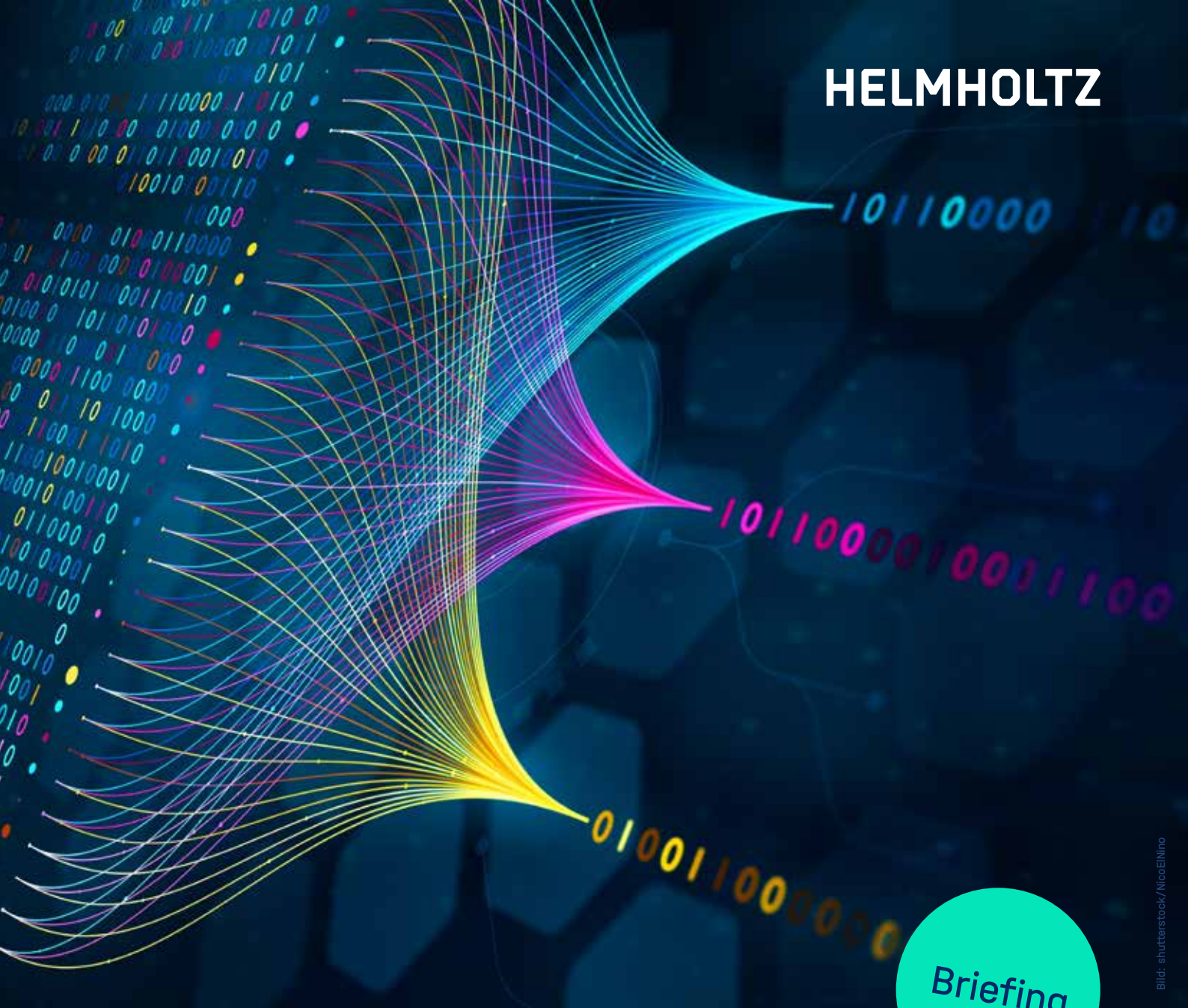


HELMHOLTZ



Briefing

# Helmholtz Foundation Model Initiative

KI-Grundlagenmodelle für Wissenschaft und Gesellschaft

Bei den sogenannten Foundation Models handelt es sich um eine neue Generation von KI-Modellen, die eine breite Wissensbasis haben und deshalb in der Lage sind, eine Reihe von komplexen Problemen zu lösen. Sie sind deutlich leistungsstärker und flexibler als herkömmliche KI-Modelle und bergen somit ein enormes Potenzial für die moderne, datengetriebene Wissenschaft. Sie können mächtige Werkzeuge werden, die eine Vielzahl von Forschungsfragen beantworten.

