] demonstrated the effectiveness of dense passage retrieval for open-domain question answering and compared it with conventional sparse retrieval. However, retrieval performance can vary across domains and evaluation datasets.

Thakur et al. [9] introduced BEIR, a heterogeneous benchmark for evaluating information retrieval systems across multiple tasks and domains. Their evaluation illustrates that no single retrieval approach exhibits uniform behavior across all settings and that retrieval effectiveness must be considered alongside computational cost. These findings motivate application-specific evaluation when selecting retrieval methods for a specialized domain such as public administration.

A further approach is to re-rank an initial set of retrieved candidates using a more expressive model. Nogueira and Cho [10] demonstrated the use of BERT for passage re-ranking, in which candidate passages are evaluated in relation to a query. Re-ranking can therefore serve as an additional stage for refining an initial retrieval result.

Building on these approaches, Minwon Itda combines vector-based semantic retrieval with keyword matching. The implemented retrieval pipeline assigns a weight of 60% to vector search using text-embedding-3-small and 40% to keyword matching. The resulting candidates are subsequently re-ranked using an LLM-based mechanism informed by domain-specific data. The retrieval weights were selected through a grid search using a 50-item evaluation dataset. This configuration is intended to accommodate both semantically expressed user needs and queries containing explicit administrative terminology.

Research Gap and Positioning of the Proposed System

The reviewed literature addresses several complementary aspects of conversational public services, including citizen access, public trust, chatbot design, knowledge-grounded response generation, and information retrieval [1]–[10]. However, these aspects represent distinct technical and service-design concerns. Conversational access alone does not provide procedural tracking, while retrieval and response generation alone do not necessarily organize the results into a sequence of administrative actions.

The contribution of Minwon Itda is an integrated implementation that connects these components within a single public-service guidance workflow. The system combines natural-language service classification, hybrid retrieval, RAG-based response generation, structured procedural instructions, personalized document checklists, and session-based progress management. It also incorporates multilingual interaction and operational mechanisms such as caching and fallback processing.

Rather than proposing a new general-purpose retrieval algorithm or a new LLM architecture, this study focuses on the system-level integration of established AI and information retrieval techniques for personalized administrative guidance. Its evaluation examines classification accuracy, response latency, and implementation behavior. Further evaluation with larger and more diverse query datasets, direct comparisons with baseline retrieval methods, and controlled user studies will be necessary to establish the system's generalizability and its effects on task completion and user experience.

The proposed smart diffuser adopts a three-layer architecture consisting of a Raspberry Pi 4, an ESP32 LoLin D32 Pro, and a user smartphone application named AromaCare. The overall system architecture is shown in Figure 1.

Proposed System

System Architecture Design

The AI-based administrative public service guidance web service proposed in this study follows a three-tier architecture consisting of a front-end client, a back-end web server, and an administrative data parsing module. According to previous research, a computation offloading strategy that delegates compute-intensive tasks to the back-end is highly advantageous for reducing client overhead [5]. This system is also designed to offload the logic for refining and parsing unstructured data to the back-end server, thereby ensuring the availability of the front-end React application [5].

In particular, users who lack experience in applying for public services prioritize intuitiveness, demanding easy accessibility and clear explanations of terminology right from the initial stages [7]. To prevent users from getting lost in complex textual guidelines, this system establishes a structure that maps the required documents, issuing agencies, and precautions for each step, enabling users to intuitively access the guide through a 'View Details' button.

