



# **AMBIENT 2024**

The Fourteenth International Conference on Ambient Computing, Applications,  
Services and Technologies

ISBN: 978-1-68558-185-5

September 29 - October 03, 2024

Venice, Italy

## **AMBIENT 2024 Editors**

Hiroshi Tanaka, Kanagawa Institute of Technology, Japan

Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain

# AMBIENT 2024

## Forward

The Fourteenth International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2024), held between September 29<sup>th</sup>, 2024, to October 3<sup>rd</sup>, 2024, in Venice, Italy, continued a series of international events devoted to a global view on ambient computing, services, applications, technologies and their integration.

On the way to a full digital society, ambient, sentient and ubiquitous paradigms lead the torch. There is a need for behavioral changes for users to understand, accept, handle, and feel helped within the surrounding digital environments. Ambient comes as a digital storm bringing new facets of computing, services and applications. Smart phones and sentient offices, wearable devices, domotics, and ambient interfaces are only a few of such personalized aspects. The advent of social and mobile networks along with context-driven tracking and localization paved the way for ambient assisted living, intelligent homes, social games, and telemedicine.

We take here the opportunity to warmly thank all the members of the AMBIENT 2024 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to AMBIENT 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the AMBIENT 2024 organizing committee for their help in handling the logistics of this event.

We hope that AMBIENT 2024 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in the field of ambient computing, applications, services and technologies.

### **AMBIENT 2024 Chairs**

#### **AMBIENT 2024 Steering Committee**

Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain

Kenji Suzuki, Tokyo Institute of Technology, Japan

Yuh-Jong Hu, National Chengchi University, Taiwan

Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France

#### **AMBIENT 2024 Publicity Chairs**

Laura Garcia, Universidad Politécnica de Cartagena, Spain

Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain

## **AMBIENT 2024 Committee**

### **AMBIENT 2024 Steering Committee**

Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain  
Kenji Suzuki, Tokyo Institute of Technology, Japan  
Yuh-Jong Hu, National Chengchi University, Taiwan  
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France

### **AMBIENT 2024 Publicity Chairs**

Laura Garcia, Universidad Politécnica de Cartagena, Spain  
Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain

### **AMBIENT 2024 Technical Program Committee**

Anas Abouelkalam, Cadi Ayyad University - ENSA Marrakesh, Morocco  
Ali Al-Badi, Gulf College, Oman  
El Arbi Abdellaoui Alaoui, Ecole Normale Supérieure – Meknes | Moulay Ismail University, Morocco  
Aaroud Abdessadek, Chouaib Doukkali University, Morocco  
Ana Alves, CISUC - Centre for Informatics and Systems of the University of Coimbra, Portugal  
Francisco Arcas Túnez, Universidad Católica de Murcia (UCAM), Spain  
Hassan Badir, Abdelmalek Essaâdi University, Morocco  
Ayoub Bahnasse, ENSAM Casablanca - University Hassan II of Casablanca, Morocco  
Hind Bangui, Masaryk University, Czech Republic  
Rachid Beghdad, Bejaia University, Algeria  
Samir Brahim Belhaouari, Hamad Bin Khalifa University, Qatar  
Siham Benhadou, National School of Electricity and Mechanics, Casablanca (ENSEM-Casablanca), Morocco  
David Bouck-Standen, Kingsbridge Research Center, UK  
Andrés Bueno-Crespo, Universidad Católica de Murcia, Spain  
Valerie Camps, IRIT | Université Paul Sabatier, France  
Juan-Carlos Cano, Universitat Politècnica de Valencia, Spain  
Dimitris Charitos, National and Kapodistrian University of Athens, Greece  
Albert M. K. Cheng, University of Houston, USA  
Stéphanie Combettes, IRIT | University of Toulouse III, France  
Rémy Courdier, University of Réunion Island, France  
Mohammad Javad Darvishi, University of Montreal, Canada  
Ruediger Dillmann, Karlsruhe Institute of Technology (KIT), Germany  
William Germain Dimbisoa, University of Fianarantsoa, Madagascar  
Paul-Eric Dossou, ICAM, France  
Karim El Bouchti, Mohammed V University in Rabat, Morocco  
Nour-Eddin El Faouzi, Gustave Eiffel University, France  
Mohamed El Khaili, Université Hassan II de Casablanca, Morocco  
Ahmed El Oualkadi, Abdelmalek Essaadi University, Morocco  
Larbi Esmahi, Athabasca University, Canada

Yousef Farhaoui, Moulay Ismail University, Morocco  
Sheikh Amir Fayaz, University of Kashmir, India  
Sebastian Fudickar, University of Oldenburg, Germany  
Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan  
Marco Furini, University of Modena and Reggio Emilia, Italy  
Aryya Gangopadhyay, University of Maryland Baltimore County, USA  
Silvia García, Institute of Engineering | Universidad Nacional Autónoma de México, Mexico  
Valentina Gatteschi, Politecnico di Torino, Italy  
John Gialelis, University of Patras / Industrial Systems Institute, Greece  
Sylvain Giroux, University of Sherbrooke, Canada  
Esteban Guerrero, Vaasa University, Finland  
Barbara Guidi, University of Pisa, Italy  
Maki Habib, American University in Cairo, Egypt  
Hedi Haddad, Dhofar University, Oman / LARIA - National School of Computer Science, Manouba, Tunisia  
Ileana Hamburg, Institut Arbeit und Technik, Germany  
Felix Hamza-Lup, Georgia Southern University, USA  
Jean Hennebert, iCoSys Institute | University of Applied Sciences HES-SO, Fribourg, Switzerland  
Tzung-Pei Hong, National University of Kaohsiung, Taiwan  
Yuh-Jong Hu, National Chengchi University, Taiwan  
Qinaat Hussain, Qatar Transportation and Traffic Safety Center | Qatar University, Qatar  
Faouzi Kamoun, ESPRIT School of Engineering, Tunis, Tunisia  
Inhi Kim, KAIST(Korea Advanced Institute of Science & Technology), Republic of Korea  
Blanka Klímová, University of Hradec Kralove, Czech Republic  
Ziad Kobti, University of Windsor, Canada  
Adam Krzyzak, Concordia University, Canada  
Markus Kucera, Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Germany  
Joo-Ho Lee, Ritsumeikan University, Japan  
Lenka Lhotska, Czech Technical University in Prague, Czech Republic  
Chengxi Li, CNPC, USA  
Ruizhi Liao, The Chinese University of Hong Kong, Shenzhen, China  
Meriem Mandar, Higher Normal School of Casablanca, Morocco  
Johann M. Marquez-Barja, University of Antwerp - imec, Belgium  
René Meier, Hochschule Luzern, Switzerland  
Alex Amir Modarresi, California State University, Northridge, USA  
Frédéric Montet, iCoSys Institute | University of Applied Sciences HES-SO, Fribourg, Switzerland  
Juan Morales-García, Catholic University of Murcia (UCAM), Spain  
Ia Mosashvili, Georgian Technical University / San Diego State University Georgia, Georgia  
Faisal Nabi, University of Southern Queensland, Australia  
Khalid Nafil, University Mohammed V, Morocco  
Gerhard Nussbaum, Kompetenznetzwerk Informationstechnologie zur Förderung der Integration von Menschen mit Behinderungen (KI-I), Austria  
Brendan O'Flynn, Tyndall Microsystems | University College Cork, Ireland  
Andrei Olaru, National University of Science and Technology POLITEHNICA Bucharest, Romania  
Eva Oliveira, 2Ai - EST -Polytechnic Institute of Cavado and Ave, Portugal  
Salome Oniani, Georgian Technical University, Georgia  
Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain  
Hugo Daniel Peixoto, Centro Algoritmi | University of Minho, Braga, Portugal  
Jorge Peña Queralta, Turku Intelligent Embedded and Robotic Systems (TIERS) Group - University of

