



Für das KI-Servicezentrum werden an den Standorten Leibniz Universität Hannover/LUIS und Georg-August-Universität Göttingen/GWDC Infrastrukturdienste in Form von HPC-Hardware, Speicherressourcen und diversen darauf aufbauenden Diensten und Softwarelösungen bereitgestellt. Weiterhin können Modelle und Daten publiziert und aus einem kuratierten Angebot bezogen werden.

Infrastruktur: Rechenressourcen

Die technische Grundlage für Training von KI-Modellen und zeitkritische Inferenz. Folgende System wurden bzw. werden bereits beschafft:

📊 **GPU-Knoten 2022 (Training):** 35 Knoten am Standort Göttingen

- 2x AMD Epyc 7513 32c, 512 GB DDR4 RAM, 2x 960 GB NVMe SSD
- 4x NVIDIA A100 (80 GB, SXM4 Sockel, 6.912/432 CUDA-/Tensorkerne)
- NVLink (HGX "Redstone"), 2x 200 GBit/s InfiniBand HDR

📊 **GPU-Knoten 2023 (Inferenz):** 16+ Knoten an beiden Standorten

- 2x Intel Xeon Platinum 8470 52c, 512 GB DDR5 RAM, 2x 3,84 GB NVMe SSD
- 4x NVIDIA H100 (80 GB, PCIe, 14.592/456 CUDA-/Tensorkerne)
- 2x 100 GBit/s Ethernet



Infrastruktur: Speicherressourcen

Bei den Storage-Systemen ist Performance und flexibler Zugriff entscheidend. Für das Training von KI-Modellen wurde folgende Lösung gewählt:

📁 **Storage 2022 (Trainingsplattform):**

- VAST All-Flash System
- 8x Stateless Server (CBox, 4x IB, 4x Ethernet), 1x NVMe Enclosure (DBox)
- 600 TB nutzbare Kapazität, POSIX+S3 Zugriff
- Flexible Anbindung InfiniBand/RoCE



Infrastruktur: Entwicklungsplattformen

Die KISSKI-Entwicklungssysteme bieten Zugriff auf Alternativen zu klassischen CPU-/GPU-Systemen. Beschaffungen erfolgen flexibler, je nach Verfügbarkeit.

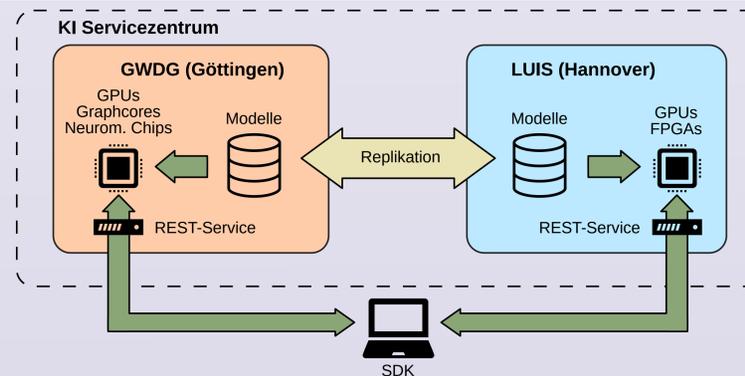
👉 **Neue Architekturen:** Graphcore, FPGAs, Neuromorphe Chips, AMD Instinct GPUs, Esperanto.ai RISC-V Beschleuniger, SambaNova, ...



Infrastruktur: Services

Die effiziente Nutzung der Hardwareressourcen, Modelle und Daten wird durch die folgenden am KISSKI entwickelten Dienste ermöglicht:

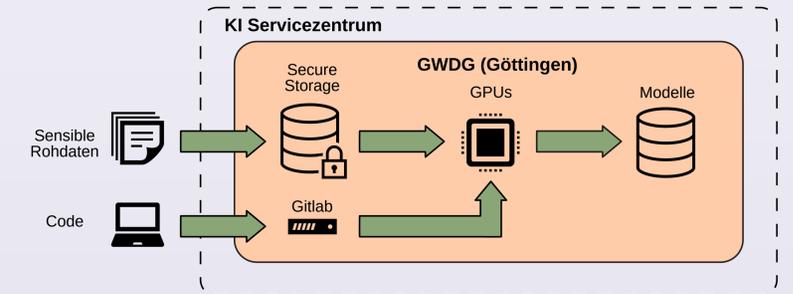
- 🔒 **Secure Scalable Container Hosting:** Verschlüsselte Images und Ein-/Ausgabe.
- 🛡️ **Fault-tolerant Inference Service:** Hohe Verfügbarkeit durch Standortredundanz.
- 📊 **Data Lake:** Datenherkunft, Datensouveränität, Reproduzierbarkeit.
- 🔄 **Datenaustausch:** Einfache Integration mit dem lokalen Arbeitsplatzrechner.
- 🌐 **Bereitstellung von Web-Plattformen:** Ermöglicht GUI-Zugriff auf APIs.
- 📦 **Bereitstellung Repositories:** Hosting interner und nutzergenerierter Container.



Software

Das Angebot umfasst die Bereitstellung und Weiterentwicklung von KI-Software:

- 📖 **Katalog:** Kuratierte Sammlung mit Webfrontend und semantischer Suche.
- 📦 **Containerdatenbank:** Einfache Softwarenutzung auch auf Kundensystemen.
- 🔒 **Secure Platform-/Software-as-a-Service (SPaaS/SSaaS):** Gehärtete Entwicklungs- und Analyseumgebungen mit einschlägigen Bibliotheken (z.B. Python-Module, TensorFlow, Dask), Nutzung der Daten und Modelle des KISSKI möglich.
- 🕒 **Container-Blaupausen für zeitkritische Inferenz:** Nutzung der fehlertoleranten Inferenzplattform, Analyse z.B. von Wetter-, Energiemarkt, oder Patientendaten, KI-gestützte Anomaliedetektion.



