AI SPRINT
AI IN SECURE PRIVACY-PRESERVING COMPUTING CONTINUUM

Future challenges of the Cognitive Cloud

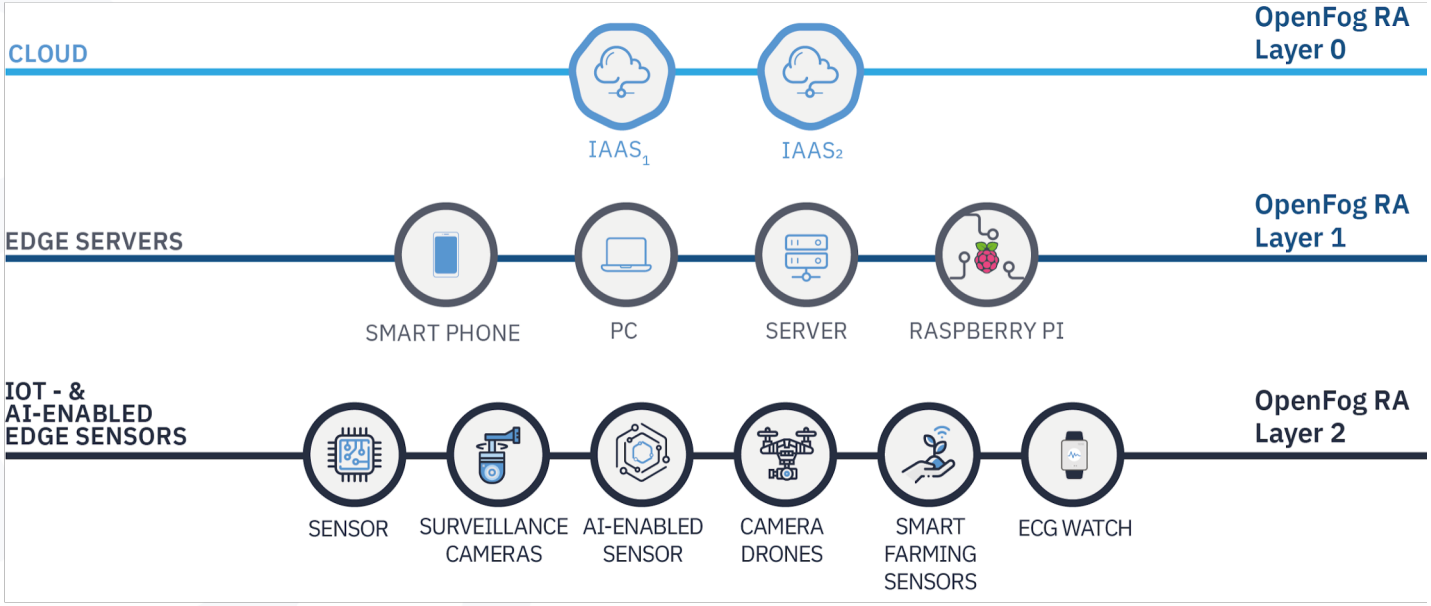
Danilo Ardagna, Politecnico di Milano

danilo.ardagna@polimi.it

AI-SPRINT project has received funding from the European Union Horizoon 2020 research and innovation programme under Grant Agreement **No. 101016577.**

Innovation for AI applications in Edge and Cloud Environments

- By 2024, AI worldwide market will approach $554.3 billion (CAGR 17.5%[1]) while edge computing will reach $250.6 billion (CAGR 12.5%[2])
- AI needs resources at the edge of the network
- New challenges from the infrastructural perspective



[1]IDC Forecasts Improved Growth for Global AI Market in 2021

[2]IDC: Edge Spending Guide September 2020

AI is hungry

Engineering hungry

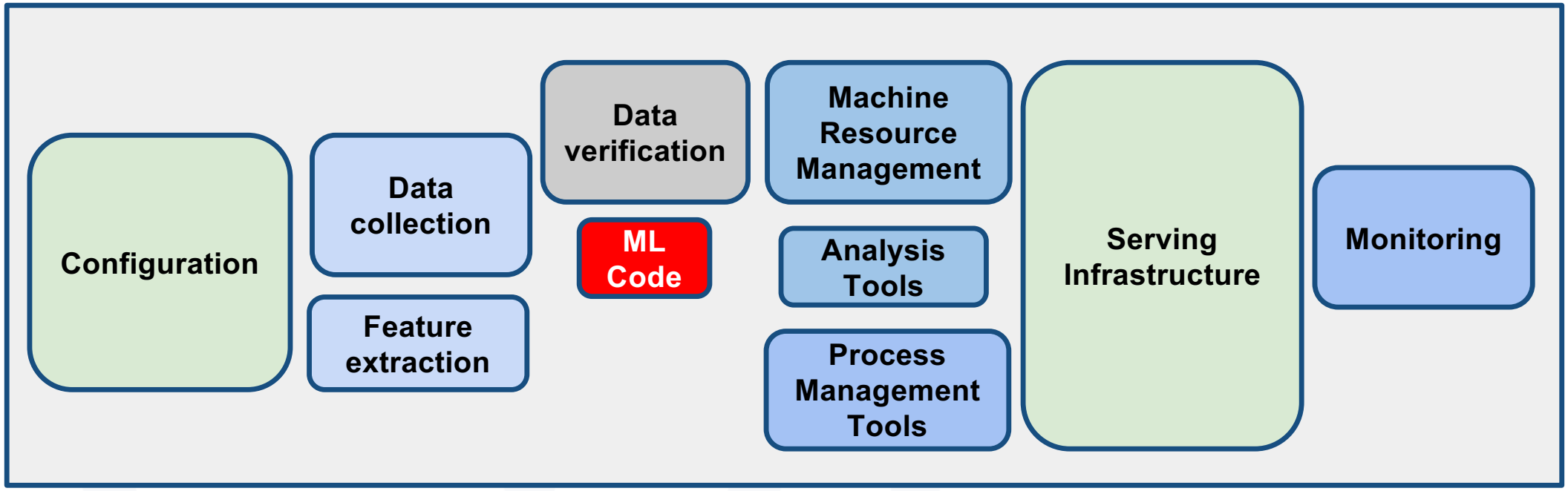
Cloud hungry

Silicon hungry

- Expertise
- Plethora of frameworks
- Trial & Error
- Data volumes
- Privacy and Security

- Heterogeneity
- Elasticity
- Efficient use of resources: costs, energy, ...

