HELMHOLTZ

Helmholtz Benchmark Projects – Call UNLOCK 2025

funded by the Helmholtz Initiative and Networking Fund

1 Background

The Helmholtz Association has a long-standing mission to address grand societal and scientific challenges in domains such as health, energy, climate, aerospace, transport, and materials science. As artificial intelligence (AI) continues to reshape scientific discovery, it is crucial to ensure that AI advancements are effectively harnessed to accelerate progress in these fields. One key mechanism for achieving this is through well-designed benchmarking datasets that define AI tasks in Helmholtz-relevant domains, motivating AI researchers to contribute solutions that directly impact scientific progress.

Challenge: The Need for Benchmarking in Reproducible and Trustworthy Science

AI's rapid advancement has created a critical demand for high-quality benchmarking datasets to ensure reproducibility, robustness, and fairness in model development. Projects like metrics reloaded have already shown the overarching need for improved benchmarking. Helmholtz's entry into this field strengthens its competitiveness in trustworthy AI, ultimately improving solutions in domains like medical imaging and climate forecasting. With the working group on benchmarking datasets, the incubator Information and Data Science has discussed this topic since 2023. Based on these discussions, the hyperUNLOCK Initiative was formed and produced the underlying concept paper for this call.

Existing AI benchmarks suffer from major limitations: they are often narrow in scope, lack standardization, and face accessibility barriers due to proprietary restrictions or insufficient metadata. The absence of standardized, diverse, and multimodal datasets hampers scientific progress and can lead to misleading conclusions about AI performance. Compared to unimodal single-domain benchmarks, multimodal cross-domain benchmarks could truly test ambitious AI models, such as foundation models. Thus, they offer a more comprehensive evaluation by assessing generalization, robustness, and the ability to handle diverse data types and contexts.

Helmholtz is uniquely positioned to address this gap. With vast, high-quality datasets spanning multiple scientific fields, it can pioneer cross-domain and multimodal benchmarks, enabling more comprehensive AI evaluations. By leveraging its data assets, Helmholtz can drive groundbreaking advancements and position itself as a leader in the emerging science of benchmarking - particularly in supporting foundation models designed for multimodal AI. To address this challenge, and to initiate a community of practice for benchmarking, and to support these first systematic benchmarking activities, the Helmholtz Association supports the hyperUNLOCK Initiative with this first call UNLOCK[1] for benchmarking projects.

[1] The Helmholtz Association is uniquely positioned to fill this gap: First, it holds vast amounts of data across its numerous research centers. These diverse and high-quality datasets, often generated for specific scientific questions, remained largely untapped for broader AI research so far. Unlocking and leveraging these data assets could catalyze groundbreaking scientific advancements in domain and AI research. ***Solution: Unlocking Helmholtz data as key benchmarks.*** By unlocking the data treasures within the Helmholtz Association as benchmarking datasets, hyperUNLOCK aims to foster innovation, improve algorithmic robustness, and promote transparency, ultimately driving transformative advancements in the socioeconomically most relevant fields.

2 Objective of funding

The UNLOCK call leverages the Helmholtz Association's data treasures to create benchmarking datasets that drive innovation, enhance algorithmic robustness, and promote transparency. This benefits both domain scientists, who gain AI-driven insights, and AI researchers, who access high-quality datasets for model development.

Creating such benchmarks requires careful preprocessing, realistic data splits, and standardized metadata to ensure usability. The UNLOCK call facilitates this by connecting Helmholtz centers and domains to generate multimodal and cross-domain datasets. Collaboration with Helmholtz AI, Helmholtz Imaging, HIFIS, and HMC ensures best practices in metadata handling, reproducible pipelines, and open access.

Beyond academia, UNLOCK's benchmarks aim to support industry and public sectors by providing real-world datasets for AI development. This can accelerate technological advancements, enhance product reliability, and inform public policy in fields like healthcare, environmental protection, and energy management.

