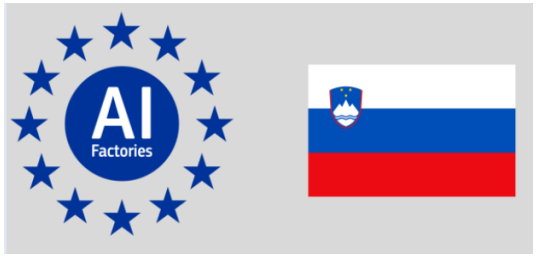
AI Lifecycle Orchestration Expertise

Data collection, pre-processing, validation, MLOps, energy footprint computation



Carolina Fortuna

https://sensorlab.ijs.si

Jožef Stefan Institute, Ljubljana Slovenia,

Email: carolina.fortuna@ijs.si

IRC AI

International Research Centre
on Artificial Intelligence
under the auspices of UNESCO

SensorLab

JSI

SNS **NANCY, OPTICON**, MSCA **GENOME**

Jozef Stefan Institute

- The institute was founded by the State Security Administration (Yugoslavia) in 1949 for atomic weapons research.

- It is named after the distinguished 19th-century physicist Josef Stefan (Slovene: Jožef Stefan), best known for his work on the Stefan–Boltzmann law of black-body radiation.

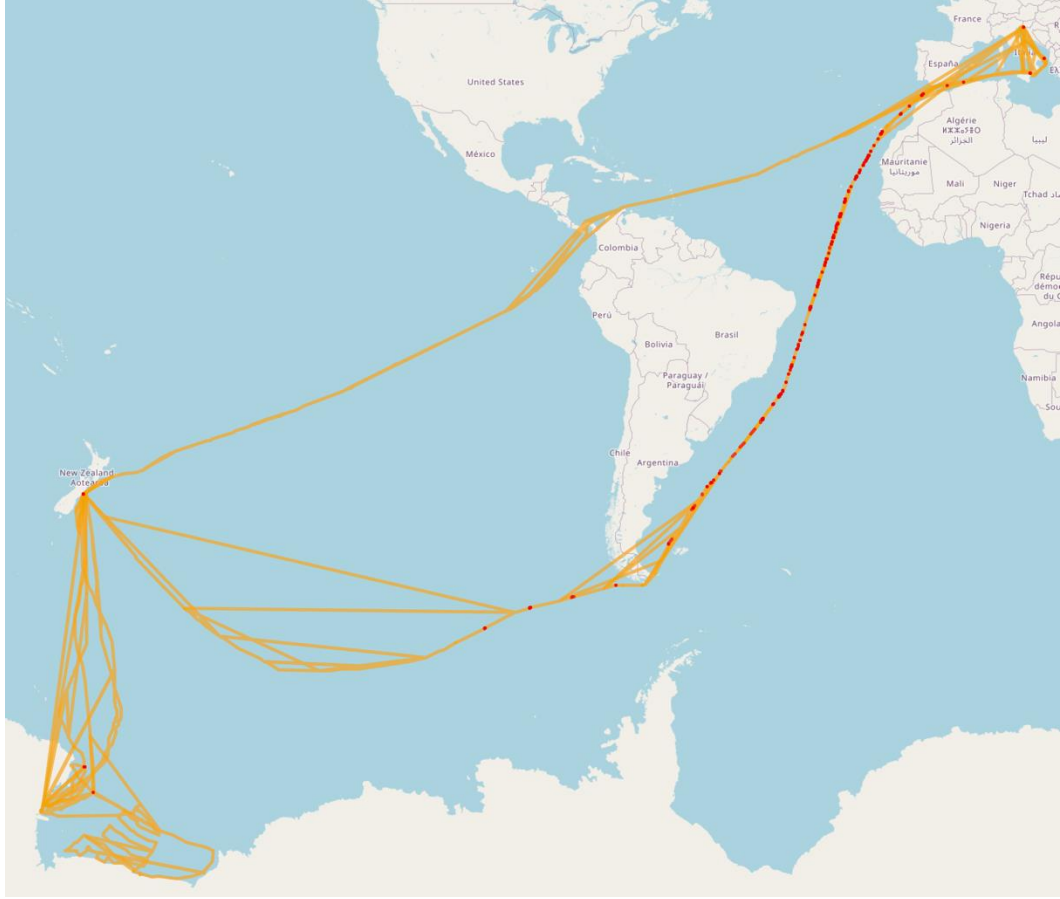
- The main research areas are physics, chemistry, molecular biology, biotechnology, information technologies, reactor physics, energy, and the environment.
- > 1,170 employees, of whom > 568 are PhD holders

