

Hessische Staatskanzlei
Hessische Ministerin für
Digitale Strategie und Entwicklung

HESSEN



Bestandsaufnahme im Auftrag der Hessischen Staatskanzlei
Ministerin für Digitale Strategie und Entwicklung

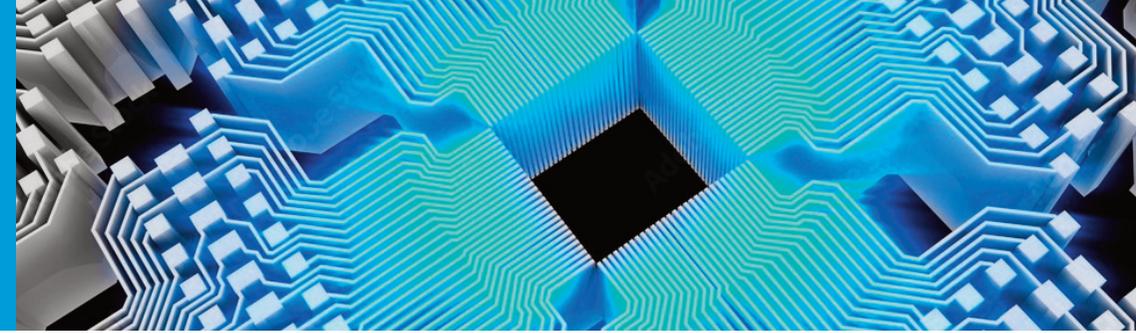
Ressourcen für Künstliche Intelligenz in Hessen – vorhandene Infrastrukturen und Handlungsempfehlungen



digitales.hessen

Prof. Dr. Michael M. Resch
Dr.-Ing. Bastian Koller
Dipl.-Inf. Oleksandr Shcherbakov

Executive Summary



Künstliche Intelligenz (KI) entwickelt sich von einer Spezialität der informationstechnischen Forschung zu einem Standardwerkzeug der Forschung und Entwicklung sowohl im wissenschaftlichen als auch im wirtschaftlichen Bereich. Ihre Anwendungsmöglichkeiten sind vielfältig. Ihr Bedarf an Rechenzeit, Software und Methodik, und damit an verfügbarer Infrastruktur, ist hoch und derzeit exponentiell steigend. Laut einem **IDC-Bericht¹** wird der Umsatz von KI-Systemen bis zum Jahr 2024 von heute etwa 50 Mrd. Euro auf dann rund 110 Mrd. Euro steigen.

Mit der vorliegenden Publikation wird der aktuelle und der absehbare zukünftige infrastrukturelle Bedarf des Landes Hessen auf dem Gebiet der KI im wissenschaftlichen und wirtschaftlichen Bereich untersucht. Unter Rückgriff auf die verfügbaren Informationen aus anderen Studien sowie auf neue Recherchen werden zunächst der Bedarf und die verfügbaren Infrastrukturen für KI in Hessen ermittelt und ein kritischer Abgleich vollzogen. Somit lassen sich Lücken in der infrastrukturellen Versorgung aufzeigen.

Die Bestandsaufnahme stützt sich auf nationale und internationale Modelle zur Infrastrukturversorgung für Wissenschaft und Wirtschaft. Ausgehend von den identifizierten kurz- und mittelfristigen Versorgungslücken in Hessen einerseits und den existierenden Optionen, diese zu schließen, andererseits werden Vorschläge für die Gestaltung der Infrastruktur im Bereich der Künstlichen Intelligenz für Wissenschaft und Wirtschaft in Hessen unterbreitet.

Als Ergebnis resultiert zunächst die Empfehlung, die verfügbaren Potenziale in Hessen stärker zu nutzen und zu koordinieren.

- Hierfür sollten vorhandene KI-Recheninfrastrukturen als Basis genutzt und ausgebaut werden.
- Eine enge Abstimmung mit dem Hessischen Zentrum für Künstliche Intelligenz (hessian.AI) würde die Verzahnung von Infrastruktur und Methodenentwicklung stärken.
- Gezielte Förderprogramme werden empfohlen, um die Wirtschaft und hier insbesondere kleine und mittlere Unternehmen an die KI und die Infrastrukturen heranzuführen. Unternehmen sollen stärker in Förderprogramme zur KI integriert werden.
- Weiterhin wäre es zielführend, Kooperationsmöglichkeiten in der gemeinsamen Beschaffung, dem Betrieb sowie der Nutzung von KI-Recheninfrastrukturen auszuloten.
- Darüber hinaus sollte eine enge Abstimmung mit bundesweiten und europäischen Konzepten und hier insbesondere mit GAIA-X angestrebt werden.

Inhalt

Executive Summary	2
1 Hintergrund und Methodik	4
2 Aktueller Stand	6
3 Bedarfsanalyse und verfügbare Technologien	18
4 Welche Konzepte sind noch zu betrachten?	23
5 Beispielmodelle der Interaktion von Wissenschaft und Wirtschaft	32
6 Potenziale und Empfehlungen	36
Anhang 1: Außeruniversitäre Forschungseinrichtungen mit Standorten in Hessen	40
Anhang 2: Fachterminologie	43
Anhang 3: Potenzielle Technologien	44
Verweise mit Web-Links zu weiteren Informationen	48
Impressum	54

1

Hintergrund und Methodik



1.1 Einführung

Der Begriff der Künstlichen Intelligenz (KI) wird für die Zwecke dieser Bestandsaufnahme breit angelegt, sodass alle Konzepte, die im weitesten Sinn auf die automatische und eigenständige Problemlösung durch digitale Systeme ausgerichtet sind, Berücksichtigung finden.

Betrachtet werden KI-Recheninfrastrukturen – also alle Aspekte der notwendigen Hard- und Software für die Erforschung, Entwicklung und Nutzung von Künstlicher Intelligenz. Bei der Realisierung von tragfähigen Lösungen im Infrastrukturbereich sind darüber hinaus Aspekte der Softwareversorgung mitzudenken.

Die vorliegende Bestandsaufnahme richtet den Fokus im Wissenschaftsbereich auf Forschung und Entwicklung. KI kann auch in der Lehre und in der Verwaltung eine große Rolle spielen. Diese beiden Aspekte werden aber hier nicht berücksichtigt, da mit ihnen weitere Herausforderungen verbunden sind, die im Bereich der Forschung weniger relevant sind, wie der Datenschutz und das persönliche Verhältnis zwischen Lehrperson und Studierenden.

KI entwickelt sich immer mehr von einem vorwiegend wissenschaftlich angewandten neuen Technologietrend hin zu einer gereiften Technologie,

die vielfältig in realen Anwendungsfällen eingesetzt wird. Dieses steigende Interesse konnte auch nicht durch die COVID-19-Pandemie gebremst werden. So hat [Gartner](#)² im September 2020 eine Analyse vorgestellt, nach der 47 Prozent der Firmen in unveränderter Höhe in KI investieren und 30 Prozent der Firmen sogar eine Erhöhung planen.

Gerade durch die Einschränkungen der Pandemie werden weltweit Geschäftsmodelle angepasst, und KI-basierte Werkzeuge wie Chatbots unterstützen die Virtualisierung der Geschäftsbereiche.

Vorhergehende Publikationen wie [Hessens Ambitionen für Künstliche Intelligenz \(2018\)](#)³ oder [KI – Künstliche Intelligenz: Wirtschaft und Innovation \(2020\)](#)⁴ haben sich bereits mit dem Technologiespektrum für Künstliche Intelligenz auf der Ebene von Algorithmen, Theoremen und Konzepten beschäftigt. Auch wurden hier bereits verschiedenste Anwendungsfelder aus Medizin, Ingenieurwissenschaften und auch aus der Finanzbranche adressiert und genauer ausgewertet.

Allerdings wird in den meisten Fällen nur auf die Wichtigkeit hingewiesen, die vorhandenen Ressourcen (Software und Hardware = Recheninfrastrukturen) zu betrachten, eine detaillierte Analyse vorhandener und potenziell notwendiger Recheninfrastrukturen blieb bisher aus.

Eine solche Recheninfrastrukturlandschaft bildet die Erfolgswahl, um Hessen von einem führenden Standort für Rechenzentren zu einem idealen Standort für die Anwendung und Entwicklung rund um Künstliche Intelligenz zu entwickeln. Diese darf konzeptuell nicht nur auf die Nutzbarkeit durch eine Anwendergruppe ausgerichtet sein, sondern sollte sowohl Wissenschaft als auch Wirtschaft adressieren, basierend auf dem vollständigen Lebenszyklus von KI-Anwendungen: vom Austesten über das Implementieren neuer Anwendungen bis hin zum produktiven Betrieb. Aktuell sind kleinere Recheninfrastrukturen vorhanden, die teilweise zur Erprobung und für erste Implementierungen genutzt werden können. Größere KI-Recheninfrastrukturen findet man eher intern bei Firmen; diese stehen daher der Allgemeinheit nicht zur Verfügung.

Ziel dieser Publikation ist eine Bestandsaufnahme von KI-Recheninfrastrukturen in Hessen. Deren potenzielle Nutzbarkeit für einen stetig steigenden Bedarf gilt es zu evaluieren und ferner, Lücken in der Abdeckung der Bedarfe aufzuzeigen, Modelle und Potenziale zur Weiterentwicklung zu präsentieren sowie schließlich daraus Empfehlungen abzuleiten.

Dazu wird eine Übersicht vorhandener Infrastrukturen gegeben, mit einem Hauptaugenmerk auf solche, die je nach Modell auch von mehreren Akteuren genutzt werden können.

Ein weiterer Überblick wird geboten hinsichtlich der KI-Aktivitäten (in Bezug auf Recheninfrastrukturen) auf Länder-, Bundes-, europäischer und internationaler Ebene, um damit den Kontext zu KI-Aktivitäten in Hessen zu markieren und eventuelle Modelle und Potenziale ableiten zu können.

Mit einer Bedarfsanalyse, der Klärung der Abdeckung durch aktuell vorhandene Infrastrukturen und Technologien, einem Einblick in den aktuellen Stand der Technik bei Recheninfrastrukturen und einem Ausblick auf Technologietrends in den nächsten Jahren wird die Basis gelegt für Handlungsempfehlungen zu Maßnahmen im Bereich Recheninfrastrukturen, welche dann abschließend in dieser Bestandsaufnahme präsentiert werden.

1.2 Methodologie

Neben einer allgemeinen Recherche in verfügbaren Informationssets (z. B. thematische Webseiten) erfolgte eine Sekundäranalyse mittels Literaturrecherche in bereits vorliegenden Studien zum Thema KI in Hessen, z. B. „Hessens Ambitionen für Künstliche Intelligenz“ (2018) oder „KI – Künstliche Intelligenz: Wirtschaft und Innovation“ (2020). Hier lag das Augenmerk darauf, welche Recheninfrastrukturen bereits benannt werden bzw. welche Planungen dafür bekannt sind. Dazu wurden auch die Anwendungsfelder analysiert, um eine Grundlage für die Identifikation der aktuell und zukünftig notwendigen Recheninfrastrukturen für KI in Hessen zu schaffen. Innerhalb dieser Bestandsaufnahme wird an entsprechenden Stellen auf die bereits durchgeführten Studien hingewiesen und um den Infrastrukturaspekt ergänzt. Daneben flossen auch ergänzende Gespräche mit den Stakeholdern wie dem Hessischen Zentrum für Künstliche Intelligenz (hessian.AI) oder dem Mittelstand-Digital Zentrum Darmstadt in die Erörterungen ein, auch ergänzende Informationen wurden erarbeitet und in die Bestandsaufnahme integriert.

2

Aktueller Stand



2.1 Vorhandene KI-Infrastrukturen in Hessen

Viele KI-Recheninfrastrukturen wurden im Rahmen existierender Angebote zu Recheninfrastrukturen als zusätzliche Technologien etabliert und sind daher noch wenig auf spezifische Anwendungsfelder abgestimmt. Dies gilt vor allem für den wissenschaftlichen Sektor, der meist an Hochschulen und Forschungseinrichtungen kleine KI-Recheninfrastrukturen installiert hat, die hauptsächlich zur Evaluierung der Anwendbarkeit dieser Technologien beiträgt. Mit diesen kleinen Recheninfrastrukturen werden dann, meist im Rahmen von Projektarbeiten, Ansätze entwickelt und verbessert. Darüber hinaus stellen Unternehmen aus der Wirtschaft (z. B. die Open Telekom Cloud, aber auch Google und ONCITE) KI-Rechenressourcen zur Verfügung, meist mit einem Infrastructure-as-a-Service-Ansatz.

2.1.1 Hochschulen

Infrastrukturen zur Unterstützung der KI-Forschung und -Entwicklung sind zunächst an den Hochschulen zu suchen. Hier gilt es insbesondere, die Ausstattung im Bereich des Hochleistungsrechnens (HPC) zu betrachten, da diese Systeme meist modular sind und durch entsprechende KI-Hardware sehr leicht aufgerüstet werden können.

Im Folgenden wird auf diejenigen großen verfügbaren Infrastrukturre-sourcen an den Hochschulen des Landes Hessen genauer eingegangen, welche auch für KI verwendet werden können.

■ Technische Universität Darmstadt

Das **Hochschulrechenzentrum der Technischen Universität Darmstadt (HRZ)**⁵ hat sich vor kurzem erfolgreich um eine Förderung im Rahmen der **Initiative Nationalen Hochleistungsrechnens (NHR)**⁶ beworben. Der aktuell installierte und verfügbare Rechner (Lichtenberg II) ist derzeit auf Platz 100 der **Weltrangliste Top500**⁷ (Stand November 2020), welche die schnellsten Rechenressourcen weltweit ausgibt. Diese Infrastruktur besitzt derzeit wenig dedizierte Hardware für KI. Momentan sind dort vier Rechenknoten mit insgesamt 16 NVIDIA V100 GPUs und vier Knoten mit insgesamt 16 NVIDIA A100 GPUs geplant, welche sukzessive verfügbar gemacht werden. Außerdem ist noch die Erweiterung um drei NVIDIA-DGX-A100-Systeme mit jeweils acht NVIDIA A100 GPUs in Planung.

■ Goethe-Universität Frankfurt am Main

An der Goethe-Universität Frankfurt am Main ist das Center for Scientific Computing (CSC) angesiedelt. Das Zentrum betreibt ein mittelgroßes Clustersystem mit aktuell ca. 600 Knoten. Der Cluster wurde 2019 installiert. Er ist zwar nicht in der offiziellen Liste der weltweit schnellsten Rechen-systeme (TOP500) gelistet, stellt aber mit einer Spitzenleistung von rund 1,5 PFLOPS (Petaflops) eine substantielle Ressource dar. Mit der Berufung von Prof. Thomas Lippert auf die Professur für Modular Supercomputing and Quantum Computing im Jahr 2020 bekommt das CSC einen neuen inhaltlichen Schub, der sich in der Zukunft sicher auch positiv auf den Standort Hessen im Bereich der KI auswirken wird.

■ Technische Hochschule Mittelhessen

An der **Technischen Hochschule Mittelhessen betreibt die Bioinformatik eine eigene Infrastruktur**⁸, zu der eine **NVIDIA-DGX-1**⁹-Lösung gehört, welche für KI verwendet werden kann. Allerdings reicht diese eine Installation nicht aus, um größere KI-Anwendungen laufen zu lassen.

2.1.2 Landesweite Einrichtungen

■ Hessisches Kompetenzzentrum für Hochleistungsrechnen (HKHLR)

Im Bereich des Hocheistungsrechnens existiert mit dem **Hessischen Kompetenzzentrum für Hochleistungsrechnen (HKHLR)**¹⁰ eine gesamt-hessische Organisation. Dieses Kompetenzzentrum bündelt die HPC-Res-

ourcen im Land. Es stellt über seine Mitglieder Recheninfrastruktur hauptsächlich für die Wissenschaft bereit. KI-Recheninfrastrukturen sind für das HKHLR zum Zeitpunkt dieser Publikation noch kein sichtbares Thema. Es wurden allerdings schon im Bereich von Schulungen u. a. die Themen KI und Deep Learning aufgegriffen.

■ Hessisches Zentrum für Künstliche Intelligenz (hessian.AI)

Das Hessische Zentrum für Künstliche Intelligenz wurde im Herbst 2020 ins Leben gerufen. Mit einer Gesamtförderung von 38 Mio. Euro stellt es eine bundesweit einzigartige Bündelung von exzellenter Forschung, Anwendungsorientierung und Transfer in Wirtschaft und Gesellschaft dar. Das Zentrum wird von 13 hessischen Hochschulen getragen und wird 20 neue Professuren im Bereich der Künstlichen Intelligenz schaffen.

Die Mittel, die dem Hessischen Zentrum für Künstliche Intelligenz im Hardwarebereich zur Verfügung stehen, sollen perspektivisch in etwa 10 DGX-Systeme investiert werden, sodass eine Grundausstattung an Hardware gegeben ist. Für größere Anwendungsszenarien soll aber auf externe Ressourcen zugegriffen werden. Um diese Bedarfe abdecken zu können, ist das Zentrum hinsichtlich der Nutzung größerer Ressourcen im Gespräch mit dem Hochschulrechenzentrum der TU Darmstadt.

■ KompAKI

Das **Kompetenzzentrum für Arbeit und Künstliche Intelligenz im Rhein-Main-Gebiet (KompAKI)**¹¹ ist ein im Oktober 2020 gestartetes Verbundprojekt, das durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert wird.

Innerhalb dieses Verbundprojektes (geleitet von der TU Darmstadt und mit verschiedenen Partnern, unter anderem acht Unternehmen) werden „neue Potenziale für menschenzentrierte KI-Anwendung und deren Geschäftsmodelle erschlossen, **neue Ansätze der kooperativen KI entwickelt, mit denen Anwender die KI transparenter und einfacher nutzen können, sowie neue Methoden zur Bewertung der Arbeit in KI-gestützten Arbeitssystemen entwickelt**“¹². Das KompAKI nutzt aktuell einzelne Server mit GPUs oder DGX-Boxen in den Fachbereichen der TU Darmstadt bzw. DGX-Anteile im Lichtenbergsystem. In naher Zukunft sind aber auch Investitionen in eigene Infrastrukturen geplant.

2.1.3 Forschungseinrichtungen

Das Land Hessen verfügt über eine hohe Anzahl an **außeruniversitären Forschungseinrichtungen**¹³, wobei hier insbesondere die Institute der Fraunhofer-Gesellschaft und der Max-Planck-Gesellschaft zu nennen sind. Das Hessische Ministerium für Wissenschaft und Kunst listet 41 solcher Einrichtungen auf. Im Folgenden wird nur die Verfügbarkeit von Infrastrukturen an den Forschungseinrichtungen betrachtet, soweit diese signifikant groß genug ist, um als Ressource auch extern genutzt zu werden. Die einzelnen Einrichtungen verfügen je nach Größe jeweils über eigene IT-Ausstattung.

