

Technische Informationsbibliothek (TIB) - Leibniz Information Centre for Science and Technology



TIB, being part of the Leibniz Research Association, is with around 600 employees one of the largest technical information centres in Germany, having three main functions: 1. Acting in the capacity of the Leibniz Information Centre for Science and Technology, our forward-looking services secure the infrastructural requirements for providing researchers in science and industry with high-quality information and literature. 2. Carrying out cutting edge research to further expand TIB's role as a German information centre for the digitisation of science and technology. 3. Offer library services for the Leibniz University Hannover. As the German National Library of Science and Technology, our future-oriented services ensure the infrastructural requirements for a high-quality supply of information and literature for research in science and industry. With the Open Research Knowledge Graph (ORKG), we are working to revolutionise the exchange and use of scientific knowledge in the digital age.

Senior AI/ML Engineer or Postdoctoral Researcher - LLM Infrastructure (m/f/d)

City: Hanover; Starting date (earliest): 01/09/26; Duration: fixed-term for three years; Remuneration: pay scale group 13 TV-L; Reference number: 16/2026; Closing date: 14/06/26

Tasks

Your role

We are seeking a Senior AI / ML Engineer interested in building, optimizing, and operating general-purpose AI assistants at scale within the European Open Science Cloud (EOSC) ecosystem. The successful candidate will have a strong background in machine learning and artificial intelligence, sound mathematical understanding, and practical experience with LLM inference optimization in cloud environments.

The role focuses on scalable training and inference pipelines, token-efficiency and model selection strategies, distributed ML systems in production, and the evaluation of open models. It also includes optimizing latency, cost and performance trade-offs, supporting parameter-efficient adaptation workflows, and ensuring reliable integration with orchestration layers in EOSC service environments.

Your responsibilities will include

- Designing and developing inference infrastructure in cloud environments, with continuous optimization for latency, throughput and cost efficiency
- Benchmarking models against task-specific and general-purpose evaluation criteria and deriving data-driven recommendations on model-task fit
- Applying quantization, distillation, parameter-efficient fine-tuning and other optimization techniques to improve efficiency, performance and cost-

effectiveness

- Building an orchestration layer that integrates seamlessly with the broader agent framework and supports emerging AI protocols such as MCP, A2A and ACP
- Embedding observability from the outset using tools such as OpenTelemetry and Grafana to make model performance, cost and reliability transparent, measurable and actionable

Requirements

Your profile

- Completed scientific university degree (Master's degree or equivalent) in computer science, software engineering, artificial intelligence, data science or a related field
- Proven experience with LLM inference optimization in cloud environments
- Practical experience with fine-tuning, quantization, parameter-efficient fine-tuning (PEFT) and deployment of large language models in production or production-like settings
- Experience with model profiling, benchmarking and establishing performance baselines
- Experience in designing, developing or operating scalable training and inference pipelines within distributed AI/ML systems
- Knowledge of multi-agent systems, agentic workflows or orchestration frameworks, such as LangGraph or AutoGen
- Experience working with open-source LLMs, such as Llama or Mistral
- Very good programming skills in Python and experience in developing reproducible code, e.g. using PyTorch or comparable frameworks
- Good knowledge of cloud-native technologies, in particular Docker and Kubernetes
- Very good written and spoken English skills

Desirable additions to your profile

- Completed PhD in a relevant field
- Experience with agentic design patterns, such as reflection, ranking, exploration/discovery or human-in-the-loop feedback loops
- Experience handling scientific datasets and applying FAIR principles to data, model outputs or research workflows
- Strong written communication skills and experience in interdisciplinary collaboration, including contributions to scientific publications
- Familiarity with the Rust programming language

What we offer

The Technische Informationsbibliothek (TIB) – Leibniz Information Centre for Science and Technology – Program Area D, Open Research Knowledge Graph, is seeking a

Senior AI/ML Engineer or Postdoctoral Researcher - LLM Infrastructure (m/f/d)

to join the Artificial Intelligence for Scholarly Communication Research Group (Prof. Dr.

Sahar Vahdati) as of 1st September 2026.

