

Sachbericht zum Verwendungsnachweis

Teil I: Kurzbericht

Hochschule Magdeburg-Stendal Breitscheidstr. 2 39114 Magdeburg	Förderkennzeichen: 13FH029KI2
Vorhabenbezeichnung: KI-Nachwuchs@FH 2-2021: KI-Infrastruktur für Forschung, Technologie und Transformation in Sachsen-Anhalt (KITT)	
Laufzeit des Vorhabens: 01.04.2023 bis 30.09.2024	

Ausgangslage und Zielsetzung

Mit dem Vorhaben *KITT* reagierte die Hochschule Magdeburg-Stendal auf den wachsenden Bedarf an leistungsfähiger Recheninfrastruktur für KI-Forschung und -Lehre. Ausgangspunkt war eine heterogene IT-Landschaft ohne zentral verfügbare GPU-Ressourcen. Ziel war der Aufbau einer hochverfügbaren, nutzerfreundlichen und skalierbaren KI-Plattform, um datenintensive Anwendungen und didaktische Szenarien zu ermöglichen. Die Maßnahme adressierte sowohl den technischen Aufbau eines zentralen KI-Clusters als auch dessen nachhaltige Integration in Forschung, Lehre und Hochschulorganisation.

Ablauf und Umsetzung

Die technische Basis bildet ein NVIDIA DGX H100 System, eingebunden in ein Kubernetes/Rancher-Cluster mit run:ai zur GPU-Ressourcensteuerung. Trotz pandemie- und ausschreibungsbedingter Verzögerungen konnte das System vollständig integriert und in Betrieb genommen werden. Die Plattform wurde über Single-Sign-On, VPN und Moodle systemseitig angeschlossen und iterativ mit zentralen Diensten wie JupyterHub, MATLAB-Containern und einem hochschuleigenen LLM-basierten Chat-System („h2 Chat“) ausgestattet.

Wesentliche Ergebnisse

- **Technische Infrastruktur:** Aufbau einer robusten, containerisierten KI-Plattform mit intelligenter GPU-Verwaltung über run:ai und kontinuierlicher Integration in hochschulinterne Dienste.
- **Lehre:** Breite Nutzung in Studiengängen (u. a. AI.Engineering, IT and Data Science), Bereitstellung von GPU-Umgebungen über Moodle/JupyterHub, Integration großer Sprachmodelle in didaktische Szenarien.
- **Forschung:** Unterstützung von Projekten wie *SynerKI* (LLM-gestützte Vereinfachung von Fachtexten), Pilotprojekt Auswertung von Luftbilddaten aus der Fernerkundung mittels U-Net-basierten Künstlichen Neuronalen Netzen,

Überflutungsprognose mittels Transformer-Netzen sowie verschiedenen studentischen Forschungsprojekten; Plattform als technische Basis in Drittmittelanträgen.

- **Governance:** Einführung eines abgestuften Nutzerrollen- und Ressourcenmanagements; Etablierung von Support-Strukturen und hochschulöffentlicher Dokumentation.
- **Verstetigung:** Verstärkung durch Multiplikator:innenstrukturen, hochschuldidaktische Schulungen und Einbindung in Drittmittelprojekte; fortlaufende Öffnung für neue Use Cases durch containerisierte Architektur.

Zusammenarbeit

Die Umsetzung erfolgte in enger Abstimmung mit internen Projekten wie *ZAKKI* (didaktische Integration) und *SynerKI* (wirtschaftsnahe Anwendung). Kooperationen mit Fachbereichen wie WUBS (Umwelt- und Bauwesen) sowie IWID (Ingenieurwissenschaften und Informatik) ermöglichten eine transdisziplinäre Verankerung. Mehrere Drittmittelanträge greifen KITT explizit als technologische Grundlage auf.

Fazit

KITT hat eine zukunftssichere Infrastruktur für KI-Forschung und -Lehre geschaffen, die bereits während der Projektlaufzeit intensiv genutzt und didaktisch wie organisatorisch verankert wurde. Die Plattform gilt nun als zentraler operativer Kern für KI-Anwendungen an der Hochschule Magdeburg-Stendal. Ihre modulare Erweiterbarkeit, strategische Anschlussfähigkeit und Einbettung in Governance-Strukturen sichern den nachhaltigen Projekterfolg.

Sachbericht zum Verwendungsnachweis

Teil II: Eingehende Darstellung

Hochschule Magdeburg-Stendal Breitscheidstr. 2 39114 Magdeburg	Förderkennzeichen: 13FH029KI2
Vorhabenbezeichnung: KI-Nachwuchs@FH 2-2021: KI-Infrastruktur für Forschung, Technologie und Transformation in Sachsen-Anhalt (KITT)	
Laufzeit des Vorhabens: 01.04.2023 bis 30.09.2024	

1. Ausgangslage und Zielsetzung

Die rasanten Fortschritte im Bereich der Künstlichen Intelligenz (KI) sowie die zunehmende Bedeutung datengetriebener Prozesse in Wissenschaft, Wirtschaft und Gesellschaft stellen Hochschulen vor neue Herausforderungen: Um Studierenden eine zeitgemäße Ausbildung und Forschenden leistungsfähige Werkzeuge zur Verfügung zu stellen, ist der Aufbau moderner KI-Infrastrukturen von zentraler Bedeutung.