STREAM B

Collection, Generation and Validation of Datasets suitable for training AI models for 6G Networks

Data Collection

- The right data for the right problem!
- Unique dataset for 3D networks:
  - a unique 200-day dataset of raw GNSS single-frequency L1 receiver
    - October 7, 2024, to April 24, 2025.
    - multi-constellation L1-band observations, including position and time (UBX-NAV-PVT), satellite metadata (UBX-NAV-SAT), RF front-end diagnostics (UBX-MON-RF), FFT results of the received signal spectrum (UBX-MON-SPAN), raw satellite pseudorange and carrier-phase data (UBX-RXM-MEASX)
    - intermittent Galileo SAR return-link messages (UBX-RXM-RLM).



Voyage to the Frozen Continent: A Comprehensive GNSS Dataset From a Ship's Expedition to Antarctica
Blaž Bertalanič, Franc Dimc, Matej Bažec, Aljaž Blatnik, Scientific Data, 2026.
https://www.nature.com/articles/s41597-025-06366-x

eCAL - the Energy Cost of Artificial Intelligence Lifecycle

- Energy consumption for data collection
- An open source, modular and extensible simulator for eCAL: https://github.com/sensorlab/eCAL

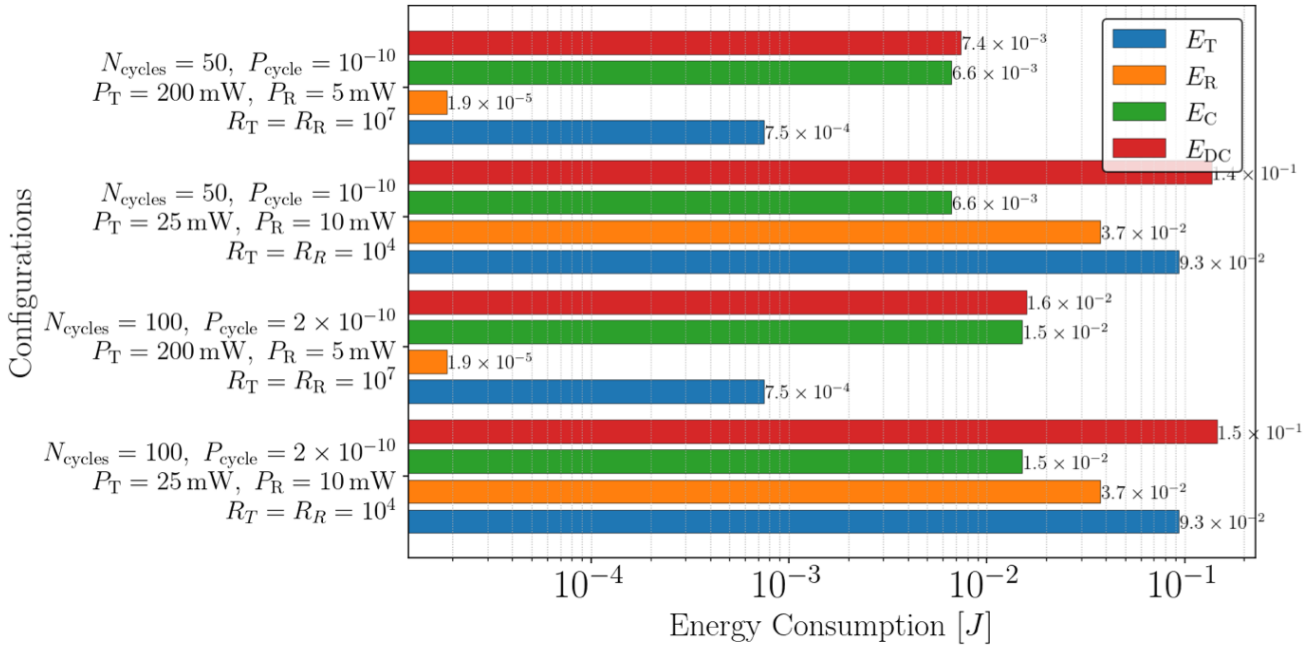


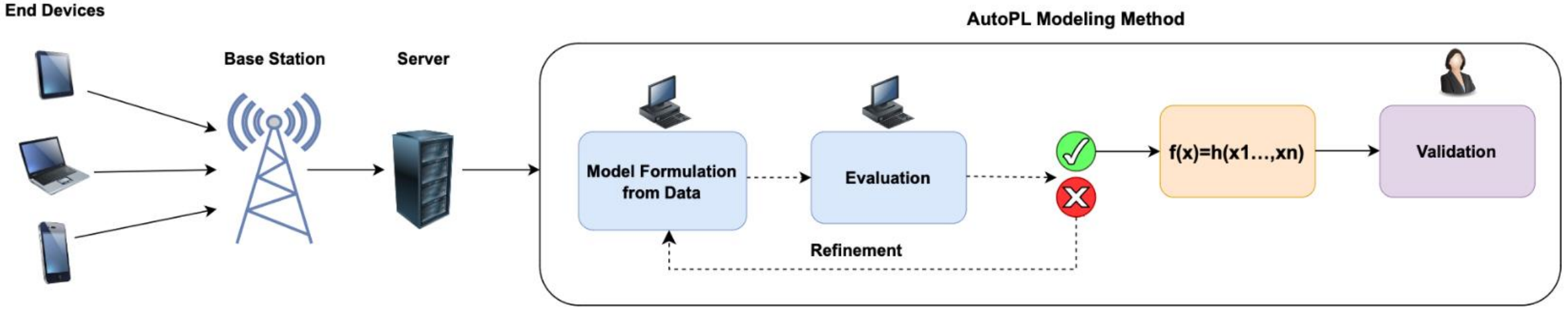
Fig. 3: Energy consumption of data collection (E_{DC}) (log scale) with its components and configurations, including transmission energy (E_{T}), receiving energy (E_{R}), and computational energy (E_{C}) with $N_{\text{S}} = 256$, $RR_{\text{DP},l} = RR_{\text{CP},l} = 1$, $OH_{\text{DP},l} = 10\%$, $OH_{\text{CP},l} = 5\%$, and $\gamma_l = 0.5$.

The Energy Cost of Artificial Intelligence Lifecycle in Communication Networks, Shih-Kai Chou, Jernej Hribar, Vid Hanžel, Mihael Mohorčič, Carolina Fortuna, IEEE Journal on Selected Areas in Communications, 2025

Data Generation

- Synthetic data generation
  - Generative AI techniques
  - Constrained AI techniques

- **Interpretable AI techniques that learn the theoretical expression and then use that to produce additional data**



Towards Automated and Interpretable Pathloss Approximation Methods, Ahmad Anaqreh, Shih-Kai Chou, Irina Barašin, Carolina Fortuna, AAAI 2025 Workshop on Artificial Intelligence for Wireless Communications and Networking: Philadelphia, Pennsylvania, USA, Mar 03, 2024.

Automated Modeling Method for Pathloss Model Discovery, Ahmad Anaqreh, Shih-Kai Chou, Mihael Mohorčič, Thomas Lagkas, Carolina Fortuna, 2025, https://arxiv.org/abs/2505.23383v2

SABER: Symbolic Regression-Based Angle of Arrival and Beam Pattern Estimator, Shih-Kai Chou, Mengran Zhao, Cheng-Nan Hu, Kuang-Chung Chou, Carolina Fortuna, Jernej Hribar, 2025, https://arxiv.org/abs/2510.26340v1

Data Validation

From the domain experts PoV

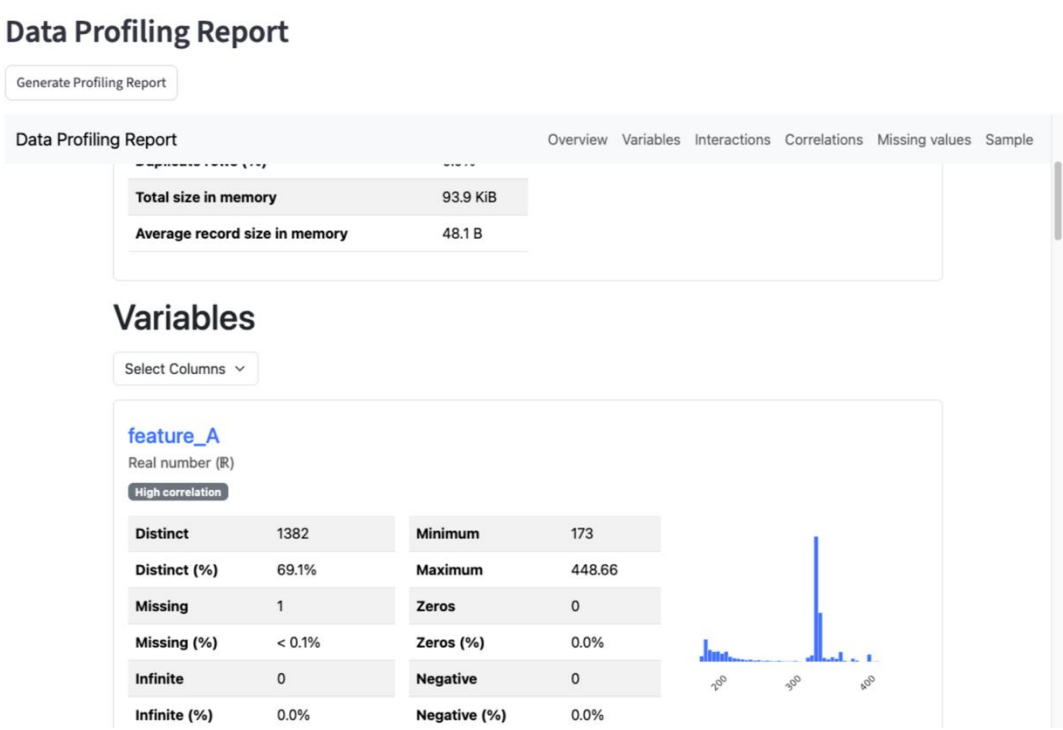
- Reported Jamming indicators
- Anomalies in wireless signals or instrumented infrastructure

