Edge AI from the scientific perspective

Examples and future trajectory

Anders Lindgren
Senior Researcher, RISE Research Institutes of Sweden
Adjunct Senior Lecturer, Luleå University of Technology
Coordinator, DAIS Project
anders.lindgren@ri.se

2024-12-05

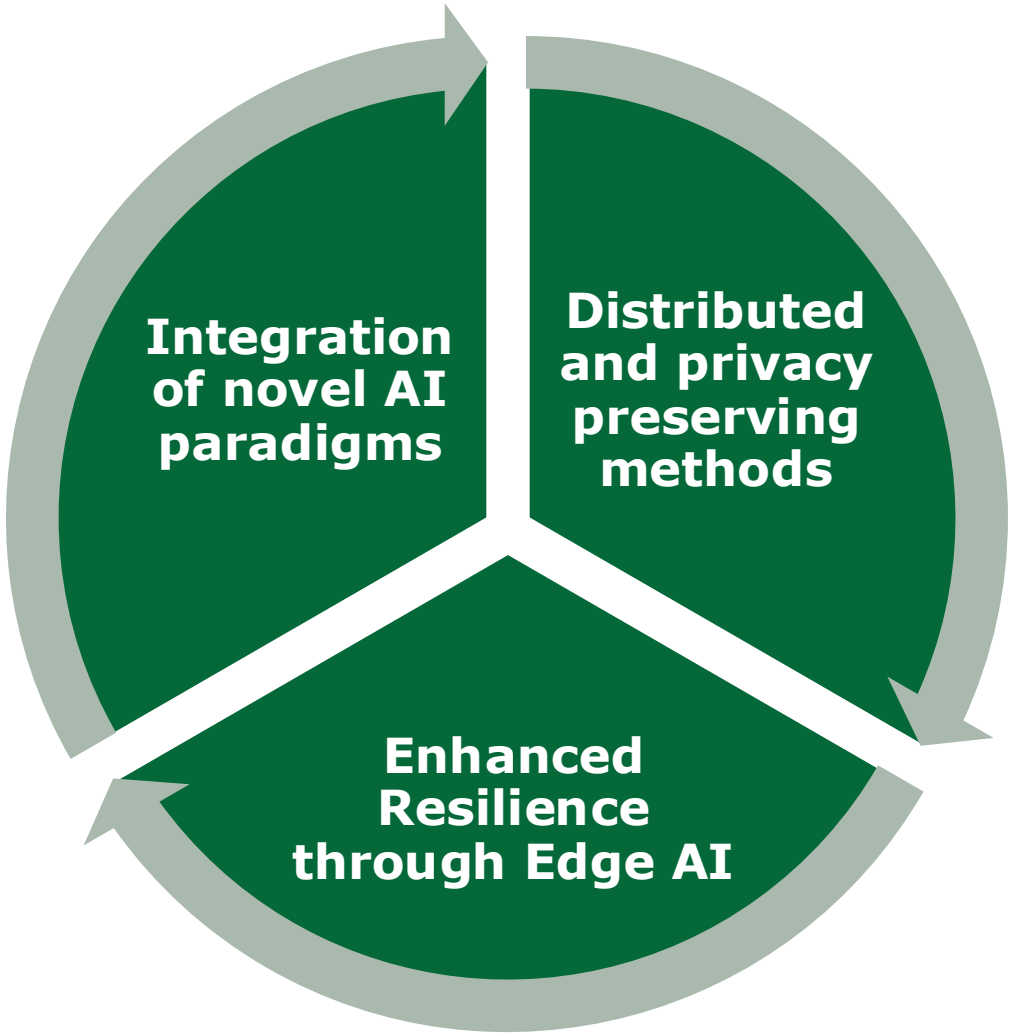
DAIS – Distributed Artificial Intelligent systems

DAIS has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No x. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Sweden, Netherlands, Germany, Spain, Denmark, Norway, Portugal, Belgium, Slovenia, Czech Republic, Turkey.

