Zertifikat seit 2008
audit familiengerechte hochschule

TECHNISCHE UNIVERSITÄT BERLIN

Reference number: IV-428/26 **Published:** 12.11.2026
Start date (earliest): Earliest possible, befristet für 3 Jahre
Salary: Salary grade 13 TV-L Berliner Hochschulen
Full/Part-time: full-time, part-time employment may be possible **Application deadline:** 26.11.2026

Research Assistant

Faculty IV - Electrical Engineering and Computer Science, BIFOLD / Machine Learning and Security

About us

The Chair of “Machine Learning and Security” at the BIFOLD institute is seeking a research associate for the research project “AMIRA”. The project investigates the security and trustworthiness of mechanistic interpretability, a family of methods for explaining neural networks, such as circuit discovery and sparse autoencoders. The goal is to characterize the vulnerabilities of these methods, design novel attacks, and develop appropriate defense mechanisms.

Your responsibility

- Security analysis of mechanistic interpretability
- Research and development of attacks against these methods
- Research and development of defense mechanisms for these methods
- Scientific publishing and the opportunity to pursue a PhD

Your profile

- Successfully completed university degree (Master, Diplom or equivalent) in Computer Science or a related technical field
- Strong background in machine learning
- Strong background in computer security
- Experience in adversarial machine learning
- Experience in mechanistic interpretability
- Good knowledge of German and/or English required; willingness to acquire the respective missing language skills
- Ability to conduct independent research is desirable

Zertifikat seit 2008
audit familiengerechte hochschule

TECHNISCHE UNIVERSITÄT BERLIN

What we offer

- Exciting and challenging research
- Collegial and collaborative work environment
- Renowned and dedicated team
- International research network

How to apply

Please send your application with the **reference number** and the usual documents (one file max. 5 MB) only via email to **hashmi@tu-berlin.de**.

All documents must be submitted as a single PDF file. The application must include at least:

- Cover letter
- Complete curriculum vitae (CV)
- Certificates and transcripts (Master's and Bachelor's)
  If available, please also include letters of recommendation and publications.

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/.

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.

Data protection:

Full job posting: