HELMHOLTZ SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN

HELMHOLTZ ARTIFICIAL INTELLIGENCE COOPERATION UNIT (HAICU)

INHALTSVERZEICHNIS

EINLEITUNG IN INFORMATION & DATA SCIENCE IN DER HELMHOLTZ-GEMEINSCHAFT

Information & Data Science – Eine globale Herausforderung

Die digitale Transformation ist eine der größten Herausforderungen für Wissenschaft, Wirtschaft und Gesellschaft zu Beginn des 21. Jahrhunderts. Sie bietet enorme Chancen in nahezu allen Bereichen des Lebens – innovative Formen von Arbeit und Zusammenleben, völlig neuartige Plattformen für Handel und Wissenschaft, ungeahnte Möglichkeiten für die Medizin, revolutionäre Ansätze für Netzwerke der Energieversorgung, zukunftsweisende Verfahren für Klimaschutz und vieles mehr.

Im Kern der digitalen Transformation steht die Wertschöpfungs- und Erkenntniskette „von Daten zu Wissen zu Innovation". Dies ist auch die entscheidende Herausforderung in allen wissenschaftlichen Disziplinen, welche die Grundlagen für den gesellschaftlichen Fortschritt der Zukunft bereiten.

Damit Deutschland diese Entwicklung mitgestalten und daran partizipieren kann, sind ambitionierte, mutige und zukunftsweisende Schritte erforderlich. Die Verbindung von Informatik, Mathematik, Statistik, Sensortechnologie, Simulation und datenintensivem Rechnen mit anspruchsvollen Anwendungsfeldern aus dem breiten Spektrum der Natur- und Ingenieurswissenschaften, der Medizin sowie den Geistes- und Sozialwissenschaften wird ein dynamischer Motor des Innovations- und Forschungsstandortes Deutschland sein. Diese innovativen Verbindungen werden völlig neue wissenschaftliche Erkenntnisse mit erheblichen Mehrwerten für Wirtschaft und Gesellschaft hervorbringen.

Deutschland hat im Feld der Hochtechnologien mit großem transformativem Potenzial eine sehr gute Ausgangsposition im weltweiten Vergleich. Um diese Position zu halten und auszubauen, sind neben gesamtgesellschaftlichen Anstrengungen auch völlig neuartige Ansätze im Wissenschaftssystem erforderlich.

Information & Data Science in der Helmholtz-Gemeinschaft

Die Helmholtz-Gemeinschaft Deutscher Forschungszentren leistet Beiträge zur Lösung großer und drängender Fragen von Gesellschaft, Wissenschaft und Wirtschaft durch interdisziplinäre, wissenschaftliche Spitzenleistungen in sechs Forschungsbereichen: Energie, Erde und Umwelt, Gesundheit, Schlüsseltechnologien, Materie sowie Luftfahrt, Raumfahrt und Verkehr. Sie ist mit über 39.000 Mitarbeiterinnen und Mitarbeiter in 18 Forschungszentren die größte Wissenschaftsorganisation Deutschlands.

Als Forschungsorganisation, die sich zum Ziel gesetzt hat große und aktuelle gesellschaftliche Herausforderungen zu adressieren, hat die Helmholtz-Gemeinschaft in den letzten Jahrzehnten ein enormes Kompetenzportfolio im Bereich Information & Data Science aufgebaut: auf Gebieten wie Informationsverarbeitung, Big Data, Datenalyse, Simulation, Modellierung, Bioinformatik, bildgebenden Verfahren, Forschungsdaten-Management, High Performance Computing, Robotik, technischen sowie biologischen Informationssystemen und vielen weiteren zukunftsweisenden Technologien. Wie kaum eine andere Forschungsorganisation verfügt sie über einen exponentiell wachsenden Schatz von Big Data.

Die Helmholtz-Zentren und Forschungsbereiche haben auch eine herausragende Ausgangsposition für eine erfolgreiche, synergetische Verbindung der Kompetenzen: als Betreiber großer Forschungsinfrastrukturen (beispielsweise Satellitenmissionen oder Großanlagen der Kern- und Teilchenphysik), als Anwender von Supercomputing der neuesten Generation sowie als Kompetenzträger für komplexe Simulationen (beispielsweise umfangreicher Erd- und Klimamodelle, virtuelles Materialdesign und Systembiologie). Sie sind daher auch schon seit Langem mit dem gesamten Data-Lifecycle vertraut: der Forschungsplanung, dem Erheben, der Handhabung und Pflege, dem Analysieren, dem Auswerten und der Nutzbarmachung sehr großer und komplexer Datenmengen. An allen Standorten und in allen Einzeldisziplinen gibt es teilweise weltweit führende Ansätze und herausragende Methodenkompetenz – aber gerade auch in ihrem Zusammenspiel ergeben sich ungeahnte Möglichkeiten.

Darauf aufbauend und darüber hinausgehend verstärkt die Helmholtz-Gemeinschaft im hochaktuellen Themenfeld Information & Data Science die eigene Kompetenz, schafft Synergien in der Forschungslandschaft, greift die Entwicklungen im nationalen, europäischen und internationalen Kontext auf und setzt neue disruptive Ansätze um. Neben der Stärkung der Einzeldisziplinen mit modernsten daten- und informationswissenschaftlichen Methoden ist es erklärtes Ziel, das Thema Information & Data Science auf Gemeinschaftsebene disziplinübergreifend voranzutreiben.

Die Helmholtz-Gemeinschaft adressiert das komplexe Themenfeld Information & Data Science auf allen Ebenen:

- Es hat hohe Priorität in der Agenda des Präsidenten.
- Helmholtz-Zentren und Forschungsbereiche berücksichtigen in ihren jeweiligen Strategien das große Potenzial dieses Themenfeldes.
- Im Rahmen der Neuausrichtung der Forschungsbereiche wird der bisherige Forschungsbereich Schlüsseltechnologien in einen neuen Forschungsbereich Information weiterentwickelt.
- Die Gemeinschaft errichtete mehrere neue Institute zu Simulations- und Datentechnologien und Cybersicherheit und plant aktuell die Aufnahme eines neuen Helmholtz-Zentrums für Informationssicherheit am Standort Saarbrücken.
- Wissenschaftlicher Nachwuchs wird in diesem Bereich auf neuen Wegen und in großer Zahl in einem neuen Netzwerk regionaler Helmholtz Information & Data Science Schools (HIDSS) ausgebildet.
- Der von der Helmholtz-Gemeinschaft initiierte Helmholtz-Inkubator Information & Data Science potenziert die einzelnen Initiativen als ein neuartiger, gemeinschaftsweiter Think-Tank und Zukunftsmotor.
- Durch Ihre intensive Interaktion mit nationalen und internationalen Partnern leistet Helmholtz auch entscheidende Beiträge zu Initiativen wie der Nationalen Forschungsdaten Infrastruktur (NFDI) und zu internationalen Allianzen im Forschungsdatenmanagement

Der Helmholtz-Inkubator Information & Data Science

Der Helmholtz-Inkubator Information & Data Science wurde im Juni 2016 vom Präsidenten der Helmholtz-Gemeinschaft ins Leben gerufen, um die vielfältige, dezentrale Expertise der Gemeinschaft im weiten Themenfeld Information & Data Science intelligent zusammenzuführen und zu potenzieren.

Dazu haben alle Helmholtz-Zentren je zwei hochkarätige Wissenschaftlerinnen und Wissenschaftler in den Helmholtz-Inkubator entsandt. Diese 36 Fachleute vertreten zusammen eine enorme fachliche Breite und decken viele innovative Kompetenzen ab. Unterstützt werden sie durch ausgewiesene Expertinnen und Experten aus forschenden Unternehmen und namhaften Forschungseinrichtungen sowie mehreren Beratungsunternehmen. Begleitet wird der Helmholtz-Inkubator von der Geschäftsstelle der Helmholtz-Gemeinschaft.

Der Helmholtz-Inkubator integriert bestehende, zukunftsweisende Initiativen der Helmholtz-Gemeinschaft durch einen gemeinschaftsweiten Bottom-up-Prozess. Die regelmäßige Zusammenführung und Verdichtung der Expertise der Helmholtz-Zentren ermöglicht die visionäre Gestaltung des Themas Information & Data Science über die Grenzen von Zentren und Forschungsbereichen hinaus. Dabei geht der Helmholtz-Inkubator völlig neue Wege, um das Zukunftsfeld dynamisch, umfangreich und fachübergreifend durch Setzung strategischer Schwerpunkte zu gestalten.

Der Helmholtz-Inkubator verfolgt derzeit folgende Ziele:

- die regelmäßige Interaktion kreativer Köpfe aus der gesamten Gemeinschaft,
- die Schaffung von Grundlagen für innovative, interdisziplinäre Netzwerke und Ansätze,
- die Identifizierung zukunftsweisender Themenfelder und disruptiver Pilotprojekte,
- die Planung und Begleitung von langfristig angelegten und gemeinschaftsweiten Plattformen.

Der Helmholtz-Inkubator hat innerhalb der Helmholtz-Gemeinschaft eine hohe Dynamik entfaltet, zahlreiche Impulse mit großer thematischer Breite formuliert, vielfältigen Austausch zwischen Digitalisierungsexpertinnen und -experten aus allen Domänen ermöglicht und so einen tiefgehenden Strategieprozess eingeleitet, der in dieser Form und in diesem Umfang einmalig im deutschen Wissenschaftssystem ist.

Über 150 beteiligte Wissenschaftlerinnen und Wissenschaftler, davon 36 Vertreterinnen und Vertreter des Helmholtz-Inkubators, sowie 10 externe Beraterinnen und Berater (u.a. IBM, SAP, Trumpf, Blue Yonder, Gauß-Allianz) haben sich in den letzten zwei Jahren im Rahmen von über 35 Workshops und AG-Treffen diesem großen Themenkomplex gewidmet. Auch die Vorstände der Helmholtz-Gemeinschaft haben sich in einem dedizierten Workshop mit den strategischen Überlegungen des Helmholtz-Inkubators auseinandergesetzt. Nach Schätzungen der Geschäftsstelle sind bis heute über 25.000 Personenstunden Arbeit in diesen Prozess geflossen.

Der bisherige Prozess

Der Helmholtz-Inkubator trat im Oktober 2016 zu seinem ersten, zweitägigen Workshop zusammen und diskutierte das Thema Information & Data Science offen und uneingeschränkt; jedes Zentrum lieferte Impulse aus seiner jeweiligen Sicht.

Obwohl zu Beginn des Prozesses die Schaffung von Konsortien zur Bearbeitung von Pilotprojekten im Vordergrund stand, formulierten die Expertinnen und Experten die klare Empfehlung, dass die Entwicklung von dauerhaften, strukturellen Aktivitäten im Themenbereich Information & Data Science auf einigen Teilgebieten auf Gemeinschaftsebene notwendig sei – zusätzlich zu neuen Impulsen durch Pilotprojekte.

Die Vertreterinnen und Vertreter erörterten die zentralen Herausforderungen der Helmholtz-Gemeinschaft auf Ebene der gesamten Gemeinschaft. Die identifizierten Themen wurden umfassend mit allgemeinen forschungspolitischen Initiativen der Allianz-Organisationen und Vorstellungen der Zuwendungsgeber sowie internationalen Entwicklungen abgeglichen. Zusätzlich wurde der Helmholtz-Inkubator dabei von führenden externen Expertinnen und Experten im Thema Information & Data Science beraten.

Die Vorschläge zu gemeinschaftsweit zu bearbeitenden Themenfeldern wurden im Frühjahr 2017 in mehreren fokussierten Arbeitsgruppen des Helmholtz-Inkubators konkretisiert. Die Arbeitsgruppen setzen sich aus den Inkubator-Mitgliedern zusammen, die sich spezifisch in diesem Thema einbringen wollen; weitere mit den Themen befasste Expertinnen und Experten der Helmholtz-Gemeinschaft nahmen beratend an den Sitzungen der Arbeitsgruppen teil. Die Ergebnisse der Arbeitsgruppen wurden in einem Inkubator-Workshop mit allen Vertreterinnen und Vertretern des Helmholtz-Inkubators und externen Expertinnen und Experten im Mai 2017 diskutiert und weiter geschärft.

Aus den diskutierten Themenvorschlägen wurden diejenigen Themen weiter verfolgt, von denen gemeinschaftsweite Mehrwerte und ein erhebliches Entwicklungspotenzial erwartet werden konnten. Der Helmholtz-Inkubator identifizierte so fünf Themenkomplexe, die er der Helmholtz-Gemeinschaft zur langfristigen Verfolgung durch die Etablierung von fünf dedizierten Plattformen vorschlug.

Die Vorstände der Helmholtz-Gemeinschaft haben sich in einer außerordentlichen, fachlichen Sitzung im September 2017 eingehend mit diesen fünf Plattform-Ansätzen befasst. Zu allen Themen wurden inhaltliche und strukturelle Leitplanken für eine detaillierte Konzepterstellung formuliert. Auf der Mitgliederversammlung haben sie diese zur detaillierten Ausarbeitung durch den Helmholtz-Inkubator und seiner Arbeitsgruppen empfohlen, um eine potenzielle Umsetzung ab September 2018 vorzubereiten. Die Mitgliederversammlung forderte die Geschäftsstelle dazu auf, die AGs bei der Ausgestaltung von Governance und Finanzierungsfragen zu unterstützen.

Die Ergebnisse des Vorstandsworkshop Information & Data Science und der Mitgliederversammlung wurden anschließend dem Senat der Helmholtz-Gemeinschaft übermittelt.

Die Arbeitsgruppen des Helmholtz-Inkubators haben darauf folgend ab Oktober 2017 die weitere Konkretisierung der Themen vorgenommen. Die Inkubator-Vertreter der Zentren wurden nochmals aufgefordert, spezialisierte Fachleute in die jeweiligen AGs zu entsenden; so wurde themenspezifisch ein noch höheres Maß an Expertise aufgebaut.

Die Arbeitsgruppen haben zur weiteren Ausarbeitung der Konzepte Berichterstatter bestimmt. Zur Unterstützung der Arbeitsgruppen und Berichterstatter wurden pro Themenfeld zwei Referenten eingestellt; hierfür stellte der Präsident Mittel aus dem Impuls- und Vernetzungsfonds bereit. Die AGs bildeten folgende Teams:

- <u>Arbeitsgruppe wissenschaftlicher Nachwuchs:</u> vertreten durch die Berichterstatter Achim Streit (KIT) und Uwe Konrad (HZDR), 28 Mitarbeitende, unterstützt durch die Projektreferentin Susan Trinitz. Erarbeitetes Konzept: Helmholtz Information & Data Science Academy (HIDA).
- <u>Arbeitsgruppe Mehrwerte aus Forschungsdaten durch Metadaten:</u> vertreten durch die Berichterstatter Rainer Stotzka (KIT), Kirsten Elger (GFZ) und Frank Ückert (DKFZ), 27 Mitarbeitende, unterstützt durch die Projektreferentinnen Nanette Rißler-Pipka und Rumyana Proynova. Erarbeitetes Konzept: Helmholtz Metadata Center (HMC).
- <u>Arbeitsgruppe Basistechnologien und grundlegende Dienste:</u> vertreten durch die Berichterstatter Ants Finke (HZB), Volker Gülzow (DESY) und Uwe Konrad (HZDR), 36 Mitarbeitende, unterstützt durch die Projektreferenten Knut Sander und Tobias Frust. Erarbeitetes Konzept: Helmholtz Infrastructure for Federated ICT Services (HIFIS).
- <u>Arbeitsgruppe bildgebende Verfahren (Imaging):</u> vertreten durch die Berichterstatter Christian Schroer (DESY) und Wolfgang zu Castell (HMGU), 44 Mitarbeitende, unterstützt durch die Projektreferenten Alexander Pichler und Murali Sukumaran. Erarbeitetes Konzept: Helmholtz Imaging Platform (HIP).
- <u>Arbeitsgruppe Künstliche Intelligenz und Maschinelles Lernen:</u> vertreten durch die Berichterstatter Fabian Theis (HMGU) und Katrin Amunts (FZJ), 45 Mitarbeitende, unterstützt durch die Projektreferentinnen Susanne Wenzel und Angela Jurik-Zeiller. Erarbeitetes Konzept: Helmholtz Artificial Intelligence Cooperation Unit (HAICU).

Im März 2018 fand der vierte Inkubator-Workshop statt. Hier wurden die bis zu diesem Zeitpunkt erarbeiteten Konzeptentwürfe präsentiert und weiterentwickelt. Dabei wurden die Inkubator-Vertreter erneut von ausgewiesenen externen Expertinnen und Experten unterstützt. Die Ergebnisse des Inkubator-Workshops wurden im Anschluss in den jeweiligen Arbeitsgruppen umgesetzt und die Konzepte von den Arbeitsgruppen finalisiert. Flankierend fanden individuelle Gespräche aller Berichterstatter mit der Geschäftsführerin der Helmholtz-Geschäftsstelle zu Fragen der Governance und Finanzierung statt.

Nach Fertigstellung der Konzepte durch die Arbeitsgruppen und Berichterstatter wurden diese im Juli 2018 allen offiziellen Inkubator-Vertretern übergeben. Der Helmholtz-Inkubator empfiehlt den Vorständen der Helmholtz-Gemeinschaft die vorliegenden fünf Konzepte zur nachhaltigen und gemeinschaftsweiten Umsetzung.

Die Vorstände der Helmholtz-Gemeinschaft können nun im Rahmen der Mitgliederversammlung im September 2018 über die Umsetzung und Finanzierung jedes Plattform-Konzepts befinden. Die Vorstände bestimmen außerdem für jedes genehmigte Konzept, welche Helmholtz-Zentren die Plattformen zum Nutzen der gesamten Helmholtz-Gemeinschaft tragen sollen oder definieren ein Verfahren zur Verortung der Plattformen.

Die wichtigsten Prozessschritte der vergangenen zwei Jahre sind in Abbildung E.1 schematisch dargestellt.

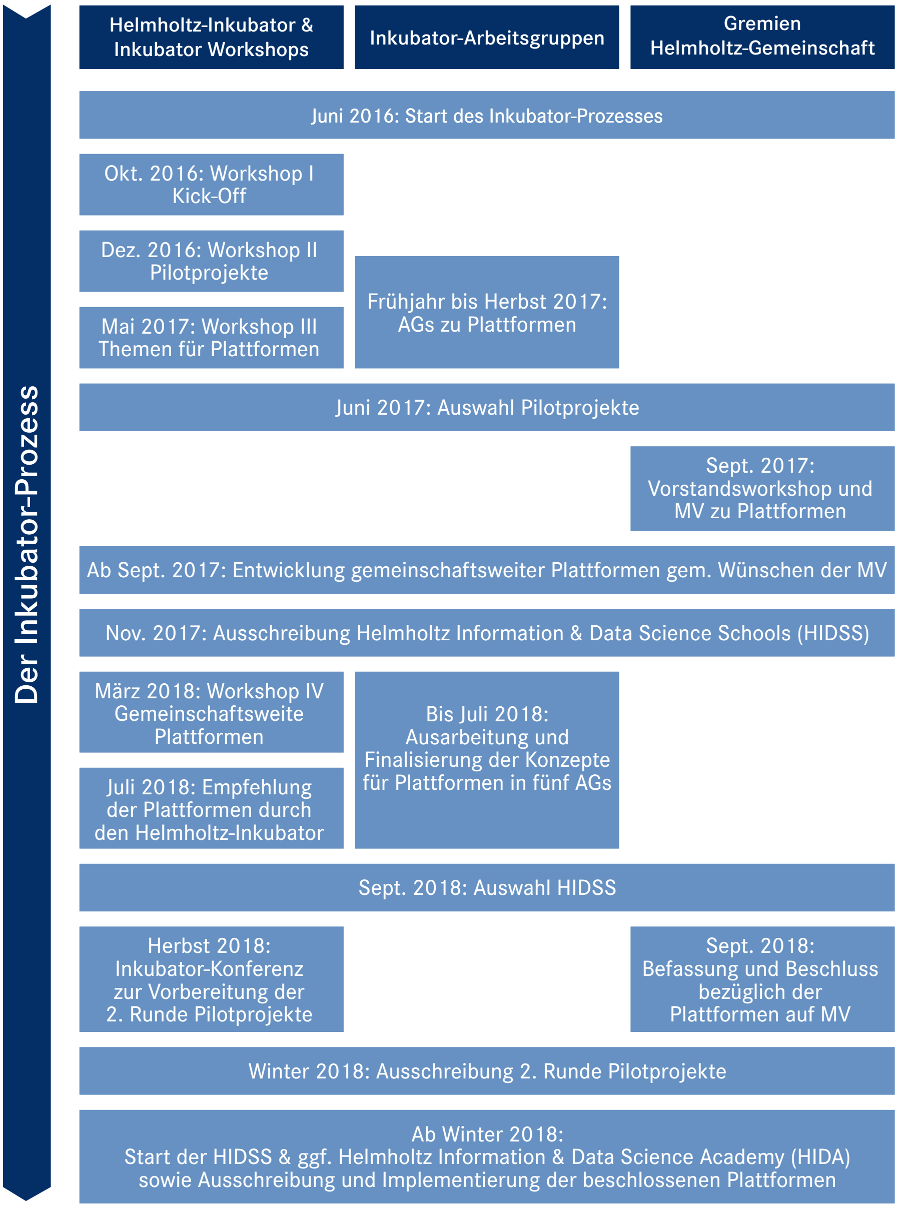

Abbildung E.1: Die wichtigsten Schritte des Inkubator-Prozesses der vergangenen zwei Jahre.

1 SUMMARY IN GERMAN AND ENGLISH

1.1 ZUSAMMENFASSUNG

Der Helmholtz-Inkubator plant, im Bereich 'Information and Data Science' interdisziplinäre Plattformen für innovative Forschung zwischen allen Helmholtz-Zentren in Deutschland aufzubauen: Dabei wurden die disruptiven Schlüsseltechnologien Künstliche Intelligenz (KI), Maschinelles Lernen (ML) und Deep Learning als Bereiche identifiziert, die eine zentrale Rolle bei der Lösung zukunftsweisender Aufgaben in der gesamten Helmholtz-Gemeinschaft haben. In diesem Papier wird ein Konzept für eine neue, dedizierte Einheit vorgeschlagen — *die Helmholtz Artificial Intelligence Cooperation Unit (HAICU).*

HAICU wird Verfahren der KI für Anwendungen über alle Helmholtz-Zentren hinweg entwickeln, umsetzen und verbreiten. KI bezieht sich dabei auf rechnergestützte Methoden, die die Automatisierung intelligenten Verhaltens und das Maschinelle Lernen adressieren. Obwohl die Anfänge der KI schon auf die 50er Jahren zurückgehen, hat KI insbesondere im letzten Jahrzehnt, bedingt durch Fortschritte im Bereich des Computings, der zur Verfügung stehenden Daten und der Algorithmen eine Reihe von Bereichen revolutioniert, wie bspw. Spracherkennung oder Computer Vision. KI und die zugrundeliegenden Technologien werden die meisten, wenn nicht alle Disziplinen sowohl in der Industrie als auch in der Wissenschaft maßgeblich und nachhaltig beeinflussen. KI wurde daher als einer der wichtigsten Investitionsbereiche in vielen Ländern identifiziert.

Die Helmholtz-Gemeinschaft strebt mit HAICU eine internationale Führungsposition in der angewandten KI an, indem sie KI-basierte Methodiken mit Helmholtz' einzigartigen Forschungsfragen und Datensätzen kombiniert. HAICU wird (i) neue KI-Methoden entwickeln und implementieren, die durch konkrete Forschungsfragen motiviert sind, und diese in Anwendungsfällen über mehrere Helmholtz-Zentren hinweg verwenden, (ii) KI-Methoden für Nutzer in den verschiedenen Zentren anwendbar machen, neue Methoden in diesem Bereich verbreiten und ausbilden und (iii) Ähnlichkeiten zwischen Anwendungen über verschiedene Forschungsbereiche hinweg identifizieren und Synergien entwickeln. Dabei baut HAICU auf die hohe Qualität und den großen Umfang der Daten in der Helmholtz-Gemeinschaft, die sich durch ihre wissenschaftliche Einzigartigkeit, Qualität und umfangreiche Dokumentation auszeichnen. HAICU wird in den verschiedenen Forschungsfeldern und Zentren als Wegbereiter für angewandte KI-Forschungsaktivitäten praktische Wirkung entfalten und gleichzeitig methodische Stärken allen Helmholtz-Zentren zugutekommen lassen.

Wir schlagen vor, HAICU als „Rad-Speichen-Modell" zu strukturieren, mit einer international sichtbaren, zentralen Einheit und einer starken Peripherie aus verschiedenen Helmholtz-Zentren. HAICUs zentrale Einheit (HAICU Central) wird fünf Forschungsgruppen beherbergen, die sich mit den für die Helmholtz-Gemeinschaft relevanten interdisziplinären Aspekten der KI befassen, sowie einem Zentren-übergreifenden wissenschaftlichen High Level Support Team (HLST) und dem Management. Die peripheren Einheiten, angesiedelt an mehreren Helmholtz-Zentren, werden aus fünf weiteren HLSTs bestehen, die den Transfer der Expertise in den unmittelbaren Forschungsbereich der jeweiligen Zentren und auch in die anderen Zentren sicherstellen, sowie HAICU-Nachwuchsgruppen, die für Flexibilität in HAICUs wissenschaftlicher Ausrichtung und Strategie und für die Förderung von Nachwuchsforschern sorgen. Sie werden dicht an den Helmholtz-relevanten Anwendungen forschen, Methoden für den Einsatz in die Domänen übersetzen und dafür eng mit HAICU Central interagieren.

Die HLSTs werden den Anwendern einen hochgradigen Support bieten, der es ermöglicht, die Anstrengungen flexibel, dynamisch und transparent auf die Bedürfnisse der verschiedenen Partner abzustimmen. Jedes Helmholtz-Zentrum kann sich für wissenschaftliche HAICU-Projekte verschiedener Größe bewerben: kleinere

Projekte (max. Laufzeit 6 Monate), die über die HSLTs durchgeführt werden, mittlere Projekte für die Dauer von bis zu drei Jahren und langfristige Forschungsprojekte, die durch die Vergabe der HAICU Young Investigator Groups umgesetzt werden und nach fünf Jahren wieder neu ausgerichtet werden können.

HAICU wird ein Helmholtz-weites Netzwerk für angewandte KI aufbauen. Ein internationales Besucherprogramm, international sichtbare Workshops und Hackathons sollen Methoden innerhalb der Helmholtz-Gemeinschaft und darüber hinaus verbreiten.

