Task-Oriented Multimodal Conversational AI for Assisting Older Adults with Daily Tasks

Xiaoxin Lu*, Ranran Haoran Zhang*, Rui Zhang, Marie Boltz
The Pennsylvania State University
PennAITech Aging Focus Pilot Core

Background & Motivation

An **aging world**: over 1 billion people aged 60 and above!

An **increasing demand**: innovative age tech solutions to improve the life quality for senior people.

An **optimistic outlook**:
task-oriented AI assistant to help older adults with real-world complex daily tasks.

An **AI era**: conversational AI assistants are broadly deployed to facilitate people in all aspects.

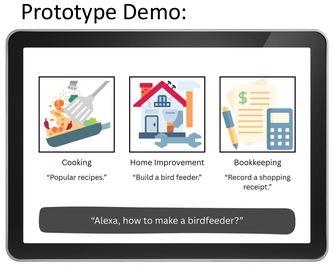
Objectives

Our aim: a **task-oriented multimodal conversational assistant** to help older adults with daily tasks spanning diverse scenarios

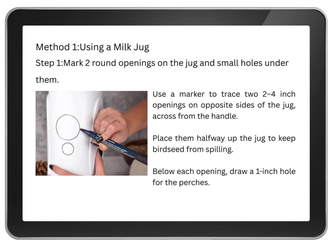
Involved scenarios:

- Online Shopping
- Meal Planning
- Home Maintenance
- Schedule Reminder

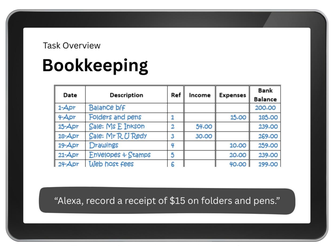
Prototype Demo:



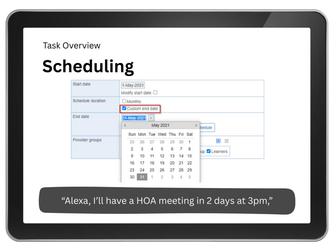
Menu: diverse scenarios



Function: task decomposition



Function: bookkeeping



Function: reminder setup

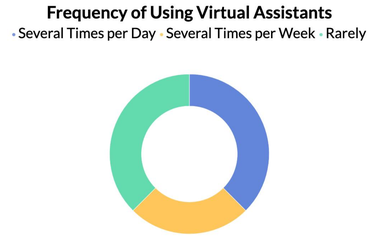
Methods & Milestones

❑ We plan to conduct our research with three specific aims:

1. Human-centered system design through diverse and inclusive survey.
2. Human-centered system development through task-oriented human-AI collaboration.
3. Human-centered system evaluation through iterative testing.

❑ Milestones

- Survey
  - **Questionnaire design**: a 2-stage survey with demographics screening and questions about participants' usage habits of electronic devices and preferences of daily tasks in need.
  - **Recruitment strategy**: 50 older adults aged 60-80 with diverse demographics (gender, socioeconomic status, mobility, etc.)
  - **Recruitment sites**: local area agencies on healthy aging and senior community centers.
  - **Polit findings**:



➢most mentioned *challenging tasks*: home maintenance; smart home devices management
➢most expected *functions*: step-by-step task decomposition; reminder
➢most needed *features*: interactive interface; data protection.

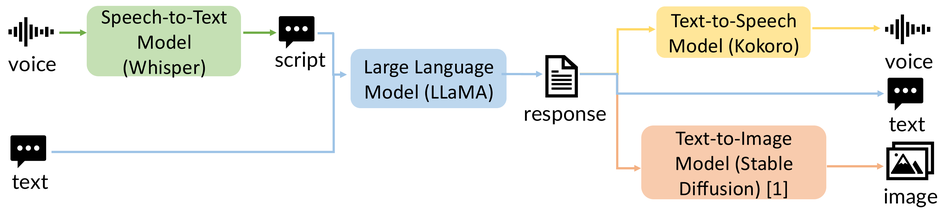
I find it hard to remember things like appointments, taking medicine, my shopping list, important dates… Really Need a reminder…

When my car breaks down, I want it to tell me what's the problem and how to fix it…

I'm worried about sharing my data with other people… Could it work locally?

❑ Framework Design:

Our interactive interface enables both text and voice as input with a Speech-to-Text model like Whisper. The Large Language Model then serves as the core module to parse user expectations and provide accordingly function calling. At last, the framework generates preferred outputs, including text, voice (with a Text-to-Speech model like Kokoro), and image (with an image generation model like Stable Diffusion model).



Our pilot study has validated the effectiveness and efficiency of the above design. With light open-sourced models, the framework can be deployed locally on users' devices while achieving a promising performance and low latency.

Under Review

[1] Lu, X., Zhang, R. H., Zhang, Y., Zhang, R. (2025). Enhance Multimodal Consistency and Coherence for Text-Image Plan Generation. *ACL Rolling Review February 2025.*

This project builds on our work in **multimodal LLM planning**, using a large language model and a fine-tuned image editing model to decompose **daily tasks** into **text-image paired steps**. We also introduce a diverse multimodal benchmark and demonstrate our framework's effectiveness through extensive experiments. This approach aligns well with the project's goal of task decomposition.

Next Steps

❑ Data collection and processing

1. Process and study the survey feedback to incorporate the unique backgrounds and preferences of older adults.
2. Amalgamate datasets from relevant domains (finance, healthcare, maintenance) to introduce expert knowledge.
3. Adopt the **WoZ** strategy to construct the dataset.

❑ System development

1. Test most recent light models as the framework **backbone.**
2. End-to-end **fine-tune** the backbone models on our dataset.
3. Apply **Direct Preference Optimization** to customize response style and accommodate users' preferences.

❑ Evaluation and refinements through iterative testing

- We have established two progressive **milestones** for our project, targeting **task completion rates of 30% and 60%**. Upon reaching each milestone and developing the corresponding prototype, we will engage potential users for evaluative feedback, ensuring human-centered refinements guide our iterative development process. Our **ultimate goal** is to achieve a **70% or higher task completion rate** and a **3.8 user satisfaction rate**.

Acknowledgments

This research was supported by the National Institute on Aging under grant P30AG073105 through the PennAITech initiative.