KI-Servicezentrum Berlin-Brandenburg
kisz@hpi.de

SPONSORED BY THE
Federal Ministry of Education and Research

KI Service Zentrum by Hasso-Plattner-Institut

AI Service Centre Berlin-Brandenburg

Our Offerings at a Glance

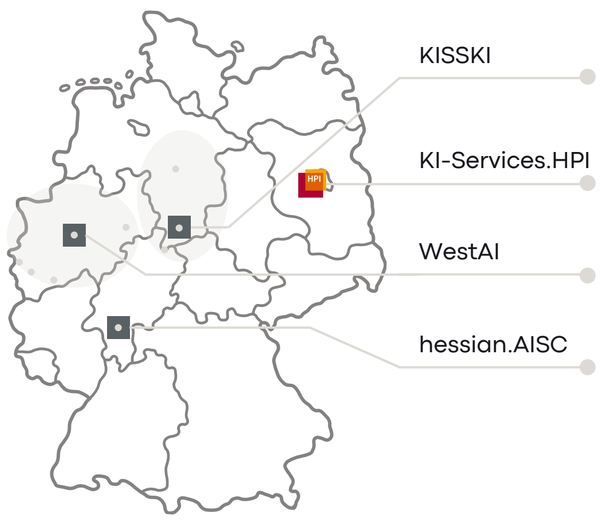
Want to find out more?
Contact us via our website
www.hpi.de/kisz

Mission

Our Goal

Our goal is to **lower the barriers** to implementing **Artificial Intelligence** in both **business and society**.

We support you in developing and professionalizing AI applications. **Scientific research** forms the basis for our **educational** and **advisory services**. In addition, our **AI infrastructure** offers platforms for model training and inference, laying the foundation for developing AI applications and integrating them into society.



Research

Contributions in AI Methods and AI Operations

Our research improves AI methods for process efficiency and optimizes datacenter AI operations.

- Multimodal AI models, addressing language biases in visual question answering, with applications in art and medicine: T. Bleidt, S. Eslami, and G. de Melo, "**ArtQuest: Countering Hidden Language Biases in ArtVQA**" in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 7326-7335.
- AI-based translation, improving multilingual accuracy and document-level translation: Y. Al Ghussin, J. Zhang, and J. van Genabith, „**Exploring Paracrawl for Document-level Neural Machine Translation**“, in Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 2023, pp. 1304–1310. doi: 10.18653/v1/2023.eacl-main.94
- Large language model adaptation and efficient training: K. Dobler and G. De Melo, „**FOCUS: Effective Embedding Initialization for Monolingual Specialization of Multilingual Models**“, in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 13440–13454. doi: 10.18653/v1/2023.emnlp-main.829
- Energy Efficient AI Model training: N. Alder and R. Herbrich, „**Energy-Efficient Gaussian Processes Using Low-Precision Arithmetic**“, 2024, in Proceedings of the 41st International Conference on Machine Learning, 2024, pp. 955–975.

Computing Resources

Our Heterogeneous AI Infrastructure

To reduce the barriers to implementing your **AI applications**, we provide access to our **AI infrastructure** , tailored for the following needs:

- **Training**, **applying** and **optimizing** AI models
- Leveraging **high-performance computing** resources
- Analyzing costs and performance across various **workloads and hardware types**

Training
- 8 x 8 NVIDIA H100 GPU

Inference
- 5 x 8 NVIDIA A30 GPU

ARM Server
- Ampere Altra Max M128-30 CPU
- 2 x NVIDIA L40 GPUs

GPU Server
- AMD Epyc CPU
- 8 x NVIDIA L40S GPU

Edge
- ARMv8 CPU
- NVIDIA Jetson AGX Module

Storage
- 1.5 PB NVRAM

Network
- 400 Gb/s Infiniband
- 200 Gb/s Ethernet

Education

Enhance and Expand Your AI Knowledge and Skills

Our educational offerings, including hands-on workshops, Massive Open Online Courses (MOOCs), and guest lectures, provide opportunities to explore the following questions:

- What does **generative AI** mean **for** our **society**?
- How can you **build** your own **neural networks**?
- What are **word** or **image embeddings**?
- How does **visual question answering** work?
- How can AI be used in industries such as **banking**?

To find out more about our workshops, please visit: **https://aimaker.community/**

Our AI Data Center with NVIDIA DGH H100 Basepods.

Advisory

Customized AI Application Advisory Services

Our expertise allows us to provide guidance on the following tasks:

- **Identifying ideas** for AI projects and applications
- Selecting **datasets**, **models**, **frameworks,** and **metrics**
- Planning and scaling with options for **cloud, on-premise, and hybrid solutions**
- Implementing proof-of-concepts, AI prototypes, and **AI pilot projects**

You can apply for our AI pilot projects every two months!

We welcome your questions and are happy to assist you during our AI Advisory hours. Please contact us at **kisz@hpi.de.**