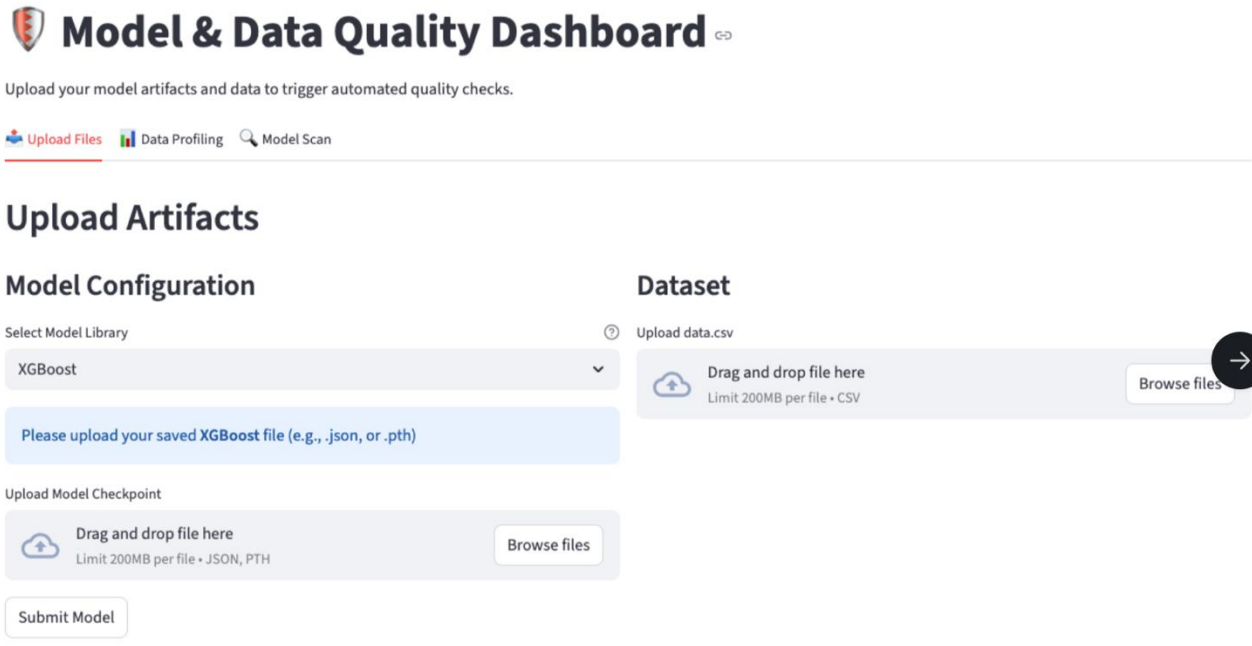
From the AI experts PoV

- Feature interaction and drift
- Correlation and causality
- Data leakage

From the regulators PoV

- AI act compliance





A Configuration-First Framework for Reproducible, Low-Code Localization, Tim Strnad, Blaž Bertalanič, Carolina Fortuna, 2025

eCAL - the Energy Cost of Artificial Intelligence Lifecycle

- Energy consumption for data validation and pre-processing
- An open source, modular and extensible simulator for eCAL: https://github.com/sensorlab/eCAL