HAICU wird die Helmholtz-Gemeinschaft im Bereich der angewandten KI erheblich bereichern, Wissenschaftler aller Zentren vernetzen, sie in der Bearbeitung ihrer konkreten Forschungsfragen aktiv unterstützen und damit transdisziplinäre und disruptive Forschungsergebnisse erzeugen.

1.2 SUMMARY

The ‘Helmholtz Incubator Information & Data Science’ initiative aims to build interdisciplinary platforms for innovative research between all Helmholtz Centers in Germany. This initiative identified key areas of development, amongst others Artificial Intelligence (AI), including Machine Learning (ML) and Deep Learning (DL). In this proposal, we summarize the concept proposed by the AI working group of the Helmholtz Incubator Information & Data Science. We propose a concept for a new, dedicated unit that compiles, develops, fosters and promotes applied AI and related methods nationwide across all Helmholtz Centers.

AI refers to the use of computation to perform tasks that normally require human intelligence. Although having been emerged in the 50s already, AI has revolutionized a series of fields such as speech recognition or computer vision in the last decade. It is broadly agreed that both in industry and science, AI and the underlying technologies will impact most if not all disciplines and has therefore been identified as an area of investment across many sites.

The *‘Helmholtz Artificial Intelligence Cooperation Unit’ (HAICU)* aims to reach an international leadership position in applied AI by combining AI-based analytics with Helmholtz’ unique research questions and data sets. HAICU will (i) develop and implement novel AI methods, including contributions to the theoretical basis of AI, motivated by concrete research questions and apply them in use cases across multiple Helmholtz Centers, (ii) provide support for applied AI activities for users in the various Helmholtz Centers and disseminate emerging methods in the field, and (iii) identify and leverage similarities between applications across different topics. In this regard, HAICU builds on the Helmholtz Association´s high quality and large data sets, characterized by their scientific uniqueness, quality and rich documentation. HAICU will have practical impact in the research fields and Helmholtz Centers as an enabler of applied AI-based research activities, with the development and distribution of AI methods across all Helmholtz Centers.

We suggest to structure HAICU in a wheel-and-spoke model, with an interdisciplinary international visible central unit and a strong applied periphery located at various Helmholtz Centers. HAICU will host a central unit, consisting of five research groups, which will address interdisciplinary aspects of AI relevant to the Helmholtz Association, as well as management and a cross-centre science support team. Peripheral units will consist of five High Level Support Teams (HLST), which will ensure translation of expertise into the immediate research domain of the respective Centers, and five HAICU Young Investigator Groups which will provide new scientific topics and strategies, and promote young talents. The HLSTs will provide high-level support to AI users considering their different levels of expertise, which will allow to scale efforts according to the needs of the different partners in a flexible, dynamic, and transparent way. Both HLSTs and Young Investigator Groups will get close to applications, translate methods and interact with HAICU Central.

Each Helmholtz Center is invited to apply for scientific HAICU projects of different scopes. Small projects (max. runtime 6 months), implemented via the HSLTs, medium projects for the duration of up to three years, and long-term research projects that will be implemented by awarding the HAICU Young Investigator Groups, which will be re-tendered again after five years.

HAICU aims to implement a Helmholtz-wide network for applied AI. An international visitor program, internationally visible workshops, and hackathons will be implemented, intended to disseminate methods across the Helmholtz Association and beyond.

To summarize, HAICU will significantly enrich Helmholtz in terms of applied AI. It will bring together scientists from all centers, provide support in their needs and thereby will foster transdisciplinary and ground-breaking research.

2 INTRODUCTION

2.1 BACKGROUND

Artificial Intelligence (AI) has established itself as a highly relevant disruptive technology over the last years, both in science as well as industry (Hassabis 2017; Ledford 2017; Russell 2017; Silver et al. 2017). The European Commission has recognized AI´s significance and will boost its investments in artificial intelligence by about 70 percent to 1.5 billion € by 2020 to catch up with Asia and the United States, which are currently each investing at least three times more than Europe. In 2016 for example, European private investments in AI totaled around 2.4-3.2 billion €, compared to Asia's almost 10 billion € and America's 18 billion € (European Commission 2018). The European Commission intends total private and public investment in AI to reach at least 20 billion € by the end of 2020 to ensure this important field retains competitiveness and does not face brain drain (Europäische Kommission 2018). Several European countries have also set up own programs and funding: France, for example, will invest 1.5 billion € in AI research and innovation by 2022, as part of a national strategy to create networks of AI research institutes in four or five places across the country (France24 2018). In the U.K., a Parliament report from April 2018 urged the government to help the country become an AI leader (Authority of the House of Lords 2018). Also, in April 2018, 25 European countries signed a statement of principles, in which governments agreed to work together on AI and to consider AI research funding 'as a matter of priority' (European Commission and Council 2018).

The German Government also agreed in its coalition agreement to push developments in AI (Bundesregierung 2018a). Federal Minister of Education and Research Anja Karliczek decided to support four competence centers for Machine Learning (ML) in Berlin, Dortmund/St. Augustin, München and Tübingen with about 30 million € (BMBF 2018). Most recently the German Government announced to publish the national strategy for AI by November 2018. For that the German Government passed a key issues paper (Bundesregierung 2018b) proposing to attract and retain AI experts in Germany, increased research and innovation funding, and expand infrastructure expansion.

A range of national and international universities as well as research facilities set up data science programs, including the Deutsche Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Max Planck Institute for Intelligent Systems, University of Toronto (Vector Institute 2018), and Universities of New York, Washington and Berkley (Moore-Sloan Data Science Environment 2018) and MIT Computer Science & Artificial Intelligence Lab (CSAIL). Moreover, the Max Planck Society and the Helmholtz Association offer a data science graduate school to attract excellent students and researchers in this field.

The Helmholtz Association, as the biggest science organization in Germany, with its broad range of research fields, unique research questions, amazing collections of research data, and centers all around the country has now the unique opportunity to act as a pioneer in the field of AI by combining different domain-specific experts with method-oriented experts. This will provide a truly unifying context, which will bring together under the same structure ground-breaking new methodology developments with a large spectrum of application fields.

2.2 MOTIVATION FOR AN AI UNIT IMPLEMENTATION IN THE HELMHOLTZ ASSOCIATION

Based on the large amount of unique data acquired from the Helmholtz Association´s large scale facilities, its outstanding research infrastructure, and the already existing domain-specific but distributed methodological strengths in AI, we aim to build an interdisciplinary platform for innovative research between all Helmholtz Centers – the *Helmholtz Artificial Intelligence Cooperation Unit (HAICU)*. HAICU will build on the current AI strengths in the Helmholtz Association, to foster interactions between the centers as well as with strategic academic partners, and push future topics of information and data science in order to enable the Helmholtz Association to take a pioneering role in the field of information and data science.

We believe that HAICU will help shaping the Helmholtz Association to become a driver for applied AI, to cope with big data and research problems through a substantial and coordinated effort of all centers. Successfully implemented, HAICU will strengthen the Helmholtz Association´s national and international visibility in applied AI and provide opportunities for cooperations with new external partners from universities, research institutions, but also industry. For this, HAICU will attract excellent experts, nationally and internationally, to build up new junior and senior research groups inside the Helmholtz Association. It will create a network of German applied AI research with attractive partners worldwide to develop future-oriented, interdisciplinary research fields. This way HAICU will become an excellent scientific network for applied AI working across disciplines that are relevant in the Helmholtz Association.

2.3 AI STATE OF THE ART IN THE HELMHOLTZ ASSOCIATION AND POTENTIAL APPLICATION AREAS

The Helmholtz Association identified the rapidly developing field of digital data processing and complex data analysis as one of the biggest challenges for the science system, and, thus, for the Helmholtz Association.

The Helmholtz Incubator Information & Data Science as an innovation platform across all research fields and Helmholtz Centers has already funded pilot projects covering ground-breaking research areas:

The Helmholtz Analytics Framework (HAF) aims on the development of common tools and methods for data analysis, e.g., the predictions of solar power, approximation of aircraft models, data mining in volumetric images, neuroscience, and structural biology. It implements a successful cooperation across centers, namely DLR, DESY, DKFZ, FZJ, KIT and HMGU (HAF 2018; FZJ 2017).

Sparse2Big connects HMGU, DZNE, DKFZ, FZJ, GFZ/HPI, HZI, MDC and UFZ. It aims on imputation and fusion for massive sparse data and the development of novel computational approaches initially for the efficient analysis of single cell genomics data and remote sensing. Afterwards, the developed ideas will be translated to other research fields.

Reduced Complexity Models (RedMod) connects the research fields Earth and Environment (HZG, GEOMAR, GFZ, UFZ), Energy (IPP, FZJ), Health (HZI) and Aeronautics, Space and Transport (DLR). The project brings together existing but distributed expertise at the involved centers to address three key challenges for the transformation from data to knowledge, these are (i) uncertainty quantification, (ii) the development of fast surrogate models and (iii) the identification of key parameters and dependencies in complex systems. The project focuses on the development of transferable techniques and, among others, also includes ML.

Further Helmholtz Incubator Information & Data Science pilot projects are Automated Scientific Discovery (IPP, DLR), and Imaging at the Limit (DESY, HMGU, DZNE, HZDR). All of them enforce transdisciplinary cooperations among multiple research fields and Helmholtz Centers.

Within the frame of the Helmholtz Association´s cross-program activities, the Earth System Knowledge Platform (ESKP) GFZ, HZG, AWI, GEOMAR, DLR, FZJ, KIT and UFZ collaborate in providing fast and comprehensive, but also well-grounded background information in cases of events, such as natural hazards or environmental risks, or slowly progressive processes such as the climate change (ESKP 2018). Another Helmholtz initiative that HAICU will naturally link to is the Digital Earth Project. It connects all eight centers within the research field Earth and Environment and aims at a better integration of data and knowledge from the different earth science disciplines and earth compartments.

In preparation of the presented concept for HAICU, the project referees recently visited researchers from almost all Helmholtz Centers, except two centers, of which one was contacted by phone. Furthermore, a video conference, a physical meeting at FZJ, bilateral meetings, and further email communication and phone calls ensured a close collaboration between a broad range of Helmholtz Association data science researchers from all domains and (almost) all centers. As an outcome of these intensive discussions and visits, we are able to provide a broad overview of existing expertise in the field of AI and the needs the researchers expressed. Table 1 and 2 (Sections 9.1 and 9.2) summarize expertise and needs (not necessarily comprehensive). Table 1 also lists the high impact papers of Helmholtz Association researchers that have been published in specific topics.

In sum, there is a strong expertise in AI, ML and DL over all research domains and Helmholtz Centers (Table 1). Especially in classical ML and Data Mining methods, a broad experience over all research fields can be proven by a large amount of high impact papers. Also, in DL, which is comparably new in the field, expertise is already available in almost all Helmholtz Centers. This represents an excellent starting point for knowledge transfer and transdisciplinary collaboration. As expected, the fields of simulation and model-based approaches are also covered in the Helmholtz Association. Finally, also a recent research task in AI, coupling data-driven approaches as ML and DL, with explicit prior knowledge in terms of model-based approaches, is already addressed in a couple of Helmholtz Association research groups.

In contrast to existing expertise, researchers identified which topics necessitate further development and support. Table 2 (Section 9.2) lists methodological needs researchers requested. Topics that repeatedly occur are: DL, explainable AI, coupling data-driven approaches with explicit prior knowledge in terms of model-based approaches, scalable and distributed ML and DL in extreme-scale computing, uncertainty quantification, quality control of ML models and real time parameter tuning, time series analysis, data fusion, efficient learning strategies and learning from sparse data. These are topics which should be addressed on HAICU's research portfolio. Those which are currently of highest interest are listed as potential topics for HAICU Central's research groups, see Section 4.

The goal of HAICU is to provide the necessary links between the existing expertise (Table 1) and needs (Table 2), as well as to increase the volume of AI research per-se. HAICU will address the researcher's requirement for missing links within the Helmholtz Association, and their demand for an increasing, especially transdisciplinary knowledge transfer as well as more possibilities for educating young researchers in the rapidly changing fields of AI.

In many domains researchers cope with legal barriers in terms of data sharing, open access and open data. Making progress in these issues at least within the Helmholtz Association will be one step forward and is essentially needed for real collaborative working across centers and research fields. Finally, in order to ensure all possibilities of collaborative working, basic tools like cloud service, cloud storages, communication channels, etc., working across the whole Helmholtz Association, are essential.

STATUS QUO AND EXEMPLARY USE CASES FROM THE PAST

As lighthouse projects for recent advances in AI in the Helmholtz Association we would like to highlight a few exemplary achievements:

- Researchers at the HMGU Institute of Computational Biology together with colleagues from ETH Zürich have developed a deep learning algorithm that is trained with millions of image patches of single blood cells. Using the algorithm, they can predict the fate of a blood stem cell, the cell type it will differentiate into, up to three generations before conventionally used surface markers (Buggenthin et al. 2017).
- In particle physics a highlight of the ATLAS and CMS experimental program involving large amounts of data, where researchers from DESY are involved, was the identification of very rare events in the measurement of Higgs boson processes, in association with top quarks. The identification of those events was only possible with advanced ML and DL based data selection methods and therefore can be regarded as a triumph of ML (ATLAS Collaboration 2018; CMS Collaboration 2018).
- In Neuroscience, researchers from FZJ use AI, especially DL methods, to support the development of the digital 3D atlas of the brain by automatically annotating different areas of the brain. It also helps to create maps that indicate the probability that a given area of the brain is found in a specific location in the brain. Furthermore, rules about the connectional organization of the brain are being extracted that may have the potential to be translated to artificial neural networks. These successes significantly enhance science's ability to better understand the structure and function of the human brain (NVIDIA 2017). The investigations are part of workflows utilizing supercomputers at Jülich's Supercomputing Center, in order to cope with the massive amounts of data and computational efforts of highly resolved brain sections.
- In space physics plasmaspheric modelling is important for GPS navigation and satellite charging. Researchers from GFZ were able to automatically infer the electron number density from plasma wave measurements using ML methods given solar wind and geomagnetic parameters and location information. In the following, they were able to combine these with point measurements into a global plasmaspheric model (Zhelavskaya et al. 2016; Zhelavskaya et al. 2017).
- Researchers at IPP have recently developed (in collaboration with colleagues in the U.S.) neural network (NN) accelerated models for predicting the radial density and temperature profiles in magnetically con-

fined fusion plasmas (Meneghini et al. 2017). With a speed-up of more than three orders of magnitude over traditional methods, these NN models can be used for real-time control applications which have been identified in a recent study by the DOE of the U.S. (FESAC 2018) as a 'First Tier Transformative Enabling Capability'.

- Researchers from the DKFZ together with the Neuropathology Department of the Heidelberg University applied ML methods to identify molecular distinct brain tumour entities by analysing DNA-methylation data and established a DNA-methylation-based classification of central nervous system tumours. They trained a ML classifier that greatly improved precision in routine diagnostics of thousands of prospectively diagnosed brain tumours and provide an online classification tool that is widely used by neuropathologists around the globe (Capper et al. 2018).
- Researchers from DKFZ also showed how advanced imaging and ML technologies can help to avoid many control biopsies following suspicious findings from mammography screening. They combined diffusion-weighted MR imaging with intelligent image analysis and ML methods to detect malignant changes in tissues and thereby were able to reduce the number of false positive detections in the study group by 70 percent. In sum they reached an accuracy which is comparable with the reliability of MRI methods that use contrast agents (EurekAlert! 2018; Bickelhaupt et al. 2018).
- Researchers at the MDC used ML to identify immune infiltration in tumours from histological images (Yuan et al. 2012) and developed novel ML techniques to infer the evolutionary history of cancer in the patient from genome-wide profiles of somatic aberrations (Schwarz et al. 2014; Schwarz et al. 2015). By applying these techniques, they could demonstrate ongoing selection and convergent evolution in the genomes of patients across thousands of clinical samples (Jamal-Hanjani et al. 2017; Turajlic et al. 2018) and trace early evolutionary events from circulating tumour DNA with broad applications for early detection and intervention (Abbosh et al. 2017).
- Other groups at the MDC are at the forefront of applying ML in ground-breaking applications for genomics and gene regulatory networks. They allow for the most comprehensive direct identification of translated mRNAs (Calviello et al. 2016), showed the first 3D reconstruction of gene expression in an animal model based on single cell genomic (Karaiskos et al. 2017), and develop ML approaches for first-in-class reconstruction of ultra-high resolution images from light and EM microscopy (Preibisch et al. 2014).
- A team of researchers from GEOMAR together with international partners applied ML methods to deep sea image and hydro-acoustic data to assess marine mineral resources. The results highlight the high spatial diversity in the mineral resource distribution, show strong links between the resource and the associated fauna and for the first time enable a direct and quantifiable observation of the anthropogenic impact in the deep sea (Peukert et al. 2018). Another GEOMAR team developed a novel clustering-based method for spatio-temporal analysis of ocean observations to separate the natural variability from sampling uncertainty which enabled mapping previously unknown ocean current systems (Koszalka et al. 2011). A unique evolutionary optimization algorithm helped to tune the parameters of biogeochemical ocean models to better understand the de-oxygenation of the oceans under a warming climate (Kriest et al. 2017).
- Coastal research offers a broad range of applications for ML. Novel approaches were presented at HZG combining advanced Earth System Modelling and observations through ML based data assimilation (Wahle et al., 2016). Current efforts in the frame of the Helmholtz Incubator Information & Data Science project RedMod focus on the application of ML to reconstruct regional climate in high resolution on centennial time scales enabling cutting edge research.

- Researchers from DLR are pioneers on deep learning in remote sensing and earth observation (Zhu et al., 2017). General challenges include very limited and noisy annotated data, integration of domain expertise into deep learning models, five-dimensional data (x-y-z-t-λ) and interpretable and quantitative AI. They have developed tailored deep nets for a variety of remote sensing problems, such as hyperspectral image analysis (Ghamisi et al., 2016, Mou et al., 2017a), scene interpretation of space videos (Mou et al., 2017b), multimodal data fusion with a particular focus on SAR and optical fusion (Hu et al., 2017, Hughes et al., 2018), geo-information extraction from social media data (Kang et al., 2018), and very high resolution scene interpretation (Mou and Zhu, 2018).

Helmholtz Centers cooperate mostly within their research field. They are densely connected within their domain (international research institutions, universities, industry), but less strong connected across domains. HAICU will function as an essential connector for more cross-domain cooperations.

During our discussion with different researchers we already identified parallel developments in different domains and therefore potential synergies for knowledge transfer and future collaborations.

Table A 1 summarizes existing expertise in different fields of AI and ML. This already highlights similar theoretical and practical expertise in the same fields. As practical examples we identified the following ideas: ML methods for detecting and counting manganese nodule in deep sea images is similarly applicable on cells in histological brain sections, and vice versa. Similarly, DL methods used for clustering of plankton are applicable to clinical data, such as single cells, potentially in order to detect cancer cells, and vice versa. Finally, methods from medical x-ray tomography for the sake of cancer diagnosis might be transferred to cracks in materials.

The Helmholtz Centers also have platforms that show that there is high demand for concepts like HAICU. For example, the Helmholtz Zentrum München has a variety of core facilities that are used for support, teaching and consultancy. Examples include the biostatistics core facility, which has been steadily increasing since its inception, and a newly established Next-Generation Sequencing (NGS) bioinformatics core facility, which was established in accordance with POF recommendation and shows high demand of the institutes and great success. Similarly, the Berlin Institute of Medical Systems Biology (BIMSB) at the MDC established several platforms working as High Level Support Teams (HLST) at its start in 2009, including a Scientific Bioinformatics Platform (MDC 2018) that offers consultancy, training in Bioinformatics and Machine Learning, and providing assistance in projects. Another example for such a structure is the National Analysis Facility (NAF) of DESY used for Computing by all German institutes involved in \`Heads of the Particle Physics Division\` (HEP); part of NAF are computing schools and user support done by the IT department for ML. DKFZ has established core structures in the area of genome informatics, medical imaging and medical informatics that support research groups and clinical applications. Simulation Laboratories (SimLabs) at FZJ and KIT provide domain-specific research and support structures with advanced support in the fields of data analysis, modeling, simulation, HPC methods and visualization. In close cooperation with scientific groups, they provide independent research contributions in several application areas of scientific computing (FZJ JSC 2018; KIT SCC 2018). HLSTs within the Human Brain Project (HBP) provide support for HBP-internal and external users from neuroscience, medical research, neuromorphic computing and neurorobotics with respect to tools, services and data science (HBP 2017). Data Lifecycle Labs work in close cooperation with Helmholtz Association scientists and process, manage, and analyze data during its whole life cycle (e.g., LSDMA 2014). Finally, we refer to the John von Neumann Institute for Computing (NIC) that manages the compute time on supercomputers (NIC 2018) on a peer review basis.

3 GOALS

Today in Germany, the efforts in AI applied to big data as in the Helmholtz Association are still in their infancy and we now see the big chance for the Helmholtz Association to become the major German driver for applied AI. Both in politics and in the general public, AI is becoming an increasingly important topic, and it is inconceivable that Germany may lag behind in one of the most important challenges of the next years. It is the right time to bring all AI experts throughout the Helmholtz Association together in order to build up an internationally competitive base for applied AI. So far, synergies in applied AI within the Helmholtz Association have hardly been realized, and HAICU will offer an exciting new strategy which will speed up applied AI research in Germany with international impact. By explicitly building on the unique datasets generated within the Helmholtz Association, and leveraging its large scale data and compute facilities for HPC-driven big data analytics, HAICU will be able to fill a unique position in the field.

3.1 GENERAL GOALS

HAICU´s overarching goals are:

I. **Develop and implement novel AI methods** motivated by concrete research questions and apply them **in use cases across multiple Helmholtz Centers**, and

II. **Provide support for applied AI activities** for users in the various centers and **disseminate and educate emerging methods** in the AI field.

Reaching these goals, HAICU will enable the Helmholtz Association to

- **Guide** as one **of the pioneers in applied AI, reach an international leadership position** and **increase visibility** by combining AI-based analytics with the Helmholtz Association´s unique research questions, unique data sets, and unique data infrastructures and simulation facilities.
- **Serve as an enabler of applied AI-based research activities** in the research fields and centers.
- Create **a powerful network** for ground-breaking new ideas and exchange in the field of applied AI, based on the substantial competencies rooted within Helmholtz Centers and programs.
- Enable responsible research in the field of AI for the **benefit of society.**

3.2 DETAILED GOALS

In order to reach the goals defined above, we describe the following concrete conceptual aspects HAICU will address

Goal I: Develop and implement novel AI methods motivated by concrete research questions and apply them in use cases across multiple Helmholtz Centers.

I-1 **Contribute to conceptual cross-discipline developments in applied AI** motivated by application examples.

I-2 **Identify and leverage similarities** between applications across different topics and implement those in generalized AI methods.

I-3 Contribute to the theoretical basis of AI and use this for better methods in **applied AI methods**.

I-4 **Create a crystallization point for excellence** by establishing a powerful network.

I-5 **Provide scalable implementations of methods on large compute systems** for scientific big data analytics.

Goal II: Provide support for applied AI activities for users in the various centers and disseminate and educate emerging methods in the field.

II-1 **Distribute methodological strengths** across all Helmholtz Centers.

II-2 **Provide support** across all Helmholtz Centers to emerging activities in applied AI.

II-3 **Educate and train** the next generation of data scientists and all Helmholtz Association employees on different scales.

II-4 **Offer a creative research atmosphere** by providing space for collaboration and unite technologies of different disciplines.

According to identified requirements and taking into account best practice examples worldwide, we identified the following cornerstones:

HAICU will have a strong focus on the development of methods that are essential for the different domain specific research fields in the Helmholtz Association. However, a solid background in fundamentals of AI, going beyond a familiarity with some of the tools, is highly relevant in doing innovative applied work. E.g., technology companies place a huge emphasis on hiring people with a strong background in the fundamentals of AI, although the main work of these companies is application driven. Similarly, for the Helmholtz Association the main long-term gains will come from integrating solid AI expertise into the centers. Along the same line of arguments, neuro-inspired research will facilitate the development of scientific research and application.

Therefore, there is a need for strong research groups around these methods, who generate use cases. A good use case does not only drive methodical development but has valuable potential for generalization to other research topics.

However, HAICU will not only be a platform for scientists doing their own research, it will also provide tools and services across centers on different levels with the aim of supporting scientists with a flexible way to consider different expertise, the volume of a research task, and other constraints. For these needs a flexible hierarchical system of cooperation possibilities has been conceptualized for which can be applied:

1. a fast reviewed and easy accessible voucher system for short projects with project runtime in the range of two weeks up to six months (small projects, feasibility studies, etc.).
2. bigger scientific projects in the range of up to three years (medium projects).
3. a HAICU Young Investigator Group program (HAICU YIG) for scientists performing application-oriented science for research periods of five years (large projects), located at HAICU sites that can be readjusted.

HAICU will drive new methods and ask new questions. Therefore, there is a need to gather and curate domain specific data sets with more abstracted cost functions and to allow for systematic benchmarking, potentially across disciplines. This enables the centers to provide and use benchmarks more quickly and professionally and administers a strong input to method groups to keep them up to date and to strengthen reproducibility of research. Further, it has promising side effects, such as high visibility and citations.

In order to ensure all of these goals there must be strong ties to other activities in the Helmholtz Incubator Information & Data Science, and in the different research fields.

A strong network between all Helmholtz Centers is essential to guarantee the success of HAICU. To ensure this, there will be different essential connecting factors for the AI community in the Helmholtz Association:

- Communication channels similar to slack channels and social media, symposia and workshops will enable the exchange of experience, an excellent scientific network and getting in touch with the members of support teams and research groups.
- International experts will give regular talks that will be video channelled to reach a broader audience.
- High Level Support Teams (HLSTs) at HAICU will provide support and assistance in the planning, preparation, realization and implementation of project collaborations on different scales. This starts with answers of questions in the case of proposal preparations for first ideas and continues with support of HLST experts during all phases of the projects.
- Training and education on different levels will on the one hand create a new highly qualified generation of data scientists and allows on the other hand to further educate all Helmholtz Association employees, in close cooperation with the proposed Helmholtz Information & Data Science Academy (HIDA).