Vor diesem Hintergrund hatte das Projekt **KITT – KI-Infrastruktur für Forschung, Technologie und Transformation in Sachsen-Anhalt** das Ziel, an der Hochschule Magdeburg-Stendal (h2) eine zentrale, hochverfügbare und effizient nutzbare KI-Infrastruktur zu etablieren. Diese sollte nicht nur als technologische Grundlage für aktuelle und künftige Forschungsprojekte dienen, sondern insbesondere auch in die Lehre integriert und für Studierende sowie Lehrende nutzbar gemacht werden.

Die Ausgangssituation war durch eine heterogene IT-Landschaft geprägt, in der spezialisierte Ressourcen für rechenintensive KI-Anwendungen entweder nicht vorhanden oder nur eingeschränkt zugänglich waren. Insbesondere fehlte eine einheitliche, skalierbare Lösung für GPU-basierte Workloads, die den vielfältigen Anforderungen der Fachbereiche gerecht werden konnte. Es bestand ein Bedarf, sowohl **klassische Machine-Learning-Workloads als auch komplexe Deep-Learning-Verfahren** in Forschung und Lehre zu unterstützen – idealerweise in einer integrierten, containerisierten und benutzerfreundlichen Umgebung. Im Projektverlauf stellte sich als wichtige Ergänzung die **Bereitstellung von Foundation Models (bspw. große Sprachmodelle) als Infrastruktur** heraus.

Die Zielsetzung von KITT war daher zweifach: Einerseits sollte durch den Aufbau eines zentralen KI-Clusters mit leistungsfähiger Hardware und intelligenter Ressourcenverteilung (mittels run:ai und Kubernetes) die Basis für Forschung und Entwicklung im Bereich KI gelegt werden. Andererseits sollte durch entsprechende didaktische und technische Maßnahmen ein möglichst niederschwelliger Zugang

geschaffen werden, der die Nutzung durch verschiedene Zielgruppen (Studierende, Lehrende, Forschende) ermöglicht und fördert.

Im Fokus standen dabei insbesondere:

- die Ermöglichung eigener Forschungs- und Anwendungsprojekte auf Basis der Infrastruktur,
- die Integration von KI-Tools in die Lehre,
- die institutionelle Verstetigung durch Governance- und Supportstrukturen sowie
- die Vorbereitung auf eine nachhaltige Nutzung über die Projektlaufzeit hinaus.

Mit dem Projekt wurde so ein zentraler Baustein für die Weiterentwicklung der KI-Kompetenzen an der Hochschule geschaffen – sowohl im wissenschaftlich-technischen als auch im organisatorisch-didaktischen Sinne.

2. Wesentliche Projektergebnisse

2.1 Aufbau der technischen Infrastruktur

Der erfolgreiche Aufbau der KI-Infrastruktur KITT basierte auf einer sorgfältig geplanten und technisch anspruchsvollen Kombination aus Hochleistungs-Hardware, Container-Orchestrierung und intelligenter Ressourcensteuerung. Ziel des Arbeitspakets war es, eine leistungsfähige und zukunftssichere Rechenumgebung für Machine Learning (ML) und Deep Learning (DL) zu schaffen, die sich zugleich flexibel in bestehende Hochschulprozesse integrieren lässt.

Auswahl und Beschaffung der Hardware

Als zentrales Hardwareelement wurde ein *NVIDIA DGX H100* System beschafft. Die Wahl fiel auf diese Plattform, da sie speziell für KI-Workloads im Hochleistungsbereich ausgelegt ist und durch ihre hohe Speicherbandbreite und Rechenleistung auch komplexe Modelle (z. B. große Transformer-Architekturen) effizient verarbeiten kann. Die Entscheidung für ein geschlossenes Komplettsystem statt einer offenen Serverarchitektur wurde getroffen, um eine robuste und validierte Plattform mit garantierter Kompatibilität für aktuelle Software-Stacks zu erhalten.

Die Beschaffung der Hardware verlief erfolgreich, war jedoch durch pandemiebedingte Marktverzögerungen und Lieferschwierigkeiten beeinträchtigt. Die ursprünglich vorgesehene Inbetriebnahme konnte daher erst verspätet realisiert werden. Diese zeitliche Verschiebung wurde dem Projektträger angezeigt und hatte keine Auswirkungen auf den Gesamtmittelabruf.