3 Goals

The overall goal of the UNLOCK call is …

to promote science of benchmarking
to encourage cross disciplinary and cross domain interaction
to foster and publish benchmarking datasets
to identify common standards and pipelines
to establish a community of practice of benchmarking in the Helmholtz Association

Given the natural diversity of possible projects, different milestones and goals can be defined depending on the starting point of each participant. Every project should clearly describe its starting point, the scientific benefit of the benchmarking data set and the achievable product after one year of funding. It is the goal of this initiative to identify and support projects in the continuum between producing a much needed benchmark data set and the expansion of an already existing data set e.g. into a multimodal benchmarking data set:

- domain scientists may begin by structuring raw data, followed by annotation and quality control, with the goal of creating a high-quality usable benchmark close to publishing,

- while data scientists aiming for AI competitions may expand existing datasets, start defining concrete competition tasks, establishing baseline models and sound evaluation pipelines, with the ultimate goal of running a high-quality AI competition and attracting a diverse group of scientists to participate.

Each project will have to produce a publicly available benchmarking data set and /or run a competition on their data.

By the end of 2026, benchmark datasets will be submitted for publication, e.g., for the NeurIPS Datasets Benchmarks[2] track or at Nature Scientific Data[3] and depending on their maturity should

[2] https://neurips.cc/virtual/2024/events/datasets-benchmarks-2024
[3] https://www.nature.com/sdata/

be submitted to kaggle[4] or similar competition platforms e.g. the Jülich Challenge.

Scientists with lacking expertise may receive additional support in preparing their data from the to-be-established Helmholtz Imaging data curation unit, as well as the Helmholtz AI Consultant teams.

3.1 Establishing a community of practice

Besides individual projects, this initiative aims to establish a community of practice in the field of benchmarking datasets. Therefore, all projects will be required to contribute to workshops, i.e., **each project should reserve a budget of 5.000€ to co-organize a workshop** together with other projects. Helmholtz Imaging will support these community driven workshops. These workshops will help to clarify the benchmarking approach, and the demands by different communities, but will also help to join forces and connect the projects with the Helmholtz Information and Data Science Framework.
A kick-off workshop (Q1 2026) will provide networking opportunities for all participants and help to define shared standards and goals. A mid-term workshop (mid 2026) will collect best practices, identify opportunities to promote the derived benchmarks and discuss the potential to combine individual projects into cross-domain and multimodal benchmarks. A finishing workshop (Q4 2026) will help to plan further activities for the projects and open to-dos for the Helmholtz-Association. Moving into 2027, the need for additional benchmarking datasets will be clearer, potentially leading to another call for projects.

4 Criteria

Applications that meet the following criteria will be evaluated based on the evaluation criteria (see below in “Evaluation process, selection, and evaluation criteria”).

4.1 Thematic Criteria

With the UNLOCK call, the Helmholtz Information and Data Science Framework wants to attract promising candidates for publishing benchmarking datasets. Hence, we envision a wide range of potential project applications with different starting points.

1. **All projects have to produce a high quality benchmarking data set, answering a demand from their research community.**
2. **Depending on the demand from the respective communities, existing benchmarking datasets could be expanded into multimodal benchmarking datasets.**
3. **All selected projects have to submit the benchmarking data set for publication/dissemination, e.g., for the *NeurIPS Datasets Benchmarks*[5] track or at *Nature Scientific Data*[6].**
4. **Depending on the starting point of the proposed project the UNLOCK call encourages applicants to run an AI competition on their data.**

4.2 General Requirements

For all proposals, the following criteria should be considered to select the best applications:

[4] https://www.kaggle.com/
[5] https://neurips.cc/virtual/2024/events/datasets-benchmarks-2024
[6] https://www.nature.com/sdata/

1. **Scientific Relevance and Data Quality**
    - Motivation: Alignment with Helmholtz's mission and the potential to address grand societal challenges. Explanation of research gap and missing data for the envisioned application.
    - Scientific advancement: Clear demonstration of how the dataset can advance scientific knowledge and technological development. This includes specifying the potential for computational innovation, e.g., in data science challenges, that have not been sufficiently addressed or the need for AI solutions.
    - Data source: Description of which data will be used and how high standards of data integrity, completeness, accuracy, and representativeness will be achieved.
2. **Benchmarking Potential**
    - Quantifiability: Clear definition and justification of benchmarking tasks, metrics, and evaluation protocols.
    - Benchmarking innovation: Potential to set new standards for AI model performance assessment.
    - Longevity: Likelihood of the dataset remaining relevant and valuable over time.
3. **Accessibility and Usability**
    - A strategy to promote the use of the data set, e.g. by hosting a competition or by other suitable means to generate impact and make the data set available to the research communities.
    - Open access: Plans for open access or well-structured sharing protocols that promote broad usage.
    - Meta data: Comprehensive documentation, definition of metadata, and tools to facilitate dataset adoption by the research community.
    - Sustainability: Outlook on how the benchmark could be maintained and extended in the future, e.g. in follow-up projects or as part of other research activities.
4. **Project Soundness**
    - Clear and realistic work plan: Measurable achievements with defined goals and milestones reaching the goal
    - Demonstration of the project consortium's capability to carry out the proposed work, e.g. through previous achievements of team members, or, in case of lacking experience, a clear plan how to obtain that expertise.
5. **OPTIONAL: Interdisciplinary and Collaborative Potential**
    - Cross-domain: Promotion of cross-disciplinary research and collaboration within and beyond the Helmholtz Association.
    - Transfer: Involvement of stakeholders from academia, industry, and the public sector.
    - Cross-center: Outline

