



## **A virtual assistant dedicated to the accompaniment of the person informed of his life context thanks to the smartphone**

Florian Szczepaniak, Jérôme Boudy, Gérard Chollet, Mossaab Hariz,  
Christophe Lohr

### **► To cite this version:**

Florian Szczepaniak, Jérôme Boudy, Gérard Chollet, Mossaab Hariz, Christophe Lohr. A virtual assistant dedicated to the accompaniment of the person informed of his life context thanks to the smartphone. 16th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '23), University of Texas at Arlington, USA, Jul 2023, Corfou, Greece. pp.662-666, 10.1145/3594806.3596594 . hal-04181551

**HAL Id: hal-04181551**

**<https://hal.science/hal-04181551>**

Submitted on 16 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A virtual Assistant Dedicated to the Accompaniment of the Person Informed of his Life Context thanks to the Smartphone

FLORIAN SZCZEPANIAK, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

JEROME BOUDY, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

GÉRARD CHOLLET, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

MOSSAAB HARIZ, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

CHRISTOPHE LOHR, IMT Atlantique, Lab-STICC UMR 6285 CNRS, France

Publications in recent years describe a strong use of wearable systems (especially smartphones) in the field of Human Activity Recognition (HAR). The smartphone can be considered both as a set of sensors and as an interface for the user. The European Horizon 2020 e-VITA project aims to promote active and healthy aging in senior populations. To this end, and thanks to Information and Communication Technologies (ICT), the current work is focused on the development of a virtual assistant to help the user in his daily tasks. One of the ICT used in this project is the smartphone. Indeed, the sensors of the smartphone make it possible to measure and reconstitute a part of the user's context, thanks to the communication with a remote server. Moreover, this server is able to offer different services (including the fusion of different modalities) in order to reconstitute the context. The fusion of multimodal data implies in its turn different machine learning modules. Finally, in order to efficiently restore the context to the user and to dialogue with him, a chatbot module under the chatGPT model developed by openAI is integrated. The communication with the user will then be established in a textual or vocal way.

CCS Concepts: • **Computer systems organization** → **Real-time system architecture**; **Client-server architectures**; • **Computing methodologies** → *Neural networks*; *Classification and regression trees*; • **Human-centered computing** → **Smartphones**; • **Applied computing** → **Health care information systems**.

Additional Key Words and Phrases: e-Health, Devices, Signal Processing, Data Fusion, Machine Learning, Network

## ACM Reference Format:

Florian Szczepaniak, Jerome Boudy, Gérard Chollet, Mossaab Hariz, and Christophe Lohr. 2023. A virtual Assistant Dedicated to the Accompaniment of the Person Informed of his Life Context thanks to the Smartphone. In *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '23)*, July 5–7, 2023, Corfu, Greece. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3594806.3596594>

## 1 INTRODUCTION

The current context of IoT development [5] (Internet of Things) makes the centralization and enrichment of these data sources crucial in many fields, especially in e-Health. This trend is also confirmed when it comes to using smartphones as a sensor, more precisely as a set of sensors [5]. Indeed, smartphones are equipped with multiple onboard sensors (GPS, 3D accelerometers, magnetometers, gyroscopes, microphones, etc.): it is a fully programmable and communicating tool. In addition, its use has been greatly democratized, which facilitates the deployment of solutions based on this

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

system. The work of V.P. Cornet and R.J. Holden [5] also informs that studies on the use of smartphones are conducted in samples of diverse populations (age), with identified conditions (neurodegenerative pathologies or stress). Thus, the diversity of use cases and native smartphone functionality make this device an excellent IoT sensor for e-Health.

In the context of the Horizon 2020 program, the European and Japanese e-VITA project aims to promote active and healthy aging among senior populations. This intercultural and multidisciplinary consortium is developing a virtual assistant (or coach) to help a user in his activities of daily living (ADL). Physically integrated into the person's environment, the virtual coach benefits from a network of IoT sensors, allowing the contextualization of the user's activities. To achieve this, the latter will have the choice of equipping himself and his home with multiple sensors, such as infrared presence detectors (DeltaDore or EnOcean), door opening detectors (DeltaDore or EnOcean), a companion robot (Nao), a hologram (Gatebox), connected thermometers (NetAtmo), etc. In addition, to ensure that the assistant is always available, the user's smartphone can be added to the list of IoT sensors presented above. This set of sensors will later be referred to as the e-VITA project's IoT ecosystem.

