Simulating Personalized Smart-Home Activity Datasets with Generative AI: A Case Study

Harditya Sarvaiya* Masaki Hasegawa* Haibo Zeng† Xinghua Gao† Na Meng†
Virginia Tech
Falls Church, VA* Blacksburg, VA†
{harditya,masakih,hbzeng,xinghua,nm8247@vt.edu}

***Abstract*—To create, evaluate, and compare different technologies of home health monitoring, people need lots of sensor data that captures (1) residents' activities of daily living (ADL) like bathing, and (2) their usage of home appliances like watching TV. Unfortunately, such datasets of real-world monitoring are quite limited and scarce, due to issues like sensor cost, technique complexity, and deployment time. Existing simulators attempt to resolve these issues, by generating synthetic data based on predefined models or interactions with users. However, they give little consideration to personas, and rarely support personalized simulation based on humans' age, lifestyles, or health conditions.**

This paper introduces our novel investigation of using ChatGPT to create usable and shareable datasets of (1) human daily activities, and (2) their usage of home appliances. Specifically, there are two phases in our investigation. First, we described personas of home residents and layouts of home appliances, in order to use ChatGPT to generate data that mimic human behaviors and schedule their usage of appliances. Second, we analyzed and visualized the generated data to check whether the data is meaningful. Our results show great promise: the daily activities of different humans match the described personas, and the simulated appliance usage resembles the typical appliance usage in real-world settings. Our work may shed light on future directions of ADL simulation. It also facilitates studies on home health monitoring, disease diagnosis, and home automation.

***Index Terms*—Generative AI, ChatGPT, simulation, activities of daily living (ADL), persona, lifestyles, personalization**

I. Introduction

Smart home is an emerging application domain of Cyber-Physical Systems [1], especially with recent advances in sensor technology, big data, and artificial intelligence. One particularly interesting scenario is home health monitoring, in which humans' activities of daily living (ADL) are captured to monitor their health conditions [2]–[4], recognize safety threats (e.g., fall detection) [5], [6], and meet people's emotional as well as psychic needs [7]. One prerequisite for the development of home health monitoring technologies is credible and sufficient sensor data that can represent the characteristics of different groups of people. Unfortunately, real-world sensor datasets are often limited and scarce for various reasons [8]. For instance, the design, implementation, and deployment of sensor networks can be challenging and costly; the lengthy time of data collection can be the biggest bottleneck for people to get enough data; home residents have privacy concerns, and refuse to get monitored by sensor networks for quite a long time or publicize their personal data for research usage.

One alternative that overcomes the limitation of sensor-data scarcity is to generate simulated smart home activity datasets, including model-based approaches and interactive approaches [8], [9]. Specifically, model-based approaches involve the specification of activity models that define the order of events, the probability of events occurring, and the time taken for each event during the performance of specific activities [10]–[15]. Interactive approaches simulate virtual environments; they support users to either manipulate virtual sensors or control embedded avatars, in order to generate activities and datasets [16]–[21].

However, these existing simulation techniques share a common limitation: they do not vary the mimicry of human behaviors with **personas** [9]. That is, their generated data do not necessarily reflect the fact that human movements and activities differ by gender, age, personality, living habit, lifestyle, and health condition. Even if the interactive approaches can be extended to intentionally inject variations when manipulating avatars, such manual injections is time-consuming and unreliable. Consequently, the resulting datasets may not represent real-world activities and can be misleading for any observation or decision based on these datasets.

In our project, we explore the usage of generative AI to simulate personalized smart-home activities. A recent work by Park et al. proposed to use generative AI in the simulation, and the resulting generative agents produce believable human behavior: the agents can sleep, cook breakfast, or head to work [22]. Inspired by this prior work, we defined two research questions for further investigation:

- **RQ1:** How well does generative AI simulate personalized ADL?
- **RQ2:** How well does generative AI simulate personalized usage of home appliances?

To explore both questions, we inspected and modified the generative agents from prior work Simulacra [22], to generate movements for humans with different personas. We then programmed Python code to extract datasets of ADL and appliance usage from the movement data. Finally, we

*The two authors contributed equally to this work.

This work was supported by Commonwealth Cyber Initiative grants CCI-PE6GLBAE and CCI-PMH7EUMV.

