Bridging the Digital Divide: Empowering Elderly Smartphone Users with Intelligent and Human-Centered Design in Agemate

Liangliang Chen
Tongji University
2250576@tongji.edu.cn

Yongzhen Mu
Tongji University
2251650@tongji.edu.cn

Abstract

As mobile devices become central to modern life, elderly users often struggle with their complexity, leading to digital divide. This paper explores how the integration of Human-Computer Interaction (HCI) principles and Natural Language Processing (NLP) techniques can enhance the way elderly users learn to use smartphones. To demonstrate this approach, we present *AgeMate*, a prototype mobile agent designed to support seniors in acquiring smartphone skills more intuitively and effectively. Specifically, we investigate how personalized feedback generated by large language models (LLMs), appropriate granularity in instructional content, and mechanisms for preventing and correcting user errors can contribute to more adaptive and user-friendly learning experiences for elderly users. Rather than focusing solely on system performance, our study emphasizes the instructional value of NLP-enhanced interaction: enabling step-by-step, conversational teaching tailored to users' real-time context. By analyzing usage patterns and interaction challenges, we propose design strategies that bridge the gap between accessibility and intelligent guidance to better support elderly users in digital environments.

1 Introduction

Recent advancements in LLMs have demonstrated strong reasoning abilities, enabling them to perform complex tasks such as recommendation and HCI (Chen et al., 2024). The growing availability of LLM has inspired research into mobile agents that operate directly on mobile devices, offering potential for practical AI applications in real-world scenarios (Zou et al., 2023).

Previous work on mobile agents primarily focused on automated task execution through predefined rules or access to system metadata such as XML layouts. While these approaches have enabled end-to-end execution, they often lack interactivity, flexibility, and adaptability to user needs (Zhang et al., 2023). Real-time, as discussed in prior work (Lu et al., 2025), signifies the system's ability to provide timely and responsive assistance integrates smoothly with ongoing user interactions, minimizing perceived delays to ensure a seamless experience—a capability that has received relatively little attention in agent–user collaboration, such as assisting users through educational guidance or real-time feedback during mobile app usage (Yu and Chattopadhyay, 2024).

To address this gap, we explore the intersection of LLMs, natural language processing (NLP), and human-computer interaction to investigate how a mobile agent can support both autonomous execution and user-guided interaction. Our central research question is: *How can a mobile agent be designed to help users—especially elderly or less tech-savvy ones—learn and perform tasks on mobile devices?*

To explore this, we first conduct a brief literature review on mobile agents and LLM-powered interaction systems. We then develop a prototype system, **AgeMate**, that leverages vision language model (VLM) to interpret UI screenshots, plan actions, and interact with users through either direct execution or tutorial-style guidance. AgeMate is capable of parsing multi-round LLM responses and dynamically adapting to user feedback.

Our contributions are threefold: (1) we analyze prior work at the intersection of NLP and HCI in the context of mobile agents; (2) we develop **AgeMate**, a dual-mode (text and speech IO) agent system for mobile interaction and education; and (3) we offer insights for designing LLM-powered agents that support personalized, explainable, and adaptive assistance.

Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025), pages 327–333

2 Related Works

In recent years, the development of mobile agents has advanced significantly, but most existing methods have focused on rule-based automation and limited interaction with users. For instance, some systems leverage pre-defined rules for task execution, relying on access to system-specific metadata like XML layouts (Zhang et al., 2023; Vu et al., 2023). These approaches, while functional, often struggle with providing dynamic and personalized experiences, especially when it comes to handling complex tasks that require real-time adaptation to user behavior.

One notable area that remains underexplored is the integration of more interactive mobile agents that can collaborate with users. In particular, research has highlighted the potential of mobile agents that provide educational guidance or real-time feedback to enhance user experience during mobile app usage (Yu and Chattopadhyay, 2024; Jin et al., 2022). Such collaboration could bridge the gap between pure automation and user-centered assistance, allowing for a more personalized and adaptive approach to mobile task execution.

