

# Understanding the Daily Lives of Older Adults: Integrating Multi-modal Personal Health Tracking Data through Visualization and Large Language Models

Jiachen Li, Justin Steinberg, Xiwen Li, Bingsheng Yao, Dakuo Wang, Elizabeth Mynatt, Varun Mishra

Northeastern University

li.jiachen4@northeastern.edu, steinberg.ju@northeastern.edu, li.xiwe@northeastern.edu, b.yao@northeastern.edu, d.wang@northeastern.edu, e.mynatt@northeastern.edu, v.mishra@northeastern.edu

## Abstract

Understanding the daily lives and routines of older adults is crucial to facilitate aging in place. Ubiquitous computing technologies like smartphones and wearables that are easy to deploy and scale, have become a popular method to collect comprehensive and longitudinal data for various demographics. Despite their popularity, several challenges persist when targeting the older adult population such as low compliance and hard to obtain feedback. In this work-in-progress paper, we present the design and development of a multi-modal sensing system that includes a phone, watch, and voice assistant. We are conducting an initial longitudinal study with one older adult participant over 30 days to explore how various types of data can be integrated through visualization techniques and large language models (LLMs). As a work-in-progress, we discussed our preliminary insights from the collected data, and conclude with a discussion of our future plans and directions for this research.

## Introduction

“Aging in place” (AiP) offers numerous benefits to older adults, such as fostering social connections and reinforcing a sense of identity, making it a preferred option for over 90% of seniors (Emlet and Mocerri 2012; Wiles et al. 2011; Farber et al. 2011). AiP, however, also presents distinct challenges for caregivers and researchers. As older adults experience decline in self-care and cognitive abilities, their families and friends often worry about their well-being (Li et al. 2023). This concern around “How is my [X] doing” where X could be parent, spouse, patient, neighbor and more, has driven researchers to develop technologies capable of monitoring daily activities to offer a peace of mind (Mynatt et al. 2001).

Ubiquitous computing technologies such as smartphones and wearable have the capability to passively capture a variety of health and behavioral information about an individual, e.g., mobility patterns, activity levels, socialization levels, heart-rate, etc. These technologies offer a scalable solution to capture daily dynamics of a person longitudinally over a period of time, thus enabling us to understand older adults’ daily activities. Interpreting this data, however, is challenging, particularly due to the reluctance to burden older adults with collecting self-reported ground truths. The

critical question of “How’s the older adult doing?” remains difficult to answer with raw sensing data from personal devices. Previous research has tried converting personal health information of older adults into a narrative format, but is often limited by rigid, rule-based systems that lack the ability to perform complex, dynamic sensemaking (Metaxas et al. 2007; Chung, Ozkaynak, and Demiris 2017; Shahid, Sagona, and Åhlund 2022). With the advent of Large Language Models (LLMs), there is potential to leverage their common-sense capabilities to generate insights from sensing data (Englhardt et al. 2024; Yang et al. 2024; Ferrara 2024; Ji, Zheng, and Wu 2024; Fang et al. 2024). However, most research has concentrated on generating simple classification (Ji, Zheng, and Wu 2024; Englhardt et al. 2024), rather than providing a holistic and qualitative overview of an older adult’s day that truly reflects the needs of caregivers and other stakeholders.

In this study, we explore the feasibility of generating summaries using LLM through a real-world deployment with older adults. We designed and developed a multi-modal sensing system incorporating a phone, watch, and voice assistant. We deployed it with an older adult for over a month, conducting various analyses to explore how we could summarize their daily activities using LLM and visualization techniques. Our work-in-progress project aims to bridge a significant gap in providing comprehensive, adaptive care support tailored to the nuanced daily lives of seniors.

## Data Collection

We collect: 1) a pre-study survey, 2) a passive mobile and wearable sensing system, and 3) a voice assistant for check-ins. Before the deployment, participants complete a series of surveys to capture their demographics and routines using the Social Participation Measure (SPM) (Pittsburgh-ColdStudy3 2007), Arizona Activity Frequency Questionnaire (AAFQ) (Staten et al. 2001), and Multidimensional Emotion Questionnaire (MEQ) (Klonsky et al. 2019), as well as their regular visit locations, personal hobbies, etc.