Fig. 1: Our simulated environment has a main room, a bathroom, and a cafe

Fig. 2: The collision map shows walls, deployed objects, and zones


```
…
"daily_plan_req": "Isabella Rodriguez goes to Hobbs
  Cafe at 8am every day, and works at the counter
  until 8pm, at which point she closes the cafe.
  Before opening the Cafe, she turns on the furnace
  and TV in the Cafe. She turns off the furnace
  after some time.",
"name": "Isabella Rodriguez",
"age": 34,
"innate": "friendly, outgoing, hospitable",
"learned": "Isabella Rodriguez is a cafe owner of
  Hobbs Cafe who loves to make people feel welcome.
  She is always looking for ways to make the cafe a
  place where people can come to relax and enjoy
  themselves. She has an active lifestyle and does
  a lot of activities.",
"currently": "Isabella Rodriguez is enjoying a cup
  of coffee in Hobbs Cafe with her customers. And
  she sometimes watches TV in the Cafe during her
  break time.",
"lifestyle": "Isabella Rodriguez goes to bed around
  11pm, waking up around 6am. She sometimes watches
  TV",
"living_area": "the Ville:Isabella Rodriguez's
  apartment:main room",
…
```


Fig. 3: A JSON snippet for Isabella's persona

analyzed the extracted datasets, and compared them with existing datasets to assess the simulation quality.

Our exploration shows that generative AI can produce promising personalized datasets of ADL and home appliance usage. With a modified version of Simulacra, we prompted ChatGPT-3.5 to generate movement data for (1) a 34-year-old woman Isabella, and (2) a 68-year-old woman Maria. By analyzing and comparing datasets of the two simulated persons, we found Maria to work less (308 vs. 658 minutes), and rest more than Isabella on average (427 vs. 83 minutes). Isabella spent more time using the following appliances: coffee machine, furnace, kitchen sink, and refrigerator, probably because they are necessary for her work in a cafe.

In this paper, we made the following research contributions:

- We novelly investigated the usage of ChatGPT-3.5, in creating personalized simulation datasets for (1) humans' ADL and (2) their usage of home appliances.
- We customized an existing approach—Simulacra—to generate movement data of avatars, and to extract datasets from the generated data.
- We applied our approach to simulate datasets for people with different personas, and got interesting results.

In the following sections, we will first introduce Simulacra [22] (Section II), as our approach is built on top of that. Then we will introduce our methodology of personalized data simulation (Section III), and experiment results (Section IV).

II. Background: Simulacra [22]

Simulacra introduces a fusion between a large language model (LLM) and computationally interactive agents in a sandbox environment, to enable believable simulations of human behavior. Specifically, the sprite-based sandbox game environment defines

1) a 2D grid-based environment map, to describe areas (e.g., houses and rooms), spatial boundaries (e.g., walls), objects (e.g., piano), and their locations (i.e., xy-coordinates),
2) a 2D grid-based collision map, to describe zones where avatars can stay to interact with defined objects, and
3) 25 unique agents with textual descriptions, sprite avatars, abilities to move and interact.

While the sandbox game environment functions as the front-end of Simulacra, the back-end server passes the sandbox information to ChatGPT-3.5 via prompts, uses LLM to periodically generate movements for agents, and adopts the generated information to control agents as well as influence their interactions. All prompts are defined to implement a generative agent architecture. Specifically, with the help of those prompts, agents first perceive their environment, and all perceptions are saved in a comprehensive record of the agent's experiences called the memory stream. Based on their perceptions, the architecture retrieves relevant memories to generate movements, form longer-term plans, and create higher-level reflections; the latter two are then integrated into the memory stream for future use.

III. Methodology

Our investigation consists of two phases: customizing Simulacra for data generation (Section III-A), and dataset extraction from the generated data (Section III-B).

A. Phase I: Customization of Simulacra for Data Generation

The original simulated environment of Simulacra includes a complex map of a village, and 25 residents in that village. Every agent is configured to start his/her day at 12 am on a specified date, to generate a movement every 10 seconds. To quickly prototype our research idea, we focused on mimicking the daily activities of a single person, and thus simplified the environment to have one apartment, one cafe, and one avatar. As illustrated in Fig. 1, the apartment includes a main room and a bathroom; the cafe is the workplace of that avatar.

- **Main Room** is the major area for personal life, providing furniture to support human activities like sleeping, reading, and putting on/off clothes.