Turku, Finland  
António Pinto, ESTG - Polytechnic of Porto / INESC TEC, Portugal  
Ivan Miguel Pires, Instituto de Telecomunicações | Universidade da Beira Interior, Covilhã, Portugal  
Sigmundo Preissler Jr., Universidade do Contestado, Brazil  
Awatif Ragmani, University Hassan II of Casablanca, Morocco  
Hajarisena Razafimahatratra, University of Fianarantsoa, Madagascar  
Diego Reforgiato Recupero, University of Cagliari, Italy  
Fernando Reinaldo Ribeiro, Polytechnic Institute of Castelo Branco, Portugal  
Valérie Renault, CREN Laboratory - Le Mans University, France  
Daniele Riboni, University of Cagliari, Italy  
Claudio Rossi, LINKS Foundation, Italy  
David Ruete Zuñiga, Universidad Andrés Bello, Chile  
Michele Ruta, Politecnico di Bari, Italy  
Mohamed Saad Azizi, ENSAM | Moulay-Ismaïl University, Morocco  
Khair Eddin Sabri, The University of Jordan, Jordan  
José Iván San José Vieco, University of Castilla-La Mancha, Spain  
Peter Schneider-Kamp, University of Southern Denmark, Denmark  
Florian Scioscia, Polytechnic University of Bari, Italy  
Halima Sema, Université Hassan premier, Morocco  
Atriya Sen, University of New Orleans, USA  
Antonio Sgorbissa, Università di Genova, Italy  
Navid Shaghghi, Santa Clara University, USA  
Syed Ihtesham Hussain Shah, Institute of High-Performance Computing and Networks (ICAR) | National Research Council (CNR), Italy  
Riaz Ahmed Shaikh, University of East Anglia, UK  
Ahmed Shehata, University of Alberta, Canada  
Sheng Shen, University of Illinois at Urbana-Champaign, USA  
José Luís Silva, ITI/LARSYS & Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Portugal  
Shihao Song, NVIDIA Corporation, USA  
Carine Souveyet, Université Paris 1 Panthéon-Sorbonne, France  
Daniela Elisabeth Ströckl, Carinthia University of Applied Science - Institute for Applied Research on Ageing (IARA), Austria  
Mu-Chun Su, National Central University, Taiwan  
Kenji Suzuki, Tokyo Institute of Technology, Japan  
John Symons, University of Kansas, USA  
Vladimir Trajkovik, University "Ss Cyril and Methodius", Skopje, Republic of Macedonia  
Markku Turunen, Tampere University, Finland  
Rob van der Mei, Vrije Universiteit Amsterdam, Netherlands  
Thomas Waas, Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Germany  
Marcin Wozniak, Silesian University of Technology, Poland  
Xiang Xiao, Google Inc., USA  
Yoonjin Yoon, KAIST, South Korea  
Kin-Choong Yow, University of Regina, Canada  
Abdellah Zaaloul, Université Ibn Zohr, Morocco  
Christos Zaroliagis, University of Patras, Greece

## Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

## Table of Contents

Basic Investigation for Sign Language Sentence Interpretation Using Acceleration Sensor Information <i>Hiroshi Tanaka, Moriyoshi Umeda, Yuusuke Kawakita, Hiromitsu Nishimura, and Jin Mitsugi</i>	1
Integration of Large Language Models into Control Systems for Shared Appliances <i>Frederic Montet, Karl Lowenmark, Marcus Liwicki, Fredrik Sandin, and Jean Hennebert</i>	6

# Basic Investigation for Sign Language Sentence Interpretation Using Acceleration Sensor Information

Hiroshi Tanaka

*Dept. of Information and Computer Sciences  
Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan  
email: h\_tanaka@ic.kanagawa-it.ac.jp*

Yoshimori Umeda

*Course of Information and Computer Sciences  
Graduate School of Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan  
email: s2385004@cco.kanagawa-it.ac.jp*

Yuusuke Kawakita

*Dept. of Information and Computer Sciences  
Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan  
email: kwkt@ic.kanagawa-it.ac.jp*

Hiromitsu Nishimura

*Dept. of Information Media  
Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan  
email: nisimura@ic.kanagawa-it.ac.jp*

Jin Mitsugi

*SFC Laboratory  
Keio University  
Fujisawa, Kanagawa, Japan  
email: mitsugi@keio.jp*

**Abstract—** This paper presents a method for segmenting a sign language sentence consisting of multiple words into individual word motions, which is an elemental technique for achieving the final goal of interpreting sign language sentences. We propose a segmentation method based on the similarity of motions, focusing on the fact that the word motion is included in the sentence motion. We selected 22 frequently occurring sign words and created 5 short sentences using them and acquired word and short sentence motion data. The results of the segmentation method using these data are presented. In addition, we show the results of word classification and confirm the feasibility of the proposed method for sentence interpretation.

**Keywords-** *Sign language; Acceleration sensor; Segmentation; LSTM; SVM; Motion classification.*

## I. INTRODUCTION

The hearing of more than 430 million people worldwide is impaired [1]. Traditionally, communication with normal-hearing people using written or text input tools has been the norm, but with the use of automatic transcription through speech recognition [2], the barriers to communication are gradually being lowered. The use of automatic interpretation has been increasing for verbal communication between different languages, and automatic interpreters have already been commercialized in Japan [3]. If automatic interpretation for sign language, which is considered to be an extension of these technologies, comes into practical use, the communication barrier between people with hearing disabilities and people with normal hearing will be eliminated.

Research on sign language interpretation has so far focused on word-level recognition of sign language motions.

A camera, an accelerometer, and a data glove with a built-in strain gauge have been proposed as sensors for detecting sign language motions [4][5]. In recent years, there have been studies using OpenPose, MediaPipe, and other applications that can extract skeletal nodal information from camera images [6], and multimodal use of multiple sensors for higher accuracy [7]. These have ensured a certain level of recognition accuracy in word count limitations.

Since sign language is composed of multiple words, similar to a normal conversational dialogue, research is currently developing away from a focus on recognition at the word level to a deeper recognition of entire sign language sentences [8]. In this context, there are initiatives to recognize signed sentences using Transformer and Conformer technologies, which have been increasingly used in the field of natural language processing and speech recognition in recent years [9][10]. These are approaches that recognize sign language sentences without splitting them into the individual words that make up the sentence. However, learning data for sign language sentences is required, and a huge amount of learning data is needed to make these approaches practically workable. Acquisition of sign language motion data also imposes a significant burden.

In contrast, we propose a method to interpret sign language sentences from word motion data, considering the situation where a database of sign word motions is provided to make the proposed method feasible [11][12]. While some papers have proposed a method for determining the segmentation point of each word in a signed sentence by the motion speed, etc. [13], we focus on the similarity between the motions of each word in sentences and the hand motions of individual words and propose a method for interpreting by segmenting the sentence into words. The proposed method is



made feasible by limiting the target domain for sign language interpretation, meaning that the number of words required can be reduced, and existing sign language word data can be used.

In Section 2, we present the final sign language interpretation sequence and the research target of this study; in Section 3, we present the method used to segment a sign language sentence into its component words and the results of our experiments; in Section 4, we present the classification results of the segmented motion data and its evaluation; and in Section 5, we discuss the results of this paper and future work.

## II. SEQUENCE OF INTERPRETATION AND INVESTIGATION TARGET

### A. Sequence of sign language interpretation

The sequence for interpretation of sign language sentences is shown in Figure 1. We focus on the fact that the motions of each word that makes up a sentence are contained in the sentence. We divide the sentence into words using a segmentation process, which will be described in Section 3. We already have the sign words' motion data and classify the segmented words using a learning model for the sign word motions.

This paper clarifies this segmentation method and attempts to classify each word based on segmentation results. The research target is the region shown by the blue rectangle in Figure 1. Modification of the classification results and composition into sentences based on a sign language linguistic model remains a future work.

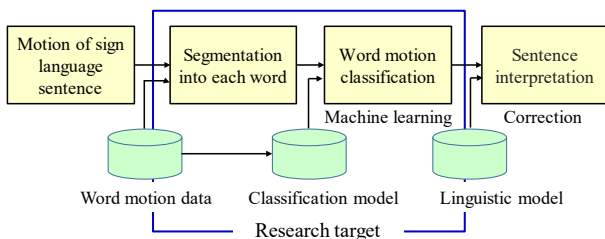


Figure 1. Sequence of sign language sentence interpretation.