We developed a robust sensing platform to collect data from the participant’s smartphone and a study-provided Garmin smartwatch (Fig. 1). This system has been deployed in prior studies (Campbell et al. 2023; Mishra et al. 2021; Künzler et al. 2019; Mishra 2022), and can track physical activity, location, call logs, phone and app usage, Wi-Fi and Bluetooth connections from phone, as well as inter-

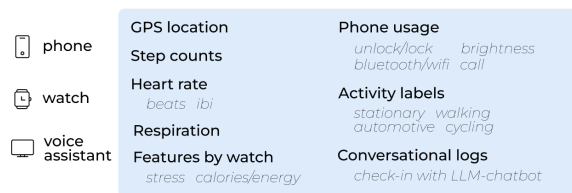


Figure 1: The multi-modal sensing system comprising a smartphone, smartwatch, and voice assistant.

beat interval (IBI) data, heart rate, raw accelerometer, and step counts from watch. The platform also generates digital biomarker features that assess aspects like mobility. The sensing system operates in the background, saving and regularly uploading data to our secure server. The watch’s battery lasts approximately 4-5 days when fully collecting data, and participants receive a reminder on their phone to charge the watch once the battery is low. Beyond passive sensing through phones and watches, we incorporated a LLM-based chatbot via Amazon Alexa designed to engage participants in casual daily conversations about their day.

### Analysis System Design

We developed a visualization system for direct representation of data with LLM-based agents for secondary inference.

#### Visualization

The visualization system is a React-based web application connected to a backend server with a MongoDB database. The system layers similar data modalities chronologically for easy comparison: heart rate and respiration data are positioned at the top, followed by phone usage details and activity levels (Fig. 2). Each modality is plotted in distinct colors and formats corresponding to its data type. Time series data are presented using scatter plots. Wi-Fi connection (0/1) is displayed using a purple line chart. Battery levels are displayed using a line chart, ranging from 0 to 100. Screen unlock periods are highlighted with rectangular overlays to indicate active phone usage. Step counts are depicted using a bar chart, alongside activities detected by the phone represented by rectangles (‘stationary’ in gray, ‘walking’ in green, ‘automotive’ in purple). Location labels derived from frequently visited addresses are color-coded as the background. Interactive features include hovering to view data points and selecting different time ranges for detailed examination.

#### LLM-Based Inference

We created a multi-agent GPT-4 system to synthesize inference from the sensing data. The **basic agent** is designed to generate hourly inferences based on multiple inputs. These inputs include a **general goal description**, **sensor data**, **data interpretation guidance**, **user persona**, and a **historical summary from GPT**. The general description outlines the overall objectives of the system. The sensor data component transforms raw sensor readings from the phone and watch into a time-ordered narrative, such as “The person’s phone is connected to a wifi named “[WIFINAME]” at

“[HH:MM:SS].” Data interpretation guidance provides expert advice on how to analyze the data, highlighting the limitations of the sensing technology to prevent over-reliance on the data. The user persona includes demographic information gathered from surveys. Lastly, the historical summary from GPT offers a recap of the data from the previous hour.

Our system returns a JSON format with four components: summary, inference, reason, and further question. ‘Summary’ provides a direct summary of the data without any inference. ‘Inference’ offers an interpretation of the data. ‘Reason’ contains justifications for the inferences drawn. ‘Further questions’ list any supplementary information that our system thinks would aid in data interpretation. In addition to the basic agent, we also deploy LLM agents that synthesize results over longer time frames, including daily and weekly summaries. One version, named ‘daily-conservative,’ compiles the ‘summary’ results from hourly agents to offer a 24-hour non-inferential summary. Another version functions similarly to the basic agent but provides daily summaries, inferences, reasons, and further questions based on the hourly inferences. We discuss these developments in detail in the preliminary results section.

### Preliminary Insights

We conducted a deployment of our sensing and data analysis systems with an older adult participant who has been diagnosed with mild cognitive impairment (MCI) and currently resides alone. We present some preliminary insights and findings derived from our multimodal system.

#### Compliance, Data Quality and Privacy

We analyzed a month’s data, from 6/25 to 7/24, and found high compliance across all modalities. The average daily phone data collection was recorded at 21.76 hours, 23.09 hours for watch ON, and 20.05 hours for wearing the watch. The average daily travel distance was 4191.39 meters calculated by the phone’s GPS, and we have valid user check-in data with Alexa for 83.3% of the days.