Eigene Hardware in sichtbarem, signifikantem Umfang existiert nur beim **Deutschen Wetterdienst (DWD)**.¹⁴ Die dort verfügbare Hardware basiert auf NEC-Aurora-Technologie. Diese ist prinzipiell für den Einsatz in der Künstlichen Intelligenz geeignet. Allerdings sind die Systeme des DWD durch die täglichen Aufgaben und die Verpflichtung, regelmäßige Wetterprognosen zu liefern, in hohem Maße bereits durch den DWD ausgelastet, sodass hier keine Nutzungspotenziale in größerem Umfang zu erwarten sind.

Bei der **Gesellschaft für Schwerionenforschung (GSI)**¹⁵ stehen ebenfalls Rechnerressourcen zur Verfügung. Das dortige Clustersystem (Kronos) wurde Ende 2020 durch ein neues Clustersystem (virgo) ersetzt. Das System basiert auf Standard-AMD-Prozessoren, welche für den Einsatz in der KI wenig geeignet sind. Darüber hinaus wird das System zur Auswertung der im **Projekt FAIR**¹⁶ anfallenden Daten eingesetzt, sodass eine Nutzung über die GSI hinaus schwierig wäre.

An allen anderen Einrichtungen existieren Hardwareressourcen nur in geringerem Umfang, sodass diese Einrichtungen weniger als Bereitsteller von Infrastruktur und mehr als Nutzer von zukünftigen größeren öffentlich verfügbaren KI-Recheninfrastrukturen infrage kommen.

2.1.4 Unternehmen mit KI-Infrastrukturen

■ Samson AG

Die Samson AG mit Hauptsitz in Frankfurt am Main agiert hauptsächlich im Bereich der Stellventiltechnik. Dafür werden die Produkte inzwischen unter Nutzung der neuesten Technologien aus dem Industrie-4.0-Bereich, inklusive Künstlicher Intelligenz, kontinuierlich weiterentwickelt. Die für KI genutzte Infrastruktur kommt unter anderem von dem Kooperationspartner InfraServ Wiesbaden und dessen Teilunternehmen **GES Systemhaus**¹⁷.

■ Continental

Die Firma **Continental hat im Sommer 2020 am Standort Frankfurt am Main einen eigenen Supercomputer in Betrieb genommen**¹⁸, welcher aus über 50 NVIDIA-DGX-Einheiten und damit mehr als 400 NVIDIA A100 GPUs besteht. Diese Infrastruktur ist für Künstliche Intelligenz optimiert und soll die technischen Entwicklungen unterstützen und beschleunigen, auch im Bereich des autonomen Fahrens.

■ ADAMOS

ADAMOS¹⁹ (ADaptive Manufacturing Open Solutions) ist ein von DMG Mori, Dürr, Zeiss und der Software AG gegründetes **Joint Venture, das im Bereich Maschinenbau vor allem Mittelständlern Zugriff auf digitale Infrastrukturen und Services für die Realisierung von Industrie 4.0 und des Industriellen Internets der Dinge (IIoT) dienen soll**²⁰. Gerade die Plattform im Bereich IIoT bietet eine gute Infrastruktur, die auch für Künstliche Intelligenz verwendet werden kann.

■ Innoplexus

Die Innoplexus AG hat sich auf den medizinischen Bereich fokussiert und bietet neben diversen Services rund um die Analyse und Behandlung von (medizinischen) Datensätzen auch Data-as-a-Service an, um einen kontinuierlich wachsenden Daten-Pool zu realisieren. Der Hauptsitz der Firma ist in Eschborn.

2.1.5 Infrastructure-as-a-Service-Anbieter mit Sitz in Hessen

Das **Kompetenznetzwerk Trusted Cloud e.V.**²¹ hat in Zusammenarbeit mit dem Bundesministerium für Wirtschaft und Energie (BMWi) eine Online-Plattform etabliert, welche die Vernetzung von interessierten Anwendern, vornehmlich aus der Wirtschaft, mit Anbietern von Cloud Services (u. a. Infrastrukturen) ermöglicht. Die Plattform prüft die Angebote auf Transparenz, Sicherheit, Qualität und Rechtskonformität. Die Trusted Cloud ist aktiv an der GAIA-X-Aktivität beteiligt.

Zahlreiche Anbieter von Recheninfrastrukturen als Infrastructure-as-a-Service (IaaS) sind bereits auf dieser Plattform registriert, mit Firmensitz in Hessen beispielsweise die folgenden:

■ Aixit Cloud

Die Firma **aixit GmbH**²² bietet unter anderem IaaS im Rechenzentrum mit Standort Frankfurt am Main an.

■ Servinga GmbH – Managed Cloud

Die Firma **servinga GmbH**²³ bietet IaaS mittels einer Managed Cloud an. Die Angebote werden durch ein Rechenzentrum mit Standort Frankfurt am Main angeboten.

■ Open Telekom Cloud

Die **Open Telekom Cloud**²⁴ ist ein Angebot von T-Systems für IaaS auf Basis der Open Source Software Open Stack. Während die Open Telekom Cloud einen Standort in Hessen hat, sind die Rechenzentren im Land Sachsen-Anhalt (Magdeburg und Biere) lokalisiert.

■ Google Region Frankfurt

Googles Cloud-Plattform ist an diversen Standorten lokalisiert. Ein Standort ist die Region Frankfurt am Main²⁵ (europe-west3), welche auch Big Data und Service im Bereich Maschinelles Lernen anbietet. Bislang sind die verfügbaren Recheninfrastrukturen durch Google nur angemietet, es gibt aber aktuell **weitere Pläne**²⁶ für Rechenzentren in Dietzenbach und Erlensee sowie für ein weiteres Rechenzentrum in Babenhausen.

■ ONCITE

ONCITE²⁷ ist ein Zusammenschluss von Unternehmen der Friedhelm Loh Group (German Edge Cloud, Rittal und der **IoTOS GmbH**²⁸) unter anderem mit **Bosch Connected Industry**²⁹ und der Fraunhofer-Gesellschaft. Der Schwerpunkt liegt auf dem Angebot einer KI-Edge-Cloud-Lösung für den Mittelstand. Dazu wird das Konzept des Edge Computing implementiert, welches die Datenbearbeitung vor Ort bei den Firmen in nahezu Echtzeit erlaubt und somit auch die Erfordernisse von Datenschutz und Datensouveränität bedient. Der Standort von ONCITE ist in Eschborn.

2.2 KI-Einsatz in Wirtschaft und Wissenschaft in Hessen

Die Einsatzmöglichkeiten von Künstlicher Intelligenz in Wirtschaft und Wissenschaft steigen stetig, immer neue Gebiete werden adressiert und evaluiert. In Hessen werden KI-Methoden und -Techniken momentan in zahlreichen Gebieten eingesetzt. So hat die Studie von Buxmann et al. (2020) bereits folgende potenzielle Anwendungsgebiete identifiziert: Industrie, Finanzwirtschaft und Fintechs, Kultur- und Kreativwirtschaft, Energiesektor, Mobilität, Logistik und Handel, Medizin und Gesundheit, Wohnen und Leben sowie Cybersicherheit.

Im Rahmen von **Industrie 4.0 nutzen hessische Firmen unter anderem im Verbund mit Wissenschaftlerinnen und Wissenschaftlern KI, um ihre Prozesse zu optimieren**³⁰ (z. B. durch KI-unterstützte Bilderkennung zur automatischen Qualitätssicherung im Produktionsablauf).

In der Finanzwirtschaft und im Fintechbereich sieht man die Aktivitäten in Hessen mit der Arbeit an dem GAIA-X-Anwendungsszenario Financial Big Data Cluster (siehe 2.3) kontinuierlich weiterentwickelt.

In der Kultur- und Kreativwirtschaft wird Kunst und Kultur mithilfe von KI kontinuierlich entwickelt und findet zunehmend Akzeptanz bei den Firmen und Schaffenden, die sich damit auseinandersetzen.

Vorhersagen über Störanfälligkeiten der Energienetze sind ein Bereich, in dem KI schon heute im Energiesektor in Hessen eingesetzt wird. Auch Energieberaterinnen und -berater sowie Energiemanagerinnen und -manager nutzen zunehmend KI-gestützte Analyseverfahren zur Optimierung ihrer Arbeit.

Im Bereich Mobilität, Logistik und Handel gewinnt die Nutzung von Künstlicher Intelligenz im Zusammenhang mit dem autonomen Fahren immer mehr an Bedeutung und wird von einigen Firmen und Wissenschaftlerinnen und Wissenschaftlern bereits eingesetzt.

Medizin und Gesundheit sind Anwendungsfelder, die schon Künstliche Intelligenz avisieren und teilweise einsetzen, vor allem im Bereich der Bildanalyse von medizinischen Aufnahmen. Aber auch im Bereich der Erforschung von Medikamenten (z. B. bei SARS-CoV-2) wird KI bereits eingesetzt, oft im Verbund von Wissenschaft und Wirtschaft.

Im Bereich Wohnen und Leben finden sich ebenfalls bereits erste Einsatzgebiete von KI durch Wissenschaft und Wirtschaft. Hier stehen bislang Smart Homes und Internet of Things im Mittelpunkt, es ist aber durchaus weiteres Potenzial gegeben.

Die Nutzung von KI im Anwendungsfeld Cybersicherheit konzentriert sich im Moment auf die Bereitstellung von intelligenten Algorithmen und KI-Methoden, um Angriffe schneller zu erkennen und passende Maßnahmen zu ergreifen. Darmstadt ist bereits einer der führenden Standorte im Bereich Cybersecurity und u. a. durch das Forschungszentrum **ATHENE**³¹ massiv in diesem Bereich aktiv.

Der **Branchenverband Bitkom hat im Oktober 2020 Ergebnisse einer Studie als Pressemitteilung**³² veröffentlicht, die unter anderem die Nutzung von Künstlicher Intelligenz in Unternehmen (bundesweit) evaluiert hat. Die Ergebnisse der Studie zeigen, dass bislang nur ein kleiner Prozentsatz der Unternehmen KI in verschiedenen Bereichen, von personalisierter Werbung bis hin zur Erstellung automatisierter Prognosen, nutzt (maximal 4%). Gravierend erscheint auch, dass KI von kaum einem Unternehmen zur Generierung neuer Produkte herangezogen wird. Woran liegt das?

Die Gründe liegen unter anderem in der immer noch existierenden Hürde, KI-gestützte Prozesse und Produkte zu realisieren, gerade bei kleinen und mittelständischen Unternehmen. Für diese Unternehmen stellt die Investition in KI-Expertise und eventuell auch dedizierte KI-Infrastruktur ein großes Hemmnis dar. Dies wird auch in einer **Studie von Keller und Lorenz**³³ deutlich, die 2019/20 15 Unternehmen aus Mittelhessen und dem Siegerland befragt haben. Dabei wurde deutlich, dass durchaus ein Verständnis der Potenziale von KI in den Unternehmen gegeben ist. Gleichzeitig weisen aber der komplexe Zugang zu den Technologien und die zu erwartenden Kosten für Inhouse-Experten dem „Onboarding“ auf Künstliche Intelligenz eher einen längeren Zeithorizont zu. Hier zeigt sich klar die Notwendigkeit, interessierte Unternehmen auf ihrem Weg zur Nutzung von KI durch Programme, Services und Infrastrukturen zu unterstützen.

2.3 KI-Infrastrukturen und Programme in anderen Bundesländern und auf Bundesebene

Im Folgenden werden bundesweite Ansätze und die Konzepte anderer Bundesländer, die im Bereich der KI angesiedelt sind, betrachtet. Der Vergleich mit diesen Aktivitäten bzw. der Bezug auf diese Infrastrukturressourcen erlaubt es zum einen, die hessische KI-Landschaft einzuschätzen, und zum anderen, die Bedarfe der hessischen KI-Landschaft im nationalen Kontext zu sehen.

2.3.1 Bundesebene

■ Das Gauss Centre for Supercomputing – GCS

Das **Gauss Centre for Supercomputing (GCS)**³⁴ vereinigt die drei Bundeshöchstleistungsrechenzentren: das **Jülich Supercomputing Centre (JSC)**³⁵, das **Leibniz-Rechenzentrum (LRZ)**³⁶ in Garching bei München und das **Hochleistungsrechenzentrum Stuttgart (HLRS)**³⁷. Im Rahmen des GCS stimmen die drei Zentren sich regelmäßig ab, um dem Bedarf der Nutzergemeinschaften von Wissenschaft wie auch Industrie gerecht zu werden. Dabei nimmt gerade das HLRS eine Vorreiterrolle bei der industriellen Nutzung von Höchstleistungsrechnen ein, da es seit mehr als 20 Jahren sein Ökosystem auch auf diese Zielgruppe abgestimmt hat.

So, wie sich die Anwendung von KI-Technologien in Zusammenhang mit Höchstleistungsrechnen und Datenanalyse ausweitet, erweitern die GCS-Zentren stetig ihre Services und Infrastrukturen um Fähigkeiten im Bereich Künstlicher Intelligenz.

Am JSC steht zum Zeitpunkt der Bestandsaufnahme ein Booster-Modul des Supercomputers JUWELS zur Verfügung. Dieser beinhaltet u. a. 3.744 NVIDIA A100 GPUs, die für KI nutzbar sind. Das LRZ führt zurzeit 64 NVIDIA Tesla V100 GPUs in seinem Systemportfolio auf. In Stuttgart ist am HLRS ein dediziertes KI-System (CS-Storm) mit 64 NVIDIA Tesla V100 GPUs vorhanden und im Quartal 1 oder Quartal 2/2021 wird dort eine Erweiterung des Supercomputers HAWK um 196 NVIDIA A100 GPUs installiert.

■ Verbund Nationales Hochleistungsrechnen – NHR

Die Gemeinsame Wissenschaftskonferenz (GWK) hat sich im November 2018 – um der zunehmenden Bedeutung und der steigenden Nachfrage nach Hochleistungsrechnern gerecht zu werden – auf die gemeinsame Förderung eines koordinierten Verbundes des **Nationalen Hochleistungsrechnens (NHR)**³⁸ durch Bund und Länder verständigt. Die Aufnahme von Rechenzentren in die Förderung erfolgte in einem wettbewerblichen und wissenschaftsgeleiteten Auswahlverfahren.

Seit Januar 2021 sind acht Hochleistungsrechenzentren in die gemeinsame Förderung von Bund und Ländern aufgenommen. Sie sollen den zukünftigen NHR-Verbund aufbauen und künftig deutschlandweit und bedarfsgerecht Rechenkapazität für die Forschung zur Verfügung stellen. Bis zu 62,5 Mio. Euro jährlich stehen über einen Zeitraum von zehn Jahren jeweils für Investitions- und Betriebskosten bereit.

Der NHR stellt seine Ressourcen deutschlandweit zur Verfügung. Da die Förderung nach Artikel 91b Grundgesetz erfolgt, ist die Nutzung der Ressourcen für wirtschaftliche Zwecke nicht zulässig. Forschung im Bereich der Künstlichen Intelligenz kann jedoch auf die Ressourcen zugreifen. Im Land Hessen hat sich die TU Darmstadt mit ihrem Hochschulrechenzentrum erfolgreich um die Teilnahme am NHR beworben. Darmstadt wird damit ein hessischer Knoten im deutschen Gesamtverbund des Hochleistungsrechnens.

Aufgrund des steigenden Bedarfs an Rechenleistung für den Bereich der Künstlichen Intelligenz ist zu erwarten, dass der NHR bei der Beschaffung auch Hardware für die KI mit einbeziehen wird. Da die Zentren des NHR für die bundesweite Nutzung geöffnet sind, stehen diese Ressourcen dann allen Wissenschaftlerinnen und Wissenschaftlern in der Bundesrepublik zur Verfügung und indirekt profitiert auch wiederum die Wirtschaft von den Erfahrungen der Nutzerinnen und Nutzer.

■ Die Gauß-Allianz

Die **Gauß-Allianz**³⁹ ist ein gemeinnütziger Verein zur Förderung von Wissenschaft und Forschung. Zu diesem Zweck unterstützt sie die wissenschaftliche Gemeinschaft in Deutschland, indem sie die Voraussetzungen für eine nachhaltige und effiziente Nutzung von Supercomputingressourcen der obersten Leistungsklassen schafft. Insbesondere bündelt und koordiniert sie die einander ergänzenden Kompetenzen und diversifizierten Rechnerarchitekturen sowie die zugehörige Zugangsstruktur.

Für die Verwirklichung ihrer Ziele setzt die Gauß-Allianz neben der Durchführung wissenschaftlicher Veranstaltungen insbesondere auf die Beratung von Wissenschaft, Wirtschaft, Bund, Ländern und Trägerorganisationen. Damit trägt sie zur optimalen Erschließung und Weiterentwicklung des Potenzials des wissenschaftlichen Rechnens im Allgemeinen und insbesondere des Hoch- und Höchstleistungsrechnens bei. Die Gauß-Allianz vermittelt Firmen Kontakte zu ihren Mitgliedszentren, die bereit sind, ihre Ressourcen der Wirtschaft gegen Kostenersatz zur Verfügung zu stellen.

■ GAIA-X

GAIA-X⁴⁰ ist eine von Bundeswirtschaftsminister Peter Altmaier initiierte Aktivität zum Aufbau einer sicheren und vertrauenswürdigen Cloud- und Dateninfrastruktur. Ursprünglich für Deutschland geplant, hat sich GAIA-X inzwischen zu einer europäischen Aktivität entwickelt. Hierbei geht es unter anderem auch um den Aufbau einer sicheren Infrastruktur für Künstliche Intelligenz, die von verschiedenen Anbietern europaweit betrieben wird. GAIA-X adressiert somit einige der gängigen Hürden, die kleine und mittelständische Unternehmen von einer Integration von Künstlicher Intelligenz in ihr Ökosystem abhalten. Durch den Zugang zu GAIA-X bieten sich gerade **kleinen und mittelständischen Unternehmen erweiterte Möglichkeiten, schnell auf Daten und Dienstleistungen zuzugreifen**⁴¹, und all dies im sicheren Datenraum.

So ergibt sich im Rahmen des GAIA-X-Betriebsmodells ein gutes Potenzial, um eine regionale Infrastruktur mit einzubinden. Beispielhaft für einen zentralen Anwendungsfall in GAIA-X steht das **Financial Big Data Cluster (FBDC)**⁴² in Hessen. Hier wird ein Finanzdaten-Cluster für das Finanzwesen aufgebaut, auf dem bereits jetzt im Rahmen des Projektes **SafeFBDC**⁴³ diverse KI-Konzepte entwickelt und evaluiert werden. Hieran wird ersichtlich, welche Möglichkeiten sich bieten, in Hessen KI innerhalb einer größeren Aktivität einzubetten.

