석 사 학 위 논 문
Master's Thesis

# 노인의 모바일 유저 인터페이스 접근성 향상을 위한 과제 지향적 다중 모달 에이전트 시스템

TOMAS: A Task-Oriented Multimodal Agent System to Enhance Mobile User Interface Accessibility for Older Adults

2024

이 상 욱 (李 商 旭 Lee, Sangwook)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

# 노인의 모바일 유저 인터페이스 접근성 향상을 위한 과제 지향적 다중 모달 에이전트 시스템

2024

이 상 욱

한 국 과 학 기 술 원

전산학부

# 노인의 모바일 유저 인터페이스 접근성 향상을 위한 과제 지향적 다중 모달 에이전트 시스템

이 상 욱

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2023년 12월 19일

심사위원장 신 인 식 (인)

심 사 위 원 김 주 호 (인)

심 사 위 원 Jia-Jun Li (인)

# TOMAS: A Task-Oriented Multimodal Agent System to Enhance Mobile User Interface Accessibility for Older Adults

Sangwook Lee

Advisor: Insik Shin

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Computer Science

Daejeon, Korea
December 19, 2023

Approved by

Professor of Computer Science

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MCS . 노인의 모바일 유저 인터페이스 접근성 향상을 위한 과제 지향적 다중 모달 에이전트 시스템. 전산학부 . 2024년. 17+iv 쪽. 지도교수: . (영문 논문)
Sangwook Lee. TOMAS: A Task-Oriented Multimodal Agent System to Enhance Mobile User Interface Accessibility for Older Adults. School of Computing . 2024. 17+iv pages. Advisor: Insik Shin. (Text in English)

**초 록**

이 연구는 노인들이 스마트폰을 사용하는 데 겪는 어려움을 해결하기 위해 개발된 과업 지향적 다중 모달 에이전트 시스템을 소개합니다. 이 시스템은 노인들이 흔히 마주치는 작은 글씨, 복잡한 메뉴 구성, 네비게이션 문제 등을 해결하는 데 중점을 두고 있습니다. 노인 스마트폰 사용자를 지원하는 경험이 있는 학생들과 사회복지사들과의 인터뷰를 통해, 인지적, 신체적, 디자인 등에서의 주요 어려움이 확인되었고, '대리 지원'과 '교육적 지원'의 두 가지 지원 방법이 발견되었습니다. 교육적 지원 방법에서 착안한 에이전트 시스템은 사용자 인터페이스를 간소화하고 음성 명령과 같은 노인 친화적 기능을 통합하여, 노인 사용자들이 스마트폰을 보다 독립적으로 자신 있게 사용할 수 있도록 지원합니다. 한국의 노인들을 대상으로 한 영화관 티켓 예매 작업에서의 사용자 테스트는 시스템의 개선이 필요한 부분, 예를 들어 응답 시간과 지시사항의 명확성 등을 드러내며, 보다 직관적이고 시각적인 도움을 제공하는 사용자 인터페이스의 필요성을 강조합니다.

**핵 심 낱 말** 노인 스마트폰 사용 지원, 사용자 인터페이스 단순화, 과업 지향적 다중 모달 에이전트 시스템

**Abstract**

This research introduces the Task-Oriented Multimodal Agent System (TOMAS), which was developed to overcome the challenges older adults face in using smartphones, such as small text size, complex menu structures, and navigation issues. Interviews with students and social workers experienced in supporting elderly smartphone users helped identify key challenges, including cognitive, physical, and design difficulties, as well as two primary methods of assistance: 'Proxy Assistance' and 'Educative Assistance.' Inspired by educative assistance, TOMAS focuses on simplifying the user interface and integrating senior-friendly features like voice commands to empower older adults to use smartphones more independently and confidently. User testing with elderly Korean participants, conducted through a movie theater ticket booking task, revealed areas where TOMAS could be improved, particularly in response time and instruction clarity, highlighting the necessity for more intuitive and visually supportive user interfaces.

**Keywords** Elderly Smartphone Usage Support, User Interface Simplification, Task-Oriented Multimodal Agent System

# Contents

# List of Tables

# List of Figures

# Chapter 1. Introduction

Recently, with the increase in the older adult population, there has been a rise in smartphone adoption among them. Now, they use smartphones to communicate with others, enjoy music and videos, and make purchases through online shopping. However, they still use fewer mobile apps and fewer features per app compared to younger populations[5], and their level of utilizing digital devices and the internet is lower than other digitally disadvantaged groups like the disabled, low-income, and rural populations[1]. One of the causes is the inconvenience of mobile user interfaces provided by smartphones. Previous studies have identified major usability issues such as small text and UI elements, complex menus with too many options, and navigation problems[2]. Due to these inconveniences, older adults typically rely on younger family members or professionals who are more familiar with technology[7, 8, 1]. However, this person-to-person support is limited by time and physical constraints and can lead to uncomfortable feelings between individuals[8], as well as privacy issues[3].

To address these issues, we developed TOMAS, a Task-Oriented Multimodal Agent System that mimics the support provided by humans in assisting smartphone use, utilizing large language models (LLM). For the interaction design of TOMAS, we conducted interview sessions with 15 students and four social workers who were experienced in assisting the elderly with smartphone usage. These interviews focused on analyzing the main difficulties the elderly encounter while using smartphones and the methods used to assist them.

The interviews revealed that older adults' primary difficulties in smartphone usage include cognitive, physical, design, and experience or knowledge-related issues. These include memory issues, fear of incorrect operations, visual discomfort, difficulty using keyboards, UI complexity, and difficulty understanding icons and terms. Also, two main support methods, 'Proxy Assistance' and 'Educative Assistance,' were identified. 'Proxy Assistance' involves performing tasks on behalf of the older adult, typically used for complex app usage or settings, whereas 'Educative Assistance' is a method of teaching them to use smartphones independently. These methods are vital in enabling elderly users to use smartphones more effectively.

From these interview insights, the interaction design of the TOMAS system focuses on enabling elderly users to use smartphones independently. The system highlights essential UI elements, simplifies the user interface, and offers elderly-friendly input interfaces like voice commands. This enables the elderly to learn and understand various smartphone functions more quickly and build confidence in their smartphone usage. Ultimately, TOMAS encourages the elderly to explore smartphones more actively and independently and find and use the necessary features.

For this study, user testing was conducted with four elderly Korean participants to evaluate the design and improvement directions of the TOMAS system. TOMAS was adapted to provide a movie theater ticket booking service using the Megabox mobile website. The results indicated that participants pointed out the slow response time of the TOMAS system and the confusion caused by separate screens. Some participants assessed that TOMAS's explanations could be helpful for beginners but needed to be more concise and clear. Additionally, we found the need for explanations with visual aids and additional information in the user interface.

# Chapter 2. Related Work

## 2.1 Smartphone Usage of Older Adults

While smartphone use among older adults has become increasingly common, many still struggle to adapt to and utilize various applications. Notably, they find it challenging to comprehend the meanings of icons and navigate interfaces comprised of menus and buttons[4]. These difficulties vary in degree based on individual factors such as age, cognitive abilities, and vision, leading the elderly to use only basic functions despite having smartphones[5]. Contrary to some perceptions, older adults prefer to learn technology autonomously and seek remote support[6], and it has been found that family members positively impact their learning as supporters, protectors, and monitors[7]. However, repeated reliance on family assistance can lead to feelings of burden for both the older adults and their helpers, potentially causing guilt or embarrassment in the elderly[8]. Our system is designed to address these identified issues by aiding elderly users in easily understanding and navigating smartphones' complex and varied UI. By providing personalized approaches that consider the cognitive and visual changes associated with aging, this system helps to reduce the elderly's dependence on family members, promoting more independent learning and use of technology.

## 2.2 Mobile UI Assistant for Older Adults

Various solutions have been researched to assist novice users, including older adults, using mobile UIs. HelpViz automatically generates tutorials with images and highlights based on text descriptions, aiding users in easily accessing mobile apps[10]. Video2Action uses deep learning models to annotate actions in tutorial videos automatically, enabling users to recognize necessary actions quickly[11]. These tools suggest that sequential explanations of complex interactions and visual cues for actions can aid in learning mobile UIs. Systems for assisting older adults with mobile UI usage have typically been proposed in forms utilizing help from others. Synapse is designed to allow helpers to easily create multimodal interactive tutorials, enabling older adults to learn apps independently through a trial-and-error mode[13]. Additionally, Meetkat allows older adults to share screenshots with explanations and visual cues with friends, family, or volunteers when they struggle with mobile UIs, making it easier to receive help[12]. Recent studies have proposed voice assistant designs that enable older adults to actively explore UI features through questions, as investigated in the Wizard of Oz study[14]. TOMAS is developed based on this research trend and is designed to help older adult users more easily understand and utilize various smartphone UIs. This system leverages the advantages of visual tutorials provided by tools like HelpViz and Video2Action and adopts approaches supporting independent learning and interaction, similar to Synapse and Meetkat. Furthermore, based on insights from voice assistant designs, TOMAS helps older adults engage in natural conversational interactions to deepen their understanding of smartphone UIs and navigate them according to their individual needs.

## 2.3 Understanding User Interface with LLMs

Recent studies utilizing large language models (LLMs) explore the potential to enhance interaction with graphical user interfaces (GUIs). Notably, LLMs have demonstrated the ability to interpret UI layout codes and exhibit exceptional natural language understanding, suggesting the possibility of transforming interactions with UI into various forms of conversational interactions[15, 16]. Consequently, UI-based mobile automation systems have been introduced recently. AutoDroid[17] is one such system that uses LLMs to automate tasks by interacting with mobile UIs. Further developments include systems that use separate image models to reflect UI visual elements[18], vision-language models[19], or advanced commercial models like GPT-4V[20], to automate interactions with mobile UIs based on natural language queries. Meanwhile, it's been recognized that LLMs effectively understand HTML with rich attribute information[16], revealing their capability to discern important elements on webpages from given natural language queries[20]. TOMAS leverages LLMs to comprehend the rich attribute information of mobile web HTML, aiding elderly users in easier smartphone use. While currently not employing multimodal models, integrating such models could further enhance the understanding and interaction with UIs, potentially improving performance.

