

Implementation and Application of a Generative AI Virtual Librarian: A Case Study of the National Library of Public Information, Taiwan

Hsiang-Ping Ma & Yun-Fan Chen

Abstract:

Since its transformation into Taiwan's first national digital library in 2013, the National Library of Public Information (NLPI) has continuously adopted innovative technologies to advance intelligent services. In response to the rise of large language models (LLMs), NLPI launched the "Generative AI Virtual Librarian" project in 2023 and developed Xiaoshu, a virtual librarian capable of natural voice interaction. Centered on generative AI, the system integrates four databases and eight intent categories, applying retrieval-augmented generation (RAG) and speech recognition to provide collection search assistance, book recommendations, service inquiry responses, and social interaction.

Xiaoshu effectively handles many repetitive yet diverse library inquiries, demonstrating the linguistic flexibility of generative AI. Compared with rule-based systems that rely on extensive pre-set Q&A pairs, generative AI reduces maintenance workload by shifting the focus from data quantity to data quality. From late 2023 to October 2025, Xiaoshu recorded over 70,000 interactions, serving about 3,000 users monthly with an accuracy rate above 80%, and reducing human librarian workloads by approximately 16 hours per month.

The project highlights the importance of defining clear service goals, user scenarios, and resource planning in the early stages. Overall, NLPI's experience shows that generative AI librarians can enhance service efficiency and create a new paradigm for human-AI collaboration in public libraries.

To cite this article:

Ma, H.-P., & Chen, Y.-F. (2025). Implementation and Application of a Generative AI Virtual Librarian: A Case Study of the National Library of Public Information, Taiwan. *International Journal of Librarianship*, 10(4), 137-148.
<https://doi.org/10.23974/ijol.2025.vol10.4.567>

To submit your article to this journal:

Go to <https://ojs.calaijol.org/index.php/ijol/about/submissions>

Implementation and Application of a Generative AI Virtual Librarian: A Case Study of the National Library of Public Information, Taiwan

Hsiang-Ping Ma, Yun-Fan Chen

National Library of Public Information, Taiwan

ABSTRACT

Since its transformation into Taiwan's first national digital library in 2013, the National Library of Public Information (NLPI) has continuously adopted innovative technologies to advance intelligent services. In response to the rise of large language models (LLMs), NLPI launched the "Generative AI Virtual Librarian" project in 2023 and developed Xiaoshu, a virtual librarian capable of natural voice interaction. Centered on generative AI, the system integrates four databases and eight intent categories, applying retrieval-augmented generation (RAG) and speech recognition to provide collection search assistance, book recommendations, service inquiry responses, and social interaction.

Xiaoshu effectively handles many repetitive yet diverse library inquiries, demonstrating the linguistic flexibility of generative AI. Compared with rule-based systems that rely on extensive pre-set Q&A pairs, generative AI reduces maintenance workload by shifting the focus from data quantity to data quality. From late 2023 to October 2025, Xiaoshu recorded over 70,000 interactions, serving about 3,000 users monthly with an accuracy rate above 80%, and reducing human librarian workloads by approximately 16 hours per month.

The project highlights the importance of defining clear service goals, user scenarios, and resource planning in the early stages. Overall, NLPI's experience shows that generative AI librarians can enhance service efficiency and create a new paradigm for human-AI collaboration in public libraries.

Keywords: artificial intelligence, LLMs, chatbot, public libraries, virtual librarian, generative

BACKGROUND

Since its transformation in 2013 into Taiwan's first national digital library, the National Library of Public Information (NLPI) has continuously adopted innovative technologies to advance

intelligent public library services. Over the years, NLPI has introduced several smart service systems, including self-service book pickup machines, 24-hour unmanned mini-libraries, automated sorting systems that streamline staff workflows, and a first-generation rule-based predictive AI chatbot capable of standardized and routine Q&A services. These initiatives demonstrate NLPI's ongoing efforts to promote intelligent library practices.

In reader consultation services, the first-generation predictive chatbot could only provide structured responses based on rules and keywords. Because the system required a large pre-defined question bank, librarians had to constantly expand it to cover users' diverse ways of expression. Even minor differences in wording could result in recognition errors, limiting both service efficiency and flexibility.