■ Digital Hubs

Die **Digital-Hub-Initiative**⁴⁴ unter Förderung des Bundesministeriums für Wirtschaft und Energie hat zum Ziel, die digitale Innovation in Deutschland voranzubringen. Dazu wurden zwölf Digital Hubs kreiert, die jeweils einen Zusammenschluss von Unternehmerinnen und Unternehmern, Gründerinnen und Gründern sowie Forscherinnen und Forschern bilden und in den verschiedensten Themengebieten digitale Lösungen entwickeln.

Unter den **zwölf Digital Hubs befindet sich auch ein dedizierter für Künstliche Intelligenz mit Standort in Karlsruhe**⁴⁵. In Hessen befindet sich der Fintech und Cyber Security Hub (Frankfurt am Main), dessen Anwendungsszenarien auch die Nutzung von KI umfassen. Im Laufe des Jahres 2021 ist es geplant, auch **europäische Digital Innovation Hubs**⁴⁶ zu installieren, die dann europaweit an Themen wie KI arbeiten sollen. Hessen beteiligt sich an den European Digital Innovation Hubs (EDIH) mit zwei Anträgen.

2.3.2 Landesebene

Der folgende Abschnitt gibt eine kurze Übersicht über die KI-Programme des jeweiligen Bundeslandes im Rahmen der **nationalen KI-Strategie**⁴⁷ und stellt dar, ob es dedizierte Infrastrukturen und Infrastrukturprogramme gibt. Die Auswahl orientierte sich an der **Plattform Lernende Systeme**⁴⁸, einer Plattform für Künstliche Intelligenz.

Die Übersicht trägt im Rahmen dieser Bestandsaufnahme nur die wesentlichen Punkte des jeweiligen Landesprogramms zusammen, Richtschnur hierfür sind die Wichtigkeit von KI innerhalb der einzelnen Strategien sowie explizite Investmentvorhaben in KI-Recheninfrastrukturen.

■ Baden-Württemberg

Künstliche Intelligenz ist ein wichtiger Teil der **Innovationsstrategie des Landes Baden-Württemberg**⁴⁹. Diverse Handlungsfelder und Maßnahmen werden hier gebündelt, um die Positionierung des Landes bezüglich KI zu festigen. Damit soll die Erforschung und die Wertschöpfung aus der Technologie sichergestellt werden, beispielsweise durch den Ausbau des **Cyber Valley**⁵⁰ zum Exzellenzzentrum oder die Umsetzung von Programmen zur Unterstützung von KMU beim Einstieg in KI-Technologien.

Recheninfrastrukturen: Innerhalb der **Landesstrategieaktivitäten bwHPC**⁵¹ ist der Ausbau der verfügbaren Infrastrukturen für KI vorgesehen. Die vorhandene Infrastruktur wird sowohl von Wirtschaft als auch Wissenschaft genutzt.

■ Bayern

Die bayerische Landesstrategie **Hightech Agenda Bayern**⁵² sieht massive Investitionen im Bereich Künstlicher Intelligenz vor. Diese umfassen zum einen die Stärkung des wissenschaftlichen Know-hows (u. a. durch Professuren, Stärkung der Forschung), zum anderen auch die Unterstützung der KI-Nutzung durch die Wirtschaft.

Recheninfrastrukturen: Im Rahmen der Strategie sind Investitionen in Recheninfrastrukturen vorgesehen. Ein Hauptaugenmerk liegt im Ausbau des Leibniz-Rechenzentrums der Akademie der Wissenschaften hin zu einem Kompetenzknoten für Big Data und KI. Die vorhandene Infrastruktur wird sowohl von Wirtschaft als auch Wissenschaft genutzt.

■ Berlin

Das Bundesland Berlin fokussiert sich im Verbund mit Brandenburg in seiner KI-Strategie **innoBB**⁵³ auf die Kompetenzen rund um das Thema Künstliche Intelligenz.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Brandenburg

Brandenburg hat Ende 2018 die Zukunftsstrategie **Digitales Brandenburg**⁵⁴ verabschiedet. Dort wird Künstliche Intelligenz als ein Themenfeld für Forschung und Entwicklung definiert.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Bremen

Bremen hat im Oktober 2020 die Strategie **BREMEN.KI – Strategie Künstliche Intelligenz**⁵⁵ verabschiedet. Dort wird ein Fokus auf die Handlungsfelder in den Bereichen Wissenschaft, Wirtschaft, Gesellschaft und Qualifizierung gelegt.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Hamburg

Im Januar 2020 wurde durch den Hamburger Senat die **Digitalstrategie 2020 für Hamburg**⁵⁶ verabschiedet. Die Strategie konzentriert sich auf Behörden, städtische Einrichtungen, Unternehmen, Wissenschaft und Gesellschaft und soll gerade im Bereich KI Verstärkungen für Hamburgs Angebot an Wissenschaft und Wirtschaft bringen.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Mecklenburg-Vorpommern

In Mecklenburg-Vorpommern wurde 2018 in einer Digitalen Agenda der Bereich Künstliche Intelligenz erwähnt. Eine eigene Agenda bzw. Strategie liegt in diesem Bundesland nicht vor.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Niedersachsen

Niedersachsen führt Künstliche Intelligenz als einen wichtigen Baustein in seinem 2018 verabschiedeten **Masterplan Digitalisierung**⁵⁷ an. Der Fokus hier liegt auf dem Ausbau anwenderorientierter Forschung und auf der Stärkung von Forschung und Lehre. Dazu wird das **Zentrum für Digitale Innovation (ZDIN)**⁵⁸ aufgebaut, das verschiedene Schwerpunkte umfasst (u. a. Agrarwirtschaft, Energie, Gesellschaft).

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Nordrhein-Westfalen

Nordrhein-Westfalen entwickelt seine im Jahr 2019 veröffentlichte **Digitalstrategie**⁵⁹ kontinuierlich weiter. Derzeit wird das Thema Künstliche Intelligenz ausgearbeitet und konkretisiert, u. a. durch die **KI-Kompetenzplattform KI.NRW**⁶⁰ zur stärkeren Vernetzung der diversen Akteure aus Wissenschaft und Wirtschaft.

Recheninfrastrukturen: In der Strategie ist eine weitere Stärkung des Jülich Supercomputing Centre des Forschungszentrums Jülich vorgesehen. Die vorhandene Infrastruktur wird sowohl von Wirtschaft als auch Wissenschaft genutzt.

■ Rheinland-Pfalz

Rheinland-Pfalz hat im September 2020 seine KI-Agenda vorgestellt, die Investitionen von 36 Mio. Euro in diverse Bereiche (Forschung, anwendungsorientierter Transfer, Qualifizierung und Weiterbildung von Fachkräften und internationale Vernetzung) vorsieht.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen.

■ Saarland

Das Saarland hat eine **Innovationsstrategie**⁶¹ entwickelt, welche sich auch mit dem Bereich der Künstlichen Intelligenz beschäftigt. KI wird in den folgenden Aktivitäten adressiert: Forschung im Bereich KI und Schaffung eines KI-Hubs, Förderung und Transfer für Vorhaben im Bereich KI und Demonstratoren zu KI-basierten Systemen.

Recheninfrastrukturen: Innerhalb der Innovationsstrategie wird die Förderung von Forschungsinfrastrukturen erwähnt, aber nicht konkretisiert. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Sachsen

Die Strategie **Sachsen Digital**⁶² wurde 2019 veröffentlicht und hat das Ziel, Sachsen bis 2025 zu einem führenden Standort für KI in Forschung und Innovation zu etablieren. Dies soll unter anderem durch den Ausbau des nationalen Kompetenzzentrums für Big Data **ScaDS (Center for Scalable Data Analytics)**⁶³ geschehen, welches zu einem erweiterten Zentrum namens ScaDS.AI entwickelt werden soll.

Recheninfrastrukturen: Es ist anzunehmen, dass die Stärkung und Erweiterung des ScaDS auch Investitionen in Recheninfrastrukturen für Künstliche Intelligenz beinhaltet. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Sachsen-Anhalt

In der **Digitalen Agenda für das Land Sachsen-Anhalt**⁶⁴ wird das Themengebiet Künstliche Intelligenz nicht direkt aufgeführt.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Schleswig-Holstein

Schleswig-Holstein hat 2019 eine **KI-Strategie**⁶⁵ entwickelt, die die Bereitstellung öffentlicher Daten, Forschung in Form eines Aufbaus des Zentrums für Künstliche Intelligenz Lübeck (ZKIL) und eines KI-Labs sowie Transfer und Vernetzung und Bildung und Kultur thematisiert. Daneben werden auch Förderprojekte zur Entwicklung von KI-Anwendungen avisiert.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

■ Thüringen

Die in Thüringen erarbeitete **Strategie für die Digitale Gesellschaft**⁶⁶ sieht die Unterstützung der Entwicklung und des Einsatzes von KI-Technologien für Wirtschaft, Wissenschaft und Gesellschaft vor.

Recheninfrastrukturen: Größere Investitionen in bzw. der Aufbau von extra Recheninfrastrukturen für KI sind aktuell nicht explizit vorgesehen. Die vorhandene Infrastruktur wird hauptsächlich von der Wissenschaft genutzt.

3

Bedarfsanalyse und verfügbare Technologien



Im Folgenden werden die Bedarfe in Wissenschaft und Wirtschaft abgeschätzt, um einen Überblick über die aktuelle und zukünftige Nutzung der KI-Methoden und -Anwendungen in Hessen zu erhalten.

Dabei wird berücksichtigt, dass die Anforderungen an die Infrastruktur für KI in den Bereichen Wissenschaft und Wirtschaft unterschiedlich sein können. Weil die Deckung des jeweiligen Bedarfs jeweils anderen Prämissen unterliegt, werden die beiden Bereiche getrennt betrachtet.

Der wesentliche Unterschied liegt darin, dass im wissenschaftlichen Bereich sowohl die Entwicklung als auch der Einsatz von Methoden der KI für wissenschaftliche Fragestellungen im Vordergrund stehen, während im wirtschaftlichen Bereich überwiegend der Einsatz der Methoden bedeutsam ist. Hierbei geht es um die qualitative Verbesserung eines angebotenen Services oder Produkts oder die Optimierung von Service- und Produktionsverfahren sowie Verbesserungen in der Produktentwicklung.

3.1 Stakeholder-Analyse: Forschungseinrichtungen

3.1.1 Universitäten und Hochschulen für angewandte Wissenschaften

An den Universitäten und Hochschulen für angewandte Wissenschaften hat sich durch die Einrichtung des Hessischen Zentrums für Künstliche Intelligenz (hessian.AI) eine Initiative gebildet, die die Aktivitäten an den beteiligten Standorten bündelt. Die 22 im hessian.AI gebündelten Gründungsprofessuren kommen überwiegend aus der Informatik und nur zu einem kleinen Teil aus Anwendungsgebieten. Sie decken ein breites Spektrum im Bereich der Methodenentwicklung und einige Bereiche der Anwendung von Methoden der Künstlichen Intelligenz ab. Die Hessische Landesregierung stellt für die fünfjährige Aufbauphase Mittel in Höhe von 38 Mio. Euro zur Verfügung, aus denen 20 weitere KI-Professuren finanziert werden sollen. Diese werden den Bedarf an KI-Recheninfrastruktur in den kommenden Jahren weiter steigen lassen.

Derzeit besteht an den Hochschulen des Landes ein kleiner bis mittlerer Bedarf an Infrastrukturen für Künstliche Intelligenz, der jedoch aufgrund der rasanten Entwicklungen in diesem Bereich schon kurzfristig ansteigen wird. Deshalb ist für die kommenden zwei bis drei Jahre zu erwarten, dass sich die Verfügbarkeit auch größerer KI-Systeme im Infrastrukturbereich als notwendig erweist. Dies betrifft sowohl den Bereich der Methodenentwicklung als auch den Bereich der Anwendungen. Dabei sind für die Methodenentwicklung kleinere bis mittlere Systeme typischerweise ausreichend. Für Testzwecke ist aber der Zugriff auf größere Systeme unerlässlich. Für den Anwendungsbereich reicht das Spektrum von kleinen bis zu sehr großen Systemen, je nach Aufgabenstellung.

Über das hessian.AI hinaus ist mittelfristig mit einem erheblichen Anstieg des Recheninfrastrukturbedarfs – insbesondere im Anwendungsbereich – zu rechnen. Nachdem durch das hessian.AI die Methodenentwicklung massiv vorangetrieben wird, werden diese Methoden entsprechend in der Breite durch die Anwendungswissenschaften aufgegriffen werden. Die Anwendungsmöglichkeiten der KI erstrecken sich dabei über alle Wissenschaftsbereiche. Es sind davon also nicht nur die in der Digitalisierung gut aufgestellten Ingenieur- und Naturwissenschaften betroffen, sondern zunehmend auch die bisher der Digitalisierung noch weniger nahestehenden Geistes- und Sozialwissenschaften sowie die Wirtschaftswissenschaften. Der aus den Anwendungsbereichen kommende Bedarf wird also im Ausmaß und abhängig von der Geschwindigkeit der Methodenentwicklung stark ansteigen. Mittelfristig und langfristig ist daher ein Ausbau der entsprechenden Infrastrukturen unerlässlich.

3.1.2 Außeruniversitäre Forschungseinrichtungen

Unter den außeruniversitären Forschungseinrichtungen wurde anhand deren öffentlicher Forschungsarbeiten eine Analyse des Bedarfs dieser Einrichtungen durchgeführt. Die Ergebnisse sind in Anhang 1 dargestellt.

Zusammenfassend ergibt sich, dass in Bezug auf die Nutzung der Künstlichen Intelligenz bei 22 der Einrichtungen von einem hohen Bedarf (Zugang zu größeren Maschinen) an Infrastruktur für Künstliche Intelligenz, vor allem in der Anwendung der Technologie, auszugehen ist. Insbesondere in den naturwissenschaftlich orientierten Einrichtungen ist dieser Bedarf schon heute deutlich sichtbar, wobei einige der Einrichtungen bereits über Arbeitsgruppen verfügen, die sich mit dem Thema befassen. Ein hoher Bedarf an Infrastruktur zur Unterstützung des Einsatzes von KI in den Forschungsarbeiten ergibt sich aber auch in Einrichtungen, die nicht aus den Naturwissenschaften kommen. Der Bedarf in den genannten 22 Einrichtungen ist kurzfristig vorhanden und wird mittelfristig erheblich ansteigen.

Für weitere acht Einrichtungen ist ein mittlerer Bedarf (Zugang zu lokalen Ressourcen) im Bereich der KI abzusehen. Auch hier ist eine große Bandbreite von Einrichtungen betroffen. Der Anstieg des Bedarfs ist bei diesen Einrichtungen schwerer abzusehen, da die KI zumeist als unterstützende Technologie eingesetzt werden wird. Daraus können sich Servicekonzepte entwickeln, auf die die Einrichtungen zurückgreifen können, ohne selbst einen steigenden Bedarf an Infrastruktur zu haben. An elf Einrichtungen besteht ein geringer Bedarf an Infrastruktur für KI, da in diesen Bereichen KI selten bzw. gar nicht eingesetzt wird.

Hierzu gehört der juristische ebenso wie der historische Bereich. Obwohl auch dort Potenziale für die Nutzung der KI vorhanden sind, sind die jeweiligen Einrichtungen in ihrer Forschung noch nicht auf diesen Einsatz ausgerichtet. Inwieweit sich mittelfristig höhere Bedarfe ergeben, hängt stark von der Forschungsrichtung dieser Einrichtungen in der Zukunft ab.

3.2 Stakeholder-Analyse: Unternehmen

Im **Monitoring-Report Wirtschaft Digital 2017: Hessen**⁶⁷ wurden innovative Wachstumsbereiche für das Land Hessen und seine Unternehmen detailliert untersucht. Als Wachstumsbereiche werden Internet der Dinge, Smart Services, Big Data, Robotik/Sensorik, Künstliche Intelligenz und Industrie 4.0 genannt.

Diese in der Studie vorgenommene Unterscheidung von Themen, die in der Informatik und im Maschinenbau oft synergetisch gedacht werden, basiert auf der Annahme, dass es sich bei der Künstlichen Intelligenz vornehmlich um eine Technologie handelt, die die Absicht hat, menschliches Verhalten perfekt zu kopieren. Im praktischen – und hier verwendeten – Sinn ist aber Künstliche Intelligenz begrifflich weiter zu fassen, sodass in der hier vorgelegten Betrachtung auch eine Reihe von Aspekten von Big Data, Smart Services sowie der Industrie 4.0 mit inbegriffen sind.

Betrachtet man die Ergebnisse der Studie, so fällt zunächst ins Auge, dass hessische Unternehmen in etwa so intensiv digitalisiert sind wie der bundesdeutsche Durchschnitt. Gerade bei den größeren Unternehmen zeigt sich aber, dass diese in Branchen angesiedelt sind (Informationstechnologie, Banken, Medizintechnik), die als Branchen mit sehr hohem Potenzial für die Nutzung von KI eingestuft werden.

Eine genauere Analyse der Studie zeigt folgende Durchdringung der hessischen Wirtschaft in den maßgeblichen Bereichen der KI.

KI-Technologie	Genutzt	Geplant
Internet der Dinge	45 %	10 %
Smart Services	37 %	12 %
Big Data	21 %	11 %
Robotik & Sensorik	9 %	5 %
Kernbereich KI	5 %	6 %
Industrie 4.0	14 %	5 %

Tabelle 1 – Genutzte und geplante KI-Technologien in der hessischen Wirtschaft nach Monitoring-Report Wirtschaft Digital 2017, S. 47

Auffallend sind die hohen Werte beim Internet der Dinge sowie bei den Smart Services. In beiden Bereichen spielen Methoden der Künstlichen Intelligenz eine wesentliche Rolle, sodass davon auszugehen ist, dass unter Verwendung eines weiter gefassten Begriffs von KI – wie er in der aktuellen Situation und in dieser Publikation genutzt wird – ein hoher Anteil von KI-Nutzung in diesen beiden Bereichen zu beobachten ist. Darüber hinaus zeigen beide Bereiche auch quantitative Ausbaupotenziale von 10 bzw. 12 Prozent. Auch der Bereich der Big Data ist einerseits schon relativ stark vertreten (21 %) und zeigt andererseits ein hohes kurzfristiges Ausbaupotenzial (11 % bzw. eine Erhöhung um 50 %), sodass hier in Zukunft hohe Nutzungspotenziale für KI-Methoden zu erwarten sind.

Die Bereiche Robotik und Sensorik sowie Industrie 4.0 sind dagegen noch eher schwach vertreten. Das kann damit zusammenhängen, dass Hessen weniger von klassischen großen Produktionsstätten geprägt ist als etwa die Region Stuttgart. Hier sind also auch in Zukunft nicht die größten Wachstumsraten in Hessen zu erwarten.

Wenig überraschend ist der relativ geringe Anteil von Unternehmen, die auf klassische KI im Sinne der Imitation menschlichen Verhaltens setzen. Auch wenn die Imitation menschlichen Verhaltens eine wichtige und spannende Aufgabe der Forschung ist, so spielt sie doch im Allgemeinen im wirtschaftlichen Bereich eine eher untergeordnete Rolle. Das wesentliche Szenario für den Einsatz menschlicher Imitationen ist letztlich nur ein Bereich, in dem unabdingbar notwendig ein menschenähnliches Artefakt mit Menschen interagieren und kommunizieren muss.