Hardest part of AI isn’t AI

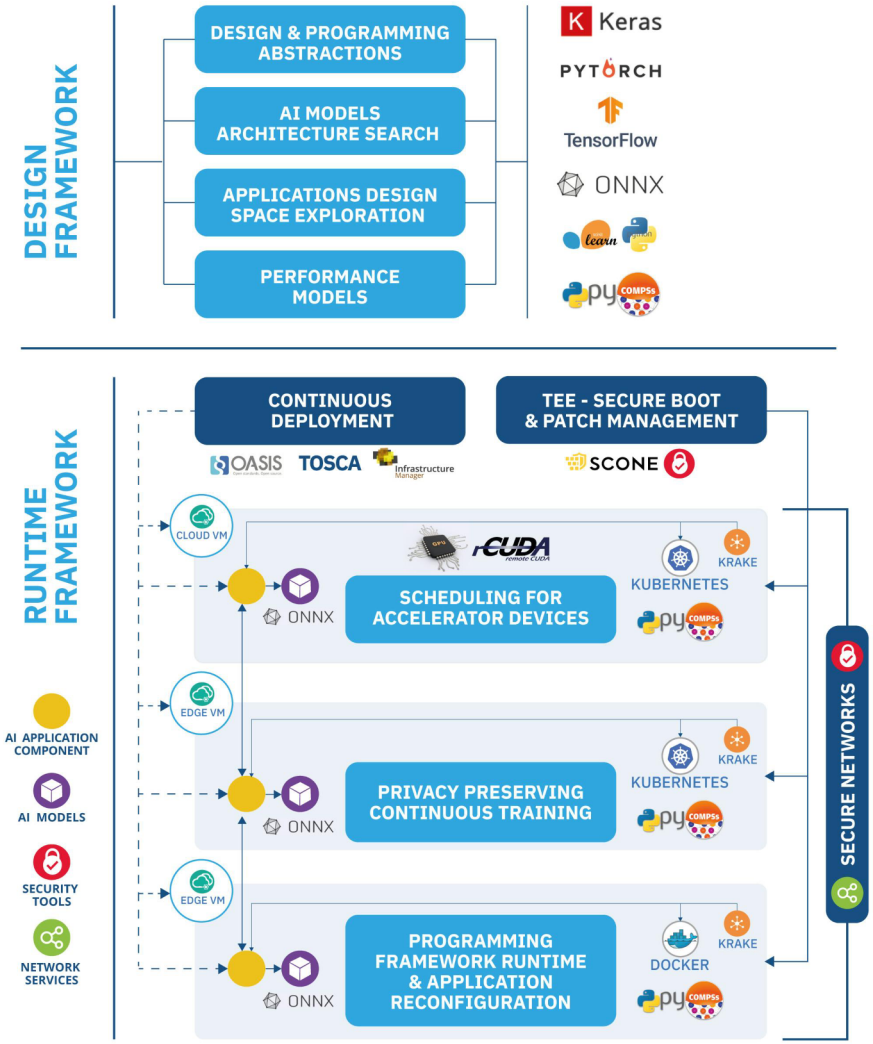


Only a small fraction of real world ML systems is composed of the ML code

Hidden Technical Debt in Machine Learning Systems, Google. NIPS 2015

AI-SPRINT promises

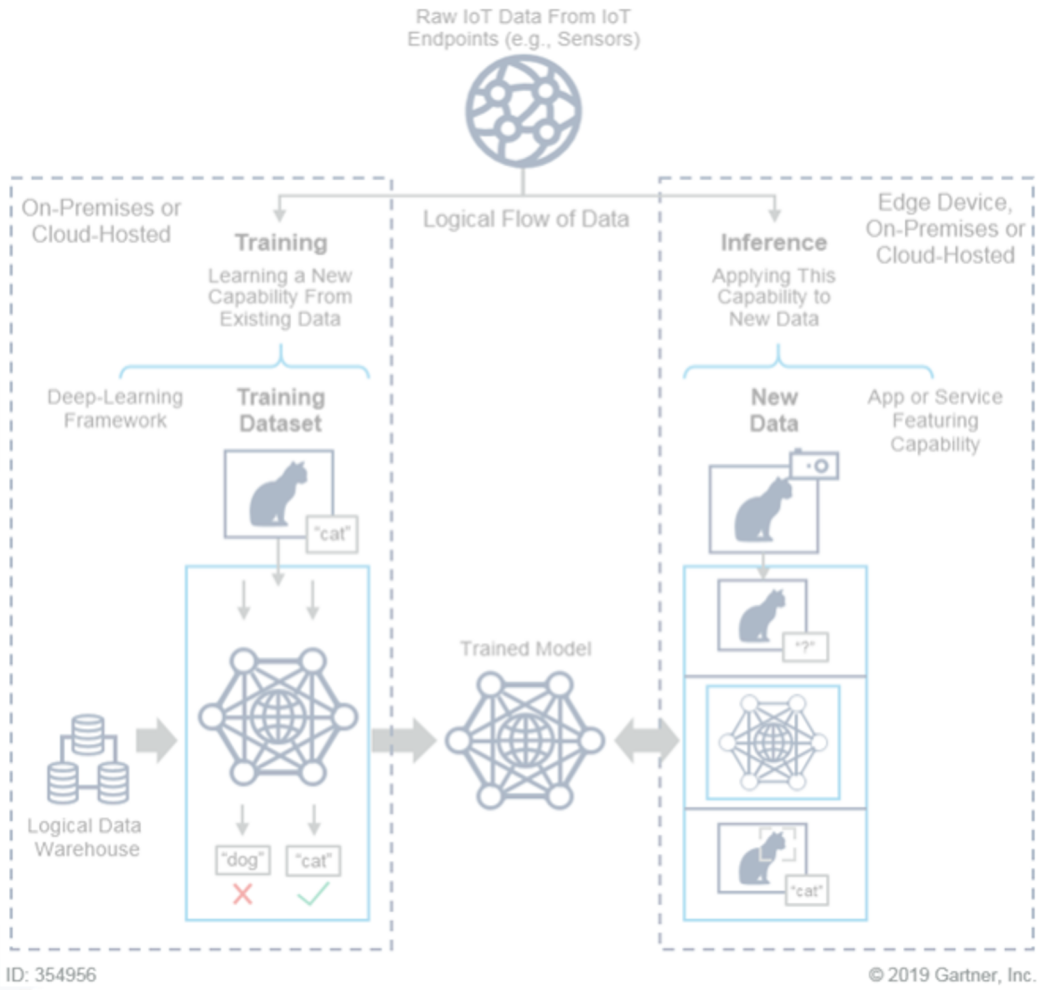
- Simplified programming models
- Automated deployment and dynamic reconfiguration
- Secure execution of AI applications
- Highly specialized building blocks for distributed training, privacy preservation and architecture enhancement