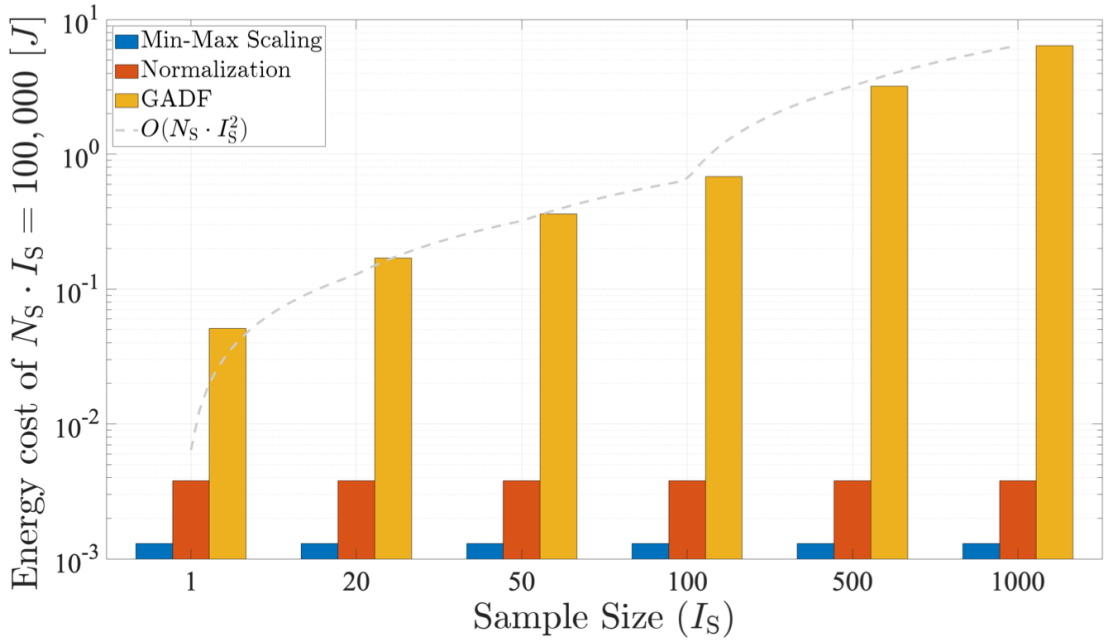


Fig. 6: Energy consumption (E_{pre}) of fixed data point ($N_S I_S$) across different preprocessing techniques and sample size (I_S).

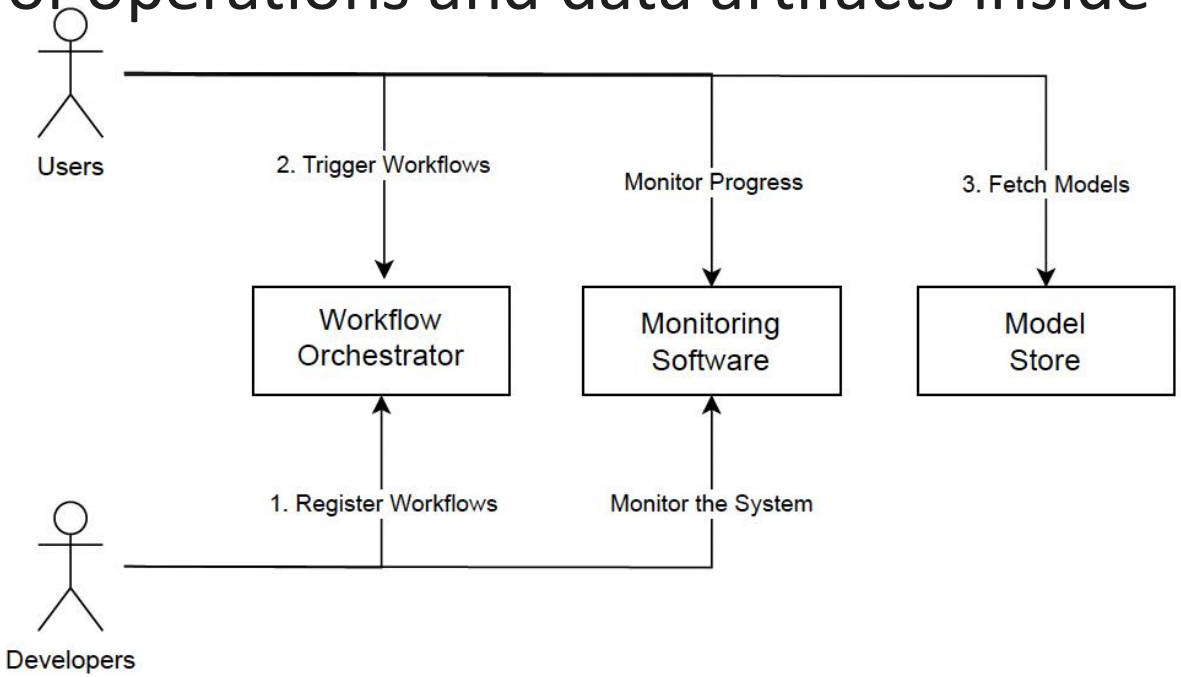
The Energy Cost of Artificial Intelligence Lifecycle in Communication Networks, Shih-Kai Chou, Jernej Hribar, Vid Hanžel, Mihael Mohorčič, Carolina Fortuna, IEEE Journal on Selected Areas in Communications, 2025