The position is initially limited to the project duration of three years with envisioned extension. An extension is possible. The regular weekly working hours are 39.8 hours (full-time). The position is generally suitable for part-time work. The position is generally expected to be carried out on site in Hanover. Mobile work is possible to a certain extent, subject to operational requirements. The remuneration is based on pay scale group 13 of the collective agreement for the public service of the German states (TV-L).

What we offer

Our mission is to keep rethinking and innovating the provision and use of research data and information. In TIB's Research and Development Department you have the opportunity to advance your career in a sizable, dynamic and excellent environment. We provide a scientifically and intellectually inspiring environment with an entrepreneurial mindset, embedded in a leading technical university and one of the largest technical information centers being part of the Leibniz Association. TIB collaborates closely with the L3S Research Center at Leibniz University Hannover, one of the world's leading research institutes in the field of Web & Data Science, within the Leibniz Joint Lab Data Science & Open Knowledge. Last but not least, we attach great importance to an open, creative and fun work atmosphere.

Furthermore, we offer

- A job in the public service oriented towards the common good on the basis of the collective agreement for the public service of the German states (TV-L) with a salary in pay scale group 13 TV-L
- A flexible workplace in terms of time and space with offers to reconcile work and family life, such as mobile and remote work options as well as flexible working time models (flexitime)
- A special annual payment at the end of the year and 30 days of vacation per year with a five-day working week as well as additional days off on Christmas Eve (December 24th) and New Year's Eve (December 31st)
- A modern workplace in a central location of Hannover with a collegial, attractive and versatile working environment
- An employer with a wide range of internal and external further education and training measures, workplace health promotion and a supplementary pension scheme for the public sector (VBL)
- Employee discount in the canteens of the Studentenwerk Hannover as well as the possibility to use the various offers of the University Sports Hannover at a discount
- Independent and future-oriented activities that offer variety and room for personal development
- Funding for necessary equipment, conference and research visit travel
- Work in the context of a national, European or international research and innovation project
- A portfolio of technology components to build on, including ORKG, OpenResearch.org, TIB AV-Portal, DBpedia.org and other

If you are interested, questions regarding the position can be addressed to Prof. Dr. Sahar Vahdati, head of the Artificial Intelligence for Scholarly Communication Research, by mail

at Sahar.Vahdati@tib.eu.

Application

We look forward to receiving your application. To submit your application, please use the application form on our website <https://tib.eu/bewerbungsformular-16-2026eng>.

Paper applications are also possible on an equal basis. In this case, please send your complete application documents with the subject 16/2026 before June 14, 2026 addressed to

Technische Informationsbibliothek (TIB)
Personalservice
Herrn Daniel Eilers
Welfengarten 1 B
30167 Hannover
Germany

or alternatively as a single PDF document file to bewerbung@tib.eu. For applications in digital form, please send a PDF file with a maximum size of 10 MB.

We invite you to include an indication of your annual salary expectations with your application.

In the case of foreign university degrees, it is mandatory to submit a Statement of Comparability from the Central Office for Foreign Education (ZAB) for the final assessment of the hiring requirements during the hiring process (link for more information). A fee will be charged for this. This obligation does not apply if the foreign university degree has already been assessed by an expert and is listed in the ANABIN database of the ZAB as an equivalent university degree (link to ANABIN database).

TIB is an equal opportunity employer, providing ideal working conditions, and continuously taking action to enable its employees to combine career interests with family life. It wants to promote equal career opportunities for women and men in particular and therefore urges qualified women to apply.

Severely disabled candidates and those with equivalent status will be given preferential consideration in the event of equal suitability. If you wish your rights as a person with a severe disability or as a person deemed equivalent to a person with a severe disability to be taken into account during the selection process, please inform us of this voluntarily at an appropriate place in your application (e.g. in the cover letter or CV, or at another clearly highlighted point) and enclose appropriate proof (e.g. a copy of the disability certificate).

Equal opportunity is one of our core organizational values. We therefore welcome applications regardless of nationality, ethnic and social origin, gender identity, gender expression and/or gender characteristics, religion and worldview, age, sexual orientation, and identity.

Please indicate in the subject of your application which job exchange brought our offer to

your attention.

Please note that your application documents will not be returned, and application and travel costs cannot be reimbursed.

More information at <https://stellenticket.de/204170/HAWK/>

Offer visible until 14/06/26

