Closing the Loop in Multimodal Edge AI for Healthcare

[Half-day tutorial] Accelerating the development of real-time closed-loop distributed sensing and edge AI processing systems

Description

In this interactive half-day tutorial, we will teach the use of a new developer-friendly framework for end-to-end edge AI system design and deployment to the wider CPS/IoT audience, using a visually entertaining use case of assisted mobility devices (lower-body exoskeleton or a leg prosthesis). The participants will gain practical experience and an intuition for balancing design decisions of real-time distributed multimodal edge AI and AIoT (Artificial Intelligence of Things) sensing systems. The session will cover real-world relevant components - multimodal distributed sensing and high-quality dataset curation, data fusion and streaming multimodal inference strategies, hardware-aware embedding of AI.

The guided hands-on part of the session will help participants pass the learning curve of adapting the framework with new sensing, actuation, and AI processing functionality, and integrating it with the streamlined AI workflows and toolchains. At the end of the tutorial, participants will (1) have a more cohesive view on the end-to-end edge AI system development, (2) know how to accelerate their prototyping efforts using HERMES[1], (3) have a locally-run tool for curation of high-quality multimodal datasets.

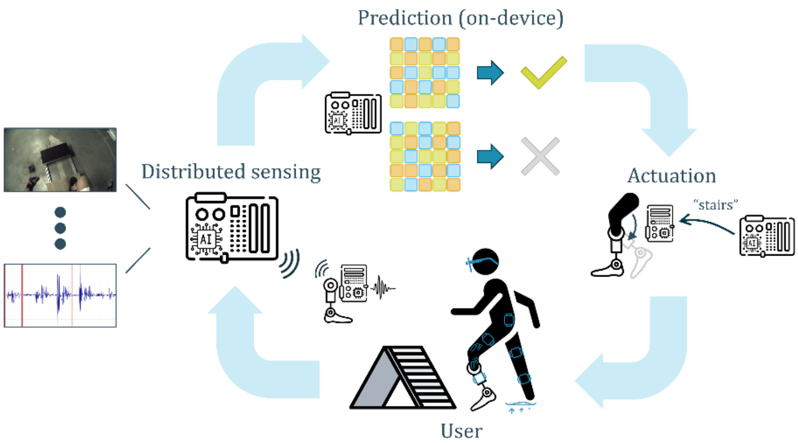

Figure 1. Overview of a closed-loop multimodal edge AI system.

Motivation

Healthcare applications increasingly demand low-latency, adaptive, and intelligent systems capable of operating at the edge. Traditional cloud-based AI approaches cannot meet these stringent requirements due to latency, connectivity, and privacy constraints. Diverse R&D use cases, from dataset curation to wearable intervention devices, in fixed labs and in the wild, share the common need for reliable continuous multimodal sensing. These multimodal signals from distributed sensors encapsulate rich information about the state of the user and the environment, exceeding the depth of insights represented by a single modality. However, multiple

[1] Our open-source framework on GitHub for physiological sensing and edge AI processing.

technical cross-disciplinary challenges and a limited number of frameworks inhibit the efforts of researchers in their core objectives. By streamlining the system design flow and leveraging reusable extensible community-driven frameworks, we can accelerate the process of developing novel edge AI and sensing systems for transformative healthcare applications.

Rationale for CPS/IoT community and target audience

The evolution from IoT to AIoT is reshaping how connected devices operate – moving from simple data collection into cloud-based buckets with batch processing, to intelligent collaborative decision-making systems at the edge. For the broader CPS and IoT community across domains, this transition introduces new challenges and a paradigm shift – both, in its design and in implementation.

Through this tutorial, participants will learn how to design an AIoT system that closes the loop in assisted mobility applications, integrating multi-modal sensing, embedded AI, and adaptive control. While focused on healthcare, the same theoretical concepts can extrapolate to similar use cases in a broader set of domains. Beyond theory, the session will demonstrate these concepts in practice with hands-on prototyping exercises, enabling attendees to gain practical experience in building closed-loop AIoT solutions to their specification.