TABLE I: An exemplar snippet of the simulated ADL dataset

Date	Time	X	Y	Room	Object/Region	Activity
2023-02-13	07:51:00	7	7	cafe	behind the cafe counter	getting ready for work (having breakfast)
2023-02-13	07:51:10	6	7	cafe	kitchen sink	getting ready for work (having breakfast)
2023-02-13	07:51:20	5	7	cafe	cooking area	getting ready for work (having breakfast)
2023-02-13	07:51:30	4	7	cafe	cooking area	getting ready for work (having breakfast)
2023-02-13	07:51:40	3	7	cafe		getting ready for work (having breakfast)
2023-02-13	07:51:50	3	8	cafe	furnace	getting ready for work (having breakfast)
2023-02-13	07:52:00	3	9	cafe	furnace	getting ready for work (having breakfast)

Empty cell means the avatar moves between labeled objects/regions

TABLE II: The allowed states for different appliances

Appliance	States
① Shower	idle, in use
② Piano	idle, in use
③ Kitchen sink	idle, in use
④ TV	turned on, turned off
⑤ Furnace	turned on, turned off
⑥ Refrigerator	open, closed
⑦ Bathroom sink	idle, in use
⑧ Coffee machine	turned on, turned off

TABLE III: An example snippet of the simulated appliance usage dataset

Date	Time	Bathroom sink	TV	Coffee machine	Kitchen sink	Furnace	Refrigerator	Piano	Shower
2023-02-13	08:11:20	idle	turned on	idle	idle	idle	closed	idle	idle
2023-02-13	08:11:30	idle	turned on	idle	idle	idle	closed	idle	idle
2023-02-13	08:11:40	idle	turned on	idle	idle	idle	open	idle	idle
2023-02-13	08:11:50	idle	turned on	idle	idle	idle	open	idle	idle

- **Bathroom** is a specialized area where the avatar can perform hygiene activities (e.g., bathing).
- **Cafe Area** is a communal space where the avatar works to provide food service and interact with customers.

In addition to ADL, we also intended to simulate people's appliance usage. Thus, we revised Simulacra to have eight appliances deployed in the above-mentioned environment: ① shower, ② piano, ③ kitchen sink, ④ TV, ⑤ furnace, ⑥ refrigerator, ⑦ bathroom sink, and ⑧ coffee machine.

To realize the customization described above, we first revised the environment map (mentioned in Section II), which is represented as xy-based grids in CSV files, to specify appliances and their locations. We also revised the collision map (i.e., CSV files) to specify regions where the avatar can move or stand to use distinct appliances. To facilitate understanding, in Fig. 2, we use gray to mark regions that are not walkable for the avatar. We use other colors to mark walkable regions; different colors imply distinct region types, such as green for cooking region and orange for TV-watching.

Furthermore, we modified the avatar's memory (i.e., JSON files), to initialize the memory of furniture/appliance deployment, and the persona. As shown in Fig. 3, in one of the JSON files we customized to define personas, we used the predefined key-value pair structure to specify a person's name, age, innate characteristics, routine, personal experience, lifestyle, and the living area. Simulacra later incorporates such information into the prompts it sends to ChatGPT, to make the LLM generate textual descriptions of human movements accordingly.

Lastly, we revised a JSON-based configuration file to start activity generation at 6 am on February 13, 2023, and stop at 12 am of the next day. This is because people typically sleep during 12 am–6 am every day. By limiting the simulation time period to the specified 18 hours, we could focus our simulation effort on more meaningful activities that help characterize the behavioral patterns of different people.

At the end of Phase I, Simulacra executes with our revision, to derive and animate a sequence of movements. Each movement is recorded with a JSON file, covering the avatar's xy-coordinate, movement description, and simulated timestamp.

B. Phase II: Dataset Extraction

From the movement data generated by Simulacra, we wrote Python scripts to extract two kinds of datasets: humans' ADL and their appliance usage.

1) Extraction of Activities of Daily Living (ADL): Our scripts enumerate all movement files, parse those files in the temporal order of their generation, and store data in a CSV file. As shown in Table I, the CSV file defines seven columns: date, time, x-coordinate of the avatar's location, y-coordinate, room where the avatar is located, object/region (i.e., furniture or appliance) where the avatar operates or stays, and activity description. The dataset includes one movement generated every 10 seconds in the simulated environment, such as *getting ready for work (having breakfast)*. Multiple consecutive entries in the dataset can share the same activity, as the avatar may take multiple actions or move around to fulfill one activity.

Five out of the seven columns were directly extracted from the original movement data. Two of the columns, **Room** and **Object/Region**, were newly created based on our analysis of the raw data. Namely, we compared the avatar's xy-coordinates with predefined coordinate scopes of rooms and objects/regions. If a coordinate falls into the scope of an object/region, our scripts fills in the object/region information; if a coordinate does not fall into any predefined scope, it means that the avatar moves between labeled objects/regions and we do not fill in any content for the **Object/Region**-column. We did similar processing for the **Room**-column.