H. Sedghani, et. al. Advancing Design and Runtime Management of AI Applications with AI-SPRINT. AIM 2021 Workshop Proceedings

AI Challenges in the Computing Continuum

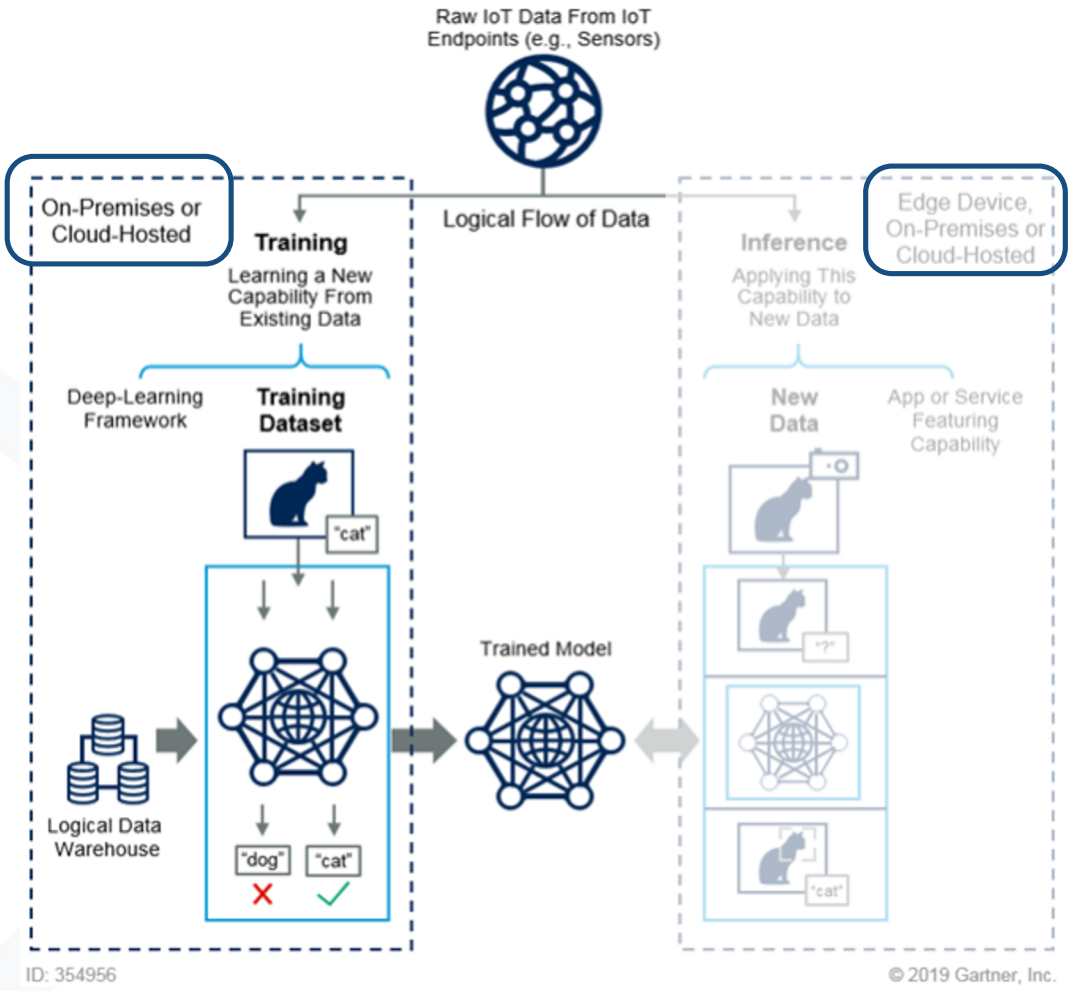
- Novel computing continua break tradition AI development paradigms



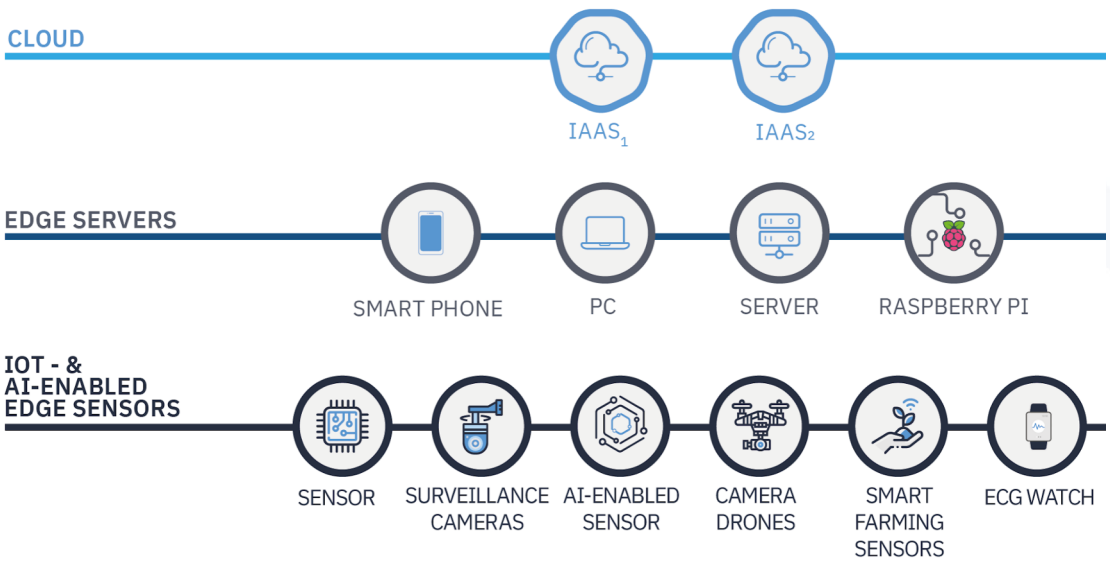
AI Challenges in the Computing Continuum