# Chapter 3. Understanding Elderly Assistance in Smartphone Usage: Insights from Students and Social Workers

We envisioned an agent system that mimics the people who respond to elderly individuals' requests for help with smartphone usage through face-to-face interaction. Typically, when older adults face difficulties in using digital devices, they tend to seek assistance from their tech-savvy children, friends, or social workers who have the time to help[8]. In most cases, these helpers receive the older adult's smartphone, understand their intent through questions, and assist them by using the smartphone on their behalf. Before designing our agent system, we investigated how these helpers actually assist older adults with smartphone use in various situations.

## 3.1 Methods

In our research, we utilized semi-structured interviews and role-playing to better understand the difficulties faced by older adults in using smartphones. Initially, we interviewed adults with experience assisting older adults with smartphones. Students were recruited through university community announcements, and social workers were recruited through a local senior welfare center. There were a total of nineteen participants (fifteen students and four social workers), eight of whom were Korean and the rest from various Asian countries(Table3.1).

Before the interviews, participants completed a pre-survey about their relationship with the older adult, the tasks they were asked to assist with, and how they helped. This survey helped participants recall their experiences. During the interviews, we asked detailed questions about their relationship with the older adult, the frequency of assistance, and the specific tasks requested based on the survey results.

A unique aspect of the interviews was the role-playing. Participants reenacted the situations in which they had assisted older adults, with the interviewer playing the role of the older adults. Throughout this process, participants used Android smartphones, and the entire session was audio-recorded and video-recorded. The smartphone screens were also recorded to capture detailed information about the participants' handling of the devices. The recorded audio was later transcribed and, along with the videos, analyzed by the paper's main author.

## 3.2 Results

### 3.2.1 Relation with Older Adults

Student participants primarily assisted family members such as grandparents, parents, or relatives. Korean students, often living apart from their families, mostly provided face-to-face assistance during home visits. They also employed remote support methods like sharing YouTube videos (P4), smartphone screen sharing (P8), and sharing screenshots (P11). In contrast, international student participants, who mostly lived with older adults in their home countries, frequently provided assistance upon request. Social worker participants mainly assisted older adults in visiting the welfare centers.

Table 3.1: Interview Participants' demographic

| Participants | Job | Nationality | Gender | Age | Relationship with older adults |
|---|---|---|---|---|---|
| P1 | Student | India | Male | 19 | Grandparent, Parent, Relative, Neighbor |
| P2 | Student | Korea | Female | 22 | Parent, Relative |
| P3 | Student | Korea | Male | 22 | Stranger |
| P4 | Student | Korea | Male | 24 | Parent |
| P5 | Student | Korea | Male | 23 | Grandparent |
| P6 | Student | India | Male | 21 | Grandparent, Relative, Neighbor |
| P7 | Student | Kazakhstan | Female | 19 | Grandparent, Parent, Relative, Neighbor |
| P8 | Student | Philippines | Female | 19 | Grandparent, Parent, Stranger |
| P9 | Student | Kazakhstan | Female | 20 | Grandparent |
| P10 | Student | Korea | Female | 22 | Grandparent |
| P11 | Student | Korea | Male | 29 | Parent, Neighbor |
| P12 | Student | Kazakhstan | Male | 19 | Grandparent, Parent |
| P13 | Student | Korea | Female | 21 | Parent, Relative |
| P14 | Student | Korea | Female | 30 | Grandparent |
| P15 | Student | Kazakhstan | Female | 21 | Grandparent, Relative, Stranger |
| S1 | Social Worker | Korea | Female | 26 | Client |
| S2 | Social Worker | Korea | Male | 28 | Client |
| S3 | Social Worker | Korea | Female | 57 | Parent, Relative, Client |
| S4 | Social Worker | Korea | Female | 31 | Client |

### 3.2.2 Key Challenges of Older Adults' Smartphone Usage

According to the interviews, the challenges in smartphone usage faced by older adults can be categorized into cognitive, physical, design, and experience/knowledge-related issues. Cognitive issues primarily arose from infrequent smartphone use, leading to memory problems and fear of incorrect operations or mistakes. Participants P1, P5, P10, P11, and P14 reported that older adults tend to forget previously learned operations. Participants P1, P11, and S3 noted that older adults often feel anxious about making mistakes or incorrect operations. Physical challenges included visual discomfort due to small screen size and difficulties with keyboard typing. Participants P2, P3, P9, P13, and P15 mentioned that older adults experience discomfort due to the small size of smartphone screens. Participants P7, P8, P9, and P12 reported that older adults find keyboard typing challenging.

Design issues were mainly related to difficulties with multi-step procedures and the complexity of user interfaces (UI) on the screen. Participants P2, P6, P10, and S1 pointed out that following multi-step procedures on smartphones is challenging for older adults. Participants P2, P3, P5, P6, P8, and P10 mentioned that older adults find the arrangement of UI elements on the screen too complex and confusing. Finally, experience and knowledge-related issues involved difficulty understanding icons and terms and challenges in using new features. Participants P6, P7, P8, and S2 reported that older adults struggle to understand the meanings of specific icons or terms. Participants P2, P3, P5, P10, P14, and P15 mentioned that older adults face difficulty adapting to new features or updated apps.

