

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 1 Intelligentes Kamerasystem zur Erkennung von Ertrinkenden in Schwimmbädern



Projektübersicht

Anzahl Studierende	1
Art	Vorlaufforschung mit externen Praxispartnern aus dem Bäderbetrieb
Projektverantwortung	Betreuung durch Dr. Schultenkämper und Prof. Brandt-Pook in Zusammenarbeit mit Anwendungspartnern aus dem Schwimmbad- und Sicherheitsumfeld
Projektkontext	Das Projekt ist im Schnittpunkt von Computer Vision, KI-gestützter Sicherheitstechnik und angewandter Forschung im öffentlichen Raum angesiedelt. Im Projektumfeld stehen typische Becken- und Videoszenarien, fachliche Expertise aus dem Bäderbetrieb, methodische Betreuung im Bereich maschinelles Lernen sowie Rechenressourcen für Training und Evaluation zur Verfügung. Perspektivisch sind Kooperationen mit kommunalen oder privaten Badbetreibern sowie Anschlussmöglichkeiten für studentische Hilfstätigkeiten im Forschungsumfeld denkbar.

Abstrakt

Das Projekt untersucht KI-gestützte Kamerasysteme zur frühzeitigen Erkennung potenzieller Ertrinkungssituationen in Schwimmbädern. Ziel ist die Entwicklung und Evaluation von Verfahren der Bild- und Videodatenanalyse, die kritische Bewegungsmuster robust erkennen und das Aufsichtspersonal durch verlässliche Warnungen unterstützen. Neben der Modellgüte werden Datenschutz, Echtzeitfähigkeit und praktische Einsetzbarkeit betrachtet.

Kurzbeschreibung

Ertrinkungsunfälle in Schwimmbädern entstehen häufig in kurzer Zeit und werden trotz Aufsicht nicht immer sofort erkannt, insbesondere bei hoher Beckenbelegung, eingeschränkter Sicht oder parallelen Ereignissen im Badebetrieb. KI-unterstützte Kamerasysteme versprechen hier eine zusätzliche

Sicherheitsebene, indem sie Videoströme in Echtzeit analysieren und auffällige Bewegungs- oder Lageänderungen im Wasser automatisch markieren. Ziel des Projekts ist es, datengetriebene Methoden zur Erkennung potenzieller Notsituationen zu untersuchen, prototypisch umzusetzen und hinsichtlich Genauigkeit, Fehlalarmrate, Robustheit und praktischer Integrierbarkeit in den Schwimmbadbetrieb zu bewerten.

Aufgabenstellung

Das Projekt analysiert den Stand der Forschung zu videobasierter Erkennung kritischer Situationen im Wasser und leitet daraus ein geeignetes methodisches Vorgehen ab. Darauf aufbauend werden relevante Ereignisklassen definiert, geeignete Bild- oder Videodaten aufbereitet, ein Auswerteverfahren für die Erkennung potenzieller Ertrinkungssituationen entwickelt und anhand nachvollziehbarer Qualitätsmaße evaluiert. Je nach Projektfortschritt können zusätzlich Verfahren zur Reduktion von Fehlalarmen, zur erklärbaren Modellbewertung, zur datenschutzgerechten Verarbeitung oder zur echtzeitnahen Alarmierung untersucht werden.

Bezug zum Thema Data Science

Das Projekt adressiert zentrale Teilbereiche der Data Science, insbesondere Datenakquise und -aufbereitung, explorative Datenanalyse, Merkmalsextraktion aus Bild- und Videodaten, maschinelles Lernen sowie die systematische Evaluation von Modellen. Methodisch stehen Verfahren aus Computer Vision und Deep Learning im Vordergrund, etwa Objekterkennung, Tracking, Aktions- bzw. Anomalieerkennung und die Analyse zeitlicher Muster in Videosequenzen. Ergänzend spielen Themen wie Datenqualität, Klassenungleichgewicht, Generalisierungsfähigkeit, Fehlalarmminimierung, Interpretierbarkeit und datenschutzgerechte Datenverarbeitung eine wesentliche Rolle.

Verfügbare Ressourcen

Für das Projekt stehen methodische Betreuung im Bereich Data Science und Computer Vision, übliche Entwicklungswerkzeuge für die Analyse von Bild- und Videodaten sowie Rechenressourcen für Training, Experimente und Evaluation zur Verfügung. Je nach Kooperationsstand können beispielhafte Video- oder Simulationsdaten aus schwimmbadnahen Szenarien, Domänenwissen aus dem Sicherheits- und Bäderumfeld sowie Feedback aus der Anwendungspraxis eingebunden werden. Zusätzlich können Open-Source-Bibliotheken und etablierte Frameworks für Deep Learning, Tracking und Videointerpretation genutzt werden.

Projektplan

Iteratives Vorgehen und frühe Prototypen

Das Projekt folgt bewusst keinem reinen sequenziellen Vorgehen, sondern kombiniert wissenschaftliche Analyse mit einer iterativen, prototypischen Entwicklung.

Bereits in frühen Projektphasen sollen einfache, datenbasierte Prototypen (Minimum Viable Models) entstehen, um ein praktisches Verständnis für die Herausforderungen der Videodatenanalyse zu entwickeln. Diese frühen Ansätze können auf vereinfachten oder öffentlich verfügbaren Beispieldaten basieren und werden im Projektverlauf schrittweise erweitert und verbessert.

Ziel ist es, theoretische Erkenntnisse kontinuierlich mit praktischen Experimenten zu verbinden, um frühzeitig Potenziale, Grenzen und typische Fehlerquellen zu identifizieren.