2) Extraction of Appliance Usage: We first defined allowed states for the eight appliances mentioned in Section III-A. As shown in Table II, each appliance has two possible states: idle/in use, turned on/turned off, or open/closed, depending on the appliance type. Next, we used these appliances as keywords to parse ADL data, to identify (1) Object/Region data matching any keywords, and (2) the corresponding timestamps. For those timestamps, we set the appliance states as either in use, turned on, or open, depending on the appliance type. For the other timestamps, we set the appliance states as either idle, turned off, or closed. Table III shows an example snippet of the dataset we generated in this way.

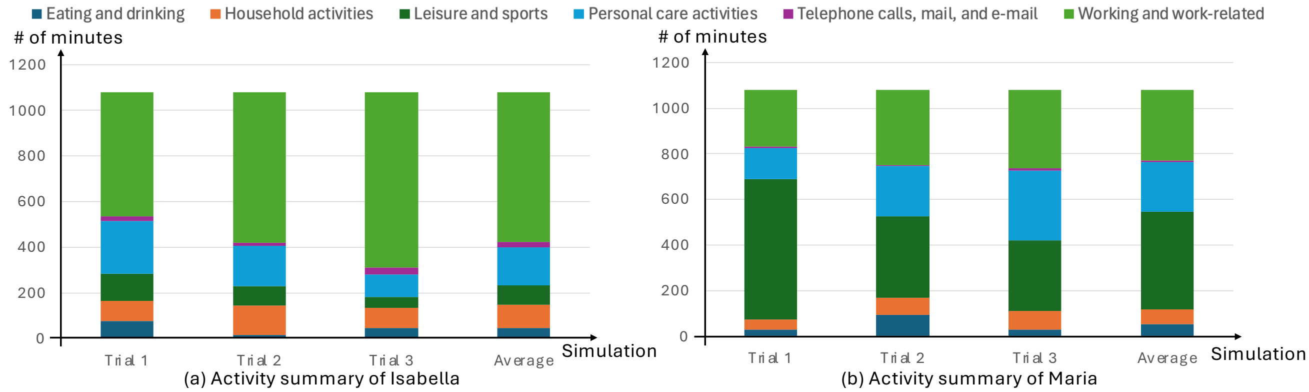

Fig. 4: The comparison of ADL datasets separately simulated for Isabella and Maria

IV. Experiment

To assess the simulation results, we conducted an experiment by generating movements for two personas: a 34-year-old woman Isabella who is active and outgoing (see Fig. 3), and a 68-year-old woman Maria who is reserved, contemplative, and gentle (see Fig. 5). We generated data for the period between 6:00 am on February 13, 2023 and 12:00 am on February 14, 2023. As there can be various differences between different runs of simulation, we ran the simulation for each person three times to observe (1) the differences between days of the same person, and (2) differences between the two people.

A. The Generated ADL Datasets

There are over 50 distinct activities generated in each simulation run, so it is almost impossible to visualize people's time-spending on each activity. To compactly present results, we used the classification method mentioned in prior work [23] to map all activities to six major categories:

C1: eating and drinking,
C2: household activities (e.g., food preparation and cleanup),
C3: leisure and sports (e.g., watching TV),
C4: personal care activities (e.g., sleep and hygiene),
C5: telephone calls, mail, and email,
C6: working and work-related.

We accumulated the time spent on each activity category for each run, and visualized the results in Fig. 4. As shown in the figure, Isabella's time spent on activities are very different among the three runs. Compared with Trial 2 and Trial 3, Trial 1 allocates less time on working but more time on the other kinds of activities. Trial 2 allocates the least time on eating and drinking (i.e., 15 minutes), but more time on household activities than the other trials. Albeit the differences, the three runs present some commonality in time allocation: they all allocate a lot of time on work-related activities (i.e., C6); they all allocate less time on remaining categories. They present similar sequential orders of different activities. Namely, according to our simulation, after waking up, Isabella typically completed her morning routine, prepared and ate breakfast, prepared to open the cafe, worked in the cafe for a long time, and enjoyed her life after closing the cafe.

Maria spent a lot of time in resting (i.e., C3), in working (i.e., C6), and on personal care activities (i.e., C4). Compared with Isabella, Maria spent a lot more time on C3 but much less time on C6. Such a comparison matches our persona description that Isabella spends a lot of time working and Maria spends much of her day resting.

Furthermore, we compared our simulated data against the real-world data reported by Americans on their daily time usage [23]. We need to exclude the comparison of C4, as our simulation is imprecise by excluding the time period 0:00 am–6:00 am when most people sleep. The comparison of other categories shows that our simulation matches the real-world data partially. It successfully simulates the facts that (1) average young women spend more time in C6, and (2) average older women spend more time in C1 and C3. However, we did not observe it to successfully simulate the comparison for C2 and C5 between the two age groups. This may be due to GPT's lack of domain knowledge about the behavioral patterns of people in different age groups, or due to the random differences presented by our limited simulation runs.