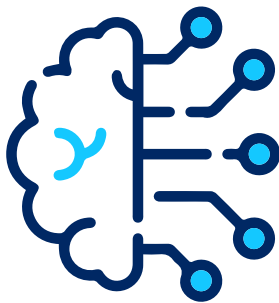
Die Helmholtz-Gemeinschaft bringt ideale Voraussetzungen mit, um solche zukunftsweisenden Anwendungen zu entwickeln: eine Fülle an Daten, leistungsstarke Supercomputer, auf denen die Modelle trainiert werden können, und eine tiefgreifende Expertise im Bereich der künstlichen Intelligenz.

Unser Ziel ist es, Foundation Models über ein breites Spektrum von Forschungsfeldern hinweg zu entwickeln, die zur Lösung der großen Fragen unserer Zeit beitragen.

## Warum Foundation Models?

In der Wissenschaft fallen riesige Mengen an Daten an. Neben der Menge und Qualität der Daten sind Rechenpower und gut ausgebildete Datenwissenschaftler:innen heute entscheidend für den wissenschaftlichen Fortschritt. Das volle Potenzial dieser Daten lässt sich jedoch nur ausschöpfen, wenn wir in der Lage sind, immer größere Datenmengen auszuwerten. Die rasante Entwicklung im Bereich der künstlichen Intelligenz (KI) eröffnet dabei völlig neue Möglichkeiten, um bislang unzugängliche Datenschätze zu erschließen – und somit auch bislang unlösbare Problemstellungen anzugehen.

Diese neue Dimension kann die Wissenschaft mit einer neuen Generation von KI-Modellen erreichen: den sogenannten Foundation Models (auf Deutsch: KI-Grundlagenmodell oder KI-Basismodell). Foundation Models sind heute im Zusammenhang mit der Verarbeitung natürlicher Sprache bekannt (zum Beispiel Chat-GPT). Sie sind aber nicht auf Sprachanwendungen beschränkt, sondern können auf jeden Bereich ausgedehnt werden. Das macht sie auch für die Forschung so interessant.



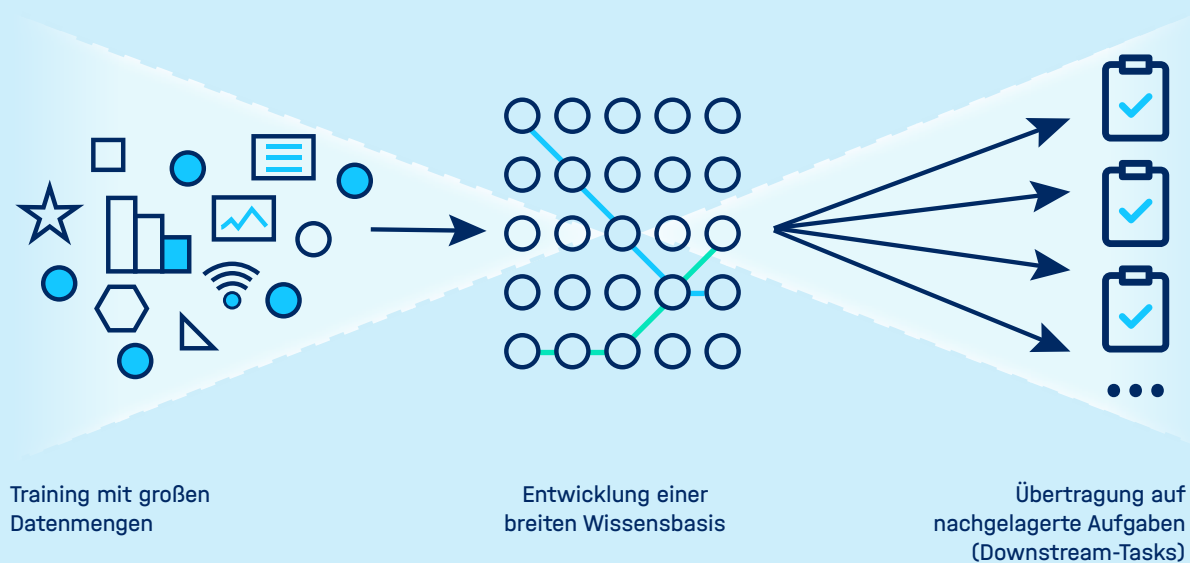
Foundation Models sind komplexe KI-Modelle, die ein kaum zu überschätzendes Potenzial für die Wissenschaft haben. Die Zutaten, die man dafür benötigt, sind riesige Datenmengen, enorme Rechenpower, naturwissenschaftliches Fachwissen sowie Know-how in den Bereichen künstliche Intelligenz und Supercomputing. Mit der Gründung der Helmholtz Foundation Model Initiative leistet Helmholtz Pionierarbeit auf diesem Gebiet.