Erstes Semester: Exposé & Grundlagen

- Strukturierte Analyse des Stands der Forschung (Computer Vision / Anomalieerkennung in Videos)
- Überblick über bestehende technische Ansätze (z. B. Object Detection, Tracking, Action Recognition)
- Marktüberblick:
 - existierende kommerzielle Systeme
 - relevante Open-Source-Ansätze
- Erste Einordnung:
 - Welche Ansätze sind für Schwimmbäder plausibel geeignet?
- Grobes methodisches Konzept für das eigene Vorgehen
- Erste einfache MVPs auf Basis von Beispiel- oder Open-Source-Daten zur frühen Validierung grundlegender Annahmen

👉 Zielbild: „Ich verstehe Problemraum + Lösungsraum“

Zweites Semester: Methodische Ausarbeitung & erste Prototypen

- Definition konkreter Use Cases / Ereignisklassen (z. B. bewegungslos im Wasser, atypische Bewegungsmuster)
- Konzeption Datenbasis:
 - Welche Daten wären ideal?
 - Welche Daten sind realistisch verfügbar?
 - Annotationskonzept
 - Klassenungleichgewicht
- Ableitung eines geeigneten Modellierungsansatzes basierend auf Datenverfügbarkeit und Use Case, z.B.
 - Überwachtes Lernen (gelabelte Daten - „Ertrinkung“ vs. „normal“, vs. ...)
 - Anomalieerkennung (lernt „normales Verhalten“ und erkennt Abweichungen)
- Abstimmung des Datenkonzepts mit dem Praxispartner (z. B. Kamerapositionen, Blickwinkel, Nutzungsszenarien) sowie Ableitung von Anforderungen an die Datenerfassung (z. B. relevante Szenarien, Datenqualität), die durch den Praxispartner umgesetzt werden können
- Baseline-Ansatz (z. B. einfache Heuristiken oder Standardmodelle)
- Weiterentwicklung erster Prototypen auf projektspezifischen Daten
- Erste Evaluation und Analyse typischer Fehlerquellen

👉 Zielbild: „Ich kann das Problem in Daten und Modelle übersetzen“

Drittes Semester: Iterative Entwicklung & Evaluation

- Entwicklung und Vergleich mehrerer Modellansätze
 - klassische CV vs. Deep Learning
 - verschiedene Architekturen oder Feature-Ansätze
- Iterative Verbesserung:
 - Hyperparameter-Tuning
 - Feature Engineering
 - ggf. Datenaugmentation

- Fokus auf:
 - Fehlalarmrate (sehr wichtig im Kontext!)
 - Robustheit (z. B. volle vs. leere Becken)
 - Generalisierbarkeit
- Systematische Evaluation:
 - sauberes Evaluationsdesign
 - geeignete Metriken (z. B. Precision/Recall, False Positives)
 - Vergleich mehrerer iterativ entwickelter Modellvarianten
 - Analyse typischer Fehlerquellen
- Reflexion:
 - Wo funktionieren die Ansätze gut / schlecht?

👉 Zielbild: „Ich verstehe, was wirklich funktioniert – und warum“

Viertes Semester: Integration & Transfer

- Konsolidierung des besten Ansatzes
- Überführung eines iterativ entwickelten Prototyps in eine konsolidierte Lösung
- Betrachtung der praktischen Umsetzung:
 - Echtzeitfähigkeit
 - Integration in bestehende Überwachungssysteme
- Bewertung:
 - technische Machbarkeit
 - praktische Einsatzfähigkeit
- Optional:
 - einfache Pipeline / Systemarchitektur skizzieren
- Masterarbeit + Kolloquium
- Wissenstransfer zum Praxispartner
 - Aufbereitung der Projektergebnisse für den Praxispartner sowie Diskussion möglicher Einsatzszenarien und Grenzen
 - Dokumentierter Programmcode

👉 Zielbild: „Ich bringe Forschung in Richtung Anwendung“

Eignungskriterien

Zwingend: Sehr gute Kenntnisse in Python sowie solide Grundlagen in Statistik, maschinellem Lernen und Datenanalyse; Interesse an Computer Vision, sorgfältiges wissenschaftliches Arbeiten und Bereitschaft zur Einarbeitung in sicherheitskritische Anwendungskontexte.

Optional: Erfahrung mit Deep-Learning-Frameworks wie PyTorch oder TensorFlow, Kenntnisse in Bild- und Videobearbeitung, Erfahrung mit Evaluationsdesign, Anomalieerkennung, Edge- oder Echtzeitsystemen sowie Grundverständnis für Datenschutz und ethische Fragen bei KI-gestützter Videoanalyse.

Erwerbbare Kompetenzen

Durch das Projekt erwirbt der/die Studierende Kompetenzen in der wissenschaftlichen Analyse eines aktuellen Anwendungsfeldes der Data Science, in der Entwicklung und Evaluation von Computer-Vision- und Machine-Learning-Verfahren sowie in der strukturierten Aufbereitung komplexer Bild- und Videodaten. Darüber hinaus stärkt er/sie Fähigkeiten in Forschungsdesign, Literaturarbeit, experimenteller Validierung, kritischer Ergebnisinterpretation und adressatengerechter wissenschaftlicher Kommunikation. Zusätzlich entwickelt er/sie ein vertieftes Verständnis für die Anforderungen sicherheitskritischer KI-Systeme, insbesondere im Hinblick auf Robustheit, Fehllarmer, Datenschutz und praktische Umsetzbarkeit.

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 2 Digitale Zwillinge lernen



Projektübersicht

Anzahl Studierende	1-3
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. Christian Schwede
Projektkontext	Projekt in Zusammenarbeit mit dem Fraunhofer-Institut für Software- und Systemtechnik (ISST) in Dortmund und im Kontext des Center for Applied Data Science (CfADS) sowie des Institute for Data Science Solutions (IDaS) ; Anstellung als Hiwi bei Fraunhofer ist möglich