Where we stand today – current outlook

- Key advancements
  - low-power hardware, federated learning, 5G integration.
- Applications
  - autonomous vehicles, smart sensors, predictive maintenance, industry automation, resilient systems
- Limitations
  - constrained resources, model optimization, and lack of standardization
  - Regulation and privacy concerns
  - Dependence on other parts of the world

Some Key Scientific Challenges and Application Opportunities on the Horizon



Integration of neuromorphic computing into digital ecosystem

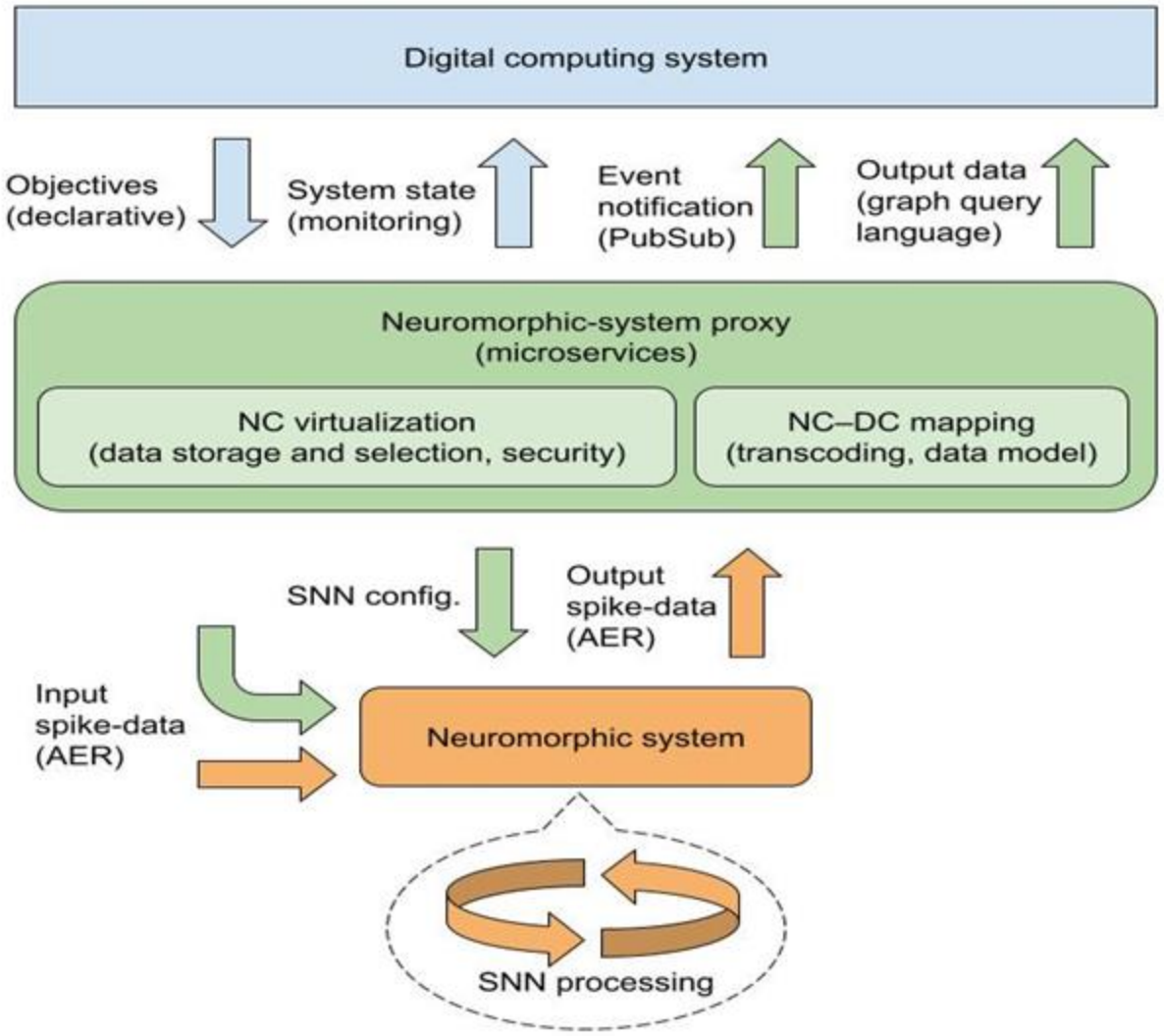
- Current AI systems extremely power hungry and unsustainable
- One approach is human-brain inspired neuromorphic computing