With the rapid development of Large Language Models (LLMs) in the field of generative artificial intelligence, the release of ChatGPT based on the Generative Pre-trained Transformer (GPT) framework in late 2022 brought revolutionary inspiration to library services (Zhao et al., 2023; Yang et al., 2024). LLMs, with their strong abilities in natural language understanding and generation, can process diverse and open-ended questions with fewer pre-set templates, producing semantically coherent responses. These capabilities allow users to enjoy more human-like, flexible, and highly interactive communications, breaking through the limitations of earlier AI systems.

Building on this vision, NLPI launched the Generative AI Virtual Librarian Project in 2023 to develop Xiaoshu, a virtual librarian capable of real-time conversation and natural voice interaction. The project aims to extend human service capacity and explore new possibilities for AI-assisted public library services, positioning the library as a pioneer in applying generative AI to enhance user experience and service innovation.

DESIGN AND IMPLEMENTATION

Development Process and Data Preparation

NLPI began developing its generative AI virtual librarian project in June 2023. Considering internal technical capacity, budget constraints, project timeline, service goals, and user scenarios, and aiming to leverage the advantages of generative AI to provide an innovative experience distinct from online predictive AI chatbots, the library planned to combine the most advanced GPT-based model with a physical installation in the library space. The system can detect approaching users through a camera, initiate voice greetings, and respond through gesture and facial expression animations, creating a highly immersive and interactive experience. This design aimed to encourage users to engage in book recommendations, collection searches, library inquiries, and casual conversations.

To ensure that the system accurately reflected users' real needs, the initial phase focused on integrating existing datasets, including archived Q&A logs from earlier predictive AI chatbot systems, librarian-curated service FAQs, and approximately 160,000 bibliographic records as the main knowledge base for Xiaoshu's responses (Hsieh et al., 2024).

The most significant early challenge was data cleaning and structuring. Because bibliographic records contain many metadata fields that can reduce retrieval efficiency, the development team redefined the data schema to retain only essential elements, such as bibliographic ID, title, author, edition, publication data, subject, classification number, book cover, and summary, while removing redundant information to improve processing speed. For records with brief or incomplete summaries, generative AI was applied to automatically expand the content and generate book tags, supporting future question answering and recommendation functions.

In addition, the team reviewed more than 10,000 Q&A records from the earlier predictive AI chatbot, eliminating duplicate and inaccurate items to compile a refined dataset of about 700 core questions for model training (Hsieh et al., 2024). To ensure that Xiaoshu could appropriately respond to diverse user intentions, the development team established eight intent categories covering library services and collections, persona-related questions about Xiaoshu (e.g., identity, age, and hobbies), real-time information (such as weather and transportation), public figures, political issues, ethical boundaries (including privacy, safety, defamation, misinformation, and legality), and open-ended casual conversation. This groundwork in data structuring, semantic classification, and intent design laid a solid foundation for Xiaoshu's ability to understand questions precisely and generate contextually appropriate responses.

System Architecture

Beyond technical capability, user experience (UX) was a key design consideration (Lew & Schumacher, 2020; Zheng et al., 2022). When developing the virtual librarian, both the interactive interface and overall system design were carefully planned to optimize the human-AI interaction experience. The NLP team designed Xiaoshu's system architecture around user-centered interaction, integrating front-end interaction with back-end knowledge generation to create a service framework that combines human warmth with technological depth (Hsieh et al., 2024). The overall framework comprises four main modules (see Figure 1):

