Human or AI Powerful? A Study on Improving Call Center Performance with Generative AI - The Case of Taoyuan City Government

Chaoming Wu*
Department of Information Management, Chung Yuan Christian University, Taiwan
*Corresponding Author: mislighter@gmail.com

Received 7 April 2025; received in revised form 23 September 2025; accepted 5 December 2025

Abstract

This research uses the "Taoyuan 1999 Citizen Hotline Generative AI Human-Machine Collaboration System" as a case study to explore how generative AI can be applied to actual citizen consultation services, and to analyze the difficulties encountered and effectiveness achieved during its implementation. Drawing upon the Technology Acceptance Model (TAM) and Human-Computer Interaction theory, this study provides theoretical insights into the adoption and effectiveness of generative AI in public service delivery. This research aims to provide insights into the potential and challenges of generative AI in improving customer service center performance through this empirical case, hoping to offer references for future applications of related technologies.

Keywords: Generative AI, call center, customer service, technology acceptance, human- computer interaction

1. Introduction

With the rapid development of technology, artificial intelligence has gradually penetrated various industries. Call centers, as important bridges for communication between organizations and customers, face challenges in improving efficiency and service quality. Traditional customer service center operation models, such as the Taoyuan City Government's 1999 Citizen Hotline, rely on service personnel to consult existing knowledge databases to respond to public inquiries. However, when facing increasingly complex and diverse citizen issues, service personnel often need to spend a significant amount of time searching across different datasets, resulting in prolonged service times and increased waiting times for citizens. To address these problems and enhance citizen service capacity, the application of generative artificial intelligence (Generative AI) is considered a potential solution.

Generative AI, as a subset of AI, can generate new content based on existing data and user input (prompts), including text, images, audio, video, or code. Compared to traditional AI operating under predefined rules, generative AI utilizes advanced algorithms, especially in the fields of natural language processing (NLP) and deep learning, to create responses and interactions that mimic human behavior (Tris, 2024). This technology demonstrates potential for optimizing interactions and service delivery in the customer service center domain, ultimately improving customer satisfaction and operational efficiency (Shulzhenko, 2025).

This research uses the "Taoyuan 1999 Citizen Hotline Generative AI Human-Machine Collaboration System," developed through cooperation between the Taoyuan City Government Research and Evaluation Commission and private technical units, as a case study to explore how generative AI can be applied to actual citizen consultation services, and to analyze the difficulties encountered and effectiveness achieved during its implementation. This research aims to provide insights into the potential and challenges of generative AI in improving customer service center performance through this empirical case, hoping to offer references for future applications of related technologies.

2. Literature Review

2.1 Application of Generative AI in Call Centers

Generative AI is revolutionizing the landscape of call centers by enhancing productivity and significantly reducing operational costs through automated and efficient service delivery (Chui et al., 2023). Research in information systems has shown that AI-powered customer service systems can improve overall customer satisfaction by reducing waiting times and providing round-the-clock service (Huang & Rust, 2021). These systems also enable personalized interactions at scale, thereby enhancing the quality of interactions between customer service personnel and customers (Davenport & Ronanki, 2018).

Key components of generative AI in call centers include natural language processing (NLP) and machine learning algorithms (Zhang et al., 2022). NLP serves as the foundation, helping machines understand and interpret human language, enabling systems to accurately respond to customer inquiries, whether in text or voice format. Through complex algorithms, NLP can identify the intent and context of queries, allowing chatbots and voice recognition systems to quickly provide relevant responses. Machine learning, as a subset of AI, enables generative AI to learn from data and enhance its predictive capabilities (Jordan & Mitchell, 2015). By analyzing past interactions, these algorithms can predict customer needs and optimize routing to the most suitable customer service personnel, thereby increasing the efficiency of customer service operations.

2.2 Benefits and Challenges of Generative AI Implementation

The literature reveals both significant benefits and considerable challenges in implementing generative AI systems in organizational contexts. Research by Brynjolfsson et al. (2023) demonstrates that generative AI can increase worker productivity by 14% in customer service roles, particularly for less experienced workers. However, successful implementation requires careful attention to organizational factors, technology infrastructure, and human factors.