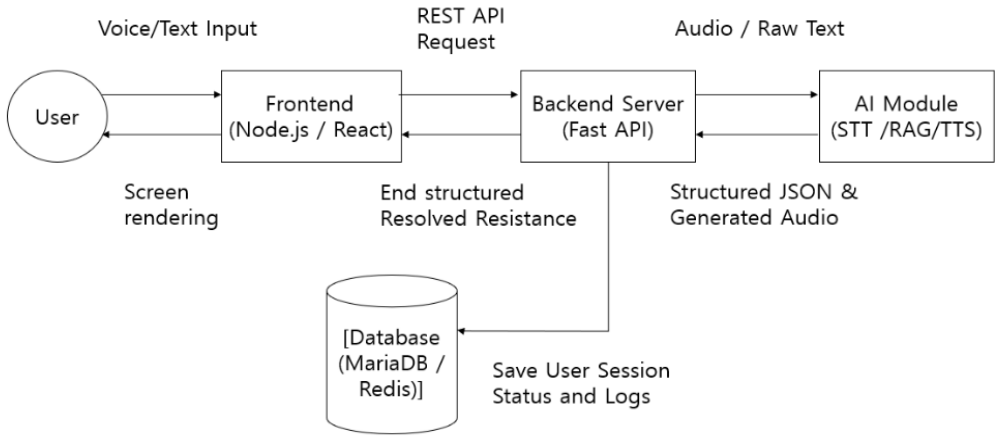


Fig.1.Overallsystemarchitectureofminwonitda

The overall system pipeline begins with the user requesting assistance for their specific public service needs through the front-end interface—utilizing either everyday natural language text or voice input—and ultimately concludes by guiding the user through the necessary administrative procedures.

Service Security, Stability, and Access Control Environment

Recently, in microservices and REST API environments, combining JWT with an in-memory database has emerged as an essential security architecture requirement to overcome the limitations of traditional token-based authentication and provide robust server protection [8][9]. This system adopts a stateless JWT authentication framework to protect the back-end server from indiscriminate API requests and unauthorized access. Additionally, a Redis-based rate-limiting algorithm is applied at the endpoint level to strictly limit the number of requests from a specific IP address to 60 per minute. This effectively prevents server overhead caused by brute-force attacks and excessive traffic.

Administrative Public Service Data Collection and Information Structure

To ensure the reliability of the unstructured data processed by the back-end server, 74 public service scenarios and 110 administrative terms are collected from five websites—including Government24, Bokjiro, and Hi Korea—using a Selenium-based integrated crawler. The extracted data is enriched with detailed procedural information via an LLM and then loaded into the database (MariaDB) through an UPSERT pipeline. To ensure the stability of the data acquisition process, features such as checkpoint-based pause and resume, an automatic

three-time retry mechanism, and link validity verification were implemented. The data model is managed across 12 core tables divided into a public zone (owned by the chatbot back-end) and a user zone (domain-owned). A read-only principle is applied to the public zone to guarantee data integrity.

AI-Based Hybrid Search and Re-ranking System

To accurately grasp the diverse intents behind users' everyday language queries and prevent hallucinations, an advanced 3-stage AI search algorithm was implemented. The core search logic follows a RAG architecture that utilizes a weighted combination of vector search (60%) based on `text-embedding-3-small` and keyword matching (40%). The retrieved Top-N candidates are then re-ranked to determine the final answer using an LLM fine-tuned on domain knowledge (751 records). The hybrid weights were set to the optimal ratio achieving the highest Hit Rate through a grid search utilizing a 50-item evaluation dataset.

Customized Multilingual and Voice Support, and Fault Tolerance Mechanisms

To eliminate the language barriers faced by foreigners and digitally marginalized groups, an advanced multilingual and voice processing pipeline was established within the AI module. Upon receiving user input, the system automatically detects the language within 1ms, performs customized eligibility filtering for 10 types of foreign visas, and executes batch translation caching into four supported languages. Voice data is converted in real-time using an STT (Speech-to-Text) model (Whisper), which boasts a 99-100% Korean recognition accuracy, along with a TTS (Text-to-Speech) system featuring six different voices. Notably, to ensure the stability of the overall system, a multiple fallback mechanism was implemented to overcome the risk of a Single Point of Failure (SPOF). This mechanism bypasses external LLM API call failures by routing to direct database-driven guidance, and transitions to local in-memory processing in the event of a Redis failure.

Database Modeling and Schema Design

The database architecture of this system is designed based on a relational database (MariaDB) to maintain the integrity of the collected unstructured administrative data and to track personalized user session states in real-time. The overall schema is broadly divided into a read-only 'Public Administrative Domain' and a write-based 'User Interaction Domain', structured as shown in <Table I>.

TABLE I KEY DATABASE TABLES BY DOMAIN

Category	Key Tables	Role
Administrative Services	Services, procedures, documents, civil_service_types	Definition of public service metadata and relational normalization of step-by-step procedures and required documents
Quality and Source	Service_souces, data_quality_issues	Tracking of external original sources and managing data quality issues (e.g., parsing, invalid links)
User Session	Users, user_session, chat_session	User authentication, storage of chatbot conversation history, and session management for maintaining multi-turn context
Progress Management	Service_progress, check_progress	Tracking individual users' progress through administrative procedures and synchronizing checklist completion rates
Barrier Resolution	Terms, service_translations, target_groups	Administrative terminology data, multilingual translation mapping, and target group (e.g., elderly, foreigners) classification

Implementation And Experimental Evaluation

Server System and Technology Stack Validation

The Minwon Itda system was implemented in an Ubuntu-based remote server environment using Python 3.11,

FastAPI 0.110, Next.js 14, MariaDB 10.11, and Redis 7.2. The OpenAI GPT-4o-mini model was used for response generation, and the database, cache, back-end, and front-end components were configured as containerized services using Docker Compose. The back-end and front-end communicate through REST API endpoints.