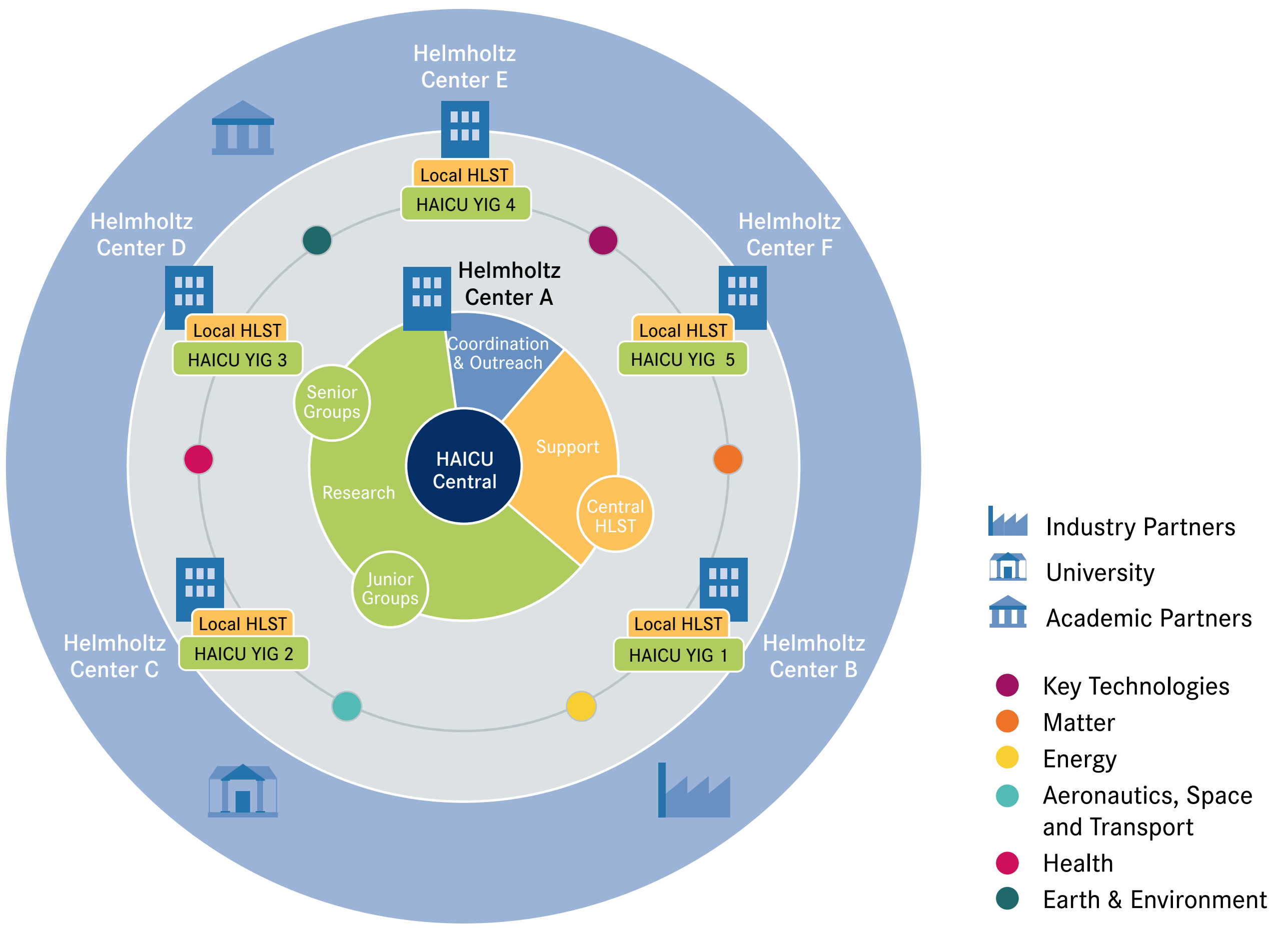

Figure 1: The Helmholtz Artificial Intelligence Cooperation Unit (HAICU): a strong, internationally visible Central Unit and multiple local Units as bridgeheads into the domain specific research fields. High Level Support Teams (HLSTs) provide expert support to users across all Helmholtz Centers.

4 HAICU STRUCTURE

The structure of HAICU has been proposed and subsequently optimized during several Helmholtz Incubator Information & Data Science meetings, telephone conferences and visits to all Centers, where we discussed it in depth with all participating researchers. Based on the formulated aims above, we therefore propose the following organizational structure (see Figure 1).

4.1 OVERVIEW

We propose to structure HAICU in a **wheel-and-spoke model**, with an interdisciplinary highly visible central unit called \`HAICU Central\` and a strong applied periphery located at various Helmholtz Centers, called \`HAICU Local\`. Corresponding finances will be 4.0 M€/a for HAICU Central (35% of overall volume) and 4.4 M€/a for HAICU Local (38%). Additionally, 2.7 M€/a (24%) will be allocated to Helmholtz-wide projects and 300 k€/a (3%) are foreseen for HAICU events.

HAICU Central will address three core themes – \`Research', 'Support' and 'Coordination and Outreach' – which use 54 %, 27 %, and 19 % respectively, of the resources. It will host five research groups, two of them senior, three of them junior groups, focusing on methodological research topics. Further, HAICU Central will host a team of experts, called High Level Support Team (HLST), supporting all centers that have needs for AI (see Section 4.2.). Distribution of manpower takes place through a voucher system (see Section 4.2).

We propose to attach HAICU Central to one Helmholtz Center yet to be selected in an international, transparent review process, to which all Helmholtz Centers can apply on the basis of clear criteria (see Section 4.2). To ensure that HAICU Central is working in the proposed direction, strong quality measurements and control mechanisms based on milestones are defined (see Section 4.4).

The **HAICU Local** units will be spread across five other Helmholtz Centers to ensure wide dissemination of methods as well as broad uptake of data-based research questions. Ideally, each research field will host one of these local sites (excluding the one which hosts HAICU Central). They will be represented by five HAICU Young Investigator Groups (HAICU YIG) funded for a period of five years and five local HLSTs. Each Helmholtz Center can apply for one of these YIG groups and one HLST. In exceptional cases, it is also possible to apply only to one of the two components. Strong in kind contribution of the center is recommended with respect to room space/infrastructure, future perspective for the groups after the 5 years funding phase, embedding of the groups into the respective Helmholtz Center etc.. Topics for HAICU Local groups will be broad, application-oriented, domain specific or potentially cross-topic. An HLST will be affiliated to each of these HAICU local sites. A strong connection to HAICU Central is essential for the local HLSTs (regular meetings, jointly decided allocation of topics for different HLSTs). Quality measurements and control mechanisms based on milestones are defined (see Sections 4.4 and 6) and will ensure cross-topic cooperations and service performance for other centers.

Strong interactions between HAICU Local and Central allow the most efficient way for creating exciting new results very fast. HAICU Central will address methodological research topics, see Section 4.2., in order to utilize them for the domain specific areas. The research will be driven by domain-specific use cases and research questions at HAICU Local sites. Specific use cases will be given by research groups across the Helmholtz Association.

4.2 STRUCTURE OF HAICU CENTRAL

HAICU Central will address the three core themes 'Research', 'Support', and 'Coordination and Outreach', which will be explained in more detail in the next sections.

RESEARCH

AI tasks within the Helmholtz Association are typically domain specific handling and analysing large and very specific datasets. These datasets have been generated by highly specialized and innovative hardware that operates often at high speed (e.g., high-throughput acquisition, streaming data applications).

As already stated in Section 3.2, a solid background in state-of-the-art AI methods is highly relevant for innovative domain-specific applications. Further, it may be a unique selling point to be able to apply AI methods on raw data generated by specialized large-scale equipment of the Helmholtz Association.

We envision the following topics to be of particular importance, thus enabling multiple, general data science applications:

- Deep Learning (DL)
- Explainable AI
- Uncertainty Quantification
- Scalable and distributed ML and DL in extreme-scale computing
- Incorporating prior knowledge, e.g., equation-constrained modelling
- Transfer learning for cross-discipline applications

These topics are based on the specific strengths, expertise and requirements of different Helmholtz Centers, which have been identified during requests and visits of all Helmholtz Centers. We will explain these points in detail below.

Deep Learning is currently the most salient set of methods in ML that uses various artificial neural network architectures, reaching ground-breaking performance applied in a diverse field of applications. Currently, a strong research trend in DL is to create methods for unsupervised, generative learning that would unify such approaches as generative adversarial networks (GAN), variational autoencoders (VAE), variants of predictive coding driven learning and reinforcement learning. Other important trends in DL address the task of demystifying how neural nets work, what we call here **explainable AI,** or to **incorporate neurobiological constraints on information processing and learning in network architectures** (Sabour et al. 2017). Artificial intelligence has a traditional link to the science of biological information processing, most particularly for the case of vision processing (Marr 2010; Cox and Dean 2014). Very often, significant advances in AI borrow inspiration from neuroscience. For example, similar to the central building block of ResNets (He et al. 2015), skip connections are a structural property of connections of brain areas (van Hartevelt et al. 2015; Schirner et al. 2018; Urbanski et al. 2014; Felleman and Van Essen 1991), as well as hierarchy of processing stages with feed-forward convergent connectivity patterns across the layers. At the same time, there is a growing need to **relate and explain the behavior of AI** algorithms taking evidence from cognitive research and neuroscience (Ritter et al. 2017; Fong et al. Cox 2017). Despite the potential benefits, there is still a large gap between biological and artificial cognitive systems, which should be closed to better understand the advantages and disadvantages of different network organisation. HAICU aims to realize this potential to innovate AI in this respect, by

combining the existing expertise in the Helmholtz Association on biological information processing, neuroscience and cognitive science, and human intelligence with AI research.

An important challenge in the current development of AI methods is the question of **dealing with explicit knowledge**, as opposed to implicit knowledge contained in and acquired from the training data by learning. This especially concerns DL methods that use neural network architectures to learn tasks from the data. It is still unresolved how to inform a neural network about existing explicit solutions to a task, for example available in form of physics equations describing a part of task relevant phenomena, and there is still a long way to go to have a methodology that incorporates and reads out explicitly accessible knowledge about acquired skills from trained neural networks, leading to transparent, **explainable AI**. The Helmholtz Association can provide crucial expertise in this line of research, as **dealing with explicit models within the simulation science** is a main part of its scientific portfolio. Further, this field of research has not been sufficiently addressed so far within the AI community and, due to the strong potential, might offer a niche where the Helmholtz Association may reach excellence – competitive at the international level.

Research areas dealing with measurements typically address the task of quantifying their measurements' accuracy and reliability which is essential for a proper assessment of experimental results. This should propagate through the whole evaluation pipeline until the final result of data interpretation. Thus, on the one hand uncertainty must be measured, e.g., for measuring instruments. On the other hand, technical uncertainty must be coupled with **subjective information**, e.g., manual mappings by experts, e.g., hydrologist or geologist. Up to now it is not clear how to integrate such information, which are obviously error-prone due to subjectivity, as ground truth information for training ML methods. Finally, for most applications it is essential to provide results of data interpretation along with a proper **quantification of its uncertainty** in order to allow a proper assessment of these results and its reliability.

Another important challenge is to provide AI methods with enough input and proper environment for training. As already mentioned, specifically for the Helmholtz Association is its large amount of high quality data, but typically these data are provided with limited annotation, which leads to a strong need for self-supervised or **unsupervised pre-training, transfer learning,** and **sparse reinforcement learning methods**. For some domains, like image recognition, there already exist data sets with manually created training labels like for instance ImageNet (Stanford Vision Lab 2016), CIFAR-100 (Krizhevsky 2009) or COCO (COCO 2018) that cover the demand. Many other domains however do not offer such large collections of data, because data collection is generally tedious and effortful. For some domains it is not even clear whether it is possible to gather data in the necessary amount from the real-world setting (e.g., data on rare, exceptional phenomena like disruptions in plasma physics, earthquakes, financial crashes, etc.). One pathway taken into direction of **providing AI methods** with appropriate input is **realistic simulation** of task relevant environments. Platforms like OpenAI Gym (OpenAI 2018) or DeepMind (Beattie et al. 2016) Lab offer a diversity of simulated environments where AI algorithms receive complex input, can influence the environment by taking actions and where their performance can be compared. Using its expertise in different areas of **scientific computation and simulation**, the Helmholtz Association can extend this effort by providing state-of-the-art simulation environments for different fields of natural science (e.g., climate modelling, Earth system simulation, plasma fusion, high energy and astroparticle physics, solar energy, proteomics, genomics and radiomics, etc.) that can **generate data for training ML algorithms**.

Machine Learning applications show increasing requirements for **training complex models on large dat**a, often demanding high communication bandwidth and very large memory storage. This applies to use cases with large size single inputs, like for instance high resolution imagery, or in case of streaming data. Our research therefore has to address how to **distribute training procedures** in cases where trivial parallelism fails. Here, the HPC facilities in the Helmholtz Association will play a strategic role: HPC architectures provide low latency and high-bandwidth interconnects that dramatically accelerate distributed training with large data or models. HPC research will also soon provide new accelerator technologies beyond GPUs in modular environments, which offer new possibilities for speeding up training and inference. One important application will be fusing the learning of forward generative models (simulation, data generation) and of inverse models (inference from data) in one single end-to-end architecture. Such DL architecture will be able to improve the initially error prone forward simulation model by detecting inconsistencies in forward-inverse loops and also by digesting data originating from real experiments. Due to the size of such a network, memory and communication demand will be extremely high. Developing **scalable AI methods**, especially those relying on DL architectures, on HPC systems will therefore be one research aspect in HAICU, with a strong link to the HIFIS.

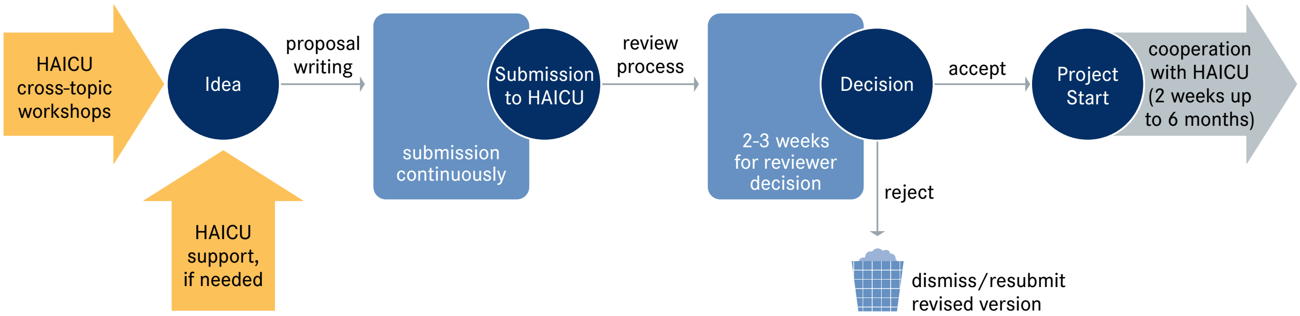

Figure 2: How to get started with a voucher

HAICU SUPPORT: COLLABORATION VIA SMALL AND LARGE SCALE PROJECTS

HAICU will provide high-level support to users in a flexible, dynamic and transparent way. While long term scientific questions will be addressed by HAICU's research groups, HAICU will fund larger research projects for the duration of up to three years. Finally, for small projects with runtimes of 2 weeks up to 6 months, the High Level Support Teams (HLSTs) will provide additional engineering and software support. These small projects managed via a voucher system will ensure minimal bureaucracy, short term collaborations and support with respect to tools, services and data science.

Users, i.e., researchers from the different Helmholtz Centers will apply for such voucher projects in order to work with HAICU on a specific AI problem. In general, centers request for transdisciplinary knowledge transfer and networking, educating young researchers, and support on different scales, including support for their local projects and support in creating and advertising high impact benchmark datasets.

Capacities of HLSTs will be managed by a voucher system. For different research questions, researchers can apply for vouchers, i.e., support at HAICU Central, which is strongly linked to local HLSTs and initiates/coordinates the contact between researchers and the respective HLST. The process from getting an idea up the actual start of such a voucher project is visualized in Figure 2. Support begins with consulting during the proposal preparation phase, gathering first ideas, and continues with support by HLST experts during all phases of different projects. Each support service will be coordinated through vouchers irrespective of whether or not the HLST is co-localized at the requesting Helmholtz Center.

Therefore, HAICU Central will implement a *quality assured voucher system* for these smaller projects with fast-track applications in line with the rules and regulations of the Initiative and Networking Fund of the

Helmholtz Association. They differ from *calls for larger research projects*, which will undergo in-depth independent peer review. For voucher systems, we can rely on existing expertise inside the Helmholtz Association, see Section 2.3.

The process of resource allocation for both types of projects is visualized in Figure 3. Both, for small and large projects, users apply by writing a short proposal, including a clear definition of task, goal, and time frame. To guarantee scientific excellence of the selected projects, the evaluation and selection procedure will be coordinated by the HAICU 'Project Coordination and Outreach' unit. Vouchers will be reviewed and decided by the scientific coordinators along a procedure agreed with the Scientific Advisory Committee (SAC), while project calls will be reviewed by the SAC (see Section 6). Review criteria will be quality of the project and research question, qualification of the applicant, job opportunities and scientific environment, goals and working program, suggestion for the scope of funding, and diversity and equal opportunity in the science system. The Steering Board will prepare funding decisions for projects, while the President of the Helmholtz Association will decide about the funding through the Helmholtz Association Initiative and Networking Fund. Such open and dynamic resource allocation scheme have also been successfully implemented by other international institutes, e.g., *The Moore-Sloan Data Science Environment* (msdse 2017).

In order to rapidly exhibit first success stories, a flying start for larger project calls right after the concept decision in September 2018 is planned and a first call for `HAICU pilot projects` will be announced: the initial criterion will be two centers from different research fields finding synergies and writing a proposal. All proposals will be evaluated and ranked by external experts in a process managed by the Head Office of the Helmholtz Association, until the Scientific Advisory Board is established. The final decision of the project award is in the hands of the President of the Helmholtz Association. Evaluation of these HAICU pilot projects by external reviewers after two years will be coordinated by the scientific coordinators and reviewed by the SAC, the final decision is in the hands of the Steering Board.

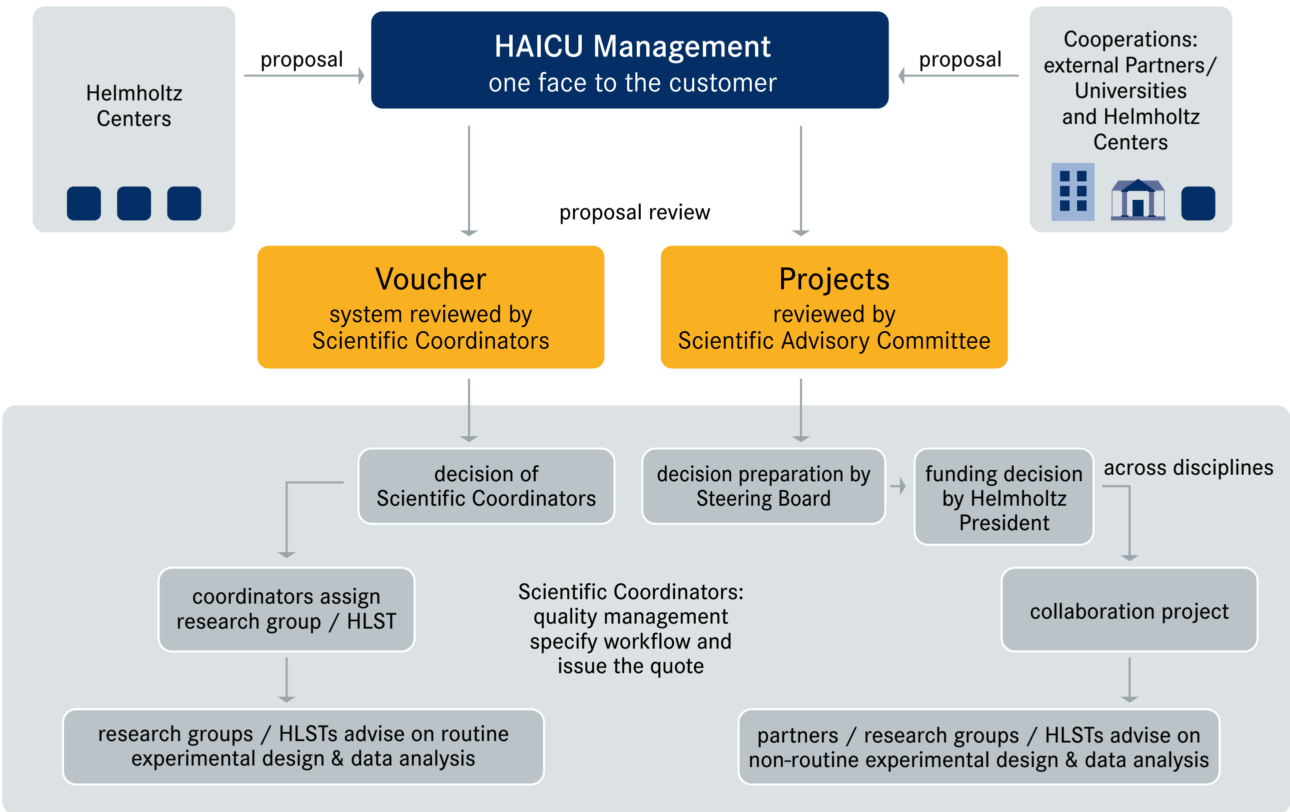

Figure 3: Resource allocation for small scale project vouchers and large scale projects

At the beginning, the major users will be researchers at all Helmholtz Centers. Following the implementation and after having reached full capacity, HAICU will be opened to collaborations of Helmholtz Centers with external partners, e.g., universities or, at a later stage, industry. External partners are not eligible to apply without cooperations with one Helmholtz Center.

Benchmarks: Providing Helmholtz Association Open Data Challenges

Helmholtz Centers have very unique datasets. Preparing such datasets as a structured Machine Learning task or a comprehensive open challenge for Helmholtz Centers and/or the international AI community, can be an effective instrument to attract external data analysts. However, it is often a difficult and time-consuming task, requiring routine and expertise in benchmark design. Publicly available benchmark data and standardized evaluation frameworks will increase the visibility of HAICU and Helmholtz research groups and will have an important contribution to the reproducibility of research results and therefore the verification of research methods. As examples we refer to the Kaggle competitions as one of the possibilities to increase visibility by participating and contributing to openly available benchmark and competition platforms in the community. As highest impact benchmarks in AI we refer to the ImageNet challenge (Russakovsky et al. 2014; Stanford Vision Lab 2010) or 'The Higgs Boson Machine Learning Challenge' which uses the ATLAS experiment to identify the Higgs boson (Higgs Challenge 2018).

As best practice examples with Helmholtz Association participation (MDC, HZI) we refer exemplarily to a benchmark workshop about 'Advancing computational biology through critical assessments, community experiments, and crowdsourcing' in conjunction with the International Conference on Intelligent Systems for Molecular Biology (ISMB 2018). Also, at the DLR´s Earth Observation Center, a big data benchmark about social media and earth observation satellite data is in preparation (So2Sat 2017).

HAICU, specifically members of the central HLST, having expertise in these tasks, will help the Helmholtz Association research groups to expose their data and analytics problems to the wider community and attract methodologists. To this end, HAICU will gather information about domain specific datasets and according analytics problems and provide support in defining and publishing benchmarks – data, annotations, task descriptions, and evaluation frameworks – in a standardized manner to the broad community.

This naturally implies a close collaboration with other Helmholtz Incubator Information & Data Science activities, Helmholtz Metadata Center (HMC), Helmholtz Infrastructure of Federated ICT Services (HIFIS), and Helmholtz Imaging Platform (HIP). Benchmark datasets will be stored at their owning Helmholtz Centers. HAICU will cooperate with HMC in order to gather information about data at one place and to ensure data to be searchable by appropriate metadata. HIFIS is expected to provide possibilities to access these data in a fast, secure and easy way and further to allow bringing AI algorithms to the data instead of vice versa. HIP is expected to be a driver for generating unique and interesting data. Finally, providing benchmark data will allow the Helmholtz Information & Data Science Academy (HIDA) to provide student challenges on specific Helmholtz datasets.

To increase outreach wrt. such Helmholtz benchmarks, HAICU plans to host open workshops and hackathons with international visibility.

COORDINATION AND OUTREACH

The 'Coordination and Outreach' unit of HAICU Central will be the managed interface to all partners and hubs.

Management team

HAICU will require a compact management team, including the Head of HAICU, a financial manager, a secretary, and four scientific coordinators.

The four scientific coordinators will guarantee fluent daily business. One of these coordinators will exclusively be responsible for public relations and outreach, which means in detail he or she sets up a webpage and maintains social media (Facebook, Twitter, LinkedIn, Xing). He or she will organize Helmholtz Association internal and external workshops and conferences, video conferences and online talks and publish video talks. This coordinator will establish public relation events, like popular scientific discourses. If requested by a domain scientist, Citizen Scientist projects might be implemented, in close collaboration with the HLSTs, as an effective opportunity for public participation in scientific research. As best practice example from the Helmholtz Association we refer to the planktonID project (Christiansen and Kiko 2016).

Further three scientific coordinators will be responsible for the coordination of all HAICU Central and Local activities and projects: they will coordinate HAICU Central and Local research groups and HLSTs through regular meetings in order to provide exchange. Besides, they set up regular meetings of all boards and coordinate all processes in the sense of governance. Vouchers and projects will also be managed by these coordinators: they will publish calls, coordinate the review processes of proposals, assign requests from Helmholtz Centers to the most suited HLST and monitor the progress of each project in accordance with the rules and procedures of the Initiative and Networking Fund. Vouchers will be reviewed by themselves for a short and easy allocation. They will organize knowledge transfer on all scales (see knowledge transfer below) and will be responsible for scouting AI experts by attending international scientific conferences and staying up to date to the newest developments. Finally, they will coordinate the evaluation process and regular reporting within HAICU to the Helmholtz Incubator Information & Data Science and to the Helmholtz Association.

In addition to that, the Management Team of HAICU Central will be in touch with strategic partners from external universities, research institutes or even industry - always in the scope of Helmholtz Association relevant topics.

Knowledge Transfer

Currently, there is a critical lack of young data scientists especially with domain-specific knowledge. During our center visits all participators stated that one of their biggest needs are more possibilities for educating young researchers in the rapidly changing fields of AI. Thus, an important part of HAICU´s activities is the transfer of knowledge, which has an essential link to HIDA. HAICU will ensure training of young researchers (first of all PhDs and PostDocs) and will have strong connections to the Helmholtz Data Science Graduate Schools that are currently under review in several locations or already implemented (HEIBRIDS in Berlin). These graduate schools provide important links between the methods-oriented local Data Science communities at universities and scientific application areas of the Helmholtz Association partners. It will offer PhD positions related to the different HAICU methods research fields and applied to one or more domain specific use cases. Thus, the next generation of excellent-qualified data scientists will be trained. In the context of larger research projects coming in through calls, additional PhD students will be allocated in the Helmholtz Centers to work closely with the coordination unit. Laboratory rotations are another attractive possibility for students to gain knowledge. This way a significant, but dynamic part of HAICU's resources will be utilized in different centers.

Knowledge transfer also requires the organization of regular Helmholtz internal workshops, as well as conferences with international scope. Training and exchange among young researches will take place in summer schools and hackathons organized by HAICU under the roof of HIDA.