Key challenges identified in the literature include:

- Technology Integration Complexity: Organizations face significant challenges in integrating AI systems with existing infrastructure (Fountaine et al., 2019). This integration requires not only technical compatibility but also organizational readiness and change management.
- Skills and Training Requirements: The effective use of generative AI requires new skills and competencies from workers (Wilson et al., 2017). Organizations must invest in comprehensive training programs to ensure successful adoption.
- Performance Variability: Research indicates that the effectiveness of AI systems can vary significantly based on task complexity, user experience, and contextual factors (Raisch & Krakowski, 2021). This variability necessitates careful evaluation and customization of AI implementations.

2.3 TAM and HCI Theory

This study is grounded in two complementary theoretical frameworks that provide comprehensive lens for understanding the implementation and effectiveness of generative AI in public service delivery. One is the Technology Acceptance Model (TAM), another is Human-Computer Interaction (HCI) Theory.

The Technology Acceptance Model, developed by Davis (1989), provides a theoretical framework for understanding how users come to accept and use technology. TAM posits that perceived usefulness and perceived ease of use are primary determinants of technology acceptance intentions (Venkatesh & Davis, 2000). In the context of our study, TAM helps explain why some call center agents demonstrated higher acceptance and effectiveness with the generative AI system, while others showed initial resistance or adaptation challenges.

According to TAM, perceived usefulness refers to "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989). Our findings align with this construct, as senior agents who perceived the AI system as useful for improving their response efficiency. Conversely, perceived ease of use, defined as "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989), may explain why junior staff initially experienced slight performance decreases, suggesting a learning curve that affects ease of use perceptions.

Human-Computer Interaction theory provides insights into how humans and computers work together to accomplish tasks effectively. Norman (2013) emphasizes that successful human- computer systems require careful consideration of human capabilities, limitations, and cognitive processes. The concept of "human-in-the-loop" systems, where humans and AI collaborate rather than compete, is particularly relevant to our study (Holzinger, 2016).

Our research demonstrates the importance of designing AI systems that augment rather than replace human capabilities. The varying effectiveness across different issue types can be understood through HCI principles that suggest different cognitive demands require different levels of human-AI collaboration (Shneiderman, 2020).

3. Research Methods

3.1 Research Design and Approach

This study adopts a single-case study methodology to investigate the implementation and effectiveness of generative AI in public sector call center operations. The case study approach is particularly appropriate for exploring contemporary phenomena within their real-life context, especially when the boundaries between the phenomenon and its context are not clearly delineated (Yin, 2018). This methodology enables an in-depth examination of the complex implementation process, organizational challenges, and performance outcomes associated with the deployment of the generative AI system within the specific operational environment of the Taoyuan City Government's 1999 Citizen Hotline.

The research addresses three primary questions that guide the investigation. First, the study examines how the implementation of generative AI technology affects overall call center performance, specifically focusing on measurable outcomes such as call duration and knowledge search efficiency. Second, the research explores the factors that contribute to differential effectiveness of generative AI across various types of citizen inquiries and among users with different levels of professional experience. Third, the study identifies and analyzes the key challenges encountered during implementation as well as the critical success factors that facilitate effective deployment of generative AI in public sector service delivery contexts.

3.2 Case Selection and Context

The selection of the "Taoyuan 1999 Citizen Hotline Generative AI Human-Machine Collaboration System" as the research case was based on several strategic considerations that ensure the relevance and validity of the findings. The operational scale of the system provides a robust foundation for analysis, as it processes over 19,700 calls per month, generating sufficient data volume to support meaningful statistical analysis and pattern identification. The system's implementation maturity was another crucial factor, as it had been operational for an adequate period to allow for comprehensive evaluation of both immediate and emerging effects on organizational performance.

Data accessibility represented a significant advantage in case selection, as the Taoyuan City Government provided comprehensive access to operational data, system logs, and personnel for research purposes. This level of organizational cooperation is essential for conducting rigorous case study research that requires multiple data sources and perspectives. Furthermore, the representativeness of the case enhances its theoretical and practical value, as municipal government services face typical challenges encountered by public sector organizations in implementing advanced AI technologies, including bureaucratic constraints, public accountability requirements, and diverse stakeholder expectations.

