universität innsbruck
Fakultät für Mathematik, Informatik und Physik

BERUFUNGSVORTRÄGE: „Informatik mit dem Schwerpunkt Edge AI / FFG (Stiftungsprofessur)“

Montag, 9. September 2024 Institut für Informatik (ICT Gebäude) - Raum 3W03	
N. N.*	
11:00 – 11:30	**Lehrevortrag und Diskussion: „Introduction to Federated Learning“**
11:30 – 12:10	**Forschungsvortrag und Diskussion: „Edge AI and Beyond: Designing Robust AI Architectures for the 6G Era“** *In this talk, I will explore cutting-edge advancements in architecting AI-intensive systems, with a particular focus on the emerging field of Edge AI. As AI systems become increasingly integral to both research and industry, the ability to design, implement, and manage distributed AI architectures is critical. This presentation will focus into key research areas, including the migration of legacy systems to microservices, the challenges of architectural degradation, and the development of robust cloud-to-edge computing paradigms within 6G frameworks.* *We will examine the role of distributed intelligent computing, emphasizing how hyper-locality, orchestration, and trust are pivotal in the evolution of the edge-to-cloud continuum. Special attention will be given to the design and implementation of distributed AI models, addressing the challenges of deploying and managing AI workloads across diverse, dynamic environments. Additionally, the presentation will cover multimodal sensing and modeling, which are essential for enabling real-time, intelligent applications in complex, data-rich environments.*
Dienstag, 10. September 2024 Institut für Informatik (ICT Gebäude) - Raum 3W03	
N. N.	
08:30 – 09:00	**Lehrevortrag und Diskussion: „Introduction to Federated Learning“**
09:00 – 09:40	**Forschungsvortrag und Diskussion: „Think Broader, Run Local*: Towards Neuro-Symbolic AI at the Edge”*** *The presentation starts by reviewing the convergence of two historically disconnected AI branches: symbolic AI of explicit deductive reasoning owning mathematical rigor and interpretability, and connectionist AI of implicit inductive reasoning lacking readability and explainability. Afterward, it makes a case of neuro-symbolic data processing in knowledge graph representation, applied for anomaly prediction in data centers using graph neural networks. Machine learning-driven graph sampling algorithms support its training and inference on resource-constrained Edge devices. The presentation concludes with an outlook into future projects targeting Edge large language models fine-tuned and contextualized using symbolic knowledge representation for regulatory AI compliance, job market, and medicine.*

*) Die Veröffentlichung von Namen und Titel erfolgt nur auf Zustimmung der Vortragenden. Stellungnahmen zu den Vorträgen der Hearing-TeilnehmerInnen richten Sie bitte bis längstens 13:00 Uhr des letzten Vortragstages an Fakultaet-MIP@uibk.ac.at.

Dienstag, 10. September 2024 Institut für Informatik (ICT Gebäude) - Raum 3W03	
N. N.*	
11:30 – 12:00	**Lehrevortrag und Diskussion:** **„Introduction to Federated Learning“**
12:00 – 12:40	**Forschungsvortrag und Diskussion:** **„Quality of Analytics and Runtime Explainability for Edge AI“** *Modern, complex edge-cloud software and new software developments in Edge-Cloud continuum increasingly utilize artificial intelligence/machine learning (AI/ML) features as a part of the software. Such AI/ML features encapsulate advanced AI/ML capabilities and require different integration, provisioning, optimization and management methods and techniques. Based on capable ML models, these features are powerful but exhibit several contractual problems for service providers and consumers in terms of operational deployment performance, quality of inferences, uncertainty, data quality effects and explainability, to name just a few. Achieving key objectives, like faster serving time, cheaper operation cost, and higher reliability of inferences, is of paramount importance in developing and operating such AI/ML software in Edge-Cloud continuum. Such key objectives can be formulated in the aspect of Quality of Analytics (QoA), in which the QoA is considered from the view of (i) the engineering process, (ii) the artifact as the output of the process, and (iii) the service encapsulating the process/artifact, acting as the key driver for multi-dimensional elasticity.* *In this talk, we present our novel concept of (QoA and its impact on AI/ML services and edge-cloud software for end-to-end ML serving. We will present the QoA4ML framework for qualifying ML services and runtime explainability for end-to-end AI/ML serving. We identify and specify contractual concerns covering different aspects of data, ML models and services. Based on that we develop policies for ML service contracts, providing runtime explainability and adaptation for AI/ML serving. We will discuss our work with realistic applications of object detection, malware detection, and predictive maintenance in the Edge-Cloud continuum. Finally, we will outline some future perspectives of QoA for the emerging multi-dimensional time, intelligence and computing continuum, the privacy-preserving, continuous edge observability and AI/ML, and the incorporation of edge/cloud Large-Language Model (LLM) utilities.*