Format and technical requirements

This hands-on interactive tutorial requires the participants to bring their personal laptop with Python installed. The organizing team will provide sensors and the final assisted mobility device. Together with the participants, they will collaboratively integrate support for a “custom” wearable sensor, the high-level controller of an assisted device, and a suitable AI model in a code-along format, with short exercises. It will rely on the released open-source HERMES framework as a launchpad for closed-loop sensing and edge AI systems prototyping. The session requires a room with WiFi, space for 20-50 people, and an open area at the conference venue of ~20m^2 with improvised obstacles to demonstrate the jointly completed prototype.

Tentative roadmap

1. **Welcome and introduction** (15 min): Introduce an end-to-end Edge AI pipeline and the HERMES framework.
2. **Multimodal data fusion** (30 min): Provide overview of key design knobs and their trade-offs in multimodal sensing and realtime AI inference.
3. **[Hands-on] Design closed-loop system** (90 min): Extend HERMES with a new sensor, perform a small-scale multimodal data collection, and design an AI model. Benchmark the resource utilization and latency of the closed-loop system. Iteratively adapt the AI model to fit the application constraints.
4. **[Hands-on] Real-time Demonstrator** (60 min): Deploy the pretrained model onto embedded hardware [Jetson, Raspberry Pi, LattePanda] through PyTorch or Edge Impulse and run it on the exoskeleton or prosthesis use case. Validate the functionality of the system.
5. **Future Directions** (30 min): Finalize the tutorial with concluding remarks, along with a discussion on open challenges and collaboration opportunities.

Organizing team

Maxim Yudayev (KU Leuven)
maxim.yudayev@kuleuven.be
in/maxim-yudayev

Maxim is a graduating PhD in the team of Prof. Bart Vanrumste at the e-Media Research Lab of KU Leuven in Belgium. He develops novel software and hardware edge AI technologies that enable real-time distributed sensing and continuous multimodal processing of sensor data for intelligent ambient healthcare applications. Prior to KU Leuven, Maxim worked at the Nokia Bell Labs research centre in Antwerp, Belgium, where he patented a look-ahead computation hardware mechanism for efficient and energy-preserving AI computer architectures. With domain expertise in digital design, networks, electronics, distributed parallel software, and embedded systems, he focuses on delivering practical solutions from the systems-level perspective to meet complex R&D challenges.

Jona Beysens (KU Leuven)
jona.beysens@kuleuven.be
in/jona-beysens

Jona is an Assistant Professor in the e-Media Research Lab at KU Leuven. His research interests include tinyML and wireless networking, with a focus on designing hardware-aware AI algorithms tailored to low-power embedded systems for innovative healthcare applications. Prior to KU Leuven, he was a Senior R&D Engineer at CSEM in Switzerland, where he led several national and international (EU) tinyML-related research projects. He obtained a PhD from KU Leuven in 2020 on Robust Visible Light Communication Networks (LiFi). With his team at CSEM, he won the first prize of the international tinyML Smart Weather Station Challenge 2022, where they built an acoustic weather station powered by embedded AI. During his time at CSEM, he was a member of the Swiss EdgeAI Foundation branch, consisting of members from both Swiss industry and academia.

Louis Flynn (Vrije Universiteit Brussel)
louis.flynn@vub.be

Louis is a leading postdoctoral researcher in human-robot interaction and rehabilitation robotics at the Robotics & Multibody Mechanics (R&MM) and BruBotics research groups of Vrije Universiteit Brussel (VUB) in Belgium. His work focuses on the mechatronic design and advanced control of lower-limb assistive technologies, including active prostheses (like the *CYBERLEG X-Leg*) and exoskeletons. Dr. Flynn is a key contributor to the design of adaptive controllers that use human-in-the-loop optimization (HILO) strategies to personalize robotic assistance of intelligent mobility devices, reduce user effort, and maximize functional ability of the wearer. His research is dedicated to translating compliant actuation and bio-inspired robotics into safe, energy-efficient, and user-accommodating devices for real-world impact.