Modelle und Daten

Das KI-Servicezentrum bereitet Artefakte aus den Themenbereichen der Projektpartner auf und stellt über seine Plattform auch nutzergenerierte Inhalte bereit:

- ⚡ **Energie**
 - Aufbereitete Daten des Energiesektors, z.B. Wetter, Energiehandel, Messzeitreihen
 - Modelle zur Lösung energiewirtschaftlicher Fragestellungen, z.B. Lastprognosen
- 🏥 **Medizin**
 - Multidimensionale stationäre Routinedaten aus dem UMG-MeDIC und der MHH
 - KI-Prädiktionsmodelle in der klinischen Medizin und Gesundheitsversorgung
 - Trainings- und Validierungsdaten für Bild- und KI-Verfahren, z.B. physionet
- 🛠️ **Dienste für alle Themenbereiche**
 - Daten- und Modell-Katalog: Referenz auf intern und extern gehostete Artefakte nach Qualitätsprüfung, bequeme Übersicht als Webfrontend.
 - Publikationsplattform: Einfache Veröffentlichung nutzergenerierter Datensätze und vortrainierter Modelle inkl. Publikationsmetadaten und persistenter DOIs.

Management

Bereitstellung

Consulting

Produkt-
entwicklung

Angebote

Skalierbare KI

Automatische
Skalierung

Heterogene
Hardware

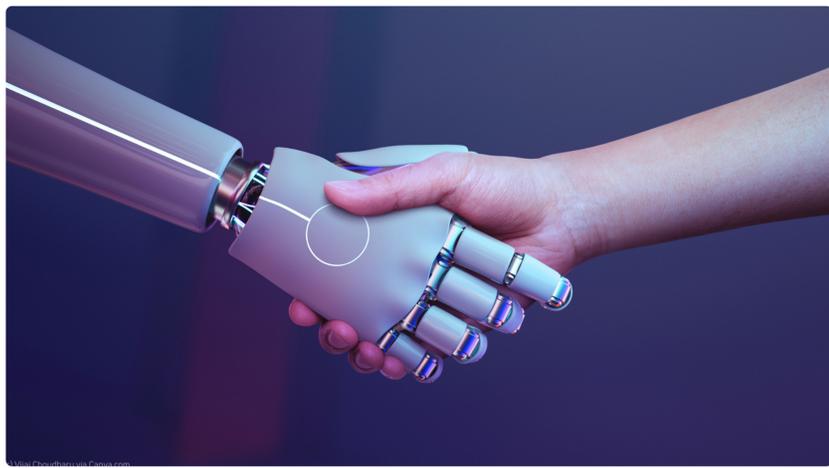
Daten-
management



Die Angebote sind von allen KISSKI Nutzenden buchbar. Neben einer Erstberatung im Bereich Medizin und Energie sollen auch Kursangebote geschaffen werden um Grundlagen auffrischen zu können

Einstiegsberatung
Die Beratungen umfassen unter anderem:

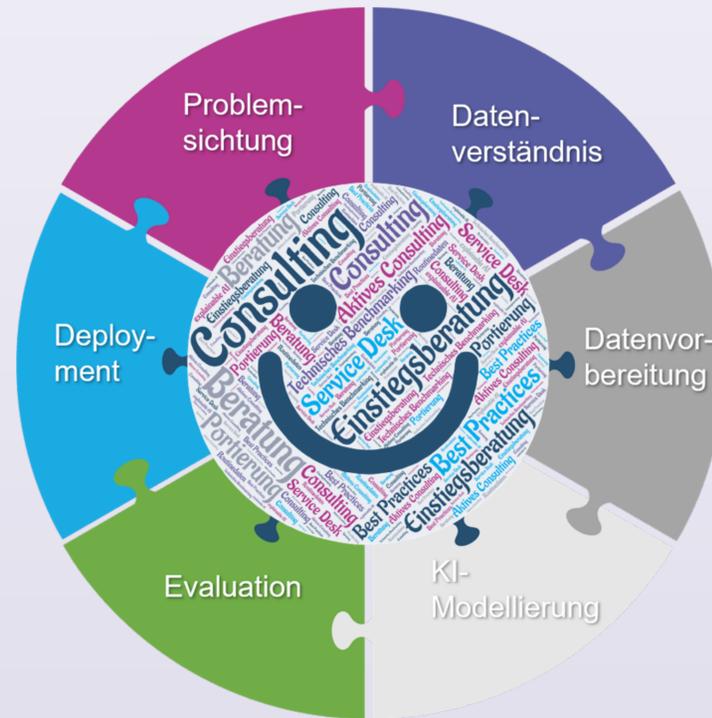
- ♥ **Medizin:** Erstberatung für die Durchführung medizinischer Experimente mit KI-Komponenten
- ♥ **klinische Studien:** Kundenspezifische Beratung mit Bedarf an Biostatistik und/oder Bioinformatik
- ♥ **Impact Analysen:** Implementierung geeigneter (adaptiver) Studiendesigns, Fallzahl und Powerberechnung
- ♥ **Medizinische Bilddaten:** Erstberatung zu geeigneten Methoden zur KI-basierten Analyse medizinischer Bilddaten. U.A. zu Faltungsnetzen und adverseriellen Beispielen
- ♥ **Routinedaten:** Erstberatung bei Routinedaten, kategorische und/oder numerische Daten oder Messreihen. Fachberatung zu Routinedaten und Datenschutzauflagen
- ♥ **Data-Science im Gesundheitssektor:** Anwendungsgebiete und besondere Herausforderungen bei der Datenanalyse.