### 3.2.3 Behaviors of Elderly Assistance

Our study identified two main methods of assisting older adults with smartphone usage: 'Proxy Assistance' and 'Educative Assistance.' 'Proxy Assistance' involves performing the task on the smartphone on behalf of older adults, often used for complex app usage or smartphone settings. For example, participants assisted with online shopping apps (P1, P6, P8, P9, S3), reservation apps (P1, P6), and administrative service apps (P15, S1). This method was also applied to elderly individuals (P8, P9, S4, P12, P13) who physically struggled or were disinterested in smartphone usage. Participants also handled one-time tasks like installing apps, registering and logging in (P6), or setting up payment methods (P2).

Conversely, 'Educative Assistance' focuses on teaching older adults to use smartphones independently. This method is commonly used in the initial stages of smartphone usage to teach basic functions like calling, texting, and photography (P5, P7, P10, P12), or advanced features of existing apps (P1, P2, P10, P15, S2) and new app usage (P2, P4, P7, P10, P13, P14, S1). Participants employed various strategies such as repetitive explanations (P2, P10), practical exercises (P5, P10), and user manuals (P10, P14) to facilitate older adults' learning. Sometimes, participants combined 'Proxy Assistance' with teaching, allowing older adults to try it themselves (P1, P6). For instance, P1 encouraged older adults to scroll the screen or click certain buttons, while P6 explained the UI's content and meaning during use.

In both methods, participants mostly sat beside older adults, viewing the smartphone together, which enabled the elderly to observe the participant's actions and voice their input easily. Some participants chose not to show the smartphone when they were fully dependent (P12, P13) or when sitting side by side was physically challenging (P3).

In Proxy Assistance, participants typically proceeded with most interactions based on their judgment without explanation. They quickly passed non-critical steps, asking for or directly inputting necessary information from older adults. When a list requiring selection appeared, participants either made choices considering the elderly's preferences or showed the screen for them to choose from.

In Educative Assistance, participants showed and explained each step of using the smartphone's features on the screen. They pointed to UI elements, explaining their shape, location, and purpose, and provided additional explanations when complex icons or technical terms appeared. This approach helped older adults better understand and independently use the various features and interactions of the smartphone.

### 3.2.4 Integrating Interview Results into System Design

The interview results indicated that older adults often forget learned smartphone operations due to infrequent usage and feel anxious about incorrect operations or mistakes. This suggests the need for an agent system that can repetitively explain the correct operation methods, even without any support. The difficulties in viewing small screens and using touch keyboards are also essential considerations. The system should offer larger UI elements and various alternative input methods to address these issues.

Furthermore, older adults feel confused by too many UI elements on one screen or by UIs composed of complex steps across multiple screens. The system should display only a few essential UI elements required by the user and minimize screen transitions to reduce psychological burden. Lastly, elderly users often struggle to understand technical terms or icons and adapt to new features or app updates. The system should provide simple explanations for icons and terms used in the UI and ensure that user experience does not change significantly with app updates.

Based on these requirements, we conceived an AI-based agent system that can be used on a larger screen separate from the smartphone. This system can be applied to various UIs displayed on the smartphone screen and is easily accessible to elderly users when needed. The agent extracts only the necessary UI elements from the smartphone screen and displays them on a larger screen with explanations of icons and terms. It also provides elderly-friendly input interfaces like voice commands and large buttons, enhancing user convenience.

In the interaction design of the system, we emulated the Educative Assistance approach. While the Proxy Assistance approach is more efficient and convenient, Educative Assistance plays a crucial role in building older adults' skills and confidence in using smartphones. Therefore, our designed system supports them in learning and understanding the various functions of the smartphone independently. The system simultaneously displays the mobile phone screen and the large system screen, highlighting necessary UI elements on the phone screen as if pointing with a finger. This approach emphasizes important aspects, clarifying which elements the user should focus on.

Moreover, the system provides detailed explanations for each screen and UI element, helping older adults understand the purpose and operation of each feature. For instance, when explaining how to use a specific app, the system highlights important buttons or menus and provides easy explanations for them to understand. This approach encourages the user to explore the smartphone more actively and to find and use the necessary features independently.

# Chapter 4. TOMAS: Task-Oriented Multimodal Agent System

TOMAS is a web application composed of frontend and backend components. The backend uses Puppeteer to access the mobile web, read HTML, and perform interactions like clicking and entering data. It sends questions based on the screen UI to the frontend and processes the user's natural language responses.

The backend consists of four main modules: the Screen Parser analyzes UI on the screen, the Action Planner determines the next UI interaction based on context, the Dialog Generator transforms the UI into a question-and-answer format, and the 'Action Executor performs interactions based on user responses (Figure 4.1). These modules are all implemented using the OpenAI API and operate on combinations of GPT models with various prompts. Each model parses information with prompts specifying the task description, required information, and output format (JSON or string).

