Human Activity Data Collection, Analysis and LLM-based Zero-Shot Reasoning

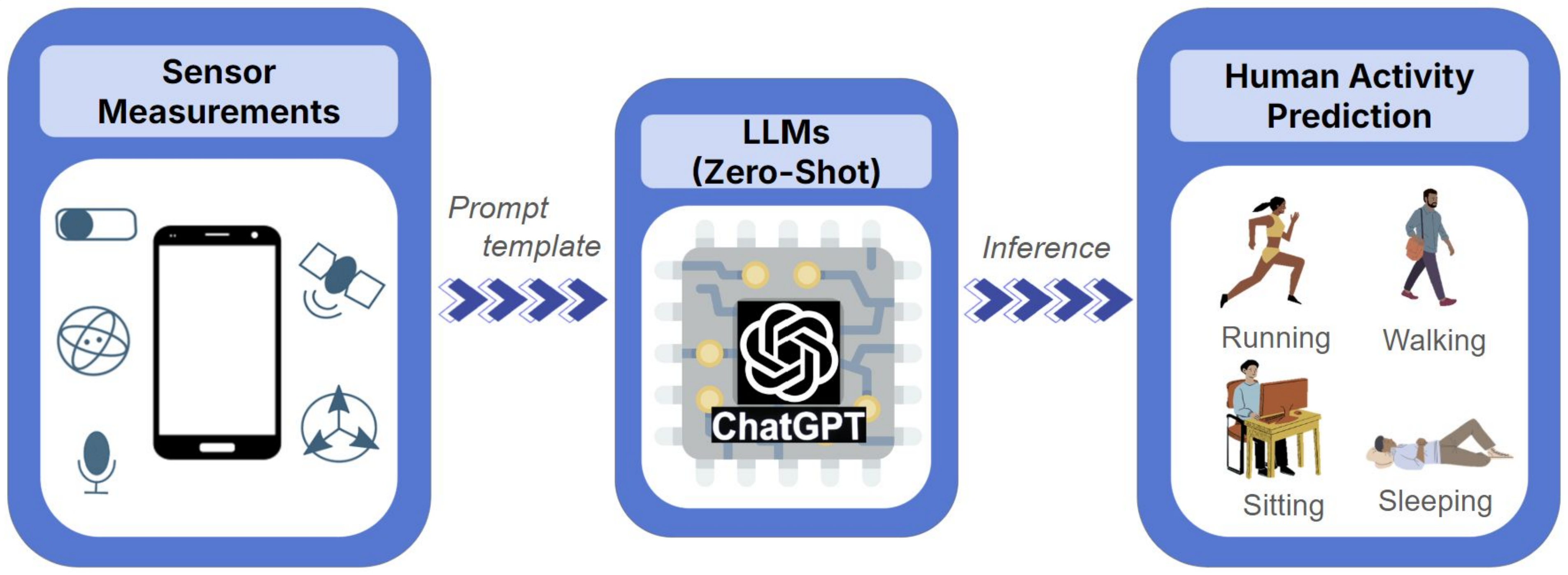
UC San Diego
JACOBS SCHOOL OF ENGINEERING
Computer Science and Engineering

Presenters: *Audrick Bolaños Linares, Cesar Daniel Ortiz Aguilera*
Xiaofan Yu[1], Tajana Rosing[1], Moshiri Niema[1], Diego Muñiz De las Casas, and Marena Isabel Schiess

Background & Motivation

Automatic recognition of **Human Activity** (HA) using Internet-of-Things (IoT) devices has become crucial to health monitoring, aging care, and human behavior analysis.

- Current Challenges: Require **large datasets**, **long training time**, and **extensive resources**.
- The proposed project presents an alternative way of using LLMs based on:
  - **Zero-shot reasoning** to analyze raw IMU data and predict HA **without training**
    - **Chain-of-thought** prompt design



Related Works

Convolutional neural network model

MultiCNN-FilterLSTM [1], and LSTM-CNN [2]

Limitations: Complex multi-head CNN-based architectures, and **expensive training**.