The experimental environment was designed to evaluate not only the technical operation of the system but also the effectiveness of its major information-processing components. In particular, the evaluation focused on service classification, hybrid retrieval, LLM-based re-ranking, response generation, and the presentation of structured administrative guidance.

The administrative knowledge base used in the evaluation contains 74 public-service scenarios and 110 administrative terms collected from five public information sources, including Government24, Bokjiro, and Hi Korea. The collected information includes service descriptions, eligibility conditions, required documents, and procedural information.

For the evaluation, user queries were constructed to represent both official administrative terminology and natural-language expressions that users may employ when seeking public services. The query set included service identification, eligibility, required-document, and procedure-related questions. Each evaluation query was associated with a reference public-service category and corresponding administrative information.

The previously reported classification accuracy of 98% was obtained from the service classification evaluation. To provide a more rigorous assessment, the expanded evaluation considers precision, recall, and F1-score in addition to accuracy. Retrieval performance is further evaluated using Top-k retrieval accuracy, Recall@k, and Mean Reciprocal Rank (MRR).



Fig. 2 Application Screen Comparison (AI Guide ON/OFF)

Application Screen Implementation Test and UX Evaluation

To verify whether the Minwon Itda service effectively applies the AI guide to the actual application interface, an in-app screen implementation test was conducted. The same user query was processed under two conditions: the baseline mode, in which the AI guide was disabled and a simple text response was provided, and the AI guide mode, in which step-by-step guidance and a customized checklist were displayed.

As shown in Fig. 2, the AI guide mode presents procedural steps, required documents, and explanations of administrative terminology in a structured form. This design is intended to reduce the amount of interpretation required from users when converting general administrative information into concrete actions.

Rather than assuming that structured presentation automatically reduces cognitive workload, the effectiveness of the interface was evaluated in terms of task comprehension and task completion. In a subsequent controlled evaluation, completion time, error rate, comprehension, satisfaction, and perceived workload can be measured to determine whether the structured guidance provides measurable benefits compared with the baseline interface.

Server Communication Logs and Data Consistency Validation

To verify the reliability of the end-to-end communication pipeline, runtime logs from the FastAPI back-end and Next.js front-end were collected and analyzed during actual system operation. Cross-tier traceability was established using browser-side network logs and structured back-end logging.

Figure 3 presents a representative communication sequence for a user query. The front-end transmits the user query, language information, and session identifier to the back-end. The back-end subsequently returns the generated answer together with the structured response, procedural steps, customized checklist, and response latency.

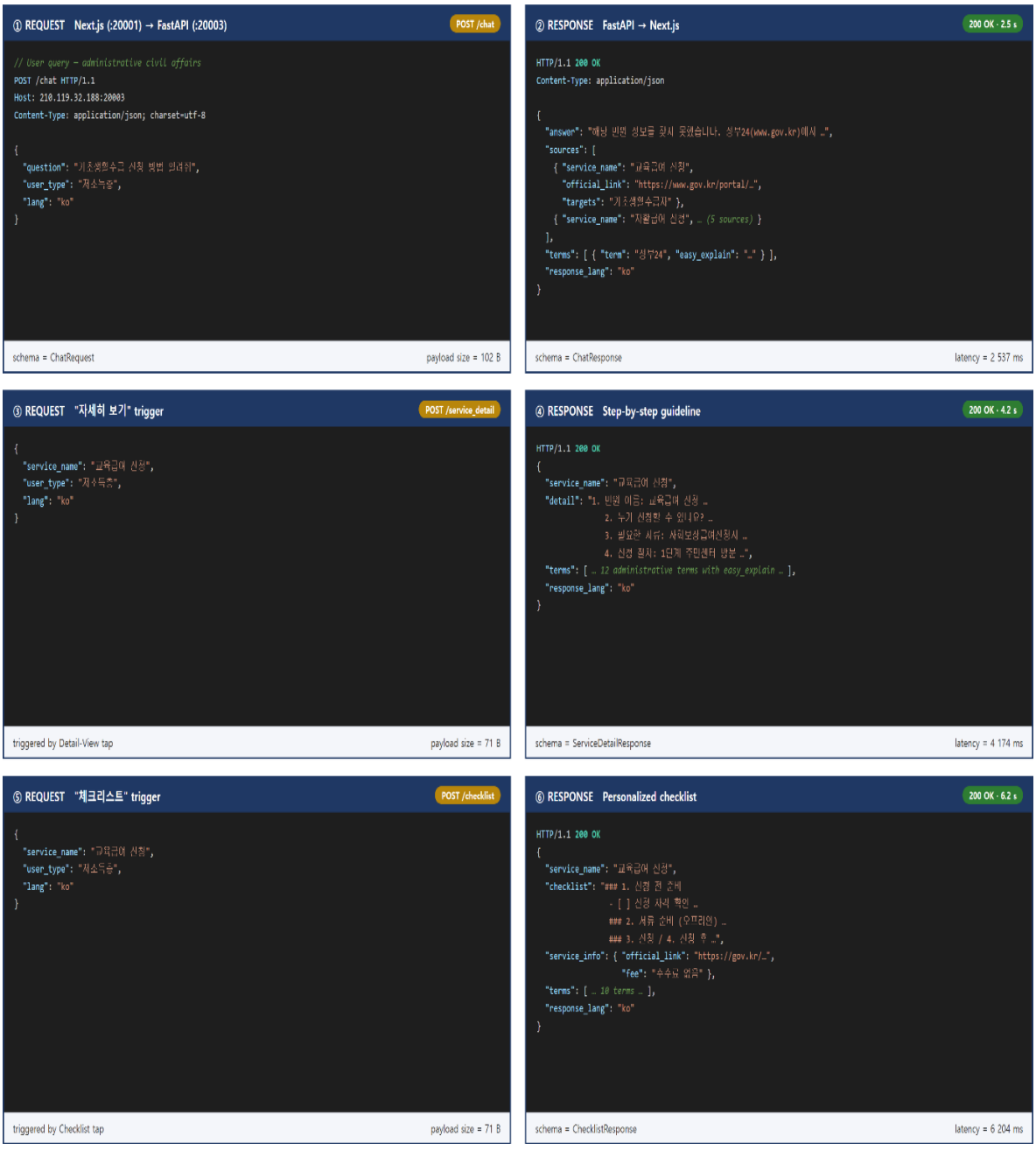


Fig. **3** Runtime Communication Log (Frontend ↔ Backend, 200 OK)

The log-based evaluation was used to verify three system-level characteristics: (i) consistency of the client-server data structure, (ii) correct invocation of the processing pipeline, including language detection, cache lookup, keyword matching, database retrieval, and LLM augmentation, and (iii) correct propagation of multilingual responses through the translation layer.

In addition to operational consistency, the experimental evaluation was extended to examine the contribution of individual processing components. The proposed hybrid retrieval and LLM re-ranking configuration was compared with keyword-only retrieval, vector-only retrieval, and hybrid retrieval without re-ranking. This comparison was designed to determine whether the integration of semantic and lexical retrieval provides measurable improvement over individual retrieval methods.

Performance Evaluation of OpenAI LLM Response Generation

To evaluate the practical performance of the LLM-based response-generation component, response latency was measured over 100 requests in the implemented remote server environment. The total latency was measured from the time a user query reached the back-end until the response was returned to the client.

The evaluation considered three operational conditions: keyword/database retrieval, LLM-augmented response generation, and cached response retrieval. The measured average latency was approximately 0.4 seconds when keyword matching was successful and approximately 1.8 seconds when an LLM augmentation call was required. Cached responses were delivered in less than 0.1 seconds on average.

In addition to latency, the quality of generated responses was considered from an administrative information perspective. Because fluent language generation alone does not guarantee the correctness of administrative guidance, the expanded evaluation examines answer correctness, completeness, faithfulness, and source-groundedness.

Faithfulness refers to whether the factual statements in the generated response are supported by the retrieved administrative information. Source-groundedness evaluates whether the retrieved source provides direct evidence for the corresponding response. Unsupported factual statements are treated as potential hallucinations.

The evaluation also compares the proposed RAG configuration with RAG configurations based on keyword-only and vector-only retrieval. This comparison is intended to determine whether hybrid retrieval and LLM-based re-ranking improve the factual reliability of generated administrative guidance.