Training and further education in AI, workshops and conferences for all Helmholtz Centers will be organised by the scientific coordinators and offered and made available throughout HIDA.

As knowledge transfer and further education for all Helmholtz Association employees were identified as requirements during our center visits, relevant offers for all Helmholtz Association employees will be implemented depending on the demands. Professors, group leaders and HLST staff will give lectures, seminars and trainings for all interested Helmholtz Association employees. These measures will be complementary to the course offers of the Data Science Graduate Schools and the experts´ talks and seminar series of HAICU and will be coordinated through HIDA.

Visiting Researchers Program

In order to gather innovative research methods, keep track of most recent research questions, and on the other hand ensure knowledge transfer, HAICU will (i) attract excellent international senior researchers for long term visits (e.g., half a year including teaching responsibilities) as guest professors and (ii) implement a short term visiting researchers program for external young researchers.

In detail, goals of both are

- Exchange of ideas in both directions
- Identification of joint research interests and most recent research directions
- Publication of results in joint publications

For the short and the long term visiting researchers program, the scientific coordinators will publish calls after consultation with the Scientific Advisory Committee (SAC) twice a year, according to specific scientific topics and recent research questions. Excellent young researchers may apply to join HAICU typically for the duration of 6-12 weeks for short term visits and half a year in the case of guest professorships. Applications will be coordinated by the scientific coordinators, reviewed by the SAC and appropriated candidates will be selected by the Steering Board. As best practice example cf., e.g., Alan Turing visiting researchers program (The Alan Turing Institute 2016) or UCLA Postdoctoral Scholars (UCLA 2018). In order to attract excellent candidates to join HAICU, the program has to offer highly attractive research questions, which is ensured by the SAC. It should offer excellent working condition and partial funding, e.g., free accommodation and travel. Finally, a mentoring program provides scientific advice by honourable fellows.

4.3 PROFILE OF REQUIREMENTS

Key for the success of HAICU is an excellent and strong HAICU Central that reaches international visibility. Thus, the following criteria should be met by the hosting Helmholtz Center.

HOSTING HELMHOLTZ CENTER FOR HAICU CENTRAL

The Helmholtz Center that hosts HAICU is expected to provide (always in accordance with the HAICU concept)

- already existing, strong and international competitive expertise in applied AI
- already existing tight connections to and scientific collaborations with research facilities and / or universities offering high quality data science education programs and research, including the possibility to award PhD titles
- realistic concept for attracting AI researchers to the site, ideally proven track record
- excellent scientific infrastructure within the hosting Helmholtz Center and collaborating universities at the site
- excellent IT infrastructure providing best conditions for scalable and distributed ML in large-scale computing
- a convincing concept for the concrete implementation of research groups and HLSTs in relation to the existing scientific structure of the center and the local scientific landscape
- a strong concept for the scientific focus of HAICU central research groups
- opportunities for long term perspectives of junior group leaders
- an attractive working environment (family-friendly, flexible working time models)
- good possibility to integrate family/ marital partner: site has critical mass to offer broad employment possibilities for marital partner and children day care
- good international accessibility of the site especially in terms of close distance to international airport and central train station
- existing links to companies with AI experience and initiatives

HOSTING HELMHOLTZ CENTERS FOR HAICU LOCAL GROUPS AND HLSTS

It is strongly recommended that every research field participates in hosting a local group and an HLST. Helmholtz Centers applying for a HAICU Local Unit are expected to provide (always in accordance with the HAICU concept)

- an excellent scientific infrastructure within the hosting Helmholtz Center
- a convincing concept for the concrete implementation of research groups and HLSTs into the existing scientific structure of the center and the local scientific landscape
- a strong concept for the scientific focus of the research group and integration into the research of the respective research field

- possibilities for collaborations with universities offering high quality data science education programs and research, including the possibility to award PhD titles
- a realistic concept for attracting AI researchers to the site, ideally proven track record
- opportunities for long term perspectives of junior group leaders
- an attractive working environment (family-friendly, flexible working time models)
- good possibility to integrate family/ marital partner: site has critical mass to offer broad employment possibilities for marital partner and children day care
- existing links to companies with AI experience and initiatives

PROFILE REQUIREMENTS HEAD OF HAICU

The profile of the Head of HAICU is crucial as she or he is supposed to contribute to the development of HAICU into the internationally recognized and visible platform we describe here. A professorship, excellent scientific track records and a strong network as well as international reputation in the AI and ML community shall be requirements for this position. Based on her or his professional experience (which we expect to be focused on applied Machine Learning), she or he needs to know the AI and ML landscape in Germany as well as on the international level in detail and needs to be able to successfully negotiate between partners of different corporate cultures (university, non-university research organizations, industry) and various interest groups and to lead and advise multi-professional teams. Within the framework provided by the Host Lab of HAICU Central, the Head of HAICU shall be able to head her or his own group directly within HAICU or independently. The financial manager will support the Head of HAICU regarding financial questions.

PROFILE REQUIREMENTS HAICU GROUP LEADERS (CENTRAL AND LOCAL)

As already stated in Section 3.2, for HAICU Central group leaders a solid background in the fundamentals of AI is highly relevant in doing innovative applied research. We believe that HAICU's main long-term gains will come from integrating solid AI expertise into the centers.

Thus, interested candidates should have a track record in statistical modelling and development of algorithms for data science, and should be internationally recognized experts in one or more of the following fields: Deep Learning, Computer Vision and Imaging, Scientific Computing, Optimization, Data Fusion, Data Management, Big Data, Graph and Network Optimization and Modelling, Natural Language Processing, Information Theory, Statistical Learning, Data Mining.

Additionally, HAICU Local group leaders should provide a strong background in the according application area.

All candidates for HAICU Central and Local are expected to have contributed outstanding research in these fields, a strong commitment to research-oriented teaching and successful acquisition of third-party funding. Moreover, candidates are expected to actively participate in the HAICU network and to make use of opportunities for interdisciplinary research cooperations within the Helmholtz Association.

ATTRACTING GROUP LEADERS (CENTRAL AND LOCAL)

Regarding top experts in the field, HAICU will be in strong competition not only with international companies but also with national and international universities and research institutions looking for leading scientists as well. We expect relatively few people to meet the skill-set needed to lead either methodological or applied AI research at a sufficiently high level. One to two senior experts in HAICU Central, where students can be sent to, will be highly important for visibility. But HAICU will only be attractive to recruit one or two senior scientists if the environment is attractive to them in terms of resources and scientific environment. Therefore, interesting research questions, most attractive working conditions, agreements, and contracts have to be offered.

Arguments to attract researchers and AI experts are the unique research questions and unique datasets available in the Helmholtz Association, the possibility to work on diversified research questions and applications and finally excellent working conditions in terms of scientific infrastructure. The candidates will have a high freedom in research and will be able to work in a highly attractive environment. Finally, we expect HAICU to gain international visibility such that researchers get attracted by its name.

To receive more flexibility in HAICU´s concept and in order to cope with expected difficulties to hire these experts, we suggest as contingency plan to start searching for excellent junior scientists, maybe even shortly after PhD, as junior group leaders. This is very common elsewhere (Cold Spring Harbor Laboratory 2018) and could complement the traditional YIG model. It would endow the local Helmholtz Centers with significant flexibility to get promising activities started quickly on a still somewhat smaller scale and see how things work out. As those groups are limited to five years, we gain flexibility to start looking for senior experts again after that period. Moreover, during that time, HAICU will have grown and gained international visibility.

PROFILE HLST EXPERTS (CENTRAL AND LOCAL)

Candidates for HLSTs should have a broad experience in AI methods, ideally combined with a domain specific background. Experts should hold a PhD from a field like computer science, engineering, environmental science, mathematics, physics, or technology directly related to data science. They will support and collaborate with all Helmholtz Centers, generate performance analysis and benchmarks for selected applications, provide consultancy to scientists as well as optimize and adapt scientific applications. Further, they will participate as trainers in training schools, hackathons and workshops.

Experts for HAICU Central should ideally have strong knowledge of general AI methods (´generalists´) independent of application area.

Experts for local HLSTs instead will ideally be domain-specific, application-oriented data scientists. It is strongly recommended that every research field participates in hosting a local HLST. Thus, there will be a specialist contact person for every research field.

In future, through close cooperations with the proposed Data Science Graduate Schools, we will educate and train our own data science experts.

ATTRACTING HLST EXPERTS (CENTRAL AND LOCAL)

As already mentioned, HAICU will be in tough competition for AI experts. Arguments to attract researchers and AI experts for joining the HLSTs include the diverse datasets involved within the Helmholtz Centers, the possibility to work on diversified research questions and applications and finally excellent working conditions in terms of scientific infrastructure. Nevertheless, it is necessary to gain experts for the HLSTs by offering attractive positions with the following conditions:

- HLST positions should be permanent positions in the long run. They should be offered as 2 years limited positions with the expectation to become permanent after that time.
- HLST experts are supposed to be up to date with state-of-the-art methods. Therefore, they should have resources to regularly participate in international conferences.
- Finally, they should be enabled to work on their own research questions for a predefined fraction of their working time, e.g., 20%.
- Education will be part of their work, they will offer training schools and seminars, in constant exchange with the Data Science Graduate Schools.

As contingency plan, in order to cope with expected difficulties to hire these experts, it is conceivable to partially hire graduates right after university (e.g., from AI study programs), limited to two years, who want to gain a track record in AI at a pulsing research environment.

4.4 EVALUATION

It is of key importance to have a strong reporting line coordinated in HAICU Central as quality measurement and control. Thus, control and monitoring mechanisms on different levels will be installed.

First, HAICU Central will manage and review yearly reports which will be recommended and sent to the Steering Board. The Steering Board in turn will report to the Helmholtz Incubator Information & Data Science which then reports to the Assembly of Members of the Helmholtz Association. These reports will include sections describing the research work of all research groups (including HAICU YIGs) and all HLST support, and thus reflect all vouchers and projects from HAICU Central and Local. Yearly, research groups will adapt their goals in discussion with the management to fit to HAICU's overall goals and current projects. In case of HLST groups, the Central HLST is expected to work at least 75 % for all other Helmholtz Centers whereas local HLSTs are expected to work at least 50 % for other Helmholtz Centers to guarantee cross-domain cooperations. To ensure equal possibilities for all centers the number of vouchers per center shall not exceed more than 15% of the total volume per year. Education, cooperations, scouting, and public relations will be further aspects of the report.

Furthermore, HAICU shall be evaluated regularly by external independent reviewers, the first physical review will be three years after start. The examination procedure will be organized by the Head Office of the Helmholtz Association (Geschäftsstelle) which selects external reviewers and with advice of the SAC. The fundamental criterion for this review is proof of work along the defined milestones (see Section 5).

In detail, for Central research groups evaluation criteria are on the one hand on the basis of a quality assessment of the research group´s scientific achievements according to international standards and an evaluation of the projects´ relevance to the AI field, on the other hand cross-domain work has to be specified. This means, participation at HAICU workshops and conferences, cooperations and education and training have to be measured.

Furthermore, HLSTs (central and local) have to verify the amount of service, support and benchmark help they have offered to the different Helmholtz Centers in the same percentage as for the yearly reviews. Another feature for a positive evaluation have to be regular informal meetings with all other HAICU HLSTs (central and local) for agreements of currently running projects and a yearly meeting for all HLST members where all projects are presented. The Steering Board will adopt measures to evaluate the activities of the HLSTs.

(HAICU pilot) projects have to deliver a final report to HAICU Central to refer to their scientific results and their success. Evaluation criteria of this report will be similar to DFG final reports (DFG 2014).

Evaluation of HAICU YIG will proceed as it is defined for Helmholtz YIG (Helmholtz Association 2017). The evaluation process will be coordinated by the Steering Board in tight agreement with the hosting center.

Scientific directions of HAICU will be monitored by the SAC.

5 IMPLEMENTATION

The described aims will be implemented along the following seven work packages. Figure 4 shows the different work packages and milestones along the schedule.

Work package 1: Initialization and Public Outreach

end of 2018 - end of 2020

The first step for HAICU´s success is building a strong network between all Helmholtz Centers. Such cooperations will create use cases and center overarching projects. As both personnel (2 project referees) and resources for travel expenses are ensured, the implementation of this work package can be started immediately after a positive decision in September 2018.

The project referees will organize internal 'HAICU workshops' for all interested Helmholtz Center researchers to get to know each other, to develop ideas for joint projects, and to find new cooperation partners. The first internal workshop is planned for the end of 2018 and will include invited external/international AI experts (milestone 1.1).

As a follow-up, international workshops will be organized, enabling close collaborations with international research groups and organizations. A first joint workshop between the Canadian Institute for Advanced Research (CIFAR) and the Helmholtz Association is already planned for January 2019 (milestone 1.2).

Constant consolidation and improvement of the network is crucial for success, visibility and ground-breaking new ideas. Therefore, intensive public outreach is one of the most important elements for the initiation phase. A website, social media (Facebook, Twitter, LinkedIN, Xing), a chat channel like Slack and a video lecture channel will be implemented until the end of 2018 (milestone 1.3).

Furthermore, a seminar series with invited internal and external speakers, which will also be accessible online, aims at making HAICU visible, spreading AI knowledge across the community and giving possibilities for fruitful scientific discussions and ideas for collaborative projects. The seminar series will be started in the beginning of 2019 (milestone 1.4).

The first international HAICU conference is of highest importance for international visibility. It is planned for summer 2020 and will give insights into the latest developments in AI and first HAICU project results (milestone 1.5).

The HAICU governance has to be established fast to guarantee quality and steering control. The various committees should be appointed latest in early 2019 (milestone 1.6).

To spread the HAICU concept and gain greater visibility the project referees will visit different conferences important for AI and do scouting in 2019 and the beginning of 2020 (milestone 1.7).

Work package 2: Establishment Phase of HAICU Central

end of 2018 - end of 2020

HAICU Central has to be implemented directly after HAICU's launch since all managing processes are directed from there. For national and international visibility, the selection of the location of HAICU Central is very important. Thus, the call for HAICU Central should be released and organized by the Head Office of the Helmholtz Association right after concept evaluation in September 2018. Every Helmholtz Center is invited to apply for hosting HAICU Central. An external review panel should evaluate the proposals and rank them (milestone 2.1).

After the decision of HAICU Central´s location the hosting Helmholtz Center immediately announces and fills the positions to set up the complete management in agreement with the Steering Board and enables HAICU Central being launched until the third quarter of 2019 (milestone 2.2).

In parallel, first calls for research groups will be announced by the hosting Helmholtz Center starting in the middle of 2019. The applications will be evaluated and ranked by the management team in agreement with the Steering Board in the framework of the Helmholtz Center that hosts HAICU Central. The research part of HAICU Central should be completed by the end of 2020 (milestone 2.3).

Furthermore, positions for HLST members will be announced at the same time as for the research groups. This is of particular relevance as research groups and the HLST will work closely together and the HLST aims at giving support/consulting for all Helmholtz Centers. The first HLST group members should be integrated in HAICU until the end of 2019 (milestone 2.4).

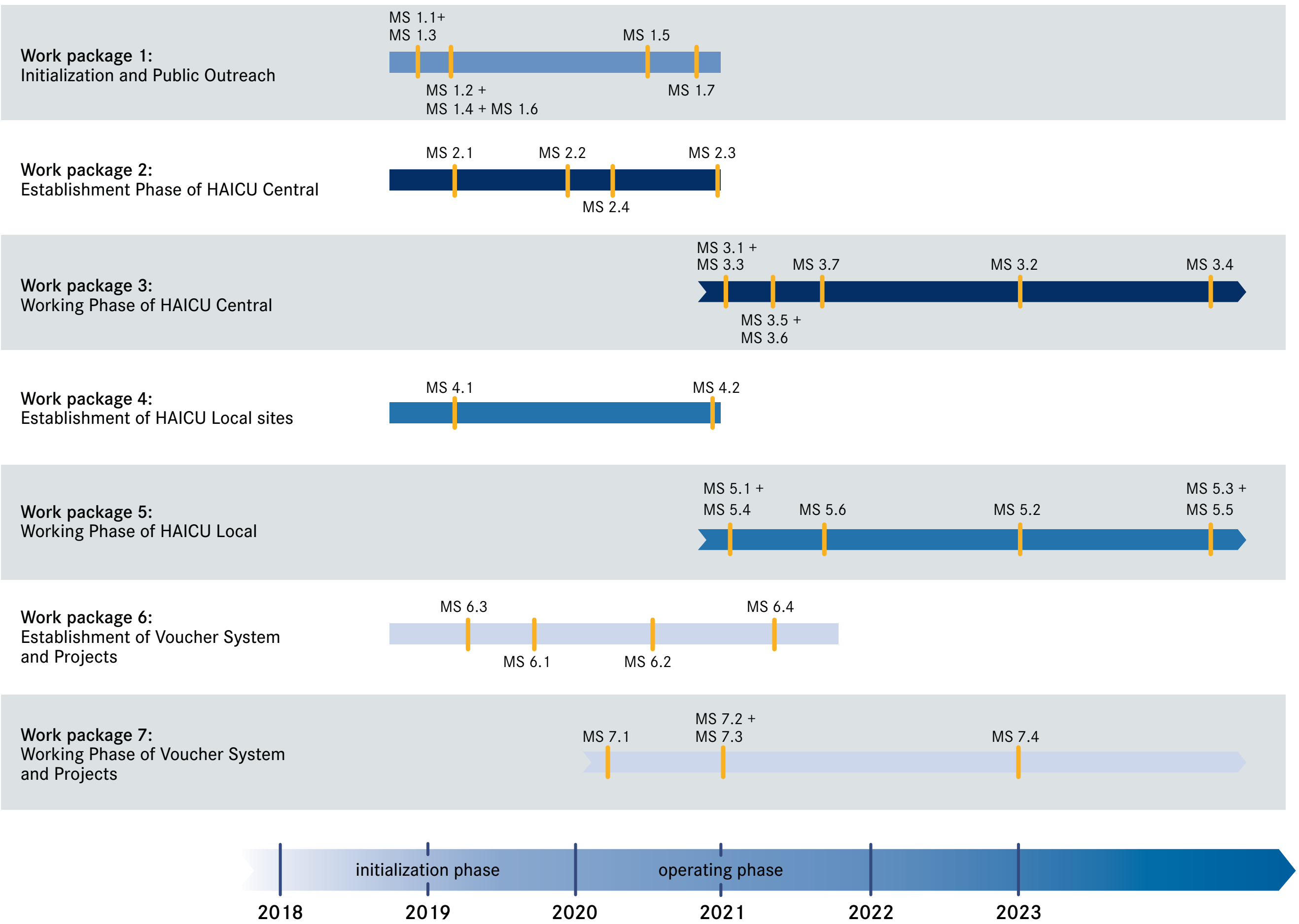

Figure 4: Schedule for work packages and milestones (MS)

Work package 3: Working Phase of HAICU Central

from 2021 on

Having successfully implemented all important components of HAICU Central, it will be able to establish overarching projects and HAICU goals. Implementation quality will be guaranteed by two different evaluation processes: First, HAICU has to write a yearly report for the Steering Board including HAICU Central and HAICU Local (milestone 3.1). Second, HAICU Central and Local together, excluding HAICU YIGs, will be evaluated every three years; HAICU YIGs will have a separate evaluation depending on the start of each group (milestone 3.2). Evaluation criteria are defined in Section 4.4.

Having set up HLST Central and some of the local HLSTs, the yearly HLST meeting will be organized by HAICU Central's Scientific Coordinators, which is important for expert exchange, networking and part of the evaluation (milestone 3.3).

After every five years, junior research groups will be newly appointed to react on possible new AI topics as flexible as possible (milestone 3.4).

An important part of knowledge transfer will be accomplished by inviting internationally recognized AI experts as guest professors starting in 2021 (milestone 3.5).

Another aspect of knowledge transfer and for achieving higher visibility is the visiting researchers program which will start in 2021 as well (milestone 3.6). These last two milestones will help to attract external AI experts to HAICU.

It is expected for all HAICU Central personnel (HLST and research group leader) to educate and train all Helmholtz staff on different levels (milestone 3.7).

Work package 4: Establishment of HAICU Local sites

end of 2018 - end of 2020

The local sites of HAICU are of equal strategic importance for HAICU as HAICU Central. Thus, the call for HAICU local HLSTs and the HAICU YIGs will be released and organized in parallel to HAICU Central by the Head Office of the Helmholtz Association after concept evaluation in September 2018. An external review panel will evaluate the proposals and rank them (milestone 4.1). The Steering Board will decide about the locations of HLSTs and HAICU YIGs. It is strongly recommended that every research field participates in the call and will be represented in a local HLST and a HAICU YIG.

After the decision about the HAICU local sites the hosting Helmholtz Centers immediately announce and fill the positions to set up their HLSTs and HAICU YIGs always in strong interaction with the Steering Board, so all local sites will be launched until the end of 2020 (milestone 4.2).

Work package 5: Working Phase of HAICU Local

from 2021 on

Having successfully implemented all HLSTs and HAICU YIGs, the local part will be able to realize the intended goals and create overarching results by the research groups cooperating closely with the HLSTs. Proof of work in the right direction will be guaranteed by two different evaluation processes as described for work package 2:

First, HAICU has to write a yearly report for the Steering Board including HAICU Central and HAICU Local (milestone 5.1). Second, HAICU (Central and Local together, excluding HAICU YIGs) will be evaluated after every three years (milestone 5.2). HAICU YIGs will have a separate evaluation depending on the start of one group (milestone 5.3). Evaluation criteria are fixed in Section 4.4.

After successful set up of HLST Central and first local HLSTs, a launch for first HLST meetings will be organized where local HLSTs have to participate (milestone 5.4).

After every five years, junior research groups will be newly appointed to react on possible new applied AI topics as flexible as possible (milestone 5.5).

It is expected for all HAICU Local personnel (HLSTs and research group leader) to educate and train all Helmholtz staff on different levels (milestone 5.6).

Work package 6: Establishment of Voucher System and Projects

2019 – 2021

Before starting the voucher system, the best implementation of a booking system has to be identified in agreement with the Initiative and Networking Fund. Thus, the Scientific Coordinators will investigate and implement the best booking system regarding realization and option (e. g., openIRIS that is used at the MDC technology platforms) until the end of 2019 (milestone 6.1).

The voucher system can be launched and first voucher proposals can be submitted as soon as the first HLST has been established (milestone 6.2).

After concept evaluation in September 2018 and establishment of the Steering Board, the first HAICU workshop and the Helmholtz Incubator Information & Data Science Conference will take place at the end of 2018, where all Helmholtz AI scientists will meet, discuss ideas and define possible projects. Directly afterwards, the project referees will announce a call for HAICU pilot projects in accordance with the Initiative and Networking Fund: the initial requirement will be two participating centers from different research fields that exploit synergies in applied AI methods (milestone 6.3).

Successful HAICU pilot projects will be evaluated after 2 years and will serve as lighthouse projects for interdisciplinary research within HAICU across research field borders (milestone 6.4).

Work package 7: Working Phase of Voucher System and Projects

from 2021 on

As soon as HAICU is running at full capacity, having successfully implemented all HLSTs, all Helmholtz scientists are invited to submit vouchers depending on their scientific needs. Vouchers will be distributed among the HLSTs considering the best HLST experts and always under the scope of cross-domain cooperations (milestone 7.1, see Section 2).

Besides allowing a flying start of HAICU pilot projects during the initialization phase, project calls will be released twice a year with a running time of three years starting in 2021 (milestone 7.2).

Evaluation of vouchers and projects will be part of the yearly reports and the evaluation every three years (milestone 7.3 and milestone 7.4).

6 GOVERNANCE

To ensure an efficient and innovative scientific and organizational leadership and based on best practice examples and Helmholtz experiences, we propose to embed HAICU as follows into the Helmholtz Association – in line with the general considerations to embed the platforms into the Helmholtz Association (these principles are described for all platforms in the general opening chapter). HAICU Central and HAICU Local will be located at Helmholtz Centers (host labs). The Helmholtz Centers (host labs) have the disciplinary responsibility for the personnel of HAICU that is located at their Center. They operate HAICU in accordance with the goals and services agreed to by the host labs within the Helmholtz Association. This is illustrated in Figure 5.

The Steering Board is responsible for the general development of the platform. It coordinates and leads the activities of HAICU Central and Local and it is responsible that HAICU acts in the interest of the whole Helmholtz Association. It reports to the Helmholtz Incubator which then reports to the Assembly of Members of the Helmholtz Association. It shall synchronize all activities of HAICU, including projects, the voucher system, the visiting researchers program, local HAICU YIGs and HLSTs and HAICU Central. The Steering Board prepares all decisions regarding funding from the Initiative and Networking Fund after a review by the SAC. The President of the Helmholtz Association makes the funding decisions (in line with the overall regulations of the Initiative and Networking Fund).

The members of the Steering Board will be selected by the Assembly of Members of the Helmholtz Association. It shall consist of 7-9 Helmholtz internal scientists; one important point is to ensure that every research field is represented. Its members shall be representatives from the Helmholtz Incubator Information & Data Science and shall include the two HAICU rapporteurs and guest representatives from the Head Office of the Helmholtz Association (Geschäftsstelle). They will meet four times a year (physically and/or via video conference).

The Scientific Advisory Committee (SAC) is responsible for steady advice of scientific development. It asserts that HAICU works along the state of the art and produces relevant scientific output. The SAC shall give advice in all cases related to calls and funding schemes, especially the proposals for the visiting researchers program, projects and vouchers. The SAC shall be able to determine scientific guidelines for the funding of small activities (like the vouchers) that do not exceed a certain amount (to be determined with the Head Office of the Helmholtz Association). In this case, the Steering Board can make the funding recommendations without the review of the SAC. However, the SAC must evaluate all larger project proposals. Also, it regularly assesses HAICU´s scientific activities based on the yearly reports. It can be included in the first review of HAICU after three years and further reviews. The members of the SAC shall be appointed by the President of the Helmholtz Association. The SAC shall consist of five to seven well known international researchers. SAC meetings will be twice a year (physical and video conference).

In addition to the embedding of all platforms into the Helmholtz Association, there shall be a further advisory board directly invited by the HAICU scientists. The Circle of Fellows shall have a commitment to the HAICU initiative, give scientific advice and directions and mentor the HAICU YIGs. This board is of great importance for HAICU in terms of advertisement and having honorable representatives for HAICU. The Circle of Fellows shall be established jointly by HAICU Central and HAICU Locals. Thereby the Circle of Fellows strengthens the HAICU network and helps to disseminate HAICU's developments and methods.