3.3 Data Collection

The data collection strategy employed a multi-method approach that combines quantitative performance measurement with qualitative insights to provide a comprehensive understanding of the system's implementation and impact. This triangulation of data sources enhances the validity and reliability of the findings while allowing for both statistical analysis of performance outcomes and nuanced exploration of user experiences and organizational dynamics.

Primary quantitative data collection focused on objective performance metrics gathered through systematic observation over a twenty-day period from October 1-20, 2023. The study involved four call center agents selected to represent different experience levels, specifically two senior agents with over seven years of experience and two junior agents with less than two years of experience. This stratified sampling approach enables analysis of how professional experience influences the adoption and effectiveness of AI-augmented service delivery. Performance metrics captured included call duration measured from initiation to completion, knowledge search time representing the duration spent locating relevant information, and issue resolution rates indicating the quality of service delivery outcomes.

Data collection procedures combined automated system logging with manual timing records to ensure accuracy and completeness. The automated systems captured detailed interaction logs, response times, and system usage patterns, while manual recording provided validation and context for automated measurements. Over the observation period, approximately 536 knowledge queries were processed daily, providing a substantial dataset for statistical analysis.

The research examined four primary service categories selected based on their call volume and varying complexity levels. Bus services encompassed route information requests, schedule inquiries, and service disruption notifications. YouBike services included rental procedures, station locations, and technical issue resolution. Citizen card services covered application processes, usage guidelines, and technical support. Social welfare services addressed benefit eligibility determinations, application procedures, and policy information dissemination. This categorization allows for analysis of how system effectiveness varies across different types of citizen service requests.

Qualitative data collection involved semi-structured interviews with eight participants representing diverse perspectives within the organization. Interview participants included call center agents with varying experience levels, supervisory personnel responsible for operational management, and IT implementation staff involved in system deployment and maintenance. Each interview lasted 45-60 minutes and was conducted during October 2023 to capture immediate experiences and perceptions following system implementation. Interview topics covered user experience with the AI system, perceived usability and effectiveness, training needs and adequacy, implementation challenges, and both benefits and limitations observed in daily operations.

Direct observation supplemented interview data by providing real-time insights into human-AI interaction patterns and workflow changes. Non-participant observation was conducted over 40 hours distributed across the twenty-day study period, allowing researchers to observe natural work behaviors without influencing operational activities. Observation focused on documenting how agents interacted with the AI system, changes in problem-solving approaches, and the integration of AI-generated suggestions into customer service workflows.

Secondary data sources provided historical context and baseline measurements essential for evaluating system impact. Historical call center performance data from January through September 2023 established pre-implementation baselines for comparison purposes. System usage statistics and error logs offered technical insights into system reliability and utilization patterns. Knowledge base utilization patterns revealed changes in information-seeking behaviors following AI implementation. Documentation review encompassed system implementation records, training materials and user manuals, and policy documents that guided the deployment process.

3.4 Data Analysis Approach

The data analysis strategy integrated quantitative statistical methods with qualitative analytical techniques to provide comprehensive insights into system effectiveness and implementation dynamics. Quantitative analysis focused on performance comparison through before-and-after statistical evaluation of key metrics. Segmentation analysis examined performance variations across user experience levels and service categories, revealing differential impacts of AI implementation.

Efficiency metrics calculation involved systematic computation of average call duration reductions, knowledge search time improvements, and daily time savings projections based on observed call volumes. These calculations provide concrete measures of operational improvement that support evidence-based evaluation of the AI system's value proposition.

Qualitative analysis employed thematic analysis techniques to identify patterns and insights from interview transcripts and observation notes. Open coding procedures systematically analyzed textual data to identify recurring themes, concepts, and relationships. Theme development involved iterative refinement of identified patterns, focusing particularly on factors related to user acceptance, implementation challenges, and success factors that influenced system effectiveness.

Content analysis of system interaction logs provided additional insights into usage patterns and behavioral changes following AI implementation. This analysis examined how knowledge base utilization changed, which types of queries benefited most from AI assistance, and how user interaction patterns evolved over time.

3.5 Taoyuan 1999 Generative AI Human-Machine Collaboration System

The technical infrastructure underlying the Taoyuan 1999 Generative AI Human-Machine Collaboration System represents a sophisticated integration of cloud computing services, artificial intelligence models, and knowledge management technologies. The system architecture was designed to seamlessly integrate with existing call center operations while providing advanced AI capabilities for information retrieval and response generation.