Recent advancements in large language models (LLMs) have opened new avenues for enhancing mobile agents with greater flexibility and adaptability. Some studies have shown the potential of LLMs in processing and interpreting complex information from various sources, offering a more natural form of communication between agents and users (Zou et al., 2023; Jiang et al., 2025). However, integrating LLMs into mobile agents remains challenging due to the limitations in ensuring consistent task execution across diverse mobile environments. This gap in the literature underscores the importance of exploring ways to combine LLMs with task execution in mobile agents, as well as improving agent adaptability to dynamic, user-driven scenarios (Salman et al., 2023).

Thus, while current research has made strides in automating mobile tasks, there is a need for further exploration of more adaptable, interactive, and user-centric mobile agents that leverage advanced AI models.

3 System

System Overview AgeMate is an LLM-powered mobile assistant designed to address smartphone operation challenges for older adults. It uses GPT-4o as its underlying LLM. AgeMate features two distinct interaction modalities:

- **Auto-Execute Mode**: Completes target tasks directly through secure API integration (e.g., sending messages via WeChat)
- **Tutorial Mode**: Provides step-by-step guidance with adjustable granularity, ranging from atomic operation prompts to high-level workflow explanations

Users can choose the mode that best meets their needs—either fast, assisted execution or detailed tutorials with varying levels of detail.

Auto-Execute Mode When users want AgeMate to execute a command on their behalf, they simply enter a question into the user interface. AgeMate then initiates a multi-round execution process. In each round, the following steps are performed:

First, AgeMate captures the screenshot and XML structure of the current phone page and stores them locally. Using the XML file, AgeMate overlays a semi-transparent numbered label on each element. Once labeling is complete, the labeled screenshot is converted to Base64 encoding. Next, AgeMate concatenates a predefined prompt with the XML structure and the encoded image, and sends this composite input to the VLM.

The response is structured into sections: Observation, Thought, Action, and Summary (Yao et al., 2023). Observation is a brief, simple description of the current screen, outlining the app/page, main content, key interactive areas, and possible user actions. Thought is the agent's reasoning process, explaining the plan to address the next task step based on the screen's observation and why a specific action is chosen, including sensitivity assessment. Action specifies one of the following operations: Back, Tap, Text, Long Press, or Swipe, applied to a particular element or coordinate context. Summary is a brief internal note for the AgeMate system to record all past actions and reflect progress toward completing the task; it is not shown to the user.

Upon receiving the response, AgeMate parses it to determine the next step, such as tapping a specific coordinate or entering text into an input box. As the rounds progress, AgeMate dynamically adjusts its decisions based on the current page content to fully address the original query.

Tutorial Mode In Tutorial Mode, users receive detailed, step-by-step tutorials enabled by our human-computer interaction technology. Like

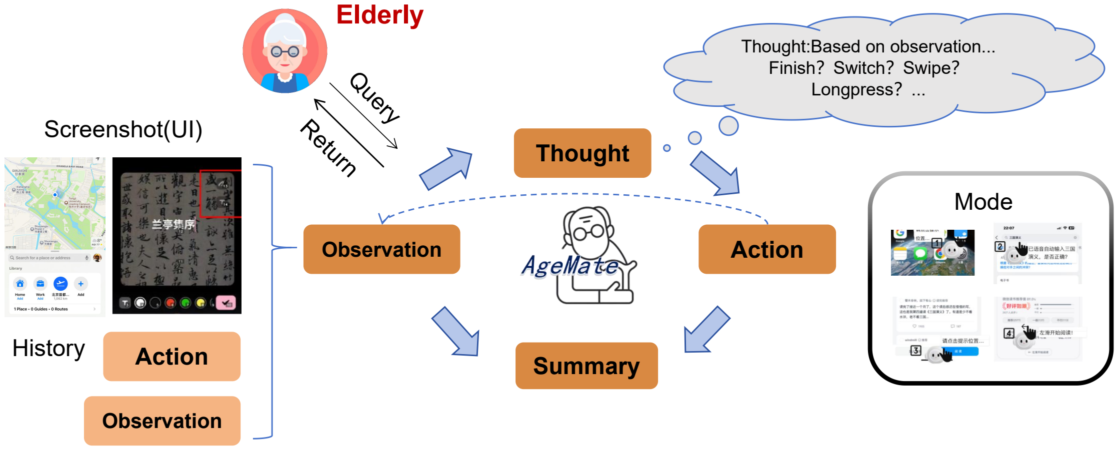

Figure 1: This figure shows how AgeMate works as a whole. The older person gives a question, AgeMate takes a screenshot of the current page in each round, and gives action as well as tutorial feedback to the older person through the four sections Observation, Thought, Action, and Summary.