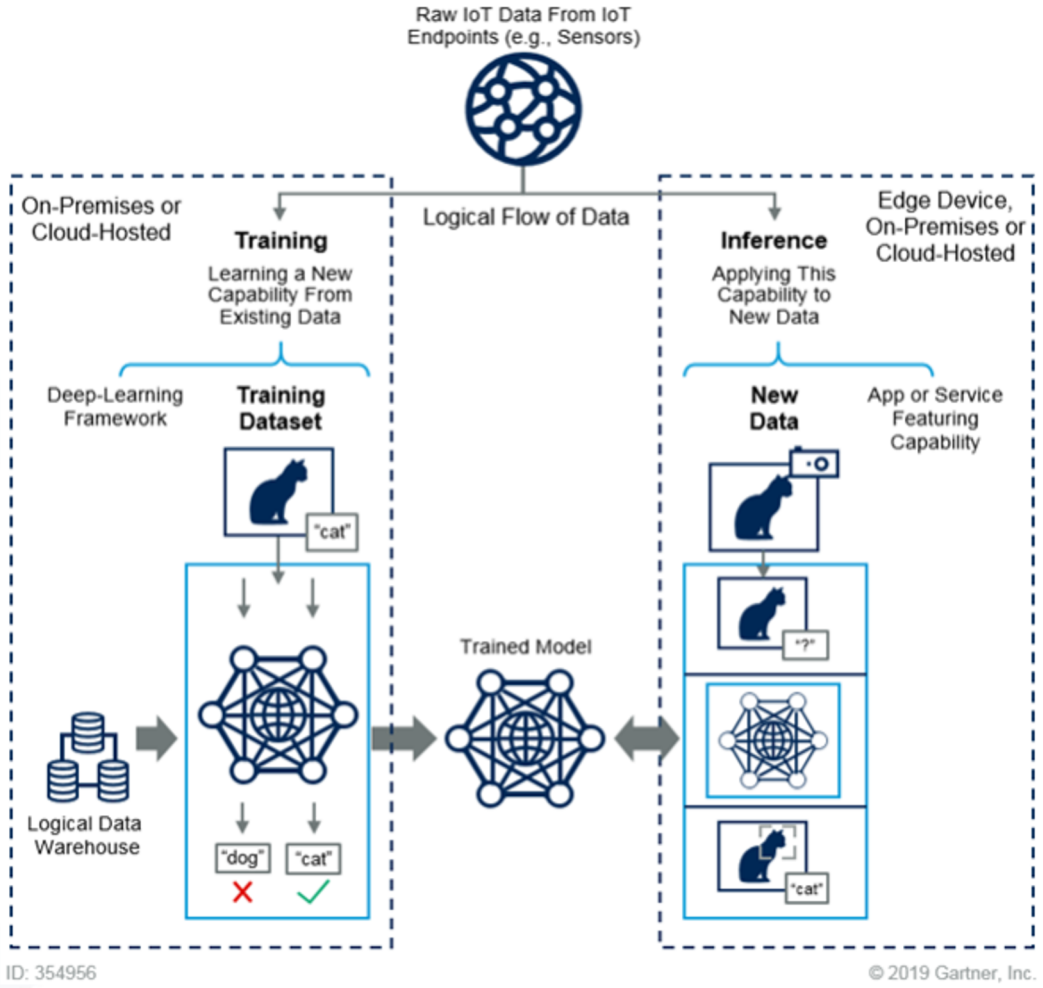
- Novel computing continua break tradition AI development paradigms

- AI development beyond the classic
  - Data from IoT
  - Train on the Cloud
  - Inference on the Edge / Cloud