## 4.1 GUI Parser

The GUI Parser identifies actions that users can perform on the screen. The Screen Parser generates natural language descriptions of the current screen's purpose and function, referencing smartphone screen HTML tags. This description serves as the current context for other modules. The UI Component Parser finds possible actions in interactive UI elements. It has four action types: Click for buttons and links, Input for forms, Select for repetitive elements, and Modify for complex UI group states. Action descriptions are provided in the format {`Action Type`} {`Element`} `to` {`Purpose`} (Figure 4.2, 4.3).

The Screen Parser uses screen descriptions, HTML code of UI components, and surrounding HTML code as inputs. This helps GPT understand the context of screen components for more accurate purpose derivation. For instance, a button's HTML code alone may not indicate its purpose, but if it's located on a payment screen and wrapped in a div with the ID `confirmPayment`, LLM can recognize the button's intended use for payment.

## 4.2 Action Planner

The Action Planner selects the next action based on the given situation: user context and action history. The user context is summarized sentences through LLM, reflecting the user's intentions and objectives. Action history includes screen descriptions and a list of performed actions, aiding in selecting the next action.

## 4.3 Dialog Generator

The Dialog Generator creates interaction questions for the next action determined by the Action Planner. Questions are based on the screen and action descriptions provided by the GUI Parser and are transformed into a format understandable to the target user using GPT. Notably, a 'volunteer explaining smartphone usage to the elderly' persona is applied to LLM to simplify complex terms and explain various examples.

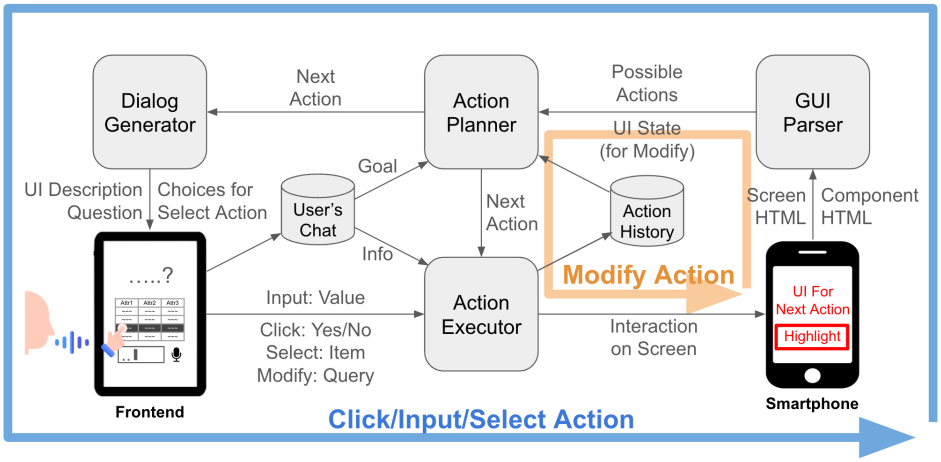


Figure 4.1: TOMAS Pipeline

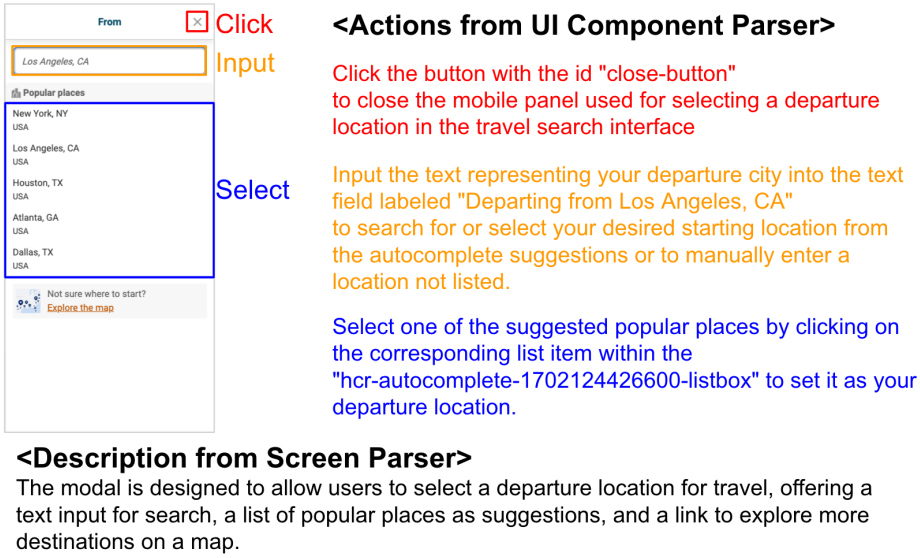


Figure 4.2: GUI Parser outputs from the departure modal of the Greyhound website

For actions like Select targeting repetitive elements, the Dialog Generator creates a JSON array of items offering choices. For instance, a date list forms a JSON array like [{"year": "2023", "month": "Dec", "day": "5"}, ...]. This array is presented in a table format on the frontend, helping users choose considering various attributes.

## 4.4 Action Executor

The Action Executor performs actions based on user responses. All actions support natural language input, with Click and Select actions implemented as 'yes/no' buttons and table format choices from the Dialog Generator, respectively. Natural language input is processed differently depending on the action type: Click actions consider user consent and Input actions attempt inputs based on information collected from the user's previous interactions. Questions are asked for additional information if information is lacking until enough is acquired.