TABLE I. TARGET SIGN LANGUAGE WORDS AND SHORT SENTENCES

Sign language words					
1. new	2. system	3. create	4. human	5. animal	6. difference
7. driving	8. license	9. update	10. family	11. put	12. work
13. prioritize	14. ordinary	15. people	16. familiar	17. shop	18. basic
19. power	20. public	21. election	22. law		
Short sign language sentences					
1. new/system/create (Create a new system.)		2. human/animal/difference (Humans and animals are different.)			
3. driving/license/new/update (Update a driving license.)		4. family/put/work/prioritize (Prioritize work over family.)			
5. ordinary/people/familiar/shop (A shop is familiar to ordinary people.)					

### B. Target sign language motions

From approximately 10,000 sign words in the “New Japanese-Sign Language Dictionary” [14], we selected 22 sign words from those with the highest number of references in the dictionary. Short sentences combining these words were created by a sign language instructor, and these sentences were used for segmentation and word classification. Table I shows the 22 sign words and 5 short sentences that are combinations of these words.

### C. Acquisition of sign language motion data

It is not necessary to attach a sensor to the person signing when using a camera, which may be advantageous from the standpoint of real use. There are applications that output body node information from camera images, such as OpenPose and Media Pipe, but they are limited to detecting motion in a plane. An acceleration sensor can measure the motions of sign language in 3D and has a higher sampling rate than a normal video acquisition camera, with each of the methods having their own advantages and disadvantages. We acquired sign language motion data using both an acceleration sensor and a camera. Figure 2 shows the data acquisition setup.

As the purpose of this investigation is to confirm the possible practical application of the proposed method, detailed finger motions were excluded. In order to acquire 3D motions including depth movements, an acceleration sensor (model: Analog Devices, ADXL362) was used to acquire motion data. The sampling rate was 10 ms, with a maximum measurement of 8 G. The sensors were attached at four locations on the elbows and wrists of both hands; sensor data from the four locations were synchronized for data reception using a backscatter communication system [15].

Acceleration data were acquired for the motions of individual words and sentences. Each word was acquired by repeating the sign language motion from a starting position in which the signer was standing still with both hands down by the sides of the body. The beginning and end of the short sentences were the same as for word acquisition. The data set used in this study is shown in Table II: 15 samples were acquired per word, for a total of 330 samples, and 5 different of sentences, 3 samples per sentence, for a total of 15 samples per short sentence. Since the purpose of this study was to verify the feasibility of the method, it was decided to use data from one signer (the aforementioned sign language instructor),

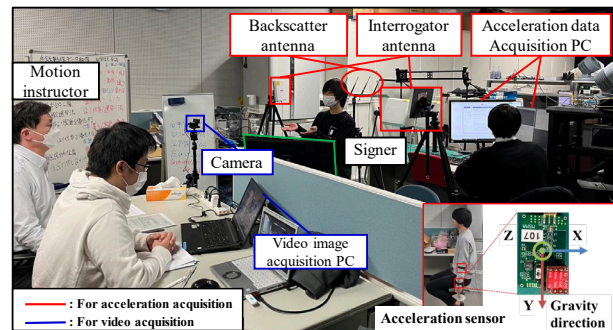


Figure 2. Data acquisition configuration. [16]

TABLE II. TARGET SIGN LANGUAGE WORDS AND SHORT SENTENCES

	No. of signers	No. of motions	No. of samples / motion	No. of total samples
Words	1	22	15	330
Sentences	1	5	3	15

whose signing motions are correct and stable. This was done to minimize the influence of differences in the motions of individual signers.

### III. SEGMENTATION METHOD AND RESULTS

In this section, we discuss the concept underlying the sentence segmentation method and the results of our experiments.

#### A. Fundamental concept

Figure 3 shows an example of acceleration data during the motions for a word and a sign language sentence (data in the x-axis direction for the left wrist). This is for the word “driving” and the sentence “driving/license/new/update” which means "I renew a driver's license". From this graph, we can see that parts of the word motion data are similar to what is seen in the sentence motion data.

Based on this concept, the segmentation method involves the extraction of similar portions of word motions from the sentence motion data. In a sentence, the interval between word actions includes a transition section, which is neither of the two motions (shown in Figure 4). This transition section is considered to be shorter than the duration of the sign word motion and was assumed to be included as part of the sign word in this investigation.

There are two methods for detecting similar parts of actions: the Dynamic Time Warping (DTW) method [17] and the method based on likelihood information using a Long

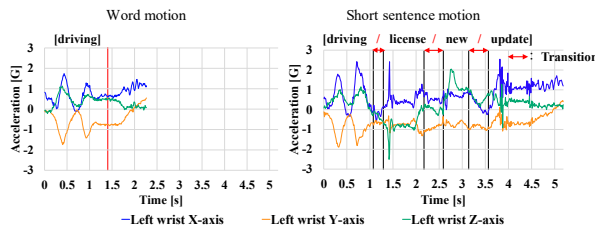


Figure 3. Example of acceleration data.

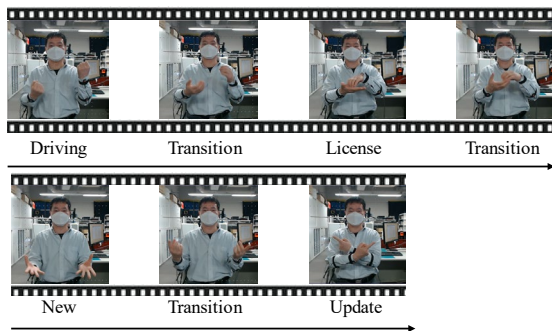


Figure 4. Sentence motions and transition section.

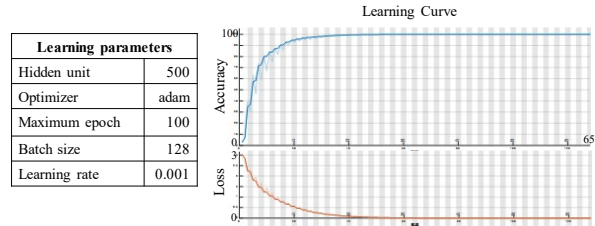


Figure 5. Parameters and learning curves for LSTM model creation.

Short Term Memory (LSTM) model [18]. When using an LSTM model, it is possible to quickly evaluate similar parts taking into account all word data by creating the LSTM model in advance, so in this study, we decided to use an LSTM model that had been trained by word motions.

#### B. Word learning model and likelihood output

To classify the 22 words, the LSTM model was created using 15 samples of each word, with motion data acquired for a total of 330 words as described above. The parameters used to create the model and the learning curve are shown in Figure 5. The convergence of the curve indicates that sufficient learning was achieved.

Figure 6 shows the likelihood output from the LSTM model when the short sentence “driving/license/new/update” is input. Here, the likelihood is the values obtained from the softmax layer of the model (a probability value ranging from 0 to 1, the sum of the elements (22 in this model) is 1). It was confirmed that the likelihoods of the words driving, license, new, and update, which make up the short sentence, are output with the highest values, and that the order of these words is also output correctly in this example.

As another example, Figure 7 shows the results when "family/put/work/prioritize" is input to the LSTM model. Unlike Figure 6, the likelihood of each word is not stable and its variation is large. In the case of Figure 6, it is considered relatively easy to divide the words that make up the sentence, but in the case of Figure 7, division into individual words is difficult and some sort of division criteria must be set.

The following two conditions were imposed for the segmentation of each word using this likelihood information. Here, we also added the condition that the segmentation time should not be less than 0.3 seconds, considering the minimum time required to make a sign language word motion.

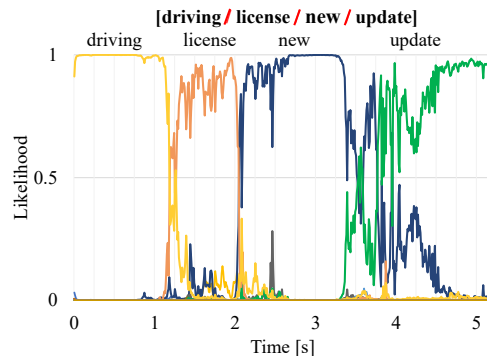


Figure 6. Example of likelihood (Case 1).