3.2.1 Großunternehmen

Für die größten Unternehmen in Hessen wird eine Übersicht gegeben, die die Firmen einer Branche zuordnet und auf der Basis dieser Zuordnung über den abgeschätzten zukünftigen Bedarf an Infrastrukturen für Künstliche Intelligenz informiert. Der Abschätzung liegen Studien von Wirtschaftsinformatikern aus Hessen (Buxmann et al.) und vom MIT⁶⁸ zum Einsatz von KI in der Wirtschaft sowie eigene Erfahrungen der Autoren in **Projekten**⁶⁹ und **Kooperationen**⁷⁰ zugrunde. Die Fokussierung auf die größten Unternehmen erfolgt, weil sie als potenzielle Partner für die KI-Recheninfrastrukturversorgung, aber auch als potenzielle große Nutzer öffentlicher KI-Recheninfrastrukturen zu betrachten sind.

Unternehmen	Branche	Umsatz (2019)	Mitarbeitende (2019)	KI-Potenzial
Fresenius (fresenius.de)	Medizintechnik	35,4 Mrd. Euro	ca. 300.000	■ Hoch
Deutsche Bank (db.com)	Bank- und Kreditgewerbe	32,2 Mrd. Euro	ca. 90.000	■ Hoch
Heraeus Holding GmbH (heraeus.com)	Mischkonzern mit Medizintechnik	22,4 Mrd. Euro	ca. 14.900	■ Mittel bis hoch
Merck KGaA (merckgroup.com/de)	Chemie und Pharma	16,2 Mrd. Euro	ca. 57.000	■ Mittel
Commerzbank (commerzbank.de)	Bank- und Kreditgewerbe	9,5 Mrd. Euro	ca. 48.000	■ Hoch
T-Systems (t-systems.com)	Information & Kommunikation	5 Mrd. Euro	ca. 12.000	■ Sehr hoch

Tabelle 2 - KI-Potenzial für hessische Großunternehmen

Es zeigt sich, dass die größten in Hessen ansässigen Unternehmen durchweg ein hohes oder sogar sehr hohes Potenzial für die Nutzung von KI aufweisen.

3.2.2 Größte Mittelständler⁷¹

Auch für die größten Mittelständler lässt sich eine vergleichbare Übersicht darstellen. Die Zuordnung der mittelständischen Unternehmen zu Branchen und die Abschätzung des Bedarfs an Infrastrukturen für Künstliche

Intelligenz ist in Tabelle 3 dargestellt. Wie bei den oben genannten Großunternehmen besteht auch hier ein Potenzial zur gemeinsamen Nutzung von KI-Recheninfrastrukturen mit öffentlichen Einrichtungen.

Unternehmen	Branche	Umsatz (2019)	Mitarbeitende (2019)	KI-Potenzial
Kion Group AG (kiongroup.com)	Fördertechnik	8,8 Mrd. Euro	ca. 34.600	■ Sehr hoch
Ineos Styrolution GmbH (ineos-styrolution.com)	Chemie	5 Mrd. Euro	ca. 3.600	■ Hoch
Viessmann Werke GmbH & Co. KG (viessmann.de)	Heiztechnik	2,6 Mrd. Euro	ca. 12.300	■ Mittel
Friedhelm Loh Group (friedhelm-loh-group.com)	Mischkonzern	2,6 Mrd. Euro	ca. 12.200	■ Hoch
Software AG (softwareag.com)	Informationstechnologie	0,9 Mrd. Euro	ca. 5.000	■ Sehr hoch

Tabelle 3 - KI-Potenzial hessischer Mittelständler

Die Auswertung der wirtschaftlichen Aktivitäten der genannten Mittelständler zeigt, dass auch hier in vier von fünf Fällen ein hohes Potenzial für die Nutzung der KI besteht. Im Fall der Kion Group ist sogar ein sehr hohes Potenzial zu sehen.

4

Welche Konzepte sind noch zu betrachten?



4.1 Generelle Überlegungen

Recheninfrastrukturen für Künstliche Intelligenz sollten sich immer am Bedarf der jeweiligen Nutzergruppe ausrichten. Je nach Zielgruppe (öffentliche Einrichtung, Wissenschaft, Industrie) oder bei überlappenden Zielgruppen benötigt man verschiedene Klassen von Infrastrukturen und darüber hinaus auch verschiedene Betriebsmodelle.

Deswegen gilt es, beim Aufbau der Recheninfrastruktur verschiedenste Fragen zu diskutieren, unter anderem:

- Wer sollen meine Nutzerinnen und Nutzer werden?
- Welchem Zweck soll das System dienen?
- Welche Ausfallsicherheit brauche ich?
- Brauche ich eine reine KI-Recheninfrastruktur oder ein hybrides System mit weiteren Funktionen (dies hängt dann auch von der erwarteten Nutzung des Systems und den Workflows der Nutzerinnen und Nutzer ab)?
- Welche Technologie-Diversität brauche ich?
- Welche Geschäftsmodelle will ich verfolgen und welche rechtlichen Fragestellungen ergeben sich dadurch?

Vor diesem Hintergrund werden kurz die potenziellen Arten von Recheninfrastrukturen, verfügbare und zukünftige Technologien (soweit schon bekannt) und Zugangsmodelle sowie weitere Aspekte betrachtet.

4.2 Die geeignete Recheninfrastruktur für den jeweiligen Zweck

Bei Betrachtung der verschiedenen KI-Aktivitäten wird zweifellos deutlich, dass das Interesse von Wissenschaft und Wirtschaft an KI steigt. Allerdings ist das spezifische Interesse immer an die jeweilige Erfahrung (falls vorhanden) der Nutzerinnen und Nutzer gebunden, sodass in Hinblick auf eine Unterstützung der möglichen Anwendungsfelder auch unterschiedliche Konfigurationen von Infrastrukturen einbezogen werden müssen.

Um alle Phasen des Entwicklungs- und Lebenszyklus von KI-Anwendungen abzudecken, braucht es eine Reihe von verschiedenen Recheninfrastrukturen, die auf die jeweiligen Phasen abgestimmt sind und in einer offenen und unterstützten Art zugänglich sind.

Dabei kann man folgende Arten von Infrastrukturen unterscheiden, auf die im Folgenden näher eingegangen wird:

- Onboarding-Infrastrukturen
- Produktionsinfrastrukturen
- Forschungs- und Entwicklungsinfrastrukturen
- Schulungsinfrastrukturen

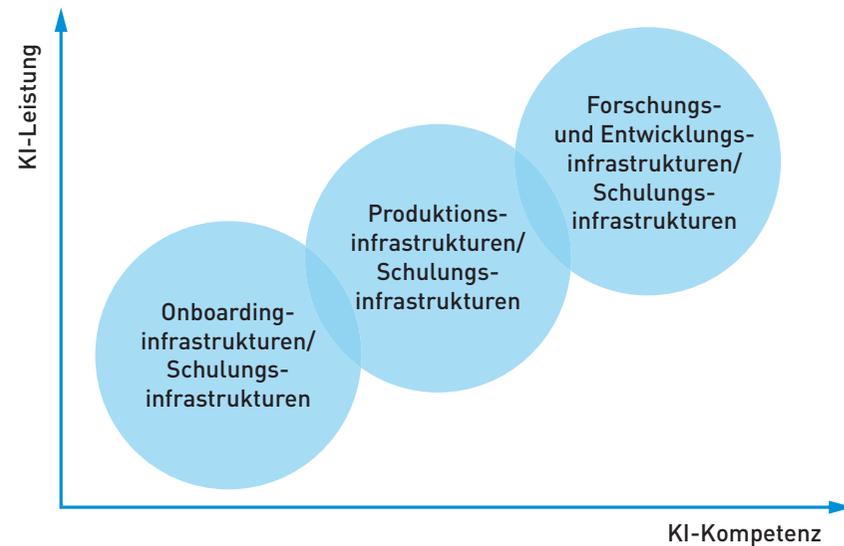


Abbildung 1: Grobe Positionierung der verschiedenen Infrastrukturarten je nach Kompetenz und benötigter Leistung

4.2.1 Onboarding-Infrastrukturen

Onboarding-Infrastrukturen sind diejenigen Recheninfrastrukturen, die in der Art einer initialen Testumgebung funktionieren. Diese Systeme können aus mehreren verschiedenen Technologien (unterschiedlich auch im Reifegrad) zur Verfügung stehen und sollen den Nutzerinnen und Nutzern ein Ausprobieren ermöglichen. Gerade für Nutzerinnen und Nutzer aus der Wissenschaft kann diese Art der Infrastruktur dazu dienen, die Potenziale aufzuzeigen, aber auch neue Ansätze zu erproben. Ein Beispiel für Forscherinnen und Forscher ist das **BEAST⁷²**-System (Bavarian Energy, Architecture and Software Testbed) des Leibniz-Rechenzentrums in Garching, welches eine Bandbreite verschiedener Technologien beinhaltet.

Generell könnten diese Onboarding-Infrastrukturen aber auch dezentral zur Verfügung gestellt werden, z. B. als Verbund mehrerer Recheninfrastrukturbetreiber.

4.2.2 Produktionsinfrastrukturen

Produktionsinfrastrukturen sind die Recheninfrastrukturen, die für den produktiven Einsatz für Wissenschaft und vor allem Wirtschaft (im täglichen Geschäftsbereich, meist kommerziell) genutzt werden können. Im Gegensatz zu Infrastrukturen, welche ausschließlich für Forschung und Entwicklung genutzt werden, müssen Produktionsinfrastrukturen klare und strenge Anforderungen an Bedingungen hinsichtlich Verfügbarkeit, Datenmanagement und Datensicherheit, Leistung des Systems und gegebenenfalls Nutzbarkeit auch durch Nicht-Experten erfüllen. So ist ein Merkmal die den Nutzerinnen und Nutzern angepasste Größe: Ein System kann so entweder auf einen Powernutzer ausgerichtet sein oder auch auf den Bedarf mehrerer Nutzerinnen und Nutzer. Gerade im wirtschaftlichen Umfeld gilt es sicherzustellen, dass diese Systeme stabil laufen und robust sind. Dies verlangt eine klare Planung und Kontrolle der Hardwareverfügbarkeit, aber auch Services wie Zugangs- und Sicherheitskonzepte müssen etabliert sein. Es hat sich hier gezeigt, dass cloudähnliche Zugänge und eine Sicherheitszertifizierung (z. B. ISO27001) von der Industrie bevorzugt werden.

Generell ist es sinnvoll, Produktionsumgebungen vor allem für den Mittelstand pro Anwendungsfeld zentral zur Verfügung zu stellen. Bei großen Firmen wird es in den meisten Fällen dezentrale Inhouse-Lösungen geben, zu deren Vorbereitung die Firmen jedoch Zugriff auf zentrale Produktionsumgebungen benötigen können.

4.2.3 Forschungs- und Entwicklungsinfrastrukturen

Forschungs- und Entwicklungsinfrastrukturen sind auf die verschiedenen Nutzergruppen und deren Können zugeschnitten. In Deutschland und auch in Europa sind diese Recheninfrastrukturen in Leistungsebenen (englisch: Tiers) unterteilt, welche die Leistungsfähigkeit und somit auch die Rechenkapazitäten der einzelnen Systeme widerspiegeln.

Oft sind diese Systeme zudem für verschiedene Anwendungsgebiete angepasst und können teils auch von der Industrie in kleinem Maßstab als Produktionsumgebung genutzt werden. Allerdings macht dies dann Anpassungen des Betriebsmodells notwendig, um den Anforderungen an Produktionsumgebungen gerecht zu werden. Eine Übersicht über die verschiedenen Anbieter (Rechenzentren) und deren Systeme kann z. B. für Ebene 1 (Tier 1) der Bundeshöchstleistungsrecheninfrastrukturen auf der Webseite des [Gauss Centre for Supercomputing](#)⁷³ gefunden werden, die Ebene 2 und 3 (Tier 2 und Tier 3) auf der Webseite der [Gauß-Allianz](#)⁷⁴.

Generell können diese Forschungs- und Entwicklungssysteme bis Tier 2 dezentral zur Verfügung gestellt werden, z. B. als Verbund mehrerer Recheninfrastrukturbetreiber. Tier 1 sollte zentral positioniert werden, da die Mechanismen rund um die Infrastruktur nicht trivial zu implementieren sind.

4.2.4 Schulungsinfrastrukturen

Ein nicht zu vernachlässigender Zweck von Recheninfrastrukturen (auch im Bereich Künstliche Intelligenz) ist die Ausbildung der Nutzerinnen und Nutzer. Gerade in Deutschland gibt es ein sehr gut ausgearbeitetes Schulungskonzept für Nutzerinnen und Nutzer aus Wissenschaft und Wirtschaft im Bereich Verstehen und Anwendung von High Performance Computing und Big Data (Analytics), welches zurzeit auch um Kurse im Bereich Künstliche Intelligenz kontinuierlich erweitert wird. Hierbei ist es absolut wichtig, nicht nur theoretische Schulungen zu bieten, sondern auch die Möglichkeit sogenannter Hands-on-Schulungen sicherzustellen. Dies geschieht teilweise mit Fernzugriff auf Recheninfrastrukturen, die z. B. den jeweiligen Technologieeigentümern oder Organisatoren der Schulungen gehören. Es sollte aber auch sichergestellt werden, dass diese Art von (kleinen) Systemen auch für lokale Schulungsaktivitäten verfügbar ist. Diese Systeme können Teile von Forschungs- und Entwicklungssystemen sein, es sollten aber aufgrund der Relevanz von Stabilität, Verfügbarkeit und Sicherheit keine Produktionssysteme per se dafür verwendet werden.

4.3 Verfügbare und potenzielle Technologien

Recheninfrastrukturen zur effizienten Nutzung von KI- und Big-Data-Anwendungen unterscheiden sich von Standard-HPC-Recheninfrastrukturen dadurch, dass Rechenoperationen hier zu einem hohen Grad parallel ausgeführt werden können. Je nach Anwendungsfeld ergeben sich weitere technische Herausforderungen, wie die Handhabung großer Datenmengen oder die Geschwindigkeiten, mit denen Daten generiert, übertragen und bearbeitet werden. Für Anwendungen, in denen Entscheidungen in Echtzeit verlangt werden, muss die Datenbearbeitung vor Ort stattfinden, dies ist aber nicht immer trivial mit Standardtechnologie machbar. Zusätzlich ist auch der Datentransfer zu Recheninfrastrukturen durch die verfügbaren Technologien limitiert. Beispielsweise schätzt man die Größe der beim autonomen Fahren generierten **Sensordaten pro Fahrzeug auf den Terabytebereich**⁷⁵ (10^{12} Byte). Allein dafür würde man schon eine Bandbreite von etwa 2 Gbit/s benötigen. Mit der neuen 5G-Technologie kann man 10 Gbit/s erreichen, aktuell ist aber das flächendeckend verfügbare LTE+ auf 1 Gbit/s limitiert. Auch wäre die aktuell theoretisch verfügbare Kapazität von 73,6 Tbit/s (Stand: Februar 2021) des weltgrößten Internetknotens DE-CIX dann bei etwa 37.000 Fahrzeugen bereits ausgeschöpft.

Mobile Systeme sind hinsichtlich des verfügbaren Platzes und verfügbarer Leistung begrenzt, daher kann auf einem solchen System nur ein Teil der Daten bearbeitet werden. Für den Rest sind stationäre Systeme vonnöten. Dabei werden die Daten in der Regel in Datenzentren geliefert, um dort vorverarbeitet, gespeichert und mehrmals verwendet werden zu können. Gerade bei mobilen Systemen, insbesondere bei autonom fahrenden oder fliegenden Fahrzeugen, muss außerdem sichergestellt werden, dass die Infrastruktur in der Lage ist, auch Netzausfälle zu kompensieren.

Nach dem heutigen Stand können viele Technologien Künstliche Intelligenz im Rahmen dedizierter Recheninfrastrukturen sehr gut unterstützen. Dabei hat jede Technologie ihre Vor-, aber unter Umständen auch Nachteile.

Folgendes Transferbeispiel lässt sich vor Augen führen:

Sowohl ein Lastkraftwagen als auch ein Rennwagen können Waren von A nach B transportieren. Der Rennwagen kann viel schneller (höhere Geschwindigkeit/Taktfrequenz) von A nach B reisen, und er kann schnell be- und entladen werden, da kleinerer Transportraum (Latenz), kann aber auch nur wenig transportieren (Bandbreite). Wenn es um große Mengen geht, braucht ein LKW längere Zeit für das Be-/Entladen, aber dank großem Transportraum wird indirekt auch eine höhere Bandbreite erreicht und die LKW-Lösung kann insgesamt schneller sein.

In der Analogie könnte der Rennwagen für einen Prozessor und der Lastwagen für eine Grafikkarte stehen. Detaillierte Ausführungen dazu folgen. Generell ist davon auszugehen, dass jede Auswahl eine klare Analyse der Anwendungsgebiete und der resultierenden benötigten Eigenschaften erfordern wird.

Eine Übersicht potenzieller Technologie Kandidaten (z. B. Prozessoren oder Grafikprozessoren) findet sich in Anhang 3 dieser Publikation.

4.4 Zugangsmodelle und generelle Überlegungen

Obwohl der Fokus dieser Bestandsaufnahme auf den Recheninfrastrukturen liegt, sollten in diesem Kontext auch wichtige andere Themenfelder wie Zugangsmodelle, Sicherheitskonzepte und weitere Randthemen kurz betrachtet werden, da diese auch die Wahl der Infrastruktur und des Anbieters beeinflussen können (siehe Abbildung 2).



Abbildung 2: Mit der Konzeption und Bereitstellung von Recheninfrastruktur eng verbundene Themen

4.4.1 Zugangsmodelle

■ Direktzugang

Der Direktzugang ist das klassische Modell, bei dem die Nutzerinnen und Nutzer direkt über Techniken wie z. B. **Secure Shell (SSH)**⁷⁶ Zugriff auf die Recheninfrastrukturen bekommen. Beim Direktzugang stehen den Nutzerinnen und Nutzern oft dedizierte Hardwareressourcen zur Verfügung. Das kann auch in beschränktem Umfang wichtige Infrastrukturkomponenten wie Netzwerke betreffen. Hier gibt es dann eine 1:1-Beziehung zwischen dem Anbieter der Infrastruktur und den Nutzerinnen und Nutzern. Diese Zugangsart ist in der Regel sicherer und performanter, bedingt aber auch mehr Aufwand und Kenntnisse dafür. Daher ist es oft gerade am Beginn der Nutzung von externen Recheninfrastrukturen für ungeübte Anwenderinnen und Anwender schwer, diese auch effizient zu nutzen.