Although this data is from only one participant, the high compliance underscores the effective design of our sensing infrastructure, providing us with an excellent opportunity to collect and analyze the data. While the participant had a smooth experience charging the watch, they had trouble in responding to random iOS security notifications that prompt them to confirm sensing APP location access. Even though we had instructed the participant on the right way to handle these notifications, they still found them confusing, leading to gaps in the data (days 10-12) that required manual intervention by researchers. This experience highlights the importance of minimizing unexpected notifications for an older adult population to prevent confusion and ensure consistent data collection. We are also aware of the ethical considerations during data collection and removed identical data from GPS locations, self-reported conversations, and Wi-Fi/Bluetooth device names.

#### Insights From the Data

When we plotted a month’s worth of data, it became evident that most days were typical for our participant, who is

retired and does not own a vehicle. Notably, on 24 out of 30 days, the participant stayed within their neighborhood. We classified these days as typical ‘normal days’. We also focused on identifying special events that could be of interest. One such ‘regular’ event for our participant is attending church, which occurs roughly every week, as shown in Figure. 2. This event was pre-reported in the survey, including the church’s address, which allowed us to label this location accurately in our visualizations. This underscores the importance of gathering information about routine events to effectively analyze the data. Another ‘special’ event occurred on July 4th which was a national holiday. On this day, there was a noticeable decrease in the time the participant spent in the neighborhood, highlighting their participation in holiday activities outside their usual environment. This variation in routine provides valuable insights into the participant’s lifestyle and mobility, emphasizing the need to capture both regular and exceptional activities for a comprehensive analysis. For older adults, past research has identified several distinct characteristics, such as reduced physical activity, more structured routines, increased leisure time, and specific socialization patterns (Ross 1990; Schrack et al. 2014; Li et al. 2023). These characteristics were largely reflected in the data we collected, and further underscores the need for longitudinal analysis of rhythms and patterns.

### Varying Levels of Inference

We experimented with different agents to obtain varying levels of inference across different time scales(hourly/daily) and inference levels(direct summaries/inferences). These approaches generated four distinct outcomes: **detailed sensor data summaries**, integrated modalities leading to **reasonable inferences**, **high-level daily activity summaries**, and **personal lifestyle inferences** (Fig. 2).

The first level focused on capturing rich details from sensor data, including trends and patterns in time-series data. The second level is built on these details, offering reasonable inferences by integrating information across modalities. These levels often provided insights that could help transfer data to basic insights, for instance, correlating fluctuating respiration with physical activity. As we expanded to daily summaries, the minutiae of individual sensor data points became less significant, and the focus shifted to describing broader daily activities. Here, self-reported data played a crucial role in bridging gaps not covered by sensor data, making the summaries more relevant to end-users. The final level of inference delved into interpreting personal lifestyle factors such as social connections and routine regularity. These interpretations, though subjective, proved to be intriguing and were significantly informed by participants’ survey responses. For example, accurate lifestyle inferences for our first participant aligned closely with their self-reported information and survey responses, demonstrating the potential of our approach to yield personalized, meaningful insights from integrated data analysis.

**Accuracy and Issues** As a work-in-progress paper, we conducted only qualitative analysis on the summaries generated by the LLM, planning more systematic evaluations for

| Role        | Utterance  |
|-------------|--|
| [User]      | Coco, I played rummikub this morning with my friends and I actually won once. I rarely win.                                      |
| [Assistant] | That’s great! Were you playing with old friends or did you meet new ones?  |
| [User]      | ... I’ve had here for a couple of months.  |
| [Assistant] | ... Did you play any other games?  |
| [User]      | I did not. I worked on my art project. I’m doing this picture...   |
| [Assistant] | How did it make you feel to work on it?  |
| [User]      | It’s been good so far. I just hope I don’t mess it up.   |
| [Assistant] | That’s normal; mistakes are also part of the journey. Besides, often they lead to amazing outcomes. Keep going!                  |
| [User]      | That’s right, you’re absolutely right. I’ll work on it some more tonight. I think, after you encouraged me, thank you very much. |

Table 1: An example of participant’s conversation with CA.

future work. However, we have identified several insights and common issues during the study. Firstly, LLM can effectively summarize sensing data, particularly excel at processing semantic information, such as linking Wi-Fi names to location categories. It also does a great job of identifying trends for numeric data. However, challenges arose in some of the inference processes. A recurring issue is the LLM’s limited understanding of the sensing system. For example, we had to explicitly detail in the ‘Data Interpretation Guidance’ that changes in phone brightness might be system auto-adjustments, instead of always being manual changes by the participants. LLM also struggled with conflicts between different data modalities, showing an overreliance on certain data without acknowledging their limitations. For instance, discrepancies between phone and watch step counts often depended on the user’s device compliance, requiring explicit guidance for the LLM to manage these conflicts. Another challenge is the LLM’s vague boundary between evidence-based and wholly inductive inferences, sometimes making reasonable guesses without recognizing or stating that these are just one of many possible reasons. While some issues may be mitigated by adjusting the prompts, they also highlight a fundamental limitation in the LLM’s understanding of the data, necessitating further quantitative evaluation.

### Self-Report: Check-In With Conversational Agent

Our first participant also demonstrated high compliance with the conversational agent (CA), providing us with rich data for analysis (Table. 1). The CA operates casually without strict rules, and the participant began exploring various possibilities, asking questions like “what year did they start making Corvettes?”, as well as personal inquiries which the CA could not answer. We also observed that the ‘further questions’ suggested by our LLM inference system contain many pertinent queries we could pose to participants, such as asking participant’s activities at church during moments of high heart rate. Moving forward, we plan to use past summaries as scripts for the CA to ask more targeted questions.

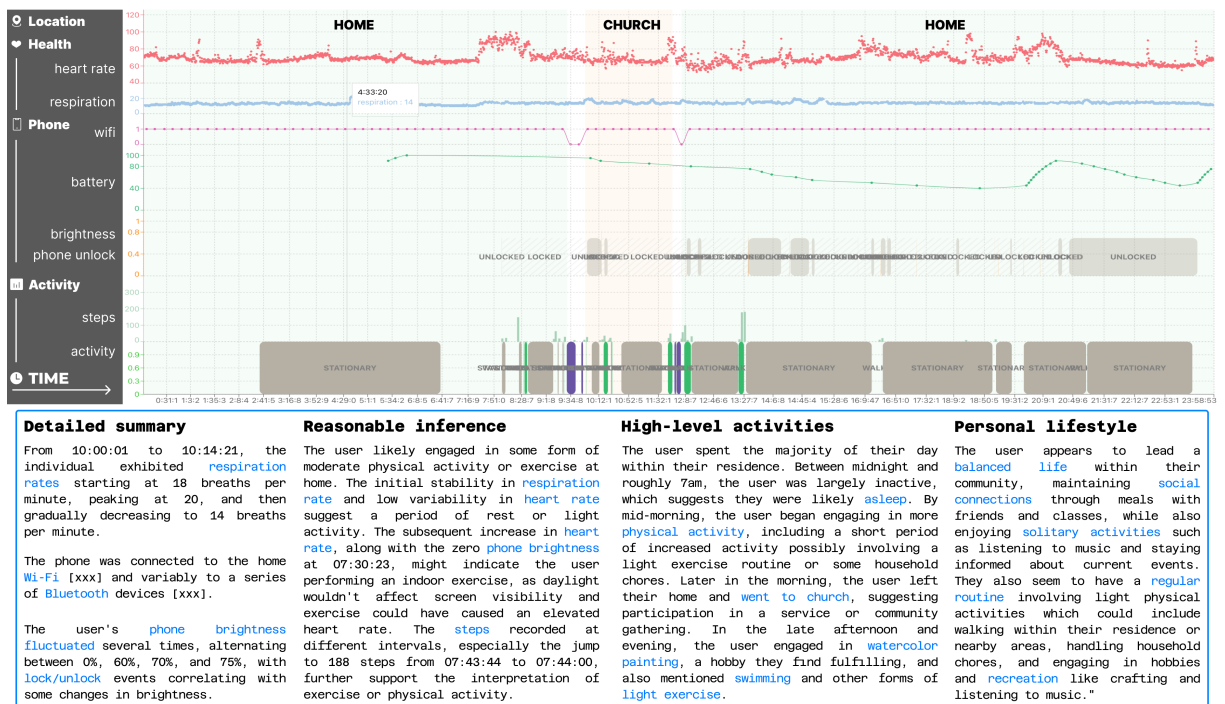


Figure 2: Top: visualization of a ‘regular day’ of the participant in 24 hours; bottom: sample of LLM inference at four levels.