TABLE IV: The average usage time of different appliances

Appliance	Isabella's Usage (minutes)	Maria's Usage (minutes)
Bathroom sink	9	30
Coffee machine	60	55
Furnace	545	209
Kitchen sink	85	48
Piano	2	26
Refrigerator	32	26
Shower	12	6
TV	687	667

B. The Generated Datasets of Appliance Usage

To compactly present the simulated appliance usage, we accumulated the usage time of difference appliances in each run, and averaged values across runs for each person. Here, with **appliance usage**, we mean an appliance is turned on (e.g., TV), opened (e.g., refrigerator), or gets used (e.g., sink).

As shown in Table IV, both Isabella and Maria used TV for the longest time, probably because they keep the TV on for customers to watch, and for their own entertainment during the leisure time. Isabella spent more time on several other appliances than Maria, such as coffee machine, furnace, kitchen sink, and refrigerator. This may be because most of these appliances were used for work-related activities. As Isabella worked for a longer time, these appliances were used more often. Interestingly, Maria spent more time on piano,

```
"daily_plan_req": "Maria Lopez starts her day slowly at 9am.
  She goes to Hobbs Cafe at 10am every day, and works at the
  counter until 1pm, at which point she closes the cafe.
  Before opening the Cafe, she turns on the furnace and TV in
  the Cafe. She turns off the furnace after some time. She
  spends much of her day resting, reading, or doing light
  activities like knitting.",
"name": "Maria Lopez",
"first_name": "Maria",
"last_name": "Lopez",
"age": 68,
"innate": "reserved, contemplative, gentle",
"learned": "Maria Lopez is the owner of Hobbs cafe. She
  cherishes quiet moments and enjoys connecting with others
  in small, meaningful ways. She values routine and
  simplicity in her life.",
"currently": "Maria Lopez is enjoying a cup of coffee in
  Hobbs Cafe with her customers. And she sometimes watches TV
  in the Cafe during her break time",
"lifestyle": "Maria Lopez goes to bed early, around 8:30pm,
  and wakes up around 6am. She starts her day with light
  stretches, takes her medication, and spends a significant
  part of her day in restful activities such as reading or
  listening to music.",
"living_area": "the Ville:Maria Lopez's apartment:main room",
```


Fig. 5: A JSON snippet for Maria's persona

probably because our persona description goes "*She ... spends a significant part of her day in restful activities such as reading or listening to music*". Finally, our simulated shower usage matches Americans' shower habits. 66% of Americans say they typically spend 15 minutes or less in the shower, and 60% of Americans typically shower in the morning [24]. Our simulated data includes shower usage in the morning, and the average values of shower time are both less than 15 minutes.

C. Reproducibility and Variability of Simulation

To better understand the simulation results, we closely examined similarities and differences between individual simulation runs for Isabella and Maria. Isabella's routine demonstrated high reproducibility, particularly in the timing and sequence of breakfast preparation, cafe opening, and cafe closing. Specifically, Isabella prepared and ate breakfast at 7:05 am–7:40 am in Trial 1, and at 7:35 am–8:00 am in the other trials. Cafe was consistently opened at exactly 8:00 am across all three trials. Cafe closing varied slightly: Trial 1 closed the cafe earlier at 9:00 pm; Trials 2 and 3 closed the cafe at 10:00 pm.

Maria's simulation runs showed greater variability, aligning with her persona's relaxed lifestyle. She did not have breakfast in any of our simulation runs, probably because senior people typically eat less or have fewer meals [25]. Her working hours varied significantly. In Trial 1, she worked at 10:30 am–1:00 pm. In Trial 2, she worked at 10:00 am–4:00 pm, with rests in between. In Trial 3, she worked at 8:30 am–2:00 pm. Despite these differences, certain activities like her morning wellness routine (consistently involving tea and music) and knitting sessions typically in the late afternoon (around 4:00–5:15 pm), show similar timings. Maria's working hours were consistently much shorter compared to Isabella.