The initiative in this paper is part of the e-VITA project ecosystem. At this stage of the experiment, only the smartphone is used as an IoT sensor. The contextualization of the user's activity is based on a quantification of the context (thanks to the IoT ecosystem) which will then be qualified (introduction of the semantic aspect). This approach is part of the data fusion research field, which can be defined as follows: *Data fusion constitutes a formal framework in which are expressed the means and techniques allowing the alliance of data coming from various sources. It aims at obtaining higher quality information; the exact definition of higher quality will depend on the application.* [16]

Numerical in nature, the formal framework can be either architectural or algorithmic, as conceptualized in the article [11]. The main difference between these two approaches lies in the layer of information centralization imposed by the architectural approach, as well as by an Edge- or Cloud-Computing type of deployment. The architectural approach allows, among other things, an asynchronous management of flows and thus, the articulation of a set of algorithmic approaches coordinated by the centralization layer. The application here is a voice assistant with knowledge about the user's context. The semantic representation of the user's life context is the valued information resulting from the data fusion.

Recent work on Transformers [15], especially for automatic text generation and conversational agents [3], are good candidates for life context generation. In order to obtain a higher quality living context, the architectural approach assumes that there are sub-information that can augment it. Thus, the study of the different information flows of the smartphone leads to the determination of these sub-contexts. The hypothesis is then that a larger number of sub-contexts, or with a higher valuation, results in a higher quality living context. The client-server mechanism, with a dedicated application on the smartphone (called *MyADL*) and a dedicated computing architecture called *Data Fusion Platform* (DFP), is the approach that has been chosen.

## 2 PROTOTYPING

### 2.1 Resources and Compatibility

The mechanism presented in the next sections works with a smartphone and a server. For the smartphone, it is a **Google - Pixel 6 Pro**. For the server, it is a **Dell - Precision 7920 Tower**, Intel Xeon Gold 6230R x2 4GHz, 64 GB RAM and NVIDIA Quadro RTX 8000. Note that the server sizing may differ from the one used here, depending on the number of users and the complexity of the analysis processing implemented. For Linux systems, with a relative complexity (maximum three services requiring more than 2 GB per use is a correct indicator) and about ten users, 16 GB of RAM, 8

cores ~3GHz are sufficient. Finally, although the smartphone chosen for the experiment is under Android, it is possible to use the application with a smartphone under iOS. Indeed, the development of the latter was carried out with the framework **Flutter** which ensures compatibility between the various platforms.

## 2.2 Smartphone

The dedicated application transmits information to the *DFP* continuously and allows the user to interact with the assistant.

The transmission concerns the inertial signals (accelerometer, magnetometer, gyroscope), the number of steps and the GPS coordinates. These data allow, through calculation and aggregation mechanisms, to determine the user's life context. Moreover, the transmission of these modalities is done in real time. Real time is defined as the processing of a quantity in a time shorter than its evolution. The first prototyped sub-context is then interested in the physical activity and the gait of the user thanks to the inertial signals [12]. The study of this real-time context imposes some specifications on the *DFP* (see section 2.3). Note that to ensure the dialogue, the transmission must be bidirectional.

The dialogue between the assistant and the user is either textual or vocal. In order to be easily usable, the functionalities *Speech To Text* (STT) and *Text To Speech* (TTS) have been added. Several SDK oriented solutions or libraries can be used for the STT part, Takenote [4], Flutter STT [6] or even the native STT of smartphones. This observation is also valid for the TTS part. The chosen solution is to use the native STT of the smartphone combined with the library Flutter TTS [7] to ensure the dialogue with the user. The different visuals are illustrated in Figure 7.

## 2.3 Data Fusion Platform

*Communication Protocol.* The first step consists of determining a communication method allowing following the evolution of the modalities that can reach an update frequency of the order of ten Hertz. Indeed, it is necessary to obtain a real-time analysis of the process. Three client-server communication protocols can be used: HTTP/1.1, HTTP/2 and WebSocket. During the experimental phase, communication via HTTP/1.1 was quickly discarded, partly because of high resource consumption on the system, and partly because of the one-way nature of the communication. Although the performance of HTTP/2 seemed interesting, especially with the gRPC framework, communication via WebSockets was chosen as a first approach in order to facilitate development. Note that both WebSockets and HTTP/2 support bidirectional communication. The latter can allow the system to easily notify the user to make the interaction dynamic.

*Collecting, Enhancing and Historizing.* The second step consists in collecting, valuing and making a history of the different modalities in order to prepare the context creation. The different modalities of the smartphone (inertial signals, GPS and number of steps) are transmitted via WebSocket communication. The GPS coordinates and the number of steps are sent to the *context broker* Fiware Orion [8] (equivalent of the centralization layer in the algorithmic approach). To track the evolution of the inertial signals and evaluate the user's gait, the data is transmitted to the Apache Kafka *message broker* [17] for consumption by the service related to the gait determination of the *DFP*. To respect centralization, this machine learning service passes the user's gait to Orion, after computation. The current entities present on Orion see their states stored in a history thanks to the integration of Fiware Cygnus.

*Aggregation and Semantic.* The third step consists of aggregating the different modalities obtained and then transforming them into a so-called semantic form. The aggregation is based on a variable temporal period, which can be adjusted to fit the user. The mechanism is based, in part, on Fiware STH-Comet allowing querying the time series, the data in the form of history, created by Fiware Cygnus. Thus, the *DFP* works largely under the Fiware standard;

```
{
  "id": "urn:ngsi-v2:ContextAggregation:0000-0000-0000",
  "type": "ContextAggregation",
  "username": {"type": "Str", "value": "Albert"},
  "age": {"type": "Int", "value": 70},
  "delta_hours": {"type": "Int", "value": 24},
  "avg_steps_count": {"type": "Int", "value": 2000},
  "avg_gait": {"type": "Dict", "value": {"sit": 0.3, "lay": 0.65, "walk": 0.05}},
  "move_outside_home": {"type": "Int", "value": 3},
  "avg_time_outside_home": {"type": "Float", "value": 0.3}
}
```

Fig. 1. Example of an Aggregation Entity

```
{
  "id": "urn:ngsi-v2:ContextAggregationSemantic:0000-0000-0000",
  "type": "ContextAggregationSemantic",
  "description": {"type": "Str", "value": "The following is a description of living habits. You are a fair and honest wellness coach and you have to write a report about the following facts. The person is named Albert and he is 70 years old. Since last 24 hours he has spent 72 minutes being physically active. Since last 24 hours his most common position has been to lay for 15.6 hours. Since the last 24 hours he has made 6000 steps. Since last 24 hours he has left his habitation for 30% of his time. Since last 24 hours he has left and returned to his habitation 3 times."}
}
```

Fig. 2. Example of a Semantic Entity

in fact, the communication between the services of the DFP and the access to the data are standardized according to the NGSI-v2 norm. The Figure 1 proposes a representation of the aggregation entity. The semantic form, on the other hand, is the result of the creation of a procedural text from the various fields available in the aggregated entity, a representation of the semantic form entity is proposed in Figure 2.

*Generate the User's Context.* The fourth and last step consists of enhancing the previous entity (Figure 2). The treatment process of this entity is carried out by transformers, in particular the **Generative Pre-trained Transformers** (GPT). Currently, it is the GPT-3 model of openAI [2]. To accelerate the development phase, the *DFP* integrates a query module to the ChatGPT API of openAI. Nevertheless, this solution is part of the use of a cloud computing solution, which is not desirable with regard to RGPD considerations (see [9]). The new entity is then a characteristic of the user's life context on the aggregation period chosen beforehand. This feature will be used in the chatbot (produced by the GPT-3 model) to provide the most relevant support to the user. Finally, it is the last step of the multimodal data fusion, allowing moving from a quantitative approach (data from sensors and sensor data fusion) to a qualitative approach on the user's life context.

The study of these different needs led to the determination of the structure of the DFP architecture, described in 4. The various components are arranged according to the Fiware, Data and Collection layers. The Fiware layer brings together all the services from the Fiware ecosystem, with the exception of the Wilma PEPs, which can be present in all the layers. The Collection layer encompasses the various services involved in data acquisition and communication with the platform. Finally, the Data layer brings together services used to enhance the value of the information.

```
{
  "id": "urn:ngsi-v2:ContextAwareness:0000-0000-0000",
  "type": "ContextAwareness",
  "characteristic": "Albert, a 70-year-old man, has been living a fairly sedentary lifestyle in the last 24 hours. He has spent 72 minutes being physically active, 15.6 hours laying down, and only 5% of his time outside of his habitation. He has taken 2000 steps and left and returned to his habitation 3 times. As his wellness coach, it is important to note that Albert's current lifestyle is not conducive to good health. He needs to increase his physical activity and reduce the amount of time he spends laying down. Additionally, he should aim to spend more time outside of his habitation, as this can help to improve his overall wellbeing. \n\nIt is recommended that Albert increase his physical"
}
```

