

Job advertisement

The High-Performance Computing Centre Stuttgart (HLRS) was founded as Germany's first federal high-performance computing (**HPC**) centre. It operates one of the fastest supercomputers in the world. It offers various HPC solutions and services for universities, research institutions, and industry. Furthermore, HLRS is a worldwide leader in engineering and global system sciences. Staff scientists at HLRS investigate emerging technologies such as Artificial Intelligence (**AI**), **Cloud** Computing, and Quantum Computing (**QC**) towards realising hybrid workflows and lowering the hurdle for non-experts using HPC technologies. In this context, HLRS is significantly involved in international and national research projects across the abovementioned research areas.

HammerHAI — The AI Factory for Industry and Science

The HammerHAI project offers to establish an AI Factory at the High-Performance Computing Center Stuttgart (HLRS), supported by a strong consortium from Germany, to successfully meet the growing demand for artificial intelligence (AI) infrastructure across Europe. HammerHAI is a one-stop shop for many AI users, focusing primarily on start-ups, small and medium-sized enterprises (SMEs), and large industrial companies, as well as supporting academic institutions and the public sector. It offers tailored services and infrastructure to accelerate AI innovation and help develop a competitive AI ecosystem in Europe. The AI Factory is located in a region that is one of Europe's powerhouses in manufacturing and engineering innovation, and it integrates into an ecosystem that promotes talent building that will be the basis of an ongoing digital transition.

The AI Factory HammerHAI will provide secure, scalable, and AI-optimised supercomputing resources to meet the needs of start-ups, SMEs, industry, and research institutions. Its infrastructure will enable users to easily migrate their AI applications from laptops or cloud environments to supercomputers, providing the computing power needed to develop large-scale AI models. Hereby, the AI Factory supports the entire AI lifecycle, from data preparation to model training, deployment, monitoring, and retraining, and provides a comprehensive package of services to ensure efficient and effective AI development and operation.

Shaping the Future of Machine Learning in HPC

We seek a highly motivated ML Application Support Engineer to assist users in effectively deploying, managing, and optimizing ML workloads within the AI Factory HammerHAI at HLRS. This role provides hands-on support to users, troubleshoot issues, and ensure best practices for ML/AI development and deployment in high-performance computing environments. The successful candidate works closely with ML researchers, software developers, and system administrators to ensure a seamless ML application experience. This position requires expertise in AI and, ideally, in high-performance computing systems.

In this context, we are looking for a

ML Application Support Engineer

(m/f/d, up to TV-L 13, 100%)

HLRS_28_2025

to work on HammerHAI with a strong team of ML experts in close collaboration with end users, system administrators and external stakeholders.

This is a temporary position. Employment is limited in accordance with the legal regulations to the project's duration, which is 31.03.2028. The salary for this position is based on your personal qualifications up to the level of TV-L 13.

Your Responsibilities

- Provide dedicated one on one support to external users to help optimize large scale training and inference workflows.
- Monitor and analyse application and system performance data, logs, and metrics to identify bottlenecks and opportunities for optimization.
- Gather and assess user requirements to design and adapt ML software architectures, tools, and runtime stacks that meet their needs.
- Benchmark and tune ML frameworks on accelerated and distributed computing architectures to improve performance and efficiency.
- Develop and maintain user focused documentation and contribute to scientific publications such as conference papers, white papers, and journal articles.

Your Profile

- A Master's or PhD in Computer Science, Computational Sciences, Engineering, or a related field.
- Demonstrated expertise in Data Science, Machine Learning, and Deep Learning, with experience applying these skills to large scale training and inference workflows.
- Strong software development skills using modern ML frameworks (e.g., PyTorch, JAX), with an emphasis on performance optimization and scalability.
- Solid understanding of high performance and distributed computing environments, including CPU, GPU, and parallel architectures, as well as collective communication patterns.
- Proficiency with Linux environments, source control and issue tracking tools, and containerized software stacks.
- Experience deploying and optimizing ML applications in multi node, distributed, or cloud native environments.
- Excellent technical communication skills, both written and verbal, for collaborating with internal and external stakeholders.
- Ability to work in a collaborative, interdisciplinary environment with technical and non-technical stakeholders.
- Problem-solving mindset with the ability to address and resolve issues effectively.
- Strong written and spoken English skills required; German is a plus.

What We Offer

- A professional working environment in a friendly, highly motivated and collaborative international team.
- Flexible working hours with a flexitime model and the possibility of compensating for time off in addition to the regular 30 days of vacation.
- Flexible work hours with currently up to 60% home office (upon request).
- Attractive social benefits of the public service.

- Subsidy of €25 per month for public transport and the possibility of job bike leasing.
- Use the wide range of further education and training opportunities (e.g., soft skills, languages, specialist courses, leadership seminars) and the sports offers of the University of Stuttgart.
- Fixed-term employment with salary and working conditions up to TV-L13.

Are you interested?

If you are passionate about AI and HPC and want to contribute to a transformative European project, we encourage you to apply.

Please send us your application, including a motivation letter, resume, certificates and references, stating the reference number of the position “HLRS_28_2025” in the subject line, preferred by e-mail to:

bewerbungen(at)hls.de,

no later than 31 October 2025, as a single PDF file.

You will be working at HLRS within the Converged Computing (C2) department, headed by Dennis Hoppe. If you have any questions about this job offer, please e-mail converged.computing@hls.de.

The University of Stuttgart invites women to apply for this job opening to strengthen the presence of female workers in the scientific areas. Full-time positions may generally be turned into part-time positions. Disabled people will have priority as long as they are equally qualified. The central administration of the University of Stuttgart will handle the recruitment process.

Information on the handling of applicant data following Art. 13 DS-GVO can be found at: <https://www.uni-stuttgart.de/datenschutz/bewerbung/>.