These findings lead us to another consideration about the granularity of check-ins. Since we did not program the CA with specific guidelines, it often did not ask precise questions, such as the exact timing of reported events. A ‘further question’ from the LLM mentioned it would be helpful to know the duration of activities that were not captured by the sensor data. This type of information, easily obtainable through follow-up CA questions, is crucial sensors alone cannot detect it. Yet, this raises a critical question about the necessary level of detail. Self-report can compensate parts of the sensing data, but do we actually need it with that level of detail? More qualitative research with various stakeholders is required to determine the optimal granularity of information needed from participants. Another limitation of the CA is its narrow focus on certain topics, while sometimes delving too deep into one specific activity. Future iterations of the design will need to balance the granularity, breadth, and duration of check-in conversations to better accommodate the diverse aspects of participants’ daily lives.

### Future Plans

We have several other tasks outlined for future development. Firstly, regarding how to effectively embed data into the LLM, we propose a shift from a strict time-sequential approach to a more dynamic method that prioritizes detecting activities, transitions, gaps, and anomalies. Another improvement is to dynamically address the retrieval of data modalities depending on the scenario. Since LLM struggles with understanding the limitations of sensor data, such as discrepancies in step counts between a watch and a phone when the participant doesn’t carry their phone, we might

need to incorporate some pre-defined rules to account for such variations. We are currently consulting with experts to refine our understanding of sensemaking procedures, aiming to enhance the LLM’s interpretative accuracy, and conduct a comprehensive evaluation of the LLM-generated results with both experts and participants. To further validate and expand our findings, we also plan to recruit additional participants with varying lifestyles. With an increased number of participants, we expect variations in user compliance. Consequently, we plan to explore different scenarios and develop corresponding solutions to reduce participant burdens. Further, we plan to engage older adults’ and their caregivers in reflective discussions to better understand their needs and concerns. This feedback will inform the customization of our summaries, such as in terms of granularity, ensuring that our approach effectively captures the older adult’s daily life and is useful/relevant for the caregiver(s).

### Broader Impact Statement

This work-in-progress project is currently in the pilot deployment stage with one older adult, where we reflect on the process and explore different analytic methods. In the future, we plan to expand our deployment and conduct human-centered studies to gather feedback on the data summaries from various stakeholders. In this study, we aim to recruit older adults with MCI. Our contribution lies in bridging the gap between collecting raw data and generating meaningful insights that are valuable to researchers and end users. These insights, including daily summaries, are designed to provide peace of mind to care partners and other stakeholders, playing a vital role in supporting aging in place.