AI Challenges in the Computing Continuum

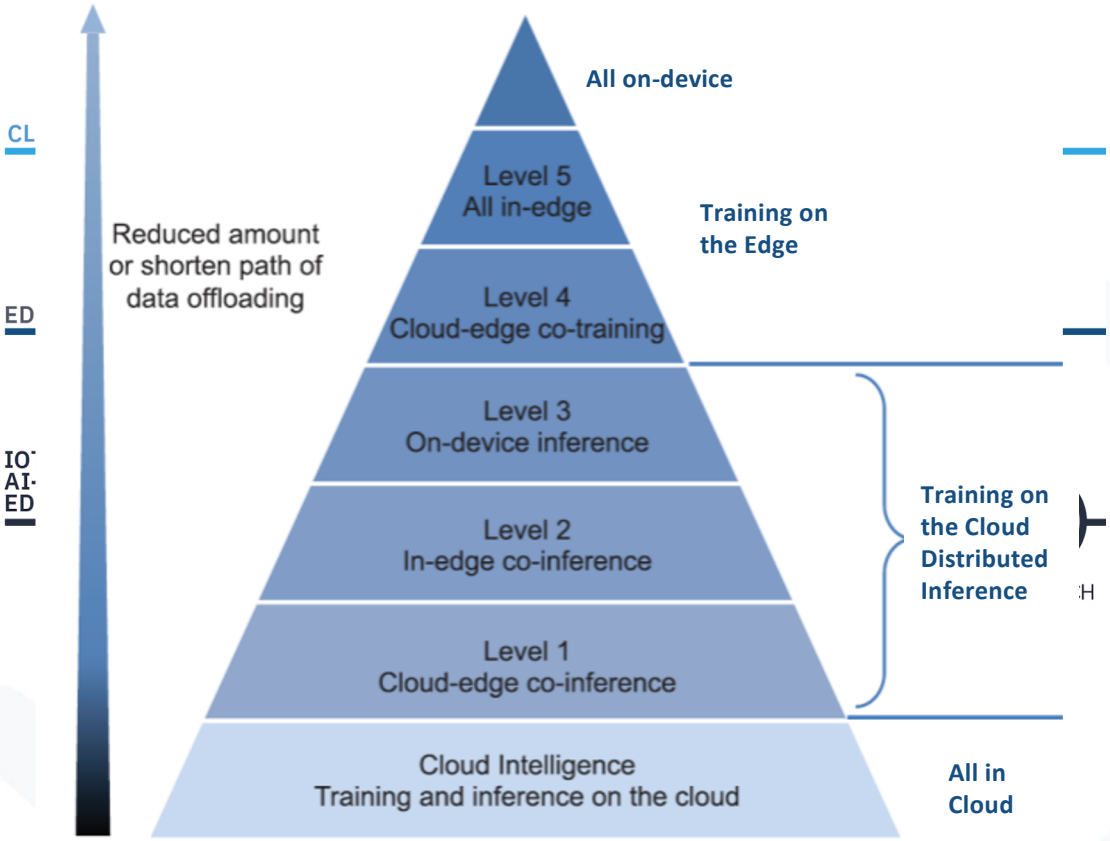


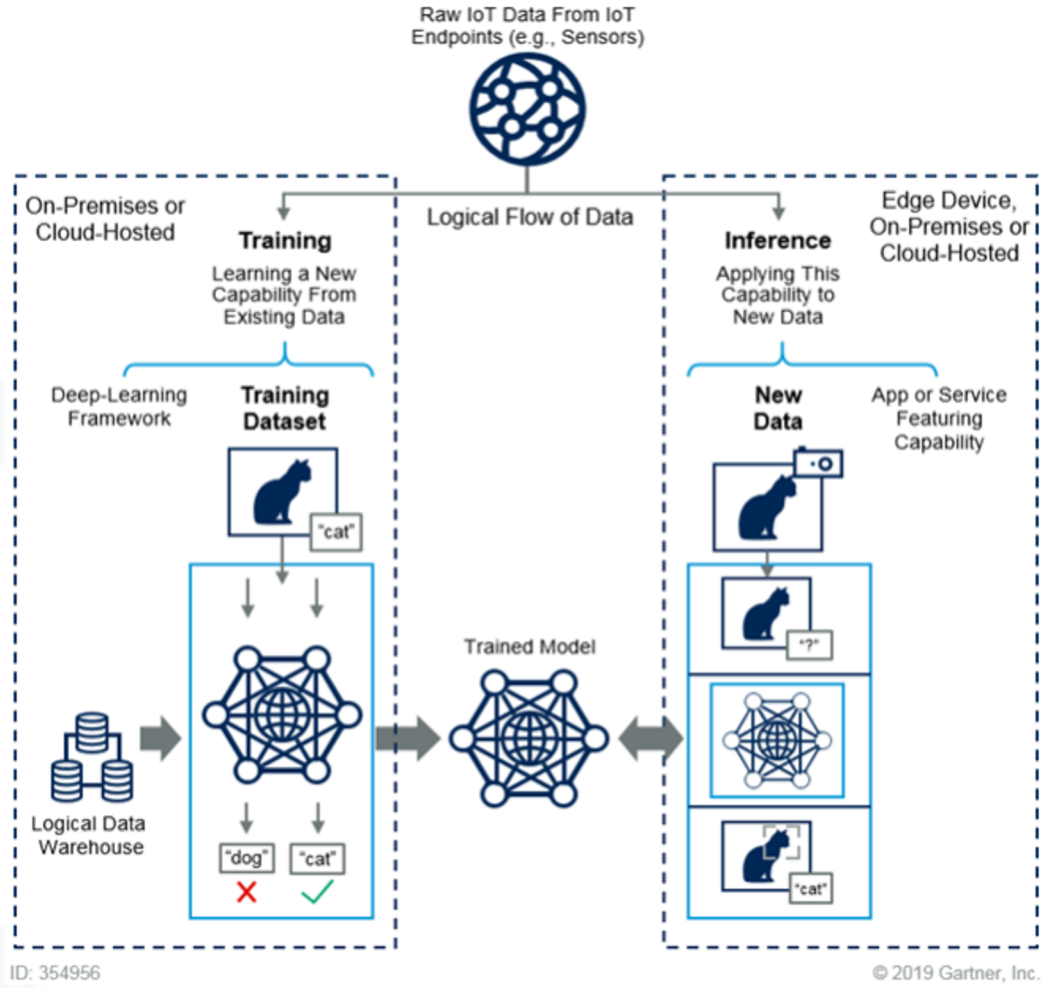


- Engineering cannot be an afterthought

AI Challenges in the Computing Continuum

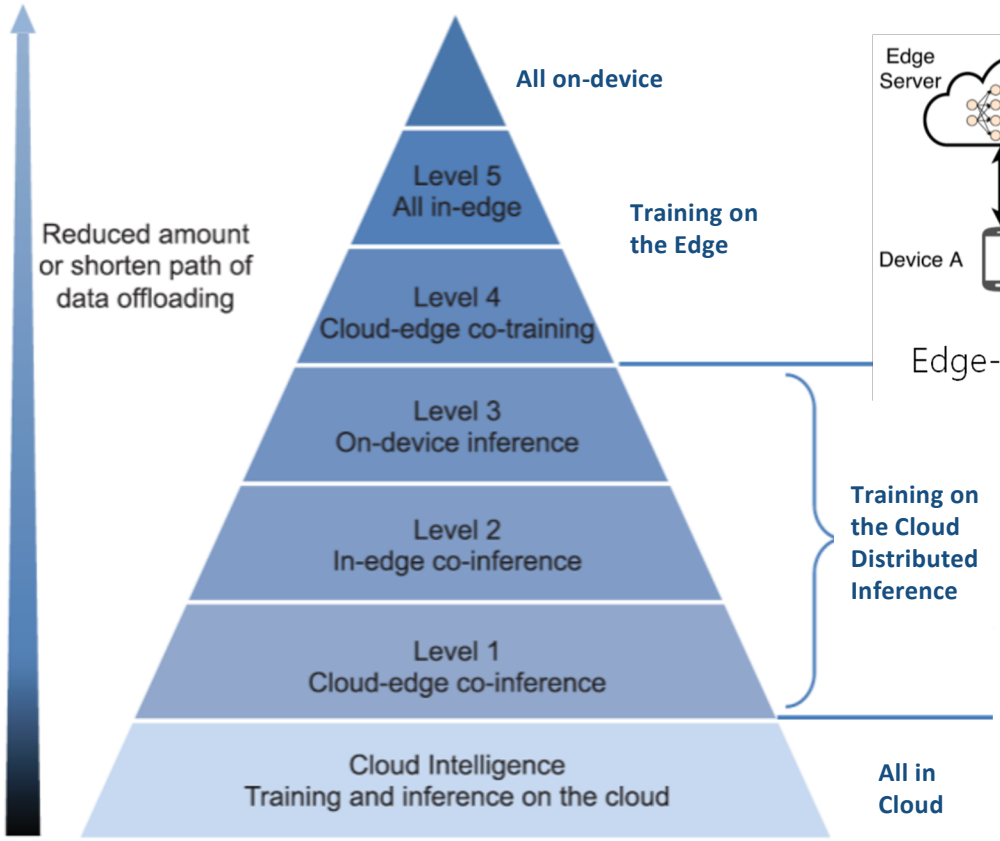
Zhi Zhou, Xu Chen, En Li, Liekang Zeng, Junshan Zhang.
"Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing"