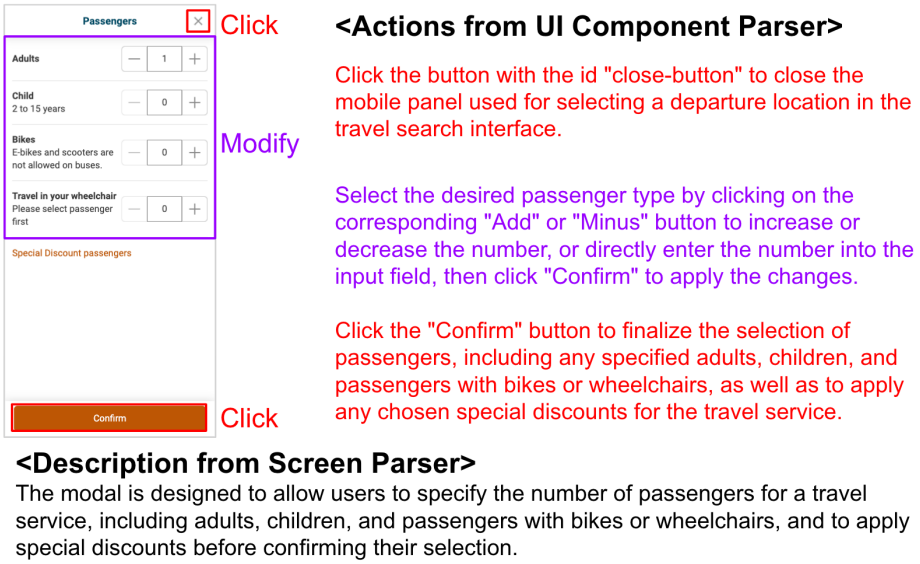


Figure 4.3: GUI Parser outputs from the passenger modal of the Greyhound website

For Select actions with natural language input, possible actions within a specific component are identified before the Action Planner decides the following action according to the user's direction. For example, when a user requests to select a specific area from a list, the system prioritizes clicking the button for the specified area within the list.

In Modify actions, GUI Parser and Action Planner, excluding Dialog Generator, are utilized. The Action Planner monitors the component's internal state of UI and continuously decides on actions. For instance, with UI elements having increase/decrease buttons for numbers, actions to click the increase button are selected until the specified number is reached. This process continues as per user input until the UI state reaches the desired value, accurately reflecting user requirements.

## 4.5 Frontend Design

The frontend is web-based, accessible from any device. Questions and choices generated by Dialog Generator are displayed to users through the GUI, and Web Speech API is used to output these questions in speech. Users can respond in natural language via voice recognition or keyboard input and interact by clicking buttons on the screen or items in tables generated by Dialog Generator. At each inquiry stage, corresponding UI elements are highlighted on the mobile screen, assisting users in easily understanding and responding to related questions. This process replicates digitally what a human assistant would do in pointing out and explaining UI elements.

# Chapter 5. User Testing

## 5.1 Methods

We conducted user testing with four elderly Korean participants to receive feedback on the system design and potential improvements. For this purpose, TOMAS was adapted to offer a movie theater ticket booking service using the Megabox mobile website. The test participants had experience using smartphones but had not previously booked movie theater tickets. Each participant engaged in testing for thirty minutes to an hour and received a compensation of 30000KRW. They first used a provided smartphone (Pixel 4XL) to book movie tickets on the Megabox website and then performed the same task using TOMAS (Figure 5.1). During the experiment, the facilitator provided appropriate assistance if participants encountered difficulties. After the experiment, the MDPQ-14[9] survey was conducted to measure the participants' proficiency with mobile devices.

## 5.2 Results

All participants were over seventy, scoring three above fifty on the MDPQ-14, similar to the median score (fifty-four) shown by elderly Americans in previous studies[8]. All participants successfully reached the payment screen on the website with minimal assistance. However, Participant P1 struggled to find UI elements, and Participant P4 misunderstood the meaning of a UI element (number of cinemas by region) at one point. Subsequently, they used TOMAS to book movie tickets.

Participants generally expressed dissatisfaction with the slow response time of the TOMAS system. The system chose one to three interactive UI elements per screen to provide information conversationally, taking one to two minutes per dialogue, resulting in up to six minutes to transition through the entire screen. This felt slower compared to the previous attempt, where they completed bookings within five to eight minutes using the Megabox website.

Participant P1 felt frustrated by the audio explanations in TOMAS, which were slower than reading written descriptions. Moreover, the participant attempted to interact with the mobile web screen displaying UI elements rather than the TOMAS screen and focused solely on TOMAS's explanations, ignoring highlights on the web screen. This suggests that having two separate screens can confuse elderly users.

Participant P2, aside from the issue of speed, gave a positive overall assessment of the user experience with TOMAS. He mentioned that providing screen explanations through TOMAS would be helpful for elderly users who struggle with smartphone screens. Additionally, P2 suggested integrating the current smartphone screen view into the frontend to aid in learning for elderly users, which could address the confusion experienced by P1. He also noted the need for more detailed information on the frontend, such as offering additional information like cast details or movie posters along with movie titles on the movie selection screen.