Softwarestack

Die technische Architektur wurde auf einem *Kubernetes-Cluster (RKE2)* mit drei Master-Knoten aufgebaut, in welches das DGX-H100 System als Worker-Node integriert ist. Die Master-Knoten werden auf existierender Hard- und Software der IT und Medientechnik (ITM) der h2 bereitgestellt und erlauben einen ausfallsicheren

Zugriff auf die KI-Infrastruktur. Zur Erleichterung der Administrierbarkeit wurde *Rancher* als Management-Plattform eingesetzt. Dies ermöglichte insbesondere eine benutzerfreundliche und zentralisierte Verwaltung des Clusters sowie die Integration der Nutzerverwaltung.

Zur Orchestrierung und Steuerung der GPU-Ressourcen wurde *run:ai* als zentrales Softwaremodul eingesetzt. Diese Entscheidung erwies sich als richtungsweisend, da *run:ai* dynamisches Scheduling, GPU-Fractioning und rollenbasierte Zugriffskontrolle ermöglicht. Durch diese Features konnte eine fein abgestimmte Zuweisung von Rechenressourcen realisiert werden – etwa durch die Definition dedizierter GPU-Kontingente für verschiedene Nutzergruppen (Studierende, Forschung, Dienste). Einige Dienste wurden mittels Helm-Charts direkt in Kubernetes umgesetzt.

Die Integration von *run:ai* in das Kubernetes-Ökosystem war technisch herausfordernd, konnte aber dank neu aufgebauter Expertise innerhalb des Projektteams erfolgreich umgesetzt werden. Die Plattform wurde in der Folge mehrfach weiterentwickelt und durch automatisierte Mechanismen ergänzt, z. B. im Hinblick auf Nutzungsüberwachung, Laufzeitbegrenzung und Nutzerverwaltung.

Systemintegration und Inbetriebnahme

Die Infrastruktur wurde in das bestehende Hochschulnetz integriert, einschließlich:

- Authentifizierung via Hochschul-SSO,
- VPN-basiertem externem Zugriff,
- Anbindung an zentrale Dienste (Dateisysteme, IBM Storage Scale, Git-Repositories, ...),
- sowie Moodle als didaktisches Frontend für bestimmte Komponenten (z. B. Jupyter Notebooks).

Nach der Inbetriebnahme der Plattform wurde die KI-Infrastruktur stufenweise freigegeben: Zunächst wurden Testworkloads durch das Projektteam eingespielt, anschließend Pilotprojekte betreut und schließlich eine breitere Nutzung durch die Hochschule ermöglicht.

Trotz der extern bedingten Verzögerung bei der Beschaffung konnte die Infrastruktur vollständig und stabil aufgebaut werden. Die Kombination aus leistungsfähiger Hardware (DGX H100), bewährter Open-Source-Orchestrierung (Kubernetes, Rancher) und intelligenter GPU-Steuerung (*run:ai*) hat sich als hochgradig leistungsfähig und zukunftssicher erwiesen. Die Plattform steht nun dauerhaft für die Nutzung in Forschung und Lehre zur Verfügung und bildet den zentralen operativen Kern des Projekts KITT.

2.2 Dienste, Plattformkomponenten und erste Use Cases

Nach dem erfolgreichen Aufbau der technischen Infrastruktur wurden erste Dienste bereitgestellt, prototypische Anwendungen implementiert und die Voraussetzungen für konkrete Nutzungsszenarien aufgebaut. Ziel war es, die Infrastruktur in Betrieb zu nehmen, anwendungsnah zu erproben und iterativ weiterzuentwickeln – sowohl im

Hinblick auf technische Anforderungen als auch hinsichtlich der Nutzerfreundlichkeit und Einbindung in den Hochschulbetrieb.

Bereitstellung zentraler Dienste

Im Zentrum der Inbetriebnahme standen drei grundlegende Dienste, die in direktem Bezug zu den Zielgruppen des Projekts standen:

- 1. JupyterHub:**
Der Dienst wurde basierend auf dem Projekt „Zero to JupyterHub“ (z2jh) über einen Helm-Chart in die Infrastruktur integriert und in das Moodle-System der Hochschule eingebunden. Dadurch entstand eine niederschwellige, browserbasierte Möglichkeit, KI-Workloads interaktiv umzusetzen – insbesondere in der Lehre. Der JupyterHub ist mit GPU-Zugriff versehen und erlaubt personalisierte Umgebungen mit Bibliotheken wie PyTorch, TensorFlow oder scikit-learn.
- 2. MATLAB-Rechenumgebungen:**
Für Fachbereiche mit stark ingenieurwissenschaftlicher Ausrichtung – insbesondere IWID – wurden spezialisierte MATLAB-Umgebungen konfiguriert, die ebenfalls containerisiert über die Infrastruktur bereitgestellt werden. Damit konnte insbesondere das methodische Spektrum erweitert und bestehende Forschungsfragen in die Infrastruktur überführt werden. Die Umgebungen werden über run:ai individuell für die Nutzung zur Verfügung gestellt.
- 3. h2 Chat – LLM-basierter Assistenzdienst:**
Als hochschuleigene Implementierung eines KI-gestützten Interfaces wurde der h2 Chat entwickelt, ein Chat-System mit Zugriff auf verschiedene Large Language Models (LLMs). Der Dienst bietet eine einheitliche Oberfläche auf Basis der Open Source Lösung „LibreChat“ zur Nutzung von Sprachmodellen in der Beratung, der Lehre und im Selbststudium. Die Implementierung beruht auf einer containerisierten Architektur mit modularer Backend-Anbindung.
- 4. run:ai für Job-Scheduling:**
Für die Umsetzung weiterer, individueller Machine-Learning Workloads steht die run:ai Umgebung zur Verfügung. Der Fokus liegt hier vor allem auf Abschlussarbeiten und Forschungsprojekte. So wurden verschiedene Bachelorprojekte sowie Entwicklungsschritte im Rahmen von Forschungsprojekten über run:ai umgesetzt.

Diese Dienste wurden als „Basisschicht“ der Plattform eingeführt und kontinuierlich erweitert, dokumentiert und hochschulöffentlich verfügbar gemacht.

Nutzung durch Pilotprojekte

Zur exemplarischen Demonstration der Leistungsfähigkeit der Infrastruktur wurden gezielt erste prototypische Anwendungen umgesetzt. Hervorzuheben ist dabei das Pilotprojekt zur Auswertung von Luftbilddaten mittels neuronaler Netze. Gemeinsam mit dem Fachbereich Wasser, Umwelt, Bau und Sicherheit (FB WUBS) wurde ein U-Net-basiertes Segmentierungsverfahren zur Ableitung von Kartierungsdaten aus Fernerkundungsbildern implementiert. Dieses Projekt diente als technischer und

didaktischer Leuchtturm und zeigte eindrucksvoll die Möglichkeiten der Infrastruktur im Umgang mit speicher- und rechenintensiven Deep-Learning-Workflows.

Die Anwendung war gleichzeitig ein Beispiel für die datengetriebene, interdisziplinäre Forschung und wurde in KI-Lehrveranstaltungen als Use Case integriert. Es steht als dokumentiertes Jupyter-Notebook mit strukturiert abgelegten Trainingsdaten und vortrainierten Gewichten zur Verfügung.

Daneben wurden mehrere experimentelle Prototypen getestet: Direkt mit Forschungspartnern aus den Fachbereichen und insbesondere mit den Projekten SynerKI und ZAKKI.

Governance und Ressourcenzuteilung

Parallel zur Inbetriebnahme der Dienste wurde eine Governance-Struktur für die Ressourcenzuteilung implementiert. In run:ai entstand eine dreistufige Struktur von Nutzergruppen:

- **h2-students:** Lehre, Übungen, Abschlussarbeiten (Bachelor-, Master- und Projektarbeiten)
- **h2-research:** Forschungsprojekte, experimentelle Workloads
- **h2-all:** gemeinsame Dienste

Diese Struktur erlaubt es, Workloads gezielt zu priorisieren, Laufzeiten zu kontrollieren und Fairness in der Ressourcennutzung sicherzustellen. Die technische Umsetzung in run:ai erfolgt über Projects und Quotas, wodurch eine kontinuierliche Überwachung und eine Automatisierung des Ressourcenmanagements möglich ist. Die Abbildung dieser Governance direkt in der technischen Plattform (run:ai) erwies sich als effizient und nachvollziehbar – und sichert die langfristige Nutzbarkeit des Systems für verschiedene Zielgruppen. Einige zentrale Dienste, darunter JupyterHub und vLLM/LiteLLM zur Bereitstellung von Sprachmodellen via API, werden als zentrale Infrastruktur neben run:ai zur Verfügung gestellt.

Weiterhin wurde eine zentrale Email-Adresse kitt@h2.de und ein zugehöriges Ticket-System in Jira eingerichtet, was die zukünftige Verarbeitung von Support-Anfragen vereinfacht.

2.3 Integration in die Lehre

Ein zentrales Ziel des Projekts KITT war die breite Integration von KI-Infrastruktur in die Hochschullehre. Dabei sollte nicht nur der Zugang zu Rechenleistung ermöglicht, sondern auch eine didaktisch sinnvolle und technisch zuverlässige Plattform geschaffen werden, mit der Lehrende wie Studierende praktische Kompetenzen im Umgang mit KI-Technologien erwerben können. Hier erfolgte ein enger Austausch mit dem Projekt ZAKKI an der h2.

