



HAL
open science

Supporting Efficient Workflow Deployment of Federated Learning Systems across the Computing Continuum

Cédric Prigent, Alexandru Costan, Gabriel Antoniu, Loïc Cudennec

► To cite this version:

Cédric Prigent, Alexandru Costan, Gabriel Antoniu, Loïc Cudennec. Supporting Efficient Workflow Deployment of Federated Learning Systems across the Computing Continuum. SC 2022 - International Conference for High Performance Computing, Networking, Storage, and Analysis (Posters), Nov 2022, Dallas, United States. hal-03878254

HAL Id: hal-03878254

<https://inria.hal.science/hal-03878254>

Submitted on 29 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



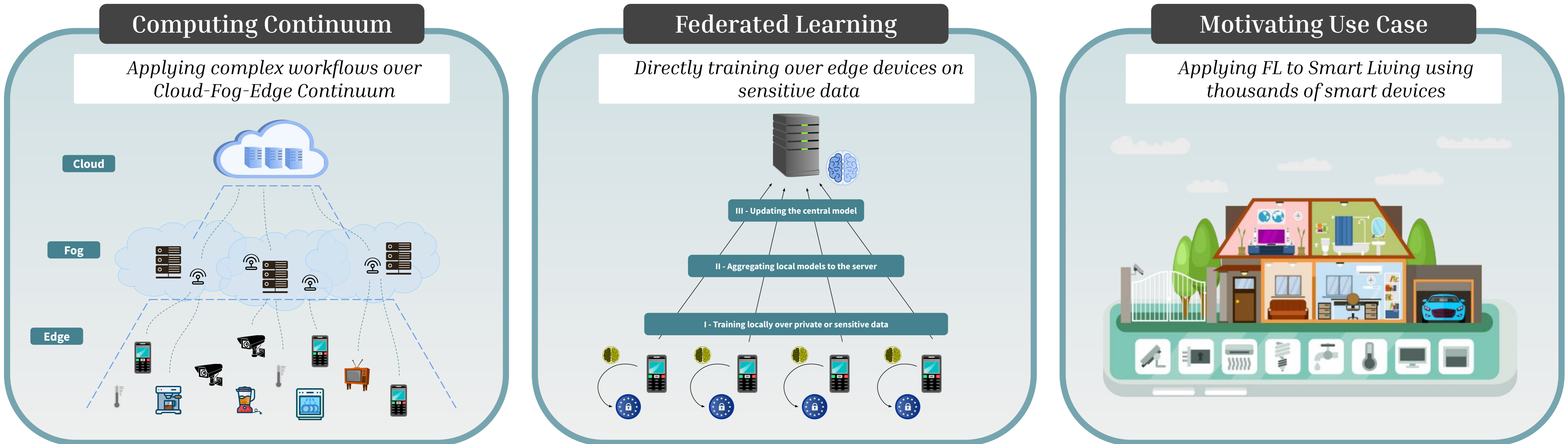
Distributed under a Creative Commons Attribution 4.0 International License

Supporting Efficient Workflow Deployment of Federated Learning Systems on the Computing Continuum

PhD Student: Cédric Prigent*, **Advisors:** Gabriel Antoniu*, Alexandru Costan*, Loïc Cudennec*
Contact: cedric.prigent@inria.fr

* University of Rennes, Inria, CNRS, IRISA - Rennes, France
 * DGA Maîtrise de l'Information - Rennes, France

Context

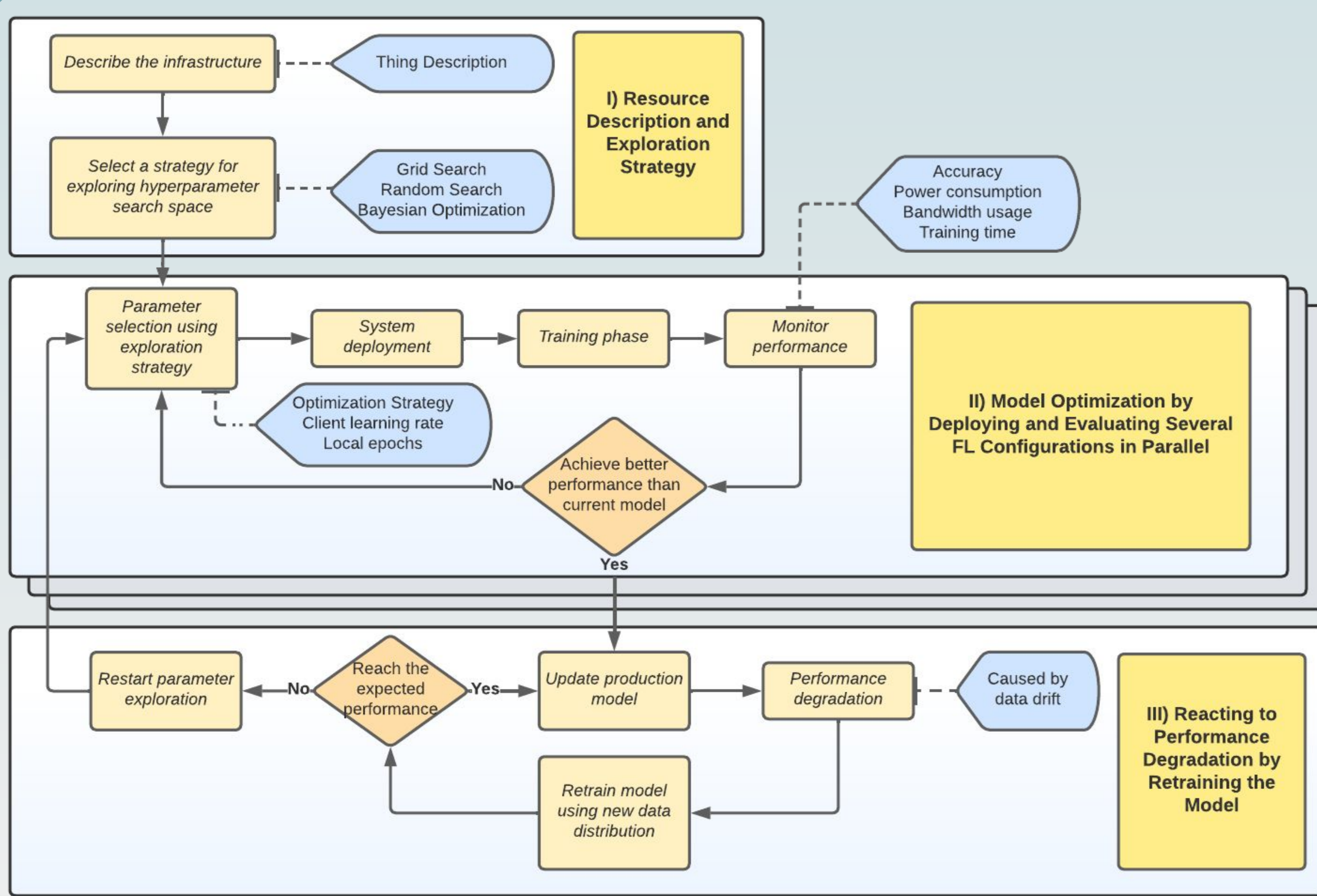


Problems

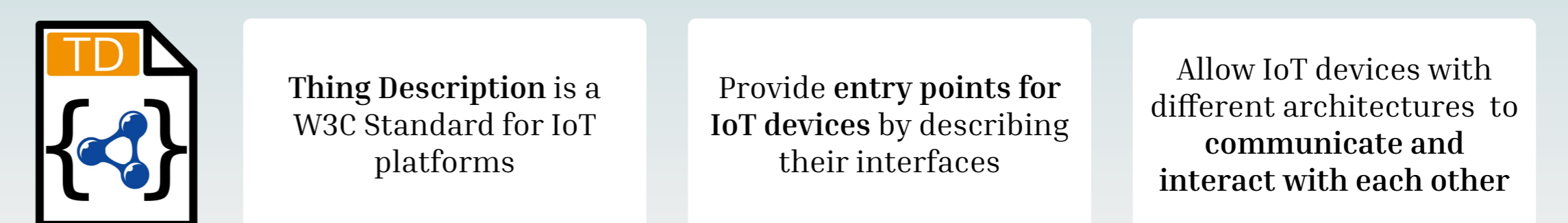
- Heterogeneous devices**: IoT objects relying on different architectures require extra efforts to be supported in a same application (supporting different standards)
- Hyperparameter tuning**: Federated Learning introduces several additional hyperparameters in contrasts with standard ML, requiring deeper optimization
- Inconsistent performance**: Devices with very different usage result in data drift and inconsistent performance of the model among clients