Fig. 3. Example of an Informed Context Entity

### 3 RESULTS AND ANALYSIS

#### 3.1 Gait Analysis

All model learning performed in this section uses the KU-HAR database [14]. The latter has been reworked to match the desired usage and is then limited to 5 labels considered sufficient (lying, running, sitting, standing, walking) for real-time gait determination. The deep learning models (Multi Layer Perceptron - MLP) have been developed with Keras & Tensorflow and the random forests (RF) have been developed with the NVIDIA Rapids project libraries. Thus, regardless of the model chosen, it is possible to have a code execution on GPU.

The results are presented in Figure 5. Note that the number of users is computed from the general computation frequency ( $f_{gen} = \frac{437269}{\text{execution time}}$ , 437269 is the number of test features), the sending frequency ( $f_{sending} = 10$  Hz), the time allocated to determine the activity in the streaming processing mechanism ( $t_{calc} = 0.05$  s) and the number of data per batch ( $data_{batch} = f_{sending} \cdot batch_{time}$ , with  $batch_{time} = 0.3$  s). Thus, the number of users is given by:  $n = \frac{f_{gen} \cdot t_{calc}}{data_{batch}}$

In terms of execution time, RF offers an efficient way to determine the user's gait in a continuous way, even for a large number of users. Nevertheless, it is clear that the MLP approach seems to be more efficient in terms of performance, but for 3 performance points more the execution time is multiplied by 10. Finally, the increase in the number of estimators only slightly increases the accuracy at the expense of an execution time that keeps increasing. In sum, the most relevant model for real-time gait analysis is the RF approach with 100 estimators.

Nevertheless, the model is only the valuation step in the process of analyzing the approach. As specified in section 2.3 the information must be collected, valorized and put in the form of history. The collection is done by sending information from the smartphone via the WebSocket channel to the collection service of the FDP. This service accepts a finite number of connections, depending on the load of the established connections. With regard to the different data transmitted, only the inertial signals can be considered. Thus, for a 2-second recording of inertial signals (accelerometer, gyroscope), with a sampling frequency of 100 Hz, transmitted at a frequency of 10 Hz, the occupancy rate of the WebSocket service is of the order of 10%, i.e. 10 users per service. In terms of exchanged packet size, it is about 200 kB which are exchanged every 0.1 s. The latency of the system would deserve a precise study, but 1 second is an upper limit, which seems correct to us. The frequency of the user's gait signal is in the order of ten hertz.

#### 3.2 Determination of the Contexts

The determination of contexts is the result of a three-step process. First, the aggregation of the various data of interest during a defined period of time must be performed. The data of interest is defined as the data that adds value to the context determination. It is the tracking of the gait, the number of steps, its movements, represented in Figure 1, to