## Welchen Beitrag leistet Helmholtz?

In der Helmholtz-Gemeinschaft möchten wir Foundation Models trainieren und sie der wissenschaftlichen Community zur Verfügung stellen. Dafür haben sich KI-Expert:innen der einzelnen Helmholtz-Zentren zusammengetan und ein Konzept für die auf drei Jahre angelegte Helmholtz Foundation Model Initiative (HFMI) entworfen.

Ziel der Initiative ist es, voll funktionsfähige Modelle unter Berücksichtigung ethischer und rechtlicher Standards zu entwickeln. Dafür wurden vier Pilotprojekte ausgewählt, an denen Wissenschaftler:innen aus zwölf Helmholtz-Zentren arbeiten. Ein wichtiges Alleinstellungsmerkmal der HFMI ist, dass KI-Expert:innen mit Fachwissenschaftler:innen in den Projekten zusammenarbeiten. Helmholtz stellt damit sicher, dass nicht nur die Modelle einen sinnvollen Beitrag zur Forschung leisten, sondern auch, dass die Ergebnisse der empirischen Überprüfung standhalten. Über einen Zeitraum von drei Jahren erhalten die Projekte eine Förderung in Höhe von 11 Millionen Euro. Weitere 12 Millionen werden in den Ausbau nötiger Infrastruktur investiert.

## Wie werden Foundation Models trainiert?



Foundation Models sind Modelle für maschinelles Lernen, die mit sehr großen Mengen aufbereiteter Daten gefüttert werden. Wichtig ist, dass die Daten gut strukturiert sind und ein relativ unspezifisches Pre-Training stattfindet. Durch dieses Pre-Training sind sie in der Lage, komplexe Zusammenhänge auf der Grundlage erlernter Muster zu verstehen, neue Zusammenhänge zu generieren sowie Prognosen zu erstellen. Die aus dem anfänglichen breiten Training gewonnene Basis kann auf viele nachgelagerte Aufgaben - die sogenannten Downstream-Tasks - übertragen werden.

Das unterscheidet Foundation Models von herkömmlichen KI-Modellen, die für eine ganz bestimmte Aufgabe trainiert werden. Mit einer so entstandenen Wissensbasis - der Foundation - lassen sich eine Reihe von wissenschaftlichen Fragen beantworten, wobei oft nur eine kleine Menge aufgabenspezifischer Daten für eine effektive Leistung erforderlich ist.

Eine **Synergy Unit** forscht zudem an interdisziplinären Fragestellungen, fördert den Wissensaustausch zwischen den einzelnen Projekten und übernimmt übergreifende Aktivitäten. Besonderes Augenmerk liegt dabei auf methodischen Fragestellungen, die das fachliche Verständnis der Funktionsweise solcher Modelle betreffen. Denn gerade das Verstehen, also das Erklären und Reproduzieren von Ergebnissen, schafft Vertrauen in die Ergebnisse der Foundation Models. Deswegen legt Helmholtz einen Schwerpunkt auf unterschiedliche Qualitätssicherungsmaßnahmen. Die geförderten Projekte sollen einen klaren Mehrwert für die Wissenschaft bieten. Darüber hinaus sind sie dem Prinzip der Open Science verpflichtet und stellen ihre finalen Ergebnisse auch der Gesellschaft zur Verfügung - vom Code über die Trainingsdaten bis hin zu den trainierten Modellen.