The foundational infrastructure utilizes Microsoft Azure Cloud Services as the primary computing platform, leveraging Azure Virtual Machines configured with 16 virtual CPUs and 64 gigabytes of RAM to ensure adequate processing capacity for real-time AI operations. Storage requirements are addressed through Azure Blob Storage systems that maintain both the original knowledge base content and the vector database representations required for semantic search functionality. Network infrastructure includes dedicated bandwidth allocation of 1 gigabit per second to support responsive AI processing and minimize latency in user interactions.

The generative AI components represent the core technological innovation of the system. The primary large language model employed is GPT-3.5 Turbo, specifically version gpt-3.5-turbo- 0613, which provides natural language understanding and response generation capabilities. Text embedding functionality utilizes Microsoft Azure OpenAI's text-embedding-ada-002 model to convert textual content into high-dimensional vector representations that enable semantic similarity calculations. Azure Cognitive Search serves as the vector database platform, supporting rapid semantic similarity matching for knowledge retrieval operations. Real-time response generation is achieved through API calls to Azure OpenAI services, maintaining average response times of less than two seconds to ensure seamless user experience.

The knowledge management system architecture supports both traditional structured data storage and modern vector-based semantic search capabilities. Azure SQL Database maintains structured knowledge entries in their original format, while Azure Cognitive Search hosts vector index representations using 1,536-dimensional embeddings generated through the text-embedding-ada-002 model. A custom web interface enables knowledge base administrators to maintain content, upload new documents, and monitor system performance. The current data volume encompasses 8,173 knowledge base entries, all of which have been converted to vector representations to support AI-powered retrieval operations.

Integration architecture ensures seamless connectivity between the AI system and existing call center infrastructure. RESTful API connections link the legacy call center system with new AI services, enabling real-time data exchange without requiring complete system replacement. The user interface adopts a Microsoft Copilot-style design paradigm, integrating AI assistance directly into existing agent workstations to minimize workflow disruption. Security measures include Azure Active Directory authentication and role-based access control to protect sensitive citizen information and ensure appropriate system access. System monitoring utilizes Azure Application Insights to track performance metrics, identify technical issues, and support continuous optimization efforts.

The data processing pipeline orchestrates the complex sequence of operations required to transform citizen inquiries into AI-generated response suggestions. Knowledge ingestion processes automatically convert documents and question-answer pairs into system-compatible formats. Vector conversion utilizes batch processing with the text-embedding-ada-002 model to generate semantic representations of all knowledge content. During operation, similarity search algorithms employ cosine similarity calculations to identify the most relevant knowledge entries for each inquiry. Response generation integrates retrieved context with the GPT-3.5 Turbo model to produce coherent, contextually appropriate suggestions for agent use. Quality assurance procedures include human validation of AI-generated responses before deployment, ensuring accuracy and appropriateness of system outputs.

4. Research Results

4.1 Overall Effectiveness Analysis

Research results show that after using the human-machine collaboration system, the overall average call time was reduced by 20.75 seconds, saving 10.21% of time. Additionally, the average knowledge search time per instance was reduced from 60 seconds to 20 seconds, representing a 66.7% improvement in search efficiency. Calculating with an average of 536 knowledge queries per day, approximately 6 hours of search time can be saved daily. This indicates that the generative AI human-machine collaboration system has significant effectiveness in improving overall service efficiency.

4.2 Effectiveness Analysis of Different Issues

Looking at the effectiveness of different issues, the average call times for YouBike and citizen card issues were both reduced after using the system (YouBike: 12.5% reduction, Citizen Cards: 8.3% reduction), while the average call time for bus issues remained statistically unchanged (0.2% increase, not significant), and the average call time for social welfare issues actually increased by 7.8%.

From a TAM perspective, this variation can be attributed to differences in perceived usefulness across issue types. Structured services like YouBike and citizen cards have standardized procedures that align well with AI capabilities, leading to higher perceived usefulness. Social welfare issues, being more complex and requiring human judgment, may have lower perceived usefulness for AI assistance, consistent with HCI theory's emphasis on matching technology capabilities to task requirements.