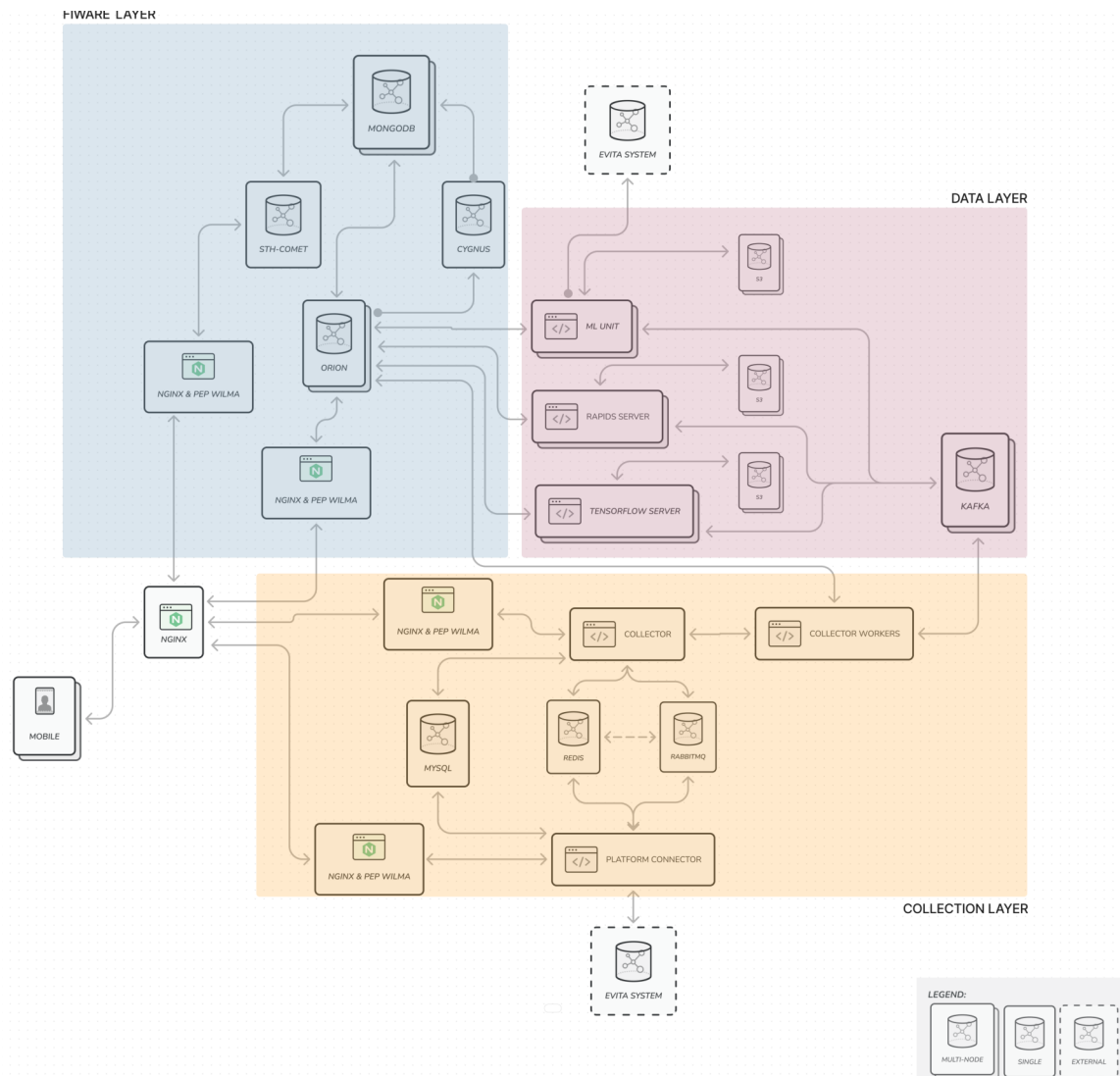


Fig. 4. DFP architecture

Model	Accuracy	Execution Time	Users
RF-100	91.1	1.734 s	≥ 8000
RF-300	91.3	2.612 s	≥ 5000
RF-600	91.4	4.772 s	≥ 2000
MLP-50	94.2	18.413 s	≥ 800

Fig. 5. Machine Learning Performances

```
{
  "id": "urn:ngsi-v2:ContextAwareness:0000-0000-0001",
  "type": "ContextAwareness",
  "characteristic": "Albert, a 70-year-old man, has been living a healthy lifestyle over the past 24 hours. He has taken 2000 steps and has left his habitation for 5% of his time. He has left and returned to his habitation 3 times in the last 24 hours. Overall, Albert is making good progress towards a healthier lifestyle. He is taking an adequate amount of steps and is getting out of his habitation regularly. However, he could benefit from increasing the amount of time he spends outside of his habitation. This would help him to get more exercise and fresh air, which can help to improve his overall health and wellbeing."
}
```

Fig. 6. Example of an Informed Context Entity

which is added information such as the first name and the age. Then by listing the various attributes, the semantic form of this entity is generated, represented in Figure 2. Finally, this text is transmitted to the API of the chatGPT in order to obtain a representation of the user's life context, represented in Figure 3.

In order to validate the starting hypothesis, the method consists of comparing the contexts from different semantic forms. The two semantic forms compared are the one in Figure 2 and the same one with the gait information removed. In the first case (Figure 3), the user had a rather unfavorable life context for his health while in the second case (Figure 6), the latter seems rather favorable. Indeed, the information related to gait shows that the user is most often lying down and does not allocate much time to active gait (walking, running, etc.). Thus, the context of the user's life is closer to reality thanks to the use of sub-contexts such as the gait context. Finally, this characterization is integrated in the service related to the dialogue so that the chatbot is informed by this information.