Niedrigschwelliger Zugang über Moodle und JupyterHub. Der Betrieb von JupyterHub als containerisierte Plattform wurde didaktisch eingebettet, indem über

das bestehende Moodle-System direkter Zugang zu interaktiven Python-Umgebungen mit GPU-Zugriff geschaffen wurde. Dies ermöglichte Studierenden eine sofortige Anwendung von Lerninhalten ohne eigene Installationen oder Konfigurationsaufwände. Die Landing-Page <https://ai.h2.de> (Zugriff über das Hochschulnetz) bündelt die wichtigsten Informationen.

Einführung von KI-Werkzeugen in der Lehre. In mehreren Lehrveranstaltungen wurden KI-bezogene Tools und Plattformdienste aktiv genutzt. Dazu zählen insbesondere:

- **KI umsetzen mit Python:** Hier konnten Studierende eigene Modelle auf JupyterHub aber auch über run:ai entwickeln und trainieren.
- **IT and Data Science:** Vermittlung von Grundlagen der IT, u.a. Nutzung von SQL und Umsetzung einfacher Data Science Projekte.
- **Informatik (für Elektrotechnik):** Nutzung für die Entwicklung einfacher C-Projekte; Planung der Einbindung von KI-Werkzeugen für das Programmieren.
- **AI Coding Club:** Extracurriculares Format für Community-Lernen mit Zugriff auf reale Modelle und Daten.
- **AI Engineering:** Im Rahmen des Studiengangs AI.Engineering fand eine erste Übung zur Studierendengewinnung statt, bei der die KITT Infrastruktur genutzt wurde.

Integration von h2 Chat als digitalem Lernassistenten. Der über KITT entwickelte Dienst „h2 Chat“ – ein LLM-basiertes System mit Schnittstelle zu unterschiedlichen Sprachmodellen – wurde in didaktische Szenarien eingebettet, u. a. als tutorielle Unterstützung, als Hilfestellung beim Programmieren und als Werkzeug zur Exploration von Textgenerierung, Prompt Engineering und Modellvergleich. Der Chat wurde in Lehrveranstaltungen und hochschuldidaktischen Kontexten (z. B. Hochschuldidaktische Woche, Hochschulforum) vorgestellt und erprobt. Konkretes Beispiel ist ein KI-basierter Tutor für das Fach „IT and Data Science“. Weiterhin wurden große Sprachmodelle in die Programmierumgebung JupyterLab integriert und didaktisch in die Lehre in den Fachbereichen WUBS, SGM und IWID eingebunden. Im Rahmen des Projektes ZAKKI wurden begleitende didaktische Studien zum Programmierunterricht unter Nutzung von KI-Unterstützung angefertigt.

Unterstützung von Bachelorarbeiten und Studienprojekten. Mehrere Bachelorarbeiten und projektbezogene Studienleistungen konnten direkt mit Hilfe der bereitgestellten Plattform durchgeführt werden. Dies betrifft sowohl klassische Machine-Learning-Projekte als auch aktuelle Deep-Learning-Ansätze. Die Infrastruktur bot dabei sowohl die notwendige Rechenleistung als auch eine standardisierte Arbeitsumgebung für Reproduzierbarkeit und kollaboratives Arbeiten.

Hochschuldidaktische Formate und interne Verbreitung. Die Infrastruktur wurde nicht nur in regulären Lehrveranstaltungen eingesetzt, sondern auch in Formaten der internen Weiterqualifikation sichtbar gemacht. In hochschuldidaktischen Wochen und internen Weiterbildungsangeboten wurden die Plattformkomponenten (z. B. run:ai, JupyterHub, h2 Chat) vorgestellt, getestet und diskutiert. Die Erfahrungen aus diesen Formaten flossen in die Weiterentwicklung des Lehrkonzepts und der Dokumentationen ein.

Zusätzlich wurden technische und didaktische Dokumentationen für Lehrende erstellt, die über das Moodle-System der Hochschule verfügbar sind. Damit wurde eine Grundlage für die dauerhafte Integration der Dienste in die reguläre Lehre geschaffen.

Die didaktische Einbettung und die Sichtbarkeit in der Hochschule ermöglichen eine dauerhafte Nutzung der Plattform über das Projektende hinaus. Allerdings ist bislang die Stabilität noch nicht umfassend getestet – dies behindert bislang die Anwendung in weiteren Fächern. Mit zunehmenden Erfahrungen aus den Pilot-Lehrveranstaltungen ist eine zunehmende Etablierung in der Breite und eine Verstetigung zu erwarten.

2.4 Nutzung in der Forschung

Ein wesentliches Ziel des Projekts KITT war neben der Integration in die Lehre auch die gezielte Öffnung der Infrastruktur für Forschungszwecke. Die flexible Nutzbarkeit der Ressourcen, insbesondere durch containerisierte Workloads und die run:ai-Umgebung, erlaubte es, forschungsseitig explorative und rechenintensive Projekte auf der Plattform umzusetzen – teilweise im Rahmen laufender Forschungsvorhaben, teilweise im Vorgriff auf geplante Drittmittelprojekte.