Auto-Execute Mode, users begin by entering a question into the input box, and AgeMate performs a multi-round execution. However, the process diverges once AgeMate parses the LLM response.

At that point, AgeMate determines both what and where operation should be performed. It then temporarily halts the backend process and displays a box on the user's mobile interface that includes a textual description of the intended action along with its target location. Three buttons are provided within the box:

- **Confirm**: Indicates that the user will perform the action manually (e.g., tapping a specific location or entering text).
- **Auto-Execute**: Instructs AgeMate to perform the action automatically.
- **Insist**: Allows users to override the system if they believe their action is correct.

Pressing either the *Confirm* or *Auto-Execute* button resumes the backend process and advances to the next round.

In addition, AgeMate monitors the touch coordinates during Tutorial Mode. If the distance between the detected touch point and the target coordinate exceeds a predefined threshold, AgeMate considers the action incorrect, ignores it, and displays a warning by flashing the box in red with an updated message. If users still believe their action was correct, they may tap the *Insist* button to proceed independently.

Granularity of Tutorials Tutorials are provided at three levels of granularity:

1. **Observation, Thought, and Action**: For novice users who rarely use smartphones or are unfamiliar with app functionalities.
2. **Thought and Action**: For users with basic knowledge of mobile phones or the app.
3. **Action Only**: For experienced users who simply need guidance on the specific operation.

This hierarchical design allows AgeMate to cater to a diverse range of user expertise by offering the level of detail that best fits each user's needs.

Cross-Application Functionality AgeMate can execute tasks across multiple apps. The LLM response is required to decompose tasks and determine whether a user's query involves cross-application execution. Moreover, AgeMate maintains a sequential log of historical actions along with brief summaries for each step. This log enables the LLM to accurately track the task's progress and decide when to switch to another application.

Local Knowledge Bases AgeMate generates task logs for both Auto-Execute and Tutorial Modes. Each log entry includes a screenshot of the current page, the XML structure, the system's interpretation of the current task, and the actions taken. AgeMate then stores these logs in the local knowledge base. When processing a user's query, the LLM analyzes the name of the app involved and queries the local knowledge base. If a record of previous operations for that app exists, it is incorporated into the prompt as supporting information, thereby enhancing the LLM's response accuracy.

Fault Tolerance and Robustness Errors can arise from two sources: the LLM and the user. The LLM may produce incorrect responses, such as non-standardized formats or misinterpretations, while users might perform erroneous actions, such as tapping the wrong location.

To mitigate LLM-related errors, we employ hierarchical and explicit prompts in conjunction with a local knowledge base, which enhances the accuracy of the LLM's responses. For tasks that involve potential risks (e.g., deletion, payment), AgeMate issues warnings to ensure users understand the risks before proceeding. Specifically, when the system issues a warning, it creates a new page and adds it to the dialog box that displays the following text in yellow: "Attention: This is a dangerous action, please be careful!". Then the user needs to click "I know" to continue. If an error occurs in the LLM's output, users can choose to ignore the suggested operations and perform the actions manually; for severe errors, they can halt AgeMate with a single click.

To prevent user errors, AgeMate provides detailed, step-by-step prompts. When AgeMate detects an incorrect tap, it disregards the action and issues a warning. This bidirectional fault tolerance mechanism enhances the robustness of AgeMate, minimizes the risk of errors, and includes a one-click emergency stop for critical situations.

4 Discussion

4.1 Learning Theory

Our learning theory is a hybrid framework that integrates **Behaviorist Design** (Skinner, 1953), **Cognitivist Support** (Bransford et al., 2000), and **Constructivist Extension** (Vygotsky, 1978). The behaviorist component involves breaking down complex processes into multiple steps and providing users with feedback upon the completion of each step. The cognitivist approach focuses on constructing explicit cognitive schemas that simplify smartphone abstractions and reduce cognitive load by hiding nonessential information by default. Finally, the constructivist perspective is employed because older adults rely heavily on experiential learning; they can better understand tutorials that are closely related to real-life situations. Nonetheless, concerns have been raised that growing reliance on AI tools may contribute to cognitive offloading and diminished critical thinking, especially among older adults (Gerlich, 2025). While our framework aligns with age-related cognitive strengths, we acknowledge the need to prevent over-reliance. Future studies will examine how AgeMate affects users' independent cognitive engagement over time.