STREAM C

SNS Experimental Infrastructure

NAOMI

- A new MLOps solution for democratizing AI workflows at the edge, developed in SNS NANCY project.
- It’s like a factory, assembly line included, for AI models
- NAOMI architecture shows a flow of operations and data artifacts inside the AI/ML workflow.

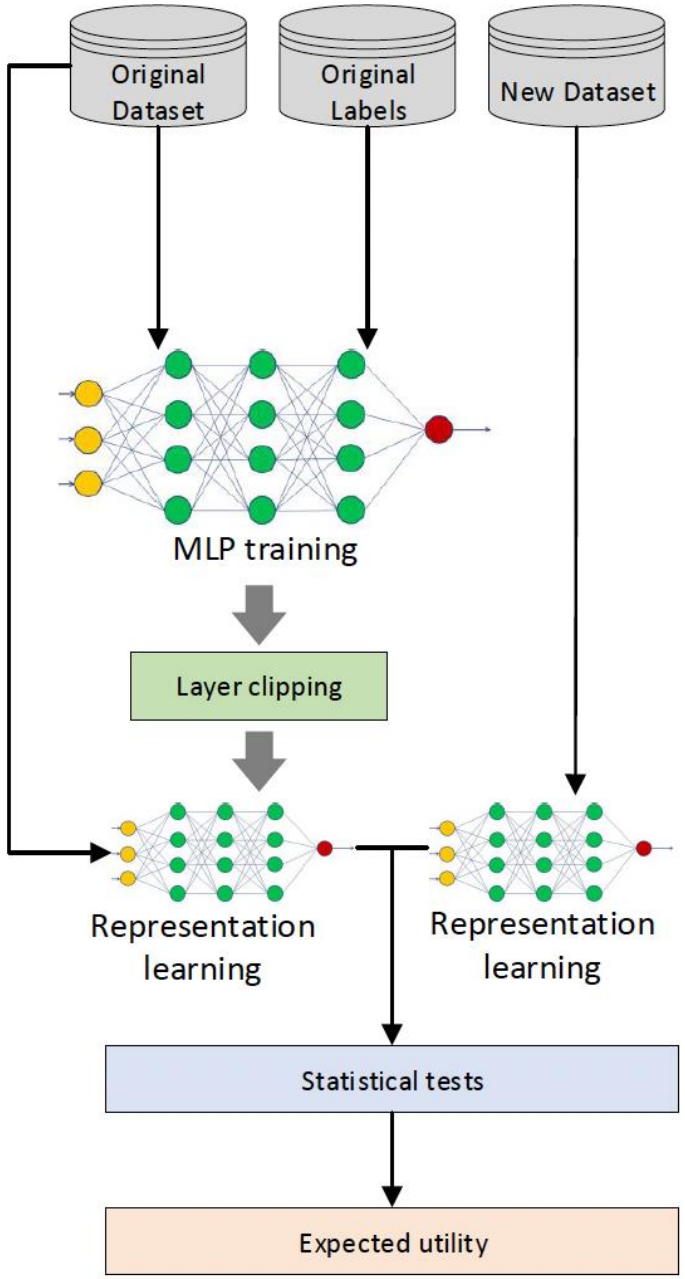


An Overview and Solution for Democratizing AI Workflows at the Network Edge. Andrej Čop, Blaž Bertalanič, Carolina Fortuna, Journal of Network and Computer Applications, 2025.