In later stages, a Board of Trustees might be integrated into the governance for steady advice for strategic development with respect to industrial and societal impact. Members of the Board of Trustees might be invited by HAICU Central and HAICU Locals. It could help HAICU to generate output for technology transfer and innovation. The Board of Trustees could consist of representatives from politics, industry, science and media (e.g., Intel, IBM, NVIDIA, SAP, Atos, Bosch, Microsoft, Hellsicht, Bayer, Max Planck, Fraunhofer) which might meet once a year.

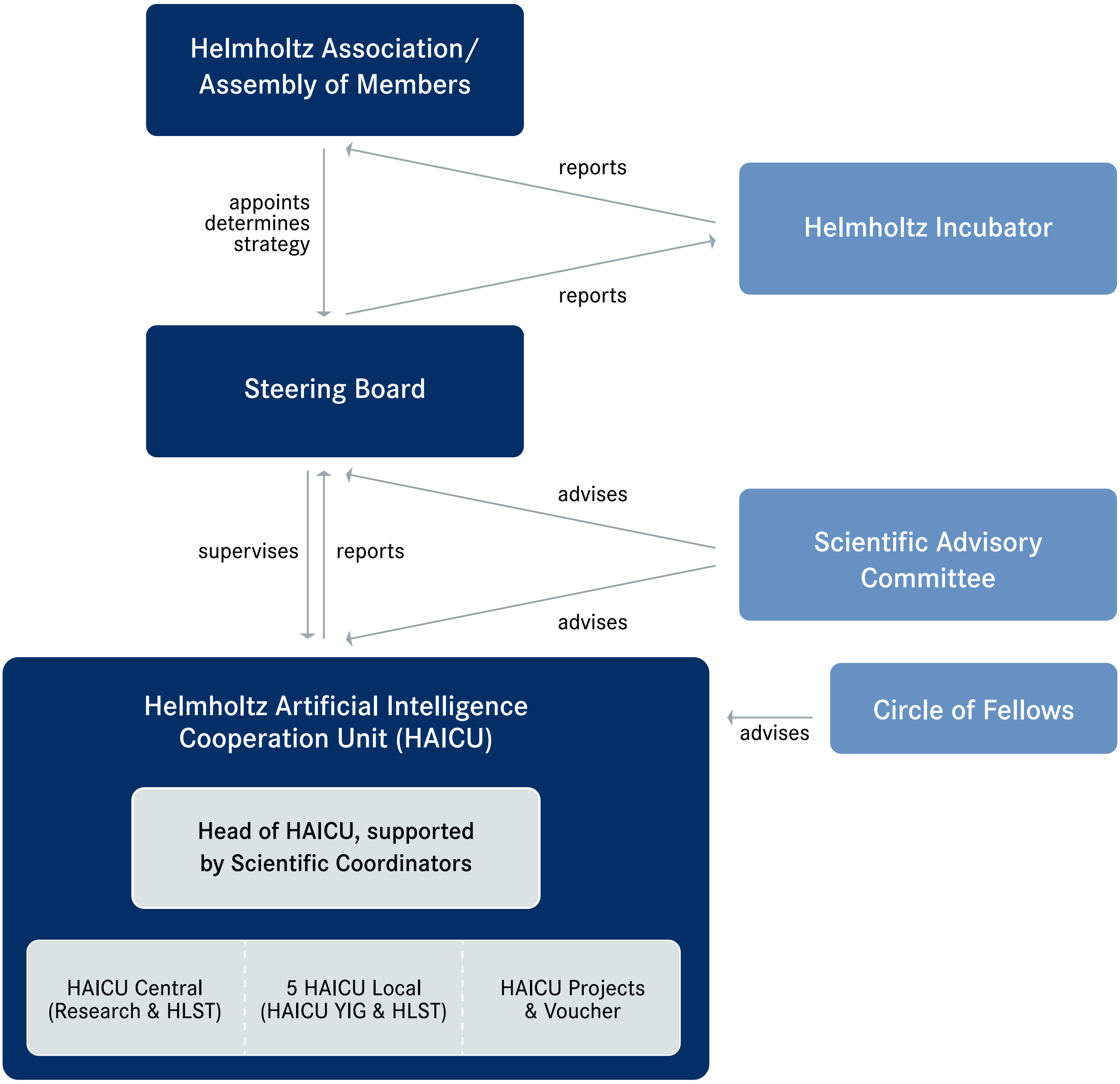

Figure 5: Schematic illustration of the integration of HAICU into the Helmholtz Association. The platform is integrated into the operational and disciplinary authority of the Helmholtz Centers that host HAICU Central and HAICU Locals (not shown in the illustration).

7 FINANCES

We propose the following financing for an excellent HAICU structure under full capacity:

The total costs for HAICU will be 11.36 M €/a at full capacity, wherefrom costs for HAICU Central will be 4.00 M €/a and costs for HAICU Local will be 4.36 M €/a. For Helmholtz-wide projects 2.70 M€/a is foreseen plus 300 k€/a for HAICU events.

The costs will be divided into two parts and separated into static costs (for HAICU Central and Local HAICU YIGs and HLSTs, respectively) and dynamic costs (including vouchers and projects). Static costs amount to a sum of 8.66 M €/a at full capacity, with personnel costs of 7.54 M €/a and material costs of 1.12 k €/a. Dynamic costs will account for 2.7 M €/a.

Table 1 summarizes all costs. Further details are given in section 9.5.

Table 1: Aggregated cost representation according to the formal specification of the Head Office of the Helmholtz Association. It shows the expected costs per year for the fully established platform, broken down into personnel (with overhead) and material costs (both basic funded) and project costs (IVF-funded). Estimates of the working group underlying this presentation are found in Section 9.5.

HAICU FINCANCES

platform part field of engagement *cost type*	**costs p.a. in k€**
HAICU central	
Management, Controlling, etc.	
nonpersonnel costs	69
personnel costs	274
Scientific Coordination	
nonpersonnel costs	98
personnel costs	350
Core Team	
nonpersonnel costs	262
personnel costs	1,982
High Level Support Team	
nonpersonnel costs	90
personnel costs	874
sum	**3,999**
HAICU Local (5 units)	
Core Teams	
nonpersonnel costs	75
personnel costs	1,875
High Level Support Teams	
nonpersonnel costs	225
personnel costs	2,184
sum	**4,359**

HAICU projects & vouchers	
HAICU Voucher System	
Project Funds	700
HAICU Projects	
Project Funds	2,000
sum	**2,700**
Events	
HAICU events	
nonpersonnel costs	300
sum	**300**
total p.a.	**11,358**
thereof:	
nonpersonnel costs (basic funding)	1,119
personnel costs (basic funding)	7,539
project funds (IVF funding)	2,700

More accurate separation between HAICU Central and HAICU Local will be explained in the next sections.

Costs for HAICU Central of 4.0 M€/a are divided into personnel costs which account for 3.5 M€/a and material costs in the range of 500 k€/a.

Personnel keys for HAICU Central are proposed as follows:

- HAICU Central: 38 FTEs
    - management: 7 FTEs
    - 2 senior research groups (5 FTEs each)
    - 3 junior research groups (3 FTEs each)
    - 10 HLST members
    - guest professorship program: 2 FTEs floating

300k€/a are planned for invitation, outreach and meetings.

Costs for HAICU Local will be 4.36 M€/a. In detail, these costs are divided into personnel costs which account for 4.06 M€/a (including overhead) and material costs in the range of 300 k€/a.

Personnel keys and project financing for HAICU Local are proposed as follows:

- HAICU Local: 40 FTEs
    - 5 HLSTs (5 FTEs each)
    - 5 HAICU YIGs (3 FTEs each)

2.70 M€/a shall be used for vouchers and projects

- 700k €/a voucher system
- 2.0 M €/a for scientific project calls (including testing /investment of innovative hardware)

An overview of the resource allocation will be given in Figure 6.

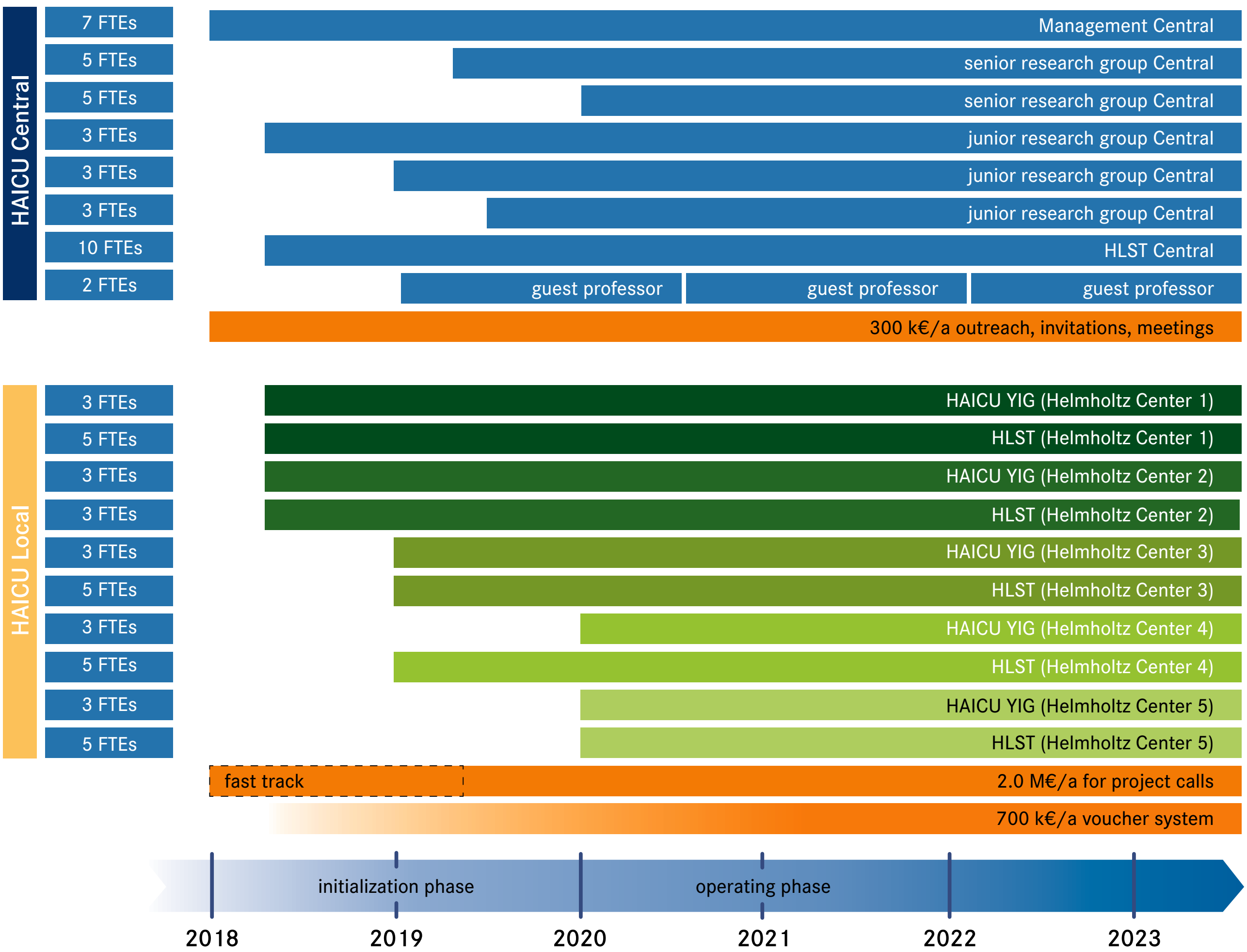

Figure 6: HAICU Resource Allocation

8 FUTURE PERSPECTIVE AND FUTURE USE CASES

With the establishment of HAICU, the Helmholtz Association aims to bring the excellent expertise in Helmholtz Centers and programs together and to achieve a pioneering role and international competitive leadership position in the field of applied AI in Germany. The international visibility of the Helmholtz Association will increase, as we plan large international cooperations with nationally and internationally distinguished experts. Following successful implementation of the HAICU structure and first project calls, universities, other research institutions, and industry will be involved in projects in the next step. Furthermore, through corporations, we will drive forward technology transfer and envisage to create spin-offs.

Throughout continuous discussion with researchers from all Helmholtz Centers, the following potential future use cases were identified, exemplarily:

- Across different domains, simulations of e.g., physical, chemical, or climate models, depend on a potentially large number of parameters, while running a suited number of simulations in order to find the best solution is very time consuming. Identification of best parameter sets using AI methods is thus beneficial for a broad range of application and research domains. Further, in model-based sciences, e.g., atmospheric modeling, radiation transport, atmospheric chemistry, weather forecast, or the Ocean's biogeochemistry, where observations are taken at distinct model positions, AI methods may help to adjust the measurement procedure such that best suited positions for those measurements are learned and predicted, respectively. Similarly, AI methods might be used for the automatic quality control of measurement data and its control parameters. Up to now this is done manually by experts, often as a post-processing step. In case of time critical applications, e.g., offshore measurements, geophysical measurements in case of natural hazards, or real time applications for particle detector operations, an online interpretation is essential. Finally, recent developments in DL will allow to exchange the computationally most expensive parts of complex simulation models by ANNs.

- In order to research the stratospheric impact on surface climate and temperature actual historical data can be combined with a simulation ensemble output. Therefore, the determination of the most likely simulation ensemble state is needed. This task might be addressed by connecting experts from climate and atmospheric simulation and experts for classical density estimation and DL methods.
- In plasma physics early detection of so-called disruptions is essential in order to ensure the reliable operation of an experiment. Experimental data on disruptions in large devices are difficult to get, and therefore disruption studies tend to be limited to comparably small devices. The question here is whether an AI model can be trained on existing limited data about disruptions (on cheap devices) and can be used for forecasting on very expensive devices. Further, additionally to the limited training data, there is even not enough understanding on disruption to build a model for simulating training data. This task can be referred to as an example for extreme transfer learning and requires experts both from plasma physics and fundamental AI methods for transfer learning.
- In the field of Earth System modeling and Remote Sensing global scale information extraction from multi-sensor data in near real time, the analysis of spatiotemporal dynamics, and finally, coupling of remote sensing analysis with modeling and other monitoring methods for more complete process understanding will be essential in order to deal with applications for multi-hazard analysis, surface characterization, or surface dynamics analysis, and presumes again transdisciplinary cooperation of domain scientists and experts in according AI methods.
- In Retro synthesis, with synthetic chemistry as the key to novel materials, intelligent procedures to reliably predict the outcome of a reaction or the feasibility of a synthesis to a target molecule do not exist so far. Here AI methods are needed for the translation of information on reaction data into suitable models. Such models will allow for the first time to reliably evaluate the synthetic feasibility of target molecules for an automated evaluation of computationally designed materials or biologically active molecules. Here a chemist / chemo-informatician and an AI expert will collaborate in order to develop the necessary chemistry models including the automated curation of the available datasets, on one hand, and the design of the appropriate DL method and its training on the other hand. Furthermore, in reaction prediction several ML based methods for structure processing were presented in the past, but none of them included the most important information for the outcome of chemical reactions and none of them fulfilled requirements in practice. Here novel models are necessary for fusing conditions like temperature, yield, and duration and an adaption of models to dependencies in their 3D structure and physical nature is needed, yielding the transition from virtual datasets to experimental data.
- Established DL methods for speech and song recognition might be transferred to a broad range of applications relevant for the Helmholtz Association: in Seismology it can be used for the detection of

earthquakes, tremors, landslides, and noise classification; in Oceanography anomalous wave patterns and ship noise can be analyzed; in Biology mammals, e.g., wales could be localized through hydrophone data; anomaly detection in wind and temperature data, shot and explosion detection in urban areas from acoustics data, or meteoroid impact detection in atmosphere data are further proposals.

9 ATTACHMENT

9.1 CONSTRUCTION PHASE

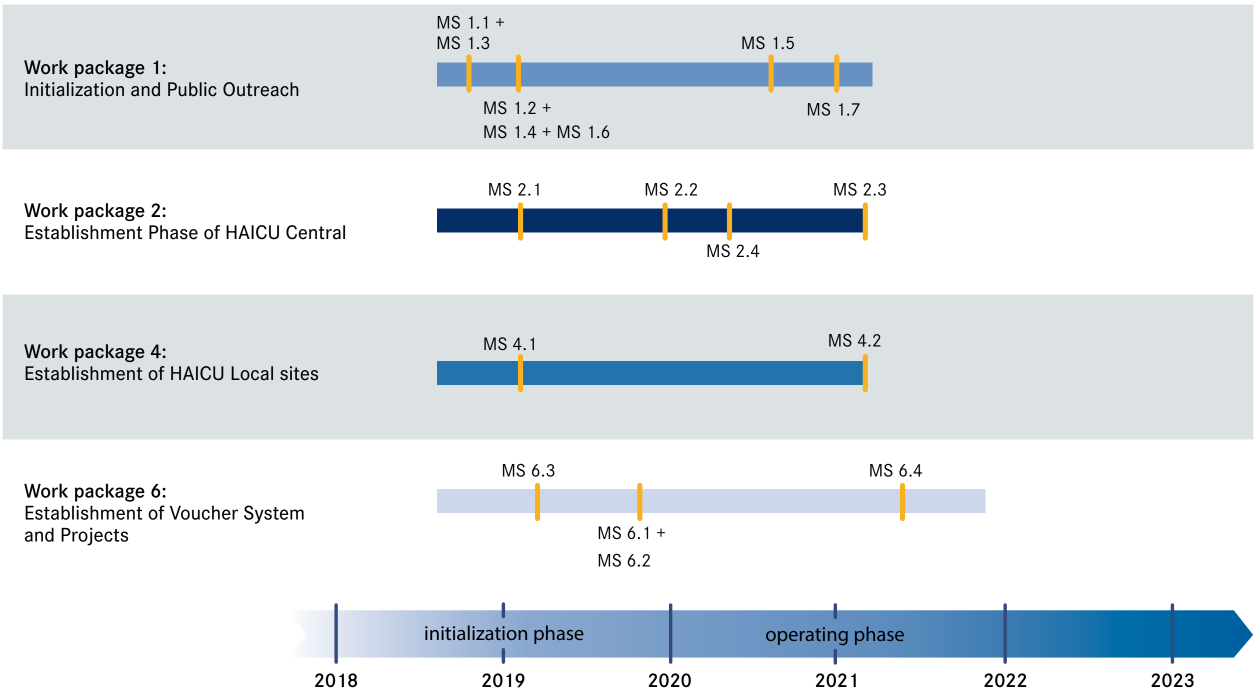

Figure 7: **Initialization phase:** Work packages and milestones (MS)

MS 1.1: internal workshop at the end of 2018; **MS 1.2:** first joint workshop between CIFAR and the Helmholtz Association in January 2019; **MS 1.3:** website, social media, a chat channel similar to Slack and a video lecture channel implementation until the end of 2018; **MS 1.4:** seminar series with start in the beginning of 2019; **MS 1.5:** first international HAICU conference in summer 2020; **MS 1.6:** governance establishment latest in early 2019; **MS 1.7:** scouting of project referees in 2019 and the beginning of 2020; **MS 2.1:** call for HAICU Central by the Helmholtz Association‘s Head Office right after concept evaluation in September 2018; **MS 2.2:** set up of HAICU´s management until the third quarter of 2019; **MS 2.3:** research groups implementation of HAICU Central until the end of 2020; **MS 2.4:** HLST implementation until the end of 2019; **MS 4.1:** call for HAICU Central by the Helmholtz Association‘s Head Office right after concept evaluation in September 2018; **MS 4.2:** HLSTs and HAICU YIGs implementation until the end of 2020; **MS 6.1:** investigation and implementation of voucher system by the Scientific Coordinators until the end of 2019; **MS 6.2:** start of voucher system realization; **MS 6.3:** call for HAICU pilot projects at the beginning of 2019; **MS 6.4:** evaluation of HAICU pilot projects after 2 years in 2021

9.2 INTERNAL AND EXTERNAL ENVIRONMENT ANALYSIS

The process for the development of the present concept goes back to the first Helmholtz Incubator Information & Data Science workshops where first needs in terms of Information and Data Science were defined. In September 2017, the Helmholtz Assembly of Members decided for the five working groups, among others HAICU, to develop comprehensive concepts for platforms. In preparation of the 4th Helmholtz Incubator Information & Data Science meeting in March 2018, a first assessment of existing expertise, needs and requirements have been requested by email and phone calls among all Helmholtz Centers and participating researchers from all research fields. As a result, the first rough concept was presented at the March‘s workshop and published as whitepaper across the Helmholtz Association.

As already mentioned throughout the paper, rapporteurs, project referees, and members of the whole working group were collaborating closely during the development of the presented concept, through center visits, video conferences, physical meetings, and email communication. This enables to provide a broad overview of existing expertise in the field of AI and the needs the researchers expressed (not necessarily comprehensive), presented in Section 2.3 and summarized in Tables 1 and 2.

Externally, an intensive research and investigation of existing research centers dealing with Data Science and Artificial Intelligence and focusing on cooperations of researchers identified institutes that are worth observing: the Canadian Institute for Advanced Research (CIFAR) and the Vector Institute Toronto in Canada; the Santa Fe Institute and the Moore-Sloan Data Science Environment, the MIT Computer Science & Artificial Intelligence Lab (CSAIL) and Center for Data Science at NYU in New York in the US; the Broad Institute and the Alan Turing Institute London, partnership of DeepMind, UCL, Oxford and Cambridge, EMBL-EBI (Hinxton) in the UK; the Institute for Advanced Study IAS of TUM, the Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), the CyberValley of Max Planck Campus Tübingen, and the European Molecular Biology Laboratory (EMBL) in Heidelberg in Germany. Recently granted competence centers for AI at Berlin, Dortmund/St. Augustin, München, and Tübingen will be established now and are funded by the BMBF with €30 million.

From our research we would like to highlight the Whitepaper from The Moore – Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington (msdse 2017), which describes in detail their experiences in creating such an environment including their requirements, their efforts to meet them, pitfalls and solutions.

Inspected research institutes mainly follow joint goals. First of all, they aim on bringing forward research by educating young researchers and attracting excellent minds to join their groups. They ensure collaboration on several levels of interaction.

For educating young researchers they partially define own study programs and offer seminar and lecture series, summer/fall schools. In order to attract excellent minds, they define hot topic research fields and create research groups related to these fields, with highly regarded staff, which are internationally visible. They offer mentoring programs through well-known fellows and ‘smooth’ conditions for working. In order to ensure collaboration, they organize workshops and conferences and visiting researchers programs. Further, for the sake of technology transfer they ensure cooperations between science and industry. All of them have very close links to universities. Further, all of them have a high visibility through webpages and social media in order to gain interest from potential new members or collaboration partners. Their leadership is mostly organized on several levels, having a board of directors, a board of trustees, a scientific advisory board, and a number of fellows. Usually they have one or two directors, either as full time and permanent positions or as double affiliation, i.e., with a professorship at a university. The board of trustees typically consists of representatives from politics, industry, science, and media and is responsible for a steady advice for strategic development. The scientific advisory board typically consists of a number of honorable scientists and is responsible for a steady input for scientific directions and ensures links to institutes all over the world. Fellows on the one hand serve as mentors for young as well as senior researchers, therefore attract excellent minds, and on the other hand may profit by themselves from the title.

What we gain from these examples is a broad experience in the setup and organization of data analysis and AI research institutes, respectively. However, we realize that hardly any nationally or internationally unifying concept/institution exists on a large-scale level; institutions are mostly method-focused or on a comparatively narrow application area, and mostly locally concentrated. We do not identify a unifying research organization across a broad range of research fields as the one the Helmholtz Association does.