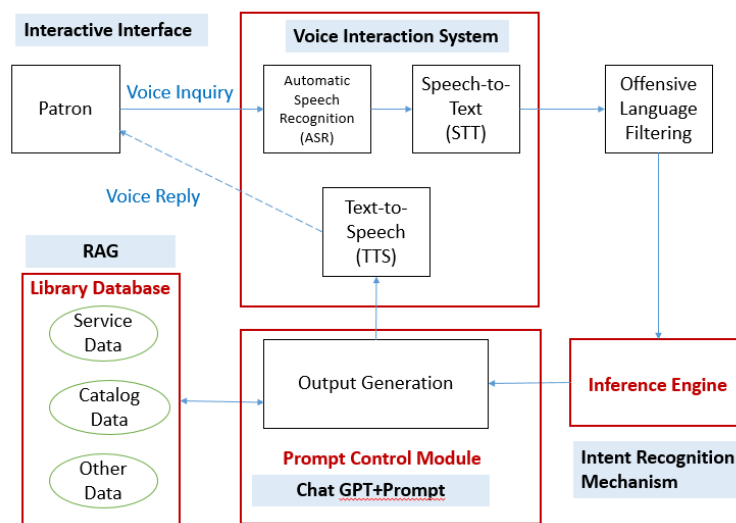


Figure 1. System Architecture of Xiaoshu

A. Front-End Interface and Virtual Character Design

To give the virtual librarian a more human-like and professional librarian image, she was named “Xiaoshu”(曉書, an abbreviation of 通曉圖書), meaning “knowledgeable about books and wise in consultation” in Chinese. Her overall appearance draws inspiration from the red British telephone booth, providing a semi-private and immersive interaction space. Inside, an LCD display shows both textual content and related information. Xiaoshu is equipped with a camera that detects users standing in front of the screen and automatically initiates a greeting. Users can interact with her through voice by pressing the red button on the right side (see Figures 2 and 3).



Figure 2. Appearance of Xiaoshu

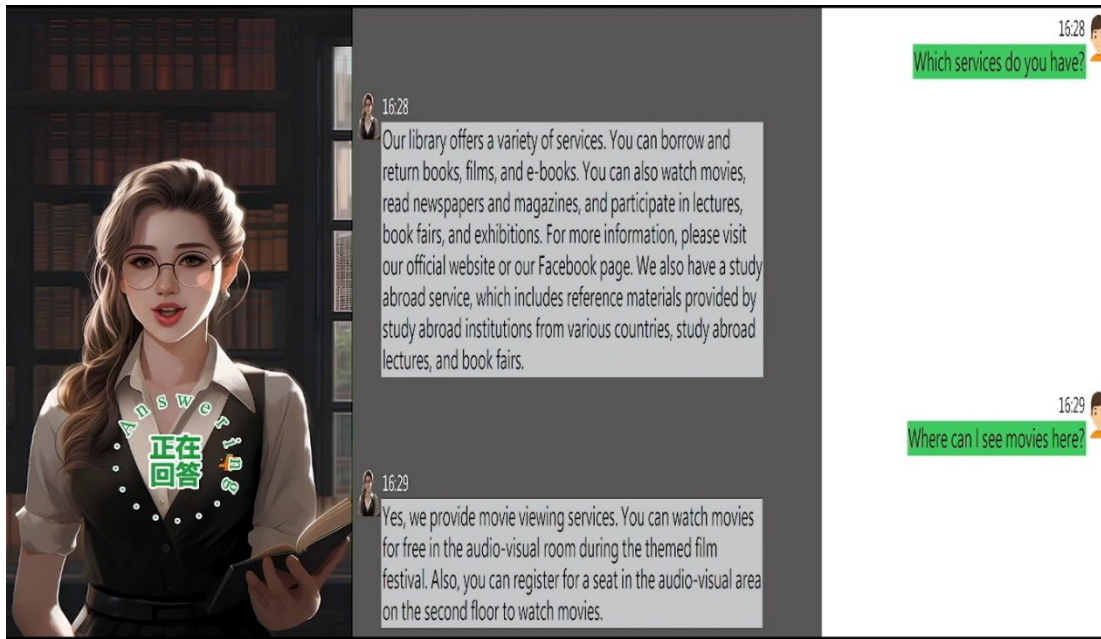


Figure 3. Interactive Interface of Xiaoshu (Left: Animation Display; Right: Touchscreen Panel)

B. Voice Interaction System

To enable real-time voice interaction in the library's physical environment, Xiaoshu integrates automatic speech recognition (ASR), speech-to-text (STT), and text-to-speech (TTS) technologies. This allows users to communicate naturally and receive instant voice responses, improving accessibility and friendliness for elderly users and children. For language support, the first year of deployment included Chinese and English STT and TTS services, which can automatically detect the input language and respond in the same one. If a question mixes both Chinese and English, Xiaoshu identifies the dominant language and replies accordingly. Starting in the second year, NLPI began testing Taiwanese Hokkien STT and TTS modules developed by local vendors to serve a broader linguistic community.