- ✂ **Energie:** Einstiegsberatung für Unternehmen aus dem Bereich Energie.
- ✂ **Neuromorphic Computing:** Beratung zur Eignung einer Implementieren von Problemstellungen auf neuromorpher Hardware.
- ✂ **Import bestehender Kurse:** Lokale Angebote der Konsortialpartner werden für Kunden des Servicezentrums geöffnet.
- ✂ **KI und Digitalisierung für KMU:** Erarbeitung von Kursen in Zusammenarbeit mit dem Mittelstand-Digital Zentrum Hannover.
- ✂ **Data-Science im Energiesektor:** Anwendungsgebiete im Energiesektor für verschiedene KI Fragestellungen.

Service Desk
Beratungsleistungen, insbesondere zur Nutzung der bereitgestellten Infrastruktur

- ⌚ **Technisches Benchmarking** Leistungsanalyse, Tuning und Optimierung eines KI-Modells zur effizienteren Nutzung
- ✔ **Best Practices** Diese werden dokumentiert und den Nutzer:innen von KISSKI zur Verfügung gestellt
- ✔ **Aktives Consulting** Fachberater:innen können dann aus ihrer Erfahrung während der Entwicklung von Best Practices fundiert beraten
- ✔ **Portierung** auf neuromorphe Systeme, GraphCore oder FPGA



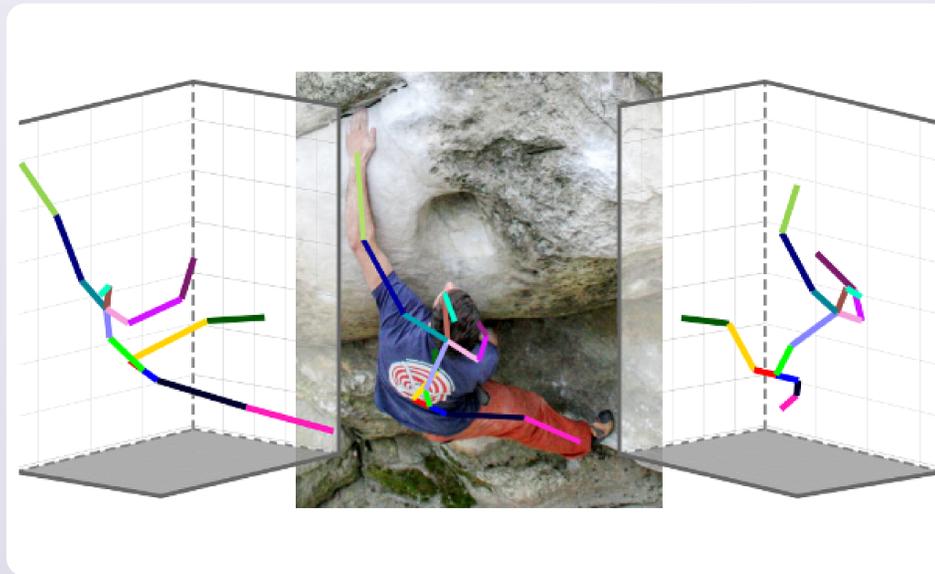
Daten

- 📄 **Beratung** zur Interpretation, Aufbereitung und Validierung von Daten
- 📄 **Datenmanagement** Datenmanagement und Datenaufbereitung im Kontext klinischer Studien und medizinischer Projekte sowie von Bild- und Biosignalen und den Bereich Energie
- 📄 **Routinedaten:** Konzeption und Zusammenstellung von Informationsmaterialien zur Aufbereitung von Routinedaten von Krankenkassen

Pakete
Geschlossene Module für Einsteiger und Fortgeschrittene

- 👍 **KI-Starter Pack:** Sie haben noch keine Erfahrung mit KI und suchen nach einem Startpunkt.
- 👍 **Innovative AI** Dieses Paket richtet sich an Unternehmen und Forschungsinstitute, die KI in Domänen einsetzen möchten, für die es bis jetzt noch keine KI-Lösungen gibt.
- 👍 **Software, Modelle und Daten** Bietet die Beratung über die Nutzung der im Servicezentrum angebotenen vor-trainierten Modelle, vorhandenen Daten und Software an.
- 👍 **AutoML** Automatisierung von KI-Pipelines

🎓 **Weiterbildung** Live-Long Learning und Workshops. Dies erfolgt in enger Absprache zu AP5



👍 **explainable AI** Erklärbarkeit ist ein wichtiger Baustein sowohl zur Erhöhung des Nutzens als auch zur Erhöhung der Akzeptanz KI-basierter Vorschläge und Entscheidungen.





Die Services rund um die Produktentwicklung richten sich an Unternehmen aus den Domänen **Medizin und Energie**, die KI-Anwendungen konkret umsetzen und dabei unterstützt werden wollen. Das Servicezentrum kann diese **Unterstützung in zwei Ausprägungen** bieten, die projektspezifisch ausgewählt werden können:

Begleitung der Projektumsetzung

Die Fachexpert:innen des Servicezentrums begleiten das Unternehmen bei der **eigenständigen Entwicklung** einer KI-Lösung.