Bereits in der Aufbauphase der Plattform wurde gezielt auf Anforderungen aus der Forschung Rücksicht genommen. Hierzu zählten unter anderem:

- spezifische Softwareumgebungen für wissenschaftliche Bibliotheken,
- Zugang zu benutzerdefinierten Containern,
- Nutzung dedizierter GPUs über längere Zeiträume,
- sowie Unterstützung für die Integration in bestehende Forschungsdaten-Workflows.

Diese Voraussetzungen erlaubten es, Forschungsprojekte mit heterogenen Anforderungen auf der Plattform zu realisieren.

Beispielhafte Anwendungsfälle

1. Übersetzung in leichte Sprache (Projekt SynerKI)

Im Rahmen erster Tests innerhalb des Projekts **SynerKI** („Synergetische KI-Lösungen für ganzheitliche Prozessoptimierung“) wurde ein konkreter Anwendungsfall umgesetzt, der die Übersetzung komplexer Fachtexte in **leichte deutsche Sprache** mit Hilfe von Sprachmodellen verfolgt. Dabei kamen verschiedene LLMs auf der KITT-Infrastruktur zum Einsatz, die hinsichtlich Verständlichkeit, Textkohärenz und semantischer Treue verglichen wurden. Der Case diente gleichzeitig als funktionales Benchmarking sowie als methodischer Einstiegspunkt für weitere wirtschaftsnahe Forschungsfragen im Kontext von Sprache, Kommunikation und Inklusion.

Die enge Anbindung an das Projekt SynerKI stellt sicher, dass der Use Case nicht nachhaltig weiterentwickelt und evaluiert wird. Die Plattform KITT wird in SynerKI als zentrale Rechenbasis für mehrere Projektteile genutzt.

2. Vorhersage von Überflutungsgebieten (FB WUBS)

Ein weiteres Beispiel für forschungsnahe Nutzung ist das Testen eines Transformer-basierten Deep-Learning-Modells zur Vorhersage von Überflutungsgebieten in Kooperation mit dem FB WUBS. Aufbauend auf einem wissenschaftlichen Preprint wurde ein bestehender Open-Source-Code adaptiert und auf hochschulinterne Geodaten angewendet. Dieses Projekt zeigt exemplarisch, wie durch die Nutzung von KITT aktuelle wissenschaftliche Entwicklungen in die angewandte Forschung an der Hochschule transferiert werden können.

Darüber hinaus wurden auf der Infrastruktur mehrere studentische Forschungsprojekte (Bachelorarbeiten und projektgebundene Studienleistungen) umgesetzt. Ein Beispiel ist das Training von Modellen für die Vorhersage von Photovoltaik-Leistungen. Die Studierenden profitierten hier nicht nur von der Rechenleistung, sondern auch von der Standardisierung der Umgebung, was insbesondere für Experimente mit KI-Modellen, Datenpipelines und Softwareentwicklungen relevant war.

Integration in forschungsseitige Projektanträge

Die Plattform KITT wurde mehrfach explizit in Drittmittelanträgen aufgegriffen. Dazu zählen unter anderem:

- *"Study and Thrive - Zusammen.Wachsen"*, Förderlinie "Lehrarchitektur", Stiftung Innovation in der Hochschullehre (*bewilligt*)
- *h2 AI Application Hub* – „Erforschung der Implementierung von KI-Lösungen in Organisationen“, EFRE Wissenschaft, Land Sachsen-Anhalt (*positiv begutachtet, auf Warteposition*)
- *KIAre* – „KI-basiertes Supportaggregat für ressourceneffiziente Pressschweißprozesse“, EFRE Wissenschaft, Land Sachsen-Anhalt als Partner mit Prof. Frank Trommer (*positiv begutachtet, auf Warteposition*)
- *KI4Aller* – „Nutzung des Maschinellen Lernens für die prädiktive Entwicklung ökologischer Trittsteine in der Agrarlandschaft: Aufwertung der Landschaftsqualität bei Wefensleben durch Anbindung des Aller-Altarms“, Bekanntmachung KI-Leuchttürme für Umwelt, Klima, Natur und Ressourcen, als Verbundpartner mit Arbeitsgruppe Ingenieurökologie (Prof. Petra Schneider), Unterhaltungsverband Aller, Umweltanalytische Produkte GmbH (*nicht bewilligt*)
- *UGI Soil* – „Urban Green Infrastructure and Soil: Stadtboden als Ressource für die Klimaanpassung“, Bekanntmachung Planungsbeschleunigung für die Klimaanpassung mit Urbanen Digitalen Zwillingen als Verbundkoordinator mit Arbeitsgruppe Ingenieurökologie (Prof. Petra Schneider), Landeshauptstadt Magdeburg, Hochschule für Technik und Wirtschaft Dresden, Hochschule Anhalt (*nicht bewilligt*)
- *HumanLight AI* – „Menschzentrierte Lichtfunktionen durch künstliche Intelligenz“, Richtlinie HAW-Forschungspraxis, mit Prof. Benedikt Lamontain, Prof. Olaf Überschär (*in Antragsstellung*)