Although we observed some level of simulation reproducibility across runs, we also noticed discrepancies between our persona description and the generated activities. For instance, although Isabella is described to work at 8 am–8 pm every day, none of the runs have the Cafe operation time exactly match that period. Although Maria's persona describes 10 am–1 pm as the working hours, none of the runs exactly matches that description. Maria's lifestyle contains "*She starts her day with light stretches, takes her medication*". However, we only observed medication intake in the morning of Trials 1 and 2, but did not observe her medication intake in Trial 3. Such discrepancies may imply LLM's great capability of introducing noise into daily routines, to better mimic people's real life (e.g., people may sometimes forget to take medicine). However, they can also imply LLM's limitation: failing to accurately follow persona as described.

To better investigate both strengths and weaknesses of LLMs, in the future, we will generate more simulation runs for each persona and include more personas. We will conduct statistical analysis to better characterize LLM capabilities. For any limitation revealed by such analysis, we will further conduct prompt engineering to refine LLM-based simulation.

V. Related Work

The related work of our research includes LLM-generation of human-like data, and generation of human activity datasets.

A. LLM-Based Generation of Human-Like Data

Researchers proposed LLM-based approaches to generate and evaluate human-like data [22], [26]–[35]. Specifically, some researchers formulated datasets of software-related artifacts, such as coding problems from LeetCode [36], technical discussions from StackOverflow [37], and code snippets from GitHub [38]. They prompted ChatGPT to generate either solutions to coding problems [33], answers to technical questions [29], code revisions in response to maintenance requests [29], or security tests to demonstrate impacts of vulnerable code [28]. These works all compared ChatGPT's outputs against human data, to assess how well the LLM can mimic human behaviors in software engineering (SE).

Some researchers explored how well LLMs simulate humans' strategies of reasoning or decision-making [31], [32], [34]. For instance, Hamilton [31] created a GPT-based multi-agent system to simulate judicial rulings of the 2010-2016 Supreme Court of the United States. Sreedhar and Chilton [34] measured LLMs' ability to simulate strategic reasoning in the ultimatum game—a classic economics bargaining experiment. Hämäläinen et al. [32] used GPT-3 to generate questionnaire responses about experiencing video games as art.

Some researchers adopted LLMs to simulate and analyze social interactions between humans [26], [27], [35]. For instance, Park et al. [27] introduced a LLM-powered social simulacra. It takes as input the designer's description of a community's design—goal, rules, and member personas; it generates thousands of distinct community members and the social interactions with each other like posts, replies, and anti-social behaviors. Similarly, Callison-Burch et al. [26] fine-tuned an LLM to play an open-ended, dialogue-based adventure game "Dungeons & Dragons (D&D)". Omirgaliyev et al. [35] deployed LLM-powered Non-Person Characters

(NPCs) in a 2D game environment, to simulate a human-like society by expanding their village and engaging in interactions.

Our work is based on Simulacra [22], which uses LLMs to simulate day-to-day human experiences and social interactions. However, different from all prior work, we focus on the application of generative AI in the area of cyber-physical systems (CPS) instead of other domains. With Simulacra customization and data analytics, we explored how well generative AI simulates personalized datasets of ADL and appliance usage. Our results are promising: the datasets match provided personas, and partially match real-world activity patterns.

B. Generation of Human Activity Datasets

Researchers proposed various approaches to generate humans' activities of daily living (ADL) [8]–[15], [39]–[41].

To create real-world datasets, Sigurdsson et al. [39] recruited people to record short videos in their homes while they act out casual everyday activities. Vaquette et al. [40] recorded sequences of actions performed by 44 distinct people during lunch time in a realistic kitchen; the recordings last for 24–64 minutes. Liu et al. [41] sampled recordings of 106 different people to cover short videos for 120 distinct action classes (e.g., put on bag). However, these real-world datasets only include video samples with short time lengths; they are insufficient to characterize the daily life of any single person.

Model-based simulation generates activities based on user-specified models [10]–[15]. For instance, to create a simulator, Li et al. [14] first acquired real-world ADL by recruiting seven people to manually record their daily activities within three months. The authors then extracted activity attributes and values from those records to define and use an activity model. Puig et al. [13] first crowd-sourced programs for activities that happen in people's homes, via a scratch-like visual programming language. Using the collected dataset, they trained a model to generate programs from natural-language descriptions or videos of people's activities. Next, the authors implemented the most common atomic actions in the Unity3D game engine, and used the generated programs to drive an artificial agent to execute tasks in a simulated household environment. However, such approaches put heavy burdens and high requirements on model creators.

Interactive approaches simulate virtual environments based on interactions with humans [16]–[21]. For instance, Musharu et al. [21] created a tool to provide a virtual space with multiple inhabitants. Users can control avatars to perform four types of activities: cooking, dressing, toileting, and sleeping. Li et al. [20] created iGibson, a simulated environment that supports the simulation of a more diverse set of household tasks. It simulates object states (e.g., temperature), updates object status based on simulated activities (e.g., cooked), and provides virtual reality (VR) interface to immerse humans in its scenes. However, such approaches require intensive involvement of users for dataset generation.