Participant P3, who no longer visits cinemas and only watches videos on smartphones, could not relate to the task of booking movie tickets. This suggests the need to offer a variety of practical tasks in future user studies that participants can find useful.

Table 5.1: User Testing Participants' demographic

| Participant | Nationality | Gender | Age | MDPQ-14[9] |
|---|---|---|---|---|
| P1 | Korea | Female | 73 | 53 |
| P2 | Korea | Male | 75 | 37 |
| P3 | Korea | Female | 87 | 54 |
| P4 | Korea | Female | 75 | 51 |

Participant P4 mentioned that using TOMAS helped accurately understand the previously confusing UI element (number of cinemas by region). While assessing that the explanations provided by TOMAS could benefit beginners, P4 noted that the explanations were overly lengthy and needed to be more concise and to the point.

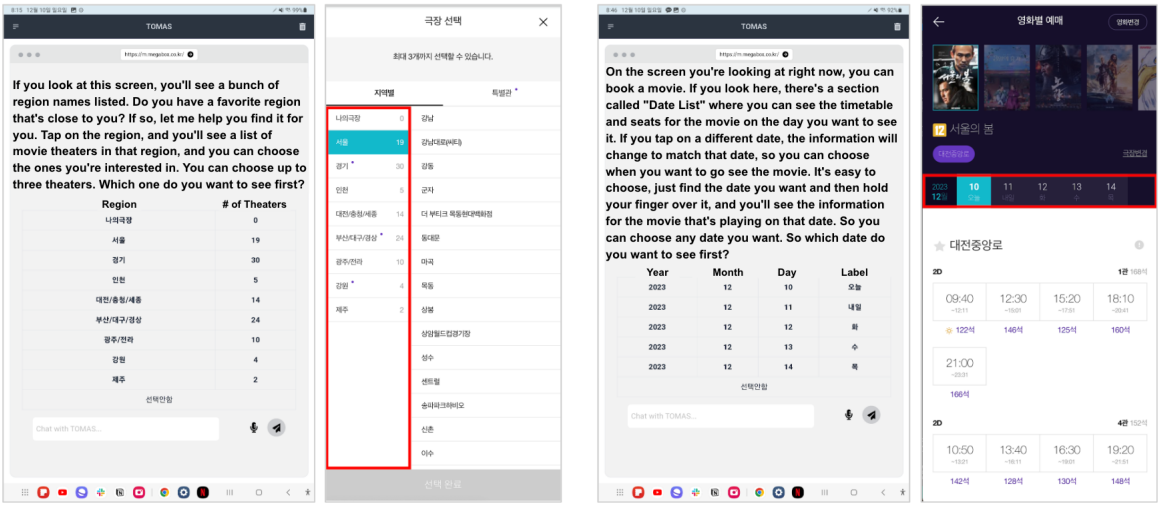


Figure 5.1: Screenshots of TOMAS adapted to Megabox website (Translated into English)

# Chapter 6. Discussion

## 6.1 Further System Improvement

According to the user testing results, one of the main issues with the TOMAS system is its speed. The GUI Parser and Dialog Generator of TOMAS generate data based on UI elements, which could have been optimized through caching. However, the Action Planner, which decides the next action based on user context and action history, has too many possible cases to make caching difficult. In fact, most of the delay in TOMAS during user testing originated from the GPT-4 based Action Planner. Nevertheless, since the range of functions used by a user within an app is limited, employing a graph-based model that predicts the next action by collecting various users' app usage patterns could reduce the time for action selection.

There is also room for improvement in TOMAS's frontend display. Currently, the system shows the frontend and the mobile webpage separately. During user testing, Participant P2 felt confused by the separation of these two screens, and Participant P3 suggested the possibility of viewing both screens together. Based on this feedback, we could consider mirroring the actual mobile screen within the frontend. Additionally, as suggested by P3, providing extra information on the frontend to aid in understanding the content is also feasible. Such improvements could enhance the user experience significantly.

Additionally, TOMAS offers a single-user experience, but options for personalization according to the elderly's situation are necessary. For example, providing options to adjust the difficulty and length of explanations could tailor the experience to the user's smartphone proficiency. Also, implementing an option to set the level of automation could realize a Proxy Assistance approach, which would help elderly users who find smartphone usage challenging to access various services more easily.

## 6.2 Further User Study

The participants in this user testing were familiar with smartphone usage and did not struggle with booking movie tickets through the Megabox website. However, there is a high likelihood that these participants do not accurately represent the actual elderly population. In fact, during the recruitment of participants from senior welfare centers, many elderly individuals hesitated to participate due to their lack of proficiency in using smartphones. Most of the participating elderly had some level of confidence in their smartphone usage. Therefore, future user studies need to include elderly individuals with varying levels of smartphone proficiency.

Additionally, including tasks in the user study that participants do not wish to perform can complicate the evaluation of the system. Thus, in the next user study, it's necessary to prepare a variety of tasks that can be performed using TOMAS, allowing participants to choose the tasks they are most interested in. In similar studies under analogous conditions, participants were given the opportunity to choose from ten different tasks.