4.3 Effectiveness Analysis of Personnel with Different Years of Experience

The effectiveness analysis for personnel with different years of experience shows that after using the human-machine collaboration system, senior personnel's average call time decreased by 80 seconds, saving 35.25% of time ($p < 0.01$). However, junior personnel's average call time increased by 3 seconds, an increase of 1.3% (not statistically significant).

These findings align with TAM predictions regarding the relationship between experience and technology acceptance. Senior staff's extensive domain knowledge enables them to quickly assess AI-generated suggestions and integrate them effectively into their workflow, leading to high perceived usefulness. Junior staff may experience cognitive overload when processing both the customer inquiry and AI suggestions simultaneously, consistent with HCI principles of cognitive load management.

4.4 User Experience

The qualitative analysis of user experience data, derived from semi-structured interviews and direct observation, reveals a complex landscape of perceptions and interactions with the generative AI system. The findings illuminate both the transformative potential and practical challenges encountered during system adoption, providing crucial insights into the human factors that influence implementation success.

The interview data analysis identified several prominent themes that characterize user experience with the AI-enhanced call center system. Efficiency enhancement emerged as the most consistently reported benefit, with 87.5% of interviewed agents acknowledging improved performance in information retrieval processes. These agents described how the AI system's ability to rapidly process natural language queries and generate contextually relevant responses significantly reduced the cognitive burden associated with traditional knowledge base searches. The semantic understanding capabilities enabled agents to pose questions in conversational language rather than relying on specific keyword combinations, creating a more intuitive and flexible information-seeking experience.

Confidence building represented another significant positive theme, particularly among senior agents who expressed increased assurance in handling complex citizen inquiries. These experienced agents reported that AI-generated suggestions often provided comprehensive coverage of relevant policy areas and procedural requirements that they might not have initially considered. The system's cross-knowledge content understanding capabilities proved especially valuable in situations where citizen inquiries spanned multiple municipal departments or service categories, enabling agents to provide more holistic and accurate responses.

Interestingly, junior agents demonstrated a different but equally important relationship with the AI system, viewing it primarily as a learning support mechanism. Despite experiencing initial performance decreases, these less experienced agents appreciated how AI suggestions exposed them to proper terminology, comprehensive response structures, and connections between different service areas. Several junior agents described the AI system as functioning like an experienced mentor, providing guidance and examples that accelerated their professional development and domain knowledge acquisition.

However, the analysis also revealed several challenges that tempered initial enthusiasm and influenced adoption patterns. Trust issues emerged as a significant concern, with 50% of agents initially questioning the accuracy and reliability of AI-generated responses. These concerns manifested in behaviors such as extensive verification of AI suggestions against traditional knowledge sources and hesitation to rely on AI guidance for complex or sensitive inquiries. The trust-building process appeared to vary considerably based on agent experience levels, with senior agents developing confidence more rapidly than their junior counterparts.

Training needs represented a universal concern across all participant groups, with every interviewed agent requesting additional instruction on optimal AI interaction strategies. Agents expressed particular interest in learning how to formulate effective queries, interpret AI response confidence levels, and integrate AI suggestions with their existing knowledge and judgment. The desire for training extended beyond technical system operation to encompass broader questions about when to rely on AI assistance, how to validate AI suggestions, and strategies for managing situations where AI recommendations conflicted with their professional intuition.

System integration challenges also emerged as a practical concern, with 25% of participants reporting occasional technical issues that disrupted workflow efficiency. These issues included system response delays during peak usage periods, inconsistent formatting of AI-generated responses, and periodic connectivity problems that prevented access to AI assistance. While not frequent enough to undermine overall system value, these technical difficulties contributed to user frustration and highlighted the importance of robust infrastructure support for AI-enhanced operations.

The comprehensive evaluation of system functions revealed generally positive attitudes toward the human-machine collaboration platform's core capabilities. Users particularly valued the Copilot portal interface, which integrated seamlessly with existing workflow patterns while providing intuitive access to AI assistance. The cross-knowledge content understanding functionality received consistent praise for its ability to synthesize information across different municipal service domains, enabling more comprehensive responses to citizen inquiries that might otherwise require transfers or callbacks.