C. Intent Recognition Mechanism

The intent recognition mechanism is one of the core components of the system, enabling automatic switching among four response modes based on user context. When a question requires answers from specific databases defined by NLPI, the system activates the Retrieval-Augmented Generation (RAG) process to retrieve and generate precise responses (Lewis et al., 2020). If the question involves real-time information, Xiaoshu accesses external open data sources such as weather or transportation websites. For general conversations, the system switches to the "social dialogue" mode, a distinctive feature of generative AI, that allows Xiaoshu to interact naturally and build user engagement and trust. When encountering sensitive, political, or unethical content, a safety filter ensures neutral and friendly replies. Through this layered intent-recognition

structure, Xiaoshu maintains both human-like interactivity and informational accuracy, with an average response time of under 20 seconds.

D. Knowledge Retrieval and Generation Engine

As mentioned earlier, to help Xiaoshu accurately answer specifically defined questions, NLPI established four main databases based on common user intents: library services, books, Xiaoshu's persona settings, and key figures. These are combined with the RAG framework and the ChatGPT language model to perform semantic analysis and deliver flexible Q&A responses. Particularly in bibliographic queries, this design marks the most significant distinction between generative AI assistants in general contexts and those in library environments. NLPI holds over 1.3 million items, but since database scale can affect system speed, approximately 550,000 frequently accessed physical book records were selected as the main dataset. This optimization enables Xiaoshu to deliver quicker responses and broader knowledge coverage than human reference librarians, enhancing the overall quality of reference services.

Challenges and Solutions

Although large language models (LLMs) demonstrate powerful language understanding and generation capabilities, they still face issues such as hallucinations, bias, and computational cost (Kaddour et al., 2023). During the development of Xiaoshu, NLPI not only had to evaluate the technical feasibility of implementation but also encountered multiple challenges in system calibration and adjustment.

First, regarding the reliability and professionalism of generated content, beyond the previously mentioned data cleaning and structuring efforts, NLPI needed to address how to prevent AI from producing biased or inappropriate responses. To achieve this, an intention recognition mechanism was introduced to identify user query types, reducing the likelihood that Xiaoshu would handle professional questions through casual conversation. Additionally, prompt engineering was applied to define language behavior and role boundaries, ensuring that responses align with public service ethics and the professional standards expected of librarians.

Second, a balance had to be maintained between response completeness and computational efficiency. To prevent answers from becoming too lengthy or repetitive, which could cause user fatigue and increase token costs unnecessarily, the system limited both the length of textual content and the duration of spoken replies. Additionally, as generative AI models evolve rapidly, NLPI needed to ensure that the deployed model remained current and compatible with future updates.

However, system adjusting introduced new challenges. The stricter the role boundaries, the less natural and engaging Xiaoshu's conversational tone became during casual interactions. Conversely, limiting response flexibility sometimes caused Xiaoshu to produce incomplete or mismatched answers when encountering atypical questions beyond the database scope. To address these issues and enhance response quality, NLPI regularly reviews user interaction logs, updates missing information, and refines datasets. Continuous retraining and adjustment are also required following each LLM upgrade to ensure that Xiaoshu consistently provides accurate, safe, and high-quality interactive reference services.

SERVICE EXPERIENCE AND USER INTERACTION ANALYSIS

Xiaoshu currently provides four main service functions: book recommendations, collection searches, library service inquiries, and social conversations. Practical observations show that, much like interactions with human librarians, social dialogue serves not only as an entry point for user engagement but also as a key factor in building trust and sustained interaction. Therefore, NLPI intentionally retained ChatGPT's conversational capabilities, allowing patrons to become familiar with Xiaoshu through relaxed, friendly exchanges that encourage continued use.

From December 1, 2023 (including the pre-launch testing period prior to the official debut on December 28) to October 30, 2025, data from the system log show that Xiaoshu has facilitated approximately 70,000 interactions. Assuming 26 working days per month and one inquiry per user, this translates to an average of about 117 users served per day.