4.2 Aging-Friendly Design

From the early stages of design, we have prioritized the needs of older adults. We developed an age-appropriate interactive interface to enhance the visual and user experience through the use of larger fonts, simple icon layouts, and high-contrast color schemes. Regarding the interaction flow, we designed the buttons to be simple with clear, precise labels so that users can quickly grasp their functions and effectively follow the tutorial, rather than being confused by complex interactions. For example, several studies have shown that interfaces with simplified layouts and larger touch targets significantly improve usability for elderly users.

4.3 Risks and Strategies

Privacy preservation is a major concern in both academia and industry. AgeMate uploads the XML structure and a screenshot of the current mobile page, which poses a potential risk to users' privacy. There is extensive literature and practice on the trade-offs between privacy preservation and contextual understanding. Although current solutions—such as dynamic redaction, hybrid data representations, and secure split learning—help to mitigate these risks, they often rely on oversimplified heuristics that inadvertently degrade task performance. A fundamental limitation of many existing privacy mechanisms is their binary approach to privacy filtering, rather than treating privacy as a context-aware, utility-preserving process. Future work should explore semantic-aware privacy operators that dynamically adjust the granularity of redaction based on task requirements. It is crucial for the NLP community to collaborate with HCI and security experts to establish standardized benchmarks for evaluating the privacy-utility tradeoffs in real-world agent deployments, moving beyond synthetic text datasets to incorporate multimodal mobile interaction traces.

4.4 Latency of Response

Currently, AgeMate exhibits a latency of approximately 20 seconds between actions, which may affect the user experience. The primary source of this delay is the call to the LLM API. Two avenues for optimization are under consideration. The first focuses on the LLM aspect: for instance, AppAgentX abstracts repetitive low-level operations into high-level, one-click actions, thereby reducing redundant reasoning steps; Mobile-Agent-v2 achieves speed improvements through multi-agent collaboration and memory units; and Apple Intelligence employs device-side models for basic tasks, resorting to cloud-based LLMs only for more complex operations. The second avenue involves improvements in interaction technology: Galaxy AI, for example, displays floating bubbles that dynamically update the task progress, and for longer tasks, it preemptively shows a "Preparing Resources" alert. These approaches can help alleviate user anxiety during waiting periods and improve the overall user experience.

5 Conclusion

We developed AgeMate, a mobile application that explores how LLM agents can effectively teach elderly users. AgeMate not only automates execution of user queries but also offers three levels of tutorial granularity, allowing users to choose tutorials based on their understanding and willingness to explore. To address potential errors from the LLM and those made by users, we incorporate both preventative and bidirectional error correction methods through advanced LLM techniques and refined user interaction strategies, thereby enhancing robustness. For potentially hazardous actions, AgeMate issues explicit warnings to ensure user safety. Through these approaches, AgeMate delivers accurate and user-friendly tutorials that support effective teaching and learning for the elderly. In addition, we discuss the use of multimodal interaction, multi-agent collaboration within mobile agents, and the advantages of browser-based forward and backward navigation for integrating tutorials with third-party applications. Altogether, AgeMate exemplifies a promising direction for integrating recent advances in LLM agents into real-world applications. We envision closer collaboration between the NLP and HCI communities to further explore how such agents can assist elderly users and help bridge the digital divide.

6 Limitations and Ethical Concerns

6.1 Limitations

Currently, AgeMate supports both text and speech input/output. While text interaction is available, speech interaction is particularly important for seniors who may find typing or reading challenging. However, the current speech system lacks sufficient support for dialects and may lead to misinterpretations. To improve accuracy, the system should be fine-tuned to better recognize common dialects and adapt to the pronunciation habits of older adults, such as variations in rhythm and intonation.