■ Cloud-Zugang

Das Cloud-Modell bietet vielseitige Möglichkeiten, um Nutzerinnen und Nutzern Zugriff zu Recheninfrastrukturen zu geben. So lassen sich mit dem Konzept Infrastructure-as-a-Service (IaaS) entweder ein Public-Cloud-Konzept, ein Private-Cloud- oder ein Hybrid-Cloud-Konzept implementieren. Letzteres ermöglicht die Nutzung der Recheninfrastrukturen auch geteilt mit verschiedenen Nutzerinnen und Nutzern.

Bei einer Public Cloud erfolgt die Nutzung von (oft mit anderen Kunden geteilten) Ressourcen über das Internet. Bei einer Private Cloud sind die Ressourcen nur einem oder wenigen Kunden dediziert und über ein meist internes Netzwerk zugänglich.

Bei IaaS werden die einzelnen Ressourcen von Recheninfrastrukturen (u.a. Prozessoren, Speicher, teilweise auch Netzwerkressourcen) in virtualisierter Form zur Verfügung gestellt. Dies bringt den Vorteil der einfachen Skalierung, kann aber auch zu Leistungsverlusten führen, da die Virtualisierungssoftware selber auch Ressourcen benötigt.

Basierend auf IaaS können auch erweiterte Lösungen angeboten werden, die nicht nur die Recheninfrastrukturen beinhalten, sondern auch z.B. die Software im Paket mit inbegriffen haben. Solche Software-as-a-Service-Lösungen (SaaS) in Kombination mit IaaS und eventuell weiteren Erweiterungen erschließen in Bezug auf Künstliche Intelligenz gerade einen Markt für **Artificial-Intelligence-as-a-Service (AlaaS)**⁷⁷. Allerdings ist es hier auch wichtig, auf Sicherheitsaspekte zu achten. Manche Cloud-Technologien können hier nicht allen Anforderungen gerecht werden, jedoch möglicherweise als Testumgebung genutzt werden.

4.4.2 Sicherheitskonzepte

Gerade bei industriellen Anwendungen sind die Sicherheitskonzepte, welche rund um Recheninfrastrukturen aufgebaut werden, immens wichtig. Dies gilt auch für KI-Recheninfrastrukturen. Hier gilt es, die Bedarfe in Bezug auf Datensicherheit, Sicherheit der Ressourcen und Netzwerke sicherzustellen und gegebenenfalls einem anderen Sicherheitskonzept als dem eines öffentlich zugänglichen Ressourcenanbieters zu folgen, bei denen z. B. keine dedizierten Netzwerkverschlüsselungen angeboten werden oder ein Resource-Sharing (Recheninfrastruktur, Dateninfrastruktur) möglich ist. Dazu gehört auch die Betrachtung von Zertifizierungen, welche im industriellen Bereich immer relevanter bei der Auswahl der Zulieferer (auch von Recheninfrastrukturen) werden.

■ Zertifizierungen

Generell ist bei der Einrichtung von Recheninfrastrukturen auf die vorhandenen Sicherheitskonzepte der Infrastrukturbetreiber zu achten und gegebenenfalls diese dann den Nutzeranforderungen anzupassen.

Gerade im Bereich der industriellen Nutzung wird immer mehr auf etablierte Zertifizierungen geachtet, z. B. [ISO/IEC 27001](#)⁷⁸. Diese Zertifizierung reguliert den Aufbau und das kontinuierliche Aktualisieren eines umfangreichen Informationsmanagementsystems für Organisationen, z. B. Recheninfrastrukturanbieter. Dabei decken diese Standards nicht nur kommerzielle Organisationen ab, sondern beispielsweise auch Non-Profit-Organisationen.

Daneben gibt es fachspezifische Zertifizierungen, die sich auf Industriezweige fokussieren, wie z. B. [TISAX](#)⁷⁹ (Trusted Information Security Assessment Exchange) oder der VDA Information Security Assessment Katalog ([VDA-ISA](#))⁸⁰, die von der Automobilindustrie definiert wurden.

■ Datensicherheit

Gerade im vorwettbewerblichen, aber auch wettbewerblichen Kontext ist bei den Recheninfrastrukturen Datensicherheit zu gewährleisten. Dies bedeutet, dass auch Konzepte wie Virtualisierung so angewandt werden müssen, dass die Daten einer Nutzerin oder eines Nutzers nicht durch andere Nutzerinnen und Nutzer einsehbar, nutzbar oder manipulierbar werden. Auch Datenverlust muss verhindert werden.

4.4.3 Netzanbindung

Je nach Anwendungsbereich und -art ist eine qualitativ hochwertige und stabile Netzanbindung für die Recheninfrastrukturen wichtig. Dies bedeutet, dass die Bandbreite der Datenübertragung zu den Nutzszenarien passen und gegebenenfalls auch durch Kontrollmechanismen gewährleistet werden muss.

4.4.4 Software-Verfügbarkeit

Hinsichtlich nutzbarer Software auf den Recheninfrastrukturen gilt es sicherzustellen, dass Lizenzen entweder vor Ort verfügbar sind oder bei Bedarf von den Nutzerinnen und Nutzern mitgebracht werden können. Auch ist es wichtig, die Recheninfrastrukturen an die Anforderungen der Software der Nutzergruppen anzupassen, um einen maximalen Nutzungsgrad zu erreichen.

4.4.5 Expertise

Im Unterschied zu den klassischen Angeboten von Amazon und anderen Anbietern stellen zielgerichtete (z. B. industrieorientierte) KI-Recheninfrastrukturen Services zur Verfügung, die über den reinen Zugang zu den Infrastrukturen hinausgehen. Gerade für die Vielfalt an potenziellen Nutzerinnen und Nutzern sind auch der Zugriff auf Expertinnen und Experten, die Unterstützung bei der Technologienutzung oder die Optimierung der Anwendungen wichtig. Sofern man ein Komplettpaket anbieten möchte, welches über die Standards der Cloud-Angebote hinausgeht (was insbe-

sondere für kleine und mittelständische Unternehmen wichtig sein kann), sollte neben der Recheninfrastruktur auch eine Anwenderunterstützung implementiert werden.

4.4.6 Datenhaltung

Im Verbund mit der Netzanbindung ist auch die Datenhaltung ein wichtiges Thema. Hier gilt es sicherzustellen, dass die Daten während der Bearbeitung an der richtigen Stelle sicher gehalten werden, dann aber je nach Bedarf entweder in Langzeitsilos kommen oder, gerade bei industriellen Anwendungen, schnell wieder von den Recheninfrastrukturen entfernt werden können.

4.5 Cloud-Anbieter

Laut der Studie **Cloud-Monitor 2020**⁸¹ von KPMG und der Bitkom nutzen drei von vier Unternehmen in Deutschland 2019 bereits Cloud-Ressourcen. Die Studie betrachtet ebenso die Verbreitung der Internet-of-Things-Anwendungen (IoT) in der Public Cloud und weitere Wachstumspotenziale für IoT, Datenspeicherung, Spracherkennung und KI-Anwendungen. Zwei von fünf Nutzenden speichern kritische Informationen in der Public Cloud und bei jedem zweiten Nutzenden liegen Kundendaten in der Public Cloud. Allerdings ist es für mehr als 80 Prozent der Befragten wichtig, dass der Hauptsitz eines Cloud-Providers im Rechtsgebiet der EU/Deutschland liegt. Generell bevorzugt ein Großteil der Unternehmen Private Clouds, um mehr Kontrolle über die Randbedingungen zu haben.

Die größten Public-Cloud-Anbieter (Amazon, Microsoft, Google) und auch deren Tochterunternehmen in der EU fallen unter den **CLOUD Act (Clarifying Lawful Overseas Use of Data Act)**⁸². Je nach Anwendungsszenario ist Vorsicht bei der Auswahl des Cloud-Providers oder -Dienstleisters geboten, mit Hauptaugenmerk auf die Bearbeitung von personenbezogenen, medizinischen oder anderen sensiblen Daten. Daher sind diese Anbieter nur bedingt für Industrieaktivitäten oder kritische Forschung nutzbar.

Es folgt eine Kurzübersicht über die prominentesten Cloud-Anbieter (in Bezug auf KI).

4.5.1 Amazon Web Services

Laut eigenen [Angaben](#)⁸³ laufen mehr als 80 Prozent der cloudbasierten KI-Anwendungen auf Amazon Web Services (AWS).

Amazon ist mit drei sogenannten Availability Zones (Datenzentren mit unabhängiger Infrastruktur und unterschiedlichen Standorten) in der Region Frankfurt am Main aufgestellt.

AWS bietet dafür IaaS-Recheninfrastrukturen in verschiedenen Konfigurationen und darauf basierend weitere Modelle wie Platform-as-a-Service (PaaS) oder AlaaS.

Als Rechenressourcen für KI (vor allem im Unterbereich neuronale Netze) werden (virtuelle) Server mit unterschiedlicher Anzahl von GPUs oder auch mit ASIC Chips, wie dem AWS Inferentia angeboten. In der AWS-Region Frankfurt sind die relevantesten zugreifbaren Technologien: NVIDIA Tesla V100, K80, T4 und AWS Inferentia. NVIDIA A100 sind zum Zeitpunkt dieser Datenerhebung (Januar 2021) in der AWS-Region Frankfurt noch nicht verfügbar.

4.5.2 Google Cloud

Google Cloud hat wie AWS drei Availability Zones in Frankfurt am Main und bietet verschiedene Cloud-Modelle wie IaaS, PaaS und AlaaS an.

Die KI-Workflows in der Region werden im Moment nur in einer Availability Zone mit NVIDIA T4 GPUs unterstützt. ASIC Chips Tensor Processing Units (TPUs) sind in der Region Frankfurt nicht verfügbar.

4.5.3 Azure

Microsoft Azure wird in Deutschland in vier Regionen bereitgestellt. Zwei davon (Frankfurt und Magdeburg) sind sogenannte Sovereign Regions, die man eher als Private Cloud betrachten kann.

Die Sovereign-Datenzentren wurden 2016 eröffnet und werden von T-Systems (im Auftrag von Microsoft) betrieben. Im Vergleich zu anderen Regionen verfügen sie nur über ein reduziertes Portfolio an Azure-Produkten. Hauptmerkmal dieser Regionen ist der Datenschutz – die Sovereign-Datenzentren werden in deutschem Rechtsgebiet und von deutschen Unternehmen betrieben. 2019 öffnete Microsoft zwei Public-Regionen (Frankfurt mit im Moment drei Availability Zones und Berlin – noch ohne Redundanz). In den Sovereign-Regionen werden keine neuen Kunden angenommen, sie werden im Oktober 2021 abgestellt.

Die GPU-unterstützten Ressourcen sind in Azure in Deutschland zum Zeitpunkt dieser Bestandsaufnahme nicht verfügbar.

5

Beispielmodelle der Interaktion von Wissenschaft und Wirtschaft



Kooperationen zwischen Wirtschaft und Wissenschaft sind auch für die Künstliche Intelligenz ein wesentlicher Schlüssel zur erfolgreichen Umsetzung wissenschaftlicher Entwicklungen in der wirtschaftlichen Praxis. Im Bereich der Recheninfrastrukturen stellt sich insbesondere die Frage, ob und wie Infrastrukturen gemeinsam beschafft, betrieben oder benutzt werden können. Weltweit gibt es dafür eine Reihe von Konzepten und erfolgreichen Beispielen. Diese sollen hier vorgestellt werden, um einen Überblick über die Kooperationsmöglichkeiten zwischen Wissenschaft und Wirtschaft zu erhalten.

5.1 AI Bridging Cloud Infrastructure

Im Jahr 2018 wurde in Tokyo mit der **AI Bridging Cloud Infrastructure (ABCI)**⁸⁴ ein Rechnersystem in Betrieb genommen, das der Industrie den Einstieg in die KI-Forschung und -Entwicklung ermöglichen sollte.

Dieses durch das National Institute of Advanced Industrial Science and Technology (AIST) initiierte Projekt AI Bridging Cloud Infrastructure (ABCI) fördert insbesondere die Kooperation zwischen Universitäten, Forschungseinrichtungen und der Industrie. Der von ABCI bereitgestellte Rechner steht derzeit auf Platz 14 der schnellsten Rechner der Welt. Er besteht aus 1.088 Knoten mit jeweils 4 NVIDIA Tesla V100 GPUs.

Die japanische Regierung hat mit ABCI eine Ressource geschaffen, die sowohl der Forschung als auch der Industrie zur Verfügung steht. Die Nutzung ist für die Industrie nicht kostenlos, sondern wird abgerechnet. Die Zugangsmodalitäten sind öffentlich einsehbar und sowohl in Japanisch als auch in Englisch verfügbar. Grundsätzlich kann also die ABCI auch von ausländischen Firmen genutzt werden.

Das Grundkonzept der ABCI besteht in einer Vorfinanzierung der notwendigen Ressourcen durch den Staat mit der Möglichkeit, die Kosten über die industrielle Nutzung zu refinanzieren. Im Vordergrund steht aber nicht die finanzielle Frage, sondern die Notwendigkeit, KI-Forschung und -Entwicklung in der Industrie anzustoßen und daher eine Anschubfinanzierung für eine industrielle Entwicklung zu geben.

Das **AIST**⁸⁵ in Japan ist eine der Fraunhofer-Gesellschaft vergleichbare Einrichtung mit dem Auftrag, die Kooperation zwischen Forschung und Industrie in Japan zu fördern. Das AIST umfasst sieben Bereiche, einer davon ist dem Thema Information Technology and Human Factors gewidmet.

5.2 Höchstleistungsrechner für Wissenschaft und Wirtschaft

Die Höchstleistungsrechner für Wissenschaft und Wirtschaft GmbH (HWW⁸⁶) ist ein Gemeinschaftsunternehmen des Landes Baden-Württemberg, des Karlsruher Instituts für Technologie, der Universität Stuttgart sowie der Unternehmen T-Systems und Porsche. Ziel der HWW GmbH ist es, Höchstleistungsrechner der Spitzenklasse und das zur Nutzung nötige Know-how für Unternehmen und Organisationen aus Wissenschaft und Wirtschaft als leicht zugänglichen Service anzubieten. Damit wird diese für die Wettbewerbsfähigkeit kritische Schlüsseltechnologie auch für Organisationen zugänglich, die die dafür notwendigen Investitionen nicht selbst tätigen können oder wollen.

Die Leistungen der HWW werden nach einem Gebührenmodell nach Nutzung abgerechnet. Die bei öffentlichen Höchstleistungsrechenzentren übliche Verpflichtung zu Begutachtung und Veröffentlichung besteht nicht. Damit können die Leistungen der HWW direkt in missionskritischen Forschungs- und Entwicklungsbereichen eingesetzt werden.

Mit der Schaffung der HWW haben das Land Baden-Württemberg und die industriellen Partner ein Konzept geschaffen, das die Nutzung sowohl öffentlicher als auch industrieller Rechenressourcen in einem Poolmodell erlaubt. Die Ressourcen werden nicht von der HWW, sondern von ihren Partnern zur Verfügung gestellt. Die Nutzung der Ressourcen ist völlig offen und wird nach wirtschaftlichen Kriterien kommerziell abgerechnet. Durch die Einbeziehung von drei Partnern als Ressource Provider (Hochleistungsrechenzentrum Stuttgart, Steinbuch Center for Computing, T-Systems) kann die HWW jederzeit eine breite Palette von Ressourcen für Simulation und Künstliche Intelligenz bereitstellen.

5.3 INCITE

Im Jahr 2004 startete das Office of Science des US Department of Energy das Programm **Innovative and Novel Computational Impact on Theory and Experiment (INCITE) mit dem Ziel, Wissenschaft und Industrie den Zugang zu dessen Supercomputern zu ermöglichen**⁸⁷. Diese Supercomputer wurden für Computational Science – also im wesentlichen Computersimulationen – beschafft. Im Rahmen des Programms können gemeinsame Forschungen zwischen Wissenschaft und Industrie unterstützt werden. In geringem Umfang (rund 1 %) können Ressourcen auch für rein industrielle Nutzung gegen Kostenerstattung genutzt werden.

Im Rahmen des Programms kann auf die Ressourcen des Argonne National Laboratory (ANL) und des Oak Ridge National Laboratory (ORNL) zugegriffen werden. Künstliche Intelligenz wurde erst kürzlich in die Liste der Anwendungen aufgenommen, für die Ressourcen bereitgestellt werden sollen.

Für das Jahr 2021 werden 38 Mio. **Knotenstunden**⁸⁸ für 51 Projekte bereitgestellt. Das entspricht einem Rechner von etwa 4.000 Knoten und damit einem System, das zu den schnellsten in Deutschland zählen würde.

Von den genannten 51 Projekten setzen laut Projektbeschreibung drei Projekte mit rund 1,12 Mio. Knotenstunden (also nur rund 3% der bereitgestellten Rechenleistung) Methoden der Künstlichen Intelligenz ein. Weitere vier Projekte mit rund 3,1 Mio. Knotenstunden (also nur rund 8%) setzen Methoden des Machine Learning (ML) ein. In Summe ist festzuhalten, dass momentan nur rund 10 Prozent der im Programm INCITE bereitgestellten Rechenleistung auf den Bereich der Künstlichen Intelligenz entfallen. Dieser Rechenzeitbedarf kann also aktuell mit einem moderat großen System abgedeckt werden. Die Nutzung von Hardwareressourcen

für KI ist also – wie die gestiegene Nachfrage und die Aufnahme von KI in die Anwendung des Supercomputings zeigt – in den USA im Steigen begriffen, wird aber zunächst noch einen geringen Anteil an der Nutzung der Rechnersysteme darstellen.

5.4 EuroHPC

EuroHPC⁸⁹ (European High Performance Computing) ist eine gemeinsame Unternehmung (Joint Undertaking – JU), die von der Europäischen Kommission mit 32 EU-Mitgliedsstaaten und assoziierten Staaten etabliert wurde. Die JU resultierte aus einer Erklärung von 2017, die von 22 Ländern unterzeichnet wurde und die darauf abzielt, europäisches HPC im internationalen Kontext wettbewerbsfähig und nutzbar für Wissenschaft und Wirtschaft zu machen. Dabei betrachtet man hauptsächlich die Bereitstellung und gegebenenfalls Neuentwicklung von Hardware und Infrastrukturen. In diesem Zuge wurden 2020/21 in einer ersten Phase fünf Petascale-Systeme (Systeme mit einer Leistung von mehr als 10^{15} mathematischen Operationen pro Sekunde) von der JU co-finanziert, wovon jeweils eines in **Bulgarien**⁹⁰, **Tschechien**⁹¹, **Luxemburg**⁹², **Portugal**⁹³ und **Slowenien**⁹⁴ installiert ist. Es werden damit sowohl HPC- als auch KI-Anwendungen avisiert (**durch hybride Recheninfrastrukturen aus Prozessoren und Grafikkarten**⁹⁵). Dazu wird es in den Ländern **Spanien**⁹⁶, **Italien**⁹⁷ und **Finnland**⁹⁸ jeweils ein Pre-Exascale-System geben. Diese Systeme gehören zu der Weltspitze an verfügbarer Rechenleistung und unterstützen ebenfalls KI-Anwendungen durch große Grafikkartensysteme. Das Ziel ist, in den nächsten Jahren über Systeme zu verfügen, die mehr als 10^{18} mathematische Operationen pro Sekunde ausführen können. Pre-Exascale bezeichnet einen Zwischenschritt (startend von 10^{15} Operationen) in diese Richtung.