Table 1: AI competences at Helmholtz with relevant high impact publications

	Deep Learning	Classical ML	Data Mining	Simulation	Model-based Approaches	Coupling model-based (Simulation) and data-driven Approaches
ML Fundamentals	**DESY, FZJ** (Jitsev and von der Malsburg 2009; Jitsev 2014), **DLR** (Zhu et al. 2016; Mou, et al. 2017)	**KIT** (Schlitter et al. 2013; Schüller and Benz 2018), **FZJ** (Krajsek and Scharr 2010; Scharr, Black, and Haussecker 2003), **GEOMAR, DESY, DLR,** **HMGU** (Theis 2006), **DKFZ** (Heim et al. 2017), **MDC** (Lee et al. 2017)	**KIT, GEOMAR, DLR** (Datcu et al. 2003)	**DLR** (Auer et al. 2017)	**DLR** (Zhu et al. 2016; Wang et al. 2016), **DKFZ** (Koch et al. 2013)	**GEOMAR, DLR**
Genomics, Transcriptomics	**MDC, KIT, UFZ, HMGU,** **HZI** (Bernard et al. 2018), **DKFZ** (Angermueller et al. 2017)	**MDC** (Calviello et al. 2016; Hafez et al. 2017), **HZDR,** **DKFZ** (Svensson et al. 2018; Argelaguet et al. 2018), **HZI** (Weimann et al. 2016; Patil et al. 2011; Asgari et al. 2018), **DZNE,** **UFZ** (Victor de Araujo Oliveira et al. 2016), **FZJ** (Lu et al. 2012), **HMGU** (Haghverdi et al. 2016)	**MDC** (Schwarz et al. 2014, 2015; Turajlic et al. 2018), **HZDR, GEOMAR, HMGU** (Sass et al. 2013), **DKFZ** (Chudasama et al. 2018; Delacher et al. 2017)	**HMGU**	**HMGU**	**HMGU, DKFZ** (Buettner et al. 2017)
Neuroscience	**DKFZ** (Kleesiek et al. 2016) **FZJ** (Morita et al. 2016; Spitzer et al. 2017; Spitzer et al. 2018)	**FZJ** (Lefort-Besnard et al. 2018; Nostro et al. 2018), **HMGU** (Viader-Llargués et al. 2018), **DKFZ** (Neher et al. 2017)	**FZJ** (Jannusch et al. 2017; Varikuti et al. 2018)	**FZJ** (Ippen et al. 2017; Kunkel and Schenck 2017)		**HMGU**
Image Analysis	**DKFZ** (Bernard et al. 2018; Ross et al. 2018), DZNE, **FZJ** (Hasasneh et al. 2018), **HMGU** (Eulenberg et al. 2017)	**DKFZ** (Norajitra and Maier-Hein 2017), **DZNE, HMGU, GEOMAR** (Schoening et al. 2016), **FZJ** (Rehme et al. 2015; Wagstyl et al. 2018; Yeo et al. 2010), **MDC** (Preibisch et al. 2014)	**GEOMAR**	**GEOMAR**	**FZJ** (Amunts et al. 2013)	**DKFZ** (Kirchner et al. 2018; Wirkert et al. 2016)

	Deep Learning	Classical ML	Data Mining	Simulation	Model-based Approaches	Coupling model-based (Simulation) and data-driven Approaches
(Bio) Medical Research	**MDC** (Telenti et al. 2018), **DKFZ** (Angermueller et al. 2016)	**HZDR, HZI, MDC** (Calviello et al. 2016; Mukherjee et al. 2017; Lippert et al. 2017), **DKFZ** (Fusi et al. 2014)	**MDC** (Schwarz et al. 2014, 2015; Turajlic et al. 2018), **HZDR, HMGU, HZI, DKFZ** (Oakes et al. 2016; Gröbner et al. 2018; Northcott et al. 2017)	HMGU	HMGU	HMGU
Drug Research		HMGU	HMGU	**FZJ** (Casasnovas et al. 2017; Sena et al. 2017; Weichsel-baum et al. 2017)	HMGU	**HMGU** (Dawidowski et al. 2017)
Materials	KIT, HZG	**KIT** (Smith et al. 2017)		HZB, KIT, HZG	HZB, KIT	HZG, KIT
Photon Science / X-ray Tomography	**DESY** (Yang et al. 2018)			HZB	HZB	
Space Physics	GFZ	**GFZ** (Zhelavskaya et al. 2017)		**GFZ** (Camporeale et al. 2016)	GFZ	**GFZ** (Shprits et al. 2013)
Particle and Astroparticle-, Plasma-, Accelerator Physics	**HZDR, IPP, DESY** (Erdmann et al. 2018; CMS Collaboration 2018)	**HZDR, IPP, DESY** (ATLAS Collaboration 2018; Belle Collaboration 2017)		HZB, HZDR, IPP, DESY	HZB, HZDR, IPP, DESY	DESY
Earth System Modelling		**AWI, DLR, HZG** (Staneva et al. 2016)	GEOMAR, HZG	AWI, KIT, DLR, GEOMAR, HZG	GFZ, KIT, DLR, AWI, GEOMAR, HZG	GEOMAR, HZG
Climate Modeling / Atmosphere	KIT, GFZ	KIT, GFZ, HZG, AWI, DLR	GEOMAR, HZG	HZG, AWI, DLR, GEOMAR	HZG, AWI, GEOMAR	GEOMAR, HZG

	Deep Learning	Classical ML	Data Mining	Simulation	Model-based Approaches	Coupling model-based (Simulation) and data-driven Approaches
Geology	GFZ	**GFZ, GEOMAR** (Peukert et al. 2018; Alevizos et al. 2018)	GEOMAR	GEOMAR		GEOMAR
Geophysics		**UFZ, GEOMAR, GFZ** (Bauer et al. 2012)	**GFZ** (Köthur et al. 2014)	UFZ, GFZ, GEOMAR	GFZ, GEOMAR	
Biology / Bioinformatics	UFZ, GEOMAR, DKFZ, HZI, DZNE, HMGU, FZJ	**UFZ, GEOMAR, DKFZ, HZI** (Gregor et al. 2016), **DZNE** (Alioto et al. 2015), **UFZ** (Hertel et al. 2014), **HMGU, MDC** (Mukherjee et al. 2017; Hafez et al. 2017)	**HMGU, HZI, DKFZ** (Werner et al. 2015)	HMGU	HMGU	
(Organic) Chemistry	KIT, UFZ, GFZ	**KIT** (Mikut et al. 2016), **UFZ, GFZ**		KIT, GFZ		
Plant Science	**FZJ** (Giuffrida et al. 2017)	HMGU	HMGU			HMGU
Biodiversity	UFZ, GEOMAR	AWI, GEOMAR	GEOMAR			
Microscopy	**KIT, HMGU** (Mishra et al. 2016; Buggenthin et al. 2017)	**KIT** (Ulman et al. 2017), **MDC** (Hörl et al. 2018)			HMGU	HMGU
3D Reconstruction	FZJ	**KIT, MDC** (Karaiskos et al. 2017)	GEOMAR	GEOMAR	GEOMAR	
Remote Sensing	**KIT, GEOMAR, GFZ, DLR** (Marmanis et al. 2016; Babaee et al. 2016; Mou et al. 2017)	**KIT, GEOMAR, GFZ** (Steinberg et al. 2016; Neumann et al. 2015), **FZJ** (Cavallaro et al. 2015; Memon et al. 2017), **DLR** (Babaee et al. 2016)	GEOMAR, UFZ, DLR			GEOMAR
Robotics		GEOMAR, DLR	GEOMAR	GEOMAR	GEOMAR	GEOMAR
Oceanography	GEOMAR	**GEOMAR** (Koszalka et al. 2011), HZG (Wahle et al. 2015)	GEOMAR	**GEOMAR** (Kriest et al. 2017)	GEOMAR, HZG	GEOMAR, HZG

Table 2: Methodological Needs

	DL	Explainable AI	Including Prior Knowledge, Model vs. Data-driven	Uncertainty Quantification
ML Fundamentals	HMGU, DLR, DKFZ, DESY, FZJ, IPP	HMGU, DLR, DKFZ, GEOMAR, DESY, FZJ	HMGU, DKFZ, DLR, GEOMAR, DESY, FZJ, IPP	HMGU, KIT, DKFZ, DLR, GEOMAR, DESY, FZJ, IPP
Genomics, Transcriptomics	MDC, GEOMAR, DKFZ	DKFZ, GEOMAR	DKFZ, GEOMAR	DKFZ, GEOMAR
Neuroscience	DKFZ, FZJ	DKFZ, FZJ	DKFZ, FZJ	DKFZ, FZJ
Image Analysis	HMGU, DKFZ, GEOMAR, FZJ, HZG, IPP	HMGU, DKFZ, GEOMAR, FZJ, HZG	HMGU, DKFZ, GEOMAR, FZJ, HZG	HMGU, DKFZ, GEOMAR, FZJ, HZG, IPP
(Bio) Medical Research	MDC, DKFZ, FZJ	DKFZ	DKFZ	DKFZ
Drug Research	MDC, HMGU, HZI, DKFZ, DZNE, GEOMAR	MDC, HMGU, HZI, DKFZ, DZNE, GEOMAR	MDC, HMGU, HZI, DKFZ, DZNE	MDC, HMGU, HZI, DKFZ, DZNE, GEOMAR
Materials	HZB, HZG, GEOMAR, FZJ	GEOMAR, HZG	HZB, KIT, HZG	HZG, GEOMAR, FZJ
Photon Science / X-ray Tomography	HZB, DLR, HZB	DLR	HZB, DLR, HZB	DLR
Space Physics	GFZ	GFZ		GFZ
Particle-, Plasma-, Accelerator Physics	HZB, HZDR, IPP	IPP	HZB, HZDR, IPP	
Earth System Modeling	GEOMAR , HZG, FZJ	AWI, GFZ, GEOMAR, HZG, FZJ	GEOMAR, HZG	GEOMAR, HZG
Climate Modeling / Atmosphere	GFZ, GEOMAR, HZG, FZJ	GFZ, AWI, GFZ, GEOMAR, HZG, FZJ	GEOMAR, HZG	HZG, GEOMAR
Geology	GEOMAR	GEOMAR	GEOMAR	GEOMAR
Geophysics	UFZ, GFZ, GEOMAR	UFZ, GFZ, GEOMAR	UFZ, GEOMAR	UFZ, GEOMAR
Seismology	GFZ	GFZ	GFZ	
Biology / Bioinformatics	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG
(Organic) Chemistry	KIT		KIT, GFZ	
Biodiversity	GEOMAR		GEOMAR	GEOMAR
Microscopy	KIT, FZJ	FZJ	FZJ	FZJ
3D Reconstruction	GEOMAR, FZJ, HZG	FZJ	GEOMAR, FZJ, HZG	GEOMAR, FZJ, HZG
Remote Sensing	GFZ, GEOMAR, DLR, FZJ, HZG	GFZ, DLR, GEOMAR, FZJ, HZG	GEOMAR, DLR, FZJ, HZG	GEOMAR, DLR, FZJ, HZG
Robotics	GEOMAR, DLR	GEOMAR, DLR	GEOMAR, DLR	GEOMAR, DLR
Oceanography	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG

Scalable and Distributed ML and DL in Extreme-scale Computing	Quality Control / Real Time Parameter Tuning	Time Series Analysis	Data Fusion	Efficient Learning Strategies / Learning from Sparse Data
KIT, DKFZ, DLR, GEOMAR, DESY, FZJ, IPP	DKFZ, DLR, GEOMAR, DESY, FZJ	HMGU, DKFZ, DLR, GEO-MAR, DESY, FZJ, IPP	DKFZ, DLR, GEOMAR, DESY, FZJ, IPP	DKFZ, DLR, GEOMAR, DESY, FZJ, IPP
FZJ, DKFZ, GEOMAR	DKFZ	DKFZ	DKFZ	DKFZ, FZJ
FZJ, DKFZ	DKFZ	DKFZ	DKFZ, FZJ	DKFZ, FZJ
HMGU, DKFZ, GEOMAR, FZJ, HZG, IPP	HMGU, DKFZ, GEOMAR, HZG	HMGU, DKFZ, GEOMAR, HZG	HMGU, DKFZ, GEOMAR, FZJ, HZG	HMGU, DKFZ, FZJ, GEO-MAR, HZG, IPP
DKFZ	DKFZ	DKFZ	DKFZ	DKFZ, FZJ
MDC, HMGU, HZI, DKFZ, DZNE, FZJ	MDC, HMGU, HZI, DKFZ, DZNE	FZJ,MDC, HMGU, HZI, DKFZ, DZNE	MDC, HMGU, HZI, DKFZ, DZNE	MDC, HMGU, HZI, DKFZ, DZNE, FZJ
HZB, KIT		HZG	HZB	FZJ
HZB		DLR	DLR	DLR
FZJ			GFZ	
IPP				
GFZ, KIT, DLR, AWI, GEOMAR, HZG, FZJ	GEOMAR, HZG	GEOMAR, HZG	DLR, GEOMAR, HZG	GEOMAR, HZG, FZJ
HZG, AWI, DLR, GEOMAR, FZJ	GEOMAR, HZG	GFZ, GEOMAR, HZG	GFZ, DLR, GEOMAR, HZG	GFZ, GEOMAR, HZG, FZJ
	GEOMAR	GEOMAR	GEOMAR	GEOMAR
	GEOMAR	GEOMAR	GEOMAR	GEOMAR
	GEOMAR			
DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG	DKFZ, GEOMAR, HZG
		GEOMAR	GEOMAR	GEOMAR
FZJ	FZJ	FZJ	FZJ	FZJ
FZJ	GEOMAR, FZJ, HZG	GEOMAR, HZG	GEOMAR, FZJ, HZG	GEOMAR, FZJ, HZG
DLR, GFZ, DLR, FZJ, HZG	GEOMAR, DLR, FZJ, HZG	GFZ, GEOMAR, DLR, FZJ, HZG	GFZ, DLR, GEOMAR, FZJ, HZG	GFZ, DLR, GEOMAR, FZJ, HZG
DLR	GEOMAR, DLR	GEOMAR, DLR	GEOMAR, DLR	GEOMAR, DLR
GEOMAR, DLR, HZG	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG	GEOMAR, DLR, HZG

9.3 SWOT ANALYSIS

For the SWOT analysis we defined our strengths, weaknesses, opportunities and risks in internal discussions and meetings and during our Helmholtz Center visits, summarized in Figure 8.

Strengths
1. Helmholtz Association and German-wide initiative to strengthen AI
2. Institutionalization of the collaboration between Helmholtz Centers
3. Enables cross-domain research and application, developing common, basic methodologies
4. Makes use of unique, high-quality and large data sets
5. Service and research under one roof

Weaknesses
1. Increases coordination and communication necessary to link local and central units
2. New efforts necessary to coordinate
3. Heterogeneous data sets in terms of quality, accessibility, format, etc.
4. High requirements to experts with the need to oversee a large area (in particular HLSTs)
5. Local and Central parts are supported in different ways

HAICU

Opportunities
1. Achieve a leadership position at the international level; new partnerships to be expected in Germany and worldwide
2. Overcomes isolated efforts and creates a critical mass in AI
3. Creates synergies between research topics from different domains
4. Scientific breakthroughs feasible
5. High international visibility of HAICU and Helmholtz Association

Risks
1. Efforts perhaps not big enough and too late
2. Collaboration between different centers on overarching topics is not successful
3. Infrastructure for exchange and analysis of large-scale data including legal and ethical requirements not available
4. Not enough qualified personal and junior staff available
5. Some centers may not benefit

Figure 8: SWOT Analysis

9.4 COOPERATION WITH OTHER HELMHOLTZ INCUBATOR INFORMATION & DATA SCIENCE ACTIVITIES

HAICU will maintain explicit links to other Helmholtz Incubator Information & Data Science activities, in order to contribute to the success of the Helmholtz Incubator Information & Data Science as a whole.

First of all, knowledge transfer and educating young researchers is closely linked to HIDA. HAICU will ensure training of young researchers by tutorials, training schools and summer schools on specific fundamental topics or on specific topics requested by the Helmholtz researchers. Furthermore, complementary courses for all HAICU researchers will be offered. Cooperation with HIDA has to be twofold: (1) HIDA coordinates all data science training offers within the Helmholtz Association, thus avoiding redundant offers. (2) HAICU courses, tutorials, training schools and summer schools will be announced to HIDA for forwarding the according information to all Helmholtz Centers. This way HAICU directly participates in HIDA by providing educational material and trainers. In terms of education HAICU also has strong connections to the actually planned Helmholtz Information & Data Science Schools (HIDSS). However, coordinating evolving educational offers is addressed by HIDA. It will turn out whether professors appointed for one of the HAICU research groups can join as principal investigator (PI) in one of the Data Science Graduate Schools.

HAICU naturally is linked to HMC, HIP, and HIFIS. Cooperation with HMC will allow to gather information about data at one place and to ensure data to be searchable by appropriate metadata. HIFIS is expected to provide possibilities to access these data in a fast, secure and easy way and further to allow bringing AI algorithms to the data instead of vice versa. Further HIFIS is expected to provide access to tools that allow for collaborative work in terms of communication, sharing data and working as a team from different locations, simultaneously on same code or text documents. HIP is expected to be a driver for generating unique and interesting data.

9.5 DETAILED FINANCIAL PLAN

The total costs for HAICU will be 11.36 M €/a at full capacity, wherefrom costs for HAICU Central will be 4.00 M €/a and costs for HAICU Local will be 4.36 M €/a. For Helmholtz-wide projects 2.70 M€/a is foreseen plus 300 k€/a for HAICU events.

The costs will be divided into two parts and separated into static costs (for HAICU Central and Local HAICU YIGs and HLSTs, respectively) and dynamic costs (including vouchers and projects). Static costs amount to a sum of 8.66 M €/a at full capacity, with personnel costs of 7.54 M €/a and material costs of 1.12 k €/a. Dynamic costs will account for 2.7 M €/a.

The detailed breakdown is given at Table 3.

All personnel costs are standardized to personnel funds of the DFG for the year 2018:

- W3 professorship: 101.100 €/a
- W2 group leader E14-15: 83.400 €/a
- Researcher/scientific employee, PostDoc E13-14: 69.900 €/a
- PhD student E13-14: 64.500 €/a
- Non-scientific employee E2-E9: 48.000 €/a

The standard DFG overhead key of 25% was applied.

Selected explanations to specific lines of the financial plan

- line 12 / line 29:
  Experts at HAICU's HLSTs are supposed to stay up to date with the rapidly changing field of AI. Therefore, not only researchers but also HLST experts will visit international conferences to identify current trends in AI.

- line 14:
  HAICU Central will regularly organize events for public relations and outreach, see Section 4.2. Further, HAICU Central's Scientific Coordinators are also supposed to visit international conferences for scouting renowned experts in the field and to stay up to date with current trends in AI.

- line 17:
  HAICU will provide a visiting researchers program as well as a program for guest professors, see Section 4.2. Therefore, funding will be provided for travelling and accommodation. Members of HAICU Governance boards will get allowance. Further corresponding social events, e.g., joint dinners, require financial resources. The same hold for expense allowance for invited guests that give talks, and corresponding social events.

- line 32-36:
  HAICU YIGs, see Section 4.1, are structured as Helmholtz YIG. Therefore, their funding is equivalent.

- line 43:
  Based on intensive discussions and visits, we are able to provide an estimate of how many voucher requests HAICU deals with. The number of HLST experts that are intended at HAICU Central and HAICU Local (35 FTE in total, 10 FTE at line 17 and 5 FTE in lines 37-41 each) fit the expected number of requests.

  Examples for vouchers are:

  - a PhD Student visits HAICU Central for two weeks in order to receive support on a specific question.
  - two HLST experts visit a Helmholtz Center to support its researchers in a specific practical or theoretical research task or to provide training on specific AI methods.
  - a PhD student moves to HAICU Central for 6 months in order to gain practical and theoretical expertise in a specific research task, advised by HAICU Central´s HLST experts and researchers. Thereby he/she gets to know young researchers from other research fields, their research questions and expertise. This way such lab rotations may act as a seed for new transdisciplinary projects.

  Additionally, to personal expenses each voucher requires additional financial expense, e.g., for travelling (PhD traveling to HAICU Central, HLST travelling to the Center), accommodation and daily allowance. Estimating these expenses w.r.t. the number of expected voucher requested, yield the financial needs given at line 60.

- line 44:
  A project with two PhD positions amounts to 135,000 €/a, if personal costs are requested solely. Thus, up to 14 of those projects might be funded per year. But within this item, additionally investments for tests and usage of innovative hardware are considered. As example we refer to recent developments in the fields of quantum computing or neuromorphic computing. If reasonable and according test settings will be available, they can be taken into account. As more recent example for such financial request we refer to the current need for powerful, i.e., HPC technology, but at the same time small hardware solutions, independently of internet connections, e.g., for offshore applications at GEOMAR, AWI or HZG. Further, researchers new in the field of ML, especially DL, might request a small GPU cluster for their first cases. Thus, HAICU´s project funding again allows for most flexible application for projects in terms of personnel or equipment.

Table 3: Detailed Financial Plan for HAICU Central and Local

				initialization				full capacity										
				year 1		year 2		year 3		year 4		year 5		year 6		year 7		
				in FTE	in €	in FTE	in €	in FTE	in €	in FTE	in €	in FTE	in €	in FTE	in €	in FTE	in €	
	HAICU Central	personnel costs																HAICU static costs
1		Management	Scientific Director	1	101.100	1	101.100	1	101.100	1	101.100	1	101.100	1	101.100	1	101.100	
2			Financial Manager	1	69.900	1	69.900	1	69.900	1	69.900	1	69.900	1	69.900	1	69.900	
3			Secretary	1	48.000	1	48.000	1	48.000	1	48.000	1	48.000	1	48.000	1	48.000	
4			Scientific Coordinators	4	279.600	4	279.600	4	279.600	4	279.600	4	279.600	4	279.600	4	279.600	
5		Research	Senior Research Group 1			5	393.900	5	393.900	5	393.900	5	393.900	5	393.900	5	393.900	
6			Senior Research Group 2					5	393.900	5	393.900	5	393.900	5	393.900	5	393.900	
7			Junior Research Group 1	3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	
8			Junior Research Group 2			3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	
9			Junior Research Group 3					3	265.800	3	265.800	3	265.800	3	265.800	3	265.800	
10		Guest professorship program	guest professor (costs in expense allowance; 25.000 € each)			1		2		2		2		2		2		
11		High Level Support Team	10 researcher/scientific employee	10	699.000	10	699.000	10	699.000	10	699.000	10	699.000	10	699.000	10	699.000	
12		travel costs (5.000 € per FTE management (excluding secretary), 20.000 € per senior research group, 15.000 € per junior research group , 5.000 € per FTE guest professor, 30.000 € per HLST group)			75.000		115.000		155.000		155.000		155.000		155.000		155.000	
13		outreach, invitations, conferences, recruiting																
14			yearly costs		300.000		300.000		300.000		300.000		300.000		300.000		300.000	
15		costs of materials and supplies (3.000 € per FTE)			60.000		87.000		114.000		114.000		114.000		114.000		114.000	
16		software licenses			80.000		114.500		150.000		150.000		150.000		150.000		150.000	
17		expense allowance (different boards, guest professor ship etc.)			50.000		100.000		100.000		100.000		100.000		100.000		100.000	
18		**HAICU Central complete**	**total sum per year**	**20**	**2.028.400**	**29**	**2.839.600**	**38**	**3.601.800**	**38**	**3.601.800**	**38**	**3.601.800**	**38**	**3.601.800**	**38**	**3.601.800**	
19			thereof personnel costs		1.463.400		2.123.100		2.782.800		2.782.800		2.782.800		2.782.800		2.782.800	
20			thereof materials/projects		565.000		716.500		819.000		819.000		819.000		819.000		819.000	
21			overhead, 25 % of personnel costs		365.850		530.775		695.700		695.700		695.700		695.700		695.700	
22			**total sum with overhead**		**2.394.250**		**3.370.375**		**4.297.500**		**4.297.500**		**4.297.500**		**4.297.500**		**4.297.500**	
23	HAICU Local	personnel costs																
24		HLST 1	researcher/scientific employee	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	
25		HLST 2	researcher/scientific employee	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	
26		HLST 3	researcher/scientific employee			5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	
27		HLST 4	researcher/scientific employee			5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	
28		HLST 5	researcher/scientific employee					5	349.500	5	349.500	5	349.500	5	349.500	5	349.500	
29		travel costs (15.000 € per HLST group)			30.000		60.000		75.000		75.000		75.000		75.000		75.000	
30		software licenses			60.000		114.500		150.000		150.000		150.000		150.000		150.000	
31		costs of materials and supplies (3.000 € per FTE)			30.000		60.000		75.000		75.000		75.000		75.000		75.000	
32		Young Investigator Group Program	Research Group 1	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	
33			Research Group 2	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	
34			Research Group 3			3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	
35			Research Group 4					3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	
36			Research Group 5					3	300.000	3	300.000	3	300.000	3	300.000	3	300.000	
37			**total sum per year**	**16**	**1.419.000**	**29**	**2.532.500**	**40**	**3.547.500**	**40**	**3.547.500**	**40**	**3.547.500**	**40**	**3.547.500**	**40**	**3.547.500**	
38			thereof personnel costs		1.299.000		2.298.000		3.247.500		3.247.500		3.247.500		3.247.500		3.247.500	
39			thereof materials/projects		120.000		234.500		300.000		300.000		300.000		300.000		300.000	
40			overhead, 25 % of personnel costs		324.750		574.500		811.875		811.875		811.875		811.875		811.875	
41			**total sum with overhead**		**1.743.750**		**3.107.000**		**4.359.375**		**4.359.375**		**4.359.375**		**4.359.375**		**4.359.375**	
42		projects																HAICU dynamic costs
43			voucher system		280.000		560.000		700.000		700.000		700.000		700.000		700.000	
44			scientific project calls (including test/invest of innovative hardware)		2.000.000		2.000.000		2.000.000		2.000.000		2.000.000		2.000.000		2.000.000	
45			**total sum per year**		**2.280.000**		**2.560.000**		**2.700.000**		**2.700.000**		**2.700.000**		**2.700.000**		**2.700.000**	
46		**HAICU Local complete**	**total sum per year**	**16**	**3.699.000**	**29**	**5.092.500**	**40**	**6.247.500**	**40**	**6.247.500**	**40**	**6.247.500**	**40**	**6.247.500**	**40**	**6.247.500**	
47			thereof personnel costs		1.299.000		2.298.000		3.247.500		3.247.500		3.247.500		3.247.500		3.247.500	
48			thereof materials/projects		2.400.000		2.794.500		3.000.000		3.000.000		3.000.000		3.000.000		3.000.000	
49			**total costs per year with overhead**		**4.023.750**		**5.667.000**		**7.059.375**		**7.059.375**		**7.059.375**		**7.059.375**		**7.059.375**	
50		**HAICU Central and HAICU Local: Static costs**	**total sum per year**	**36**	**3.447.400**	**58**	**5.372.100**	**78**	**7.149.300**	**78**	**7.149.300**	**78**	**7.149.300**	**78**	**7.149.300**	**78**	**7.149.300**	
51			**thereof personnel costs**		**2.762.400**		**4.421.100**		**6.030.300**		**6.030.300**		**6.030.300**		**6.030.300**		**6.030.300**	
52			**thereof materials/projects**		**685.000**		**951.000**		**1.119.000**		**1.119.000**		**1.119.000**		**1.119.000**		**1.119.000**	
53			**total costs per year with overhead**		**4.138.000**		**6.477.375**		**8.656.875**		**8.656.875**		**8.656.875**		**8.656.875**		**8.656.875**	
54		**HAICU complete costs**	**total sum per year**	**36**	**5.727.400**	**58**	**7.932.100**	**78**	**9.849.300**	**78**	**9.849.300**	**78**	**9.849.300**	**78**	**9.849.300**	**78**	**9.849.300**	
55			**thereof personnel costs**		**2.762.400**		**4.421.100**		**6.030.300**		**6.030.300**		**6.030.300**		**6.030.300**		**6.030.300**	
56			**thereof materials/projects**		**2.965.000**		**3.511.000**		**3.819.000**		**3.819.000**		**3.819.000**		**3.819.000**		**3.819.000**	
57			**total costs per year with overhead**		**6.418.000**		**9.037.375**		**11.356.875**		**11.356.875**		**11.356.875**		**11.356.875**		**11.356.875**	

9.6 REFERENCES

Abbosh, Christopher, Nicolai J. Birkbak, Gareth A. Wilson, Mariam Jamal-Hanjani, Tudor Constantin, Raheleh Salari, John Le Quesne, et al. 2017. “Phylogenetic CtDNA Analysis Depicts Early-Stage Lung Cancer Evolution.” Nature 545 (7655): 446–51. *https://doi.org/10.1038/nature22364.*

Alevizos, Evangelos, Timm Schoening, Kevin Koeser, Mirjam Snellen, and Jens Greinert. 2018. “Quantification of the Fine-Scale Distribution of Mn-Nodules: Insights from AUV Multi-Beam and Optical Imagery Data Fusion.” Biogeosciences Discussions, February, 1–29. *https://doi.org/10.5194/bg-2018-60.*

Alioto, Tyler S., Ivo Buchhalter, Sophia Derdak, Barbara Hutter, Matthew D. Eldridge, Eivind Hovig, Lawrence E. Heisler, et al. 2015. “A Comprehensive Assessment of Somatic Mutation Detection in Cancer Using Whole-Genome Sequencing.” Nature Communications 6 (December): 10001. *https://doi.org/10.1038/ncomms10001.*

Amunts, K., C. Lepage, L. Borgeat, H. Mohlberg, T. Dickscheid, M.-E. Rousseau, S. Bludau, et. al. 2013. „BigBrain: An Ultrahigh-Resolution 3D Human Brain Model“. Science 340 (6139): 1472–75. *https://doi.org/10.1126/science.1235381.*

Angermueller, Christof, Heather J. Lee, Wolf Reik, and Oliver Stegle. 2017. “DeepCpG: Accurate Prediction of Single-Cell DNA Methylation States Using Deep Learning.” Genome Biology 18 (April): 67. *https://doi.org/10.1186/s13059-017-1189-z.*

Angermueller, Christof, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. 2016. “Deep Learning for Computational Biology.” Molecular Systems Biology 12 (7): 878. *https://doi.org/10.15252/msb.20156651*.

