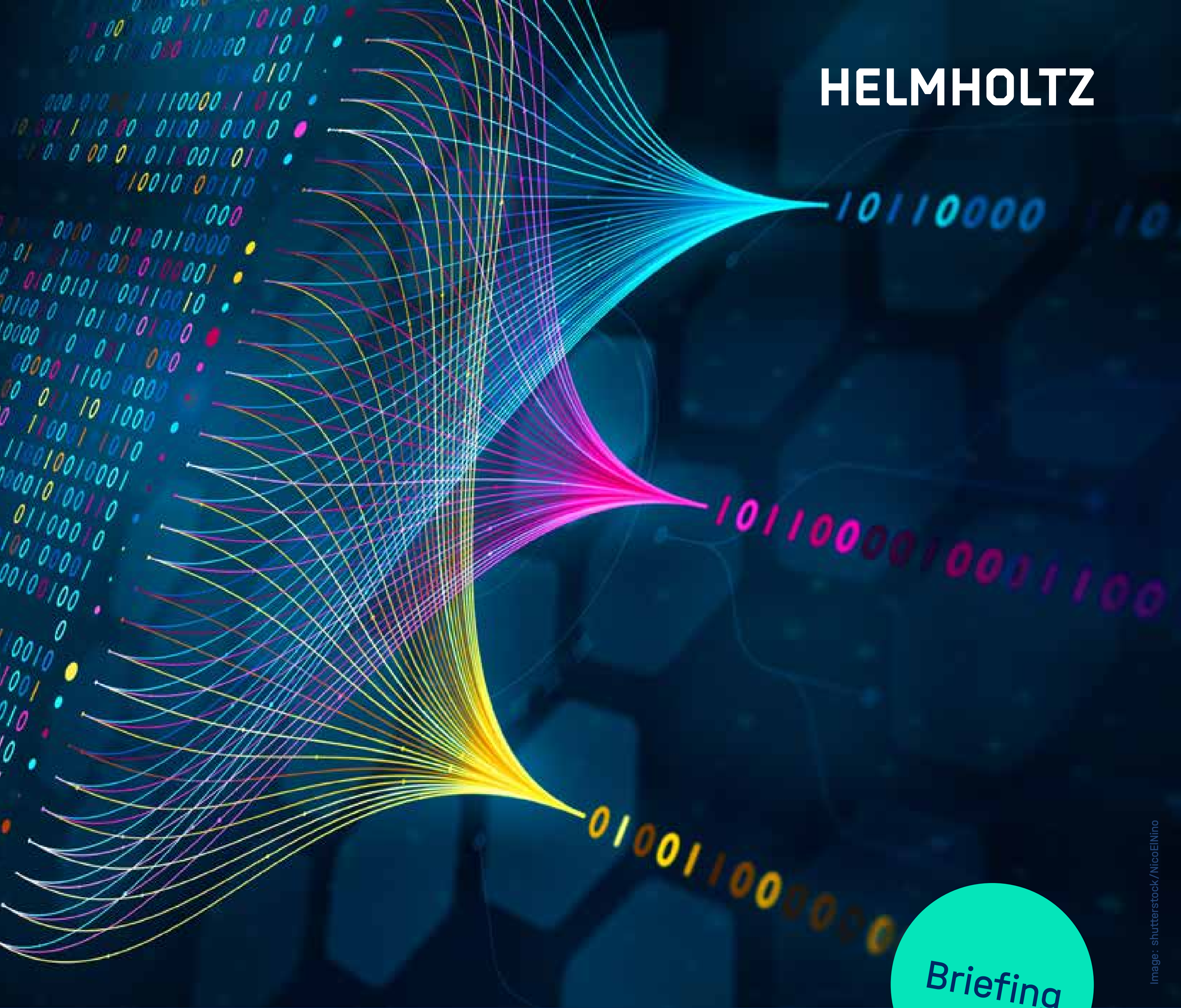


Briefing

Helmholtz Foundation Model Initiative

Foundation Models for Science and Society

Foundation Models

Foundation Models represent a new generation of AI models characterized by a broad knowledge base, enabling them to solve a range of complex problems. These models are significantly more powerful and flexible than traditional AI models, offering enormous potential for modern, data-driven science. They can become powerful tools capable of answering a variety of research questions.

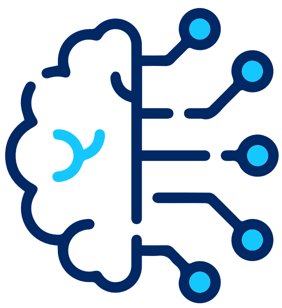
The Helmholtz Association provides ideal conditions for developing such forward-looking applications: an abundance of data, powerful supercomputers for training the models, and extensive expertise in artificial intelligence.

Our goal is to develop Foundation Models across a wide spectrum of research fields to address the major questions of our time.

Why Foundation Models?

Today, research generates enormous amounts of data. In addition to the quantity and quality of the data, computing power and highly trained data scientists are crucial for scientific progress. However, the full potential of this data can only be realized by evaluating increasingly large datasets. The rapid development in the field of artificial intelligence (AI) opens up entirely new possibilities for unlocking previously inaccessible treasure troves of data and tackling hitherto unsolvable problems.

This new dimension can be achieved with a new generation of AI models: the so-called Foundation Models. Currently, Foundation Models are known in the context of natural language processing (for example, Chat-GPT). However, they are not limited to language applications and can be extended to any field. This versatility makes them particularly interesting for research.



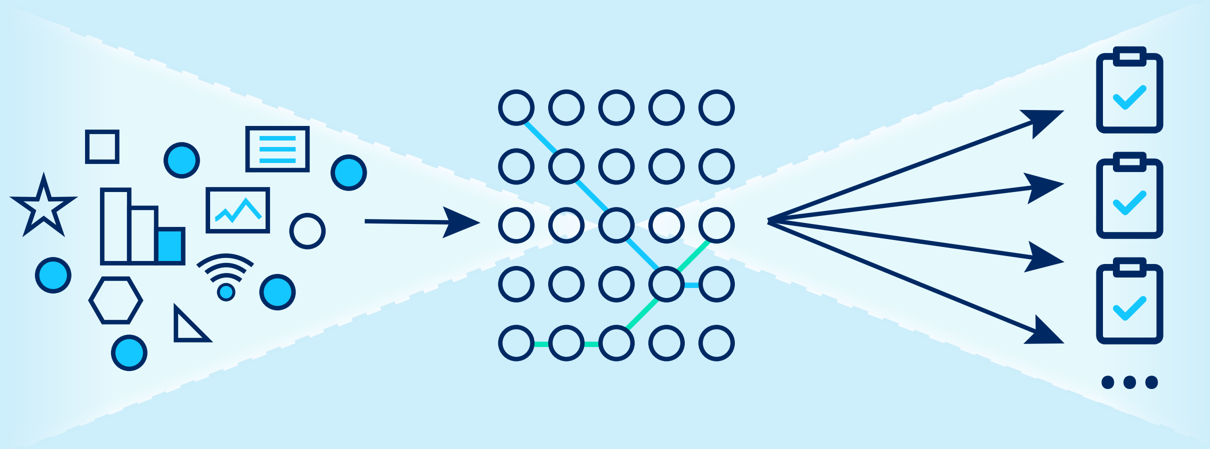
Foundation Models are complex AI models with an immeasurable potential for science. The essential ingredients for these models include vast amounts of data, immense computing power, scientific expertise, and knowledge in artificial intelligence and supercomputing. By establishing the Helmholtz Foundation Model Initiative, Helmholtz is pioneering work in this field.

What is Helmholtz's Contribution?

At the Helmholtz Association, we aim to train Foundation Models and make them available to the scientific community. AI experts from various Helmholtz Centers have come together to design a concept for the three-year Helmholtz Foundation Model Initiative (HFMI).

The initiative's goal is to develop fully functional models while considering ethical and legal standards. Four pilot projects have been selected, involving scientists from twelve Helmholtz Centers. A key unique feature of the HFMI is the collaboration between AI experts and domain scientists within these projects. Helmholtz ensures that not only do the models make a meaningful contribution to research, but also that the results withstand empirical scrutiny. Over three years, the projects will receive funding of 11 million Euros, with an additional 12 million Euros invested in expanding the necessary infrastructure.

How Are Foundation Models Trained?



Foundation Models are machine learning models that are trained with very large amounts of processed data. It is crucial that the data is well-structured and that relatively unspecific pre-training takes place. Through this pre-training, they are able to understand complex relationships based on learned patterns, generate new connections, and make predictions. The foundation developed from this initial broad training can be applied to many downstream tasks.

This distinguishes Foundation Models from traditional AI models, which are trained for a very specific task. With such a foundational knowledge base, a range of scientific questions can be addressed, often requiring only a small amount of task-specific data for effective performance.

HELMHOLTZ

In addition, a Synergy Unit researches interdisciplinary questions, promotes the exchange of knowledge between individual projects, and carrys out overarching activities. Special emphasis is placed on methodological questions that relate to the understanding of how such models function. Understanding, including explaining and reproducing results, is crucial for building trust in the outcomes of Foundation Models. Therefore, Helmholtz places a strong emphasis on various quality assurance measures. The funded projects should provide a clear added value for science. Furthermore, they are committed to the principle of Open Science and make their final results available to society - from the code to the training data and the trained models.

A Revolution in Science

Foundation Models undoubtedly have the potential to fundamentally change the way research is conducted. Never before in the history of science has it been possible to analyze such large and complex datasets. The application of Foundation Models is as diverse as science itself, and their potential is far from exhausted. In the coming years, there will be numerous new developments.

The Helmholtz Foundation Model Initiative positions Helmholtz as a global leader in this field. The four selected pilot projects will use artificial intelligence to make radiological diagnoses more reliable, improve the understanding of the global carbon cycle, enhance climate models and weather forecasts, and accelerate the development of a new generation of photovoltaic modules.

> "We are convinced that with Foundation Models, we can push the boundaries of science. Helmholtz brings together not only outstanding talents and comprehensive datasets from various research areas, but also a unique computer infrastructure."

Otmar D. Wiestler
President of the Helmholtz Association

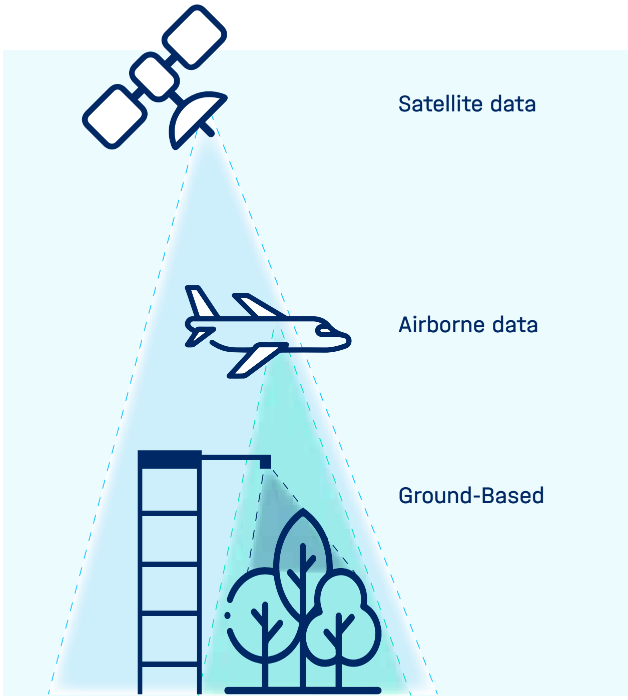
HELMHOLTZ

Foundation Models for Science and Society

3D-ABC

Calculation and Visualization of the Global Carbon Budget of Vegetation and Soils

Measurement of the structure of tropical forests and their carbon stock using terrestrial lasers – location: Ghana (image: Christian Budach/GFZ).



Data collection through satellites, drones, and local CO_2 capture stations (illustration: Helmholtz).

To mitigate the consequences of global climate change, we need in-depth knowledge of the global carbon budget, consisting of CO_2 sources and CO_2 sinks such as peatlands, forests, and permafrost soils. Until now, researchers have struggled to quantify how changes in land areas, vegetation, or soils affect the carbon cycle due to the heterogeneity and dispersion of data. The Foundation Model 3D-ABC will integrate model data from various sources, such as satellites, drones, and local CO_2 capture stations. This approach will allow key parameters of the global carbon cycle in vegetation and soils to be captured, quantified, and characterized with high spatial resolution.

Participating Helmholtz Centers:

- Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI)
- Forschungszentrum Jülich
- Helmholtz Center Potsdam – GFZ German Research Center for Geosciences
- Helmholtz-Zentrum Dresden-Rossendorf (HZDR)
- Helmholtz Center for Environmental Research – UFZ
- German Aerospace Center (DLR)

The Human Radiome Project

3D Radiology Data for a Fundamentally Better Understanding of Anatomy and Pathology

The more accurately tumors can be localized and marked, the more successful radiation therapy will be. However, medicine has so far faced significant challenges in this area because it has been very labor-intensive to correlate results from different imaging modalities and present them in a three-dimensional format. The precise localization and marking of tumors is one of many procedures that the "Human Radiome Project" aims to improve in the field of medical imaging. The project combines the world's most extensive and diverse collection of 3D radiological images, such as MRI and CT scans, into a Foundation Model. This allows researchers to gain deep insights into human anatomy and pathology, as well as a comprehensive overview of the full spectrum of radiological information. The "Human Radiome Project" not only enhances personalized medicine but also increases diagnostic efficiency by reducing the need to manually annotate complex medical images.

Participating Helmholtz Centers:
- German Cancer Research Center (DKFZ)
- German Center for Neurodegenerative Diseases (DZNE)
- Max Delbrück Center

HClimRep

Capturing Interactions Between Atmosphere, Ocean, and Sea Ice in a Novel Climate Model

What if we could make climate predictions more accurately, much faster, and more efficiently? Could we better combat the causes of climate change and mitigate its effects? Could we make the impacts of global warming impressively visible to everyone? The HClimRep project aims to answer precisely these questions. By building one of the first Foundation Models for climate research that combines data from the atmosphere, the ocean, and sea ice, researchers are developing one of the world's most precise weather and climate models. This deep learning model, with billions of parameters, will be trained on Europe's first exascale computer. This will enable it to conduct complex "what-if" experiments and other modeling tasks for the ocean and atmosphere.

Participating Helmholtz Centers:
- Forschungszentrum Jülich
- Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI)
- Karlsruhe Institute of Technology (KIT)
- Helmholtz-Zentrum Hereon

SOL-AI

Development and Optimization of Photovoltaic Materials

Photovoltaics is a key technology for the energy transition. For it to be adopted worldwide on a sufficient scale, innovative solar cell concepts need to be implemented much faster. Research and development activities in this area are increasing rapidly, resulting in a wealth of scientific publications. However, the sheer volume of data limits the implementation of the latest findings. SOL-AI aims to create a Foundation Model that will fundamentally reform materials informatics in this field. SOL-AI will be capable of integrating the diverse experimental data and research results on photovoltaic materials to drive innovations in various areas: from accelerating component development and optimization to the discovery of new solar materials. SOL-AI aims to develop solutions that have practical relevance for both research and industry.

Participating Helmholtz Centers:

- Forschungszentrum Jülich
- Karlsruhe Institute of Technology (KIT)
- Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)
- Helmholtz-Zentrum Hereon

Synergy Unit – Developing, Providing, and Connecting Foundation Models

While the individual projects focus on their specific issues, a Synergy Unit concentrates on overarching questions that are relevant to all participating projects. For example, it addresses questions related to the scalability of the models or training with datasets. It is not simply about exchanging solutions but, crucially, about how research on Foundation Models can be advanced as quickly as possible across disciplinary boundaries. Thus, the Synergy Unit ensures the long-term impact of the Helmholtz Foundation Model Initiative for the benefit of society.

Participating Helmholtz Centers:

- German Cancer Research Center (DKFZ)
- Helmholtz Munich
- Forschungszentrum Jülich
- Max Delbrück Center

Bild: Forschungszentrum Jülich/Ralf-Uwe Limbach

JUPITER –
Perhaps the Most Powerful AI Supercomputer in the World

Foundation Models are not possible without the computing power of supercomputers. Helmholtz possesses these large-scale computers and makes them available to the research community. A key role is played by the first European exascale computer at the Forschungszentrum Jülich. This supercomputer will break the barrier of one quintillion calculations per second. This supercomputing infrastructure is complemented within the Helmholtz Association by a number of smaller, powerful computers and computer clusters, known as "Helmholtz AI computing resources" (HAICORE). With the HAICORE infrastructure, Helmholtz strategically expands its supercomputing capabilities, providing not only a first-class research environment but also ensuring that supercomputer resources are only utilized when success is guaranteed.

Published by:

Helmholtz Association of
German Research Centers e.V.
Anna-Louisa-Karsch-Str. 2
10178 Berlin
www.helmholtz.de

Contact:

Florian Grötsch
Manager Information and
Data Science
Helmholtz Association
Tel. +49 30 206 329-669
florian.groetsch@helmholtz.de

More information at:

www.helmholtz.de/ki

As of:
June 2024

HELMHOLTZ