Durch die Co-Finanzierung ist auch das Zugriffs- und Benutzungsmodell von EuroHPC-Infrastrukturen etwas komplexer im Vergleich zu Systemen, die innerhalb eines Staates beschafft werden. Bei den großen Systemen (Finnland, Italien und Spanien) zahlt die Joint Undertaking bis zu 50% der Beschaffungs- und Betriebskosten, die damit dann auch als Besitzerin des Systems gilt. Bei den kleineren Systemen (Bulgarien, Tschechien, Luxemburg, Portugal und Slowenien) sind es bis zu 35% Co-Finanzierung der Beschaffungskosten. EuroHPC gilt hier als mitbesitzende Entität.

Gerade bei den Pre-Exascale-Systemen wird die Gegenfinanzierung zu EuroHPC meist von einem Konsortium mehrerer Länder erbracht. So besteht beispielsweise das Leonardo-Konsortium, welches das System mit Standort Italien beschaffen und betreiben wird, aus den Ländern Italien, Österreich, Slowakei, Slowenien und Ungarn. Je nach Beitrag zu den Gesamtkosten des Systems verfügen diese Länder über einen entsprechenden Anteil an der Nutzungszeit der Ressourcen des Systems und können diese somit an Wissenschaft und Industrie in ihrem jeweiligen Land weitergeben.

5.5 CCRT

Eine besondere Form der Kooperation zwischen Wirtschaft und Wissenschaft besteht in Frankreich. Hier wurde 2003 eine gemeinsame Initiative von Industrie und Forschung ins Leben gerufen, um gemeinsam Rechner zu beschaffen und zu nutzen. Unter dem Namen Centre de Calcul Recherche et Technologie (**CCRT**⁹⁹) wurden seit 2003 die Ressourcen des französischen Commissariat à l'énergie atomique et aux énergies alternatives (CEA) gemeinsam von CEA und Industrie beschafft und genutzt. Dabei entsprach der Nutzungsanteil von öffentlicher Forschung und privater Wirt-

schaft dem erbrachten Kostenanteil. CCRT wird innerhalb der CEA betrieben und nicht als eigenständige Einrichtung geführt. Die industriellen Partner sind de facto nur französische Firmen, da die nationale Bedeutung der CEA eine Beteiligung ausländischer Firmen zumindest schwierig macht. Das Konzept bietet damit die Möglichkeit, finanzielle Ressourcen aus Wissenschaft und Industrie zu bündeln. Im konkreten Fall der CEA existiert ein organisatorischer Rahmen, der dies auch politisch absichert.

5.6 Beispiel zur Unterstützung von KMU zur Erbringung von KI-Serviceleistungen

Ein gutes Beispiel für die Beschleunigung der KI-Anwendung zur Erbringung gewerblicher Angebote ist das 2019 in Stuttgart gegründete Startup **LandesCloud¹⁰⁰**. Die LandesCloud bietet KMU, mittelständischen Unternehmen sowie Großunternehmen Services in Form von Data-Analytics-as-a-Service an.

Grundvoraussetzung für die angebotenen Leistungen sind die Garantie eines geregelten Umgangs mit Daten, eine sichere Plattform und sicherer Datenaustausch. Die LandesCloud bietet momentan Kunden eine sichere und lokale (die Daten bleiben im Land) Cloud-Umgebung mit vorinstallierten Softwarepaketen und Entwicklungsumgebungen. Außer Datenanalyse wird auch eine Entwicklungsumgebung für das Erstellen und Trainieren von KI-Modellen angeboten.

Die Konzepte und Möglichkeiten der LandesCloud wurden bereits vor ihrer Gründung durch das Zusammenspiel mit dem Höchstleistungsrechenzentrum Stuttgart evaluiert, welches eine Test- und Onboarding-Recheninfrastruktur sowie Expertise unterstützend anbot. Dabei wurden die KI-basierten Aktivitäten zunächst nur auf CPU-basierten Recheninfrastrukturen ausgeführt. Nach dieser Phase wurde die LandesCloud als Produktivumgebung etabliert.

6

Potenziale und Empfehlungen



6.1 Ist-Zustand in Hessen und was noch benötigt wird

Die Untersuchung der Bedarfe einerseits und der verfügbaren Hardware andererseits zeigt grundlegend folgende Situation:

Der Bedarf an Rechenleistung und damit an einer Infrastruktur, die für Künstliche Intelligenz im Land Hessen zur Verfügung steht, wird kurzfristig sowohl in der Wissenschaft als auch in der Wirtschaft steigen und mittelfristig sogar sehr stark ansteigen.

In einem Zeitraum von zwei bis drei Jahren ist damit zu rechnen, dass Forscherinnen und Forscher im Bereich der Künstlichen Intelligenz auf große Systeme zurückgreifen müssen, auch wenn sie solche Systeme nicht alleine auslasten werden.

Aufgrund der Potenziale in den großen Unternehmen des Landes (hier vor allem Banken, IT und Medizin), aber auch aufgrund der im [Monitoring-Report Wirtschaft Digital 2017](#)¹⁰¹ dargestellten Situation bei den KMU, ist auch im wirtschaftlichen Bereich ein hoher Anstieg des Bedarfs zu erwarten.

Diesem steigenden Bedarf stehen zumindest erste Planungen gegenüber. Zum einen ist bei Cloud Services ein Ausbau von KI-fähiger Infrastruktur zu beobachten. Zum anderen ist für die Beschaffung eines Hochleistungsrechners am Hochschulrechenzentrum der TU Darmstadt mit der Berücksichtigung von KI-Anforderungen zu rechnen.

Darüber hinaus gibt es jedoch keine belastbaren und fest geplanten Beschaffungskonzepte. Weder im Rahmen der NHR noch im Rahmen von GAIA-X sind derzeit konkrete und belastbare Planungen für KI-Rechnerinfrastrukturen zu sehen. Es tut sich also eine Lücke in der Versorgung auf, die die Gefahr birgt, die Wissenschaft in ihrer Forschung zu verlangsamen und im internationalen Wettbewerb zu benachteiligen. Darüber hinaus kann diese Lücke auch im Bereich der Wirtschaft aufgrund der mangelnden Ressourcen zu einem nur verzögerten Ausbau der KI-Nutzung führen. Somit ist auch im wirtschaftlichen Bereich eine Verzögerung im internationalen Wettbewerb zu erwarten.

6.2 Handlungsbedarf und Potenzial für Wirtschaft und Wissenschaft

Das Land Hessen verfügt über ein deutlich sichtbares Potenzial, um in Wissenschaft und Wirtschaft eine führende Rolle im Bereich KI einzunehmen. Dabei ist das Potenzial für beide, Forschungseinrichtungen und Unternehmen, sowohl im Bereich der KI-Entwicklung als auch in der Anwendung von KI zu sehen. Grundlagen dafür sind schon gelegt, z. B. durch die Einrichtung von hessian.AI, die erfolgreiche hessische Forschungslandschaft, Hessens wichtige Rolle als Rechenzentrumsstandort und die bestehenden Initiativen im Breitbandausbau. Erwähnung finden sollte auch die allgemein stetig wachsende Zahl von Unternehmen, die im Bereich KI aktiv sind oder KI schon heute nutzen. Im Bereich der verfügbaren KI-Recheninfrastrukturen an den Standorten der öffentlichen Forschung und der Wirtschaft ist das Land mit dem heutigen Stand zwar gut für eine initiale positive Entwicklung positioniert, jedoch noch nicht so weit, um international als Spitzenregion angesehen zu werden.

Zur Umsetzung der Potenziale wird daher die weitere Verbesserung der KI-Recheninfrastrukturen im Land Hessen empfohlen. Dabei müssen Möglichkeiten der Zusammenarbeit zwischen Wissenschaft und Wirtschaft geprüft werden, da sie die Grundlage schaffen können, nicht nur die notwendige Rechenleistung und Software zur Verfügung zu stellen, sondern darüber hinaus auch als Kristallisationskerne für inhaltliche Kooperationen zwischen Wissenschaft und Wirtschaft zu wirken.

Als bereits bestehende Kernelemente einer hessischen KI-Recheninfrastruktur sind zunächst das Hochschulrechenzentrum der TU Darmstadt und das CSC an der Goethe-Universität Frankfurt am Main – aufgrund ihrer herausragenden Rolle in der NHR bzw. im Hessischen Kompetenzzentrum

für Hochleistungsrechnen – sowie das hessian.AI wegen seiner gesamthessischen Ausrichtung anzusehen.

6.2.1 Empfehlungen

Empfehlung 1

In einer ersten Phase sollte das bereits bestehende Potenzial der vorhandenen KI-Recheninfrastruktur auch über die Wissenschaft hinaus genutzt werden. Dazu gehört insbesondere die Nutzung der bestehenden und zukünftigen Recheninfrastruktur am HRZ der TU Darmstadt, die für Wissenschaft und Wirtschaft zugänglich sein sollte. Dazu gehören aber auch die Nutzung der Infrastruktur am CSC an der Goethe-Universität Frankfurt am Main sowie weitere kleinere Ressourcen, die im Rahmen des Hessischen Kompetenzzentrums für Hochleistungsrechnen koordiniert werden. Ziel muss es sein, den Bedarfen der KI – wie sie in der Bestandsaufnahme im öffentlichen Bereich (mehrere PFLOPS) und im privaten Bereich erkannt wurden – Rechnung zu tragen. Aus dem CSC heraus ergibt sich ein Potenzial, das Thema des Hochleistungsrechnens und damit auch der Künstlichen Intelligenz sowohl in der Forschung als auch im Bereich der Infrastruktur zu stärken. Mittelfristig rechtfertigt der in der Bestandsaufnahme genannte Bedarf von mehreren PFLOPS auch die Beschaffung einer dedizierten KI-Maschine im Rahmen eines dedizierten KI-Recheninfrastrukturzentrums.

Empfehlung 2

Mit dem hessian.AI wurde ein starker Kristallisationspunkt für die KI in Hessen geschaffen. Es sollte daher eine enge Abstimmung der KI-Recheninfrastrukturen mit hessian.AI stattfinden, um durch die enge Verzahnung von Bedarf und Ressourcen eine optimale Steuerung beider Bereiche zu erreichen, ohne die wissenschaftliche Freiheit zu gefährden. Ein Ausbau von KI-Recheninfrastrukturen sollte sich immer am mittelfristig absehbaren Bedarf orientieren. Zugleich muss ein Ausbau der methodischen Forschung und Entwicklung immer auch die Verfügbarkeit der dafür notwendigen KI-Recheninfrastrukturen im Auge haben.

Empfehlung 3

Während große Firmen über die finanziellen Möglichkeiten verfügen, um sich entweder inhouse oder in der Cloud mit Zugang zu KI-Recheninfrastrukturen zu versorgen, ist die Versorgung der KMU und hier insbesondere der kleineren Unternehmen ungleich schwieriger.

Es sollte ausgelotet werden, wie über gezielte Förderprogramme des Landes insbesondere den KMU die Möglichkeit des Zugangs zu, aber auch der zielgerichteten Nutzung von KI-Recheninfrastrukturen durch Bereitstellung des notwendigen Know-hows zu verschaffen wäre. Darüber hinaus sollte geprüft werden, ob dies im Rahmen bestehender Förderprogramme wie **Distr@l**¹⁰² vorstellbar ist. Zusätzlich sollte der Know-how-Transfer insbesondere in Richtung KMU durch Förderprogramme des Landes angeschoben werden. Um vor allem für KMU die Potenziale der KI sichtbar zu machen, wird die Etablierung eines KI-Evangelisten empfohlen, der zwischen Wissenschaft und Wirtschaft vermitteln soll und den Fluss von Wissen sowie die Kommunikation zwischen den beiden Bereichen stärken soll.

Empfehlung 4

Da KI-Recheninfrastruktur und KI-Wissen eng miteinander verzahnt werden müssen, um im internationalen Wettbewerb erfolgreich zu bestehen, sollte grundsätzlich eine entsprechende Verzahnung auch bei allen Angeboten an Wissenschaft und Wirtschaft angestrebt werden, wenn es um Erweiterung der KI-Recheninfrastrukturen geht. Das Land Hessen kann dazu die Förderung von KI-Recheninfrastrukturen an die Verpflichtung knüpfen, diese – im Rahmen der gesetzlichen Bestimmungen – für die private Nutzung zugänglich zu machen. Das Land kann darüber hinaus die Vergabe von Forschungsfördermitteln in diesem Bereich in Verbundprojekten bündeln, sodass eine Einbindung der Wirtschaft schon in der Forschung und Entwicklung erfolgt und nicht später erst ergänzt werden muss. Damit sollten auch Unternehmen die Möglichkeit bekommen, in öffentlich geförderten Programmen im Bereich der KI mitzuwirken.

Empfehlung 5

Die Konzeptualisierung der bereitzustellenden Recheninfrastrukturen für KI sollte bedarfsorientiert sein und mit der Etablierung passender Mechanismen einhergehen, z. B. beim Zugangsmodell und der Sicherheit. Auch sollten Aspekte wie Technologievielfalt und Ausbaufähigkeit der Kapazitäten betrachtet werden. Der grundsätzliche Bedarf ist dabei klar, spezifische Aussagen zu den verschiedenen Stakeholdern (z. B. kleine und mittelständische Unternehmen) sollten im Rahmen einer weiteren Erhebung gewonnen werden.

Empfehlung 6

Die Nutzung von KI-Recheninfrastrukturen ist für Wissenschaft und Wirtschaft von zentraler Bedeutung. Beide Bereiche sind auf die Verfügbarkeit entsprechender Strukturen angewiesen. Lösungen, die nur auf einen Bereich abzielen, haben den Nachteil, keine Synergieeffekte zwischen Wissenschaft und Wirtschaft zu erzielen.

Vor diesem Hintergrund wird empfohlen, die in Kapitel 5 dargestellten unterschiedlichen Möglichkeiten der Kooperation zwischen öffentlichem und privatem Bereich vor dem Hintergrund der Wirtschafts- und Wissenschaftspolitik daraufhin zu prüfen, ob und wie eine Versorgung der Wissenschaft und Wirtschaft mit KI-Recheninfrastrukturen in einem synergetischen Konzept für das Land Hessen und seine wissenschaftlichen und wirtschaftlichen Player sinnvoll ist und umgesetzt werden kann.

Empfehlung 7

Auf bundesweiter Ebene sowie auf europäischer Ebene werden eine Reihe von Aktivitäten vorangetrieben, um Recheninfrastrukturen zu schaffen, die sowohl der Wissenschaft als auch der Wirtschaft zur Verfügung stehen sollen. Von besonderer Bedeutung ist hier GAIA-X, das als deutsche Initiative in der Zwischenzeit zu einer europäischen Aktivität geworden ist, und in dessen Rahmen insbesondere das Thema der Datenverfügbarkeit sowie der Verfügbarkeit von Methoden und Hardware bearbeitet wird. Hessen sollte sich daher in GAIA-X bewusst etablieren und bestehende Aktivitäten weiter ausbauen. Da GAIA-X in einer Definitionsphase ist, sollte sich Hessen in die laufenden Diskussionen von GAIA-X zeitnah einbringen.

6.2.2 Zusammenfassung

Die vorliegende Bestandsaufnahme weist für Hessen ein hohes Potenzial für die Nutzung von KI sowohl im wissenschaftlichen Bereich als auch in der Wirtschaft aus. Es könnte gehoben werden, wenn vorhandene und zukünftige KI-Recheninfrastrukturen stärker verzahnt werden und dies durch Förderprogramme des Landes unterstützt wird. Die Hürden zwischen den Bereichen Wissenschaft und Wirtschaft müssen dabei proaktiv angegangen werden, um in Summe eine optimale Weiterentwicklung beider Bereiche in den kommenden Jahren zu ermöglichen.

Anhang 1: Außeruniversitäre Forschungseinrichtungen mit Standorten in Hessen

Bundesinstitut für Bevölkerungsforschung, Wiesbaden

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen der KI ergeben sich im Bereich der Auswertung von Daten

Center for Financial Studies an der Goethe-Universität Frankfurt am Main, House of Finance, Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen der KI ergeben sich aus Datenauswertung und Anwendungen der Financial Studies

DECHEMA-Forschungsinstitut, Frankfurt am Main

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen ergeben sich dort, wo Datenauswertungen Forschungen beschleunigen können

Deutscher Wetterdienst (DWD), Offenbach

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich sowohl in der Datenauswertung als auch in der Auswertung von Simulationen

Deutsches Archäologisches Institut, Römisch-Germanische Kommission des (RGK), Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**: Anwendungen können sich in der Datenauswertung und in der Analyse von Texten ergeben

Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**: Anwendungen ergeben sich in der Datenauswertung

Deutsches Polen-Institut (DPI), Darmstadt

Gering

Deutsches Konsortium für Translationale Krebsforschung (DKTK), Heidelberg

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse

Deutsches Zentrum für Lungenforschung (DZL), Gießen

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse

Deutsches Zentrum für Infektionsforschung (DZIF), Gießen – Marburg – Langen

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse

Deutsches Zentrum für Herz-Kreislauf-Forschung e.V. (DZHK), Berlin

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse

Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in Datenanalyse und Simulation; eigener Schwerpunkt Künstliche Intelligenz

Fraunhofer-Institut für Betriebsfestigkeit und Systemzuverlässigkeit (LBF), Darmstadt

Einschätzung Bedarfspotenzial **Hoch**: Datenanalyse und Simulation; eigener Bereich Smart Digital Solutions

Fraunhofer-Institut für Graphische Datenverarbeitung (IGD), Darmstadt

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse sowie in der Visualisierung

Fraunhofer-Institut für Sichere Informationstechnologie (SIT), Darmstadt

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse sowie in Simulationen

Fraunhofer-Institut für Translationale Medizin und Pharmakologie (ITMP), Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse sowie in Simulationen

Fraunhofer-Institut für Energiewirtschaft und Energiesystemtechnik (IEE), Kassel

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse sowie in Simulationen

Fraunhofer-Projektgruppe für Wertstoffkreisläufe und Ressourcenstrategie (IWKS), Hanau und Alzenau

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen ergeben sich in der Datenanalyse

Freies Deutsches Hochstift – Frankfurter Goethe-Haus, Frankfurt am Main

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen in der Textanalyse

Fritz Bauer Institut (Studien-/Dokumentationszentrum zu Geschichte und Wirkung des Holocaust), Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**

Frobenius-Institut an der Goethe-Universität Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**

Georg-Speyer-Haus, Institut für Tumorbiologie und experimentelle Therapie, Frankfurt am Main

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen in der Datenauswertung sowohl externer Daten als auch interner (Histologie-)Daten

GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und Simulation; eigene Rechnerhardware

Herder-Institut für historische Osteuropaforschung, Marburg

Einschätzung Bedarfspotenzial **Mittel**: Auswertung von Daten und Analyse von Texten

Hessisches Landesamt für geschichtliche Landeskunde, Marburg

Einschätzung Bedarfspotenzial **Gering**: evtl. Datenanalyse

Institut für sozial-ökologische Forschung (ISOE) GmbH, Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und Simulation

Institut für Sozialforschung an der Goethe-Universität Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in der Datenanalyse

Institut für Steinkonservierung e.V., Mainz

Einschätzung Bedarfspotenzial **Gering**: evtl. Datenanalyse

Institut Wohnen und Umwelt GmbH (IWU), Darmstadt

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen ergeben sich in Datenanalyse und Simulation

Institute for Law and Finance an der Goethe-Universität Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**: evtl. Auswertung von Rechtsdatenbanken

Kommission für Archäologische Landesforschung in Hessen e. V.

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen im Bereich der Datenanalyse

Kriminologische Zentralstelle e.V., Wiesbaden

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse

Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK), Frankfurt am Main

Einschätzung Bedarfspotenzial **Mittel**: Anwendungen im Bereich der Datenanalyse

Max-Planck-Institut für Biophysik, Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und Simulation

Max-Planck-Institut für europäische Rechtsgeschichte, Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**: Anwendungen in der Textanalyse sind vorstellbar

Max-Planck-Institut für Herz- und Lungenforschung (W.G. Kerckhoff-Institut), Bad Nauheim

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und in der Simulation

Max-Planck-Institut für terrestrische Mikrobiologie, Marburg

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und in der Simulation

Max-Planck-Institut für Hirnforschung, Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und in der Simulation

Paul-Ehrlich-Institut, Bundesamt für Sera und Impfstoffe, Langen

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse und in der Simulation

Senckenberg Forschungsinstitute und Naturmuseen (SFN), Frankfurt am Main

Einschätzung Bedarfspotenzial **Hoch**: Anwendungen in der Datenanalyse

Sigmund-Freud-Institut, Forschungsinstitut für Psychoanalyse und ihre Anwendungen, Frankfurt am Main

Einschätzung Bedarfspotenzial **Gering**: Das Institut kann aber ein Begleitpartner bei der Untersuchung der Auswirkungen des Einsatzes von KI in der Gesellschaft sein.

Stiftung Archiv der deutschen Frauenbewegung, Kassel

Einschätzung Bedarfspotenzial **Gering**: Eventuell Analyse von Texten

Anhang 2: Fachterminologie

Die Fachbegriffe dieser Bestandsaufnahme wurden, soweit möglich, von vorherigen Studien über das Thema KI übernommen, um eine Kontinuität der Begriffsverwendung sicherzustellen. Begriffsdefinitionen, die mit (*) gekennzeichnet sind, wurden aus der Studie KI – Künstliche Intelligenz: Wirtschaft und Innovation von P. Buxmann übernommen.

Autonomes Fahren (*)

umfasst ein Anwendungsfeld der KI mit steigender Relevanz. Autonome Fahrzeuge sind in der Lage, mithilfe diverser Sensoren, wie Kamera, Radar, Lidar und GPS, selbstständig die Umgebung und den Verkehrsfluss zu detektieren. Auf Basis dieser Informationen sollen Fahrzeuge in Zukunft zuverlässig und ohne Kontrolle durch menschliche Fahrer sicher durch sämtliche im Straßenverkehr auftretenden Situationen navigieren können.

Big Data (*)

bezeichnet eine Datenmenge, die oftmals zu groß ist, um sie mit herkömmlichen Datenverarbeitungsmethoden auszuwerten. Big Data bietet die Grundlage für viele KI-Anwendungen, da die zugrunde liegenden Algorithmen mit einer größeren Menge von Daten besser lernen und so Vorhersagen und Analysen verbessern können.

Deep Learning (*)

bezeichnet eine Methode des maschinellen Lernens und beschreibt eine Klasse von künstlichen neuronalen Netzen, die eine große Anzahl an Zwischenlagen (hidden layers) aufweisen und von der Funktionsweise eines menschlichen Gehirns inspiriert sind. Dabei wird eine große Datenmenge verwendet.

Fintech (*)

setzt sich aus den Begriffen Financial Services und Technology zusammen und umschreibt neue Geschäftsprozesse und Technologien, welche im Finanzsektor implementiert werden. Fintech-Unternehmen, häufig Start-ups, versuchen, traditionelle Tätigkeiten der Finanzwirtschaft mit Technologie zu verbessern oder sogar zu ersetzen.

Industrie 4.0

steht hauptsächlich für die Digitalisierung des produzierenden Gewerbes. Im Vordergrund stehen dabei die intelligente Vernetzung von Produktionsprozessen und Wertschöpfungsketten mithilfe von Informations- und Kommunikationstechnologien.

Infrastructure-as-a-Service

steht für das Anbieten von (Rechen-)Infrastrukturen mit einem Cloud-Geschäftsmodell. Hierbei kann sich der Kunde/die Kundin die gewünschte Recheninfrastruktur aussuchen, welche dann von den Anbietenden zur Verfügung gestellt wird.

Künstliche Intelligenz (*)

ist eine Kombination von Technologien, die es Maschinen ermöglichen, gewisse Probleme selbstständig durch lernende Algorithmen immer besser zu lösen. Unterschieden wird dabei in starke und schwache KI: Bei einer starken KI werden menschliche Vorgänge im Gehirn imitiert, bei einer schwachen KI werden ausschließlich speziell abgegrenzte Probleme gelöst.

KI-Recheninfrastrukturen

umfassen alle Aspekte der notwendigen Hardware für die Erforschung, Entwicklung und Nutzung von Künstlicher Intelligenz. Bei der Realisierung von tragfähigen Lösungen im Infrastrukturbereich sind darüber hinaus Aspekte der Softwareversorgung mitzudenken.

Maschinelles Lernen / Machine Learning (ML) (*)

basiert auf Algorithmen, die lernen, Muster und Regelmäßigkeiten in Datensätzen zu erkennen und daraus Vorhersagen und Regeln abzuleiten.

Simulation

wird verstanden als computergenerierter Entwurf eines Realitätsmodells, um dessen Verhalten zu studieren.

In diesem Anhang werden potenzielle Technologie Kandidaten in Ergänzung zu Kapitel 4.3 vorgestellt.

GPUs – Grafikprozessoren

Ursprünglich hatten Grafikkarten nur eine Funktion, nämlich die Darstellung der grafischen Information. Daher verfügten sie über keine umfangreiche Rechenleistung. Mit der Entwicklung von 3D-Grafik in den 1990er Jahren wurden sowohl Leistung als auch Videospeicher erweitert, um die Darstellung auch anpassen zu können. Damit startete bereits die Entwicklung von Grafikkarten zu Graphics Processing Units – GPUs. Im Jahr 2007 stellte die Firma NVIDIA CUDA (Compute Unified Device Architecture) eine entsprechende Programmierschnittstelle vor, dadurch konnten GPUs ab sofort als Co-Prozessoren eingesetzt werden. Der Einsatz von GPUs für nicht grafische Anwendungen wird als „general-purpose computing on graphics processing units“ (GPGPU) bezeichnet bzw. dazu fähige GPUs als „general-purpose GPUs“. Im Gegensatz zu CPUs sind GPUs für die Grafikbearbeitung optimiert, sie unterstützen Vektor- und Matrixoperationen und haben eine viel höhere Anzahl von Rechenkernen. Beispielsweise hat die NVIDIA Tesla V100 GPU 5.120 Rechenkern, während aktuell reine CPUs zwischen 12 und 128 Rechenkern umfassen.

GPUs sind inzwischen sehr populär für Deep Learning geworden, weil die mathematischen Grundlagen der neuronalen Netze Ähnlichkeiten mit der Grafikbearbeitung aufweisen und die einfache Parallelisierung der Aufgaben mit Matrizen möglich ist.

GPUs werden von verschiedenen Firmen angeboten:

NVIDIA

Die NVIDIA Corporation ist aktuell der größte Hersteller von dedizierten GPUs. Im zweiten Quartal 2020 stammten 80 Prozent der GPUs auf dem Markt von NVIDIA. Von den in der [TOP500-Liste¹⁰³](#) dargestellten Supercomputern verfügen 144 über GPGPUs oder ähnliche Chips, 140 der Großrechner basieren auf Grafikkarten von NVIDIA.

Relevante Produktreihen sind:

- GeForce und Quadro – Consumer Cards: Diese sind weniger performant als Servervarianten. Die Verwendung von GeForce GPUs in Servern ist durch die NVIDIA-Lizenz untersagt, einige Modelle (wie RTX 3090) in der Produktreihe basieren aber auf der gleichen Architektur wie die Serverkarten.
- Tesla – GPGPUs, die für den Einsatz in Servern bzw. Datacentern gedacht sind. [NVIDIA gab inzwischen den Markennamen Tesla auf, um mögliche Verwechslungen mit dem Autohersteller zu vermeiden¹⁰⁴](#), weitere Produkte aus dieser Reihe (wie NVIDIA A100) werden ohne das Tesla-Label veröffentlicht.
- Tegra – System-on-Chip-Reihe, die für mobile Systeme und IoT-Lösungen gedacht ist, und wenig Energie verlangt. Hier sind CPU, GPU und RAM in einem Chip verbaut, bei den CPU-Kernen handelt es sich um ARM-Kerne.

Für [Serverhersteller¹⁰⁵](#) bietet NVIDIA vorgefertigte Plattformen aus miteinander verbundenen GPUs an, die der Hersteller mit nötigen Serverkomponenten (wie CPU, RAM) ergänzen kann.

NVIDIA hat auch eigene Server im Angebot: DGX-1 und DGX-2 mit jeweils 8 oder 16 P100/V100 interconnected GPUs pro Server und Intel Xeon Platinum CPUs; sowie den DGX A100 mit 8 A100 interconnected GPUs pro Server und AMD Rome GPUs.

Im Jahr 2020 annoncierte NVIDIA die EGX-A100-Karten mit integriertem Netzwerkadapter ConnectX-6 Dx SmartNICs der Firma Mellanox. Vorgehene Anwendungsbereiche sind: KI für IoT und auch Entlastung der CPUs durch Übernahme der Netzwerk- und Verschlüsselungsoperationen.

Im September 2020 verkündete NVIDIA den Kauf von ARM Ltd., einer Firma, die Blaupausen von neuen, oft energieeffizienten Prozessoren entwickelt. Diese können dann von Prozessorherstellern lizenziert und selbst weiterentwickelt werden. Bislang wurden ARM CPUs oft auf Smartphones installiert, inzwischen aber auch auf Rechenservern. Der schnellste Supercomputer der Welt (laut Rangliste [TOP500¹⁰⁶](#) vom November 2020 [Fugaku, ein japanisches Rechensystem](#)) [basiert auf dem ARM-Prozessordesign. Der Kauf wird zum Zeitpunkt dieser Studie aber noch durch Behörden überprüft^{107/108}](#).

AMD

Advanced Micro Devices (AMD) ist der zweitgrößte Hersteller von dedizierten GPUs, außerdem produziert AMD auch CPUs. Auf dem Servermarkt sind AMD GPUs bislang schlechter positioniert als der Konkurrent NVIDIA, nur ein Supercomputer aus der aktuellen TOP500-Liste hatte AMD-Grafikkarten.

Allerdings geraten die neuen Generationen der AMD GPUs aktuell wieder mehr in den Fokus der Nutzerinnen und Nutzer von großen Recheninfrastrukturen (auch für KI), da bereits bei der neuen GPU Instinct MI100, die im November 2020 präsentiert wurde, eine Konkurrenzfähigkeit mit NVIDIA sichtbar wird. Der neue AMD-Prozessor ist der NVIDIA-A100-Karte auch in Punkto Preis und Energiebedarf überlegen und schlägt die NVIDIA A100 mit mehr Rechenleistung, kleinerem Energiebedarf und günstigerem Preis. Die **neue Generation von AMD GPUs verbessert die Unterstützung für Matrixoperationen (als Antwort auf NVIDIA Tensor Cores), was die Karte attraktiver für KI-Einsätze machen kann**¹⁰⁹.

NEC Vektortechnologie – Aurora

Die NEC Corporation ist ein japanischer Elektronikkonzern, der in den Bereichen Serversysteme, Supercomputer, ASICs (application-specific integrated circuit – anwendungsspezifische integrierte Schaltung) und auch Telekommunikation agiert. NEC hat eine lange Historie in Vektorrechner-Systemen. So war der Vektorcomputer Earth Simulator nach seiner Inbetriebnahme 2002 über zwei Jahre lang der schnellste Supercomputer der Welt. Aktuell sind zehn Supercomputer von NEC in der TOP500-Liste zu finden, zwei davon haben die eigenen NEC Vektorprozessoren als Beschleuniger integriert.

Das jüngste Modell von Systemen auf Basis der Vektortechnologie – NEC SX-Aurora TSUBASA – wird zusammen mit einem CPU-basierten Vektorhost und der TSUBASA-Karte als Vector Engine (VE) gebaut. Diese Kombination von Technologien erfordert eine andere Programmierung als gängige GPU-basierte Systeme, ist aber ebenfalls für die Nutzung als KI-System geeignet.

Prozessoren

Klassische CPUs (Central Processing Unit oder allgemein: Prozessor) sind eine Hauptkomponente fast aller Rechensysteme, die nach einem vorprogrammierten Algorithmus das System steuern und die Datenverarbeitung ausführen.

Moderne Prozessoren bestehen aktuell meist aus mehreren Prozessorkernen, die unabhängig voneinander funktionieren und somit Problemlösungen parallelisieren können.

Während neuronale Netze von der Leistungsfähigkeit der GPUs profitieren, laufen andere Machine-Learning- und Big-Data-Algorithmen immer noch effizient auf CPUs.

Zur am weitesten verbreiteten Familie von CPUs gehören x86-64 und ARM. Darüber hinaus sind auf manchen Server- und HPC-Plattformen POWER-Prozessoren verarbeitet, eine Eigenentwicklung von IBM.

KI- und Big-Data-Anwendungen sind gut parallelisierbar, weswegen sich dafür Architekturen mit vielen Prozessorkernen eignen. Einige Big-Data-Anwendungen (z. B. Apache Hadoop, Apache Spark Frameworks) sehen auch die Möglichkeit der Ausführung auf mehreren, über das Netzwerk verbundenen „schwachen“ Rechnern vor. Auch für sie ist die hohe Kernanzahl in einem Server von Vorteil.

Die x86-64-Architektur dominiert inzwischen im Serverbereich. x86-64-Prozessoren sind in der Regel in der Performance optimiert, Servermodelle verfügen über einige Dutzend Rechenkerne (die AMD-EPYC-ROME-Serie hat beispielsweise bis zu 64 Rechenkerne). Prozessoren dieser Architektur werden hauptsächlich von Intel und AMD hergestellt. Auch sind momentan

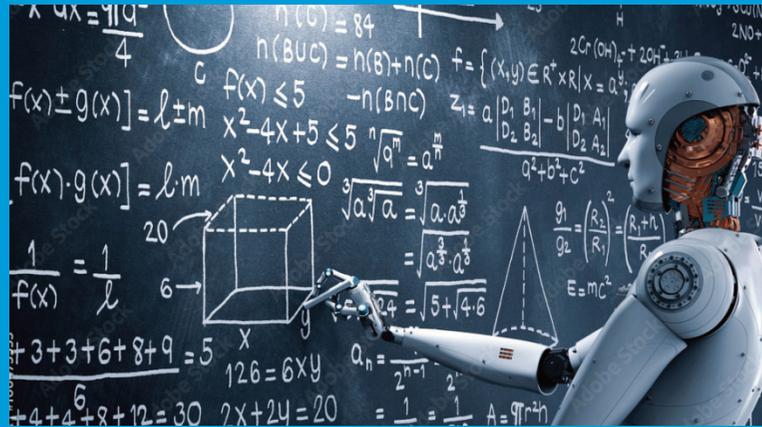
noch die meisten Anwendungen und Softwarepakete für x86-64 implementiert und optimiert.

Auf mobilen Plattformen sind ARM-Prozessoren aktuell marktführend, zunehmend werden diese aber auch als Serverprozessoren eingesetzt. Der schnellste Supercomputer der Welt (November 2020) ist ARM-basiert (Fujitsu A64FX). Die Architektur ist energieoptimiert, da viele ARM-Modelle sich auf mobile Plattformen fokussieren. Die ARM-Architektur wird von ARM Ltd. entwickelt und an andere Hersteller lizenziert. ARM-Prozessoren werden oft in der Ausführung System-on-a-Chip (SoC) hergestellt, bei der CPU-Kerne, Speicher, Schnittstellen für Hardwareanschlüsse und oft auch GPUs auf einem Chip integriert sind. Diese Bauform ist für mobile Geräte zweckmäßig, da ein Prozessorchip in diesem Fall schon die meisten Systemkomponenten beinhaltet.

POWER-Prozessoren haben eine ähnliche Architekturgrundlage wie die ARM-Prozessoren (RISC-Architektur), der Annahme folgend, dass bessere Performance dank einfacheren Prozessoroperationen erreicht werden kann. Die POWER-Prozessoren sind für HPC optimiert. So unterstützen POWER9 CPUs z. B. die schnelle Verbindung mit NVIDIA GPUs (über die NVLink-Technologie). Da die aktuell zweit- und drittgrößten Supercomputer (Summit und Sierra) auf Basis der POWER9-Prozessortechnologie gebaut sind, ist die Plattformunterstützung seitens KI-Bibliotheken und Frameworks in den letzten Jahren massiv verbessert worden.

ASICS

ASICS¹¹⁰ (application-specific integrated circuits) sind integrierte Schaltkreise, die zur Lösung konkreter Aufgaben projektiert sind. Relevant für KI-Bereiche sind ASICs, die Matrixoperationen effizienter als GPUs beschleunigen oder die solche Beschleunigung auf mobile Plattformen bringen. Beispiele dafür sind die TPU (Tensor Processing Unit) von Google, die seit 2017 auch im Rahmen der Google Cloud angeboten wird, oder der Beschleuniger AWS Inferentia von Amazon, der in der Amazon Cloud angeboten wird.