$\times 10^6$

~20 Watt
(10^{11} neurons, 10^{15} synapses)

~20 Megawatt
(1.2 exaflops, 8.7M cores)

Neuromorphic computing is great in many ways – but we live in a world of mostly digital Turing based computing systems and most tools are adapted for such systems. Need to be able to integrate NC systems into existing computing ecosystem.

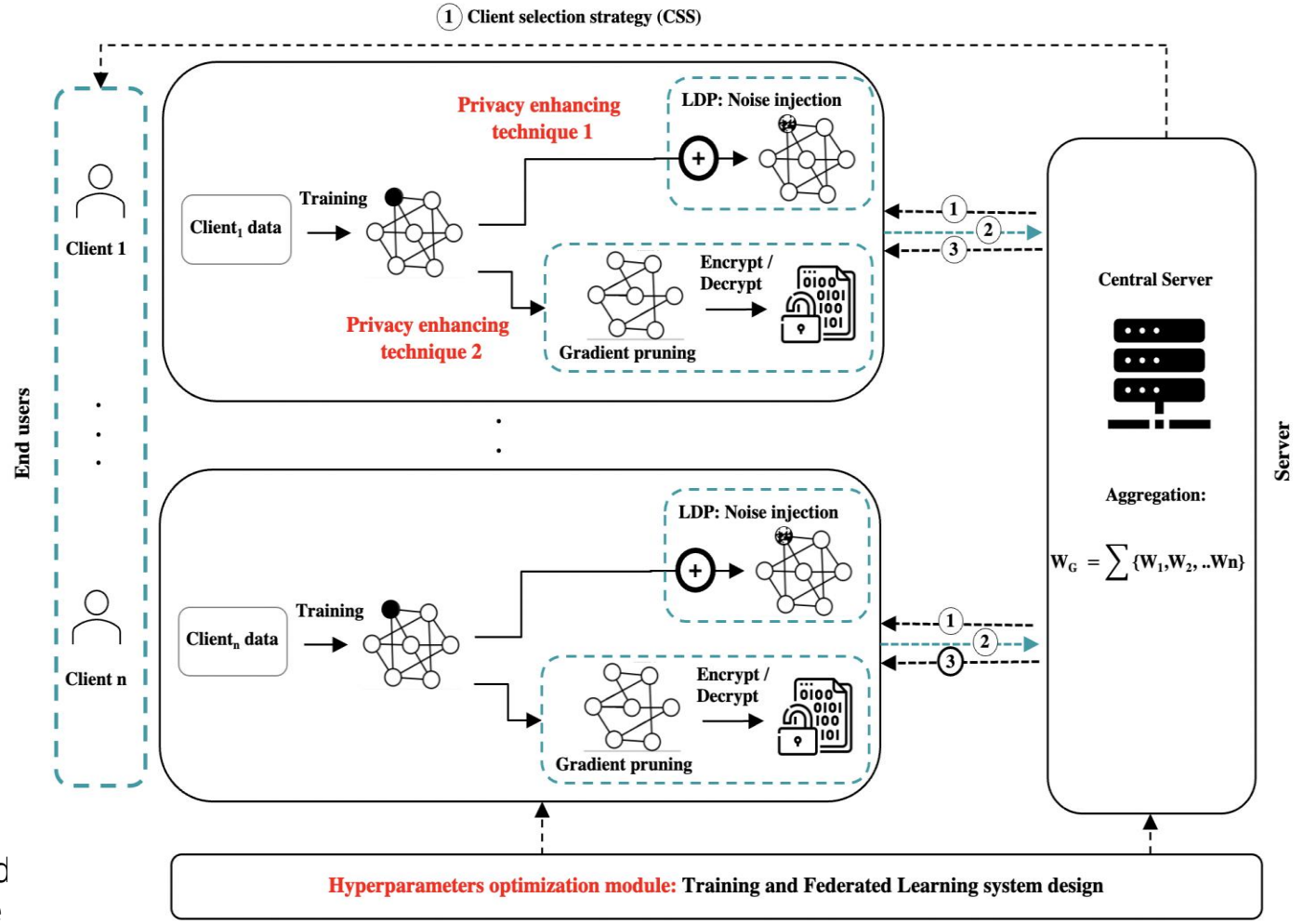


Distributed and federated learning for privacy enhancing

- **Edge AI is key for providing enhanced personal integrity and privacy preserving AI**
- **Vital for maintaining European sovereignty and our position in relation to other players**
  - **AI Act, GDPR, etc…**

Example framework:

- ✓ Combine **local differential privacy** with a novel **client selection strategy** to bolster **privacy** and **accuracy** in Federated Learning systems.
- ✓ This framework ensures **Secure and Efficient Federated Learning** by encrypting model parameters using **Paillier homomorphic encryption** before sharing them with the server. We optimize edge device performance through **gradient pruning** before encryption, reducing communication and computation overhead.
- ✓ Create module for hyperparameter optimization by automating the selection of optimal parameters in federated learning while considering multiple objectives and resource constraints.



Publication:

1. Balancing Privacy and Accuracy in Federated Learning for Speech Emotion Recognition, accepted at the 18th Conference on Computer Science and Intelligence Systems.
2. Secure and Efficient Federated Learning by Combining Homomorphic Encryption and Gradient Pruning in Speech Emotion Recognition, accepted at the 18th International Conference on Information Security Practice and Experience.
3. Hyperparameters Optimization for Federated Learning System: Speech Emotion Recognition Case Study, accepted at the 8th IEEE International Conference on Fog and Mobile Edge Computing.

Enhanced Resilience through Edge AI

How can society, industry and service operators keep their key services running in a period of significantly increased variability in the availability of basic resources and critical infrastructure (energy, communication networks, security) that current systems are designed to take for granted?

Aim: Make critical infrastructure more resilient to events across the spectrum, such as those related to:

- Power fluctuations and sudden changes
- Human-induced operational incidents
- Atmospheric, solar and weather events
- Cyber attacks and other activity

By developing edge-based computing solutions to significantly increase the robustness and resilience of critical infrastructure, many of these issues can be addressed. Examples of infrastructure that can be improved are:

- Digital and computational infrastructure, such as data centres and networks, Power production, transmission and distribution, Transport/mobility digital infrastructure, Industry and manufacturing

Supporting technologies to enable the advanced solutions in industrially relevant scenarios need to be developed in three pillars:

- Chips or semiconductor technology
- AI
- Communication and security

Where do we need to go next? Future trajectories

- Scalability of Edge AI models and efficient Edge hardware
- Trust and explainability in decision-making.
- Cybersecurity and resilience against attacks.
- Energy efficiency in diverse environments.
- Building the future of Edge AI by:
  - Investing in research and infrastructure.
  - Fostering interdisciplinary partnerships.
  - Prioritising ethical, sustainable development.
  - Empowering a skilled workforce – education is crucial

What are the opportunities for Europe?

- Strong scientific advances are done in other parts of the world, so what are the opportunities for Europe?
- Strategic alignment with Green and Digital Transitions.
  - Make sure Europe is leading the scientific forefront for the green transition supported by current policies.
- Collaboration between academia, industry, and governments.
  - The European research landscape encourages a large set of collaboration across borders, sectors, and domains. Important to maintain this advantage through funding schemes rewarding both scientific excellence as well as innovation.
  - Chips JU excellent example and opportunity of this, covering the full value chain from hardware design to applications and services.

Thank you

DAIS – Distributed Artificial Intelligent systems

DAIS has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No x. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Sweden, Netherlands, Germany, Spain, Denmark, Norway, Portugal, Belgium, Slovenia, Czech Republic, Turkey.