## Eine Revolution in der Wissenschaft

Foundation Models haben zweifelsfrei das Potenzial, die Art und Weise, wie Wissenschaft betrieben wird, grundlegend zu verändern. Nie zuvor in der Geschichte der Forschung war es möglich, derart große und komplexe Datensätze zu analysieren. Der Einsatz von Foundation Models ist so vielfältig wie die Wissenschaft selbst und ihr Potenzial noch längst nicht ausgeschöpft. In den nächsten Jahren wird es zahlreiche Neuentwicklungen geben.

Die Helmholtz Foundation Model Initiative trägt dazu bei, dass Helmholtz auf diesem Gebiet auch global gesehen eine führende Rolle einnehmen kann. Die vier ausgewählten Pilotprojekte werden mithilfe von künstlicher Intelligenz radiologische Diagnosen zuverlässiger machen, das Verständnis des globalen Kohlenstoffkreislaufs verbessern, Klimamodelle und Wetterprognosen präziser machen und die Entwicklung einer neuen Generation von Photovoltaikmodulen beschleunigen.



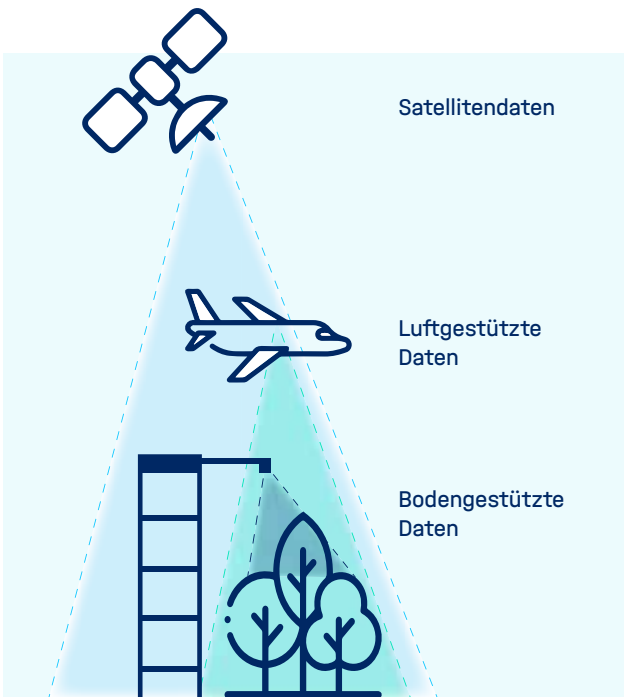
„Wir sind davon überzeugt, dass wir mit Foundation Models die Grenzen der Wissenschaft verschieben können. Helmholtz bringt dafür nicht nur herausragende Talente und umfassende Datensätze aus verschiedenen Forschungsbereichen, sondern auch eine einzigartige Computer-Infrastruktur zusammen.“

Otmar D. Wiestler  
Präsident der Helmholtz-Gemeinschaft

## KI-Grundlagenmodelle für Wissenschaft und Gesellschaft

### 3D-ABC

Berechnung und Visualisierung des globalen Kohlenstoffbudgets  
von Vegetation und Böden



Messung der Struktur tropischer Wälder und ihres Kohlenstoffbestands mit terrestrischen Lasern – Standort Ghana (Bild: Christian Budach/GFZ).

Datenerhebung durch Satelliten, Drohnen und lokale CO<sub>2</sub>-Auffangstationen (Grafik: Helmholtz).

