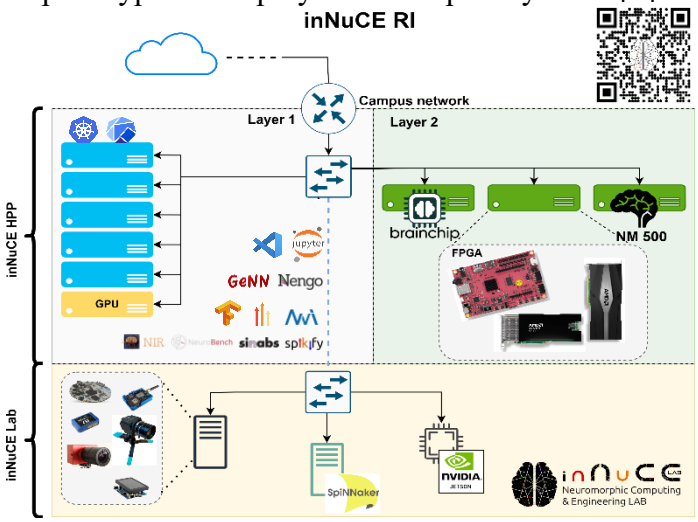
The inNuCE Research Infrastructure for Neuromorphic AIoT Apps Prototyping

Gianvito Urgese, Vittorio Fra, Andrea Pignata, Giuseppe Fanuli, Walter Gallego Gomez, Riccardo Pignari, Michelangelo Barocci, Benedetto Leto, Salvatore Tilocca, Nicola Cassetta, Enrico Macii
inNuCE Lab within the EDA Group @Politecnico di Torino, Turin, Italy

Industry 5.0 and AIoT increasingly require low-power, low-latency, and privacy-preserving computation. Neuromorphic hardware (HW) provides a brain-inspired and energy-efficient paradigm, but remains costly, fragmented, and difficult to access [1-3]. To address this barrier, we propose NMLOps, a set of practices inspired by MLOps that standardizes the development of Neuromorphic AIoT applications across the entire lifecycle, from data acquisition to model deployment and operation [4]. To support this methodology, we developed the inNuCE Research Infrastructure (RI), a facility that enables researchers and industry to prototype and deploy neuromorphic systems [4].

The inNuCE RI (in Fig.) consists of two main components: (i) the Heterogeneous Prototyping Platform (HPP), a cloud-based Prototype-as-a-Service environment for neuromorphic/digital co-design, and (ii) a dedicated laboratory hosting neuromorphic boards, sensors and embedded system solutions for direct experimentation.



The HPP supports multiple platforms, including Brainchip Akida, SynSense Speck/Xylo, SpiNNaker 2, and Intel Loihi 2, as well as Nvidia Jetson, AMD FPGAs, ST microcontroller, and GPU servers. These resources are containerized with their SDKs to ensure reproducible environments.

The NMLOps workflow covers model definition, training, simulation, hardware deployment, and results analysis, and supports both expert tools (VSCode, Jupyter, Kubeflow pipelines) and simplified interfaces.

The approach is demonstrated on several AIoT use cases, including human activity recognition [5], Braille reading [6], event-based gesture recognition, memory semantization, navigation tracking, and constraint satisfaction [7].

Results demonstrate that the inNuCE-RI lowers the entry barrier for neuromorphic computing by providing scalable cloud access and physical laboratory integration. By enabling direct comparison of models across diverse hardware, it accelerates prototyping and supports reproducibility. Integration with EBRAINS-Italy ensures sustainability, nationwide access, and alignment with the European EBRAINS ecosystem. The infrastructure thus represents a key enabler for bridging neuroscientific insights with industrial AIoT applications.

[1] Urgese, Gianvito, et al. "Powering the next-generation IoT applications: new tools and emerging technologies for the development of Neuromorphic System of Systems." FiN 17 (2023): 1197918. DOI:10.3389/fnins.2023.1197918

[2] Muir, Dylan Richard, and Sadique Sheik. "The road to commercial success for neuromorphic technologies." NC 16.1 (2025): 3586. DOI:10.1038/s41467-025-57352-1

[3] Kudithipudi, Dhireesha, et al. "Neuromorphic computing at scale." NC 637.8047 (2025): 801-812. DOI:10.1038/s41586-024-08253-8

[4] Urgese, Gianvito, et al. "The inNuCE Research Infrastructure and the Neuromorphic MLOps for AIoT prototyping." IEEE Internet of Things Journal (2026). DOI:10.1109/JIOT.2026.3663902

[5] Leto, Benedetto, et al. "Variable-precision neuromorphic state space model for on-edge activity classification." *Future Generation Computer Systems* (2025): 108193. DOI:10.1016/j.future.2025.108193

[6] Müller-Cleve, Simon F., et al. "Braille letter reading: A benchmark for spatio-temporal pattern recognition on neuromorphic hardware." FiN 16 (2022): 951164. DOI: 10.3389/fnins.2022.951164

[7] Pignari, Riccardo, et al. "Efficient solution validation of constraint satisfaction problems on neuromorphic hardware: the case of Sudoku puzzles." IEEE Transactions on Artificial Intelligence (2025). DOI: 10.1109/TAI.2025.3536428