After the first three months of operation, weekend usage began to surpass weekday usage, averaging about 3,000 interactions per month (Figure 4). The personality design of Xiaoshu, including her self-introduction, tone modulation, and conversational style, gives her a friendly and relatable demeanor, similar to that of a real librarian, making her particularly popular among young readers and senior citizens. This pattern indicates that the AI librarian not only attracts patrons but also fosters curiosity and motivation for learning about libraries and AI technologies.

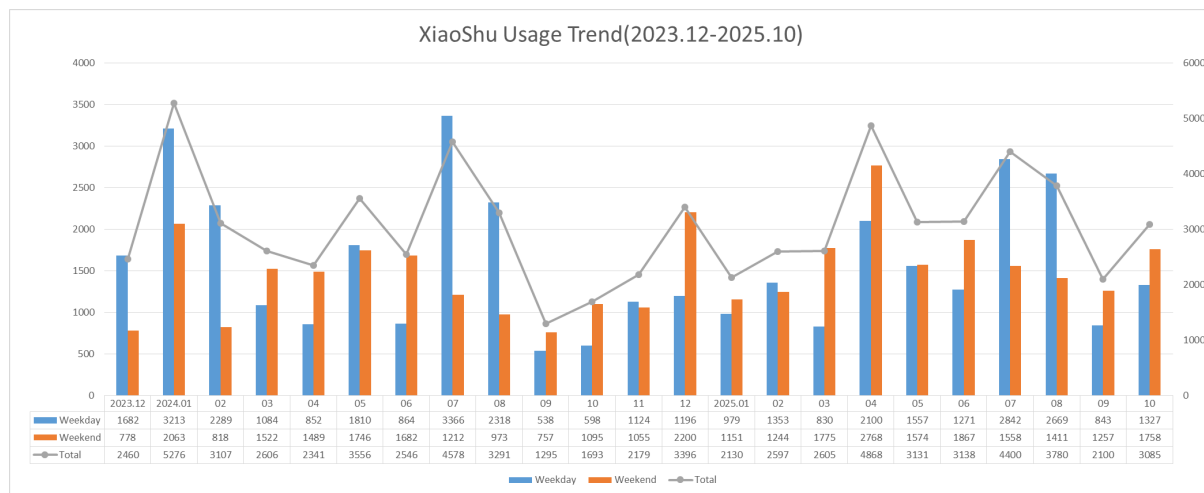


Figure 4. Usage Trend of Xiaoshu (Dec. 2023–Oct. 2025)

Regarding overall query distribution, Figure 5 shows that 31.87% of all interactions involved core library consultations (books and library services), 21.74% related to Xiaoshu's persona, and 37.2% were social interactions. As illustrated in Figure 6, the proportion of core library consultations remained stable month to month and reached a new peak in September-October 2025, excluding the December 2023 testing period. Meanwhile, total interactions increased from

around 2,000 to nearly 5,000 per month, indicating that as user volume grew, so did the number of inquiries involving core library consultations. This trend shows that Xiaoshu initially attracted users through its novelty and conversational charm but gradually earned their trust as a reliable reference tool for addressing real library needs.