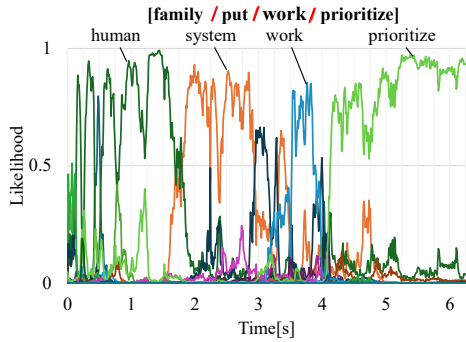


Figure 7. Example of likelihood (Case 2).

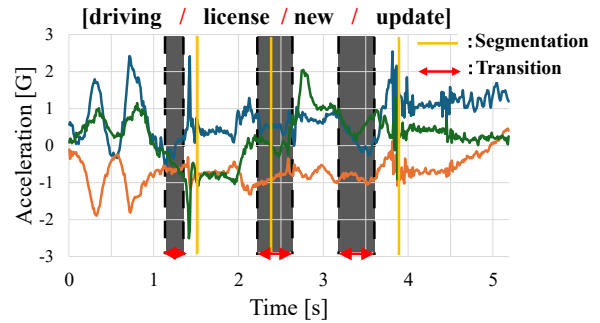


Figure 8. Example of a short sentence segmentation result.

- (a) Selection of the word that has the maximum value of the 22-word likelihood integration value
- (b) Saturation of the integrated value of the likelihood of that word

### C. Segmentation results

An example of the segmentation results from the likelihood output in Figure 6 is shown in Figure 8. The number of words composing the sentence is 4, and the number of segments is 4. The position of the segmentation as well as the number of segments is important, and in this study, the segmentation results were evaluated from the following perspective. The segmentation position was evaluated based on whether it was in the transition section or not. Since a larger number of segments generates a higher number of segmentation positions included in the transition section, the segmentation position index was calculated according to the number of segmentation positions in the transition section and the number of segments. Here, the transition section was determined visually by a person familiar with the word motions.

- (a) Number of segments
- (b) Segment position index: Number of segmentation positions in the transition section / number of segments

The result for 5 short sentences, 3 sentences each, for a total of 15 sentences is shown in Table III. The number of segments tends to be larger than the actual number of constituent words. This indicates that the risk of missing a constituent word is small, and from the perspective of sign language sentence interpretation, it tends to be better than under-division. The segment position index, which indicates the probability that a segment position falls within the transition section, was 0.31 on average. This leads to a decrease in word classification accuracy, and more correct segmentation remains an issue.

### IV. WORD CLASSIFICATION USING SEGMENTATION RESULTS

The segmentation results were used to classify word motions in that segment. We compared the classification accuracy between LSTM and Support Vector Machine (SVM) in a preliminary study and found that SVM performed better with the current number of data for training. Therefore,

TABLE III. EVALUATION RESULT FOR SEGMENTATION

Short sentences	No. of words	No. of Segment			Segment position index		
1. new/system/create	3	4	3	3	0.25	0.33	0.67
2. human/animal/difference	3	3	3	4	0.00	0.33	0.25
3. driving/license/new/update	4	4	4	4	0.25	0.50	0.50
4. family/put/work/prioritize	4	4	3	7	0.00	0.33	0.43
5. ordinary/people/familiar/shop	4	4	5	5	0.25	0.20	0.40

(three sample sentences for each short sentence)

TABLE IV. WORD CLASSIFICATION RESULT BASE ON SEGMENTATION

Sentence	1 <sup>st</sup> seg.	2 <sup>nd</sup> seg.	3 <sup>rd</sup> seg.	4 <sup>th</sup> seg.
driving/license/new/update	driving	familiar	new	system
	people	license	people	update
	law	law	driving	driving

TABLE V. WORD CLASSIFICATION RESULTS FOR SEGMENTED SECTIONS

Short sentences	First place only			Up to 3 <sup>rd</sup> place		
1. new/system/create	0.50	0.33	0.66	0.50	0.66	1.00
2. human/animal/difference	0.33	0.66	0.25	0.33	0.66	0.75
3. driving/license/new/update	0.50	0.75	0.75	1.00	1.00	1.00
4. family/put/work/prioritize	0.50	0.66	0.28	1.00	1.00	0.42
5. ordinary/people/familiar/shop	0.25	0.80	0.40	0.75	0.80	0.60

(three sample sentences for each short sentence)

we used SVM, which has proven performance as an accurate classifier, taking into account the small number of samples for learning. Based on the results of our previous studies [19], the acceleration data for each segmentation section was divided into 10 parts, and the mean value and standard deviation in this region were used as the feature values. Then, a normalization parameter of 10 as the SVM parameter and RBF as the kernel were set for the classification model by SVM. Table IV shows the results of each word classification for each of the 4 segments in Figure 8. The top three classification results are shown in this table. Here, the number of words to be classified is 22.

Word classification of the segmented sections was performed using the segmentation results for five different sentences, a total of 15 sentences. The results are shown in

Table V. As a measure of the classification performance of words in a segmented section, the Evaluation Index ( $EI$ ) was defined as expression (1). The order of words was not considered, and multiple occurrences of a correct word were counted as one.

$$EI = \alpha / \beta \quad (1)$$

where  $\alpha$  is the number of words correctly classified, and  $\beta$  is the number of segments.

The classification performance was evaluated for two cases by assuming that (a) only the first place out of 22 words was correct, and that the classification was correct, and if (b) it was included in the third place. Although there were cases in which words not included in the sentence were classified, basically the words that constituted the sentence were reliably classified. It was confirmed that the proposed segmentation method enables the classification of words that make up sentences, although this is partly because the number of words in our experience was 22.

## V. CONCLUSION

This paper proposed a method for segmenting short sentences into their component words and evaluated its performance. The proposed method is based on the likelihood information obtained from LSTM models learned on word motions. Twenty-two words and 5 different of sentences consisting of those words were created, and sign language data were obtained to evaluate the proposed method. The word classification rate after segmentation was approximately 50% for the first-place criterion and approximately 76% when the third-place criterion was applied, confirming the feasibility of the method.

In future work, it will be necessary to examine methods for eliminating the effect of transition sections to achieve highly accurate word classification, create word learning models for more accurate segmentation, and evaluate the results using sign language data from a large number of sign language users. In parallel, we will collect linguistic information, such as word frequency and word order of a signed language restricted to a specific field, and investigate more accurate methods of interpretation that incorporate such information in the segmentation and word classification results.

## ACKNOWLEDGMENT

This work includes the results of the research and development for expanding radio resources (JPJ000254) of the Ministry of Internal Affairs and Communications. It is also supported by “The Telecommunication Advancement Foundation”, Japan.