- *AI-Centric Customer Data Platform*, EFRE Wirtschaft, mit aiiio GmbH (in Antragsstellung)

In vielen dieser Projekte fungiert KITT als Methoden- und Rechenplattform, auf der die eigentlichen Forschungsfragen aufsetzen. Die Plattform ermöglicht so eine schnelle Anlaufphase, eine standardisierte Infrastruktur und eine zuverlässige Integration in bestehende IT-Prozesse der Hochschule.

2.5 Community-Bildung und Verstetigung

Die langfristige Wirksamkeit technischer Infrastrukturprojekte wie KITT hängt maßgeblich davon ab, ob es gelingt, dauerhafte organisatorische Strukturen, tragfähige Communities und realistische Verwertungspfade zu etablieren. Im Projektverlauf wurden daher gezielt Maßnahmen zur institutionellen Verstetigung, zur internen Community-Bildung und zur externen Verwertung umgesetzt.

Organisatorisch ist ein Commitment zur weiteren Betreuung der Infrastruktur gegeben durch das ITM der h2 sowie die Professur „Künstliche Intelligenz und Technische Informatik“ am FB IWID.

Die Plattform wurde aktiv in den Hochschulbetrieb eingebunden, u. a. durch:

- Regelmäßige hochschulinterne Veranstaltungen, wie Hochschuldidaktische Wochen und das Hochschulforum, in denen KITT, JupyterHub und der h2 Chat vorgestellt wurden.
- Nutzung in freiwilligen Lernformaten wie dem AI Coding Club, der Studierenden Raum für eigene Experimente mit KI-Technologien gab.
- Multiplikator:innen-Strukturen, in denen Lehrende frühzeitig eingebunden wurden und anschließend ihre Kolleg:innen bei der Integration der Plattform unterstützten.
- Die Integration technischer Dokumentation und Selbstlernmaterialien in Moodle, wodurch ein dauerhafter Wissenstransfer sichergestellt ist.

Die enge Verzahnung mit den Hochschulprojekten SynerKI und ZAKKI und die Einbindung in eine ganze Reihe von Drittmittelprojekten unterstreichen die Anschlussfähigkeit und strategische Bedeutung.

Darüber hinaus wurde KITT bewusst offen konzipiert, um mit neuen Anforderungen aus Forschung und Lehre mitwachsen zu können. Die Containerisierung und die auf Kubernetes basierende Architektur erlauben eine schnelle Anpassung an neue Tools, Bibliotheken oder Workloads.

3. Zusammenfassung und Einschätzung

Das Projekt KITT konnte seine gesetzten Ziele in weiten Teilen erfolgreich umsetzen und damit zentrale Impulse für den Auf- und Ausbau KI-bezogener Kompetenzen an der Hochschule Magdeburg-Stendal liefern. Der Aufbau einer hochverfügbaren und effizient nutzbaren KI-Infrastruktur mit NVIDIA DGX H100 als Hardwarekern, Kubernetes-Cluster und run:ai als Ressourcensteuerung stellt nicht nur eine

zukunftssichere technische Basis dar, sondern ermöglicht auch erstmals die strukturierte und niederschwellige Nutzung leistungsfähiger GPU-Ressourcen in Forschung und Lehre.

Wichtigste Positionen des zahlenmäßigen Nachweises:

Die Mittelverwendung entspricht den geplanten Zielstellungen des Projekts. Die Beschaffung der DGX-Hardware, die Implementierung der Software-Stacks (Rancher, run:ai, Helm-basierte Dienste) sowie die Integration in hochschulinterne Systeme (SSO, Moodle, VPN) bilden die größten Posten der Mittelverwendung und sind vollständig realisiert. Trotz pandemiebedingter Verzögerungen wurde der Zeitplan insgesamt eingehalten und der Mittelabruf sachgerecht durchgeführt.