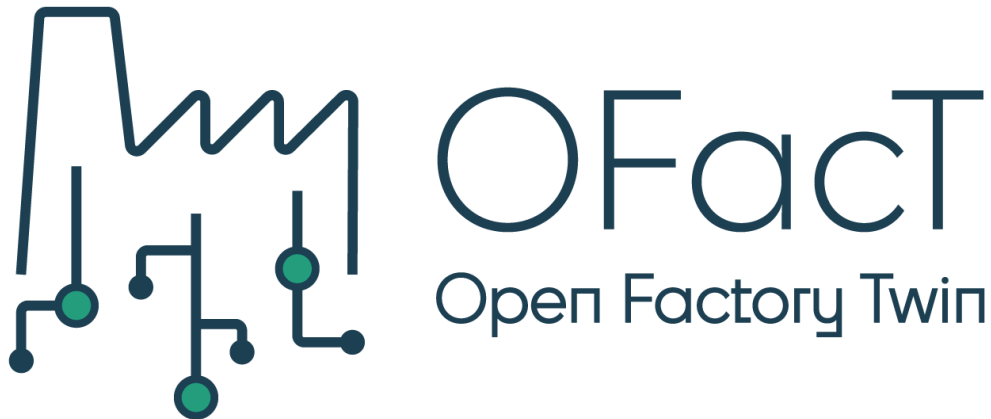
Abstrakt

Ziel des Projektes ist es simulationsbasierte Digitale Zwillinge zur Planung und Steuerung von Produktions- und Logistikumgebungen aus IoT-Daten zu lernen. Hierbei wird auf dem vom der HSBI und dem Fraunhofer ISST erstellten Open-Source-Framework OFacT aufgebaut und dieses erweitert. Die Daten stammen aus Kundenprojekten von Fraunhofer und der IoT-Factory des CfADS. Zum Einsatz kommen neben Process Mining Methoden auch generative Modelle wie Large Language Models (LLMs).

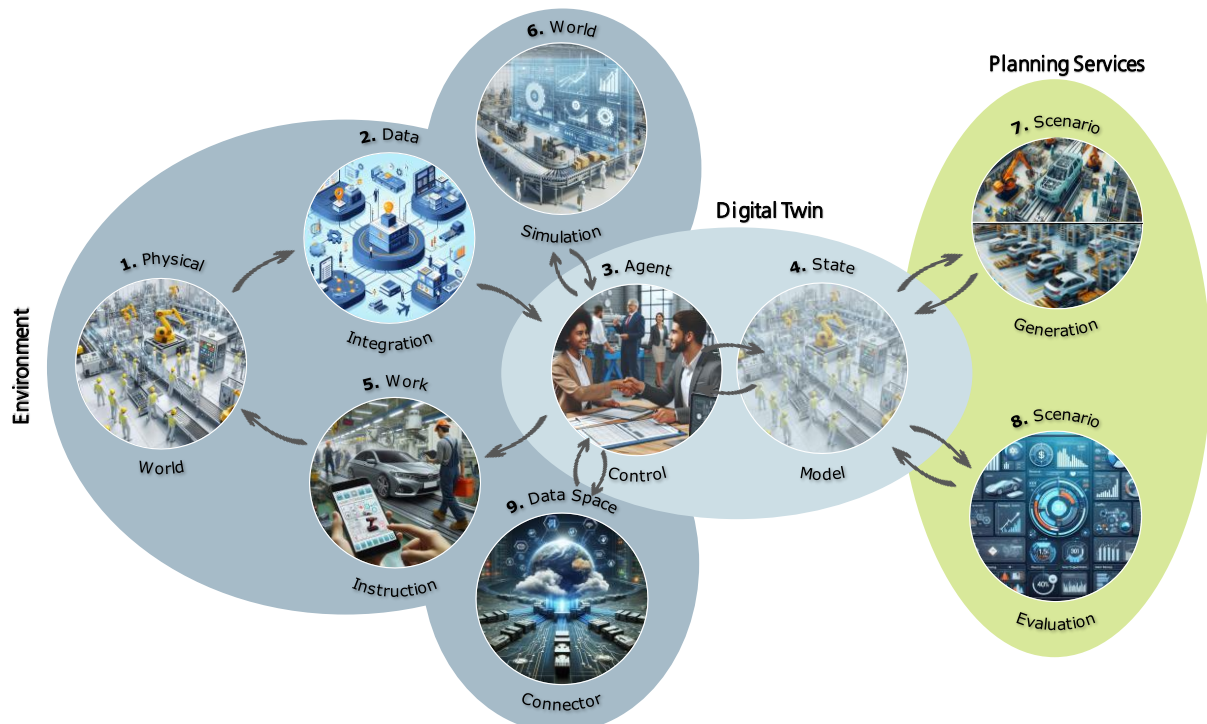
Kurzbeschreibung

Digitale Zwillinge sind der Schlüssel zur optimalen Gestaltung, Planung und Steuerung von Produktions- und Logistikumgebungen. Diese Softwareframeworks erlauben den Zugriff auf ein stets aktuelles digitales Abbild des Systems, ermöglichen die Optimierung und Simulation von Szenarien in dieser virtuellen Umgebung und vermitteln Änderungen mittels Arbeitsanweisungen direkt zurück an den Shop Floor. So können Entscheidungen in allen Lebensphasen des Systems evaluiert und optimiert werden, bevor sie getroffen werden. Sowohl langfristige Veränderungen am System, als auch Reaktionen auf kurzfristige Störungen können optimal umgesetzt werden. Das Potential für Unternehmen wird hierbei als sehr hoch eingeschätzt (prognostiziertes Marktvolumen von 10

Milliarden US \$ in 2028¹). Das Fraunhofer ISST und die HSBI haben das Open-Source-Framework OFact (Open Factory Twin) erstellt, das es insbesondere kleinen und mittelständische Unternehmen erlaubt ihre Produktionssysteme zu digitalisieren.



OFact basiert auf einem standardisierten Zustandsmodell, mit dem beliebige diskrete Materialflusssysteme abgebildet werden können. Über das Datenintegrationsmodul können beliebige Datenschnittstellen angesprochen werden. Die Steuerung des Systems wird durch ein Multi-Agenten-System übernommen, so dass Steuerungsregeln hochflexible und getrennt von Zustandsmodell abgebildet und verändert werden können. Die Agenten übernehmen auch die Steuerung des Materialflusssystemes mittels Arbeitsanweisungen.



¹ Bain & Company: Global Machinery & Equipment Report 2024

Als Umgebung kann das Framework auf eine simulierte Realität zurückgreifen, um z.B. Szenarien zu evaluieren. Planungsservices erlauben die Erstellung und Optimierung von Handlungsalternativen und ein Dashboard die Auswertung anhand von Standard-Kennzahlen.