Our Proposal

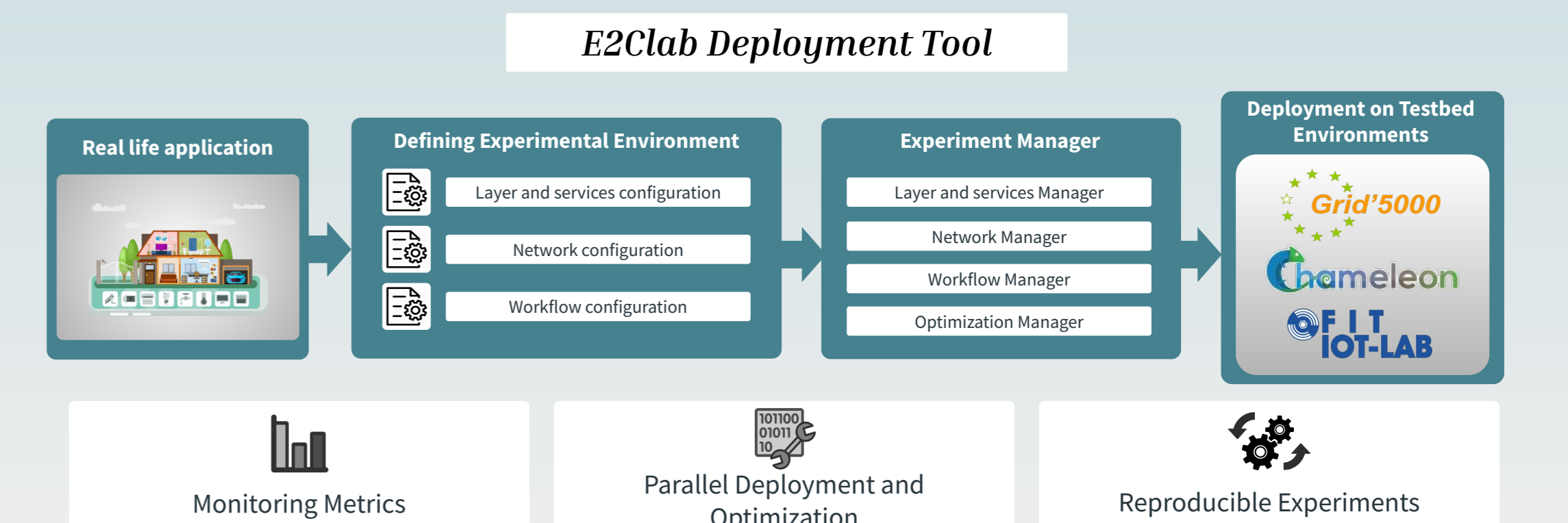
A Workflow for Efficient FL Deployment and Model Optimization



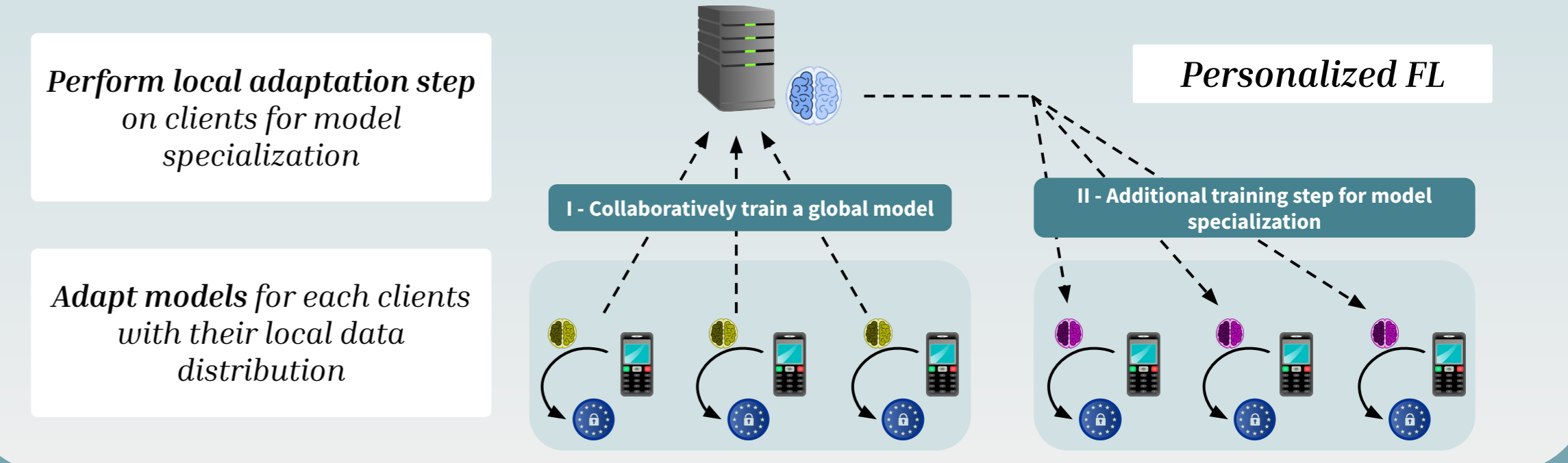
How to improve interoperability across IoT platforms ?