## References

- Campbell, C. I.; Chen, C.-H.; Adams, S. R.; Asyyed, A.; Athale, N. R.; Does, M. B.; Hassanpour, S.; Hichborn, E.; Jackson-Morris, M.; Jacobson, N. C.; et al. 2023. Patient Engagement in a Multimodal Digital phenotyping study of opioid Use Disorder. *Journal of Medical Internet Research*, 25: e45556.
- Chung, J.; Ozkaynak, M.; and Demiris, G. 2017. Examining daily activity routines of older adults using workflow. *Journal of biomedical informatics*, 71: 82–90.
- Emlert, C. A.; and Mocerri, J. T. 2012. The importance of social connectedness in building age-friendly communities. *Journal of aging research*, 2012.
- Englhardt, Z.; Ma, C.; Morris, M. E.; Chang, C.-C.; Xu, X. O.; Qin, L.; McDuff, D.; Liu, X.; Patel, S.; and Iyer, V. 2024. From Classification to Clinical Insights: Towards Analyzing and Reasoning About Mobile and Behavioral Health Data With Large Language Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2): 1–25.
- Fang, C. M.; Danry, V.; Whitmore, N.; Bao, A.; Hutchison, A.; Pierce, C.; and Maes, P. 2024. PhysioLLM: Supporting Personalized Health Insights with Wearables and Large Language Models. *arXiv preprint arXiv:2406.19283*.
- Farber, N.; Shinkle, D.; Lynott, J.; Fox-Grage, W.; and Harrell, R. 2011. Aging in place: A state survey of livability policies and practices.
- Ferrara, E. 2024. Large Language Models for Wearable Sensor-Based Human Activity Recognition, Health Monitoring, and Behavioral Modeling: A Survey of Early Trends, Datasets, and Challenges. *arXiv preprint arXiv:2407.07196*.
- Ji, S.; Zheng, X.; and Wu, C. 2024. HARGPT: Are LLMs Zero-Shot Human Activity Recognizers? *arXiv preprint arXiv:2403.02727*.
- Klonsky, E. D.; Victor, S. E.; Hibbert, A. S.; and Hajcak, G. 2019. The multidimensional emotion questionnaire (MEQ): Rationale and initial psychometric properties. *Journal of Psychopathology and Behavioral Assessment*, 41: 409–424.
- Künzler, F.; Mishra, V.; Kramer, J.-N.; Kotz, D.; Fleisch, E.; and Kowatsch, T. 2019. Exploring the state-of-receptivity for mHealth interventions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4): 1–27.
- Li, J.; Zong, B.; Cheng, T.; Li, Y.; Mynatt, E. D.; and Dhekne, A. 2023. Privacy vs. Awareness: Relieving the Tension between Older Adults and Adult Children When Sharing In-home Activity Data. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2): 1–30.
- Metaxas, G.; Metin, B.; Schneider, J.; Markopoulos, P.; and De Ruyter, B. 2007. Daily activities diarist: supporting aging in place with semantically enriched narratives. In *Human-Computer Interaction—INTERACT 2007: 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10-14, 2007, Proceedings, Part II 11*, 390–403. Springer.
- Mishra, V. 2022. *Towards Effective Delivery of Digital Interventions for Mental and Behavioral Health*. Ph.D. thesis, Dartmouth College.
- Mishra, V.; Künzler, F.; Kramer, J.-N.; Fleisch, E.; Kowatsch, T.; and Kotz, D. 2021. Detecting receptivity for mHealth interventions in the natural environment. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 5(2): 1–24.
- Mynatt, E. D.; Rowan, J.; Craighill, S.; and Jacobs, A. 2001. Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, 333–340. New York, NY, USA: Association for Computing Machinery. ISBN 1581133278.
- PittsburghColdStudy3. 2007. Social Participation Measure. ”<https://www.cmu.edu/common-cold-project/measures-by-study/psychological-and-social-constructs/social-relationships-loneliness-measures/social-participation.html>”.
- Ross, M. M. 1990. Time-use in later life. *Journal of Advanced Nursing*, 15(4): 394–399.
- Schrack, J. A.; Zipunnikov, V.; Goldsmith, J.; Bai, J.; Simon-sick, E. M.; Crainiceanu, C.; and Ferrucci, L. 2014. Assessing the “physical cliff”: detailed quantification of age-related differences in daily patterns of physical activity. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 69(8): 973–979.
- Shahid, Z. K.; Saguna, S.; and Åhlund, C. 2022. Detecting anomalies in daily activity routines of older persons in single resident smart homes: Proof-of-concept study. *JMIR aging*, 5(2): e28260.
- Staten, L. K.; Taren, D. L.; Howell, W. H.; Tobar, M.; Poehlman, E. T.; Hill, A.; Reid, P. M.; and Ritenbaugh, C. 2001. Validation of the Arizona Activity Frequency Questionnaire using doubly labeled water. *Medicine and science in sports and exercise*, 33(11): 1959–1967.
- Wiles, J. L.; Leibing, A.; Guberman, N.; Reeve, J.; and Allen, R. E. S. 2011. The Meaning of “Aging in Place” to Older People. *The Gerontologist*, 52(3): 357–366.
- Yang, B.; Jiang, S.; Xu, L.; Liu, K.; Li, H.; Xing, G.; Chen, H.; Jiang, X.; and Yan, Z. 2024. DrHouse: An LLM-empowered Diagnostic Reasoning System through Harnessing Outcomes from Sensor Data and Expert Knowledge. *arXiv preprint arXiv:2405.12541*.