Eine wichtige Hürde bei der Implementierung von Digitalen Zwillingen in kleinen und mittelständischen Unternehmen sind die Kosten, die durch das manuelle modellieren durch Experten entstehen. Diese Kosten fallen nicht nur bei der Erstellung, sondern auch bei der regelmäßigen Aktualisierung an. Das Ziel dieses Projektes ist es deshalb Digitalen Zwillinge aus Daten der Unternehmen zu lernen.

Aufgabenstellung

Der/die Studierende soll in dem Projekt Verfahren aus dem Bereich des Maschinellen Lernens nutzen, um Modelle von OFacT auf Basis von Daten von Produktionsumgebungen zu lernen. Die IoT-Daten oder auch Event-Logs vom Shop Floor der Unternehmen enthalten häufig schon viele der benötigten Informationen. Genutzt werden können Daten aus existierenden Industrieprojekten vom Fraunhofer ISST oder von der IoT-Factory am Campus Gütersloh. Neben Methoden des Process Mining, werden Regressions- und Klassifikationsverfahren z.B. auf Basis von tiefen Neuronalen Netzen eingesetzt. Zudem soll untersucht werden, in wie weit Teile der Modelle auch mit aktuellen Large Language Models (LLM) erstellt werden können. Hierzu kann es sinnvoll sein, Retrieval-Augmented Generation (RAG) zu nutzen um die LLM mit domainspezifischen Informationen zu versorgen.

Bezug zum Thema Data Science

Die verwendeten Methoden der Regression und Klassifikation, sowie LLM und RAG sind Bestandteil der Veranstaltung im Forschungsmaster und Kernbereiche der Data Science.

Verfügbare Ressourcen

- Daten werden über das Fraunhofer ISST und die IoT-Factory bereitgestellt
- Das Open-Source-Framework OFacT wird an der HSBI entwickelt und ist verfügbar
- Hardware für Maschine Learning ist über das Data Science Lab, das CfADS, sowie dem KI-Rechencluster yourAI der HSBI verfügbar

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die Aufgabenstellung und das Framework OFacT.

Zweites Semester: Literaturrecherche zur Generierung von Modellen mithilfe von Verfahren des maschinellen Lernens, insbesondere LLMs und RAG. Sichtung der Datensätze und Anwendung erster einfacher Verfahren des Process Mining und existierender Sprachmodelle. Erstellung eines Papers, das einen Überblick über das Forschungsgebiet gibt ist Prüfungsleistung.

Drittes Semester: Praktische Umsetzung der eigenen Lösung und Generierung eines OFacT-Modells für einen konkreten Anwendungsfall. Veröffentlichung eines Papers mit den ersten Ergebnissen ist Prüfungsleistung.

Viertes Semester: Masterarbeit und Kolloquium. Finale Evaluierung durch Vergleich verschiedener Verfahren.

Eignungskriterien

Zwingend:

- Programmierkenntnisse (vzw. Python)

Optional:

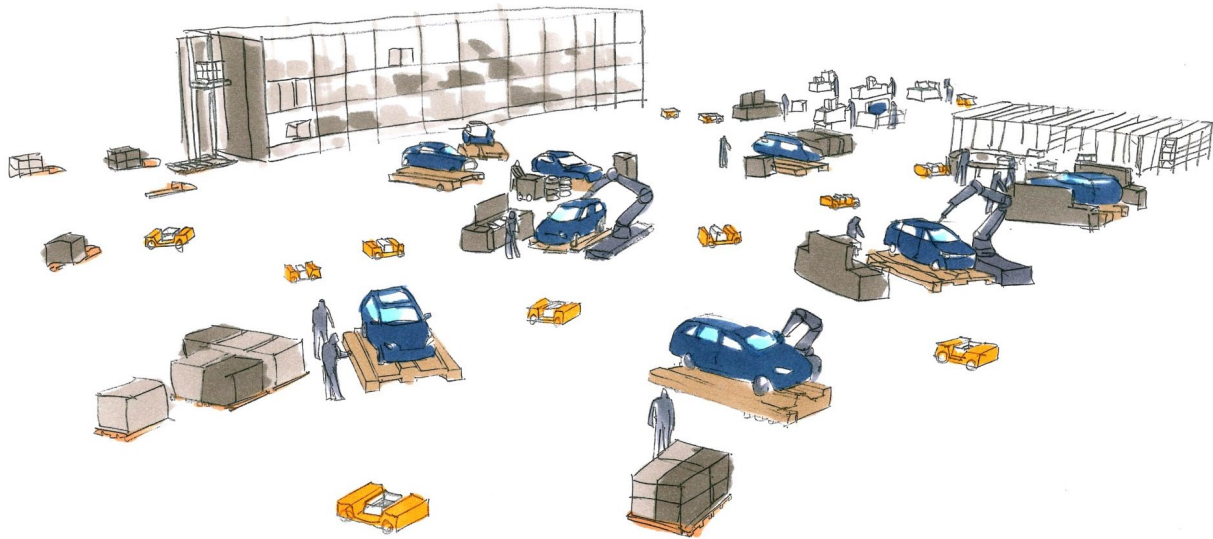
- Erfahrung mit Maschinellem Lernen
- Erfahrung mit LLMs

Erwerbbarer Kompetenzen

- Entwicklung und Verwendung von Digitalen Zwillingen
- Generierung von Modellen mittels LLMs und RAG
- Publikation von wissenschaftlichen Veröffentlichungen auf internationalen Konferenzen
- Kompetenzen in der Arbeit im Team
- Open-Source Entwicklung

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 3 KI für die hochflexible variantenreiche Massenproduktion



Projektübersicht

Anzahl Studierende	1
Art (gefördertes Projekt, Projekt mit externen Partnern, Projekt, Studienprojekt)	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt in Zusammenarbeit mit dem Fraunhofer-Institut für Software- und Systemtechnik (ISST) in Dortmund und im Kontext des Center for Applied Data Science (CfADS) sowie des Institute for Data Science Solutions (IDaS) ; Anstellung als Hiwi bei Fraunhofer ist möglich;

Abstrakt

Die stetige Individualisierung der Produkte stellt nicht nur die deutsche Automobilproduktion vor Herausforderungen, die ein vollständiges Umdenken in der Produktionsorganisation erzwingen. Weg von getakteten Fließfertigung hin zur dezentralen Matrixproduktionssystemen. Schwärme von autonomen Roboter koordinieren hier die Wertschöpfungsprozesse und passen sich flexible an die aktuelle Situation an. Während die Grundidee des Konzeptes klar umrissen ist, ergeben sich im Detail noch sehr viele offene Forschungsfragen, die insbesondere das Verhalten der einzelnen Roboter betrifft. Hier setzt das Projekt an und erforscht unterschiedliche Strategien der KI in einem virtuellen Testbed.