How to automatically deploy and monitor FL solutions ?



How to ensure good model performance on each client ?



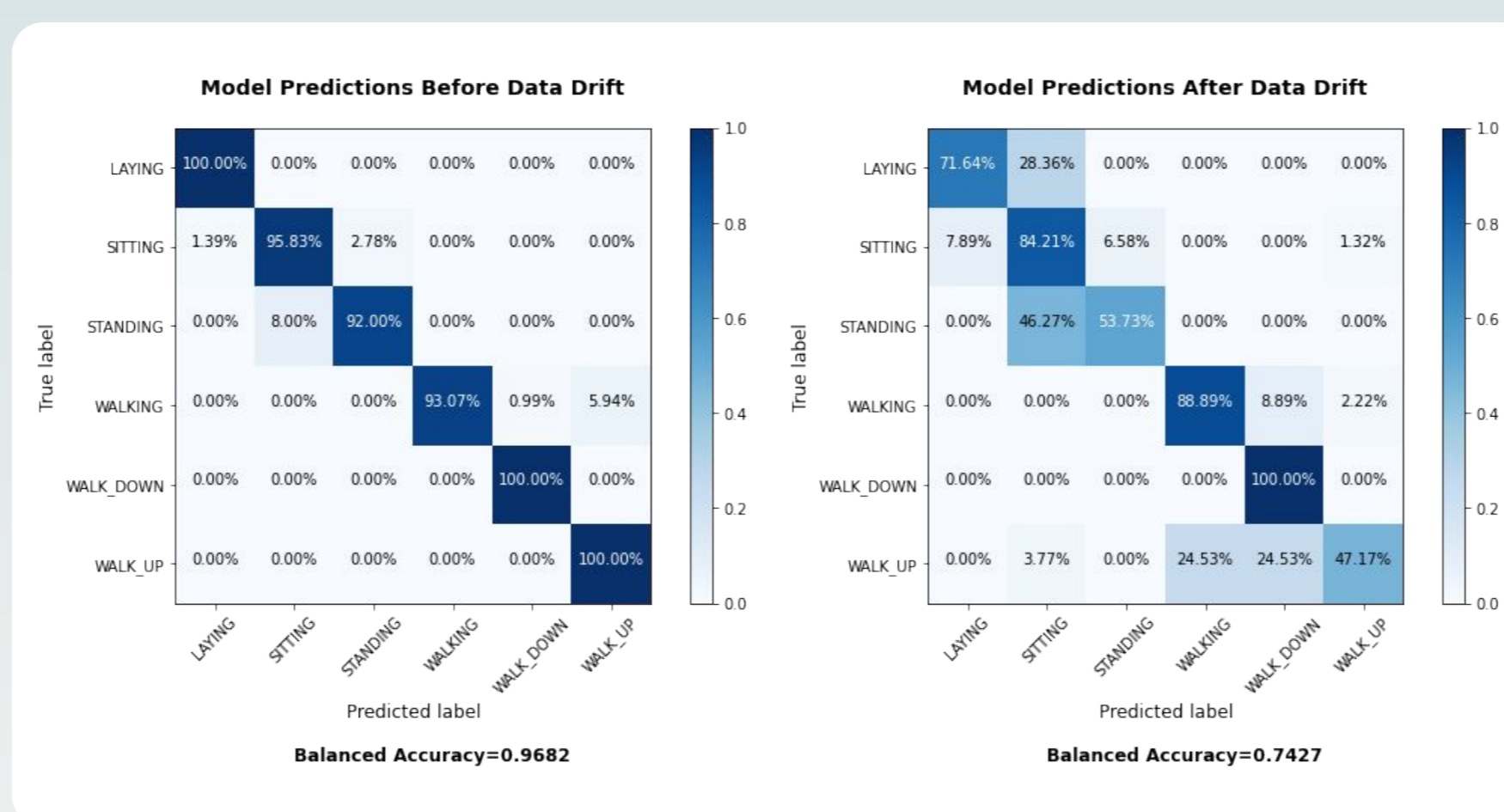
Early Results

Performing FL training over smartphone data from 8 users for recognition of Daily Living Activities

Impact of client learning rate and number of local epochs on training



Impact of data drift on model performance



Takeaways

We provided a solution to automatically deploy and optimize FL workloads in heterogeneous environments using formal description of the underlying infrastructure, hyperparameter optimization and model retraining in case of data drift.

Open Questions

How to determine which model is performing better considering several conflicting metrics and clients with different data distribution? (e.g optimizing energy consumption, accuracy of the model and bandwidth usage)

Considering constrained-devices, how often should models be trained? (minimizing battery usage of the device)