Self-Evolving Models

- Regular, scheduled
- On monitored metrics (NAOMI)
- **Feature Drift: ALERT (**SNS NANCY**)**
- Concept Drift
- Dataset drift

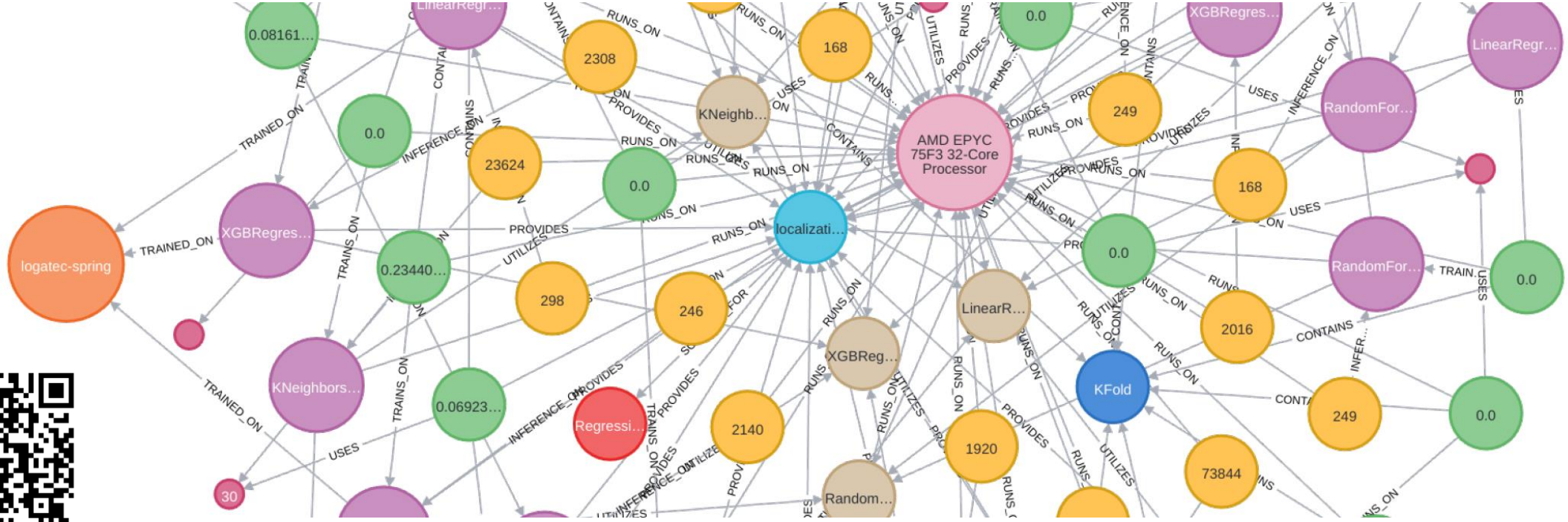
$$U = \frac{U_{KS} + U_{PSI}}{2}$$



A Representation Learning Approach to Feature Drift Detection in Wireless Networks
Athanasios Tziouvaras, Blaz Bertalanic, George Floros, Kostas Kolomvatsos, Panagiotis Sarigiannidis, Carolina Fortuna, 2025, IEEE Transactions on Emerging Topics in Computing, 2025, https://doi.org/10.48550/arXiv.2505.10325

MRM3: Machine Readable ML Model Metadata

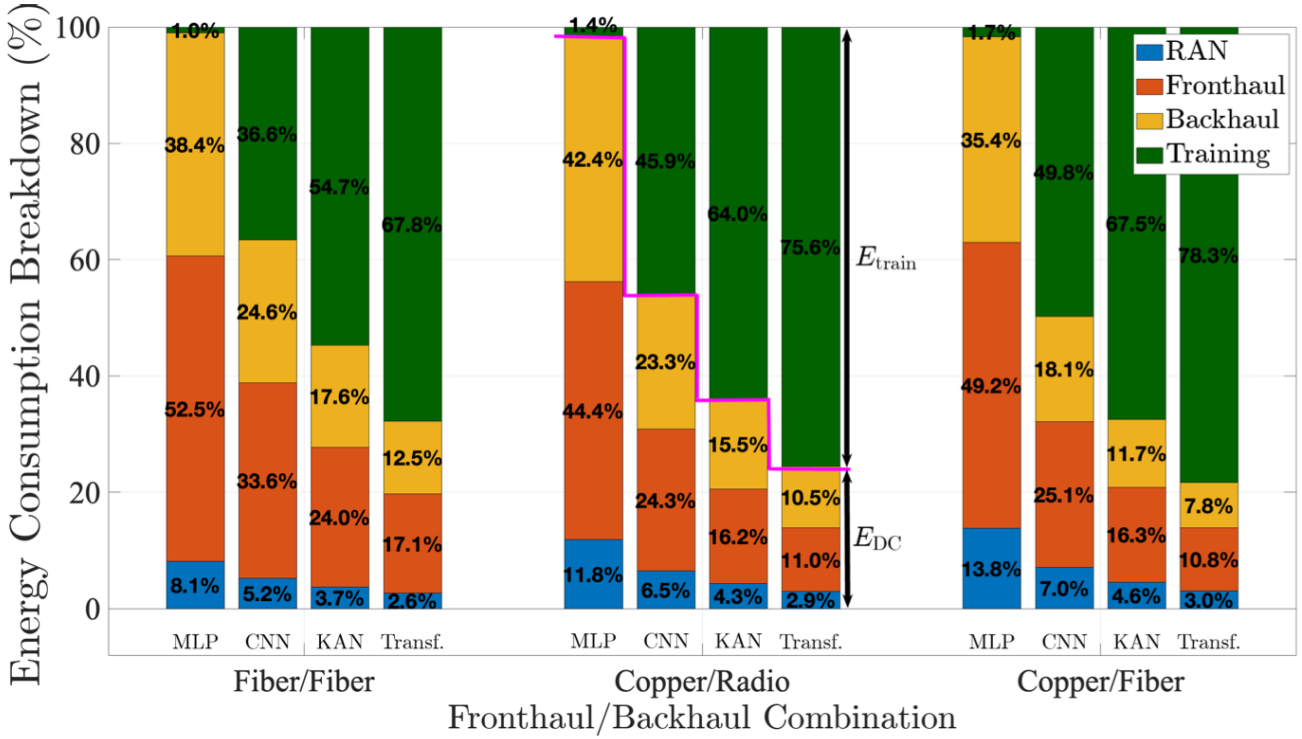
- Enables more expressive model description in the model repository
- Support for better informed and/or multi-objective model selection