## REFERENCES

- [1] Deafness and hearing loss. Available from: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> (accessed June 06, 2024).
- [2] A. Radford et al., “Robust Speech Recognition via Large-Scale Weak Supervision,” Proceedings of the 40th Int. Conf. on Machine Learning, PMLR 202:28492-28518, pp. 1-27, 2023.
- [3] PokeTalk. Available from: <https://pocketalk.jp/>(accessed June 06, 2024).
- [4] M. J. Cheok, M. Jin, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," Int. Journal of Machine Learning and Cybernetics 10 pp. 131-153, 2019.
- [5] M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep learning for sign language recognition: Current techniques, benchmarks, and open issues," IEEE Access 9, 126917-126951, 2021.
- [6] S. Qiao, Y. Wang, and J. Li, "Real-time human gesture grading based on OpenPose," 10th Int. Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1-6, 2017.
- [7] Q. Miao et al., "Multimodal gesture recognition based on the resc3d network," In Proceedings of the IEEE int. conf. on computer vision workshops, pp. 3047-3055. 2017.
- [8] S. B. Abdullahi, S. Bala, and K. Chamnongthai, "American sign language words recognition using spatio-temporal prosodic and angle features: A sequential learning approach," IEEE Access 10: 15911-15923, 2022.
- [9] P. Villegas and L. Francisco, "Sign Language Segmentation Using a Transformer-based Approach." Available from: <https://hdl.handle.net/20.500.14468/14662>, 21 pages, 2022.
- [10] J Huang et al., "Video-based sign language recognition without temporal segmentation." Proc. of the AAAI Conf. on Artificial Intelligence. Vol. 32. No. 1. pp. 2257-2264, 2018.
- [11] SmartDeaf -- Video dictionary for learners of sign language. Available from: <https://www.smartdeaf.com/>(accessed June 06, 2024)(in Japanese).
- [12] Sign language shower. Available from: <https://www.nhk.or.jp/school/tokkatsu/syuwashower/>.(accessed June 06, 2024)(in Japanese).
- [13] K. Renz, N. C. Stache, S. Albanie, and G. Varol, "Sign Language Segmentation with Temporal Convolutional Networks," IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 2135-2139, 2021.
- [14] A. Yonekawa, “New Japanese Sign Language Dictionary,” Japan Federation of the Deaf and Dumb, 2011.
- [15] J. Mitsugi, Y. Kawakita, K. Egawa, and H. Ichikawa, “Perfectly Synchronized Streaming from Multiple Digitally Modulated Backscatter Sensor Tags,” IEEE J.RFID, vol. 3, no. 3, pp. 149-156, 2019.
- [16] Y. Umeda et al., “Proposal and Evaluation of Sign Language Sentence Segmentation Method based on Acceleration Information,” RISP Int. Workshop on Nonlinear Circuits, Communications and Signal Processing, pp. 360-363, 2024.
- [17] W. Li, Z. Luo, and X. Xi, "Movement trajectory recognition of sign language based on optimized dynamic time warping," Electronics 9, no. 9, 1400, 15 pages, 2020.
- [18] Sequence-to-Sequence Classification Using Deep Learning. Available from: <https://www.mathworks.com/help/deeplearning/ug/sequence-to-sequence-classification-using-deep-learning.html> (accessed June 06, 2024).
- [19] T.Wakao et al., “Application of multimodal methods to sign language motion classification and its effectiveness evaluation,” RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, pp. 269-272, 2023.

# Integration of Large Language Models into Control Systems for Shared Appliances

Frédéric Montet<sup>\*,‡</sup> , Karl Löwenmark<sup>†,‡</sup> , Marcus Liwicki<sup>†</sup> , Fredrik Sandin<sup>†</sup> , Jean Hennebert<sup>\*</sup> 

<sup>\*</sup>Institute of AI and Complex Systems, iCoSys HEIA-FR, HES-SO, Fribourg, Switzerland

<sup>†</sup>Machine Learning, Department of Computer Science, Electrical and Space Engineering,  
Lulea University of Technology, 97187 Lulea, Sweden,

e-mail(s): frederic.montet@hefr.ch, karl.lowenmark@ltu.se

<sup>‡</sup>Equal contribution of the authors

**Abstract**—Large Language Model (LLM)-powered chatbot agents have proven to be immensely useful in tasks, such as writing and generating essays, code, and academic text. By using frameworks such as LangChain, agents can be equipped with tools to access and analyse custom data, which facilitates bespoke applications, such as customer service agents with access to internal documents and tailored reasoning. While the focus of such applications has mainly centered around textual content, custom toolboxes could also enable agents to act in completely different use cases, for instance control theory. Nevertheless, given the non-deterministic nature of LLMs, merging them with deterministic software implies challenges in applied contexts such as privacy, multi-user interactions, and consistency. To pave the way to reliable LLM usage in various contexts, this work provides the foundation for expanding the use of LLM agents to the domain of control systems and human-centric automation. An agent-based architecture is proposed, which is then implemented within the context of a shared space heating system controlled by three personas. Finally, we evaluate the capacity of the system to deal with scenarios such as normality, erratic user behavior, conflicts of interest, and system limitations. The findings of this study highlight the potential benefits and challenges of using LLMs for appliance control.

**Keywords**—control; large language models; shared appliance; LangChain; human-robot interaction; social robotics; generative models.

## I. INTRODUCTION

Large Language Models (LLMs) have transformed the field of Artificial Intelligence (AI), enabling text understanding and generation on levels that mimic human capabilities [1][2] and leading some researchers to hypothesise that modern LLMs are at an early stage of artificial general intelligence [3]. The rapid evolution of LLMs has been realised in the domain of language understanding and generation in text-based tasks, such as summarising text [4], evaluating essays [5], and generating code [6]. Furthermore, LLMs have been trained with Reinforcement Learning from Human Feedback (RLHF) [7] to act as chatbot assistants, such as ChatGPT, which are estimated to have significant impact in a variety of fields such as education [8], medicine [9], and legal practice [10]. Chatbots like ChatGPT can also act as reasoning *agents* through frameworks such as LangChain [11] and HuggingFace agents, where they are equipped with *tools* that enable them to call custom built functions that implement, for instance, retrieval augmented generation (RAG) [12], loading of custom data, and advanced data analysis (e.g., llmath [13]).

The agent thus becomes able to interpret human requests, call custom-built functions, access external data, interpret results, and return the conclusions to the user [14], which has been used for a variety of tasks, such as literature reviews [15], customer service [16], mental healthcare [17], and document management [18]. Given these impressive characteristics, LLM usage in other contexts than question answering seems promising [19], but only a few cases have been explored due to the youth of these underlying technologies [20].

### A. Cyber-physical heating systems

One such unexplored use case is LLM-integrated control loops, such as optimally controlling the indoor temperature in a building, in particular with multiple agents or users. Indoor heating might seem like a trivial problem, solved with a basic control system. However, increasing sustainability demands and limitations in energy production require modernisation of heating control to minimise emissions and cost; for instance, residential and commercial heating constitute around 40% of primary energy use in the EU and US. As a consequence, this has motivated research into predictive temperature control based on sensors and models in a Cyber-Physical Systems (CPS) framework.

As the complexity of such systems increases, so does the scope of human-machine interactions. For example, most office spaces currently feature employees as human-in-the-plant [21], where each employee is affected by the temperature control but has limited power over it. By giving customisability and control to users, the system switches to a humans-in-the-loop system [21], where each employee now participates in the control and is affected by the control. Such systems necessitate a human-centric design [22] that takes new questions into account, such as:

**Explainability** Does every user understand how to use the new system, or do some feel excluded?

**Mediation** How does the system resolve conflicts between preferences such as indoor temperature?

**Robustness** How can the system safely and efficiently take user parameters into account?

These are difficult questions to answer. For example, in explainability, there is a clear trade-off between transparency and simplicity. Transparency describes how exactly an agent decision is motivated, and simplicity describes how few data that the user must interact with. Consider the most transparent

case, where the control system explains all algorithms used to control the heating on the screen where temperature is controlled. This is certainly an overload of information for most users and an undesirable solution. Consider, on the other hand, a system in which only the temperature setting is ever divulged. This is an arcane system for the user, which could lead to frustration and resentment.

All these considerations place considerable load on the designer, who has to design a system both flexible and simple to use. Ideally, the high-level control of the CPS would be managed by a human mediator, who can take all types of context into account and make fair assessments, while users only require natural language to interact with the system through the mediator. Although human experts for every CPS would be far too expensive, the aforementioned advances in LLM chatbot technology could facilitate a human-esque mediator, serving as the controller of the heating system in a way similar to voice-controlled homes in the Internet of Things domain [23].

### B. Research questions

Given the previously highlighted issues, this paper seeks to answer *how can LLM agents be used as mediators and controllers in CPS control systems?* More specifically, the aim is to investigate:

- 1) What architectures can be used to leverage LLMs for improving user interfaces within control systems?
- 2) How can LLMs incorporate user preferences and circumstances into a control system?
- 3) How can LLMs mediate between users with varying preferences or constraints?

In the following sections, we describe the theory and method used to implement an LLM agent as a user interface in a control loop for a case study of indoor heating, describe and discuss the results, and present a path for future work in this direction.