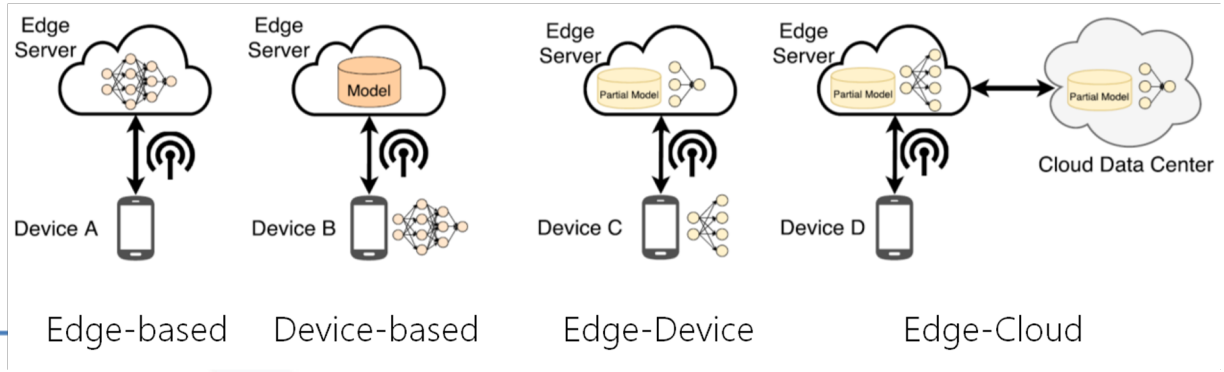


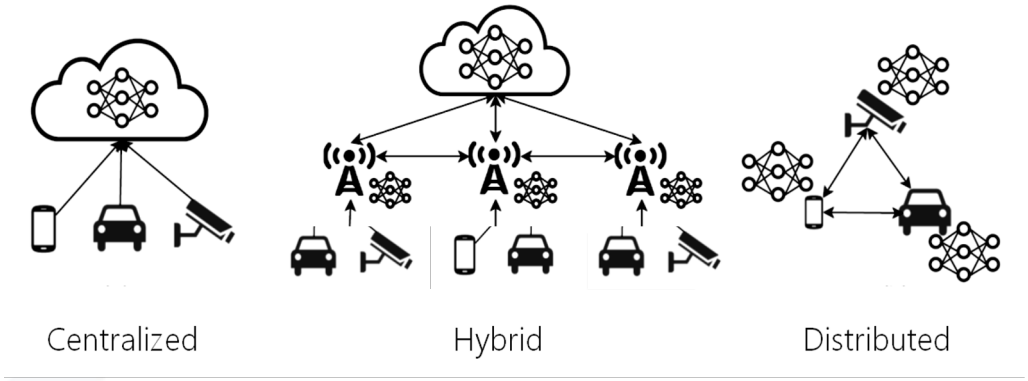


AI Challenges in the Computing Continuum

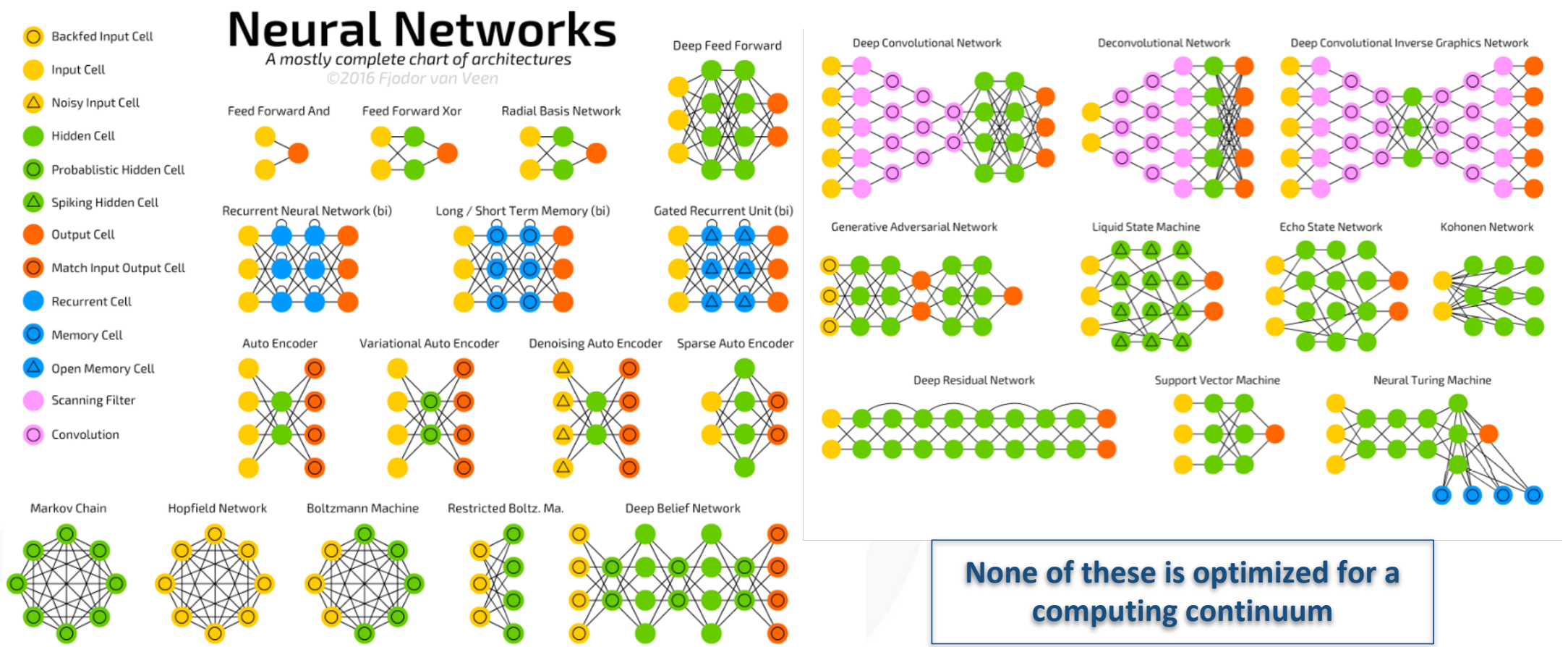
Zhi Zhou, Xu Chen, En Li, Liekang Zeng, Junshan Zhang.
"Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing"