5 Applications

Helmholtz Imaging, HMC and Helmholtz AI offer support for projects, which can also be very helpful at the application stage. If, for example, expertise in specific areas is missing for the implementation of the project, it can be checked together with the different support units hosted by Helmholtz Imaging, Helmholtz AI and HMC (s. below).

It is strongly recommended to check the need for such an addition of expertise before submitting the application.

The application consists of:

a) Cover page including English abstract (max. 0.5 page)
   - Important elements on the cover page include Title and acronym,
   - Up to 10 keywords,
   - Names and contact information (including department/institute/section and research group) of primary contact Helmholtz Center and person, coordinator(s) and Principal Investigator(s),
   - Names of participating centers and external partners,
   - budget

b) Main body (max. 5 pages),

c) Appendix (work packages, deliverables, milestones, budget, CVs),

d) The cover page, the main body and the appendices shall be submitted as one pdf-document.

e) The application must be written in English and formatted in DIN A4, Arial, 11pt, 2.5 cm margins, and single-line spacing. A full template specifying the structure and formatting is provided in Appendix 1 to this call.

f) The application's main body's structure is:

1. **Scientific Relevance and Data Quality (**starting point, envisioned goal)
2. **Benchmarking Potential (**Uniqueness, community demands)
3. **Accessibility and Usability (**datasets, provenience, availability, ethics evaluation)
4. **Project Soundness (**Implementation and management, feasibility)
5. **OPTIONAL: Interdisciplinary and Collaborative Potential**

CVs of the principal investigators should be included in the appendix, highlighting relevant expertise for the project. Please note that the CVs should also contain references to relevant accomplishments beyond scientific papers and citation metrics; such accomplishments may for instance be software packages, policy papers, standards, datasets, patent filings, entrepreneurship, and industry collaborations.

g) A budget plan has to be included.

h) A declaration, usually a signed letter, by the board of directors of participating centers must be included, guaranteeing that their own funds will match at least the amount applied for from the INF.

i) The application may include a list of six unbiased independent experts with expertise relevant to the application (experts could be invited to participate in the selection process). The list should be gender balanced.

j) An incomplete application may lead to the exclusion of the application from the selection procedure.

Use of funding

Each project selected and funded under the UNLOCK call will be funded with up to 150.000€ Initaitive and Networking Fund contribution. **The INF contribution has to be meet with matching funds accordingly.** Each project will produce a public benchmarking data set. Funding can support a wide range of activities, including data processing and annotation support, implementing benchmarks online etc. Each project will be required to support a community workshop, see below for details. Projects spanning domains and centers, e.g., to achieve multimodal or cross-domain

benchmarks, are allowed to apply for a larger INF contribution than projects with a narrower focus. All funded projects must comply with the INF regulations on the use of funds and the corresponding annual reporting.

Helmholtz UNLOCK benchmarks will be part of the strategic endeavor within the Helmholtz Association to build a community of practice and a benchmarking data set ecosystem. Together with Framework Information and Data Science and the services and support provided by Helmholtz Imaging, Helmholtz AI, Helmholtz Foundation Model Initiative, Helmholtz Metadata Collaboration and HIFIS the selected and funded projects under this call will benefit from the rich portfolio of services in addition to financial support.