### 3.3 Vocal Assistant - MyADL

The vocal assistant is a feature of the MyADL application. This application works in conjunction with the *DFP*. The latter takes advantage of the ambivalent character of the smartphone, sensor / human-machine interface (HMI). Indeed, as described previously, the objective is first to transmit information. However, the creation of an HMI allows an incentive to use the complete system and thus a more present accompaniment of the assistant. Thus, *MyADL* is composed (see Figure 7) of a homepage summarizing the important information, a dashboard to ensure transparency between the collected and transmitted data, and the chatbot module for the assistant. Although the quality of the voice synthesis of the TTS module is often monotonous, the verbal or textual conversation with the chatbot is functional.

## 4 CONCLUSION

In conclusion, the architectural approach of data fusion having as data source a smartphone hosting multiple actimetric sensors is a way to identify the user's living context. Data enrichment, through gait analysis, shows that the study of sub-contexts is a functional method to improve the quality of the generation of this life context. Inspired by the lambda architecture paradigm [1], the *DFP* is able to track the user in his activities in real time and aggregate the results over time periods. Finally, the STT and TTS support of *MyADL* allows a dialogue with the assistant and offers a set of functionalities to the user. Thus, thanks to the smartphone, it is possible to follow the activity with the perspective of determining the context of the user while being a voice assistant dedicated to the person.

For the context study or for the chatbot, it is the chatGPT API. The use of integrated transformers *DFP* could allow a better management of RGD issues. Moreover, fine-tuning [10](fine-tuning of models in order to specialize) on models like GPT-2 [13] could result in the generation of better life contexts and an edge computing oriented platform. Still in



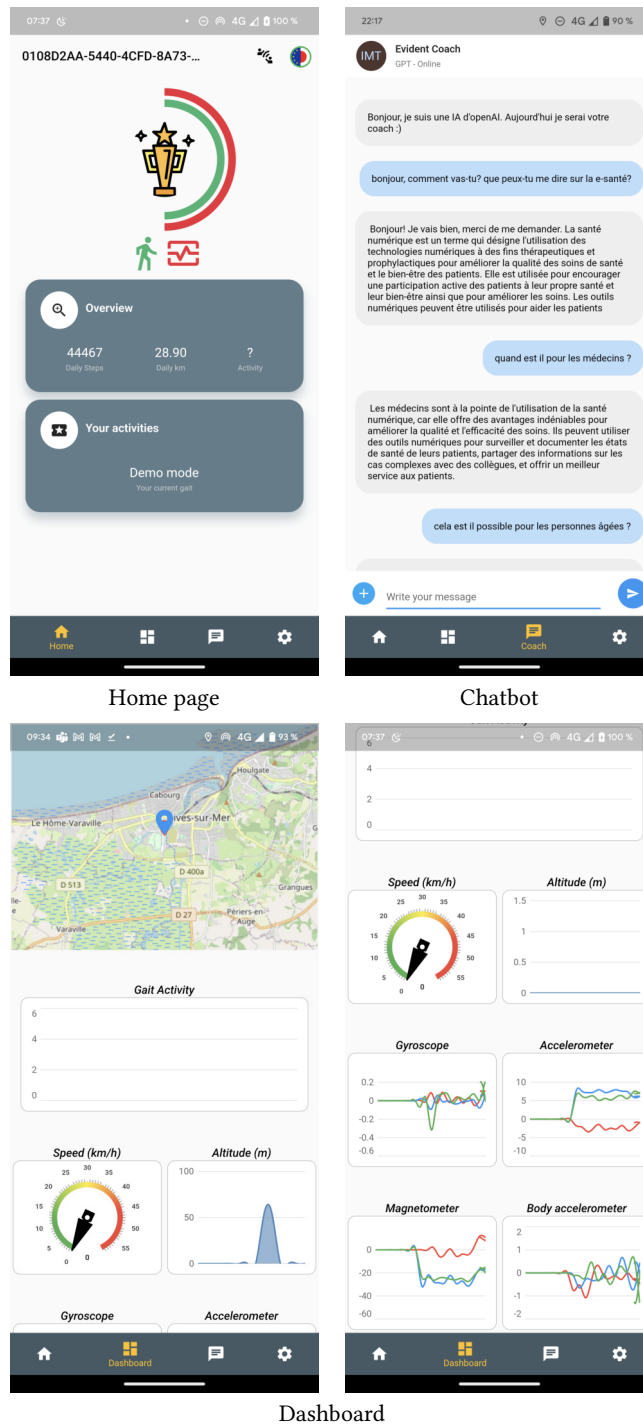


Fig. 7. MyADL Views

this improvement perspective and considering the IoT ecosystem of the e-VITA project, many sensors such as those embedded in the home can be new sources giving rise to new sub-contexts to study.