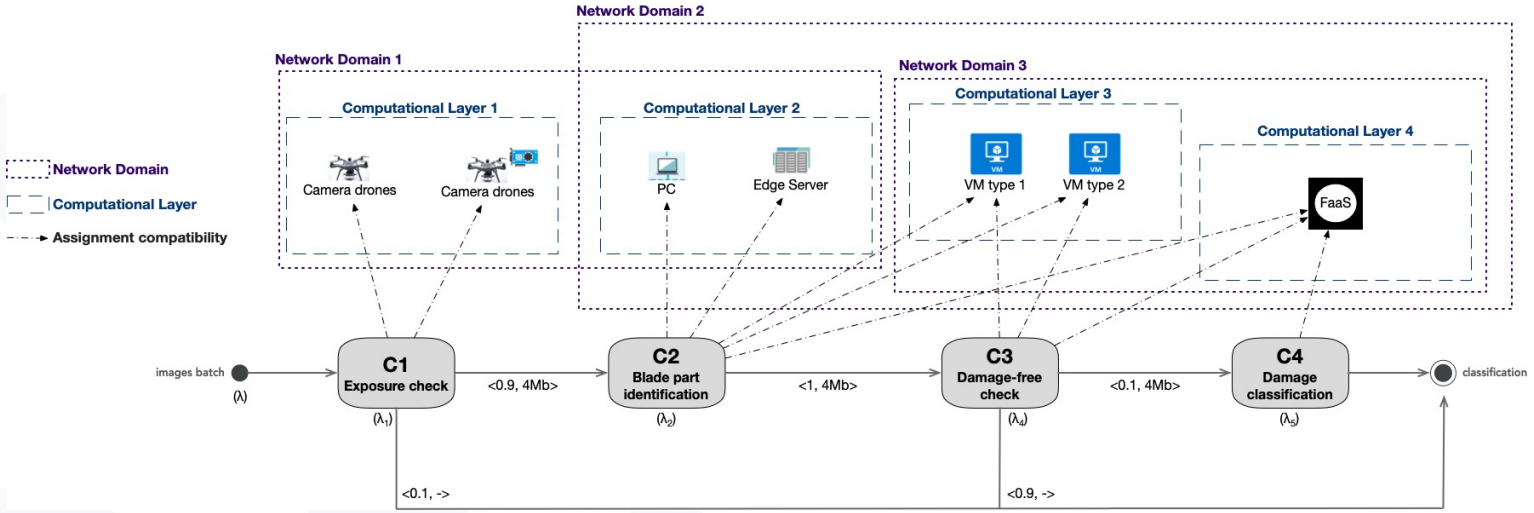
AI Challenges in the Computing Continuum



None of these is optimized for a computing continuum

Automated deployment and …

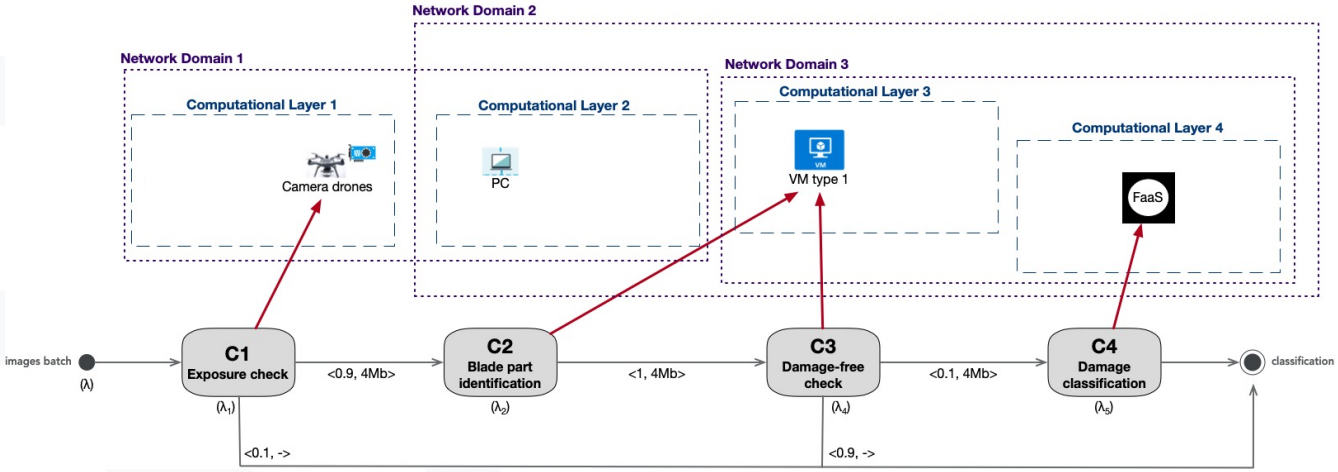
- Constraints:
  - Privacy
  - Performance
  - Energy
  - …



H. Sedghani, F. Filippini, D. Ardagna. A Random Greedy based Design Time Tool for AI Applications Component Placement and Resource Selection in Computing Continua. IEEE Edge 2021. To Appear.

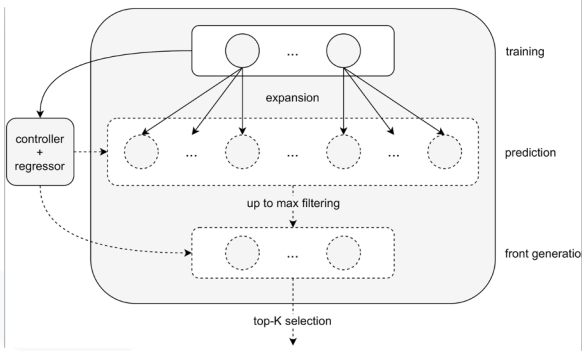
… and dynamic reconfiguration

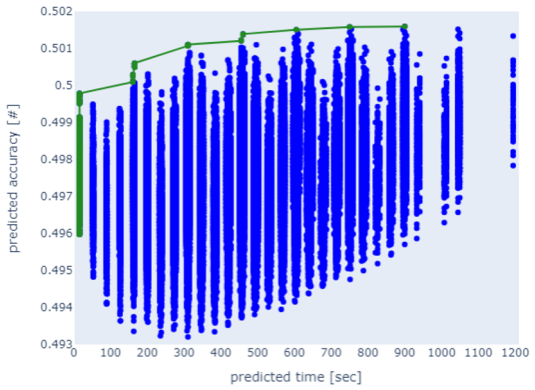
- Components migration
- Changing the deployment
- Scaling out / in cloud VMs