- WISDM dataset: 1,098,207 samples of physical activities (6 activities), 972 minutes sampled at 20 Hz
- Have **high time complexity** and the multi-head attention layer results in **many parameters**

Large language model

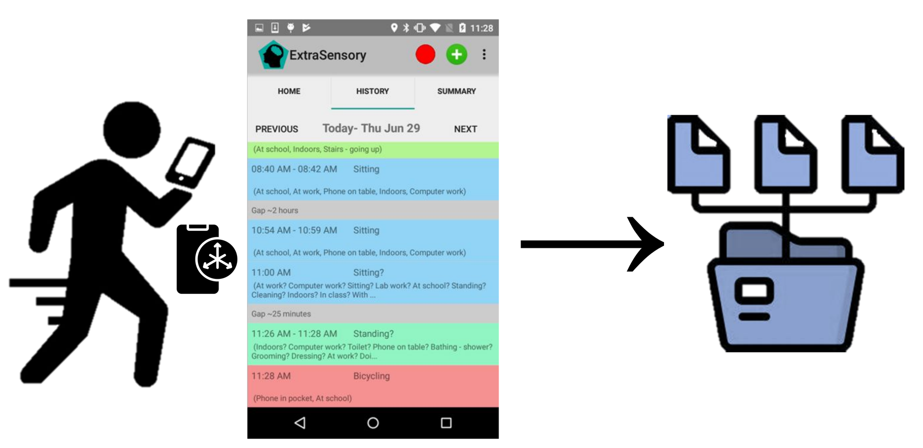
ChatGPT 4-based HAR [3]

Limitations: Limited activity collection (4 attributes), prompt **length restriction**, and low F1-score.

Methods

(1) Sensor data collection:

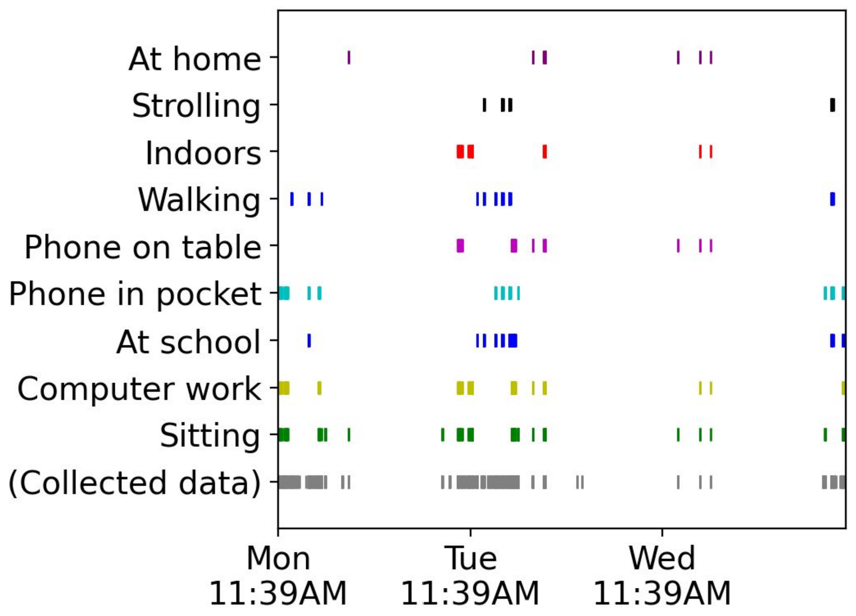
Android phone with the ExtraSensory App.