In addition, we aim for AgeMate to support forward and backward navigation similar to a web browser. However, through our research and technical validation, we found that it is relatively easy to realize backward in the interface of third-party applications on phones (meaning applications other than AgeMate), but it is difficult to realize forward accurately because the system doesn't provide APIs similar to the “Forward Stack”, and some applications are based on security considerations such as progressing the user to modify their code, etc. In future research, we hope to explore stable and reliable ways of forwarding and backwarding for third-party applications, based on which older adults can conveniently go back and forth to view tutorials at each step of the interaction and can roll back the interface to the correct position if they make an error.

6.2 Ethical Concerns

We take screenshots, save them to local albums, and upload them to public LLM API, which may contain some private information that threatens users' privacy and security.

Acknowledgments

We thank Chen Ye for the guidance on our project, Ruishi Zou for the patient assistance with our work and the advice on the paper's formatting, Xiyan Huang for designing the teaser icon, Ruogu Lang for the improvements to the frontend, and Jiarui Guan for assisting with tasks related to our testing set. Additionally, we extend our thanks to the anonymous reviewers for their constructive feedback.

References

John Bransford, Ann Brown, and Rodney Cocking. 2000. *How People Learn: Brain, Mind, Experience, and School*. National Academy Press.

Zhiyu Chen, Jing Ma, Xinlu Zhang, Nan Hao, An Yan, Armineh Nourbakhsh, Xianjun Yang, Julian McAuley, Linda Ruth Petzold, and William Yang Wang. 2024. A survey on large language models for critical societal domains: Finance, healthcare, and law. *Transactions on Machine Learning Research*.

Michael Gerlich. 2025. Ai tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1).

Wenjia Jiang, Yangyang Zhuang, Chenxi Song, Xu Yang, and Chi Zhang. 2025. Appagentx: Evolving gui agents as proficient smartphone users. *arXiv preprint arXiv:2503.02268*.

Xiaofu Jin, Xiaozhu Hu, Xiaoying Wei, and Mingming Fan. 2022. 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(3):1–24.

Junting Lu, Zhiyang Zhang, Fangkai Yang, Jue Zhang, Lu Wang, Chao Du, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, and Qi Zhang. 2025. Axis: Efficient human-agent-computer interaction with api-first llm-based agents. *Preprint*, arXiv:2409.17140.

Hasanin Mohammed Salman, Wan Fatimah Wan Ahmad, and Suziah Sulaiman. 2023. A design framework of a smartphone user interface for elderly users. *Universal Access in the Information Society*, 22(2):489–509.

B.F. Skinner. 1953. *Science and Human Behavior*. Macmillan.

Minh Duc Vu, Han Wang, Zhuang Li, Gholamreza Haffari, Zhenchang Xing, and Chunyang Chen. 2023. Voicify your ui: Towards android app control with voice commands. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–22.

L.S. Vygotsky. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Ja Eun Yu and Debaleena Chattopadhyay. 2024. Reducing the search space on demand helps older adults find mobile ui features quickly, on par with younger adults. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–22.

Chi Zhang, Zhao Yang, Jiaxuan Liu, Yucheng Han, Xin Chen, Zebiao Huang, Bin Fu, and Gang Yu. 2023. Appagent: Multimodal agents as smartphone users. *arXiv preprint arXiv:2312.13771*.

Ruishi Zou, Zi Ye, and Chen Ye. 2023. itutor: A generative tutorial system for teaching the elders to use smartphone applications. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–3.