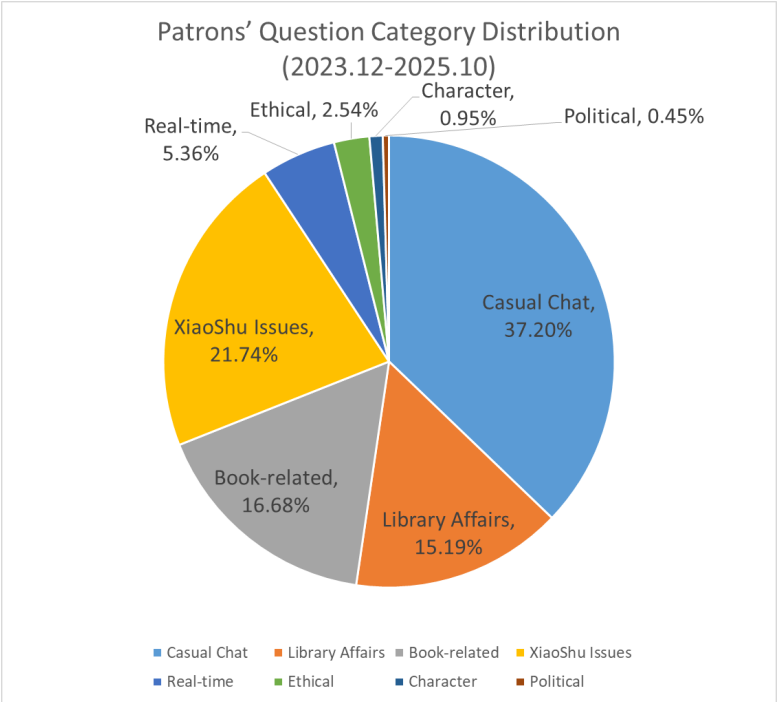


Figure 5. Overall Distribution of User Inquiry Categories with Xiaoshu

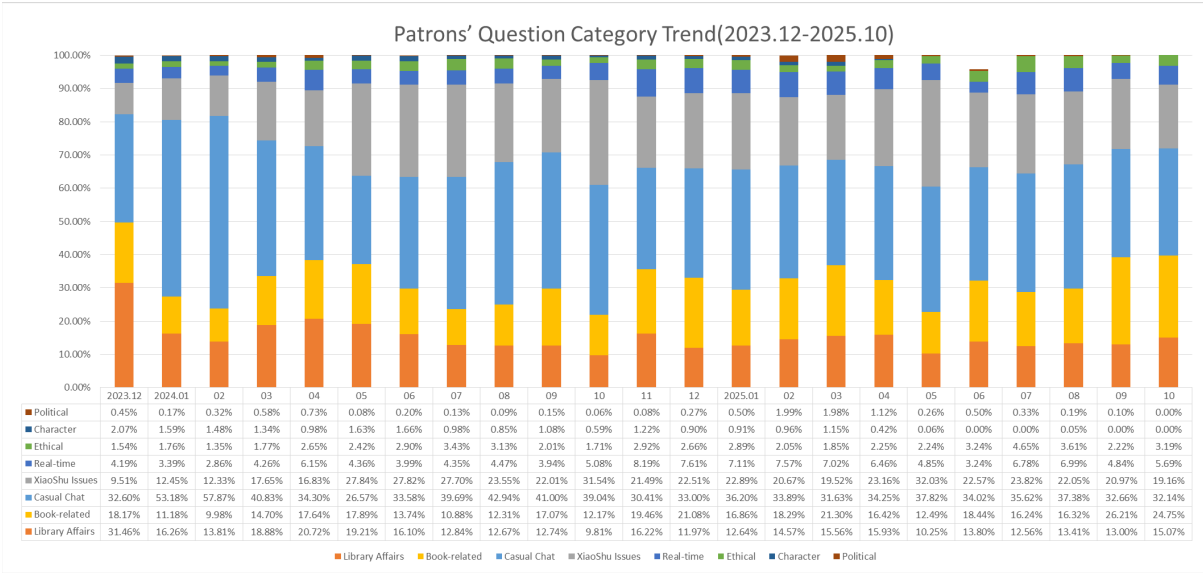


Figure 6. Monthly Trends in User Inquiry Categories

Further analysis reveals that while book searches and book recommendations vary widely depending on individual needs, most library service inquiries consist of repetitive and common questions (e.g., loan rules, card registration, and facility locations). However, patrons often phrase these questions in diverse and unstructured ways. This highlights the advantage of generative AI librarians built upon LLMs, which possess semantic understanding and contextual reasoning abilities, over predictive AI chatbots that rely solely on rigid, predefined keyword matching. With its language generalization capacity, Xiaoshu can interpret various expressions of the same intent, offering natural and accurate responses that compensate for the linguistic rigidity of earlier systems.