## REFERENCES

- [1] Rubayyi Alghamdi and Martine Bellaiche. 2023. An ensemble deep learning based IDS for IoT using Lambda architecture. *Cybersecurity* 6, 1 (March 2023), 5. <https://doi.org/10.1186/s42400-022-00133-w>
- [2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates, Inc., 1877–1901. <https://papers.nips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>
- [3] Guendalina Caldarini, Sardar Jaf, and Kenneth McGarry. 2022. A Literature Survey of Recent Advances in Chatbots. *Information* 13, 1 (Jan. 2022), 41. <https://doi.org/10.3390/info13010041> Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- [4] Verbit Company. 2023. *Takenote*. <https://takenote.co/the-best-speech-to-text-software-the-ultimate-guide>
- [5] Victor P. Cornet and Richard J. Holden. 2018. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics* 77 (Jan. 2018), 120–132. <https://doi.org/10.1016/j.jbi.2017.12.008>
- [6] Corner Software Development Corp. 2023. *Flutter STT*. [https://pub.dev/packages/speech\\_to\\_text](https://pub.dev/packages/speech_to_text)
- [7] Corner Software Development Corp. 2023. *Flutter TTS*. [https://pub.dev/packages/flutter\\_tts](https://pub.dev/packages/flutter_tts)
- [8] Alireza Hassani, Alexey Medvedev, Arkady Zaslavsky, Pari Delir Haghighi, Prem Prakash Jayaraman, and Sea Ling. 2019. Efficient Execution of Complex Context Queries to Enable Near Real-Time Smart IoT Applications. *Sensors (Basel, Switzerland)* 19, 24 (Dec. 2019), 5457. <https://doi.org/10.3390/s19245457>
- [9] Bastien L. [n. d.]. ChatGPT et vie privée : ne confiez surtout pas vos secrets à l'IA ! Voici pourquoi. <https://www.lebigdata.fr/chatgpt-vie-privee> Section: Intelligence artificielle.
- [10] Jieh-Sheng Lee and Jieh Hsiang. 2019. *Patent Claim Generation by Fine-Tuning OpenAI GPT-2*.
- [11] Michael McTear, Kristiina Jokinen, Mirza Mohtashim Alam, Qasid Saleem, Giulio Napolitano, Florian Szczepaniak, Mossaab Hariz, Gérard Chollet, Christophe Lohr, Jérôme Boudy, Zohre Azimi, Sonja Dana Roelen, and Rainer Wieching. 2023. Interaction with a Virtual Coach for Active and Healthy Ageing. *Sensors* 23, 5 (Jan. 2023), 2748. <https://doi.org/10.3390/s23052748> Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.
- [12] Ayan Kumar Panja, Adityar Rayala, Abhay Agarwala, Sarmistha Neogy, and Chandreyee Chowdhury. 2023. A hybrid tuple selection pipeline for smartphone based Human Activity Recognition. *Expert Systems with Applications* 217 (May 2023), 119536. <https://doi.org/10.1016/j.eswa.2023.119536>
- [13] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. [n. d.]. Improving Language Understanding by Generative Pre-Training. ([n. d.]).
- [14] Niloy Sikder and Abdullah-Al Nahid. 2021. KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognition Letters* 146 (June 2021), 46–54. <https://doi.org/10.1016/j.patrec.2021.02.024>
- [15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. <http://arxiv.org/abs/1706.03762> arXiv:1706.03762 [cs].
- [16] L. Wald. 1999. Definitions and Terms of Reference in Data Fusion. *International Archives of Photogrammetry and Remote Sensing* 32 (1999).
- [17] Roman Wiatr, Renata Słota, and Jacek Kitowski. 2018. Optimising Kafka for stream processing in latency sensitive systems. *Procedia Computer Science* 136 (Jan. 2018), 99–108. <https://doi.org/10.1016/j.procs.2018.08.242>

Received 1 April 2023