Development of Home Manager for Smart Home 4.0 Services

Hyunyong Park
Advanced Technology
QUBER Co., Ltd.
Seoul, South Korea
hypark@quber.net

Youngmoon Na
Advanced Technology
QUBER Co., Ltd.
Seoul, South Korea
ymna@quber.net

Youngjun Ko
Advanced Technology
QUBER Co., Ltd.
Seoul, South Korea
yjko@quber.net

Yongjin Kim
CTO/EVP
QUBER Co., Ltd.
Seoul, South Korea
cap@quber.net

***Abstract*— The integration of AI and IoT technologies is opening the era of 'Smart Home 4.0.' This evolution makes possible not only home automation but also human care, particularly for vulnerable groups like the elderly and those living alone. Despite the growing demand, existing systems often lack comprehensive solutions that combine intelligent monitoring, personalized interaction, and seamless device control. To address these challenges, this paper presents the development of the Smart Home Manager, an innovative system designed to enhance daily living and quality of life by integrating IoT platforms, AI Human interface, and conversational LLM-based chatbots. The Smart Home Manager actively monitors and manages household environments for recognizing user needs and delivering tailored services. This paper describes the system architecture of the Smart Home manager, the integration of advanced IoT technologies including IoT platforms such as OCF and Matter for interoperability, and the application of GPT-4-based conversational AI for proactive user interaction. Usability evaluations validate the system's effectiveness, offering significant insights into future innovations in smart home services and their practical implementation.**

***Keywords*— *Smart home, Home manager, AIoT, AI Human, LLM, Care Service* (*key words*)**

I. Introduction

The rapid development of AI and IoT technologies is reshaping the smart home landscape, ushering in the era of 'Smart Home 4.0,' where intelligent systems provide users with enhanced convenience, safety, and personalized services. Unlike earlier iterations, this new phase of smart home development focuses not only on routine automation but also on addressing complex human needs such as emotional care, social interaction, and proactive support. Leong et al. (2023) [1] demonstrated how AI-integrated IoT systems can significantly enhance user experience by automating device management and enabling intelligent, user-centric interactions.

In many societies, demographic shifts such as aging populations and the increasing prevalence of single-person households have brought new challenges. Issues such as loneliness, social isolation, and the need for personalized care are particularly pressing for vulnerable groups, including the elderly and individuals living alone. Smart home technologies, with their capacity for health monitoring, emotional care, and adaptive environmental control, offer a promising solution. Vrančić et al. (2024) [2] highlighted the critical role of these technologies in improving safety, health, and social interaction for older adults. However, despite their potential, many existing smart home systems fall short of delivering holistic solutions. Current systems often focus narrowly on basic automation or limited interaction, lacking the ability to understand and proactively respond to user needs. Furthermore, interoperability challenges between devices from different manufacturers continue to hinder the seamless integration required for an efficient smart home ecosystem.

To address these challenges, this paper introduces the Smart Home Manager, a system designed to integrate IoT platforms, an AI Human interface, and conversational LLM-based chatbots to provide intelligent monitoring, personalized interactions, and seamless device control. By leveraging cutting-edge technologies, the Smart Home Manager aims to enhance quality of life for all users, with a particular focus on vulnerable groups such as the elderly and those living alone. This paper details the architecture and key technologies of the proposed system, explores its services and real-world applications, and provides insights into how such a system can inform the future direction of smart home technologies.

II. Key Technologies for Home Manager

The Smart Home Manager leverages a combination of IoT technologies, a virtual avatar-based AI Human interface, and a chatbot powered by a Large Language Model (LLM) to create an intelligent and autonomous home management system. These core technologies enable seamless integration, efficient control, and personalized interaction, forming the foundation for an advanced smart home environment.

The **IoT platform** provides an integrated environment to connect, manage, process, and analdyze data from IoT devices. It employs the **OCF (Open Connectivity Foundation)** standard to ensure interoperability across IoT products and services, while the **Matter** standard ensures compatibility among devices from various manufacturers. The **AI Human interface**, using advanced generative algorithms, synthesizes realistic virtual avatars capable of engaging in natural conversations. This is further enhanced by an LLM-powered chatbot that acts as a virtual assistant, understanding and executing user requests effectively.