## II. METHOD

In this section, we introduce the general components of an LLM controller agent and describe a case study featuring a custom agent connected to a simulated building heater.

### A. LLM controller agent definitions

Figure 1 illustrates an LLM controller agent CPS system with users, an agent, and a system.

1) *Users*: Users represent any human willing to have an interaction with the system. The users can chat with the chatbot and ask it to update the control parameters, as well as ask questions about the system and the controller. Information about the users can be stored so that the agent can load chat history and user preferences to personalise the service.

2) *Agent*: The agent represents the LLM chatbot with connected tools. It has three main functions: user interaction, mediation/decision making, and system control. User interaction with one or more users is achieved with a user interface, which can be a Graphical User Interface (GUI), a chatbot interface, or a voice interface. The agent can store, load, and

TABLE I. LIST OF TESTS RUN ON THE DEVELOPED SOFTWARE.

#	Description	Type	User
1	Normal case	Normal operation	single
2	Normal case	Normal operation	multiple
3	Insistent user input	Normal operation	single
4	Insistent user input	Normal operation	multiple
14	User information gathering	Normal operation	single
15	User information gathering	Normal operation	multiple
5	Erratic user input	Disturbance	single
6	Erratic user input	Disturbance	multiple
11	Chaos	Disturbance	multiple
7	Exaggerated input, < lowest set point	System limitation	single
8	Exaggerated input, < lowest set point	System limitation	multiple
9	Exaggerated input, > highest set point	System limitation	single
10	Exaggerated input, > highest set point	System limitation	multiple
12	Unreachable system	Error	single
13	Unreachable system	Error	multiple

update user preferences and user information in a database to facilitate personalised user interactions. Preferences and user information, such as heat preference and health status, also act as constraints that affect the agent's decision making. Furthermore, the agent can answer questions related to the documentation of the system, itself, or other relevant metadata that is accessible in its knowledge base through RAG. Finally, the agent interacts with the Proportional–Integral–Derivative (PID) controller to set or get system parameters, such as temperature or heater effect.

3) *System*: The system is the component representing the (cyber)physical system. It can be any physical system having a software interface such as Representational State Transfer (REST), OPC Unified Architecture (OPC-UA), etc. When no user requests changes, the system must remain in a stable state that is satisfactory to most users.

### B. Case study: Simulation of residential building heating

Based on the previous definitions, a space heating case study is performed, where multiple users interact with a chatbot which controls a simulated boiler for indoor heating. The simulation implements the basic physical properties of a building and its heating system, where the heat is regulated by a PID controller, which in turn is controlled by the agent. The agent is implemented in LangChain, using GPT-4 as the LLM, and custom built tools to interact with a database with user information, RAG to access a vectorbase of Q&As related to the heater, and another set of tools to interact with the controller via JSON requests (REST API) to set or get system parameters.

1) *Test protocol with users*: Table I shows the test protocol for 15 different tests evaluating normality, disturbance, system limitations, and system error. The tests use a collection of three personas with varying age, preferred temperature, and a weekday and weekend schedule, visible in Table II.

The complete runs are available in the supplementary materials.

## III. RESULTS

The results of the study are split into two parts:

- 1) a software implementation given the schema in Figure 1,

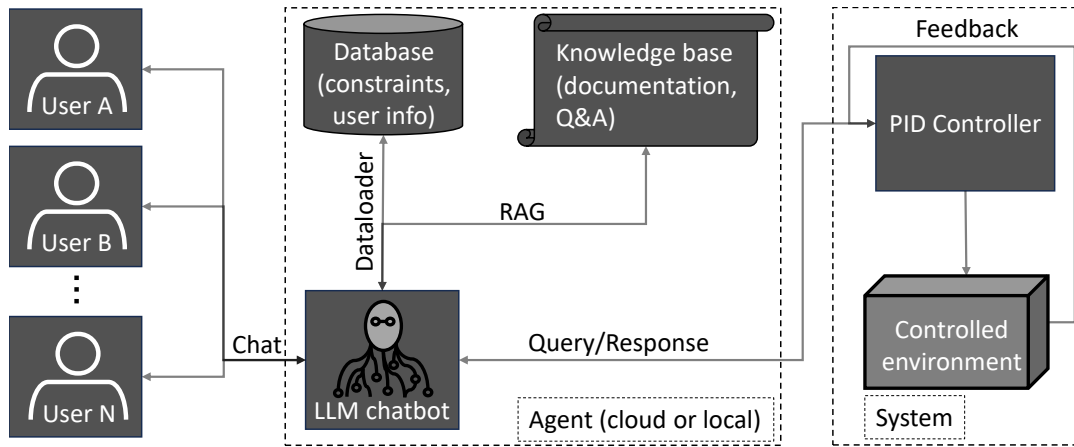


Figure 1. Conceptual block diagram describing the information flow of the Humans-in-the-Loop CPS with an agent powered by an LLM chatbot.

2) a simulated case study where the series of tests summarised in Table I are presented.

A. Proposed architecture

The conceptual schema from Figure 1 led to an implementation following the Unified Modeling Language (UML) component schema from Figure 2.

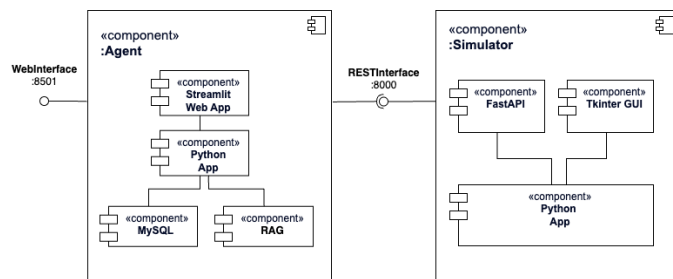


Figure 2. UML Component Schema of the control system using LLMs.

This system architecture enables all the interactions required to facilitate the different flows of information from the conceptual schema. The REST interface enables two-way communication between the Simulator and the agent via HTTP. The user then gets all the necessary feedback through the web interface; a screenshot of the latter is visible in Figure 3. Furthermore, the simulator values can be followed with a desktop Graphical User Interface (GUI) implemented using Tkinter presented in Figure 3. The code for the agent and the simulation is available on GitHub [24].

B. Case study: Simulation

The following sections present a summary of each type of test case visible in Table I using personas from Table II.

TABLE II. SHORT DEFINITION OF THE PERSONAS USED IN THE TEST CASES.

Name	Age	Temperature	Weekday	Weekend
John Smith	31	20	07:00 - 17:00	09:00 - 23:00
Ronnie Coleman	58	22	06:30 - 18:00	08:30 - 22:30
Robinson Crusoe	25	21	09:00 - 19:00	10:00 - 23:00

1) *Normal Operations:* For both single and multi-user contexts, normal operations of the heater did work successfully given test #1 and #2. However, in insistent scenarios #3 and #4, the temperature was set without taking into account other users, both for the single and multi-user context. Furthermore, the temperature set point has reached values that are much too cold for a building. In tests #14 and #15, the agent has been reluctant to provide user information to the current user if the query did not include the name of the user; otherwise, preferences were gathered successfully. In some cases, unplanned behaviour did happen such as confirming the change of temperature set points without the latter being effectively updated. Also, when asking for user preferences in a query, the next query simulating a user login did induce an answer including the user preferences of this user.

2) *Disturbance:* When dealing with erratic cases in tests #5, #6 and #11, the agent behaved successfully in both single- and multi-user scenarios. When faced with disturbances or unclear queries, the agent did not stop the flow of conversation but instead asked for more precision; even when facing chaotic situation with multiple users. In multi-user context with only one erratic user, other users could continue having normal interactions with the agent.

3) *System limitations:* Represented by tests #7, #8, #9 and #10, exaggerated queries were formulated above and below thresholds from normal set points from a house temperature, within single- and multi-user contexts. When requests went above the system threshold, the agent asked the user for another value within the system limits, thus complying with the expected behaviour to refuse to set such set points. In other cases where the user asked for temperatures in the range of the system limitations but with unrealistic temperature for a house, the agent set these temperatures without any warnings. When faced with an unrealistic set point, other users could update the value to any other set point.