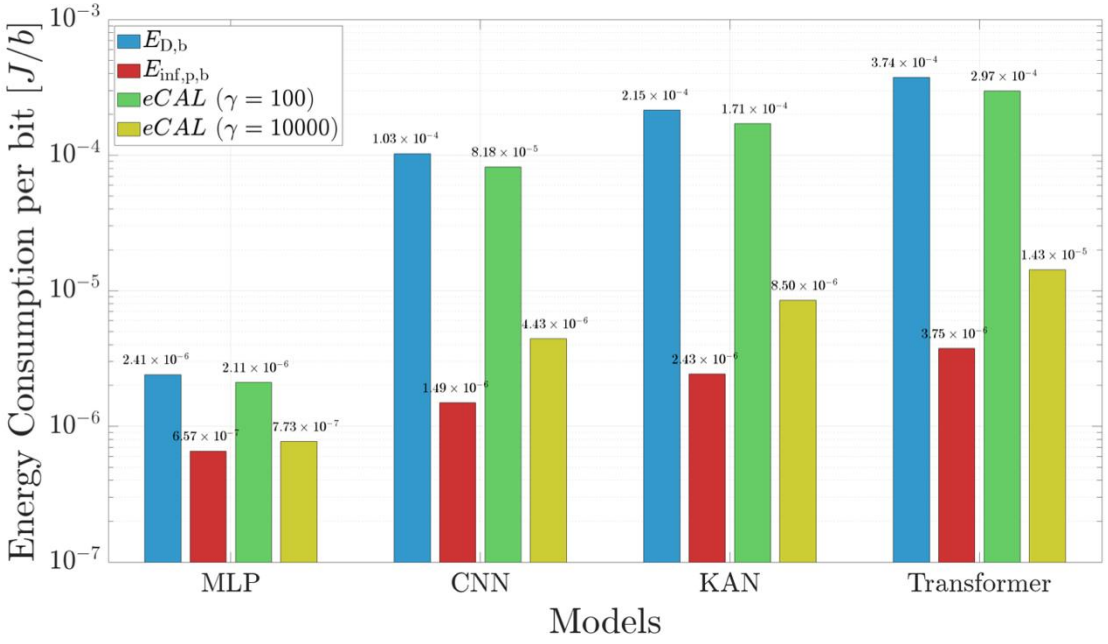


MRM3: Machine Readable ML Model Metadata, Andrej Čop, Blaž Bertalanič, Marko Grobelnik, Carolina Fortuna, MobiSys '25: Proceedings of the 23st [I. E. 23rd] ACM International Conference on Mobile Systems, Applications, and Services, 2025

eCAL - the Energy Cost of Artificial Intelligence Lifecycle

- An open source, modular and extensible simulator for eCAL: https://github.com/sensorlab/eCAL





The Energy Cost of Artificial Intelligence Lifecycle in Communication Networks, Shih-Kai Chou, Jernej Hribar, Vid Hanžel, Mihael Mohorčič, Carolina Fortuna, IEEE Journal on Selected Areas in Communications, 2025

Thank you!

Email: carolina.fortuna@ijs.si