The detailed technologies of the proposed system are as follows:

A. IoT Platform

The Internet of Things (IoT) integrates internet connectivity into devices, enabling users to conveniently access and manage various services. In this research, the IoT platform implementation is based on the OCF (Open Connectivity Foundation) and Matter standards, which ensure interoperability and seamless communication among devices from diverse manufacturers. OCF supports the integration of IoT devices regardless of their operating systems or network protocols and includes the open-source framework IoTivity, which facilitates device connectivity. IoTivity utilizes the Constrained Application Protocol (CoAP) to provide high-quality service while minimizing computing power and network overhead. CoAP adopts a lightweight RESTful architecture for message transmission and functions efficiently over UDP in low-power, limited-energy environments. The architecture of IoTivity is illustrated in Fig 1.

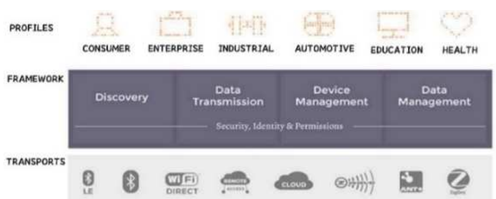

Fig. 1. IoTivity Architecture, adapted from [9]

Matter, developed by the Connectivity Standards Alliance (CSA), is an open standard designed to enhance interoperability and reliability among smart home devices. Matter supports stable and efficient connections, enabling a unified ecosystem where devices from various platforms can operate seamlessly. Built on the IP protocol, each device in the smart home is assigned a unique IP address, ensuring reliable communication over IP networks. Matter's approach significantly improves connectivity and device compatibility, as highlighted in [10]. The standard architecture of Matter is shown in Fig 2.

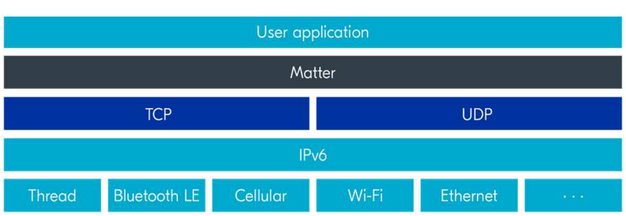

Fig. 2. Matter standard architecture

The Smart Home Manager integrates the strengths of both OCF and Matter standards, assigning specific roles to each. OCF is optimized for managing low-power devices such as ZigBee-based sensors and actuators, while Matter is used for high-performance devices relying on Wi-Fi technology. This hybrid approach enables different network technologies and devices to function cohesively, delivering a robust and customizable smart home environment. The architecture of the Smart Home Manager is depicted in Fig 3.

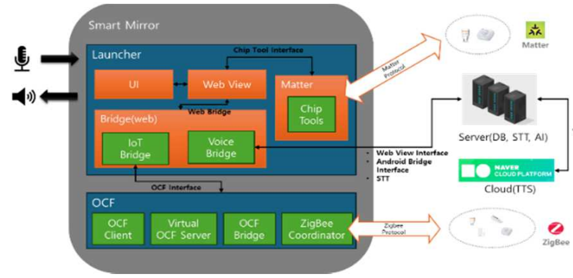

Fig. 3. Smart home manager architecture

This integrated IoT platform enhances system efficiency and improves the user experience by enabling seamless interoperability and reliable performance. Through this approach, the Smart Home Manager provides a scalable solution tailored to the evolving needs of modern smart homes.

B. AI Human Interface

The AI Human Interface developed in this research utilizes the Thin-Plate Spline (TPS) motion model [3] and the wav2lip algorithm [4] to naturally synchronize the movements and lip shapes of virtual avatars. Fig. 4 summarizes the process of creating the AI Human Interface.