Um die Folgen des weltweiten Klimawandels einzudämmen, benötigen wir fundiertes Wissen über das globale Kohlenstoffbudget, das sich aus CO<sub>2</sub>-Quellen und CO<sub>2</sub>-Speichern, wie Mooren, Wäldern oder Permafrostböden, zusammensetzt. Bislang konnten Forschende schwer beziffern, wie Veränderungen von Landflächen, Vegetationen oder Böden den Kohlenstoffkreislauf beeinflussen, da die Daten zu heterogen und verstreut waren. Das Foundation Model 3D-ABC wird Daten unterschiedlichster Quellen zusammenbringen und modellieren, beispielsweise von Satelliten, Drohnen oder lokalen CO<sub>2</sub>-Auffangstationen. Auf diese Weise können Schlüsselparameter des globalen Kohlenstoffkreislaufs der Vegetation und der Böden mit hoher räumlicher Auflösung erfasst, quantifiziert und charakterisiert werden.

#### Beteiligte Helmholtz-Zentren:

- Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)
- Forschungszentrum Jülich
- Helmholtz-Zentrum Potsdam – Deutsches GeoForschungsZentrum GFZ
- Helmholtz-Zentrum Dresden-Rossendorf (HZDR)
- Helmholtz-Zentrum für Umweltforschung – UFZ
- Deutsches Zentrum für Luft- und Raumfahrt (DLR)

## The Human Radiome Project

### 3D-Radiologiedaten zum fundamental besseren Verständnis von Anatomie und Pathologie

Je genauer Tumore lokalisiert und markiert werden können, desto erfolgreicher fällt eine Strahlentherapie aus. Allerdings stieß die Medizin genau hier bisher an ihre Grenzen, weil es sehr aufwendig war, Ergebnisse verschiedener Bildgebungsverfahren ins Verhältnis zueinander zu setzen und im dreidimensionalen Format auszuspielen. Die genaue Lokalisierung und Markierung von Tumoren ist eins von vielen Verfahren, die das „Human Radiome Project“ im Bereich der medizinischen Bildgebung voraussichtlich verbessern wird. Es vereint die weltweit umfangreichste und vielfältigste Sammlung von 3D-radiologischen Bildern, wie MRT- und CT-Scans, in einem Foundation Model. Dadurch erhalten Forschende tiefe Einblicke in die menschliche Anatomie und Pathologie sowie eine Übersicht über das gesamte Spektrum an radiologischen Informationen. Das „Human Radiome Project“ verbessert nicht nur die personalisierte Medizin, sondern auch die Effizienz der Diagnostik, indem es die Notwendigkeit verringert, komplexe medizinische Bilder manuell zu beschriften.

#### Beteiligte Helmholtz-Zentren:

- Deutsches Krebsforschungszentrum (DKFZ)
- Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE)
- Max Delbrück Center

## HClimRep

### Wechselwirkungen zwischen Atmosphäre, Ozean und Meereis in einem neuartigen Klimamodell erfassen

Was wäre, wenn wir Vorhersagen über das zukünftige Klima noch genauer und viel schneller und effizienter treffen könnten? Könnten wir die Ursachen des Klimawandels dadurch besser bekämpfen und seine Folgen abmildern? Könnten wir die Auswirkungen der Erderwärmung dadurch für alle eindrücklich sichtbar machen? Das Projekt HClimRep hat sich zum Ziel gesetzt, genau solche Fragen zu beantworten. Mit dem Bau eines der ersten Foundation Models für die Klimaforschung, das Daten aus der Atmosphäre, dem Ozean und dem Meereis miteinander kombiniert, entwickeln Forschende eines der präzisesten Wetter- und Klimamodelle der Welt. Dieses Deep-Learning-Modell mit mehreren Milliarden von Parametern wird auf Europas erstem Exascale-Computer trainiert. Dadurch wird er in der Lage sein, komplexe „Was-wäre-wenn“-Experimente sowie andere Modellierungsaufgaben des Ozeans und der Atmosphäre durchzuführen.