Argelaguet, Ricard, Britta Velten, Damien Arnol, Sascha Dietrich, Thorsten Zenz, John C. Marioni, Florian Buettner, Wolfgang Huber, and Oliver Stegle. 2018. “Multi Omics Factor Analysis—a Framework for Unsupervised Integration of Multi omics Data Sets.” Molecular Systems Biology 14 (6): e8124. *https://doi.org/10.15252/msb.20178124.*

Asgari, Ehsaneddin, Kiavash Garakani, Alice Carolyn McHardy, and Mohammad R. K. Mofrad. 2018. “MicroPheno: Predicting Environments and Host Phenotypes from 16S RRNA Gene Sequencing Using a k-Mer Based Representation of Shallow Sub-Samples.” BioRxiv, January, 255018. *https://doi.org/10.1101/255018.*

ATLAS Collaboration, M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, et al. 2018. “Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector.” Physical Review D 97 (7): 072016. *https://doi.org/10.1103/PhysRevD.97.072016.*

Auer, S., I. Hornig, M. Schmitt, and P. Reinartz. 2017. “Simulation-Based Interpretation and Alignment of High-Resolution Optical and SAR Images.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (11): 4779–93. *https://doi.org/10.1109/JSTARS.2017.2723082.*

Authority of the House of Lords. 2018. “AI in the UK: Ready, Willing and Able?”

Babaee, Mohammadreza, Stefanos Tsoukalas, Gerhard Rigoll, and Mihai Datcu. 2016. “Immersive Visualization of Visual Data Using Nonnegative Matrix Factorization.” Neurocomputing 173 (January): 245–55. *https://doi.org/10.1016/j.neucom.2015.03.121.*

Bauer, K., G. Muñoz, and I. Moeck. 2012. “Pattern Recognition and Lithological Interpretation of Collocated Seismic and Magnetotelluric Models Using Selforganizing Maps.” Geophysical Journal International 189 (2): 984–98. *https://doi.org/10.1111/j.1365-246X.2012.05402.x.*

Beattie, Charlie, Joel Leibo, Stig Petersen, and Shane Legg. 2016. “Open-Sourcing DeepMind Lab.” March 12, 2016. *https://deepmind.com/blog/open-sourcing-deepmind-lab/.*

Bernard, O., A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. A. Heng, I. Cetin, et al. 2018. “Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?” IEEE Transactions on Medical Imaging, 1–1. *https://doi.org/10.1109/TMI.2018.2837502.*

Bickelhaupt, Sebastian, Paul Ferdinand Jaeger, Frederik Bernd Laun, Wolfgang Lederer, Heidi Daniel, Tristan Anselm Kuder, Lorenz Wuesthof, et al. 2018. “Radiomics Based on Adapted Diffusion Kurtosis Imaging Helps to Clarify Most Mammographic Findings Suspicious for Cancer.” Radiology 287 (3): 761–70. *https://doi.org/10.1148/radiol.2017170273.*

Bishop, Christopher M. 2006. Pattern Recognition and Machine Learning. Springer.

BMBF. 2018. „Künstliche Intelligenz Intelligent Nutzen,“ March 29, 2018. *https://www.bmbf.de/foerderungen/bekanntmachung-1367.html.*

Buettner, Florian, Naruemon Pratanwanich, Davis J. McCarthy, John C. Marioni, and Oliver Stegle. 2017. “F-ScLVM: Scalable and Versatile Factor Analysis for Single-Cell RNA-Seq.” Genome Biology 18 (November): 212. *https://doi.org/10.1186/s13059-017-1334-8.*

Buggenthin, Felix, Florian Buettner, Philipp S. Hoppe, Max Endele, Manuel Kroiss, Michael Strasser, Michael Schwarzfischer, et al. 2017. “Prospective Identification of Hematopoietic Lineage Choice by Deep Learning.” Nature Methods 14 (4): 403–6. *https://doi.org/10.1038/nmeth.4182.*

Bundesregierung. 2018a. „Koalitionsvertrag zwischen CDU, CSU und SPD.“

Bundesregierung. 2018b. „Eckpunkte Der Bundesregierung Für Eine Strategie Künstliche Intelligenz.“ *https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf.*

Calviello, Lorenzo, Neelanjan Mukherjee, Emanuel Wyler, Henrik Zauber, Antje Hirsekorn, Matthias Selbach, Markus Landthaler, Benedikt Obermayer, and Uwe Ohler. 2016. “Detecting Actively Translated Open Reading Frames in Ribosome Profiling Data.” Nature Methods 13 (2): 165–70. *https://doi.org/10.1038/nmeth.3688.*

Camporeale, Enrico, Yuri Shprits, Mandar Chandorkar, Alexander Drozdov, and Simon Wing. 2016. “On the Propagation of Uncertainties in Radiation Belt Simulations.” Space Weather 14 (11): 982–92. *https://doi.org/10.1002/2016SW001494.*

Capper, David, David T. W. Jones, Martin Sill, Volker Hovestadt, Daniel Schrimpf, Dominik Sturm, Christian Koelsche, et al. 2018. “DNA Methylation-Based Classification of Central Nervous System Tumours.” Nature 555 (March): 469.

Casasnovas, Rodrigo, Vittorio Limongelli, Pratyush Tiwary, Paolo Carloni, and Michele Parrinello. 2017. „Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations“. Journal of the American Chemical Society 139 (13): 4780–88. *https://doi.org/10.1021/jacs.6b12950.*

Cavallaro, Gabriele, Morris Riedel, Matthias Richerzhagen, Jon Atli Benediktsson, and Antonio Plaza. 2015. „On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector

Machine Methods". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (10): 4634–46. *https://doi.org/10.1109/JSTARS.2015.2458855.*

Christiansen, Svenja, and Rainer Kiko. 2016. "PlanktonID." 2016. *https://planktonid.geomar.de/de*.

Chudasama, Priya, Sadaf S. Mughal, Mathijs A. Sanders, Daniel Hübschmann, Inn Chung, Katharina I. Deeg, Siao-Han Wong, et al. 2018. "Integrative Genomic and Transcriptomic Analysis of Leiomyosarcoma." Nature Communications 9 (1): 144. *https://doi.org/10.1038/s41467-017-02602-0.*

CMS Collaboration. 2018. "Search for $t\bar{t}$H production in the H$b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s}$= 13 TeV, April. *http://arxiv.org/abs/1804.03682.*

COCO. 2018. "COCO Common Objects in Context." 2018. *http://cocodataset.org/#home.*

Cold Spring Harbor Laboratory. 2018. "Quantitative Biology Fellows Program." 2018. *https://www.cshl.edu/research/fellows-programs-at-cshl/quantitative-biology-fellows-program/.*

Cox, David Daniel, and Thomas Dean. 2014. "Neural Networks and Neuroscience-Inspired Computer Vision." Current Biology 24 (18): R921–29. *https://doi.org/10.1016/j.cub.2014.08.026.*

Datcu, M., H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia. 2003. "Information Mining in Remote Sensing Image Archives: System Concepts." IEEE Transactions on Geoscience and Remote Sensing 41 (12): 2923–36. *https://doi.org/10.1109/TGRS.2003.817197.*

Datcu, M., and K. Seidel. 2000. "Image Information Mining: Exploration of Image Content in Large Archives." In 2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484), 3:253–64 vol.3. *https://doi.org/10.1109/AERO.2000.879853.*

Dawidowski, M., L. Emmanouilidis, V. C. Kalel, K. Tripsianes, K. Schorpp, K. Hadian, M. Kaiser, et al. 2017. "Inhibitors of PEX14 Disrupt Protein Import into Glycosomes and Kill Trypanosoma Parasites." Science (New York, N.Y.) 355 (6332): 1416–20. *https://doi.org/10.1126/science.aal1807.*

Delacher, Michael, Charles D. Imbusch, Dieter Weichenhan, Achim Breiling, Agnes Hotz-Wagenblatt, Ulrike Träger, Ann-Cathrin Hofer, et al. 2017. "Genome-Wide DNA-Methylation Landscape Defines Specialization of Regulatory T Cells in Tissues." Nature Immunology 18 (10): 1160–72. *https://doi.org/10.1038/ni.3799.*

DFG. 2014. „Leitfaden Mit Hinweisen Zu Abschlussberichten von Forschungsgruppen." 2.015-04/14. *http://www.dfg.de/formulare/2_015/2_015.pdf.*

DFG. 2016. „Hinweise Für Die Schriftliche Begutachtung." 10.20-02/16. *http://www.dfg.de/formulare/10_20/10_20_de.pdf.*

Erdmann, Martin, Lukas Geiger, Jonas Glombitza, and David Schmidt. 2018. "Generating and Refining Particle Detector Simulations Using the Wasserstein Distance in Adversarial Networks." ArXiv:1802.03325 [Astro-Ph, Physics:Hep-Ex], February. *http://arxiv.org/abs/1802.03325.*

ESKP. 2018. "Earth System Knowledge Platform." 2018. *https://www.gfz-potsdam.de/en/research/crossprogramme/eskp-earth-system-knowledge-platform/overview/.*

Eulenberg, Philipp, Niklas Köhler, Thomas Blasi, Andrew Filby, Anne E. Carpenter, Paul Rees, Fabian J. Theis, and F. Alexander Wolf. 2017. "Reconstructing Cell Cycle and Disease Progression Using Deep Learning." Nature Communications 8 (1): 463. *https://doi.org/10.1038/s41467-017-00623-3.*

EurekAlert! 2018. “Breast Cancer: How Advanced Imaging Technologies Will Help Avoid Unnecessary Biopsies.” EurekAlert! February 20, 2018. *https://www.eurekalert.org/pub_releases/2018-02/gcrc-bch021818.php.*

Europäische Kommission, Pressemitteilung. 2018. „Künstliche Intelligenz: Kommission Beschreibt Europäisches Konzept Zur Förderung von Investitionen Und Entwicklung Ethischer Leitlinien.” April 25, 2018. *http://europa.eu/rapid/press-release_IP-18-3362_de.htm.*

European Commission. 2018. “Artificial Intelligence for Europe – communication from the commission to the european parliament, the european council, the council, the european economic and social committee and the committee of the regions”
https://www.kowi.de/Portaldata/2/Resources/fp/2018-COM-Artificial-Intelligence.pdf.

European Commission and Council. 2018. “Declaration - Cooperation on Artificial Intelligence.”

Felleman, D. J., and D. C. Van Essen. 1991. “Distributed Hierarchical Processing in the Primate Cerebral Cortex.” Cerebral Cortex (New York, N.Y.: 1991) 1 (1): 1–47.

FESAC. 2018. “fusion energy sciences advisory committee report – Transformative Enabling Capabilities for Efficient Advance Toward Fusion Energy.” U.S. Department of Energy, Office of Science Fusion Energy Science. *https://science.energy.gov/~/media/fes/fesac/pdf/2018/TEC_Report_15Feb2018.pdf.*

Fong, Ruth, Walter Scheirer, and David Cox. 2017. “Using Human Brain Activity to Guide Machine Learning.” ArXiv:1703.05463 [Cs], March. *http://arxiv.org/abs/1703.05463.*

France24. 2018. “France to Invest €1.5 Billion in Artificial Intelligence by 2022.” March 29, 2018. *http://www.france24.com/en/20180329-france-invest-15-billion-euros-artificial-intelligence-AI-technology-2022.*

Fusi, Nicolo, Christoph Lippert, Neil D. Lawrence, and Oliver Stegle. 2014. “Warped Linear Mixed Models for the Genetic Analysis of Transformed Phenotypes.” Nature Communications 5 (September): 4890. *https://doi.org/10.1038/ncomms5890.*

FZJ. 2017. “Helmholtz Analytics Framework.” July 18, 2017.
https://www.fz-juelich.de/SharedDocs/Meldungen/IAS/JSC/EN/2017/2017-07-helmholtz-analytics-framework.html?nn=2066128.

FZJ JSC. 2018. “Simulation Laboratories.” 2018.
http://www.fz-juelich.de/ias/jsc/EN/Expertise/SimLab/simlab_node.html.

Ghamisi, P., Y. Chen, and X. X. Zhu. 2016. “A Self-Improving Convolution Neural Network for the Classification of Hyperspectral Data.” IEEE Geoscience and Remote Sensing Letters 13 (10): 1537–41. *https://doi.org/10.1109/LGRS.2016.2595108.*

Ghamisi, P., B. Höfle, and X. X. Zhu. 2017. “Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (6): 3011–24. *https://doi.org/10.1109/JSTARS.2016.2634863.*

Giuffrida, Mario Valerio, Hanno Scharr, and Sotirios A. Tsaftaris. 2017. “ARIGAN: Synthetic Arabidopsis Plants Using Generative Adversarial Network.” ArXiv:1709.00938 [Cs], September. *http://arxiv.org/abs/1709.00938.*

Goetz, M., C. Weber, F. Binczyk, J. Polanska, R. Tarnawski, B. Bobek-Billewicz, U. Koethe, J. Kleesiek, B. Stieltjes, and K. H. Maier-Hein. 2016. “DALSA: Domain Adaptation for Supervised Learning From Sparsely Annotated MR Images.” IEEE Transactions on Medical Imaging 35 (1): 184–96. *https://doi.org/10.1109/TMI.2015.2463078.*

Gregor, Ivan, Johannes Dröge, Melanie Schirmer, Christopher Quince, and Alice C. McHardy. 2016. "PhyloPythiaS+: A Self-Training Method for the Rapid Reconstruction of Low-Ranking Taxonomic Bins from Metagenomes." PeerJ 4 (February): e1603. *https://doi.org/10.7717/peerj.1603.*

Gröbner, Susanne N., Barbara C. Worst, Joachim Weischenfeldt, Ivo Buchhalter, Kortine Kleinheinz, Vasilisa A. Rudneva, Pascal D. Johann, et al. 2018. "The Landscape of Genomic Alterations across Childhood Cancers." Nature 555 (7696): 321–27. *https://doi.org/10.1038/nature25480.*

Gunning, David. 2017. "Explainable Artificial Intelligence (Xai)." Defense Advanced Research Projects Agency (DARPA), Nd Web.

HAF. 2018. "Helmholtz Analytics Framework." 2018. *http://www.helmholtz-analytics.de/helmholtz_analytics/EN/Home/home_node.html.*

Hafez, Dina, Aslihan Karabacak, Sabrina Krueger, Yih-Chii Hwang, Li-San Wang, Robert P. Zinzen, and Uwe Ohler. 2017. "McEnhancer: Predicting Gene Expression via Semi-Supervised Assignment of Enhancers to Target Genes." Genome Biology 18 (1): 199. *https://doi.org/10.1186/s13059-017-1316-x.*

Haghverdi, Laleh, Maren Büttner, F. Alexander Wolf, Florian Buettner, and Fabian J. Theis. 2016. "Diffusion Pseudotime Robustly Reconstructs Lineage Branching." Nature Methods 13 (10): 845–48. *https://doi.org/10.1038/nmeth.3971.*

Hartevelt, Tim J. van, Joana Cabral, Arne Møller, James J. FitzGerald, Alexander L. Green, Tipu Z. Aziz, Gustavo Deco, and Morten L. Kringelbach. 2015. "Evidence from a Rare Case Study for Hebbian-like Changes in Structural Connectivity Induced by Long-Term Deep Brain Stimulation." Frontiers in Behavioral Neuroscience 9 (June). *https://doi.org/10.3389/fnbeh.2015.00167.*

Hasasneh, Ahmad, Nikolas Kampel, Praveen Sripad, N. J. Shah, and Jürgen Dammers. 2018. "Deep Learning Approach for Automatic Classification of Ocular and Cardiac Artifacts in MEG Data." Journal of Engineering, 1. Hassabis, Demis. 2017. "Artificial Intelligence: Chess Match of the Century." Nature 544 (7651): 413–414.

HBP. 2017. "The HBP Joint Platform (HBP-JP)." 2017. *https://www.humanbrainproject.eu/en/hbp-platforms/hbp-joint-platform/.*

HBP open calls. 2018. "HBP Open Calls." 2018. *https://www.humanbrainproject.eu/en/open-ethical-engaged/open-calls/.*

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Deep Residual Learning for Image Recognition." ArXiv:1512.03385 [Cs], December. *http://arxiv.org/abs/1512.03385.*

Heim, E., A. Seitel, F. Isensee, J. Andrulis, C. Stock, T. Ross, and L. Maier-Hein. 2017. "Clickstream Analysis for Crowd-Based Object Segmentation with Confidence." IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1. *https://doi.org/10.1109/TPAMI.2017.2777967*.

Helmholtz Association. 2017. „Ausschreibungsrichtlinien Für Helmholtz-Zentren Sowie Kandidatinnen Und Kandidaten." 2017. *https://www.helmholtz.de/fileadmin/user_upload/01_forschung/Forschungsfoerderung/Nachwuchsfoerderung/Nachwuchsgruppen/Nachwuchsgruppen_2018_de/A1_Ausschreibungsrichtlinien.pdf.*

Helmholtz Association. 2018. "Information & Data Science." 2018. *https://www.helmholtz.de/en/research/information_data_science/.*

Hertel, Jana, David Langenberger, and Peter F. Stadler. 2014. "Computational Prediction of MicroRNA Genes." In RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods, 437–56.

Methods in Molecular Biology. Humana Press, Totowa, NJ. *https://doi.org/10.1007/978-1-62703-709-9_20.*

Higgs Challenge. 2018. "Higgs Boson Machine Learning Challenge." 2018.
https://www.kaggle.com/c/higgs-boson.

Hinton, Geoffrey E, Simon Osindero, and Yee-Whye Teh. 2006.
"A Fast Learning Algorithm for Deep Belief Nets." Neural Comput. 18 (7): 1527–1554.

Hörl, David, Fabio Rojas Rusak, Friedrich Preusser, Paul Tillberg, Nadine Randel, Raghav K. Chhetri, Albert Cardona, et al. 2018. "BigStitcher: Reconstructing High-Resolution Image Datasets of Cleared and Expanded Samples." BioRxiv, June, 343954. *https://doi.org/10.1101/343954.*

Hu, J., L. Mou, A. Schmitt, and X. X. Zhu. 2017. "FusioNet: A Two-Stream Convolutional Neural Network for Urban Scene Classification Using PolSAR and Hyperspectral Data." In 2017 Joint Urban Remote Sensing Event (JURSE), 1–4. *https://doi.org/10.1109/JURSE.2017.7924565.*

Hughes, L. H., M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu. 2018. "Identifying Corresponding Patches in SAR and Optical Images With a Pseudo-Siamese CNN." IEEE Geoscience and Remote Sensing Letters 15 (5): 784–88. *https://doi.org/10.1109/LGRS.2018.2799232.*

Human Brain Project Joint Platform. 2018. "Human Brain Project Joint Platform." 2018.
https://www.humanbrainproject.eu/en/hbp-platforms/hbp-joint-platform/.

Ippen, Tammo, Jochen M. Eppler, Hans E. Plesser, and Markus Diesmann. 2017. "Constructing Neuronal Network Models in Massively Parallel Environments." Frontiers in Neuroinformatics, 30.

ISMB. 2018. "Workshop on Advancing Computational Biology through Critical Assessments, Community Experiments, and Crowdsourcing." 2018.
https://www.iscb.org/ismb2018-program/ismb2018-special-sessions#sst04.

Jamal-Hanjani, Mariam, Gareth A. Wilson, Nicholas McGranahan, Nicolai J. Birkbak, Thomas B. K. Watkins, Selvaraju Veeriah, Seema Shafi, et al. 2017. "Tracking the Evolution of Non–Small-Cell Lung Cancer." Research-article.
http://Dx.Doi.Org/10.1056/NEJMoa1616288. April 26, 2017. *https://doi.org/10.1056/NEJMoa1616288.*

Jitsev, J., A. Morrison, and M. Tittgemeyer. 2012. "Learning from Positive and Negative Rewards in a Spiking Neural Network Model of Basal Ganglia." In The 2012 International Joint Conference on Neural Networks (IJCNN), 1–8. *https://doi.org/10.1109/IJCNN.2012.6252834.*

Jitsev, Jenia. 2014. "Self-Generated Off-Line Memory Reprocessing Strongly Improves Generalization in a Hierarchical Recurrent Neural Network." In Artificial Neural Networks and Machine Learning – ICANN 2014, 659–66. Lecture Notes in Computer Science. Springer, Cham. *https://doi.org/10.1007/978-3-319-11179-7_83.*

Jitsev, Jenia, and Christoph V. Der Malsburg. 2009. "Experience-Driven Formation of Parts-Based Representations in a Model of Layered Visual Memory." Frontiers in Computational Neuroscience 3.
https://doi.org/10.3389/neuro.10.015.2009.

Jannusch, Kai, Christiane Jockwitz, Hans-Jürgen Bidmon, Susanne Moebus, Katrin Amunts, and Svenja Caspers. 2017. „A Complex Interplay of Vitamin B1 and B6 Metabolism with Cognition, Brain Structure, and Functional Connectivity in Older Adults". Frontiers in Neuroscience 11 (Oktober).
https://doi.org/10.3389/fnins.2017.00596.

John von Neumann Institute for Computing. 2018. “John von Neumann Institute for Computing.” 2018. *http://www.john-von-neumann-institut.de/nic/DE/Home/home_node.html.*

Kang, Jian, Marco Körner, Yuanyuan Wang, Hannes Taubenböck, and Xiao Xiang Zhu. 2018. “Building Instance Classification Using Street View Images.” ISPRS Journal of Photogrammetry and Remote Sensing, March. *https://doi.org/10.1016/j.isprsjprs.2018.02.006.*

Karaiskos, Nikos, Philipp Wahle, Jonathan Alles, Anastasiya Boltengagen, Salah Ayoub, Claudia Kipar, Christine Kocks, Nikolaus Rajewsky, and Robert P. Zinzen. 2017. “The Drosophila Embryo at Single-Cell Transcriptome Resolution.” Science (New York, N.Y.) 358 (6360): 194–99. *https://doi.org/10.1126/science.aan3235.*

Kennedy, James. 2006. “Swarm Intelligence.” In Handbook of Nature-Inspired and Innovative Computing: Integrating Classical Models with Emerging Technologies, edited by Albert Y Zomaya, 187–219. Boston, MA: Springer US.

Kirchner, Thomas, Janek Gröhl, and Lena Maier-Hein. 2018. “Context Encoding Enables Machine Learning-Based Quantitative Photoacoustics.” Journal of Biomedical Optics 23 (5): 1–9. *https://doi.org/10.1117/1.JBO.23.5.056008.*

KIT SCC. 2018. “Simulation Laboratories (SimLabs).” 2018. *https://www.scc.kit.edu/forschung/5960.php.*

Kleesiek, Jens, Gregor Urban, Alexander Hubert, Daniel Schwarz, Klaus Maier-Hein, Martin Bendszus, and Armin Biller. 2016. “Deep MRI Brain Extraction: A 3D Convolutional Neural Network for Skull Stripping.” NeuroImage 129 (April): 460–69. *https://doi.org/10.1016/j.neuroimage.2016.01.024.*

Koch, Yvonne, Thomas Wolf, Peter K. Sorger, Roland Eils, and Benedikt Brors. 2013. “Decision-Tree Based Model Analysis for Efficient Identification of Parameter Relations Leading to Different Signaling States.” PLOS ONE 8 (12): e82593. *https://doi.org/10.1371/journal.pone.0082593.*

Kokel, David, Timothy W. Dunn, Misha B. Ahrens, Rüdiger Alshut, Chung Yan J. Cheung, Louis Saint-Amant, Giancarlo Bruni, et al. 2013. “Identification of Nonvisual Photomotor Response Cells in the Vertebrate Hindbrain.” Journal of Neuroscience 33 (9): 3834–43. *https://doi.org/10.1523/JNEUROSCI.3689-12.2013.*

Koszalka, I., J. H. LaCasce, M. Andersson, K. A. Orvik, and C. Mauritzen. 2011. “Surface Circulation in the Nordic Seas from Clustered Drifters.” Deep Sea Research Part I: Oceanographic Research Papers 58 (4): 468–85. *https://doi.org/10.1016/j.dsr.2011.01.007.*

Köthur, P., M. Sips, H. Dobslaw, and D. Dransch. 2014. “Visual Analytics for Comparison of Ocean Model Output with Reference Data: Detecting and Analyzing Geophysical Processes Using Clustering Ensembles.” IEEE Transactions on Visualization and Computer Graphics 20 (12): 1893–1902. *https://doi.org/10.1109/TVCG.2014.2346751.*

Krajsek, K., and H. Scharr. 2010. “Diffusion Filtering without Parameter Tuning: Models and Inference Tools.” In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2536–43. *https://doi.org/10.1109/CVPR.2010.5539959.*

Kriest, I., V. Sauerland, S. Khatiwala, A. Srivastav, and A. Oschlies. 2017. “Calibrating a Global Three-Dimensional Biogeochemical Ocean Model (MOPS-1.0).” Geosci. Model Dev. 10 (1): 127–54. *https://doi.org/10.5194/gmd-10-127-2017.*

Krizhevsky, Alex. 2009. “CIFAR-100: Object Recognition.” 2009. *https://www.cs.toronto.edu/~kriz/cifar.html.*

Kunkel, Susanne, and Wolfram Schenck. 2017. "The NEST Dry-Run Mode: Efficient Dynamic Analysis of Neuronal Network Simulation Code." Frontiers in Neuroinformatics, 40.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." Nature 521 (7553): 436–444.

Ledford, Heidi. 2017. "Artificial Intelligence Identifies Plant Species for Science." Nature News, August.