In addition, a 2024 user interview study conducted by a graduate researcher from the Department of Library and Information Science at National Chung Hsing University (Liu, 2024), found that most participants agreed that generative AI-based assistants are “more flexible and natural” compared to traditional predictive systems, expressing high satisfaction with Xiaoshu’s response quality and overall interaction experience. Beyond these qualitative findings, NLPI plans to roll out a real-time feedback mechanism, similar to the “thumbs up/down” feature in ChatGPT, by the end of 2025. This feature will gather direct user reactions to both positive and negative responses, providing valuable data for system improvement, error correction, and ongoing evaluation.

PRACTICAL PERFORMANCE EVALUATION

NLPI developed its generative AI virtual librarian to enhance service efficiency and provide patrons with diverse technological experiences. Based on this objective, the following section evaluates both the resources invested and the actual service outcomes, analyzing whether Xiaoshu effectively reduces the workload of human librarians while maintaining reliable and accurate responses.

Resources Invested in Xiaoshu

A common misconception about generative AI systems is that they can self-train, thereby minimizing maintenance efforts. In reality, while generative AI possesses broad linguistic capabilities, transforming it into a knowledgeable and reliable virtual librarian still requires continuous input of localized data and professional expertise. This ensures the accuracy and consistency of the Retrieval-Augmented Generation (RAG) process.

To establish a robust data foundation, NLPI initiated data cleaning and integration in June 2023, a process that lasted approximately six months and involved cross-departmental collaboration to review and test the quality of interactions. After system deployment, reference librarians assumed responsibility for maintaining and refining the databases, while IT personnel provided technical support. Frontline librarians were tasked with monitoring the kiosk’s operational status to promptly report and resolve any malfunctions.

The database maintenance librarians play a crucial role. They regularly review and update interdepartmental data to ensure accuracy and consistency, while also participating in ongoing AI literacy training to understand how emerging technologies impact library services. Based on

operational experience, maintaining Xiaoshu requires approximately 3-4 hours per month to manually review thousands of user interactions, correct misclassifications or incomplete responses, and collaborate with vendors to optimize database quality.

Xiaoshu's Response Accuracy

According to evaluations by librarians responsible for maintaining Xiaoshu, it maintains an average monthly accuracy rate of over 80%. Accuracy depends primarily on two factors: input quality (such as clear voice capture, correct language detection, and accurate speech-to-text conversion) and completeness of backend data (availability of relevant answers retrievable through RAG). While backend accuracy can be enhanced through continuous database updates, input quality remains limited by users' speaking habits and the current capabilities of generative AI speech models, such as language support restricted to Mandarin and English, without Taiwanese Hokkien, and the inherent accuracy rate of these commercial LLM-based systems (e.g., ChatGPT-4o, ChatGPT-5). Future advances in LLM technology are likely to significantly improve semantic understanding.

Xiaoshu's Service Effectiveness

Currently, NLPI operates a single Xiaoshu unit, located in the library's first-floor lobby, providing real-time voice interaction services. With an average of 3,000 monthly interactions, of which approximately 30% are professional consultations, Xiaoshu handles around 1,000 core library consultations per month. Assuming each response takes approximately one minute, this equates to roughly 16.6 hours of human librarian time saved per month.

By providing continuous on-site assistance, Xiaoshu improves the library's service accessibility, reduces staff workload during peak hours, and effectively handles repetitive inquiries and precise book recommendations. This allows human librarians to focus on tasks that require technical skills and human judgment, such as exhibition curation, reading promotion, information literacy education, and professional advising. Therefore, expanding the number or variety of Xiaoshu installations across different floors could further lessen consultation workload for librarians and enhance the library's overall service capacity.

LESSONS LEARNED AND FUTURE DIRECTIONS

The experience of NLPI demonstrates that generative AI is not a threat to library services but a valuable tool for addressing staff shortages and service limitations. It helps reduce repetitive or data-intensive tasks, allowing human librarians to focus on services that require professional judgment or human empathy.

For libraries planning to implement a generative AI-powered virtual librarian, it is crucial to conduct a thorough assessment and develop a clear implementation plan early on. The preparatory phase should explicitly define the service goals and user scenarios, such as database maintenance mechanisms, whether data management will be centralized within a single department or involve cross-unit collaboration, the design of front-end and back-end data integration, and methods for collecting and analyzing usage logs to support future improvements and avoid additional

operational burdens. Additionally, considerations like database types, update frequency, cybersecurity measures, and debugging mechanisms are critical for ensuring long-term stability of the system.