Compared with interactive approaches, our investigation is more similar to model-based approaches, as we need to customize memory files to specify persona, such as a person's age, routine, and lifestyle. The personas described in such ways are similar to activity models, although they do not need to detail on all attribute/parameter settings. The work by Almashor et al. [30] is closely relevant to ours. The researchers customized Similacra by (1) replacing ChatGPT-3.5 with Mistra-7B, and (2) deriving household energy consumption datasets from the raw data generated by Similacra. Different from Almashor et al., we simulated another two kinds of datasets: (1) humans' ADL and (2) their appliance usage.

Different from all prior work mentioned above, our research focuses on personalized simulation of ADL and appliance usage. No prior work explores such personalized data generation.

VI. Conclusion

Our research demonstrates great potentials of generative AI, in creating comprehensive and contextually-rich personalized datasets of ADL and appliance usage. The datasets will facilitate future research in CPS, smart home automation, behavioral modeling, and home health monitoring. In the future, we will improve simulation via LLM fine-tuning, Retrieval-Augmented Generation (RAG), and multi-agent interactions. We will also experiment with diverse map layouts, environmental variations, and multiple personas to gain more insights.

Acknowledgment

We thank reviewers for their insightful comments.

References

[1] S. Munir and J. A. Stankovic, "Depsys: Dependency aware integration of cyber-physical systems for smart homes," in *ACM/IEEE International Conference on Cyber-Physical Systems*, 2014, pp. 127–138.

[2] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz, and A. Rios Rincon, "Smart homes and home health monitoring technologies for older adults: A systematic review," *International Journal of Medical Informatics*, vol. 91, pp. 44–59, 2016.

[3] J. Wang, N. Spicher, J. M. Warnecke, M. Haghi, J. Schwartze, and T. M. Deserno, "Unobtrusive health monitoring in private spaces: The smart home," *Sensors*, vol. 21, no. 3, 2021.

[4] P. Morita, K. Sundar Sahu, and A. Oetomo, "Health monitoring using smart home technologies: Scoping review," *JMIR mhealth and uhealth*, vol. 11, 04 2023.

[5] R. Alesii, F. Graziosi, S. Marchesani, C. Rinaldi, M. Santic, and F. Tarquini, "Short range wireless solutions enabling ambient assisted living to support people affected by the down syndrome," in *Eurocon*, 2013, pp. 340–346.

[6] M. C. Abeledo, F. Bruschetti, G. Aguilera, P. Iriso, M. Marsicano, and A. Lacapmesure, "Remote monitoring of elderly or partially disabled people living in their homes through the measurement of environmental variables," in *IEEE Conference on Computer Sciences*, 2016, pp. 1–5.

[7] I. Marcelino, R. Laza, P. Domingues, S. G. Meire, F. F. Riverola, and A. Pereira, "Active and assisted living ecosystem for the elderly," *Sensors*, vol. 18, 2018.

[8] J. Synnott, C. Nugent, and P. Jeffers, "Simulation of Smart Home Activity Datasets," in *Sensors*, 2015.

[9] S. N. N. Htun, S. Egami, and K. Fukuda, "A survey and comparison of activities of daily living datasets in real-life and virtual spaces," in *IEEE/SICE Symposium on System Integration*, 2023, pp. 1–7.

[10] A. Mendez-Vazquez, A. Helal, and D. Cook, "Simulating events to generate synthetic data for pervasive spaces," in *Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*, 2009.

[11] K. Bouchard, A. Ajroud, B. Bouchard, and A. Bouzouane, "Simact: A 3d open source smart home simulator for activity recognition," in *Advances in Computer Science and Information Technology*, T.-h. Kim and H. Adeli, Eds. Springer Berlin Heidelberg, 2010, pp. 524–533.