*) Die Veröffentlichung von Namen und Titel erfolgt nur auf Zustimmung der Vortragenden. Stellungnahmen zu den Vorträgen der Hearing-TeilnehmerInnen richten Sie bitte bis längstens 13:00 Uhr des letzten Vortragstages an Fakultaet-MIP@uibk.ac.at.

Donnerstag, 12. September 2024 Institut für Informatik (ICT Gebäude) - Raum 3W03	
DRAGI KIMOVSKI	
08:30 – 09:00	**Lehrevortrag und Diskussion:** **„Introduction to Federated Learning“**
09:00 – 09:40	**Forschungsvortrag und Diskussion:** **„The Non-Von Neumann Edge: Revolutionizing AI-Driven Computing”** *The advent of artificial intelligence (AI) has catalyzed a significant paradigm shift in computing, enabling the development and deployment of novel types of applications, such as autonomous vehicles, smart cities and interactive chatbots. However, the conventional Von Neumann architecture, which has dominated the computing landscape for decades, is not well suited for the demands imposed by AI workloads. This architecture's limited parallel processing capabilities, complex memory hierarchy, and reliance on general-purpose Central Processing Units (CPUs) render it inefficient for AI tasks' highly distributed, data-intensive, and latency-sensitive nature.* *In this talk, we delve into the emergence of AI-driven computing, marked by the ascendance of non-Von Neumann architectures, particularly at the Edge. These architectures, encompassing Neural Processing Units (NPUs), neuromorphic processors, Graph Processing Units (GPUs), and Field-Programmable Gate Arrays (FPGAs), are engineered to exploit parallelism, distributed processing, and specialized hardware to accelerate AI computations significantly. We will discuss the principal advantages of non-Von Neumann architectures, especially in the Edge, including reduced latency, enhanced throughput, and improved energy efficiency. Furthermore, we will address the challenges and opportunities associated with integrating these architectures with traditional Von Neumann-based systems and assess the broader implications for AI-driven applications.*
N. N.*	
11:00 – 11:30	**Lehrevortrag und Diskussion:** **„Introduction to Federated Learning“**
11:30 – 12:10	**Forschungsvortrag und Diskussion:** **„Large-Scale Distributed Systems for Edge AI”** *The increasing dependence of AI systems on geographically distributed and streaming data is a reflection of the vast amounts of data generated across various distant locations. Traffic management, environmental monitoring, and energy grids are just a few examples of such applications. Next-generation distributed systems, such as edge computing platforms, are excellent prospects for processing data close to its origin, thereby reducing network latency, achieving scalability, and improving privacy. Correspondingly, Edge AI aims to make the most of the widespread edge resources to gain AI insight on a large scale.* *In this talk, I will introduce several research questions in the context of Edge AI that I have been focusing on in the last seven years, as well as my contributions to the field from the distributed systems perspective. Particularly, I will demonstrate the non-functional challenges (i.e., beyond model accuracy) for Edge AI, including failure resilience, system performance, communication optimization, and energy efficiency. Furthermore, I will briefly introduce two case studies of Edge AI with substantial societal implications from my recent international projects. One of them concerns an IoT-driven water quality monitoring and early warning system for European rivers, whereas the other focuses on the cybersecurity of monitoring systems for critical energy infrastructures and water supply networks in the EU and the UK.*

*) Die Veröffentlichung von Namen und Titel erfolgt nur auf Zustimmung der Vortragenden. Stellungnahmen zu den Vorträgen der Hearing-TeilnehmerInnen richten Sie bitte bis längstens 13:00 Uhr des letzten Vortragstages an Fakultaet-MIP@uibk.ac.at.