Dem Unternehmen steht dabei die Nutzung der KISSKI-Infrastruktur offen.

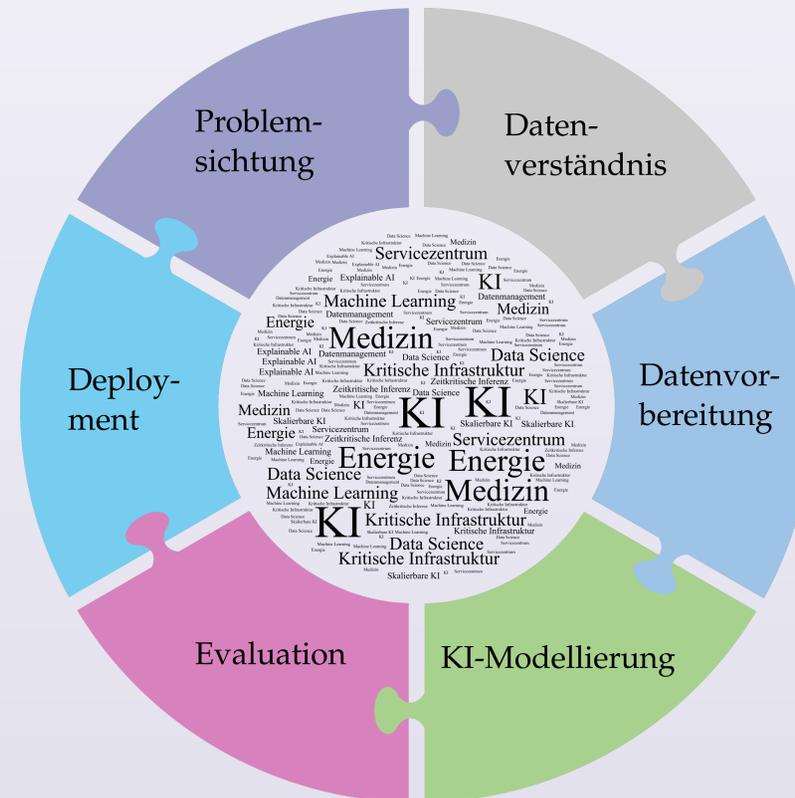
Daten
handling

Modell-
evaluation

Auswahl
KI-
Methode

Modell-
inter-
pretation

Planung
KI-
Konzept



Entwicklungsleistung

Data
Analysis
& Data
Science

Opti-
mierung
KI-
Modell

Methoden-
Kali-
brierung

Entwick-
lung KI-
Methode

Transfer
in oper-
ativen
Betrieb

Die **Fachexpert:innen** des Servicezentrums **entwickeln** für Unternehmen KI-Lösungen oder Module für KI-Anwendungen.

Produkte und Projekte können in die **KISSKI-Infrastruktur** integriert werden und bereitgestellte Hard- und Software nutzen.

Die Serviceleistungen können den gesamten Prozess einer KI-Anwendung sowie einzelne Teile davon abdecken.

Management

Bereitstellung

Consulting

**Produkt-
entwicklung**

Angebote

Skalierbare KI

Automatische
Skalierung

Heterogene
Hardware

Daten-
management



Die Angebote sind von allen KISSKI Nutzenden buchbar. Die Nutzercommunity wird auf die Angebote und besondere Ereignisse aufmerksam gemacht.

Schulungen

Die Schulungen werden zu festen Terminen mehrmals jährlich angeboten und umfassen unter anderem:

? Einführung des Servicezentrums: Vorstellung des KISSKI mit der bereitgestellten Infrastruktur und Software.

🔗 Modelle/Daten: Kundenspezifische Schulung der angebotenen Daten und Modelle.

⌚ Zeitkritische Inferenz: Schulungen zu Lösungen für zeitkritische und sichere Anwendungsfälle aus Medizin und Energie.

🔗 Heterogene Systeme: Anwendungsbezogene Kurse für heterogene Systeme und die KISSKI Hardwareplattformen.

📁 Datenmanagement: Kurse zu effektivem Datenmanagement, Datensicherheit und Datenaufbereitung, sowie Rahmenbedingungen für kritische Energieinfrastruktur und Gesundheitsdaten.

📊 Data-Science im Gesundheitssektor: Anwendungsgebiete und besondere Herausforderungen bei der Datenanalyse.

Öffentlichkeitsarbeit

Im Bereich der Öffentlichkeitsarbeit umfasst das Angebot u.A.

👥 Community und Kooperationsnetzwerk: Aufbau einer offenen Community durch Etablierung von Kommunikationsinfrastrukturen, gemeinsame Events und regionale und deutschlandweite Workshops.

👥 KISSKI-Kongress: 18. & 19. September 2024 in Göttingen. Für einen weiten und direkten Austausch.

Wissenschaftskommunikation

Unterstützung der Nutzenden bei der Kommunikation ihrer KI-Ansätze und -Lösungen.

🗣️ Kommunikationsservice: Aufarbeitung und Veröffentlichung der KI-Lösungen für die breite Öffentlichkeit sowohl auf den KISSKI Publikationskanälen als auch über weitere Presseorgane.

👍 Akzeptanz von KI: Information über Grenzen und Möglichkeiten der KI-Anwendungen zum Abbau von Misstrauen der Gesellschaft und Erhöhung der Akzeptanz.



📅 Organisation von Outreach Events: Beteiligung an öffentlichkeitswirksamen Veranstaltungen. Es werden neue Formate geschaffen wie Workshops zum Thema "KI für Jedermann".