Lee, Seunghak, Nico Gornitz, Eric P. Xing, David Heckerman, and Christoph Lippert. 2017. "Ensembles of Lasso Screening Rules." IEEE Transactions on Pattern Analysis and Machine Intelligence, November. *https://doi.org/10.1109/TPAMI.2017.2765321.*

Lefort-Besnard, Jérémy, Danielle S. Bassett, Jonathan Smallwood, Daniel S. Margulies, Birgit Derntl, Oliver Gruber, Andre Aleman, et al. 2018. „Different Shades of Default Mode Disturbance in Schizophrenia: Subnodal Covariance Estimation in Structure and Function". Human Brain Mapping 39 (2): 644–61. *https://doi.org/10.1002/hbm.23870.*

Lippert, Christoph, Riccardo Sabatini, M. Cyrus Maher, Eun Yong Kang, Seunghak Lee, Okan Arikan, Alena Harley, et al. 2017. "Identification of Individuals by Trait Prediction Using Whole-Genome Sequencing Data." Proceedings of the National Academy of Sciences 114 (38): 10166–71. *https://doi.org/10.1073/pnas.1711125114.*

LSDMA. 2014. "Large-Scale Data Management and Analysis." 2014. *https://www.helmholtz-lsdma.de/.*

Lu, Ake Tzu-Hui, Steven Bakker, Esther Janson, Sven Cichon, Rita M. Cantor, and Roel A. Ophoff. 2012. „Prediction of Serotonin Transporter Promoter Polymorphism Genotypes from Single Nucleotide Polymorphism Arrays Using Machine Learning Methods": Psychiatric Genetics 22 (4): 182–88. *https://doi.org/10.1097/YPG.0b013e328353ae23.*

Marmanis, D., M. Datcu, T. Esch, and U. Stilla. 2016. "Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks." IEEE Geoscience and Remote Sensing Letters 13 (1): 105–9. *https://doi.org/10.1109/LGRS.2015.2499239.*

Marr, David. 2010. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. The MIT Press. *http://mitpress.universitypressscholarship.com/view/10.7551/mit-press/9780262514620.001.0001/upso-9780262514620.*

McHardy, Alice Carolyn, and Andreas Kloetgen. 2017. "Finding Genes in Genome Sequence." In Bioinformatics, 271–91. Methods in Molecular Biology. Humana Press, New York, NY. *https://doi.org/10.1007/978-1-4939-6622-6_11.*

MDC. 2018. "Scientific Bioinformatics Platform." 2018. *https://www.mdc-berlin.de/de/node/5572.*

Memon, Mohammad Shahbaz, Gabriele Cavallaro, Morris Riedel, and Helmut Neukirchen. 2017. Facilitating Efficient Data Analysis of Remotely Sensed Images Using Standards-Based Parameter Sweep Models. IEEE.

Meneghini, O., S. P. Smith, P. B. Snyder, G. M. Staebler, J. Candy, E. Belli, L. Lao, et al. 2017. "Self-Consistent Core-Pedestal Transport Simulations with Neural Network Accelerated Models." Nuclear Fusion 57 (8): 086034. *https://doi.org/10.1088/1741-4326/aa7776.*

Middel, Volker, Lu Zhou, Masanari Takamiya, Tanja Beil, Maryam Shahid, Urmas Roostalu, Clemens Grabher, et al. 2016. "Dysferlin-Mediated Phosphatidylserine Sorting Engages Macrophages in Sarcolemma Repair." Nature Communications 7 (September): 12875. *https://doi.org/10.1038/ncomms12875*.

Mikut, Ralf, Serge Ruden, Markus Reischl, Frank Breitling, Rudolf Volkmer, and Kai Hilpert. 2016. “Improving Short Antimicrobial Peptides despite Elusive Rules for Activity.” Biochimica et Biophysica Acta (BBA) – Biomembranes, Antimicrobial peptides, cell membrane and microbial surface interaction, 1858 (5): 1024–33. *https://doi.org/10.1016/j.bbamem.2015.12.013.*

Mishra, M., S. Schmitt, L. Wang, M. K. Strasser, C. Marr, N. Navab, H. Zischka, and T. Peng. 2016. “Structure-Based Assessment of Cancerous Mitochondria Using Deep Networks.” In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 545–48. *https://doi.org/10.1109/ISBI.2016.7493327.*

Moore-Sloan Data Science Environment. 2018. “Moore-Sloan Data Science Environment.” 2018. *http://msdse.org/.*

Morita, Kenji, Jenia Jitsev, and Abigail Morrison. 2016. Corticostriatal Circuit Mechanisms of Value-Based Action Selection: Implementation of Reinforcement Learning Algorithms and Beyond. Vol. 311. *https://doi.org/10.1016/j.bbr.2016.05.017.*

Mou, L., P. Ghamisi, and X. X. Zhu. 2017. “Deep Recurrent Neural Networks for Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 55 (7): 3639–55. *https://doi.org/10.1109/TGRS.2016.2636241.*

Mou, L., X. Zhu, M. Vakalopoulou, K. Karantzalos, N. Paragios, B. Le Saux, G. Moser, and D. Tuia. 2017. “Multitemporal Very High Resolution From Space: Outcome of the 2016 IEEE GRSS Data Fusion Contest.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (8): 3435–47. *https://doi.org/10.1109/JSTARS.2017.2696823.*

Mou, Lichao, and Xiao Xiang Zhu. 2018. “Vehicle Instance Segmentation from Aerial Image and Video Using a Multi-Task Learning Residual Fully Convolutional Network.” ArXiv:1805.10485 [Cs], May. *http://arxiv.org/abs/1805.10485.*

msdse. 2017. “Creating Institutional Change in Data Science The Moore-Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington.” *http://msdse.org/files/Creating_Institutional_Change.pdf.*

Mukherjee, Neelanjan, Lorenzo Calviello, Antje Hirsekorn, Stefano de Pretis, Mattia Pelizzola, and Uwe Ohler. 2017. “Integrative Classification of Human Coding and Noncoding Genes through RNA Metabolism Profiles.” Nature Structural & Molecular Biology 24 (1): 86–96. *https://doi.org/10.1038/nsmb.3325.*

Neher, Peter F., Marc-Alexandre Côté, Jean-Christophe Houde, Maxime Descoteaux, and Klaus H. Maier-Hein. 2017. “Fiber Tractography Using Machine Learning.” NeuroImage 158 (September): 417–29. *https://doi.org/10.1016/j.neuroimage.2017.07.028.*

Neumann, Carsten, Gabriele Weiss, Sebastian Schmidtlein, Sibylle Itzerott, Angela Lausch, Daniel Doktor, and Maximilian Brell. 2015. “Gradient-Based Assessment of Habitat Quality for Spectral Ecosystem Monitoring.” Remote Sensing 7 (3): 2871–98. *https://doi.org/10.3390/rs70302871.*

NIC. 2018. „John von Neumann-Institut Für Computing (NIC).” 2018. *http://www.john-von-neumann-institut.de/nic/DE/Home/home_node.html.*

Norajitra, T., and K. H. Maier-Hein. 2017. “3D Statistical Shape Models Incorporating Landmark-Wise Random Regression Forests for Omni-Directional Landmark Detection.” IEEE Transactions on Medical Imaging 36 (1): 155–68. *https://doi.org/10.1109/TMI.2016.2600502.*

Northcott, Paul A., Ivo Buchhalter, A. Sorana Morrissy, Volker Hovestadt, Joachim Weischenfeldt, Tobias Ehrenberger, Susanne Gröbner, et al. 2017. “The Whole-Genome Landscape of Medulloblastoma Subtypes.” Nature 547 (7663): 311–17. *https://doi.org/10.1038/nature22973.*

Nostro, Alessandra D., Veronika I. Müller, Deepthi P. Varikuti, Rachel N. Pläschke, Felix Hoffstaedter, Robert Langner, Kaustubh R. Patil, and Simon B. Eickhoff. 2018. „Predicting Personality from Network-Based Resting-State Functional Connectivity“. Brain Structure and Function 223 (6): 2699–2719. *https://doi.org/10.1007/s00429-018-1651-z.*

NVIDIA. 2017. “How the Human Brain Project Maps the Brain Faster with Deep Learning.” June 20, 2017. *https://blogs.nvidia.com/blog/2017/06/20/human-brain-project/.*

Oakes, Christopher C., Marc Seifert, Yassen Assenov, Lei Gu, Martina Przekopowitz, Amy S. Ruppert, Qi Wang, et al. 2016. “DNA Methylation Dynamics during B Cell Maturation Underlie a Continuum of Disease Phenotypes in Chronic Lymphocytic Leukemia.” Nature Genetics 48 (3): 253–64. *https://doi.org/10.1038/ng.3488.*

Oca, A. Murillo Montes de, R. Bahmanyar, N. Nistor, and M. Datcu. 2017. “Earth Observation Image Semantic Bias: A Collaborative User Annotation Approach.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (6): 2462–77. *https://doi.org/10.1109/JSTARS.2017.2697003.*

OpenAI. 2018. “Gym.” 2018. *https://gym.openai.com/.*

Patil, Kaustubh R., Peter Haider, Phillip B. Pope, Peter J. Turnbaugh, Mark Morrison, Tobias Scheffer, and Alice C. McHardy. 2011. “Taxonomic Metagenome Sequence Assignment with Structured Output Models.” Nature Methods 8 (3): 191–92. *https://doi.org/10.1038/nmeth0311-191.*

Peukert, A., T. Schoening, E. Alevizos, K. Köser, T. Kwasnitschka, and J. Greinert. 2018. “Understanding Mn-Nodule Distribution and Evaluation of Related Deep-Sea Mining Impacts Using AUV-Based Hydro-acoustic and Optical Data.” Biogeosciences 15 (8): 2525–49. *https://doi.org/10.5194/bg-15-2525-2018.*

Preibisch, Stephan, Fernando Amat, Evangelia Stamataki, Mihail Sarov, Robert H. Singer, Eugene Myers, and Pavel Tomancak. 2014. “Efficient Bayesian-Based Multiview Deconvolution.” Nature Methods 11 (6): 645–48. *https://doi.org/10.1038/nmeth.2929.*

Rehme, A. K., L. J. Volz, D.-L. Feis, I. Bomilcar-Focke, T. Liebig, S. B. Eickhoff, G. R. Fink, and C. Grefkes. 2015. „Identifying Neuroimaging Markers of Motor Disability in Acute Stroke by Machine Learning Techniques“. Cerebral Cortex 25 (9): 3046–56. *https://doi.org/10.1093/cercor/bhu100.*

Ritter, Samuel, David G. T. Barrett, Adam Santoro, and Matt M. Botvinick. 2017. “Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study.” ArXiv:1706.08606 [Cs, Stat], June. *http://arxiv.org/abs/1706.08606.*

Ross, Tobias, David Zimmerer, Anant Vemuri, Fabian Isensee, Manuel Wiesenfarth, Sebastian Bodenstedt, Fabian Both, et al. 2018. “Exploiting the Potential of Unlabeled Endoscopic Video Data with Self-Supervised Learning.” International Journal of Computer Assisted Radiology and Surgery 13 (6): 925–33. *https://doi.org/10.1007/s11548-018-1772-0.*

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2014. “ImageNet Large Scale Visual Recognition Challenge.” ArXiv:1409.0575 [Cs], September. *http://arxiv.org/abs/1409.0575.*

Russell, Stuart. 2017. “Artificial Intelligence: The Future Is Superintelligent.” Nature 548 (7669): 520–521.

Russell, Stuart, and Peter Norvig. 2009. The Artificial Intelligence, 3e Preview Edition. Pearson Education, Limited.

S. Wehle et al. (Belle Collaboration). 2017. “Lepton-Flavor-Dependent Angular Analysis of $B\ensuremath{\rightarrow}{K}^{*}{\ensuremath{\ell}}^{+}{\ensuremath{\ell}}^{\ensuremath{-}}$.” Physical Review Letters 118 (11): 111801. *https://doi.org/10.1103/PhysRevLett.118.111801.*

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. 2017. “Dynamic Routing Between Capsules.” ArXiv:1710.09829 [Cs], October. *http://arxiv.org/abs/1710.09829.*

Sass, Steffen, Florian Buettner, Nikola S. Mueller, and Fabian J. Theis. 2013. “A Modular Framework for Gene Set Analysis Integrating Multilevel Omics Data.” Nucleic Acids Research 41 (21): 9622–33. *https://doi.org/10.1093/nar/gkt752.*

Scharr, H., M. J. Black, and H. W. Haussecker. 2003. “Image Statistics and Anisotropic Diffusion.” In Proceedings Ninth IEEE International Conference on Computer Vision, 840–47 vol.2. *https://doi.org/10.1109/ICCV.2003.1238435.*

Schirner, Michael, Anthony Randal McIntosh, Viktor Jirsa, Gustavo Deco, and Petra Ritter. 2018. “Inferring Multi-Scale Neural Mechanisms with Brain Network Modelling.” ELife 7 (January): e28927. *https://doi.org/10.7554/eLife.28927.*

Schlitter, Nico, Tanja Falkowski, and Jörg Lässig. 2013. “DenGraph-HO: A Density-Based Hierarchical Graph Clustering Algorithm.” Expert Systems 31 (5): 469–79. *https://doi.org/10.1111/exsy.12046.*

Schoening, Timm, Thomas Kuhn, Daniel O. B. Jones, Erik Simon-Lledo, and Tim W. Nattkemper. 2016. “Fully Automated Image Segmentation for Benthic Resource Assessment of Poly-Metallic Nodules.” Methods in Oceanography, Computer Vision in Oceanography, 15–16 (April): 78–89. *https://doi.org/10.1016/j.mio.2016.04.002.*

Schüller, Peter, and Mishal Benz. 2018. “Best-Effort Inductive Logic Programming via Fine-Grained Cost-Based Hypothesis Generation.” Machine Learning 107 (7): 1141–69. *https://doi.org/10.1007/s10994-018-5708-2.*

Schwarz, Roland F., Charlotte K. Y. Ng, Susanna L. Cooke, Scott Newman, Jillian Temple, Anna M. Piskorz, Davina Gale, et al. 2015. “Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis.” PLOS Medicine 12 (2): e1001789. *https://doi.org/10.1371/journal.pmed.1001789.*

Schwarz, Roland F., Anne Trinh, Botond Sipos, James D. Brenton, Nick Goldman, and Florian Markowetz. 2014. “Phylogenetic Quantification of Intra-Tumour Heterogeneity.” PLOS Computational Biology 10 (4): e1003535. *https://doi.org/10.1371/journal.pcbi.1003535.*

Sena, Diniz M., Xiaojing Cong, Alejandro Giorgetti, Achim Kless, and Paolo Carloni. 2017. „Structural Heterogeneity of the μ-Opioid Receptor’s Conformational Ensemble in the Apo State“. Scientific Reports 7 (1). *https://doi.org/10.1038/srep45761.*

Shprits, Yuri, Adam Kellerman, Dmitri Kondrashov, and Dmitriy Subbotin. 2013. “Application of a New Data Operator-Splitting Data Assimilation Technique to the 3-D VERB Diffusion Code and CRRES Measurements.” Geophysical Research Letters 40 (19): 4998–5002. *https://doi.org/10.1002/grl.50969.*

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. “Mastering the Game of Go without Human Knowledge.” Nature 550 (7676): 354–359.

Smith, J. S., O. Isayev, and A. E. Roitberg. 2017. “ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost †Electronic Supplementary Information (ESI) Available. See DOI: 10.1039/C6sc05720a Click Here for Additional Data File.” Chemical Science 8 (4): 3192–3203. *https://doi.org/10.1039/c6sc05720a.*

So2Sat. 2017. “ERC Starting Grant So2Sat.” 2017. *http://www.so2sat.eu/.*

Spitzer, Hannah, Katrin Amunts, Stefan Harmeling, and Timo Dickscheid. 2017. “Parcellation of Visual Cortex on High-Resolution Histological Brain Sections Using Convolutional Neural Networks.” ArXiv:1705.10545 [Cs], April, 920–23. h*ttps://doi.org/10.1109/ISBI.2017.7950666.*

Spitzer, Hannah, Kai Kiwitz, Katrin Amunts, Stefan Harmeling, and Timo Dickscheid. 2018. “Improving Cytoarchitectonic Segmentation of Human Brain Areas with Self-Supervised Siamese Networks.” ArXiv:1806.05104 [Cs], June. *http://arxiv.org/abs/1806.05104*.

Staneva, J., K. Wahle, H. Günther, and E. Stanev. 2016. “Coupling of Wave and Circulation Models in Coastal–Ocean Predicting Systems: A Case Study for the German Bight.” Ocean Sci. 12 (3): 797–806. *https://doi.org/10.5194/os-12-797-2016.*

Stanford Vision Lab. 2010. “ImageNet - Summary and Statistics.” April 30, 2010. *http://image-net.org/about-stats.*

Stanford Vision Lab. 2016. “ImageNet.” 2016. *http://www.image-net.org/.*

Steinberg, Andreas, Sabine Chabrillat, Antoine Stevens, Karl Segl, and Saskia Foerster. 2016. “Prediction of Common Surface Soil Properties Based on Vis-NIR Airborne and Simulated EnMAP Imaging Spectroscopy Data: Prediction Accuracy and Influence of Spatial Resolution.” Remote Sensing 8 (7): 613. *https://doi.org/10.3390/rs8070613.*

Svensson, Valentine, Sarah A. Teichmann, and Oliver Stegle. 2018. “SpatialDE: Identification of Spatially Variable Genes.” Nature Methods 15 (5): 343–46. *https://doi.org/10.1038/nmeth.4636.*

Telenti, Amalio, Christoph Lippert, Pi-Chuan Chang, and Mark DePristo. 2018. “Deep Learning of Genomic Variation and Regulatory Network Data.” Human Molecular Genetics 27 (R1): R63–71. *https://doi.org/10.1093/hmg/ddy115.*

The Alan Turing Institute. 2016. “Call for Visiting Researchers at The Alan Turing Institute.” The Alan Turing Institute (blog). September 21, 2016. *https://www.turing.ac.uk/media/news/call-visiting-researchers-alan-turing-institute/.*

Theis, Fabian. 2006. “Towards a General Independent Subspace Analysis.” Proceedings of the 2006 Conference, Adv. Neural Inf. Process. Syst. 19 (January): 1361–68.

Turajlic, Samra, Hang Xu, Kevin Litchfield, Andrew Rowan, Stuart Horswell, Tim Chambers, Tim O’Brien, et al. 2018. “Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal.” Cell 173 (3): 595-610.e11. *https://doi.org/10.1016/j.cell.2018.03.043.*

UCLA. 2018. “Postdoctoral Scholars | UCLA Graduate Programs.” 2018. *https://grad.ucla.edu/funding/working-at-ucla/postdocs/.*

Ulman, Vladimír, Martin Maška, Klas E. G. Magnusson, Olaf Ronneberger, Carsten Haubold, Nathalie Harder, Pavel Matula, et al. 2017. "An Objective Comparison of Cell-Tracking Algorithms." Nature Methods 14 (12): 1141–52. *https://doi.org/10.1038/nmeth.4473.*

Urbanski, Marika, Olivier A. Coubard, and Clémence Bourlon. 2014. "Visualizing the Blind Brain: Brain Imaging of Visual Field Defects from Early Recovery to Rehabilitation Techniques." Frontiers in Integrative Neuroscience 8 (September). *https://doi.org/10.3389/fnint.2014.00074.*

Varikuti, Deepthi P., Sarah Genon, Aristeidis Sotiras, Holger Schwender, Felix Hoffstaedter, Kaustubh R. Patil, Christiane Jockwitz, et al. 2018. „Evaluation of Non-Negative Matrix Factorization of Grey Matter in Age Prediction". NeuroImage 173 (Juni): 394–410. *https://doi.org/10.1016/j.neuroimage.2018.03.007.*

Vector Institute. 2018. "Vector Institute." 2018. *https://vectorinstitute.ai/.*

Viader-Llargués, Oriol, Valerio Lupperger, Laura Pola-Morell, Carsten Marr, and Hernán López-Schier. 2018. "Live Cell-Lineage Tracing and Machine Learning Reveal Patterns of Organ Regeneration." ELife 7 (March). *https://doi.org/10.7554/eLife.30823.*

Victor de Araujo Oliveira, João, Fabrizio Costa, Rolf Backofen, Peter Stadler, Maria Walter, and Jana Hertel. 2016. "SnoReport 2.0: New Features and a Refined Support Vector Machine to Improve SnoRNA Identification." BMC Bioinformatics 17 (December). *https://doi.org/10.1186/s12859-016-1345-6.*

Wagstyl, Konrad, Claude Lepage, Sebastian Bludau, Karl Zilles, Paul C Fletcher, Katrin Amunts, and Alan C Evans. 2018. „Mapping Cortical Laminar Structure in the 3D BigBrain". Cerebral Cortex 28 (7): 2551–62. *https://doi.org/10.1093/cercor/bhy074.*

Wahle, Kathrin, Joanna Staneva, and Heinz Guenther. 2015. "Data Assimilation of Ocean Wind Waves Using Neural Networks. A Case Study for the German Bight." Ocean Modelling, Waves and coastal, regional and global processes, 96 (December): 117–25. *https://doi.org/10.1016/j.ocemod.2015.07.007.*

Wang, Y., and X. X. Zhu. 2016. "Robust Estimators for Multipass SAR Interferometry." IEEE Transactions on Geoscience and Remote Sensing 54 (2): 968–80. *https://doi.org/10.1109/TGRS.2015.2471303.*

Weichselbaum, Ewald, Maria Österbauer, Denis G. Knyazev, Oleg V. Batishchev, Sergey A. Akimov, Trung Hai Nguyen, Chao Zhang, et al. 2017. „Origin of Proton Affinity to Membrane/Water Interfaces". Scientific Reports 7 (1). *https://doi.org/10.1038/s41598-017-04675-9.*

Weimann, Aaron, Kyra Mooren, Jeremy Frank, Phillip B. Pope, Andreas Bremges, and Alice C. McHardy. 2016. "From Genomes to Phenotypes: Traitar, the Microbial Trait Analyzer." MSystems 1 (6): e00101-16. *https://doi.org/10.1128/mSystems.00101-16.*

Werner, Stefan, Benedikt Brors, Julia Eick, Elsa Marques, Vivian Pogenberg, Annabel Parret, Dirk Kemming, et al. 2015. "Suppression of Early Hematogenous Dissemination of Human Breast Cancer Cells to Bone Marrow by Retinoic Acid–Induced 2." Cancer Discovery 5 (5): 506–19. *https://doi.org/10.1158/2159-8290.CD-14-1042.*

Wirkert, Sebastian J., Hannes Kenngott, Benjamin Mayer, Patrick Mietkowski, Martin Wagner, Peter Sauer, Neil T. Clancy, Daniel S. Elson, and Lena Maier-Hein. 2016. "Robust near Real-Time Estimation of Physiological Parameters from Megapixel Multispectral Images with Inverse Monte Carlo and Random Forest Regression." International Journal of Computer Assisted Radiology and Surgery 11 (6): 909–17. *https://doi.org/10.1007/s11548-016-1376-5.*

Yang, Xiaogang, Vincent De Andrade, William Scullin, Eva L. Dyer, Narayanan Kasthuri, Francesco De Carlo, and Doğa Gürsoy. 2018. “Low-Dose x-Ray Tomography through a Deep Convolutional Neural Network.” Scientific Reports 8 (1): 2575. *https://doi.org/10.1038/s41598-018-19426-7.*

Yeo, B. T. Thomas, Mert R. Sabuncu, Tom Vercauteren, Daphne J. Holt, Katrin Amunts, Karl Zilles, Polina Golland, and Bruce Fischl. 2010. „Learning Task-Optimal Registration Cost Functions for Localizing Cytoarchitecture and Function in the Cerebral Cortex“. IEEE Transactions on Medical Imaging 29 (7): 1424–41. *https://doi.org/10.1109/TMI.2010.2049497.*

Yuan, Yinyin, Henrik Failmezger, Oscar M. Rueda, H. Raza Ali, Stefan Gräf, Suet-Feung Chin, Roland F. Schwarz, et al. 2012. “Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling.” Science Translational Medicine 4 (157): 157ra143-157ra143. *https://doi.org/10.1126/scitranslmed.3004330.*

Zhelavskaya, Irina S., Yuri Y. Shprits, and Maria Spasojević. 2017. “Empirical Modeling of the Plasmasphere Dynamics Using Neural Networks.” Journal of Geophysical Research: Space Physics 122 (11): 11,227-11,244. *https://doi.org/10.1002/2017JA024406.*

Zhelavskaya, Irina, M Spasojevic, Yuri Shprits, and William Kurth. 2016. “Automated Determination of Electron Density from Electric Field Measurements on the Van Allen Probes Spacecraft: AUTOMATED ELECTRON DENSITY DETERMINATION.” Journal of Geophysical Research: Space Physics 121 (May). *https://doi.org/10.1002/2015JA022132.*

Zhu, X. X., C. Grohnfeldt, and R. Bamler. 2016. “Exploiting Joint Sparsity for Pansharpening: The J-SparseFI Algorithm.” IEEE Transactions on Geoscience and Remote Sensing 54 (5): 2664–81. *https://doi.org/10.1109/TGRS.2015.2504261.*

Zhu, X. X., D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. 2017. “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources.” IEEE Geoscience and Remote Sensing Magazine 5 (4): 8–36. *https://doi.org/10.1109/MGRS.2017.2762307.*

DANKSAGUNG

Wir bedanken uns bei allen Expertinnen und Experten aus der Helmholtz-Gemeinschaft die bisher am Inkubator-Prozess mitgewirkt haben und durch den Einsatz das vorliegende Dokument ermöglicht haben. Insbesondere möchten wir den Teilnehmerinnen und Teilnehmern der Arbeitsgruppe Künstliche Intelligenz und Maschinelles Lernen danken:

- Katrin Amunts (FZJ)
- Richard Bamler (DLR)
- Bank Beszteri (AWI)
- Stefan Bonn (DZNE)
- Mihai Datcu (DLR)
- Timo Dickscheid (FZJ)
- Frederic Effenberger (GFZ)
- Stephan Frickenhaus (AWI)
- Markus Götz (KIT)
- Volker Gülzow (DESY)
- Jörg Hackermüller (UFZ)
- Rolf Hempel (DLR)
- Jens-Uwe Hoffmann (HZB)
- Matthew Robert Huska (MDC)
- Frank Jenko (IPP)
- Jenia Jitsev (FZJ)
- Guido Juckeland (HZDR)
- Angela Jurik-Zeiller (HMGU)
- Judith Katzy (DESY)
- Jeffrey Kelling (HZDR)
- Rainer Kiko (GEOMAR)
- Inga Monika Koszalka (GEOMAR)
- Dirk Krücker (DESY)
- Thomas Lippert (FZJ)
- Klaus H. Maier-Hein (DKFZ)
- Alice McHardy (HZI)
- Lars Mehwald (Geschäftsstelle)
- Andreas B. Meyer (DESY)
- Sach Mukherjee (DZNE)
- Emmanuel Müller (GFZ)
- Uwe Ohler (MDC)
- Alexander Pichler (DESY)
- Hanno Scharr (FZJ)
- Timm Schoening (GEOMAR)
- Corinna Schrum (HZG)
- Roland Schwarz (MDC)
- Martin Siggel (DLR)
- Achim Streit (KIT)
- Fabian Theis (HMGU)
- Frederik Tilmann (GFZ)
- Susanne Wenzel (FZJ)
- Martin Werner (DLR)
- Xiaoxiang Zhu (DLR)
- Eduardo Zorita (HZG)

www.helmholtz.de