[12] S. Helal, J. W. Lee, S. Hossain, E. Kim, H. Hagras, and D. Cook, "Persim - simulator for human activities in pervasive spaces," in *International Conference on Intelligent Environments*, 2011, pp. 192–199.
[13] X. Puig, K. K. Ra, M. Boben, J. Li, T. Wang, S. Fidler, and A. Torralba, "Virtualhome: Simulating household activities via programs," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8494–8502, 2018.
[14] J.-H. Li, H.-C. Wang, and C.-T. Chan, "Adl simulation with abnormal behavior of mild cognitive impairment sufferers," in *10th International Conference on Applied System Innovation*, 2024, pp. 208–210.
[15] "SketchUp," https://www.sketchup.com/en, 2024.
[16] J. Lertlakkhanakul, J. W. Choi, and M. Y. Kim, "Building data model and simulation platform for spatial interaction management in smart home," *Automation in Construction*, vol. 17, no. 8, pp. 948–957, 2008.
[17] C. Stahl and T. Schwartz, "Modeling and simulating assistive environments in 3-d with the yamamoto toolkit," in *International Conference on Indoor Positioning and Indoor Navigation*, 2010, pp. 1–6.
[18] Q. Fu, P. Li, C. Chen, L. Qi, Y. Lu, and C. Yu, "A configurable context-aware simulator for smart home systems," in *International Conference on Pervasive Computing and Applications*, 2011, pp. 39–44.
[19] A. Roitberg, D. Schneider, A. Djamal, C. Seibold, S. Reiß, and R. Stiefelhagen, "Let's play for action: Recognizing activities of daily living by learning from life simulation video games," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8563–8569, 2021.
[20] C. Li, F. Xia, R. Martín-Martín, M. Lingelbach, S. Srivastava, B. Shen, K. Vainio, C. Gokmen, G. Dharan, T. Jain *et al.*, "igibson 2.0: Object-centric simulation for robot learning of everyday household tasks," *arXiv preprint arXiv:2108.03272*, 2021.
[21] T. Musharu and D. Vogts, "Extended sesim: A tool to support the generation of synthetic datasets for human activity recognition," in *Conference on e-Business, e-Services and e-Society*. Springer, 2022, pp. 148–163.
[22] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023.
[23] "AMERICAN TIME USE SURVEY — 2019 RESULTS," https://www.bls.gov/news.release/archives/atus_06252020.pdf, 2020.
[24] "Examining Americans' Shower Habits," https://theharrispoll.com/briefs/shower-habits/, 2024.
[25] L. Dismore, A. Sayer, and S. Robinson, "Exploring the experience of appetite loss in older age: insights from a qualitative study," *BMC Geriatrics*, vol. 24, no. 1, p. 117, 2024. [Online]. Available: https://doi.org/10.1186/s12877-024-04732-9
[26] C. Callison-Burch, G. S. Tomar, L. Martin, D. Ippolito, S. Bailis, and D. Reitter, "Dungeons and dragons as a dialog challenge for artificial intelligence," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 9379–9393.
[27] J. S. Park, L. Popowski, C. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Social simulacra: Creating populated prototypes for social computing systems," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.
[28] Y. Zhang, W. Song, Z. Ji, N. Meng *et al.*, "How well does llm generate security tests?" *arXiv preprint arXiv:2310.00710*, 2023.
[29] M. M. A. Kabir, S. A. Hassan, X. Wang, Y. Wang, H. Yu, and N. Meng, "An empirical study of chatgpt-3.5 on question answering and code maintenance," *arXiv preprint arXiv:2310.02104*, 2023.
[30] M. Almashor, Y. Miyashita, S. West, and T. V. D. Dong, "Can private llm agents synthesize household energy consumption data?" in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, 2024, pp. 664–668.
[31] S. Hamilton, "Blind judgement: Agent-based supreme court modelling with gpt," *arXiv preprint arXiv:2301.05327*, 2023.
[32] P. Hämäläinen, M. Tavast, and A. Kunnari, "Evaluating large language models in generating synthetic hci research data: a case study," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2023.
[33] N. Nascimento, P. Alencar, and D. Cowan, "Comparing software developers with chatgpt: An empirical investigation," *arXiv preprint arXiv:2305.11837*, 2023.
[34] K. Sreedhar and L. Chilton, "Simulating human strategic behavior: Comparing single and multi-agent llms," *arXiv preprint arXiv:2402.08189*, 2024.
[35] R. Omirgaliyev, D. Kenzhe, and S. Mirambekov, "Simulating life: The application of generative agents in virtual environments," in *IEEE AITU: Digital Generation*. IEEE, 2024, pp. 181–187.
[36] "LeetCode," https://leetcode.com, 2025.
[37] "StackOverflow," https://stackoverflow.com, 2025.
[38] "GitHub," https://github.com, 2025.
[39] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," in *14th European Conference on Computer Vision*. Springer, 2016, pp. 510–526.
[40] G. Vaquette, A. Orcesi, L. Lucat, and C. Achard, "The daily home life activity dataset: a high semantic activity dataset for online recognition," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 497–504.
[41] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, and A. C. Kot, "Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2684–2701, 2019.