4) *Error:* In scenarios simulating an error such as the unavailability of the system, the agent provided simple answers to communicate the issue to the user. This happened in both single- and multi-user tests.

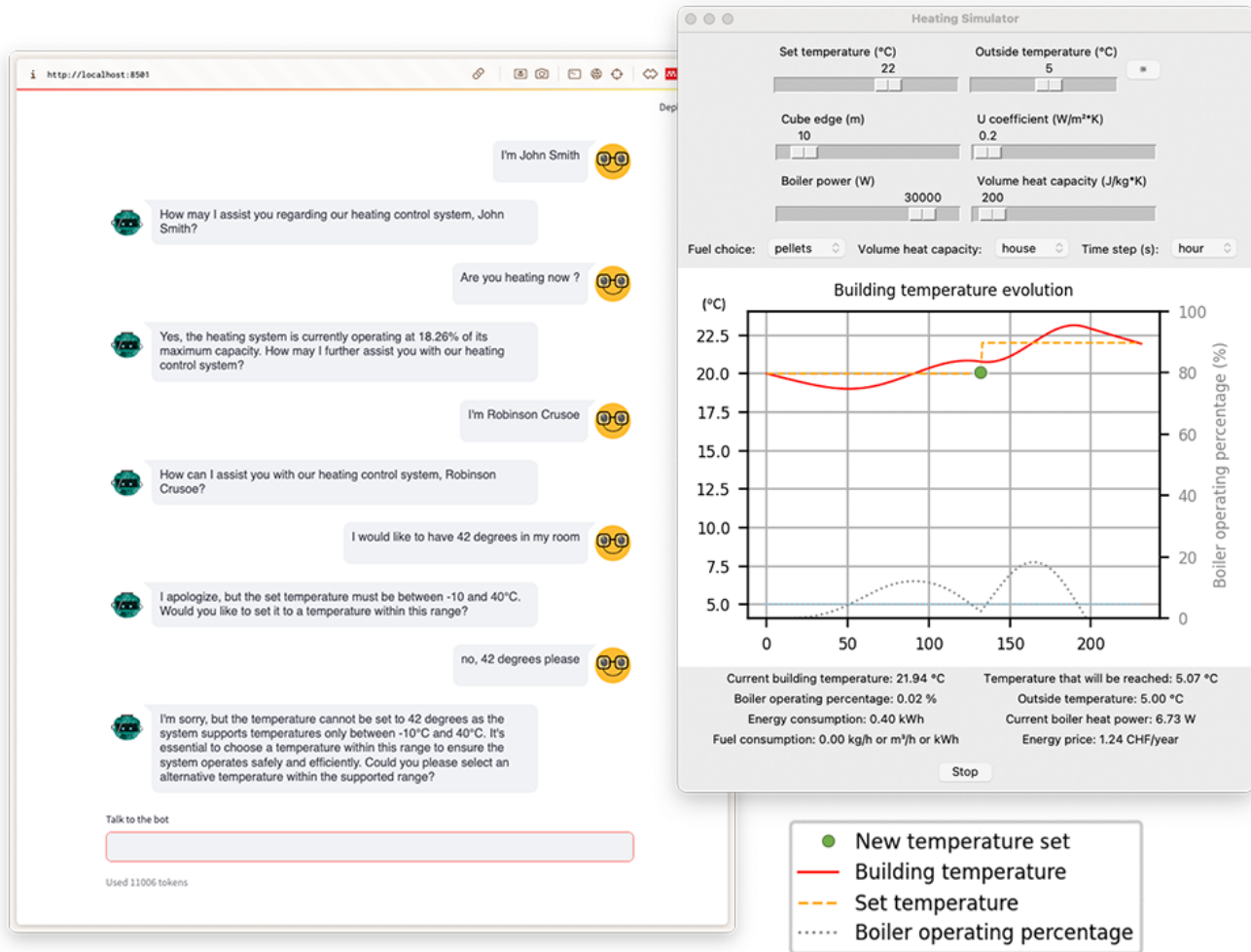


Figure 3. Interfaces of the chatbot on the left and the simulator on the right.

#### IV. DISCUSSION

Structured around three research questions, this study introduces the use of LLM-based chatbot agents to act as user interfaces and mediators in control systems. To answer the research questions cited in the introduction, a method with an implemented software and two use cases yielded qualitative results, given multiple test scenarios. Following, the three answers to the research questions will be detailed.

##### A. What architectures can be used to leverage LLMs for improving user interfaces within control systems?

Presented in Figure 2, the proposed architecture composed of an agent and a simulator remains straightforward, thus making it flexible for all research purposes. This architecture has been implemented and allowed the suite of test cases in Table I to be run with various scenarios. For example, in test case #14, the agent executed queries in the database, successfully used RAG, and set a temperature set point in the simulator. Therefore, with this diversity of queries within a single chat conversation, the first research question is answered. Nevertheless, a successful control system needs more than an architecture, which will be discussed in the next sections.

##### B. How can LLMs incorporate user preferences and circumstances into a control system?

To answer the second research question, the case study simulation and the mediator case study have been performed. For all test cases in Table I, the results show that given a decision to make, the agent is only able to gather part of the information it needs, despite being informed of its available functions. Nevertheless, when a query is precise enough, the agent can successfully gather one type of external information, synthesise its result in natural language, and deliver the answer to the user, while calling the appropriate function to, e.g., set the simulator to a desired state.

Therefore, this behaviour already enables simple queries to be executed as seen in the successful test cases, but has limitations. This implies that a better reasoning component might facilitate more complex queries and context to be answered. For instance, at each query, more additional context could be injected directly in the prompt, or RAG could be involved when large documents are required.

In the test cases in Section II-B, only the last two messages were included to generate an answer. Despite providing enough



information to answer the request, this drastically limited the reasoning capabilities of the LLM in its understanding of multi-user interactions and evolution of one or many personas needs.

Given these tests, the second research question can only be partly answered; user preferences and circumstances were incorporated into the agent, but only for simple queries. Since multi-user contexts already involve more complex queries, such a scenario has been tested and is discussed in the next section.

A potential next step in user assistance is to incorporate external information, such as energy price and environmental impact, by adding tools for real-time data querying. This would enable users to take such information into account when making decisions, and might also enable direct optimisation based on such parameters to facilitate cheaper and more sustainable operations.

### *C. How can LLMs mediate between users with varying preferences or constraints?*

The results of the case study clearly show the limitations of a chatbot that is not set up properly to mediate between conflicting interests; all tests with the custom agent show how it always conforms to the latest request, unless it is completely inappropriate, such as 42°C, as in test 10 in the supplements. This is likely due to the LangChain agent implementation being built on very early LangChain functions, and we believe that by updating the agent framework to the modern chat chains this will drastically improve. Hence, the next iteration of the custom agent should incorporate tools to facilitate chat history between users and carefully engineered prompts to prevent an overly accommodating controller.

This could be achieved with a multi-agent system [25][26], where each user has access to their own personal agent, while a controller/mediator agent interacts with the personal agents and the system. The politeness of LLM agents can thus be leveraged to create an LLM "filter" between the user and the controller, hopefully preventing abuse by overly persistent and aggressive users. User privacy concerns can also be addressed by having chat histories and complete user preferences stored so that only the personal agent can access them, while the controller agent receives only relevant information from the personal agent when necessary.

### *D. Ethical considerations*

The ethical concerns present in this line of work are related to privacy concerns and value-based judgements.

Privacy concerns are a risk due to the sharing of the agent, and the storing of personal data where the shared agent can access them. With good data practice and an agent accurately prompted to access only the data pertaining to the current user, this risk can be mitigated, though this is not something we specifically investigated. Alternatively, a multi-agent framework as described in Section IV-C can be used to keep data storage unique to each personal agent, and thus each user.

Value-based judgements are a consideration in any work with conflict resolution and mediation aspects; who decides what is right when two parts have conflicting interests that cannot be

jointly satisfied? For instance, if one user is freezing and another is sweating indoors, is it worse to sweat or to freeze, and is it easier to put on a jacket than to dress cooler? Furthermore, is it more ethical to have a cool indoor environment for the sake of reducing emissions, or is it more ethical to have an indoor environment that most users favour? These questions are difficult to answer, and not all humans will agree.