Kurzbeschreibung

In der sogenannten Matrixproduktion wird die Produktion dezentral von Software-Agenten in einem Multiagentsystem (MAS) koordiniert. Die Ressourcen-Agenten verhandeln hierbei bspw. mit dem Auftragsagenten über die Erbringung von Wertschöpfungsprozesse, die für die Realisierung des Auftrages nötig sind. Die Freiheit der Reihenfolge der Fertigungsschritte wird hierbei nur durch

technische Eigenschaften des Produktes beschränkt. Auf diese Weise kann sich das gesamte System flexibel an nichtgeplante Vorfälle, wie bspw. einen Lieferengpass oder einen Maschinenausfall anpassen. Während dem Konzept in der Automobilindustrie zuerst mit Ablehnung und Skepsis begegnet wurde, gibt es seit ein paar Jahren intensive Bestrebungen, eine Operationalisierung zu erforschen und voranzutreiben. Grundsätzlich sind zu einem effizienten Einsatz noch zahlreiche Fragen offen die es zu erforschen gilt.

In der Forschungsgruppe der HSBI wird seit einigen Jahren in Kooperation mit dem Fraunhofer ISST in Dortmund das Open-Source Digital Twin Framework OFacT (Open Factory Twin) entwickelt, das es unter anderem erlaubt mit agentenbasierter Simulation die hochflexiblen Matrixproduktionen zu modellieren und verschiedene Strategien und Designvarianten dynamisch zu bewerten. Das Tool ist außerdem als Digitaler Zwilling einsetzbar und kann über Sensordatenströme direkt mit der Produktion verbunden und zur Steuerung der operativen Prozesse genutzt werden.

Aufgabenstellung

Der/ die Studierende hat die Möglichkeit an einem der folgenden Themen zu arbeiten, um die Möglichkeiten und Limitationen der Matrixproduktion zu erforschen und neue Lösungskonzepte auf der Realisierung zum Stand der Technik beizusteuern.

Agentensteuerung für die dezentrale Produktion

Eine der entschiedensten offenen Fragen ist die der optimalen Steuerung der Agenten. Insbesondere die Synchronisation der Materialflüsse mit dem Hauptprodukt ist eine entscheidende, um Wartezeiten zu verhindern und gleichzeitig die Flexibilität zu erhalten. Die Entwicklung und Implementierung effizienter Steuerungsstrategien für das MAS ist hier die Kernaufgabe, die unter anderem die Frage beinhaltet ob und in wie weit in die Zukunft geplant und reserviert werden sollte. Methoden der Optimierung, Such und Planalgorithmen aus der KI oder Reinforcement Learning können hier eingesetzt werden.

Layoutdesign und -optimierung

Auch wenn der Begriff Matrixproduktion eine Gitternetzformige Anordnung der Produktionsstationen andeutet, so ist die Frage nach der optimalen Anordnung der verschiedenen Elemente der Matrixproduktion (Stationen, Lagerorte für Teile, Ladestationen der Transportfahrzeuge, Parkplätze zur Zwischenpufferung von Karosse usw.) noch nicht final entschieden. Optimierungsalgorithmen, die in Kombination mit der Simulation als Evaluierungsmethode eingesetzt werden müssen hier entwickelt werden, um für ein konkretes Produktionssetup das optimale Layout zu bestimmen. Auf die Frage wie stark das optimale Layout vom konkreten Auftragsmix abhängt ist relevant und könnte die Notwendigkeit einer regelmäßigen Überplanung des Layouts implizieren.

Routing und Stauanalyse

Die Entstehung von Staus vor den Produktionsstationen oder auf den verschiedenen Wegen ist ein zentraler Aspekt der Effizienz des Gesamtsystems. Hierzu muss die Simulationsumgebung um ein Element der Routenplanung erweitert werden, das auch temporäre Routenkonflikte erkennt. KI-Suchalgorithmen aus dem Bereich der Wegfindung müssen implementiert und verschiedenen Strategien getestet werden. Schließlich müssen existierende Steuerungsstrategien und Layouts vor dem Hintergrund von Staus im System untersucht und bewertet werden.

Beschleunigung der Verhandlungen mittels Reinforcement Learning

Eine Schwachstelle von Multiagentensystemen ist das Kommunikationsaufkommen, dass durch die dezentralen Verhandlungen zwischen den Agenten entsteht. Hier gilt es Verhandlungsstrategien zu untersuchen, implementieren und zu vergleichen und insbesondere unnötige Kommunikation zu

vermeiden. Reinforcement Learning kann eingesetzt werden, um zu erkennen, wann Anfragen von Agenten nicht zielführend sind und so im Vorhinein vermieden werden können.

Agentenbasierte Wertschöpfungsnetzwerke

Die Einbindung der logistischen Prozesse aus dem Wertschöpfungsnetzwerk rund um die Produktion ist ein weiteres offenes Thema. Neben der Entwicklung geeigneter Ver- und Entsorgungsstrategien, die mit den Bedürfnissen der flexiblen Matrixproduktion übereinstimmen müssen soll die Visionäre Frage beantwortet werden, ob die dezentrale, autonome Wertschöpfung auch über die Fabrikgrenzen hinaus erweitert werden kann. Hierzu müssen die Agenten auf Basis von Geld als Zahlungsmittel Verträge mit Lieferanten und Dienstleistern verhandeln. Die entsprechenden Verhandlungsalgorithmen und Szenarien müssen umgesetzt und evaluiert werden.