📈 Skalierbare KI: Kurse zum Thema (automatisiertes) Performance Engineering.

📁 Import bestehender Kurse: Lokale Angebote der Konsortialpartner werden für Kunden des Servicezentrums geöffnet.

📁 KI und Digitalisierung für KMU: Erarbeitung von Kursen in Zusammenarbeit mit dem Mittelstand-Digital Zentrum Hannover.

📊 Data-Science im Energiesektor: Anwendungsgebiete im Energiesektor für verschiedene KI Fragestellungen.

📶 Online-Präsenz: Veröffentlichung und Bereitstellung von Informationen.

📅 Koordination von Outreach Events: Koordination der Nutzenden und Unterstützung bei der Beteiligung an Veranstaltungen.

📁 Publikations-Service: Beratung und Hilfestellung bei der Publikation von am KISSKI durchgeführten Projekten und Forschungsarbeiten.

📚 Material für Schüler:innen: Angebot von interaktivem Material, um die Möglichkeiten von KI-Anwendungen insbesondere in den Bereichen Energie und Gesundheit aufzuzeigen.

Management

Bereitstellung

Consulting

Produkt-
entwicklung

Angebote

Skalierbare KI

Automatische
Skalierung

Heterogene
Hardware

Daten-
management



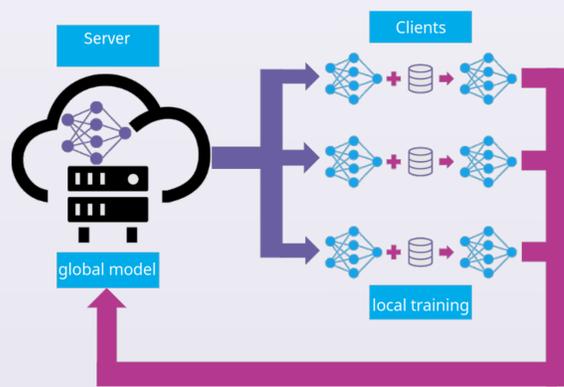
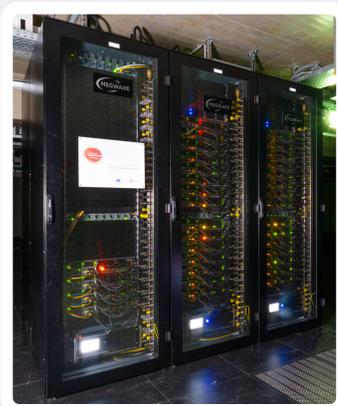
Skalierbare KI beinhaltet die Aspekte der Sicherheit und Skalierbarkeit für KI-Training und Inferenz. Inferenz muss hochverfügbar unter der Beachtung von Deadlines sein und es dürfen keine Daten exponierbar oder inferierbar sein.

Skalierbare KI

Aufgrund steigender Anforderungen der KI Modelle werden verschiedene Ansätze zur effizienten Skalierung von Training und Inferenz evaluiert.

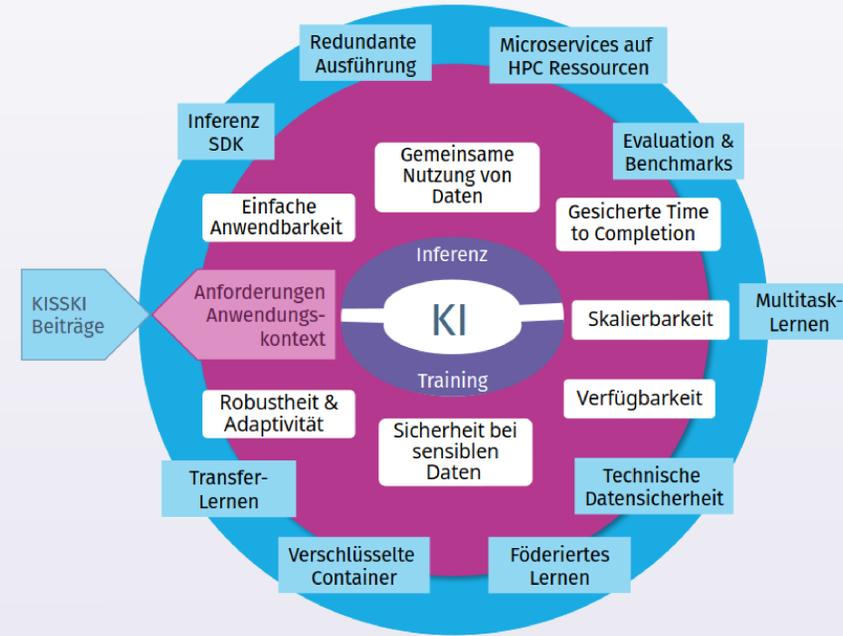
Rechenkapazität: Die notwendige Rechenkapazität für das Training von KI Modelle verdoppelt sich alle 3,4 Monate - Zeitraum der Analyse 2012 - 2018 (OpenAI 2018).

Energieverbrauch: Das Training eines Transformer Modells mit 6 Milliarden Parametern verbraucht ~10M CO₂ (Dodge et al. 2022). Dies entspricht dem Energieverbrauch eines US-amerikanischen Haushalts pro Jahr.



Hardwarebeschleuniger: Sowohl für das Training als auch für die Inferenz ist der Einsatz von Hardwarebeschleunigern unerlässlich. Daher werden für diese Aufgabe optimierte GPUs und IPUs zur Verfügung gestellt und evaluiert.

Parallelisierung: Sowohl die zunehmende Größe von KI-Modellen, als auch die für das Training erforderlichen Datenmengen machen verteiltes Training und verteilte Inferenz notwendig. Für die effiziente Nutzung verteilter Systeme wird Parallelisierung sowohl auf Model- ("model parallelism") als auch auf Datenebene ("data parallelism") eingesetzt.



Sicheres Training

Gemeinsame Nutzung von Daten und Modellen kann potentielle Wettbewerbsnachteile durch Datenarmut entgegenwirken und die Ergebnisse für alle Teilnehmer verbessern.

Multi-Task und Transfer Learning: Beim Multi-Task Learning lernt ein Modell simultan mehrere Aufgaben gleichzeitig, während beim Transfer Learning ein vortrainiertes Modell für eine spezifische neue Zielaufgabe eingesetzt wird.

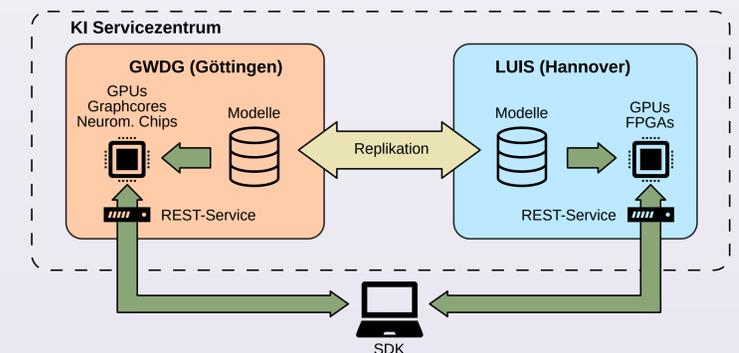
Federated Learning: Federated Learning nutzt dezentrale Datenquellen, um so hohen Sicherheits- und Datenschutzanforderungen zu entsprechen. Der zentrale Server aggregiert nur die Gewichte der Teilnehmer in einem Modell und erhält keine Rohdaten.

Inferenz Mit Deadlines

Nutzer:innen können KI Modelle auf unseren Servern installieren lassen. Jedes Modell ist über eine eigene URL nutzbar.

Automatische Skalierung: Modelle werden über Inferenz Software in Containern ausgerollt. Eine Kubernetes Plattform orchestriert diese und passt die Anzahl an Containern für jedes Modell basierend auf dessen Nutzung aktueller Nutzung an.

GPU Beschleunigung: Viele KI Modelle profitieren von GPU Ressourcen für schnellere Inferenz. Durch Virtualisierung der GPUs können verschiedene Modelle parallel auf der selben GPU rechnen.

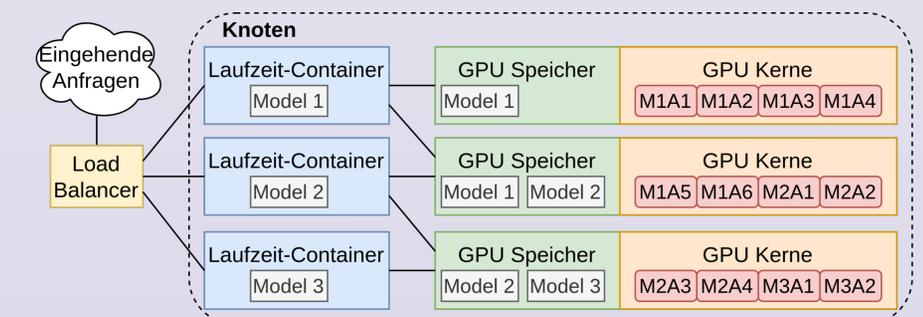
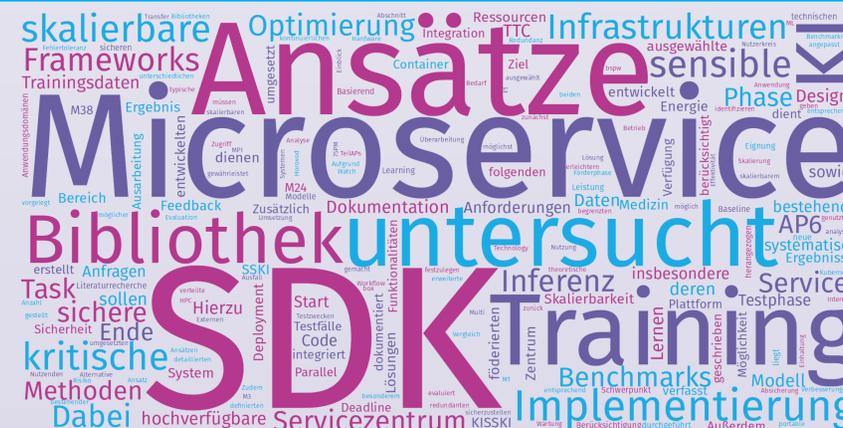


Effiziente GPU Nutzung

Die Nutzung von Grafikkarten wird für Auslastung, Skalierbarkeit und schnelle Antworten optimiert.

Parallele Nutzung: Modelle werden bei Bedarf in Laufzeit-Container geladen und von dort in den GPU Speicher. Mehrere Anfragen für ein Model können parallel auf der gleichen GPU bearbeitet werden.

GPU Teilung: GPU Ressourcen werden aufgeteilt, sodass verschiedene Modelle in den Speicher der selben GPU geladen werden können. Dies erlaubt es mehr Modelle gleichzeitig im GPU Speicher zu halten und so auf Anfragen sofort zu reagieren.