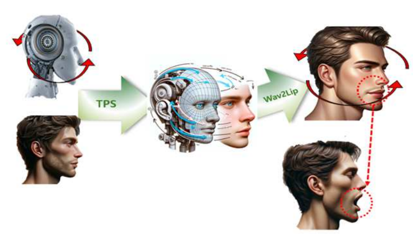

Fig. 4. Process For Generating AI Human Interface

In order to respond adequately to people with several level of age and gender, we compose the AI Human with six identities based on virtual character images generated through generative AI (Fig. 5).

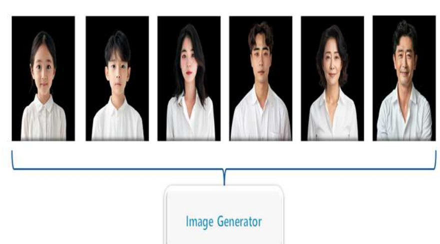

Fig. 5. AI Human types

The TPS motion algorithm is a powerful tool for Image Animation in computer vision, enabling selective movement of specific parts of an image, such as a person's face, by transforming images into desired shapes. TPS achieves this by mapping keypoints from the original image to new locations, modeling deformations similar to those of a thin plate subjected to physical forces. This process ensures smooth transformations, resulting in vibrant and natural animation effects. For instance, mapping the positions of the mouth, eyes, or entire face to corresponding positions in subsequent frames can create realistic motion. As Zhao and Zhang (2022) [3] describe, the TPS transformation function $\mathcal{T}_{tps}$ must accurately map the source keypoints P_i^S to the driving image keypoints P_i^D. The warping is governed by minimizing the overall curvature distortion across the entire image area R^2, calculated through second partial derivatives (Eq. 1):

$$min \iint_{R^2} \left(\left(\frac{\partial^2 \mathcal{T}_{tps}}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 \mathcal{T}_{tps}}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 \mathcal{T}_{tps}}{\partial y^2}\right)^2 dxdy \right. \quad (1)$$

$$s.t. \quad \mathcal{T}_{tps}(P_i^S) = P_i^D, \quad i = 1,2,\dots,N,$$

According to research [2], Wav2lip can be used to synchronize lip movements and voice with the animated images. It implements natural speaking animations for avatars based on voice input by synchronizing lip movements of the animated image with the voice data. By adjusting the mouth shapes and minimizing the discrepancy between voice and video, high-quality voice-video synchronization is achievable.

The Wav2lip algorithm is designed to naturally align the mouth movements of the character in the video according to the given voice input. The input audio data is transformed into a two-dimensional matrix on a Mel scale containing both temporal and frequency information, based on the Mel Spectrogram, and is represented as a vector through an audio encoder. The video frame data corresponding to the Mel scale is transformed into a vector through a Face encoder, and these two vectors are concatenated and then expressed as a speaking video through a Face decoder. The overall structure follows that of a GAN (Generative Adversarial Network), and the Lip-sync discriminator evaluates how well the mouth shapes in the generated video frames are synchronized with the actual audio. As shown in Eq. (2), the synchronization probability of the input audio-video pair can be calculated by computing the dot product between the video embeddings v and speech embeddings s :

$$P_{sync} = \frac{v \cdot s}{\max(\|v\|_2 \cdot \|s\|_2,\ \epsilon)} \quad (2)$$

The synchronization probability is based on cosine similarity, and a new loss called cosine-similarity with binary cross-entropy is used alongside binary cross entropy loss.

C. STT(Speech-To-Text)

In our system, the voice recognition feature uses a Whisper-based algorithm as the STT model. According to [5], the Whisper model is an STT model based on the Transformer architecture, pre-trained on a Weakly Supervised Dataset and evaluated in a zero-shot manner, making it a robust and general algorithm. The Whisper-based model can be operated on-premise, allowing for direct customization of features such as silence duration, data format, and fine-tuning. It is implemented considering optimized communication in real-time since it communicates with the client through PCM format based on Websocket protocol. It is not limited to listening to short phrases but can also listen to long voices in real-time, and it is effective for interaction by sending intermediate recognition results to the client as a certain buffer accumulates.