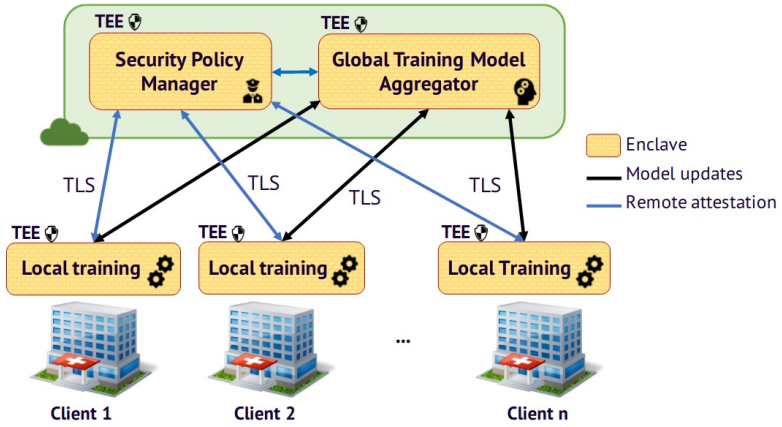
AI-SPRINT: Opportunities Beneath the Challenges

- Novel AI automatic design patterns to
  - Design new architectures optimized for the Computing Continuum
  - Reduce resource demand for hyperparameter tuning
  - Go beyond AI designer intuition





- Exploit Computing Continuum to
  - Distribute training algorithms
  - Leverage on edge computing resources which are close to data
  - Increase data privacy at training
  - Execution in secure enclaves



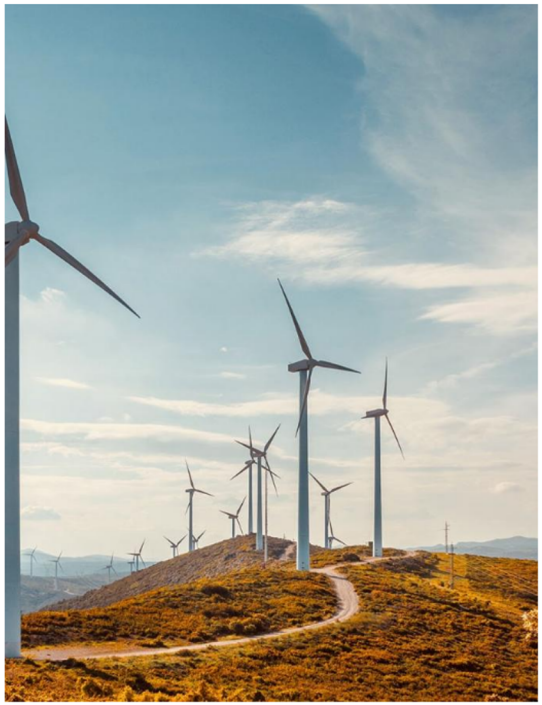
E. Lomurno, S. Samele, M. Matteucci, D. Ardagna. Pareto-Optimal Progressive Neural Architecture Search. ACM Workshop on NeuroEvolution@Work 2021

Do Le Quoc, Christof Fetzer: SecFL: Confidential Federated Learning using TEEs. arXiv

AI-SPRINT Use Cases Overview

Personalised Healthcare

Developing an automated system for personalised stroke risk assessment and prevention.



Maintenance & Inspection

Creating an infrastructure that reduces downtime and revenue losses caused by degenerative asset performance.

Farming 4.0

Delivering edge and intelligent sensors to optimise phytosanitary treatments.

Thanks for your attention….

COORDINATOR

POLITECNICO MILANO 1863

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

GREGOIRE

Beck et al. work. together.

CLOUD & HEAT

7bulls.com

AirFusion

Trust-IT Services
Communicating ICT to markets

IDC
ANALYZE THE FUTURE

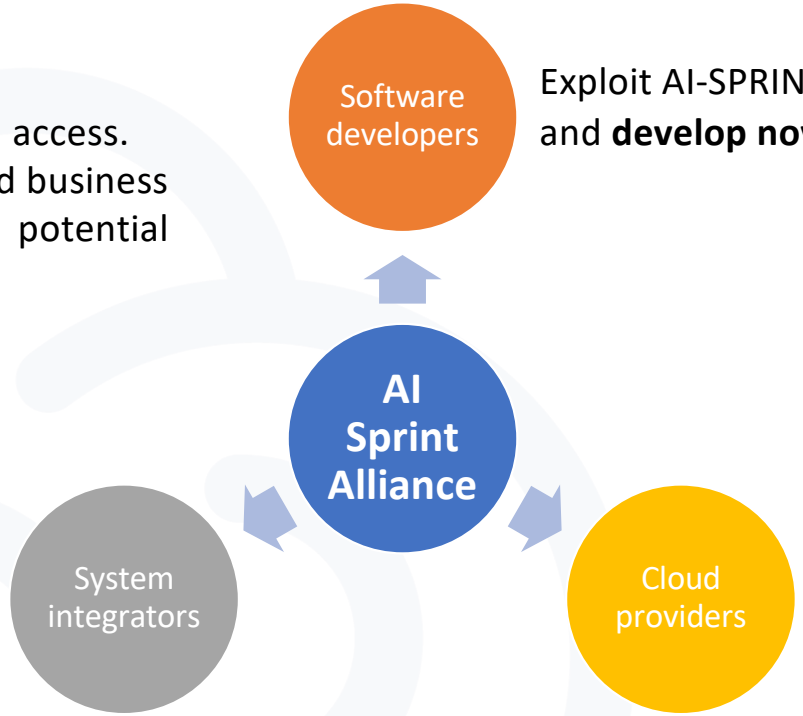
…. any questions?

https://www.ai-sprint-project.eu/

AI-SPRINT Alliance and Adopter Acceleration Club

June 2021

- Set up and launch of the Alliance
- Collaboration set-up with customised access.
- Regular briefings on technological and business innovations, including market outlook, potential competitors and adopters.



Exploit AI-SPRINT technology to **design** and **develop novel AI applications**

Offer **new services** through the **integration** of **AI-SPRINT technology stack**

Include AI-SPRINT technology **into their service catalogue** to offer easy to use **design frameworks**

Join us with a letter of support by writing at:
Niccolò Zazzeri n.zazzeri@trust-itservices.com
Danilo Ardagna danilo.ardagna@polimi.it