➔ Data collected: Continuous sensor logs and **manual reporting** of activities by users for **daily activities monitoring**

➔ Phone IMU (sensors used):
- Magnetometer (Magnet)
- Accelerometer (Acc)
- Gyroscope (Gyro)
- Microphones (Audio)

(2) Label analysis dataset:



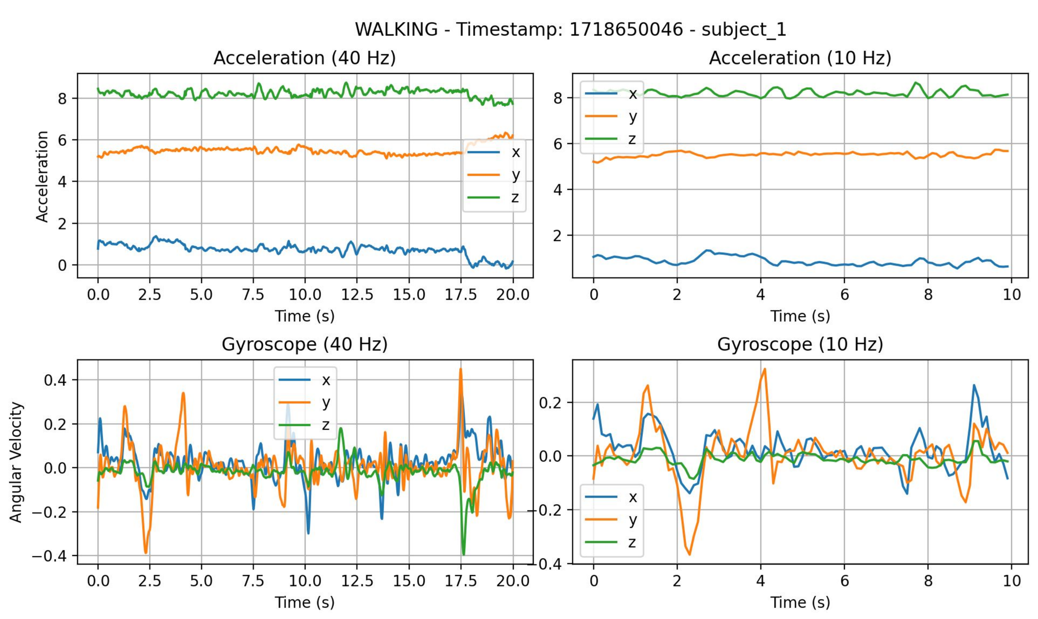
➔ Visualization: Graph the label reported by users to **visualize activities** over time and identify the main activities:

"Walking" (W)
"Sitting" (S)
"Lying down" (LD)
"Standing" (ST)

➔ Pattern Identification: Identify patterns and **correlations in sensory data** associated with different activities

(3) Data extraction & pre-processing:

Obtaining sensor data for analysis.



➔ Normalization: **Adjust the raw data** to have a uniform scale

➔ Segmentation: Divide the data into time windows consistent with the reported activities

➔ **Analytical analysis**: Compute the mean, standard deviation, and accuracy for raw IMU data

(4) Prompt generation for GPT:

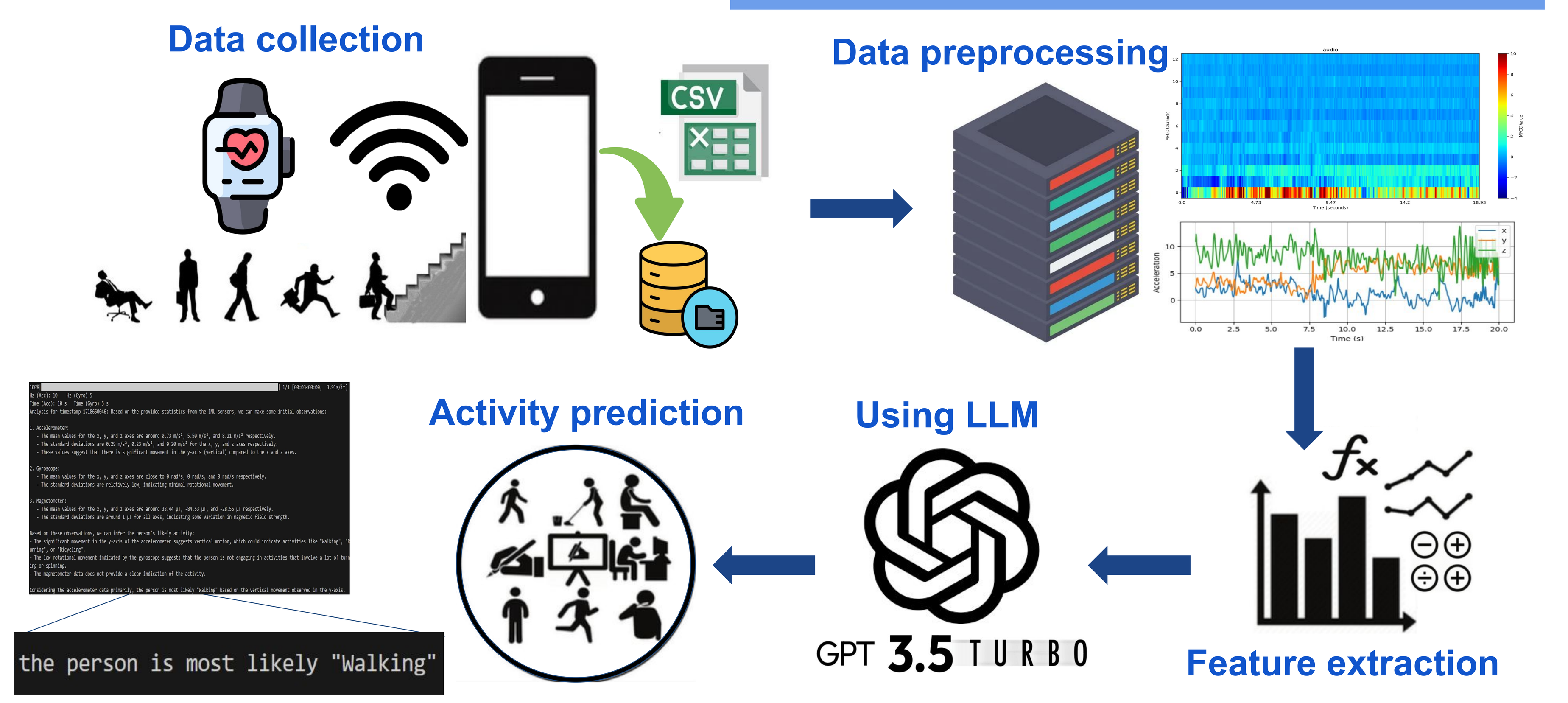
Prompt input

###Instruction: You are an expert in IMU-based human activity analysis.
###Question: The IMU data is collected from sensors attached to the user's phone with a sampling rate of **{Hz_reduced}** Hz. The IMU data is given in the IMU coordinate frame. The **three-axis** accelerations and gyroscopes are given below.
Accelerations: Gyroscope:
x-axis: {...}, y-axis: {...}, z-axis: {...} x-axis: {...}, y-axis: {...}, z-axis: {...}
**** Magnetometer/ Audio MFCC **Statistics**:
Three-axis / Channel # : **mean** = {...}, **std** = {...}, **var** = {...}
The person's action belongs to one of the following categories: **"Lying down", "Sitting", "Standing", "Walking"**. Could you please analyze and determine the person's action based on the given IMU readings? Please make an detailed **analysis step by step** and give me only the action.
Response: **{answer}**

➔ Individual Prompts: **Design different prompts** for GPT, incorporating Acc, Gyro, Magnet, and Audio values selectively

➔ Activity Prediction: Use of **ChatGPT-3.5 turbo** to predict activities based on raw HA data from all users