# Bibliography


[1] National Information Society Agency. The Report on the Digital Divide. (Ministry of Science and ICT, Government of South Korea,2022)

[2] Awan, M., Ali, S., Ali, M., Abrar, M., Ullah, H. & Khan, D. Usability barriers for elderly users in smartphone app usage: an analytical hierarchical process-based prioritization. *Scientific Programming*. **2021** pp. 1-14 (2021)

[3] Latulipe, C., Dsouza, R. & Cumbers, M. Unofficial Proxies: How Close Others Help Older Adults with Banking. *Proceedings Of The 2022 CHI Conference On Human Factors In Computing Systems*. pp. 1-13 (2022)

[4] Li, Q. & Luximon, Y. Older adults' use of mobile device: usability challenges while navigating various interfaces. *Behaviour & Information Technology*. **39**, 837-861 (2020)

[5] Li, Q. & Luximon, Y. Understanding older adults' post-adoption usage behavior and perceptions of mobile technology. *International Journal Of Design*. **12** (2018)

[6] Pang, C., Collin Wang, Z., McGrenere, J., Leung, R., Dai, J. & Moffatt, K. Technology adoption and learning preferences for older adults: evolving perceptions, ongoing challenges, and emerging design opportunities. *Proceedings Of The 2021 CHI Conference On Human Factors In Computing Systems*. pp. 1-13 (2021)

[7] Tang, X., Sun, Y., Zhang, B., Liu, Z., LC, R., Lu, Z. & Tong, X. " I Never Imagined Grandma Could Do So Well with Technology" Evolving Roles of Younger Family Members in Older Adults' Technology Learning and Use. *Proceedings Of The ACM On Human-Computer Interaction*. **6**, 1-29 (2022)

[8] Sharifi, H. & Chattopadhyay, D. Senior Technology Learning Preferences Model for Mobile Technology. *Proceedings Of The ACM On Human-Computer Interaction*. **7**, 1-39 (2023)

[9] Petrovčič, A., Boot, W., Burnik, T. & Dolničar, V. Improving the measurement of older adults' mobile device proficiency: results and implications from a study of older adult smartphone users. *IEEE Access*. **7** pp. 150412-150422 (2019)

[10] Zhong, M., Li, G., Chi, P. & Li, Y. HelpViz: Automatic Generation of Contextual Visual Mobile Tutorials from Text-Based Instructions. *The 34th Annual ACM Symposium On User Interface Software And Technology*. pp. 1144-1153 (2021)

[11] Feng, S., Chen, C. & Xing, Z. Video2Action: Reducing Human Interactions in Action Annotation of App Tutorial Videos. *Proceedings Of The 36th Annual ACM Symposium On User Interface Software And Technology*. pp. 1-15 (2023)

[12] Mendel, T. & Toch, E. Meerkat: A Social Community Support Application for Older Adults. *CHI Conference On Human Factors In Computing Systems Extended Abstracts*. pp. 1-4 (2022)

[13] Jin, X., Hu, X., Wei, X. & Fan, M. Synapse: Interactive Guidance by Demonstration with Trial-and-Error Support for Older Adults to Use Smartphone Apps. *Proceedings Of The ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies.* **6**, 1-24 (2022)

[14] Yu, J., Parde, N. & Chattopadhyay, D. "Where is history": Toward Designing a Voice Assistant to help Older Adults locate Interface Features quickly. *Proceedings Of The 2023 CHI Conference On Human Factors In Computing Systems.* pp. 1-19 (2023)

[15] Wang, B., Li, G. & Li, Y. Enabling conversational interaction with mobile ui using large language models. *Proceedings Of The 2023 CHI Conference On Human Factors In Computing Systems.* pp. 1-17 (2023)

[16] Gur, I., Nachum, O., Miao, Y., Safdari, M., Huang, A., Chowdhery, A., Narang, S., Fiedel, N. & Faust, A. Understanding html with large language models. *ArXiv Preprint ArXiv:2210.03945.* (2022)

[17] Wen, H., Li, Y., Liu, G., Zhao, S., Yu, T., Li, T., Jiang, S., Liu, Y., Zhang, Y. & Liu, Y. Empowering llm to use smartphone for intelligent task automation. *ArXiv Preprint ArXiv:2308.15272.* (2023)

[18] Zhan, Z. & Zhang, A. You only look at screens: Multimodal chain-of-action agents. *ArXiv Preprint ArXiv:2309.11436.* (2023)

[19] Jiang, Y., Schoop, E., Swearngin, A. & Nichols, J. ILuvUI: Instruction-tuned LangUage-Vision modeling of UIs from Machine Conversations. *ArXiv Preprint ArXiv:2310.04869.* (2023)

[20] Huq, F., Bigham, J. & Martelaro, N. "What's important here?": Opportunities and Challenges of Using LLMs in Retrieving Information from Web Interfaces. *NeurIPS Workshop On Robustness Of Foundation Models.* (2023)

# Curriculum Vitae in Korean

이　　름: 이 상 욱

생 년 월 일: 1996년 6월 27일

## 학　　력

2012. 3. – 2014. 2.　대전과학고등학교 (2년 수료)

2014. 3. – 2020. 8.　포항공과대학교 창의IT융합공학과 (학사)

2021. 9. – 2023. 2.　한국과학기술원 전산학부 (석사)