Verweise mit Web-Links zu weiteren Informationen

Stand: Februar 2021

- 1 Cognitive and artificial intelligence (AI) systems market revenue worldwide from 2015 to 2024; <https://www.statista.com/statistics/694638/worldwide-cognitive-and-artificial-intelligence-revenues/>
- 2 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, September 2020; <https://www.gartner.com/smarterwithgartner/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020/>
- 3 Pfannes, P. u. A. (2018). Hessens Ambitionen für Künstliche Intelligenz: Ein Beitrag zur nationalen KI-Strategie am Beispiel des Finanzsektors. Accenture. Auftraggeber: Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung.
- 4 Buxmann, P (2020). KI – Künstliche Intelligenz: Wirtschaft und Innovation. Herausgeber: Hessische Staatskanzlei.
- 5 Webseite des Hochschulrechenzentrums der Technischen Universität Darmstadt; https://www.hrztu-darmstadt.de/hrz_aktuelles/index.de.jsp
- 6 Webseite der Initiative Nationales Hochleistungsrechnen (NHR); <https://www.nhr-gs.de/nationales-hochleistungsrechnen>
- 7 The Top500 list; <https://top500.org/>
- 8 Infrastruktur der Bioinformatik der Technischen Hochschule Mittelhessen; <https://www.thm.de/mni/eric-hartmann/infrastruktur-der-bioinformatik>
- 9 NVIDIA DGX-1; <https://www.nvidia.com/de-de/data-center/dgx-1/>
- 10 Webseite des Hessischen Kompetenzzentrums für Hochleistungsrechnen – HKHLR; <https://www.hkhlr.de/de>

-
- 11 Webseite des Kompetenzzentrums für Arbeit und Künstliche Intelligenz im Rhein-Main-Gebiet; <https://kompaki.de/>
 - 12 Kooperative KI für die Arbeitswelt von morgen; https://www.tu-darmstadt.de/universitaet/aktuelles_meldungen/archiv_2/2020/2020quartal4/news_archiv_de_274176.de.jsp
 - 13 Außeruniversitäre Forschungseinrichtungen; <https://wissenschaft.hessen.de/forschen/ausseruniversitaere-forschungseinrichtungen>
 - 14 Webseite des Deutschen Wetterdienstes; <https://www.dwd.de>
 - 15 Webseite der Gesellschaft für Schwerionenforschung; <https://www.gis.de>
 - 16 Webseite des Projektes Fair; <https://fair-center.eu/>
 - 17 Webseite von GES Systemhaus; <https://www.ges-systemhaus.de/>
 - 18 Continental beschleunigt Entwicklung mit Supercomputer, CIO Online, Juli 2020. <https://www.cio.de/a/continental-beschleunigt-entwicklung-mit-supercomputer,3639661>
 - 19 Webseite von ADAMOS - <https://www.adamos.com/>
 - 20 Verbund von Maschinenbaufirmen bietet Zulieferern & Herstellern eigene Plattform an; <https://www.hessenmetall.de/themen/kuenstliche-intelligenz.html>
 - 21 Webseite der Trusted-Cloud-Aktivität; <https://www.trusted-cloud.de/>
 - 22 Aixit Cloud; <https://www.trusted-cloud.de/de/cloudservices/2574/aixit-cloud>; Webseite der Aixit GmbH; <https://www.aixit.com/de/index.php>
 - 23 Webseite der Servinga GmbH; <https://www.servinga.com>; Servinga Managed Cloud - <https://www.trusted-cloud.de/de/cloudservices/2111/managed-cloud>
 - 24 Webseite der Open Telekom Cloud; <https://open-telekom-cloud.com/de>; Open Telekom Cloud – Trusted Cloud; <https://www.trusted-cloud.de/de/cloud-services/2667/open-telekom-cloud-based-on-openstack->
 - 25 Google Cloud Platform Standorte; <https://cloud.google.com/about/locations?hl=de#europe>
 - 26 Google baut in Hessen; https://www.journal-frankfurt.de/journal_news/Wirtschaft-7/Neue-Rechenzentren-im-Rhein-Main-Gebiet-Google-baut-in-Hessen-36780.html
 - 27 Webseite der ONCITE – All In One EDGE; <https://oncite.io/oncite>
 - 28 Webseite der IoTOS GmbH; <https://iotos.io/>
 - 29 Webseite der Bosch Connected Industry; <https://www.bosch-connected-industry.com/de/de/index>
 - 30 Viertes IT Forum Hessenmetall; <https://www.hessenmetall.de/newsroom/itforen/4-it-forum.html>
 - 31 ATHENE: größtes Forschungszentrum für Cybersicherheit und Privatsphärenschutz in Europa; <https://www.athene-center.de/>
 - 32 KI wird in Unternehmen viel seltener genutzt als gedacht; <https://www.bitkom.org/Presse/Presseinformation/KI-wird-in-Unternehmen-viel-seltener-genutzt-als-gedacht>
 - 34 Webseite des Gauss Centre for Supercomputing; <https://www.gauss-centre.eu>

-
- 33 Keller, Lorenz – Studie; http://digdok.bib.thm.de/volltexte/2020/5352/pdf/THM_Hochschulschriften_12_Endfassung.pdf
- 35 Webseite des Jülich Supercomputing Centre; https://www.fz-juelich.de/ias/jsc/EN/Home/home_node.html
- 36 Webseite des Leibniz-Rechenzentrums der Bayerischen Akademie der Wissenschaften; <https://www.lrz.de/>
- 37 Webseite des Höchstleistungsrechenzentrums Stuttgart; <https://www.hlrs.de/>
- 38 Webseite der NHR Geschäftsstelle für Nationales Hochleistungsrechnen; <https://www.nhr-gs.de/>
- 39 Webseite der Gauß-Allianz; www.gauss-allianz.de
- 40 GAIA-X, Offizielle Webseite des BMWI; <https://www.bmwi.de/Redaktion/DE/Dossier/gaia-x.html>
- 41 Was bringen Datenökosysteme wie GAIA-X für den Mittelstand?; <https://blog.iao.fraunhofer.de/was-bringen-datenoesysteme-wie-gaia-x-fuer-den-mittelstand/>
- 42 Financial Big Data Cluster; <https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/GAIA-X-Use-Cases/financial-big-data-cluster-fbdc.html>
- 43 SafeFBDC - https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/KuenstlicheIntelligenzProjekte/KuenstlicheIntelligenzProjekte_ZweiterFoerderaufruf/ki-projet_SafeFBDC.html
- 44 Webseite de:hub; <https://www.de-hub.de/>

-
- 45 <https://digitalhub-ai.de/>
- 46 European Digital Innovation Hubs in Digital Europe Programme; <https://ec.europa.eu/digital-single-market/en/european-digital-innovation-hubs-digital-europe-programme-0>
- 47 Nationale Strategie für Künstliche Intelligenz; <https://www.ki-strategie-deutschland.de/home.html>
- 48 KI-Landkarte der Plattform Lernende Systeme; <https://www.plattform-lernende-systeme.de/ki-landkarte.html?SUD=1>
- 49 Innovationsstrategie Baden-Württemberg; https://www.baden-wuerttemberg.de/fileadmin/redaktion/dateien/PDF/200204_Innovationsstrategie_BW_Fortschreibung_2020.pdf
- 50 Cyber Valley – Europas größtes KI-Cluster; <https://cyber-valley.de/de>
- 51 Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS2 DM); https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/pdf/Forschung/Umsetzungskonzept_HPC_DIC_LS2DM_1.0.pdf
- 52 Hightech-Agenda Bayern; https://www.bayern.de/wp-content/uploads/2019/10/hightech_agenda_bayern.pdf
- 53 InnoBB 2025 – Gemeinsame Innovationsstrategie der Länder Berlin und Brandenburg; <https://innobb.de/de/innobb-2025-eine-neue-strategie-fuer-neue-zeiten>

-
- 54 Zukunftsstrategie Digitales Brandenburg, Beschluss der Landesregierung vom 11.12.2018; <https://digitalesbb.de/wp-content/uploads/2018/12/181211-Strategiepapier.pdf>
- 55 BREMEN.KI – Strategie für Künstliche Intelligenz; <https://www.senatspressestelle.bremen.de/detail.php?gsid=bremen146.c.346055.de&asl=bremen02.c.730.de>
- 56 Digitalstrategie für Hamburg;
<https://www.hamburg.de/contentblob/13508768/703cff94b7cc86a2a12815e52835accf/data/download-digitalstrategie-2020.pdf>
- 57 Masterplan Digitalisierung des Bundeslandes Niedersachsen;
https://www.mw.niedersachsen.de/download/135324/Masterplan_Digitalisierung_Niedersachsen.pdf
- 58 Zentrum für digitale Innovation Niedersachsen; <https://www.zdin.de/>
- 59 Strategie für das digitale Nordrhein-Westfalen;
https://www.wirtschaft.nrw/sites/default/files/asset/document/digitalstrategie_nrw_endfassung_final.pdf
- 60 Kompetenzplattform KI.NRW; <https://www.ki.nrw/>
- 61 Innovationsstrategie des Saarlands; https://www.saarland.de/DE/medien-informationen/medienservice/pressearchiv/stk/stk-medieninfo-archive/2021/Q4_2021/pm_2021-12-14-innovationsstrategie.html
- 62 Sachsen Digital;
<https://www.digitales.sachsen.de/die-strategie-sachsen-digital-3916.html>
- 63 Center for Scalable Data Analytics and Artificial Intelligence – ScaDS;
<https://www.scads.de/de/>
- 64 Digitale Agenda für das Land Sachsen-Anhalt; https://digital.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/StK/Digital/DigitaleAgenda_Sachsen-Anhalt_Lesefassung.pdf
- 65 KI-Strategie Schleswig-Holstein; https://www.schleswig-holstein.de/DE/Landesregierung/Themen/Digitalisierung/Digitalisierung/KI_Strategie/_documents/ki_strategie_download.pdf?__blob=publicationFile&v=1
- 66 Thüringer Strategie für die Digitale Gesellschaft;
https://wirtschaft.thueringen.de/fileadmin/user_upload/Publikationen/Pub-Thueringer_Strategie_fuer_die_Digitale_Gesellschaft.pdf
- 67 Monitoring-Report Wirtschaft Digital 2017: Hessen, Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung, März 2018.
- 68 Erik Brynjolfsson und Andrew McAfee (2014), The second Machine age.
- 69 Unter anderem CATALYST
<https://www.hlrs.de/about-us/research/current-projects/catalyst/>
- 70 Die Autoren kooperieren seit 25 Jahren mit Firmen im Bereich der Simulation, Datenanalyse und künstlichen Intelligenz.
- 71 Der Begriff des Mittelständlers wird hier definiert nach dem Institut für Mittelstandsforschung <https://www.ifm-bonn.org/definition> „Der Begriff ‚Mittelstand‘ stellt eine Besonderheit im deutschsprachigen Raum dar und wird definiert durch die Einheit von Eigentum und Leitung. Für die Zugehörigkeit eines Unternehmens zum Mittelstand ist also nicht dessen Größe ausschlaggebend. Vielmehr sind es dessen qualitative Merkmale.“

-
- 72 Pressemitteilung des Leibniz-Rechenzentrums – „Kommende Technologien schon heute erproben“; https://www.lrz.de/presse/ereignisse/2020-11-19_Kommende-Technologien-schon-heute-erproben/
- 73 Webseite des Gauss Centre for Supercomputing; <https://www.gauss-centre.eu/>
- 74 Webseite der Gauß-Allianz; <https://gauss-allianz.de/>
- 75 <https://datacenterfrontier.com/rolling-zettabytes-quantifying-the-data-impact-of-connected-cars/>
- 76 https://de.wikipedia.org/wiki/Secure_Shell
- 77 Blick auf die Entwicklung des AI-as-a-Service-Marktsegments; <https://tiorientering.ch/global-artificial-intelligence-ai-as-a-service-market-2022-to-2028-by-top-major-players-google-amazon-web-services-ibm-microsoft-sap-salesforce-intel-baidu-fico-sas-bigml-etc/>
- 78 ISO/IEC 27001 Zertifizierung; https://de.wikipedia.org/wiki/ISO/IEC_27001
- 79 TISAX – Trusted Information Security Assessment Exchange; <https://www.enx.com/en-US/tisax/>
- 80 VDA ISA-Katalog (Information Security Assessment); <https://www.vda.de/de/themen/digitalisierung/daten/informationssicherheit>
- 81 KPMG AG, Bitkom Research GmbH „Cloud-Monitor 2020: Gut eingerichtet in der Cloud“; <https://hub.kpmg.de/studie-cloud-monitor-2020>
- 82 H.R.4943 – CLOUD Act; <https://www.congress.gov/bill/115th-congress/house-bill/4943>

-
- 83 Nucleus Research, Inc. „Guidebook Deep Learning on AWS“; <https://d1.awsstatic.com/whitepapers/Deep%20learning%20on%20AWS.pdf>
- 84 Webseite von AI Bridging Cloud Infrastructure - <https://abci.ai/>
- 85 Webseite von AIST; https://www.aist.go.jp/index_en.html
- 86 Webseite der HWW-GmbH; <https://www.hww.de>
- 87 Webseite des INCITE-Programms; <https://www.doeleadershipcomputing.org/>
- 88 Der Begriff Knotenstunde bezeichnet die Nutzung eines Rechnerknotens für die Dauer einer Stunde. Er wird für die Vergabe von Kontingenten auf Rechnern genutzt. Die Zahl der benötigten Knotenstunden berechnet sich aus der Anzahl der genutzten Knoten und der Dauer der Nutzung. Werden 64 Knoten für 8 Stunden benötigt, so ergibt das einen Bedarf von 512 Knotenstunden.
- 89 Webseite des EuroHPC-Joint Undertaking; <https://eurohpc-ju.europa.eu/>
- 90 PetaSC: a new EuroHPC world-class supercomputer in Bulgaria; <https://eurohpc-ju.europa.eu/press-release/discoverer-powers-bulgarian-eurohpc-supercomputer-inaugurated>
- 91 Hewlett Packard Enterprise selected to build Czech Republic’s most powerful supercomputer to advance R&D in science and engineering; <https://www.hpe.com/us/en/newsroom/press-release/2020/10/hewlett-packard-enterprise-selected-to-build-czech-republics-most-powerful-supercomputer-to-advance-rd-in-science-and-engineering.html>
- 92 MeluXina: a new EuroHPC world-class supercomputer in Luxembourg; <https://eurohpc-ju.europa.eu/press-release/meluxina-new-eurohpc-world-class-supercomputer-luxembourg>

-
- 93 Deucalion; <https://eurohpc-ju.europa.eu/press-release/deucalion-new-eurohpc-world-class-green-supercomputer-portugal>
- 94 Atos' Bull Sequana XH2000 chosen for the EuroHPC peta-scale system in Slovenia; https://www.izum.si/doc/Vega-izbor_en.pdf
- 95 EU-Kommission gibt acht Standorte für erste europäische Supercomputer bekannt; https://ec.europa.eu/germany/news/20190607eu-kommission-gibt-acht-standorte-fuer-erste-europaeische-supercomputer-bekannt_de
- 96 MareNostrum 5: Europäischer 200-Petaflops-Supercomputer startet Ende 2020; <https://www.zdnet.de/88362181/marenostrum-5-europaeischer-200-petaflops-supercomputer-startet-ende-2020/>
- 97 LEONARDO: a new EuroHPC world-class pre-exascale supercomputer in Italy; <https://eurohpc-ju.europa.eu/press-release/leonardo-new-eurohpc-world-class-pre-exascale-supercomputer-italy>
- 98 LUMI; <https://www.lumi-supercomputer.eu/>
- 99 Webseite der Einrichtung „Centre de Calcul Recherche et Technologie“ <http://www-ccrt.cea.fr/>
- 100 Webseite der Landescloud; <https://www.landes.cloud/>
- 101 Monitoring-Report Wirtschaft Digital 2017: Hessen, Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung, März 2018.
- 102 Webseite des Disträdl-Förderprogramms; <https://digitales.hessen.de/digitale-zukunft/distral-foerderprogramm>
- 103 TOP500; <https://www.top500.org/>
- 104 <https://www.heise.de/newsticker/meldung/GPU-Beschleuniger-Nvidia-gibt-Markennamen-Tesla-auf-4723754.html>
- 105 <https://images.nvidia.com/content/pdf/gpu-accelerated-server-platforms.pdf>
- 106 <https://www.top500.org/system/179807/>
- 107 <https://www.businessinsider.de/politik/deutschland/politisch-brisanter-deal-bsi-ueberprueft-verkauf-von-mikroprozessor-firma-an-nvidia/>
- 108 <https://www.gov.uk/government/news/cma-to-investigate-nvidia-s-takeover-of-arm>
- 109 <https://www.heise.de/news/Server-Beschleuniger-AMD-Instinct-MI100-Ohne-Radeon-aber-mit-11-5-FP64-TFlops-4960973.html>
- 110 <https://cloud.google.com/tpu/>

Impressum

Autoren

Prof. Dr. Michael M. Resch
Dr.-Ing. Bastian Koller
Dipl.-Inf. Oleksandr Shcherbakov

Herausgeber

Hessische Staatskanzlei, Ministerin für Digitale Strategie und Entwicklung
Georg-August-Zinn-Straße 1, 65183 Wiesbaden
www.digitales.hessen.de

Der Herausgeber übernimmt keine Gewähr für die Richtigkeit, die Genauigkeit und die Vollständigkeit der Angaben sowie für die Beachtung privater Rechte Dritter. Die in der Veröffentlichung geäußerten Ansichten und Meinungen müssen nicht mit der Meinung des Herausgebers übereinstimmen.

© Hessische Staatskanzlei,
Ministerin für Digitale Strategie und Entwicklung

Vervielfältigung und Nachdruck – auch auszugsweise – nur nach vorheriger schriftlicher Genehmigung.

Ausschluss Wahlwerbung: Dieses Dokument wird im Rahmen der Öffentlichkeitsarbeit der Hessischen Landesregierung herausgegeben. Sie darf weder von Parteien noch von Wahlbewerberinnen und -bewerbern oder Wahlhelferinnen und -helfern während eines Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für Landtags-, Bundestags-, Europa- und Kommunalwahlen. Missbräuchlich sind insbesondere eine Verteilung dieser Druckschrift auf Wahlveranstaltungen oder an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl darf das Dokument nicht in einer Weise verwendet werden, die als Parteinahme der Landesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte. Die genannten Beschränkungen gelten unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Druckschrift dem Empfänger zugegangen ist. Den Parteien ist es jedoch gestattet, das Dokument zur Unterrichtung ihrer eigenen Mitglieder zu verwenden.

Projektträger

Hessen Trade & Invest GmbH im Auftrag der Hessischen Staatskanzlei
Ministerin für Digitale Strategie und Entwicklung
Konradinallee 9, 65189 Wiesbaden

Kontaktdaten

Hessen Trade & Invest GmbH
Konradinallee 9, 65189 Wiesbaden
Telefon +49 611 95017-80 /-85 , Fax +49 611 95017-8466
info@htai.de
www.htai.de

Lektorat

Uta Marini, Warstein

Gestaltung

Theißen-Design, Lohfelden

Abbildungen

stock.adobe.com: icafreitas (Titelseite); Science RF (S.2); arthead (S.4); Scanrail (S.6); greenbutterfly (S.18); sitthiphong (S.23); agsandrew (S.32); denisismagilov (S.36); phonlamaiphoto (S.48)

Stand: Februar 2021