Bezug zum Thema Data Science

Softwareagenten und Multiagentensysteme, Suchalgorithmen und Reinforcement Learning sind Kerngebiete der KI und werden in verschiedenen Veranstaltungen des Forschungsmasters thematisiert.

Verfügbare Ressourcen

- Das Framework OFacT wird von des HSBI und dem Fraunhofer ISST bereitgestellt
- Industriennahe Szenarien einer individuellen Fahrradproduktion existieren zur Validierung der Entwicklungen
- Szenarien von Industriekunden des ISST und der IoT-Factory in Gütersloh existieren ebenfalls zu Evaluierung einzelner Verfahren
- Hardware für das komplexere Maschine Learning ist über das Data Science Lab oder das CfADS der verfügbar

Projektplan

Erstes Semester: Formulierung des Forschungsexposees, Einarbeitung in die OFacT-Framework, Auswahl des Themas.

Zweites Semester: Modellierung eines ersten Szenarios, Recherche zu relevanten Arbeiten in Themenfeld und Erstellung eines Reports zu Stand der Technik.

Drittes Semester: Implementierung und Evaluierung erster Verfahren, Auswertung erster Ergebnisse auf Basis des Szenarios, Verfassen eines Papers für eine (inter-)nationale Fachkonferenz.

Viertes Semester: Entwicklung weiterer Verfahren und Optimierung des Gesamtansatzes, Auswertung der Güte und Evaluierung der Ergebnisse in der Masterarbeit

Eignungskriterien

Zwingend:

- Programmierkenntnisse (optimal in Python)

Optional:

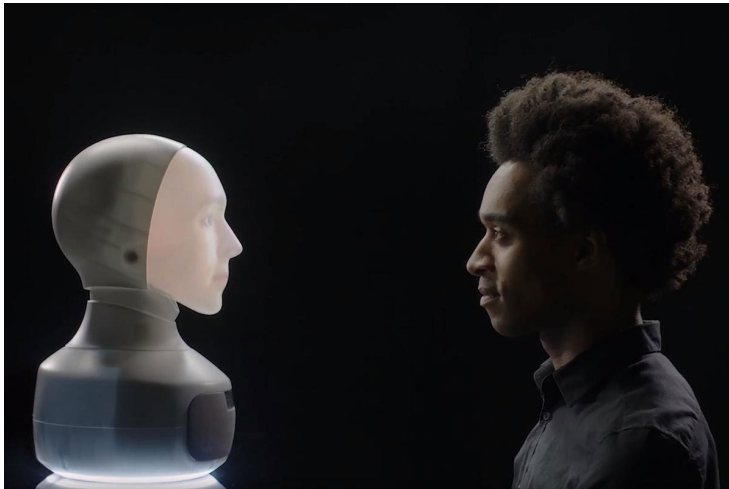
- Erfahrung mit Agentensystemen
- Erfahrung mit Simulation und Optimierung
- Hintergrundwissen zu Produktionslogistik und Produktionssystemen
- Erfahrung mit der Implementierung von Reinforcement Learning / ML-Verfahren

Erwerbbarer Kompetenzen

- Einsatz von Verfahren des Reinforcement Learnings
- Entwicklung und Bewertung von MAS
- Entwicklung von intelligenten Agenten
- Arbeiten mit Agentensimulationen
- Planung von Produktions- und Logistiksystemen insbesondere von
Matrixproduktionssystemen und Industrie 4.0

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 4 Mensch-Roboter Interaktion: Design, Implementierung und Nutzung eines humanoiden Roboters als individuellen Gesprächspartner



Projektübersicht

Anzahl Studierende	1 bis 2
Art	Studienprojekt
Projektverantwortung	Prof. Dr. Thomas Süße und Dr. Maria Kobert
Projektkontext	Mögliche Anstellung als HiWi ist abhängig von vorhandenen finanziellen Ressourcen. Zusammenarbeit mit Unternehmen ebenfalls denkbar.

Abstrakt

Die Interaktion zwischen Menschen und sozialen humanoiden Robotern ist ein zentraler Bestandteil der Mensch-Maschine Interaktion der Zukunft. Soziale Roboter können mit uns Menschen so kommunizieren, wie wir es miteinander tun - indem sie sprechen, zuhören, Emotionen zeigen und Augenkontakt halten. Sie können zum Beispiel eingesetzt werden, um Kunden zu bedienen, Gesellschaft zu leisten, Mitarbeiter*innen zu schulen oder eine Sprache zu unterrichten. Ein Beispielszenario mit dem am Campus vorhandene Roboter Furhat wird in diesem Video gezeigt:

<https://www.youtube.com/watch?v=3IEQDf9Cv4s>

Trotz der vielen Einsatzmöglichkeiten und der großen Nachfrage stellen sich bei der Entwicklung und dem Einsatz dieser vielversprechenden Technologien neue Herausforderungen bezüglich der Passung bzw. Kompatibilität zwischen Bedürfnissen und Anforderungen der Nutzenden und humanoiden Robotern dar. Eine Herausforderung liegt darin, die Roboter bezüglich Sprache, Mimik und Gestik so zu konfigurieren, dass Sie den individuellen Bedürfnissen der Nutzer gerecht werden. An diesem Punkt setzt das vorliegende Projekt an. Das übergeordnete Ziel ist es, den humanoiden Roboter Furhat, der dem Fachbereich Ingenieurwissenschaften und Mathematik am Campus Gütersloh bereits zur Verfügung steht, so zu konfigurieren, dass er in zu definierenden praktischen Anwendungsszenarien angemessen mit Menschen interagieren kann. Hierzu soll Furhat zunächst in die Lage versetzt werden,

in einem bestimmten praktischen Szenario mit Menschen kommunizieren und kollaborieren zu können. Das praktische Szenario wird gemeinsam mit den Projektbetreuenden und den Studierenden definiert. Zudem soll mit Hilfe des maschinellen Lernens erreicht werden, dass der Roboter aus den Interaktionen mit seinen menschlichen Gesprächspartnern lernt und sich an dessen Sprache, Mimik, Gestik, Vorlieben und Gesprächsinhalte erinnert. Furhat soll somit „seine Fähigkeiten ausbauen“ und als akzeptierter, angenehmer sowie hilfreicher Gesprächspartner von seinen menschlichen Gegenüber wahrgenommen werden. Begleitend zur technischen Weiterentwicklung von Furhat sollen die Interaktionen zwischen den menschlichen Probanden und Furhat auch aus der Perspektive des Menschen untersucht werden. Hierbei soll ein experimentelles Setting aufgebaut werden, das spezifische Varianten von Konfigurationen hinsichtlich des „Auftretens“ (insb. Mimik, Sprache, Optik) von Furhat und deren Wirkung auf den menschlichen Gegenüber untersucht.