Die Erforschung heterogener Hardware erschließt das Potential von KI-Applikationen durch schnellere, energieeffizientere und kostengünstigere Umsetzung künstlicher Intelligenz.

Heterogene Hardware
Wissen über die verfügbaren Hardwarearchitekturen aufbauen und vermitteln sowie Vor-/Nachteile erkennen.

Benchmarks
Über Benchmarks werden die heterogenen Hardwareplattformen auf deren Eigenschaften untersucht.

Kosten-/Performance-Modell
Über ein Modell können Vorhersagen für Inferenzen Neuronaler Netze getroffen und die optimale Hardware bestimmt werden.

CPU-/GPU-Systeme
CPU-/GPU-Systeme sind die am weitesten verbreiteten Hardwareplattformen zum Berechnen neuronaler Netze. GPUs eignen sich insbesondere durch ihre Verfügbarkeit, softwareseitige Flexibilität und hohem Parallelisierungsgrad zum Berechnen regulärer Operationen, wie sie in neuronalen Netzen verwendet werden.

Graphcore-Systeme
Graphcore-Systeme nutzen Intelligent Processing Units (IPU), spezialisierte Hardware, zum effizienten Berechnen neuronaler Netze. Während der Parallelisierungsgrad und die Energieeffizienz im Vergleich zu GPUs höher sind, sind sie in der Flexibilität begrenzt.

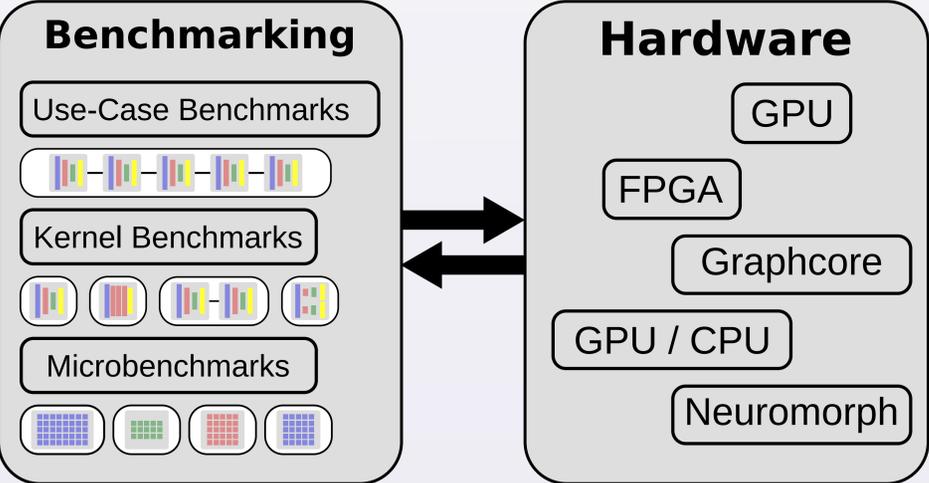
Neuromorphe Systeme
Neuromorphe Systeme basieren auf künstlichen Neuronen, welche dem Prinzip biologischer Neuronen ähneln. Durch Datenkodierung mit zeitlichen Impulsen (Spikes) sind sie energieeffizient und insbesondere nützlich für zeitabhängige Datenreihen. Einige gängige Machine-Learning-Algorithmen sind allerdings schwer darauf zu implementieren bzw. zu trainieren [1].



Xilinx Alveo U280 FPGA [2]



USB-Träger mit zwei neuromorphen Chips/Intel Loihi (Abbildung aus [1])



Benchmarks
Microbenchmarks bilden grundlegende Operationen und Elemente neuronaler Netze ab. Über Microbenchmarks kann die Performance verschiedener Hardwarearchitekturen miteinander verglichen werden, jedoch ist ein Transfer auf angewendete Modelle erschwert.
Kernel-/Synthetische Benchmarks bilden parametrisierte Anteile, wie Blöcke, und Variationen größerer Modelle ab, sodass Vorhersagen für Netze bestehend aus den ähnlichen Anteilen getroffen und ein Vergleich der Hardwarearchitekturen für Applikationen durchgeführt werden können.
Use-Case Benchmarks bilden vollständige Netze und Anwendungen auf den Architekturen ab, sodass die Hardwarearchitekturen direkt miteinander verglichen werden können. Ein einfacher Transfer der Ergebnisse und eine Prädiktion für andere Netze ist nicht direkt möglich. Zusätzlich können Modelle und Anwendungen aus den anderen Arbeitspaketen verwendet werden um Beispiele für Use-Cases auf den heterogenen Architekturen zu implementieren.

Modellbasiertes Framework

- Relevante Metriken definieren
- Hardware untersuchen
- Prädiktionsframework erstellen
- Entwurfsraum für Use-Cases aufstellen

FPGA-Systeme
FPGA-Systeme ermöglichen durch rekonfigurierbare Hardware das Abbilden verschiedenster Beschleunigerarchitekturen. Dadurch können andere Hardwarearchitekturen emuliert und getestet werden. Durch Optimierung der abgebildeten Architekturen kann eine hohe Energieeffizienz erreicht werden.

Literatur

[1] Mike Davies et al. "Advancing neuromorphic computing with Loihi: a survey of results and outlook". In: *Proceedings of the IEEE* 109.5 (2021), pp. 911–934.
[2] URL: <https://www.xilinx.com/products/boards-and-kits/alveo/u280.html> (visited on 05/23/2023).