D. LLM-based Chatbot

In this paper, we apply chatbot technology based on the Large Language Model (LLM) to provide human-like interactions through highly advanced natural language processing capabilities. The chatbot is equipped with features such as complex question answering, context understanding, and language generation abilities. To implement these functionalities, we utilized GPT-4, accessed via the OpenAI API, as the foundational LLM for this system. GPT-4, known for its robust natural language understanding and generation capabilities, serves as the backbone for creating realistic and contextually accurate interactions. While traditional rule-based chatbots offer only predetermined responses, GPT-4's capacity for nuanced understanding allows it to generate vibrant, human-like replies across various tones, styles, and contexts.

To tailor the chatbot for smart home applications, we employed **prompt engineering** techniques rather than additional fine-tuning, effectively guiding GPT-4's behavior through detailed instructions and scenarios within input prompts. Prompts were designed to handle specific tasks such as IoT device control, emotional care dialogues, and context-based long-term memory queries, eliminating the need for resource-intensive fine-tuning while optimizing GPT-4's responses for the intended use cases. A "Prompt" serves as input data provided to the Language Model, instructing it to perform specific tasks. **According to research [6]**, techniques like **Zero-shot prompting** (performing new tasks without pre-labeled data), **Few-shot prompting** (providing examples to clarify tasks), and **Chain-of-Thought prompting** (guiding step-by-step reasoning) can significantly enhance the accuracy and adaptability of LLM responses without additional training. Prompt engineering was also used to assign roles, define response limitations, and craft detailed instructions, enabling the chatbot to deliver essential smart home services effectively. **Figure 6** illustrates examples of prompt creation.

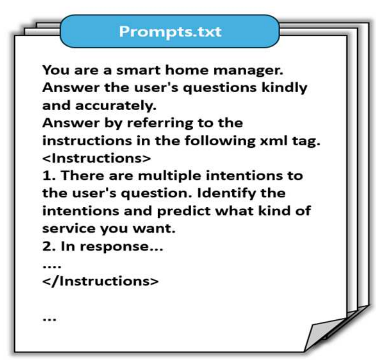

Fig. 6. Prompt examples

Effective Prompt Engineering applied to LLM can identify the intent and key entities in user sentences, leading to transitions to variously prepared prompts. Furthermore, with effective design of Prompt Engineering, real-time services that call various external functions through slight code design can also be provided [7]. Additionally, the RAG (Retrieval-augmented generation) methodology [8] is applied, which not only incorporates knowledge of the distribution of pre-trained data but also external knowledge into LLM based on prompt engineering, enabling it to answer based on new domains or the latest data. Thus, Prompt Engineering alone, without additional training such as fine-tuning, can create various prompts tailored to the desired domain and optimize them for the desired service.

For real-time deployment, the chatbot interacts with users via a conversational interface integrated with the Smart Home Manager. The system processes user inputs in natural language, formulates structured JSON commands for IoT devices, and provides meaningful feedback, ensuring seamless and intuitive communication.

III. SERVICES OF SMART HOME MANAGER

A. Smart home management services

AIoT (Artificial Intelligence of Things) used in the smart home manager utilizes the IoT platform and an AI-based LLM-based chatbot to analyze and control, providing convenience to the user. Unlike traditional IoT control, which involves direct clicking on a screen UI or rule-based voice control, the AIoT service in this implementation allows for device control through various voice conversations with an AI Human Interface-based LLM-based chatbot.

For IoT control, our system exchanges not just sentences but data in Json format during client-server conversation messages, sharing explicit information about the user's current intentions, the devices that can be controlled, and the current state of the devices. Through prompt engineering, instructions on how to interpret, analyze, predict the Json's key-values, and how to configure parameters for device control are given. Based on this, the LLM delivers an appropriate response to the user along with the message for IoT control. The client then conducts the final IoT control based on this message and delivers the final answer to the user. The flow of this process is shown in Fig. 7.