Notwendigkeit und Angemessenheit der Projektarbeiten:

Die realisierten Projektarbeiten erwiesen sich sowohl in technischer als auch in didaktischer Hinsicht als notwendig und sinnvoll. Der Aufbau einer robusten Infrastruktur inklusive Governance-, Support- und Nutzungsstrukturen war entscheidend, um die Integration in den laufenden Hochschulbetrieb zu ermöglichen. Für die Umsetzung der Pilotprojekte haben sich im Projektverlauf deutliche Anpassungen ergeben. Dadurch konnten Anwendungen identifiziert und umgesetzt werden, die eine deutlich bessere Verwertung in Forschung und Lehre versprechen, als zum Zeitpunkt der Antragsstellung absehbar war. Um trotz der Verzögerungen in der Beschaffung die Projektziele zu erreichen, wurde gezielt mit Hochschulprojekten zusammengearbeitet, insb. mit dem Projekt ZAKKI zur Sicherung und Verankerung der ursprünglich geplanten Arbeiten in AP4 – KI-Kurse und AP5 – KI-Repositoryum.

Voraussichtlicher Nutzen und Verwertbarkeit:

Die Verwertbarkeit des Projektergebnisses ist bereits während der Projektlaufzeit deutlich sichtbar geworden und zeigt sich insbesondere in folgenden Aspekten:

- **Kurzfristig:** Einsatz in der Lehre, Durchführung von Bachelorarbeiten und Studienprojekten, Pilotprojekte in Forschung und Didaktik.
- **Mittelfristig:** Integration in Drittmittelprojekte (z. B. SynerKI, KIAre, AI Hub) und curriculare Lehrangebote.
- **Langfristig:** Verstetigung der Infrastruktur über ITM und Professuren, Integration in Governance- und Supportstrukturen sowie hohe Anschlussfähigkeit für neue Tools, Modelle und Forschungsprojekte.

Im Sinne des fortgeschriebenen Verwertungsplans ist vorgesehen, die Infrastruktur als zentrale Serviceplattform dauerhaft hochschulweit bereitzustellen und sowohl technisch (z. B. durch LLM-APIs) als auch organisatorisch (z. B. über Moodle und Jira) weiterzuentwickeln. KITT fungiert dabei zunehmend als Shared Research and Learning Infrastructure mit strategischer Bedeutung.

Bekannter Fortschritt bei anderen Stellen:

Im Projektverlauf wurden kontinuierlich externe Entwicklungen beobachtet – insbesondere im Hinblick auf andere KI-Infrastrukturprojekte an Hochschulen. Die eingesetzte Lösung (Kombination aus DGX, Kubernetes, run:ai) ist auch in anderen Kontexten erprobt (z. B. Helmholtz AI, NFDI4DS) und wurde bewusst auf Open-Source- und Industrie-Standards (Kubernetes / RKE2) aufgebaut.

Die gewählte Software-Plattform run:ai wurde im Projektverlauf von der Firma NVIDIA aufgekauft – hierdurch ist zu vermuten, dass auch ein langfristiger Support der Software sichergestellt ist. Zudem wurde die zukünftige Umstellung von run:ai auf Open Source-Lizenzen angekündigt, so dass auch die langfristige Finanzierung des Betriebs sichergestellt ist. Im Hinblick auf die Hardware sind im Laufe des Projektverlaufs bereits aktueller Hardware-Versionen angekündigt und verfügbar gewesen (insb. NVIDIA H200 Architektur). Dies zeigt deutlich das vorhandene Tempo in den technischen Entwicklungen von Grafikkarten-Hardware – und ist somit ein nicht vermeidbares Risiko. Die gesammelten Erfahrungen können aber teilweise für den zielgerichteteren Ablauf künftiger Vergabeverfahren genutzt werden.

Dies beeinträchtigt jedoch in keinster Weise die langfristige Nutzbarkeit und Anschlussfähigkeit der KITT-Plattform, da kompatible Softwareschnittstellen genutzt werden.

Erfolgte Veröffentlichungen:

Erste Dissemination der Projektergebnisse erfolgte im Rahmen interner Veranstaltungen (Hochschulforum, Hochschuldidaktische Woche) sowie durch hochschulöffentliche Dokumentation und KI-Kurse (Moodle). Darüber hinaus sind mehrere begutachtete Drittmittelanträge mit direkter Referenz auf KITT entstanden. Für die Zukunft ist jedoch mit weiteren Fach-Veröffentlichungen von Projektergebnissen aufbauend auf der KITT-Infrastruktur zu rechnen.

Zusammenfassend lässt sich festhalten, dass mit KITT ein robuster, leistungsfähiger und hochschulweit anschlussfähiger Grundstein für KI-gestützte Lehre und Forschung gelegt wurde. Die Verankerung in internen Strukturen, die Anschlussfähigkeit an nationale Entwicklungen sowie die zunehmende Einbindung in Drittmittelkontexte sichern die nachhaltige Verwertung des Projekts über seine Laufzeit hinaus. Einzelne Aspekte – wie etwa die Stabilitätsprüfung unter Last, der breitere Rollout in weniger KI-affinen Fachbereichen sowie die langfristige Personalverankerung im Support – bleiben als Entwicklungsfelder bestehen und bieten Anknüpfungspunkte für die Weiterführung und Verstetigung.