



Technische Universität Berlin



Technische Universität Berlin offers an open position:

Research Assistant - salary grade E13 TV-L Berliner Hochschulen

under the reserve that funds are granted; part-time employment may be possible

Faculty IV - Institute of Telecommunication Systems / Distributed and Operating Systems (DOS)

Reference number: IV-195/25 (starting at 01/07/25 / until 30/06/27 / closing date for applications 30/05/25)

Working field:

Research and Teaching at the Chair of Distributed and Operating Systems; Publication of research results.

Large Language Models are in trend, however, increasing the models size implies the design and deployment of more complex infrastructures. Distributed training is needed due to the memory constraints, increasing the infrastructure size and the likelihood of a failure of any device, which increase operating expenses and resource waste. Therefore, an effective monitoring of failures demands a thorough understanding of the infrastructure considering the interplay of metrics belonging to inter/intra-host network metrics, CPUs, NPUs, GPUs, communication patterns as well as specifics of an LLM training. The objective of this project is to develop a framework for detecting and predicting failures in Large Language Models, specifically Mixture-of-Experts architectures based on gaining in-depth analysis and understanding of failure mechanisms in communication, computation and storage components, during training and inference.

We focus on the following topics: understanding and analyzing signals generated during LLM training, simulating scenarios through failure injection, understanding cross-effects between components in large AI infrastructures, monitoring and interpreting data from physical layer (hardware), data layer (storage and transfer), computational layer and application layer (models). We aim at learning joint representations from the multiple sources of system data to detect anomalies and their root-causes. All these will entail designing a general method, implementing a prototype in the context of existing open-source systems, and experimentally evaluating the prototype with a test data using experimental and production data.

The possibility of a PhD is given.

Requirements:

- Successfully completed university degree (Master, Diplom or equivalent) in Computer Science, with specialization in data science and machine learning
- Experience with statistical software, monitoring tools, operating systems
- Experience with ML methods for detection and classification tasks
- Experience in working with large cluster systems
- Building and operation of containers (e.g. Singularity, Docker)
- Experience with TensorFlow/PyTorch/keras
- Good knowledge of German and/or English is required; willingness to acquire the respective missing language skills

Desirable:

- Interest in system development and operation of large-scale software architecture, as well as enthusiasm to establish recent research results in practice
- Previous experience in writing and publication of scientific papers
- Familiar in working with methods and methodologies from the domain of time-series analysis
- Experience and interest in the topics of AI and AI infrastructures
- Experience in working with explainable machine learning methodologies and data from heterogeneous sources
- Experience developing accessible technologies
- Interest in project management and agile development methodologies

Please send your written application with the **reference number** and the usual documents (CV, list of grades, language certificates) to Technische Universität Berlin, Prof. Odej Kao: odej.kao@tu-berlin.de.

By submitting your application via email you consent to having your data electronically processed and saved. Please note that we do not provide a guaranty for the protection of your personal data when submitted as unprotected file. Please find our data protection notice acc. DSGVO (General Data Protection Regulation) at the TU staff department homepage: https://www.abt2-t.tu-berlin.de/menue/themen_a_z/datenschutzerklaerung/ or quick access 214041.

To ensure equal opportunities between women and men, applications by women with the required qualifications are explicitly desired. Qualified individuals with disabilities will be favored. The TU Berlin values the diversity of its members and is committed to the goals of equal opportunities. Applications from people of all nationalities and with a migration background are very welcome.

Technische Universität Berlin - Die Präsidentin - Institut für Telekommunikationssysteme, FG Verteilte Systeme und Betriebssysteme, Prof. Dr. Odej Kao, Sekr. EN 22, Einsteinufer 17, 10587 Berlin

The vacancy is also available on the internet at
<https://www.personalabteilung.tu-berlin.de/menue/jobs/>