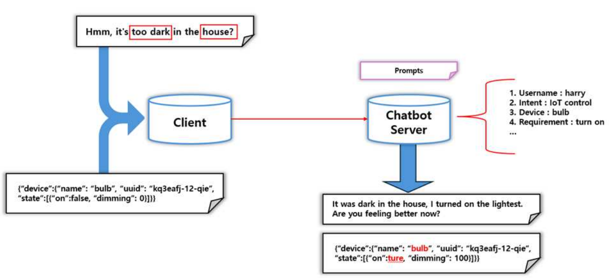

Fig. 7. IoT control flow examples

Such methods are possible thanks to the application of prompt engineering. The prompts specify the role of the chatbot, the structure of the prompt composition such as xml format, dash, triple backticks, and apply the principles of Few-shot prompting [6] and Chain of thought [6] to provide detailed operational instructions using appropriate examples.

Additionally, the AIoT service adjusts environments such as sleep mode and homecoming mode based on the user's patterns through IoT deployment control. Logical control is possible, not just processing sequential commands, but through checking IoT states. For instance, assuming a scenario where the user is returning home, if the door sensor detects homecoming and the presence sensor recognizes the user's presence for a certain period, our home manager, after recognizing the user's face, greets and proceeds with IoT control according to the pre-set homecoming mode. An example of this AIoT deployment control process is shown as Fig. 8.

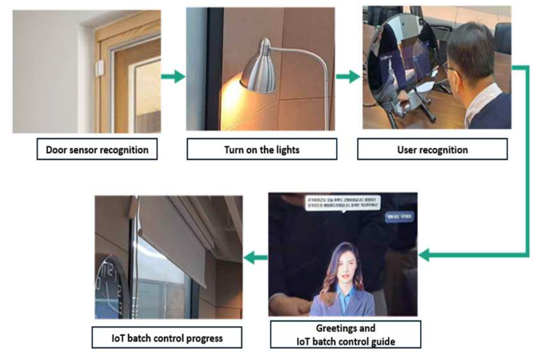

Fig. 8. AIoT batch control examples

B. Family member care service

As the number of single-person households and elderly living alone increases in the era of Smart Home 4.0, the importance of human emotional care, beyond just systems for house, appliance, and information provision, is growing. Therefore, there is an increasing need for personalized services for individual users. In our system, the Care service recognizes users and provides conversational services that can inquire about their well-being, offer advice or comfort based on what was discussed in the past, any incidents that occurred, and whether the user had any ailments, all using long-term memory conversation history.

Since LLM has not been trained on long-term memory conversation history data, this data cannot be utilized in simple conversation. Fine-tuning based on this historical data would not only require excessive resources but is also impractical in real-time. Therefore, our system applies RAG [8] to integrate external knowledge into LLM, enabling it to generate responses based on the linked data. As conversations between the user and the chatbot progress, each conversation is tokenized and vectorized using a text embedding model, and the transformed vectors are stored in vector databases such as Chroma DB and FIASS, along with metadata. Likewise, sentences from the user's questions in each conversation are vectorized, similarities are calculated with vectors of long-term memory conversation history in the vector database, and the most similar contents are retrieved. If the user's question sentence is represented as $q \in \mathbb{R}^d$, and the embedding data existing in the vector database as $V = \{v_1, v_2, v_3, \dots, v_N\} \in \mathbb{R}^{N \times d}$, the process of retrieving similar documents can be expressed as follows in Eq. (3):

$$cosine\ similarity\ (q, V) = \frac{q \cdot V^T}{\|q \cdot V^T\|} \in \mathbb{R}^N \tag{3}$$

In our system, through conversation, not only does it inquire about wellbeing or offer comfort based on past conversations, but it also classifies emotions displayed in the user's questions into five categories (very good, good, normal, bad, very bad) and manages them. We apply prompt engineering techniques to the LLM, instructing it on how to predict emotions, provide instructions, and present few-shot examples on how to analyze sentences. The analyzed emotions are visualized over periods such as daily, monthly, or quarterly, providing an emotional analysis dashboard that offers various care services such as self-diagnosis or alerts to guardians. Fig. 9 shows an example of the emotional analysis dashboard application.