- Each project accepted for funding can receive scientific and/or technical advice and benefit from the expertise offered by the Support and Research Units from Helmholtz Imaging, HMC and Helmholtz AI. Helmholtz Imaging Support Hub *or the* Helmholtz AI consultant units or the Helmholtz Metadata Collaboration Hub Matter whether gaps can be closed, e.g. via a Helmholtz Imaging Collaboration (at no additional cost). Additions to the project expertise found with the Helmholtz Imaging Support Hub can then be specified in the project proposal.
- The scientific results and software tools obtained by the UNLOCK Benchmarking projects will be made available as open source at the Helmholtz Imaging Solutions and Helmholtz Imaging Connect. To ensure sustainability for their dissemination and reusability projects are encourage using the HIFIS dCache for storing.
- There is as well the opportunity to access and employ the computing resources of HAICORE.
- Projects planning to host a challenge on their data are invited to use the Jülich Challenges platform for hosting their challenges

6 Evaluation process, selection, and evaluation criteria

Proposals are subject to the following evaluation procedure. Proposals are checked with regard to formal requirements. The evaluation panel performs a pre-assessment of all applications. Proposals are evaluated by a panel of independent experts (no presentation). Based on this evaluation, a ranked list of projects for funding will be produced. The president will review the ranked list of proposals and decide which projects shall be funded.
Proposals will be evaluated on the basis of the evaluation criteria ‘Scientific Relevance and Data Quality’,‘Benchmarking Potential’, ‘Accessibility and Usability’, ‘Project Soundness’, and the optional criteria ‘Interdisciplinary and Collaborative Potential’.

Evaluation criteria	Aspects
Scientific Relevance and Data Quality	○ Motivation: Alignment with Helmholtz’s mission and the potential to address grand societal challenges. Explanation of research gap and missing data for the envisioned application. ○ Scientific advancement: Clear demonstration of how the dataset can advance scientific knowledge and technological development. This includes specifying the potential for computational innovation, e.g., in data

	science challenges, that have not been sufficiently addressed or the need for AI solutions. ○ Data source: Description of which data will be used and how high standards of data integrity, completeness, accuracy, and representativeness will be achieved.
Benchmarking Potential	○ Quantifiability: Clear definition and justification of benchmarking tasks, metrics, and evaluation protocols. ○ Benchmarking innovation: Potential to set new standards for AI model performance assessment. ○ Longevity: Likelihood of the dataset remaining relevant and valuable over time.
Accessibility and Usability	○ A strategy to promote the use of the data set, e.g. by hosting a competition or by other suitable means to generate impact and make the data set available to the research communities. ○ Open access: Plans for open access or well-structured sharing protocols that promote broad usage. ○ Meta data: Comprehensive documentation, definition of metadata, and tools to facilitate dataset adoption by the research community. ○ Sustainability: Outlook on how the benchmark could be maintained and extended in the future, e.g. in follow-up projects or as part of other research activities.
Project Soundness	○ Clear and realistic work plan: Measurable achievements with defined goals and milestones reaching the goal. ○ Demonstration of the project consortium's capability to carry out the proposed work, e.g. through previous achievements of team members, or, in case of lacking experience, a clear plan how to obtain that expertise

OPTIONAL: Interdisciplinary and Collaborative Potential	○ Cross-domain: Promotion of cross-disciplinary research and collaboration within and beyond the Helmholtz Association. ○ Transfer: Involvement of stakeholders from academia, industry, and the public sector. ○ Cross-center: Outline

The evaluation will be managed by the Helmholtz Association Head Office. Neither Helmholtz Imaging nor Helmholtz AI nor HMC is not involved in the evaluation process.

7 Schedule

Date	Event or action
April 1, 2025	Publication of the call for Helmholtz Benchmarking Projects.
April 10, 2025, 10:00 AM - 12:00 PM	Information meeting on the application process; for agenda and registration see https://events.hifis.net/event/2234/
May 30, 2025	Application deadline, Submissions are accepted via the ProMeta platform, https://ivf.helmholtz.de/
until end of June 2025	Assessment by the panel members, funding decision by the president.
until end of September 2025	Funding contracts between Helmholtz Association and submitting centers are being drawn up.
Starting January 2026	Start of Helmholtz Benchmark Projects.

8 Contact information

For further inquiries, please contact:

- Helmholtz Head Office: Florian Grötsch via florian.groetsch@helmholtz.de

9 Appendices

- Application template (Proposal template for proposal submission, Templates for budget tables, Template for Gantt Chart, Template for CV)
- Criteria for bias/ conflict-of-interest (for suggested reviewer),
- Data protection information