Evaluation of Hybrid Retrieval and Re-ranking

The implemented retrieval pipeline combines vector similarity based on `text-embedding-3-small` and keyword matching. The current implementation uses a 60% vector and 40% keyword weighting scheme, followed by LLM-based re-ranking of the retrieved candidates.

The initial weighting configuration was obtained through a grid search using a 50-item evaluation dataset. To improve the methodological rigor of this configuration, the expanded evaluation treats the weighting process as a validation procedure rather than as part of the final test. Several vector-to-keyword weighting combinations are evaluated and compared using retrieval metrics such as Recall@k and MRR.

The final test queries are kept separate from the queries used for selecting the retrieval weights. This separation reduces the possibility of overestimating system performance due to parameter selection on the same data used for evaluation.

The contribution of the major retrieval components is additionally examined through an ablation comparison. Specifically, vector retrieval, keyword retrieval, and LLM-based re-ranking are independently removed while the remaining components are kept unchanged. This allows the performance contribution of each component to be examined within the same experimental environment.

Evaluation of Administrative Guidance and Source Consistency

Because Minwon Itda is intended to provide actionable administrative guidance rather than simple conversational responses, the correctness of generated guidance was evaluated against the retrieved administrative sources.

For each evaluated response, the generated information was compared with the corresponding source information regarding service eligibility, required documents, procedural steps, and other administrative conditions. Responses were classified according to whether the information was supported by the retrieved source.

The evaluation distinguishes between answer correctness and source-groundedness. A response may be linguistically appropriate while still containing information that is not supported by the retrieved source. Therefore, source consistency was considered an essential component of the evaluation.

The administrative knowledge base is primarily constructed from official public-service information sources. The system also maintains source-related information and data-quality records in the database, as described in Table I. When multiple sources provide overlapping administrative information, the information provided by the responsible government agency is treated as the primary reference. When available, the most recently updated official information is preferred.

Multilingual and Voice Processing Evaluation

The system provides multilingual interaction and voice input to improve accessibility for users with different language backgrounds. The multilingual processing pipeline automatically detects the input language, performs language-specific processing, and provides translated responses for the supported languages.

The expanded evaluation examines multilingual performance separately rather than assuming that successful translation automatically produces equivalent administrative guidance across languages. Retrieval accuracy, answer correctness, terminology preservation, and response latency are considered for each supported language.

For voice processing, Whisper is used as the speech-to-text component. The previous description of Whisper as providing “99–100% Korean recognition accuracy” has been revised because recognition performance depends on the evaluation dataset, acoustic conditions, speakers, and evaluation metric. The revised evaluation uses Word Error Rate (WER) or Character Error Rate (CER) when an independent Korean speech test set is available.

Statistical Evaluation and Experimental Reliability

The expanded evaluation reports accuracy, precision, recall, F1-score, Recall@k, and MRR rather than relying on classification accuracy alone. For response-quality evaluation, correctness, completeness, faithfulness, and source-groundedness are additionally reported.

For comparative experiments, mean values and standard deviations are reported where repeated measurements are available. Statistical significance is evaluated at the 0.05 level using an appropriate statistical test according to the experimental design. For paired comparisons, a paired statistical test is used, while independent groups are evaluated using an independent-sample test. When distributional assumptions are not satisfied, a corresponding non-parametric test is applied.

These procedures are intended to distinguish observed differences between retrieval and guidance configurations from differences that may arise from random variation in the evaluation queries.

Overall, the evaluation framework extends the original implementation-oriented validation by separately examining classification performance, retrieval effectiveness, response quality, source consistency,

multilingual processing, and system latency. This provides a more comprehensive evaluation of whether the proposed hybrid retrieval and RAG architecture can reliably support administrative service guidance.

CONCLUSION

This study presented *Minwon Itda*, an AI-based personalized public service guidance system designed to improve access to administrative information and simplify the process of navigating public services. To address the limitations of conventional keyword-based search and predefined service categories, the proposed system integrates natural-language interaction, hybrid information retrieval, and large language model (LLM)-based response generation within a unified service framework.

The system incorporates a two-stage classification mechanism that combines rule-based keyword mapping with an LLM, together with a hybrid retrieval pipeline that integrates semantic vector search and keyword matching. Retrieved administrative information is used to generate context-grounded responses through retrieval-augmented generation (RAG). In addition to conversational responses, the system provides step-by-step procedural instructions, personalized document checklists, administrative terminology explanations, and multilingual support. Session-based progress tracking enables users to manage their application procedures, while caching and fallback mechanisms support efficient and resilient service delivery.

The system was implemented using a three-tier web architecture comprising a Next.js frontend, a FastAPI backend, MariaDB, and Redis. Experimental results showed a public-service classification accuracy of 98%. In a 100-request latency evaluation, the average response time was approximately 0.4 s for keyword-matched queries and 1.8 s for queries requiring LLM augmentation. Cached responses achieved an average response time of less than 0.1 s. These results demonstrate the technical feasibility of integrating AI-based service classification, hybrid retrieval, and structured procedural guidance into a single administrative service platform.

The principal contribution of this study lies in the integration of conversational AI with actionable, procedure-oriented public service guidance. Rather than limiting the interaction to information retrieval or free-form question answering, the proposed system organizes administrative information into structured steps and document checklists. This approach provides a practical foundation for developing accessible and personalized e-government services. Nevertheless, the current evaluation primarily addresses system implementation and technical performance; further empirical investigation is required to establish its effectiveness in improving user comprehension, task completion, and overall satisfaction.

Future Work

Several directions will be explored to improve the reliability, scalability, and practical applicability of the proposed system.

First, the coverage and freshness of administrative information will be improved through more systematic integration with official government data sources. Where supported by the relevant agencies, future versions will investigate mechanisms for retrieving updated service requirements and application-processing status. This would reduce dependence on periodically collected information and help ensure that users receive current administrative guidance.

Second, the retrieval and response-generation components will be evaluated using larger and more diverse datasets. Future experiments will compare alternative retrieval configurations and assess retrieval effectiveness, response accuracy, source consistency, and latency. In particular, the contribution of hybrid retrieval and LLM-based re-ranking will be examined through controlled baseline comparisons. Such evaluations will help determine the extent to which each component contributes to the overall performance of the system.

Third, user-centered evaluations will be conducted with participants representing different age groups, language backgrounds, and levels of digital literacy. These studies will examine usability, perceived clarity,

trust, cognitive workload, and administrative task-completion rates. The results will provide empirical evidence of whether structured procedural guidance and personalized checklists improve the accessibility of public services compared with conventional information-search interfaces.

Finally, a continuous system-monitoring and improvement framework will be developed. This framework will track query patterns, retrieval failures, response quality, and service usage while supporting the identification of missing or outdated administrative information. Appropriate privacy and data-protection measures will be incorporated into the collection and analysis of user interaction data.

Through these developments, *Minwon Itda* aims to evolve into a more reliable, maintainable, and user-centered platform for personalized public administrative guidance. The proposed framework provides a basis for further research into the integration of conversational AI, information retrieval, and structured workflow support in digital government services.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation) in 2026 (No. 2024-0-00023).

REFERENCES

1. M. E. Cortés-Cediel, A. Segura-Tinoco, I. Cantador, and M. P. Rodríguez Bolívar, “Trends and challenges of e-government chatbots: Advances in exploring open government data and citizen participation content,” Government Information Quarterly, vol. 40, no. 4, Art. no. 101877, 2023.
2. A. G. Larsen and A. Følstad, “The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers,” Government Information Quarterly, vol. 41, no. 2, Art. no. 101927, 2024.
3. N. Aoki, “An experimental study of public trust in AI chatbots in the public sector,” Government Information Quarterly, vol. 37, no. 4, Art. no. 101490, 2020.
4. E. Adamopoulou and L. Moussiades, “An overview of chatbot technology,” in Artificial Intelligence Applications and Innovations (AIAI 2020), pp. 373–383. Springer, 2020.
5. A. Følstad and P. B. Brandtzaeg, “Chatbots and the new world of HCI,” Interactions, vol. 24, no. 4, pp. 38–42, 2017.
6. P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in Advances in Neural Information Processing Systems, vol. 33, 2020.
7. Y. Gao et al., “Retrieval-augmented generation for large language models: A survey,” arXiv preprint arXiv:2312.10997, 2023.
8. V. Karpukhin et al., “Dense passage retrieval for open-domain question answering,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6769–6781, 2020.
9. N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, and I. Gurevych, “BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models,” in Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, vol. 1, 2021.
10. R. Nogueira and K. Cho, “Passage re-ranking with BERT,” arXiv preprint arXiv:1901.04085, 2019.