Thus, when using an LLM as a mediator, it is important to consider who decides GPT-4's values? A causal language model, such as GPT-4, trained with self-supervised learning on vast amounts of text will learn to mirror the values it has seen during training. Supervised learning is then used to align the model with our desired outputs, and finally reinforcement learning from human feedback is used to optimise the chatbot output based on training participant evaluation. Consequently, there are three levels of learnt values: the hidden beast of self-supervised learning, the indirect adjustment of supervised fine-tuning, and the conforming to desires of RLHF participants.

The question of which values to enshrine is particularly interesting during RLHF, where the selection of participants will implicitly alter the "ethics" of the LLM. Hence, the ideal solutions for custom controller agents would be to set up pre-trained LLMs locally, fine-tune them on custom tasks, then define RLHF rewards so that the agents can be further optimised while running. This would also ensure maximum privacy, as everything can run locally within protected networks. However, it is also the most expensive solution in both energy and development, so different scenarios will likely require different levels of local versus cloud-based approaches.

### *E. Future work*

In order to improve the usage of LLMs for control systems, the following improvements are proposed.

First, the ability to understand context coming from multiple sources. This could be enabled with the implementation of an algorithm such as Recursion of Thoughts, an improved version of Chain of Thoughts [27], [28]. This added step would allow the agent to progressively build an answer given provided sources of information, thus creating a bigger context.

Second, an improved multi-user ability. So far the aforementioned feature has been simulated, and to enable it would require a dedicated architecture able to alleviate privacy issues. A suggested possibility would be to divide the agent in two parts: a user-agent that is personal and a system-agent that is part of the control system. This division would increase privacy since the user-agent would be isolated from the other agents, which would naturally prevent private information leakage between agents. Also, with this improvement, the system-agent could have the possibility to request information from all user-agents to build its context before taking a decision. In the case of this study, the system-agent would act as a mediator taking all user preferences and inputs to propose a new set temperature.

Third, a scheduler to take into account time-related inputs. Since controlling a system can include timely inputs, the system-agent should have some scheduling capabilities. For instance,

this could be used to adapt set points to user preferences and inputs dynamically, or take the weather into account.

## V. CONCLUSIONS

This study examined the integration of LLMs into control systems, specifically targeting shared appliances within a CPS context. The exploration centred around a framework that facilitated the use of LLMs to manage the heating system of a shared building, addressing user preferences, and mediating conflicts. Key findings underscore the LLM's capability to function beyond traditional applications, serving as a robust mediator and decision-maker within a control system. However, challenges such as handling erratic input and dependency on user interaction precision were identified, suggesting areas for future improvement. In addition, ethical considerations highlighted the need for user data management and the importance of aligning the LLM's operational parameters with ethical standards, ensuring that privacy and fairness are maintained. Further research should expand the application of LLMs within CPS by exploring multi-agent and time-dependent frameworks, to further explore the potential aspects of human-machine interactions in shared environments.

## ACKNOWLEDGEMENTS

The authors would like to thank Philippe Marziale for his remarkable contribution to the code base, as well as Sébastien Rumley and the company yord sàrl (yord.ch) for their guidance during this project.

## REFERENCES

- [1] T. Brown *et al.*, "Language models are few-shot learners", *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [2] H. Touvron *et al.*, "Llama 2: Open foundation and fine-tuned chat models", *arXiv preprint arXiv:2307.09288*, 2023.
- [3] S. Bubeck *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4", *arXiv preprint arXiv:2303.12712*, 2023.
- [4] M. G. Madden, B. A. McNicholas, and J. G. Laffey, "Assessing the usefulness of a large language model to query and summarize unstructured medical notes in intensive care", *Intensive care medicine*, vol. 49, no. 8, pp. 1018–1020, 2023.
- [5] P. U. Rodriguez, A. Jafari, and C. M. Ormerod, "Language models and automated essay scoring", *arXiv preprint arXiv:1909.09482*, 2019.
- [6] M. Chen *et al.*, "Evaluating large language models trained on code", *arXiv preprint arXiv:2107.03374*, 2021.
- [7] D. M. Ziegler *et al.*, "Fine-tuning language models from human preferences", *arXiv preprint arXiv:1909.08593*, 2019.
- [8] E. Kasneci *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education", *Learning and individual differences*, vol. 103, p. 102274, 2023.
- [9] T. Dave, S. A. Athaluri, and S. Singh, "Chatgpt in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations", *Frontiers in artificial intelligence*, vol. 6, p. 1169595, 2023.
- [10] M. Ajevski, K. Barker, A. Gilbert, L. Hardie, and F. Ryan, "Chatgpt and the future of legal education and practice", *The Law Teacher*, vol. 57, no. 3, pp. 352–364, 2023.
- [11] H. Chase, "Langchain", 2022, <https://github.com/hwchase17/langchain> [Accessed: 2024-06-13].
- [12] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks", *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [13] paulalesius, *Large language math - the mathematics of llm foundational models - for beginners*, <https://llmath.unnservice.com/> [Accessed: 2024-06-13], 2022.
- [14] O. Topsakal and T. C. Akinci, "Creating large language model applications utilizing langchain: A primer on developing llm apps fast", in *International Conference on Applied Engineering and Natural Sciences*, vol. 1, 2023, pp. 1050–1056.
- [15] R. Asyrofi, M. R. Dewi, M. I. Lutfhi, and P. Wibowo, "Systematic literature review langchain proposed", in *2023 International Electronics Symposium (IES)*, IEEE, 2023, pp. 533–537.
- [16] K. Pandya and M. Holia, "Automating customer service using langchain: Building custom open-source gpt chatbot for organizations", *arXiv preprint arXiv:2310.05421*, 2023.
- [17] A. Singh, A. Ehtesham, S. Mahmud, and J.-H. Kim, "Revolutionizing mental health care through langchain: A journey with a large language model", in *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2024, pp. 0073–0078.
- [18] A. Pesaru, T. S. Gill, and A. R. Tangella, "Ai assistant for document management using lang chain and pinecone", *International Research Journal of Modernization in Engineering Technology and Science*, 2023.
- [19] F. Dou *et al.*, "Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges", *arXiv preprint arXiv:2309.07438*, 2023.
- [20] M. U. Hadi *et al.*, "A survey on large language models: Applications, challenges, limitations, and practical usage", *Authorea Preprints*, 2023.
- [21] T. Samad, "Human-in-the-loop control: Applications and categorization", *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 311–317, 2020, 3rd IFAC Workshop on Cyber-Physical & Human Systems CPHS 2020, ISSN: 2405-8963.
- [22] Y. Fang, Y. Lim, S. E. Ooi, C. Zhou, and Y. Tan, "Study of human thermal comfort for cyber-physical human centric system in smart homes", *Sensors*, vol. 20, no. 2, p. 372, 2020.
- [23] M. A. Torad, B. Bouallegue, and A. M. Ahmed, "A voice controlled smart home automation system using artificial intelligent and internet of things", *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 4, pp. 808–816, 2022.
- [24] *GitHub code repository llm for cyber-physical systems featuring an agent controller*, <https://github.com/fredmontet/llm-cps-ac>, Accessed: 2024-09-25.
- [25] J. Qin, Q. Ma, Y. Shi, and L. Wang, "Recent advances in consensus of multi-agent systems: A brief survey", *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4972–4983, 2017.
- [26] A. Amato, A. Quarto, and V. Di Lecce, "An application of cyber-physical system and multi-agent technology to demand-side management systems", *Pattern Recognition Letters*, vol. 141, pp. 23–31, 2021, ISSN: 0167-8655.
- [27] S. Lee and G. Kim, "Recursion of thought: A divide-and-conquer approach to multi-context reasoning with language models", *arXiv preprint arXiv:2306.06891*, 2023.
- [28] M. Nye *et al.*, *Show your work: Scratchpads for intermediate computation with language models*, 2021. arXiv: 2112.00114 [cs.LG].