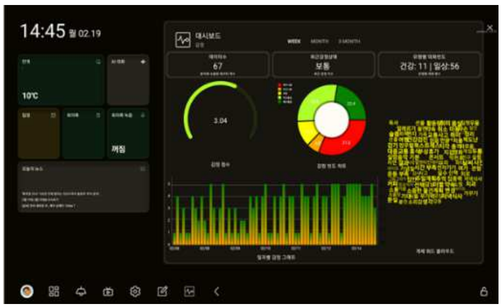

Fig. 9. Emotion Analysis Dashboard Application

C. *Office business supporting service*

Our system also provides services in office environments as well as smart homes. The office service consists of two main features. The first is a meeting summary function, which is configured by incorporating the RAG function into the LLM. Multiple users request the home manager to record a meeting, which is conducted nearby. During the meeting, all voices are recognized in real-time and converted to text by our system's STT function. Once the meeting is over, the home manager stops recording the meeting, inputs the detected texts into the LLM, and summarizes the meeting content in a certain format. Based on the summarized content, users can conduct real-time queries and responses. For example, if you ask the home manager, "When is the next week's meeting?" it uses the RAG algorithm to measure the similarity between vectors of the summarized and original documents, retrieves similar documents, and then produces an accurate response through the LLM, such as "The meeting has been scheduled for next Friday." The flow of this process is shown in Fig. 10.

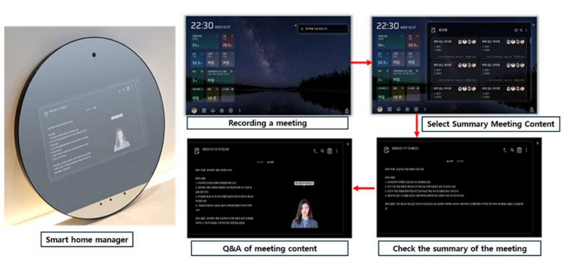

Fig. 10. Office business supporting service flow examples

Another feature of the office service is the schedule registration function. This function registers and notifies all schedules, including users' meeting schedules and private appointments. Prompt Engineering is used to instruct which essential information needs to be obtained from conversations with users, presenting questions that can secure this information in Json format. The Json data obtained for schedule registration is stored in a relational database to register the schedule, and an application sends a popup notification 30 minutes before the schedule starts to help the user remember their schedule.

IV. EVALUATION

We have conducted external usability evaluations to objectively assess the quality, commercial viability, and usability of our system. The focus of the external usability evaluation is to understand how effectively our system interacts with users and to identify the system's benefits and limitations. The results of the external usability evaluation are as follows:

A. *UBC usability evaluation Configuration*

We set up a test living lab at UBC (The University of British Columbia) for an in-depth product evaluation. The living lab was designed not just with device installation but also with interior decoration to simulate a naturalistic home environment, as shown in Fig. 11.

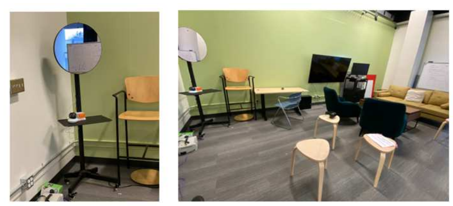

Fig. 11. Smart home manager and Living Lab at UBC

The evaluation process and methodology began with a pre-evaluation introduction to the participants about the product, and the evaluation method was divided into two parts. Initially, participants were asked to interact directly by touching to perform specific tasks. Then, after additional explanation about the AI Human-based home manager, the tasks were repeated through voice conversations. Feedback was collected through subsequent surveys and interviews. The detailed process is shown in Fig. 12.