The system's response reference information capabilities proved essential for maintaining service quality and supporting agent decision-making processes. Users appreciated the transparent attribution of information sources, which enabled rapid verification of AI suggestions and supported their professional responsibility for accuracy. The knowledge base maintenance functions streamlined the traditionally labor-intensive process of content updates and additions, while the semantic understanding and vector conversion capabilities ensured that new information became immediately searchable and accessible through natural language queries.

These user experience findings demonstrate that successful AI implementation requires careful attention not only to technical capabilities but also to the human factors that influence adoption, trust development, and effective utilization. The varying experiences across different user groups underscore the importance of tailored training approaches and ongoing support mechanisms that address the specific needs and concerns of different agent populations.

5. Conclusion

This research investigated the implementation and effectiveness of a generative AI human-machine collaboration system within the Taoyuan City Government's 1999 Citizen Hotline through an intensive case study approach. The findings provide both practical insights into AI deployment in public service delivery and theoretical contributions to understanding technology acceptance and human-computer interaction in organizational contexts.

The research findings provide substantial validation for established theoretical frameworks while extending their application to contemporary AI implementation contexts. The Technology Acceptance Model receives strong empirical support through the differential adoption patterns observed between senior and junior staff members. Senior agents' significant performance improvements clearly demonstrate high perceived usefulness of the AI system, while their ability to rapidly integrate AI suggestions into their workflow indicates favorable perceived ease of use. The contrasting experience of junior staff, who initially struggled to effectively utilize AI assistance, highlights how perceived ease of use can vary based on existing domain expertise and technological competency.

These findings extend TAM theory by suggesting that in AI implementation contexts, perceived usefulness and perceived ease of use may be more interdependent than traditionally conceptualized. Senior agents' extensive domain knowledge enabled them to quickly evaluate and integrate AI suggestions, creating a virtuous cycle where high perceived usefulness reinforced positive ease of use perceptions. Junior agents, lacking this foundational knowledge, found it more difficult to assess AI suggestion quality, potentially leading to lower confidence in system usefulness.

The research also provides valuable insights for Human-Computer Interaction theory, particularly regarding the critical importance of matching technology capabilities to task characteristics and human cognitive processes. The varying effectiveness across service categories validates HCI principles emphasizing that successful human-AI collaboration requires careful consideration of cognitive task demands. Structured, information-retrieval tasks (YouBike, citizen cards) aligned well with AI strengths in rapid data processing and pattern recognition, while complex, judgment-intensive tasks (social welfare) required human capabilities that AI could not adequately supplement.

This study contributes to the growing body of literature on human-AI collaboration by demonstrating that effectiveness depends on the dynamic interaction between individual factors (experience level, domain knowledge), task characteristics (complexity, structure), and system design factors (integration, usability). The research suggests that successful AI implementation requires moving beyond simple automation paradigms toward more sophisticated models of human-AI partnership that leverage complementary capabilities.

The findings offer several important insights for organizations considering generative AI implementation in customer service contexts. First, the research demonstrates that a differentiated implementation approach based on user experience levels and task complexity is essential. Organizations should expect AI systems to provide maximum benefit when deployed for structured tasks with experienced users, while recognizing that complex tasks and novice users require additional support mechanisms and training interventions.

The critical importance of tailored training programs emerges as a key practical implication. Senior staff may benefit most from advanced AI collaboration techniques that help them leverage their domain expertise more effectively, while junior staff require foundational training that builds both technical competency and confidence in AI system utilization. The research suggests that training programs should address not only technical system usage but also cognitive strategies for effectively evaluating and integrating AI-generated suggestions.

Organizations must also carefully consider task-technology fit in their AI deployment strategies. The research demonstrates that AI systems excel in supporting structured, information-intensive tasks but may provide limited value or even create inefficiencies for complex, judgment-intensive activities. This suggests the need for selective deployment approaches that match AI capabilities to appropriate use cases rather than pursuing comprehensive automation strategies.

Several limitations must be acknowledged in interpreting these findings. The sample size of four primary participants, while appropriate for intensive case study methodology, limits the statistical generalizability of quantitative findings. Future research would benefit from larger sample sizes that enable more robust statistical analysis and greater confidence in performance outcome measurements.

The twenty-day observation period, though sufficient for capturing immediate implementation effects, may not adequately represent longer-term adaptation processes or learning curve developments. Longitudinal studies extending over several months would provide more comprehensive insights into how human-AI collaboration patterns evolve as users develop greater familiarity with system capabilities and limitations.