#### Beteiligte Helmholtz-Zentren:

- Forschungszentrum Jülich
- Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)
- Karlsruher Institut für Technologie (KIT)
- Helmholtz-Zentrum Hereon

## SOL-AI

### Entwicklung und Optimierung von Photovoltaikmaterialien

Photovoltaik ist eine Schlüsseltechnologie für die Energiewende. Damit sie sich weltweit in ausreichendem Maße verbreitet, müssen innovative Solarzellkonzepte schneller umgesetzt werden. Die Aktivitäten in Forschung und Entwicklung hierzu nehmen rasant zu und führen zu einer Fülle an wissenschaftlichen Publikationen. Doch durch die schiere Menge an Daten stößt die Umsetzung der neuesten Erkenntnisse an ihre Grenzen. Mit SOL-AI wird ein Foundation Model geschaffen, das die Materialinformatik auf diesem Gebiet grundlegend reformieren soll. Es ist in der Lage, die vielfältigen experimentellen Daten und Ergebnisse in der Forschung zu Photovoltaikmaterialien zusammenzuführen, um Innovationen in verschiedenen Bereichen voranzutreiben: angefangen bei der beschleunigten Bauteilentwicklung und -optimierung bis hin zur Entdeckung neuer Solarmaterialien. SOL-AI soll Lösungsansätze entwickeln, die praktische Relevanz für Forschung und Industrie besitzen.

#### Beteiligte Helmholtz-Zentren:

- Forschungszentrum Jülich
- Karlsruher Institut für Technologie (KIT)
- Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)
- Helmholtz-Zentrum Hereon

## Synergy Unit – Foundation Models entwickeln, bereitstellen und vernetzen

Während sich die einzelnen Projekte auf ihre jeweils spezifischen Problemstellungen konzentrieren, fokussiert sich eine Synergy Unit auf übergeordnete Fragestellungen, die für alle teilnehmenden Projekte von Relevanz sind. Beispielsweise beschäftigt sie sich mit Fragen, die die Skalierbarkeit der Modelle oder das Training mit den Datensätzen betreffen. Dabei geht es jedoch nicht einfach nur um den Austausch von Lösungsansätzen, sondern ganz zentral um die Frage, wie die Forschung zu Foundation Models über Disziplingrenzen hinweg möglichst schnell vorangetrieben werden kann. Somit sichert die Synergy Unit eine langfristige Wirkung der Helmholtz Foundation Model Initiative zum Wohle der Allgemeinheit.

#### Beteiligte Helmholtz-Zentren:

- Deutsches Krebsforschungszentrum (DKFZ)
- Helmholtz Munich
- Forschungszentrum Jülich
- Max Delbrück Center





## JUPITER – Der vielleicht stärkste KI-Supercomputer der Welt

Foundation Models sind ohne die Rechenpower von Supercomputern nicht möglich. Helmholtz hat diese Großrechner und stellt sie der Forschungs-Community zur Verfügung. Eine Schlüsselrolle spielt dabei der erste europäische Exascale-Rechner am Forschungszentrum Jülich. Der Supercomputer wird die Grenze von einer Trillion Rechenoperationen pro Sekunde brechen. Diese Supercomputerinfrastruktur wird dabei in der Helmholtz-Gemeinschaft von einer Reihe kleinerer, leistungsstarker Rechner und Rechnerverbünde flankiert, den sogenannten „Helmholtz AI computing resources“ (HAICORE). Mit der HAICORE-Infrastruktur erweitert Helmholtz strategisch seinen Supercomputing-Bereich und bietet damit nicht nur eine erstklassige Forschungsumgebung, sondern stellt auch sicher, dass die Supercomputer-Ressourcen nur dann eingesetzt werden, wenn ein Erfolg garantiert werden kann.

The background features a complex network of thin, curved lines in shades of blue, cyan, magenta, and yellow, creating a sense of dynamic movement and connectivity. Interspersed among these lines are various binary code sequences (0s and 1s) in matching colors, some appearing as if they are floating or being processed. The overall aesthetic is futuristic and data-oriented.

Herausgegeben von:

Helmholtz-Gemeinschaft  
Deutscher Forschungszentren e.V.  
Anna-Louisa-Karsch-Str. 2  
10178 Berlin  
[www.helmholtz.de](http://www.helmholtz.de)

Kontakt:

Florian Grötsch  
Manager Information and  
Data Science  
Helmholtz-Gemeinschaft  
Tel. +49 30 206 329-669  
[florian.groetsch@helmholtz.de](mailto:florian.groetsch@helmholtz.de)

Mehr Informationen unter:

[www.helmholtz.de/ki](http://www.helmholtz.de/ki)

Stand:

Juni 2024