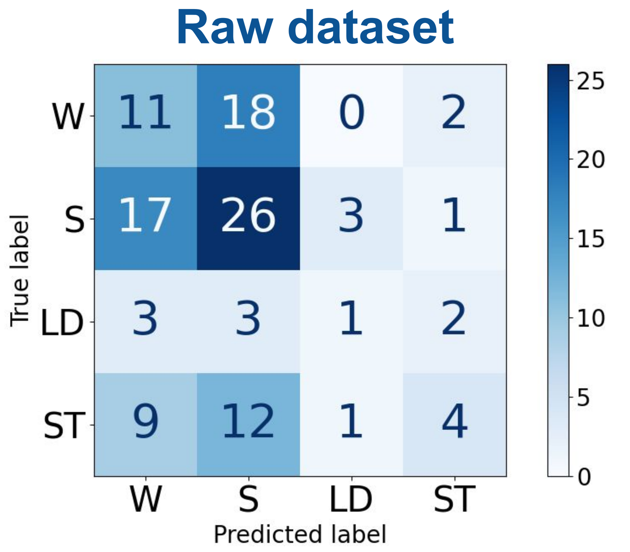
Workflow of HA prediction



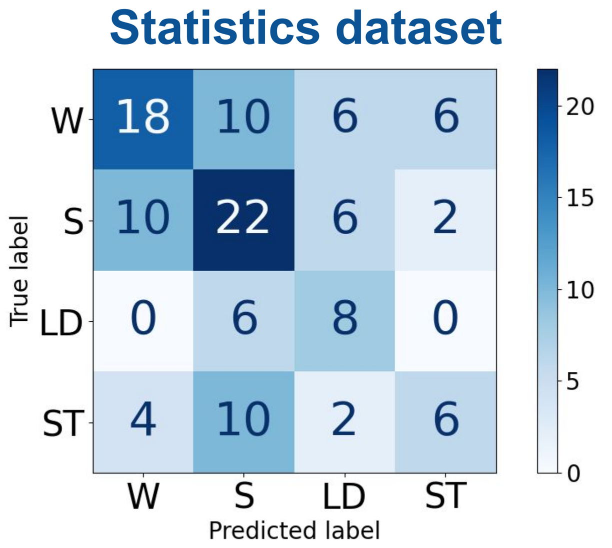
Results

	Combination of sensors	Precision	Recall	F1	Accuracy
Raw	Acc	0.400	0.842	0.542	0.495
	Acc and Gyro	0.465	**0.976**	0.630	0.548
	Acc and Magnet	0.486	0.857	0.621	0.532
	Acc and Audio	0.351	0.929	0.510	0.468
	Acc, Gyro and Magnet	**0.544**	0.804	**0.649**	**0.565**
	Acc, Gyro and Audio	0.360	0.939	0.521	0.452
	Acc, Magnet and Audio	0.337	0.806	0.475	0.385
	Acc, Gyro, Magnet and Audio	0.365	0.795	0.500	0.392
Statistics	Acc	0.302	0.839	0.444	0.375
	Acc and Gyro	0.561	**0.871**	0.667	0.569
	Acc, Gyro and Magnet	**0.588**	0.857	**0.696**	**0.588**
	Acc, Gyro and Audio	0.487	0.800	0.522	0.470
	Acc, Gyro, Magnet and Audio	0.385	0.786	0.485	0.446

Raw dataset



Statistics dataset



Experimental Setup
- **116 minutes** of data collected in-the-wild (down-sample at 10 Hz)
- 5 users: 4 main activity-labels
- Metric: **F1-score** and **accuracy**

Key takeaways
- Statistical features reduce input size for GPT and increase accuracy
- Including **Gyro** and **Magnet sensor features** significantly enhance the accuracy

Conclusions

- Zero-shot reasoning and chain-of-thought prompts decrease dependency on large datasets and extensive training times and reduce resource requirements
- Statistical features improve prediction accuracy and reduce computational cost

Future Works

- Use other LLMs to compare prediction accuracy with previous analyses
- Design a virtual assistant that takes advantage of this information to provide personalized recommendations about people's lifestyles and implement with other wearable technology

References

[1] Park, H., Kim, N., Lee, G. H., & Choi, J. K. (2023). *MultiCNN-FilterLSTM: Resource-efficient sensor-based human activity recognition in IoT applications.* Future Generation Computer Systems, 139, 196-209.

[2] Xia, Kun, Jianguang Huang, and Hanyu Wang (2020). LSTM-CNN architecture for human activity recognition. IEEE Access.

[3] Ji, S., Zheng, X., & Wu, C. (2024). *HARGPT: Are LLMs Zero-Shot* Human Activity Recognizers?. arXiv preprint arXiv:2403.02727.