The real-world operational environment, while providing authentic context for the research, introduced contextual variables that could not be controlled or systematically analyzed. Individual differences in technology skills, learning capabilities, and motivation levels among participants may have contributed to variation in system effectiveness beyond the specific factors examined in this study. Future research employing more controlled experimental designs could help isolate the effects of specific variables on AI system performance.

The case study methodology, while providing rich contextual insights, limits the transferability of findings to other organizational contexts. The specific characteristics of municipal government service delivery, the particular AI system architecture employed, and the unique organizational culture of the Taoyuan City Government may influence the applicability of findings to other settings.

The research identifies several important avenues for future investigation that could advance both theoretical understanding and practical applications of generative AI in organizational contexts. Longitudinal studies representing the most critical need, as they could capture learning effects, adaptation processes, and long-term performance impacts that were beyond the scope of this initial evaluation. Extended observation periods would enable researchers to understand how human-AI collaboration patterns evolve and whether initial performance differences between user groups persist or diminish over time.

Cross-organizational comparative studies represent another valuable research direction, enabling examination of how organizational factors, cultural contexts, and implementation strategies influence AI system effectiveness. Comparative analysis across different public sector agencies, private sector customer service operations, and international contexts could provide insights into the generalizability of findings and identify contextual factors that moderate AI implementation success.

Investigation of advanced AI models and their differential impacts represents an important technological research frontier. As generative AI capabilities continue to evolve rapidly, research examining how newer models with enhanced reasoning capabilities, multimodal processing abilities, and improved domain-specific knowledge perform in customer service contexts would provide valuable insights for future implementation decisions.

The development and evaluation of targeted training interventions emerges as a critical applied research area. Experimental studies examining different training approaches, support mechanisms, and skill development programs could provide evidence-based guidance for optimizing human-AI collaboration effectiveness. Such research could investigate both individual-level interventions (training programs, decision support tools) and organizational-level interventions (workflow redesign, performance management systems).

This research demonstrates that while generative AI holds substantial promise for enhancing public service delivery, successful implementation requires careful attention to the complex interplay between technology capabilities, task characteristics, and human factors. The findings emphasize that AI is not a universal solution but rather a powerful tool that must be thoughtfully integrated into existing organizational systems and workflows.

The central insight emerging from this study is that human-machine collaboration, rather than automation, represents the key to realizing AI's potential value in customer service contexts. Success depends not merely on technical system capabilities but on the effective guidance, application, and integration provided by human users. This perspective aligns with human-centered design principles and suggests that future AI implementations should prioritize augmenting rather than replacing human capabilities.

For government agencies and organizations considering generative AI adoption, this research provides evidence that significant operational improvements are achievable, but only through systematic attention to implementation strategy, user training, and ongoing system optimization. The findings suggest that organizations should focus not only on technological advancement but equally on knowledge base improvement, personnel development, and human-machine collaboration model refinement to realize the full value of AI technology in enhancing citizen service quality.

References

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of

Economic Research. https://www.nber.org/papers/w31161
Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., ... Yee, L. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey Global Institute.
Davenport, T., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.
Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, *97*(4), 62–73.
Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics*, *3*(2), 119–131.
Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50.
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.
Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.
Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.
Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe, and trustworthy. *International Journal of Human–Computer Studies*, *137*, Article 102385.
Shulzhenko, M. (2025). *Generative AI in contact centers: The ultimate guide for 2025*. VoiceSpin. https://www.voicespin.com/blog/generative-ai-in-contact-centers/
Tris, A. (2024). *How generative AI impacts customer support?* NobelBiz. https://nobelbiz.com/blog/generative-ai-customer-support/
Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204.
Wilson, H. J., Daugherty, P. R., & Bianzino, N. (2017). The jobs that artificial intelligence will create. *MIT Sloan Management Review*, *58*(4), 14–21.
Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.
Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., ... Perrault, R. (2022). *The AI index 2022 annual report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University.

About Author

Chaoming Wu is a Professor in the Department of Information Management at Chung Yuan Christian University in Taiwan. His research interests include Internet Marketing, AI and Data Mining, Smart City, E-Government, and Bridging Digital Divide