Kurzbeschreibung

Gegenwärtig nimmt die Bedeutung der Interaktion zwischen Menschen und sozialen Robotern fortwährend zu. Der Einsatz von sozialen Robotern verbreitet sich immer mehr, im Gesundheitsbereich, im Bildungsbereich, in Hotellerie und Gastronomie, in Shopping Malls, in der Industrie und im Haushalt. Mögliche Einsatzfelder sind u.a. die Information sowie die Aktivierung von älteren Personen, der Support in pflegenahen Dienstleistungen, die Förderung von Kindern und Jugendlichen z.B. mit autistischen Krankheitsbildern, die Hausaufgabenbetreuung bzw. generell die Wissensvermittlung im Aus- und Weiterbildungsbereich oder der Empfang und die Begleitung von Besucherinnen und Besuchern in Firmen, Organisationen oder Einkaufszentren. Trotz der vielen Potenziale, die solche Roboter mit sich bringen, sind jedoch auch zahlreiche Herausforderungen mit deren Einsatz verbunden. Beispielsweise weisen erste Studien mit menschlichen Probanden eine geringe Akzeptanz der Roboter bei den Menschen auf. Um die Akzeptanz zu erhöhen, besteht eine zentrale Herausforderung darin, die Roboter bezüglich Sprache, Mimik, Gestik und Optik so zu konfigurieren, dass Sie den Bedürfnissen der Nutzer gerecht werden. Dies soll das übergreifende Ziel des vorliegenden Projektes sein. Hierzu soll ein lernender Algorithmus entwickelt werden, der Furhat in die Lage versetzt, sich an die Besonderheiten der menschlichen Gesprächspartner*innen zu erinnern und sich diesen anzupassen.

Die forschungsleitende Frage des Projektes ist: Wie kann Furhat Personen kennenlernen und sich an diese erinnern?

Zu diesem Zweck soll mit dem sozialen humanoiden Roboterkopf Furhat gearbeitet werden, der vom schwedischen Unternehmen Furhat Robotics entwickelt wurde und vertrieben wird. Der Furhat ist ein Roboterkopf ohne Gliedmaße mit einem auf die Innenseite projizierten Gesicht, bei dem die äußere Erscheinung wie die Hautfarbe, die Position der Augenbrauen und das Make-up angepasst werden kann. Der Furhat Roboter realisiert dadurch sanfte Gesichtszüge und kann sich durch natürliche Bewegungen wie Kopfschütteln und Kopfnicken direkt an einem Gespräch beteiligen. Mit einer modernen Spracherkennung, einem fortschrittlichen Konversationssystem sowie einer automatisierten Lippensynchronisation werden ausdrucksstarke Charaktere geschaffen.

Zunächst steht der Furhat den Studierenden als Rohversion zur Verfügung. Dieser soll in einem ersten Schritt anhand eines praktischen Szenarios (Use Case) konfiguriert und interaktionsfähig gemacht werden. Im zweiten Schritt sollen regelmäßige Interaktionen mit menschlichen Probanden durchgeführt werden, z.B. wöchentliche Treffen. Mit Hilfe des maschinellen Lernens soll Furhat im Zuge der gemeinsamen Treffen eine Gesprächshistorie aufbauen und sich so an vergangene Gespräche mit bestimmten Personen erinnern können. Hierzu soll ein lernender Algorithmus entwickelt werden, der es Furhat ermöglicht, sich an sein jeweiliges Gegenüber, dessen Sprache, Mimik, Gestik, Vorlieben und Gesprächsinhalte zu erinnern und diese wieder abrufen zu können. Furhat soll demnach die

Fähigkeit entwickeln, als "Gesprächspartner" für Menschen an Informationen und Ergebnisse vergangener Gespräche anzuknüpfen und diese weiterzuführen. Dies kann an einem praktischen Szenario (Use Case), bei dem eine gemeinsame Aufgabe zwischen Mensch und Furhat über mehrere Gesprächstermine hinweg gelöst werden soll, erfolgen. Begleitend dazu wird mit Hilfe von Interviews und Beobachtungen die Mensch-Roboter Interaktion aus menschlicher Perspektive untersucht.

Aufgabenstellung

1. Der als Rohversion bereitgestellte humanoide Roboterkopf Furhat wird anhand eines praktischen Szenarios (Use Case) konfiguriert und interaktionsfähig gemacht.
2. Ein lernender Algorithmus wird entwickelt, der es Furhat ermöglicht, im Zuge regelmäßig stattfindender Interaktionen mit bestimmten Personen im Zuge des unter 1. entwickelten Use Case sich an sein jeweiliges Gegenüber, dessen Sprache, Mimik, Gestik, Vorlieben und Gesprächsinhalte zu erinnern.
3. Mit Hilfe von Interviews und Beobachtungen zu den stattgefundenen Interaktionen wird die Mensch-Roboter Interaktion aus menschlicher Perspektive untersucht.

Bezug zum Thema Data Science

Es wird ein lernender Algorithmus entwickelt.

Verfügbare Ressourcen

Der Roboterkopf Furhat und die notwendige Software um den Roboter zu programmieren wird den Studierenden zur Verfügung gestellt.

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés ist Prüfungsleistung.

Zweites Semester: Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, ist Prüfungsleistung.

Drittes Semester: Erstellung eines Papers, mit ersten quantitativen Ergebnissen ist Prüfungsleistung.