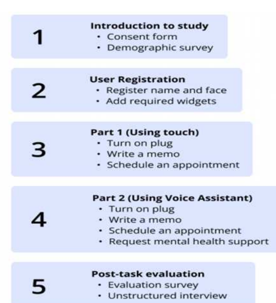

Fig. 12. A visual representation of the evaluation procedure

B. Evaluation Result

The usability evaluation focuses primarily on feedback through face-to-face interviews rather than quantitative assessments. The evaluation and interview contents include:

- Participants highly rated the ease of using voice conversations with the home manager to control smart plugs, lighting, etc., appreciating the ability to operate without direct manipulation. Moreover, many responded that they would frequently use these convenient and straightforward features in real life.
- The ability to issue voice commands in busy situations was found to be very useful from a multitasking resource perspective.
- The AI Human Interface, which replicates a human appearance, received widespread feedback for providing a realistic feeling of conversing with a person; however, some responses noted that certain movements seemed robotic.
- Participants mentioned that the emotional care feature could be beneficial, helping regulate emotions in everyday situations, and could be particularly useful in scenarios where direct human interaction is difficult.
- Some participants noted that simple conversations through AI could be beneficial when direct human interaction is challenging, suggesting that while AI cannot replace real human conversation, it could serve as an alternative for some interactions.

V. CONCLUSION

Our research introduces a smart home manager that not only provides AIoT services and a user-friendly AI Human Interface in the Smart Home 4.0 era but also offers diverse home care services. The home manager is designed to deliver convenient and innovative smart home experiences for all users, including single-person households and elderly individuals living alone. Its effectiveness and practicality have been validated through real usability evaluations.

Future plans include implementing core AI functionalities such as LLM on-device into Edge AI, which will not only enable services to operate without a network but also address critical concerns related to privacy and security by ensuring that sensitive data remains within the local environment. Additionally, we aim to enhance the stability and diversity of the IoT platform, incorporating various Matter-compatible devices to maximize user convenience and system interoperability.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MOTIE/KEIT (20026230, Development of AIoT device utilizing Channel State Information (CSI) for AI-based lifestyle recognition).

REFERENCES

[1] Y. M. Leong, E. H. Lim, and L. K. Lim, "A Review of Potential AI-Based Automation for IoT-Enabled Smart Homes," in Proc. 2023 IEEE 13th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, Oct. 2023, DOI: 10.1109/ICSET59111.2023.10295156.

[2] A. Vrančić, H. Zadravec, and T. Orehovački, "The role of smart homes in providing care for older adults: A systematic literature review from 2010 to 2023," Smart Cities, vol. 7, no. 4, pp. 1502-1550, 2024. https://doi.org/10.3390/smartcities7040062

[3] J. Zhao and H. Zhang, "Thin-Plate Spline Motion Model for Image Animation," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, June 18-24, 2022, pp. 2-4. DOI: 10.1109/CVPR52688.2022.00364.

[4] K. R. Prajwal, R. Mukhopadhyay, V. Namboodiri, and C. V. Jawahar, "A Lip Sync Expert Is All You Need for Speech to Lip Generation In The Wild," arXiv preprint arXiv:2008.10010, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2008.10010

[5] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," arXiv preprint arXiv:2212.04356, Dec. 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2212.04356

[6] P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications," arXiv preprint arXiv:2402.07927, Feb. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2402.07927

[7] S. He, "Achieving Tool Calling Functionality in LLMs Using Only Prompt Engineering Without Fine-Tuning," arXiv preprint arXiv:2407.04997, Jul. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2407.04997

[8] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Proc. of NeurIPS, Dec. 2020.

[9] J. Seo, S. Hwangbo, J. Chun, H. So, and M. Kim, "Designing and Implementing the Smart Home Service Architecture Based on IoTivity Open Source," in Proc. of the Korean Institute of Communications and Information Sciences (KICS) Summer Conference, 2020, pp.1.

[10] OSS, "[November Monthly Briefing] Trends in IoT Technology Using Open Source," posted on support page, Nov. 28, 2022. [Online]. Available: https://www.oss.kr/oss_guide/show/ce634345-2181-4c72-8d03-5e4e2a1d11e2. [Accessed: Oct. 5, 2023].