In a rapidly evolving technological environment, AI platforms and models continue to advance rapidly, with providers such as OpenAI, Microsoft Azure, and Google Gemini engaging in ongoing competition and innovation. Libraries must choose whether to develop their own models or adopt existing commercial technologies. Building their own allows for greater customization but requires significant computing power and staffing costs. Adopting existing solutions allows faster deployment but needs localization to match institutional needs and data structures. The key is balancing the cost of self-development with the efficiency of adopting mature technologies to ensure long-term sustainability. Because public library services differ greatly from commercial service systems, there is no one-size-fits-all model to replicate. Each institution must take innovation risks and invest time in working with vendors.

Looking ahead, NLPI will continue to enhance Xiaoshu's capabilities by expanding multilingual and localized speech modules, including Taiwanese Hokkien and Hakka, to serve a broader range of user groups. The system will also be upgraded with real-time events and diverse collection promotion features, enabling Xiaoshu to proactively share relevant information based on user needs. In addition, efforts will focus on improving Chinese speech recognition and semantic understanding to enhance accuracy in processing homophones and domain-specific terminology.

Ultimately, the goal is for Xiaoshu to become a new model of human-AI collaboration in public libraries, promoting reading across the country and spreading knowledge. As NLPI continues to follow AI developments, it will adaptively incorporate advanced technologies into various library services, going beyond Xiaoshu, to further expand the role of AI in public access to information and lifelong learning.

References

- Hsieh, H.-Y., Ko, C.-C., Yu, L.-K., Chen, T.-E., Jwo, J.-S., Li, Y.-C., Chang, H.-C., & Ma, H.-P. (2024). 智慧圖書館：基於 GPT-4 的智慧館員 [Smart Library: Intelligent Librarian Powered by GPT-4]. *公共圖書館研究*, 19, 10-14.
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and applications of large language models*. *arXiv preprint arXiv:2307.10169*. <https://arxiv.org/abs/2307.10169>
- Lew, G., & Schumacher, R. M. (2020). *AI and UX: Why artificial intelligence needs user experience*. Apress.

- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Liu, Z.-E. (2024). 聊天機器人之讀者體驗研究：以國立公共資訊圖書館為例 [*A study of reader experience with chatbots: A case study of the National Library of Public Information*]. In 2024 圖書資訊學術與實務研討會會議論文集 (pp. 199-206). Library Association of the Republic of China (Taiwan).
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., & Hu, X. (2024). Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. *ACM Transactions on Knowledge Discovery from Data*, 18(6), 1–32. <https://doi.org/10.1145/3649506>
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., & Wen, J. R. (2023). *A survey of large language models*. *arXiv preprint arXiv:2303.18223*. <https://arxiv.org/abs/2303.18223>
- Zheng, Q., Tang, Y., Liu, Y., Liu, W., & Huang, Y. (2022, April). UX research on conversational human-AI interaction: A literature review of the ACM Digital Library. In S. Barbosa & C. Lampe (Eds.), *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-24). Association for Computing Machinery. <https://doi.org/10.1145/3491102.3501855>

About the authors

Hsiang-Ping Ma [Email: sophiama@nlpi.edu.tw]

Director General of the National Library of Public Information (Taiwan) and holds a Ph.D. in Education from National Taiwan Normal University. With extensive experience in educational administration and international collaboration, she has been devoted to advancing smart transformation and nationwide reading initiatives. Under her leadership, the library fosters innovation, accessibility, and lifelong learning through technology and knowledge, creating an inclusive and intelligent learning environment for all.

Yun-Fan Chen [Email: a24021@nlpi.edu.tw]

A librarian at the National Library of Public Information (Taiwan) and holds a Master's degree in Public Administration from National Chengchi University. She previously worked in several public agencies, including the Coast Guard Administration, the Health Bureau, and the Tourism and Information Bureau. Her professional interests include public policy and administration, civic engagement, and lifelong learning. She is responsible for AI librarian maintenance, youth reading promotion, and sustainability education projects.