

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), which is supported by a strong consortium from Germany, to successfully meet the growing demand for artificial intelligence (AI) infrastructure across Europe. HammerHAI will be a one-stop shop for many AI users, focusing primarily on start-ups, small and medium-sized enterprises (SMEs), large industrial companies, and supporting academic institutions and the public sector. It will offer tailored services and infrastructure to accelerate AI innovation and help develop a competitive AI ecosystem in Europe. The AI Factory will be located in a region that is one of Europe's powerhouses in manufacturing and engineering innovation, and it will be integrated 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 will support the entire AI lifecycle, from data preparation to model training, deployment, monitoring, and retraining, and will provide a comprehensive package of services to ensure efficient and effective AI development and operation.

[Shaping the Future of AI in HPC](#)

We are seeking a highly motivated AI/ML Innovation Engineer to support the deployment, monitoring, and optimization of AI infrastructure within the AI Factory HammerHAI at HLRS. This role focuses on ensuring scalable, secure, and high-performance AI services for various users, including start-ups, SMEs, industry, and research institutions. The successful candidate will work on integrating AI pipelines, deploying AI workloads in HPC environments, and developing monitoring and benchmarking frameworks. This position requires expertise in AI, ML operations, cloud-native technologies, and high-performance computing systems.

In this context, we are looking for a

AI/ML Operations Engineer
(m/f/d, up to TV-L 13, 100%)
HLRS_09_2025

to work on HammerHAI with a strong team of AI 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 up to the project's duration, which is currently scheduled to run 3 years. The salary for this position is based on your personal qualifications up to the level of TV-L 13.

Your Responsibilities

- Collect and analyse user requirements to tailor AI software architectures and stacks.
- Assess AI software components, evaluating security, compatibility, and performance requirements.
- Design, deploy, and optimise AI/ML pipelines in AI-optimised supercomputing environments.
- Develop and implement best practices for MLOps, including automation, version control, and containerization.
- Test, deploy, benchmark, integrate, and monitor AI system services and pipelines with OpenStack and Kubernetes.
- Analyse monitoring data, logs, and system metrics, setting up the monitoring systems when necessary.
- Provide technical support and guidance to users on deploying and optimizing AI workloads.
- Contribute to technical documentation, user guides, and best practice reports.

Your Profile

- A Master's or PhD in Computer Science, Artificial Intelligence, Computational Sciences, Engineering, or a related field.
- Strong expertise in AI, Machine Learning, and Deep Learning.
- Hands-on experience with relational and non-relational databases, distributed data processing, and AI/ML libraries (e.g., Spark, Flink, Ray).
- In-depth knowledge of Linux OS, Linux security, containers, CI/CD pipelines, IaC tools, logging tools, version control and issue tracking.
- 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.
- A good understanding of CPU, GPU, parallel computing architectures, and collective communications is a plus.
- Problem-solving mindset with the ability to address and resolve issues effectively.
- You are fluent in English, both written and spoken.

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, MLOps, 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_09_2025**” in the subject line, preferred by e-mail to: bewerbungen@hls.de, no later than **March, 12th 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/>.