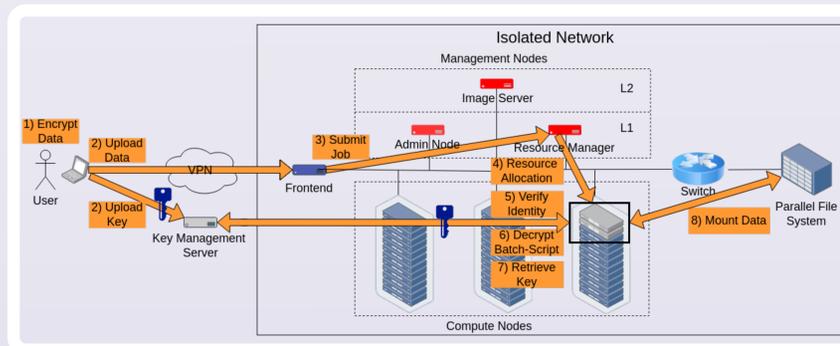
Das Ziel dieser Forschungsarbeit ist die Entwicklung von Methoden für das sichere und effiziente Datenmanagement auf heterogenen Systemen, welches für manuelle und automatisierte Prozesse gleichermaßen genutzt werden kann.

Sichere Workflows

Zum Verarbeiten DSGVO-geschützter Gesundheitsdaten wurde bei der GWDG ein Prozess entwickelt, eine sichere Partition auf einem geteilten HPC-System zur Verfügung zu stellen. Diese Umgebung muss nun insbesondere aus einem Datenmanagementsystem heraus zugreifbar gemacht werden. Grundsätzlich werden Daten stets verschlüsselt gespeichert, übertragen und verarbeitet. Zu berücksichtigen sind eine Reihe von Anforderungen:

Wahrung der Datensouveränität: Jeder Datenanbieter:in soll im Besitz der Schlüssel bleiben, die zum Entschlüsseln der Daten benötigt werden. Daher ist eine dezentrale Schlüsselverwaltung notwendig, bei der sich einzelne Klienten unterbrechungsfrei in die übergreifende Architektur hinzu- und herauschalten lassen.

Datenmanagement: Es muss ein Konzept für ein Datenmanagementsystem entwickelt und implementiert werden, das clientseitig verschlüsselte Daten annimmt und für die Verarbeitung in der SecureHPC-Umgebung bereitstellt.



Skalierbarkeit: Alle erarbeiteten Konzepte und Lösungen müssen sich beliebig skalieren lassen. Dazu muss ein automatisiertes Ausrollen der Komponenten möglich sein.

Training für Anwender:innen: In regelmäßig stattfindenden Tutorials werden Anwender:innen im Umgang mit der sicheren Verarbeitungsumgebung, dem sicheren Datenmanagement und des reproduzierbaren Arbeitens geschult.

Lokal entwickelte Komponenten: Durch die aktive Entwicklung der benötigten Komponenten innerhalb des Konsortiums, sowie des Fokus auf FOSS kann ein Vendor Lock-in verhindert und die Komponenten in Zukunft zuverlässig angeboten werden.

Data-Lake

Datengesteuerte Algorithmen erfordern nicht nur einen möglichst großen Datensatz, sondern brauchen diese Daten auch in einer integrierten Form. Zusätzlich werden so eine hohe Wiederverwendbarkeit der Daten, sowie ein großes Maß an Flexibilität in den zu nutzenden Datenquellen erzielt.

Zentrales Datenrepositorium: Es wird ein flach-hierarchisches Datenrepositorium mit verschlüsselten Daten aufgebaut, auf das mittels domänenspezifischer Attribute zugegriffen werden kann.

Kollaborationen: Die Kombination aus zentralem Datenmanagement und dezentralem Schlüsselmanagement ermöglicht das sichere Zusammenarbeiten verschiedener Institutionen.



FAIR-Prinzipien: Die Speicherung der Daten erfolgt gemäß der FAIR-Prinzipien transparent für alle Nutzer:innen.

Suche auf verschlüsselten Metadaten: Die manuell und automatisch extrahierten Metadaten werden verschlüsselt indiziert. Dadurch ist eine effiziente und sichere Suche möglich.

Datenlebenszyklus: Der Lebenszyklus von Daten kann im Voraus, z.B. in einem Datenmanagementplan, definiert und automatisiert durchgesetzt werden.

Provinienz

Um reproduzierbare Wissenschaft zu ermöglichen, werden Methoden und Laufzeitumgebungen entwickelt, die automatisiert die Datenherkunft von maschinellem Lernen erfassen.

Integration: Die automatisiert erfasste Datenherkunft wird transparent in den Metadaten des Datalakes integriert.

Interoperabilität: Die verschiedenen Methoden und Laufzeitumgebungen werden auf ihre Interoperabilität geprüft, was insbesondere vor dem Hintergrund der forcierten Datenlebenszyklen notwendig ist.



Leistungsanalyse

Um später einen effizienten Betrieb und Skalierbarkeit zu gewährleisten, werden alle Komponenten isoliert und im Zusammenspiel ausgiebig bezüglich ihrer Leistungsfähigkeit getestet. Dazu müssen standardisierte Benchmarks entwickelt werden, die in einem modularen Design zusammengeführt werden können.

Definition einer Standardlast: Um einen Benchmark zu entwickeln, bedarf es zuerst der Definition einer Standardlast. Das heißt, es müssen einzelne Funktionen definiert werden, die den realen Anforderungen möglichst gut entsprechen.

Modularer Ansatz: Um sowohl einzelne Komponenten zu benchmarken, als auch im späteren Betrieb in der Lage zu sein, dynamisch auf neue Umstände zu reagieren, wird ein modularer Ansatz verfolgt.

