EdgeAI: A Vision for Distributed Artificial Intelligence on the Edge

Lauri Lovén, Ella Peltonen, Teemu Leppänen, Juha Partala, Erkki Harjula, Pawani Porambage,
Mika Ylianttila, Jukka Riekki
first.last at oulu.fi, University of Oulu, Finland

Edge and fog computing, prominent parts of the upcoming **5G mobile networks** and future **6G technologies**, promise to reduce applications' latencies, improve controls on privacy, and reduce network bandwidth usage. The promises are delivered by pulling computations from the remote cloud to close to the devices, where data is generated and applications are used. In stark contrast, current **artificial intelligence** (AI) and in particular **machine learning** (ML) methods mostly assume all computations are conducted in a homogeneous cloud with ample computational and data transmission resources available.

The integration of edge computing with AI/ML methods – an endeavour we call EdgeAI – promises to improve both fields in a variety of aspects. **Our research aims to identify the challenges and detail the potential benefits of EdgeAI**, building a coherent and overarching vision of what distributed artificial intelligence means in the context of edge and fog computing. **Further, we aim to find the methods of realizing those benefits**, testing our hypotheses in a real-world setting on the edge-based computational platform we're building upon the 5G test network (http://5gtn.fi). The vision will be realized during the 8-year time span of the 6Genesis Flagsip research program.

Bringing edge computing and AI/ML together is **challenging** due to the fundamental difference in the premises of AI and edge computing. While edge computing is by design distributed and fog leans towards decentralization, modern AI/ML methods are only beginning to allow for distributed, let alone decentralized, computations. Further, intermittent connectivity may corrupt or slow down current ML model training algorithms [1].

Yet, clear **benefits** can be identified from the interplay of AI and edge computing. Following Park et al. [1], we divide the interplay into *edge computing for AI* (Edge4AI) and *AI for edge computing* (AI4Edge). Fig. 1 illustrates the benefits provided by EdgeAI for both AI functionality and edge computing operation, further dividing edge computing to applications, privacy, security, management, and communication aspects.

Indeed, AI methods may improve communication networks in many ways. Scalability and platform KPIs such as latency and reliability can be improved while user overall Quality of Experience (QoE) increases. Further, the management of an edge-based computational platform stands to benefit from AI, with AI providing new capabilities such as predictive or decentralized control and system orchestration, and predictive maintenance of network infrastructure.

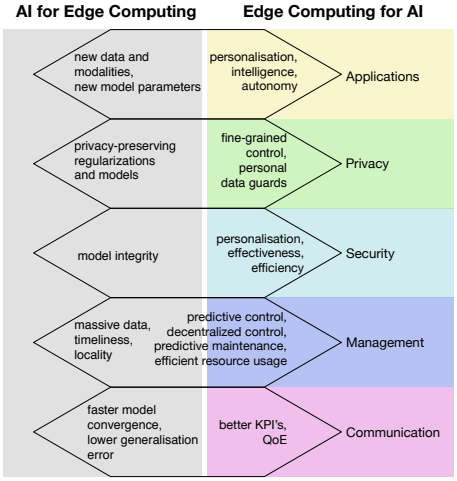

Fig. 1. Benefits provided by EdgeAI for both AI and edge computing.

Conversely, as detailed by Park et al. [1], AI models converge faster with a lower generalization error when the architectures of the underlying communication network are built to support AI workflows. Distributed edge platforms with IoT connectivity further provide AI models with massive amounts of local data in a timely fashion.

AI improves Edge security with personalized security systems adapted to each individual location or user particulars. Further, threat detection and prevention stand to gain from AI-enhanced pattern and anomaly recognition. AI models, on the other hand, are ensured of data integrity. Whereas AI methods may improve user privacy by decentralizing trust, personal privacy guards and fine-grained control on consent management and data ownership, privacy concerns enhance AI methods in turn by providing novel privacy-preserving regularisation and data-generating unsupervised or semi-supervised ML models.

Finally, whereas edge-native AI methods provide applications with unprecedented access to personalized and localized prediction and control models, enhancing application intelligence and autonomy, the applications themselves provide AI models with new data and modalities as well as well as new model parameters.

Acknowledgements: This research is financially supported by Academy of Finland 6Genesis Flagship (grant 318927).

References

[1] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless Network Intelligence at the Edge," *arXiv preprint*, 2018.