A Application Interface

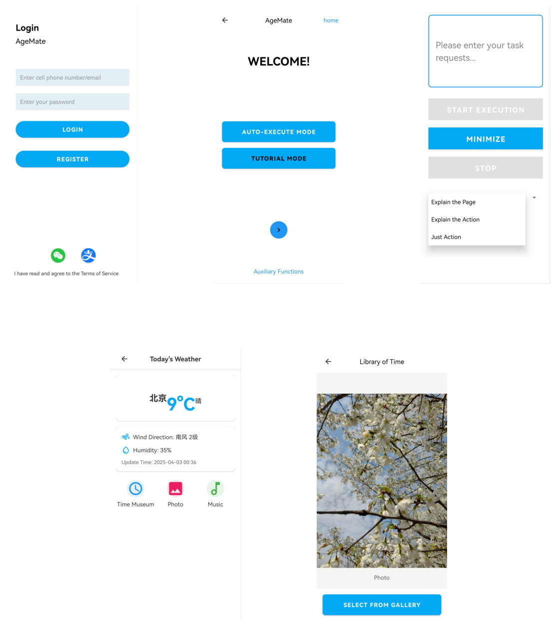

Figure A.1: Application interface screenshots

Figure A.1 shows the **Login** page, the **Home** page, the **Tutorial Mode** page, and the **Auxiliary** page of the system. The auxiliary functions include a **Time Recorder** (user can upload and save pictures), **Camera**, and **Music**, which are developed to enrich the leisure time of the elderly.

B Use Case

Round 1

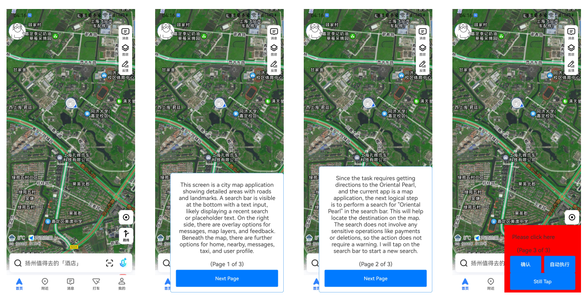

Figure B.1: Round 1: Observation and guidance; user error and AgeMate intervention.

Round 1 AgeMate gives an observation of the current page on the first page of the dialog, and then on the second page, it explains what needs to be done and why to complete the target question. On the second line, the user clicks on the wrong location, and AgeMate consumes the action, blinks the dialog box, and gives a hint. If the user insists that he or she is right, he or she can click on the Still Tap button, where the user clicks on Auto Execute, and AgeMate immediately and automatically helps the user to complete the click and enter the search screen.

Round 2

Figure B.2: Round 2: Auto entry of the destination.

Round 2 AgeMate automatically enters the name Oriental Pearl.

Round 3

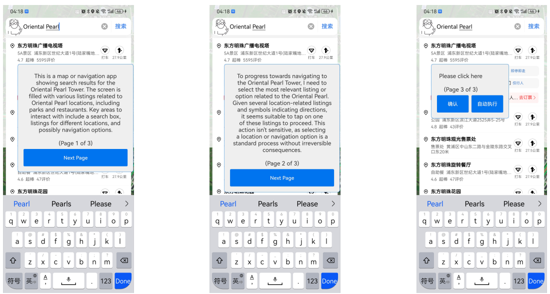

Figure B.3: Round 3: User decision-making with AgeMate's support.

Round 3 AgeMate also gives an observation of the current page on the first page, explains the action on the second page, and tells the user that they should click the search button on the third page. The user clicks on the search button and proceeds to the next round of tutorials.

Round 4

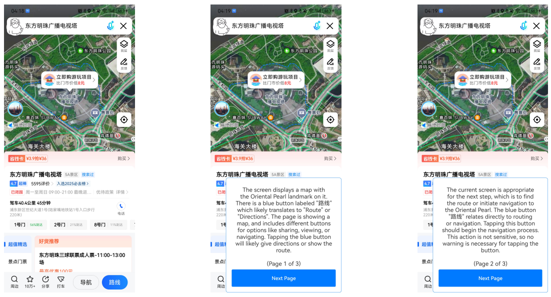

Figure B.4: Round 4: System prompting and user behavior.

Round 4 Now the map comes to the location of the Oriental Pearl. AgeMate shows page observation and task explanation again, then teaches the user to click the Nevigation button.

Round 5

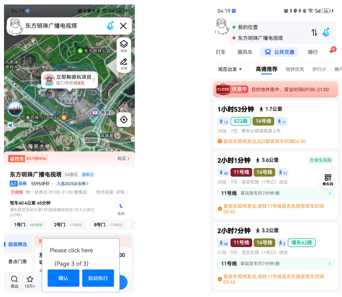

Figure B.5: Round 5: Final task execution and successful completion.

Round 5 The user clicks the Confirm button, which means the user will do this action personally. Then, the user comes to the route interface, the map application automatically selects public transportation, and the tutorial ends.