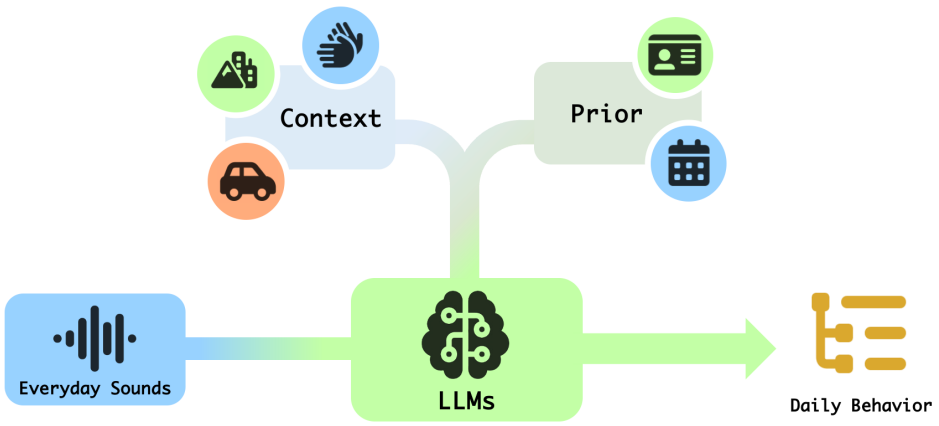
Behavior Prediction from Everyday Sounds via LLMs with Multi-sensor Context and Priors

Abstract:

Sounds are ubiquitous in our daily life—from the wind during a run to tableware clinking at meals. With context and prior knowledge, humans can easily infer ongoing activities. Machines, however, have long lacked this ability—until recent advances in large language models (LLMs), which enable real-world multimodal behaviour understanding.

Our work explores how LLMs can infer fine-grained user behaviour from everyday audio, supported by multi-sensor context and user priors. We use smartwatch-recorded audio as the main input, enhanced by signals from IMU, GPS, and other sensors, along with user profiles and routines. These cues help distinguish acoustically similar events and support higher-level reasoning—such as differentiating walking to class from walking in a mall by GPS and activity data. Our system aim of fuses these inputs into LLM-based reasoning to generate daily activity summaries and behaviour reports, enabling intelligent prediction and self-reflection.



Dataset:

Given the specificity of data required, no public dataset fits our project needs. Therefore, we will construct a custom dataset via smartwatches, mainly consisting of:

- User Questionnaire: Gathers basic demographics and routine behaviors as prior knowledge.
- IMU Data: Captures motion signals to identify specific limb movements.
- GPS Location: Provides spatial data to infer environment and mobility states (e.g., walking, running, commuting).
- Heart Rate: Records physiological responses to assist in activity classification.
- Raw Audio: Continuously records sound for behavioral inference by the model.

Author: Xingjian Tian, Fangfei Gou, Xin Tang, Zeyu Wang, Yingke Ding

Tsinghua University