Viertes Semester: Masterarbeit und Kolloquium

Eignungskriterien

Zwingend: Programmierkenntnisse, vorzugsweise in Python

Optional: Kenntnisse im Durchführen von sozialwissenschaftlichen Studien (z.B. Durchführung von Interviews)

Erwerbbarer Kompetenzen

- Erprobung eines humanoiden Roboters
- Verfahren der künstlichen Intelligenz zur Problemlösung
- Mensch-Roboter-Interaktion

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 5 Maschinelle Intelligenz für die Erkennung von Anomalien, sowie die Prädiktion von Interaktionen, anhand von Bewegungsinformationen im Smart Home



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. Thorsten Jungeblut
Projektkontext	Projekt in Zusammenarbeit mit dem KogniHome - Technikunterstütztes Wohnen für Menschen e.V. in Bielefeld; Datensätze und Hardwareressourcen werden gestellt; Anstellung als wissenschaftliche Hilfskraft ist möglich; Über die Projektbearbeitung besteht die Möglichkeit der engen Zusammenarbeit mit Vereinsmitgliedern des KogniHome e.V., u.a. Bethel, Steinel, Hettich, HUM Systems, C&S GmbH uvm.

Abstrakt

Ziel des Projektes ist die Mitentwicklung eines Systems, das mittels KI-Verfahren das Nutzungsverhalten der Bewohner*innen eines Smart Homes erlernt, um anschließend, geeignete Handlungsempfehlungen bereitzustellen oder, bei der Erkennung von Anomalien, geeignete Maßnahmen einzuleiten. Wissenschaftliche Herausforderung des Projektes ist die Anwendung und Evaluation von maschinellen Lernverfahren zum Lernen von Interaktionsmustern basierend auf den gelieferten Daten der Minimalsensorik, sowie der Prädiktion zukünftiger Interaktion und der Anomalieerkennung.

Kurzbeschreibung

Die Digitalisierung dringt in immer weitere Lebensbereiche vor, und die damit verbundene Vernetzung von vielfältigen Komponenten des täglichen Umfelds führt zu einer Situation, in der ein stetig höherer Anspruch an die alltägliche Technik gestellt wird. Die Systeme sollen smarter werden, automatisiert

und eigenständig handeln. Im Idealfall antizipiert die Technik die Bedürfnisse des Menschen und eine Korrektur durch diesen ist nicht mehr nötig. Bedarfsgerechte Mensch-Technik-Interaktion erfordert die Anpassung des Intelligenten Technischen Systems (ITS) an den Nutzungskontext und nicht umgekehrt. Voraussetzung für eine intuitive Interaktion ist daher zuallererst die sichere Erkennung des Nutzungskontextes, d.h. wo befindet sich der Nutzer und welche Handlung führt er gerade durch. Das Wissen über regelmäßig auftretende Interaktionsmuster ermöglicht es dem ITS zukünftige Interaktionen vorherzusagen und Assistenzfunktionen prädiktiv zu steuern. Bildgebende Sensorik (z.B. Kameras, hochauflösende Time-Of-Flight-Sensoren) ermöglichen durch Personen-, Objekt-, Gesten- oder gar Gesichtserkennung eine leistungsfähige Erkennung des Handlungskontextes, bringen aber die Problematik der Erfassung personenbezogener Daten mit sich. Gerade im privaten häuslichen Umfeld, aber auch in Bürogebäuden oder der Produktion kann dies aus datenschutzrechtlichen Gründen unerwünscht sein.

Minimalsensorik, wie z.B. Bewegungs- oder Präsenzmelder erzeugen nicht unmittelbar personenbezogene Daten, sondern erfassen nur punktuelle Informationen über die Anwesenheit von Personen oder Objekten in einem räumlich eingeschränkten Bereich. Für eine umfassende Erfassung des Nutzungskontextes verspricht die Kombination einer Vielzahl von einfachen Sensoren eine ausreichende Erfassung des Nutzungskontextes bei gleichzeitiger Wahrung der Privatsphäre. Ist für eine Assistenzfunktion komplexere Sensorik notwendig (z.B. für eine Sprach- oder Gestenerkennung), dann braucht sie nur (und auch nur genau dann) aktiviert zu werden, wenn die Nutzung des Assistenzsystems zumindest absehbar ist. Die kontinuierliche Erfassung des Nutzungskontextes ermöglicht aber auch das Lernen von regelmäßigen Handlungsmustern. Aus diesen gelernten Handlungsmustern kann das ITS eine zukünftige zu erwartende Interaktion ableiten und Empfehlungen für die Aktivierung von Assistenzfunktionen geben bzw. diese vorbereiten (Prädiktive Steuerung). Darüber hinaus lassen sich aber auch Abweichungen vom potenziellen Regelzustand (Anomalien) erkennen und geeignet darauf reagieren.

Die Ableitung von Handlungsempfehlungen durch das ITS kann in einfachster Implementierung regelbasiert erfolgen (z.B. „Wenn der Bewegungsmelder im Flur aktiviert wird, dann soll das Licht angeschaltet werden“). In Systemen mit einer hohen Anzahl an Sensoren ist dieses aber sehr aufwendig, nicht flexibel anpassbar und schlecht auf wachsende Umgebungen skalierbar. Darüber hinaus liefern verschiedene Sensoren unterschiedlich abstrakte Informationen (z.B. lokale Bewegung (PIR, Ultraschall) oder Bewegung über mehrere Räume hinweg (HF/Mikrowellen), binär „an/aus“ oder Entfernungen) oder gar komplexere Informationen kamerabasierter Systeme (z.B. Anzahl erkannter Personen/Haustiere/Objekte (Staubsaugerroboter/Transportplattform)). Daher gilt es die unterschiedlichen Informationen zu abstrahieren, ggf. zu anonymisieren und in einem ganzheitlichen Ansatz zum Lernen des Nutzungskontextes, zum Abschätzen zukünftiger Interaktion und zum Erkennen von Anomalien zu nutzen. Wissenschaftliche Herausforderung des Projekts ist daher die Anwendung und Evaluation von maschinellen Lernverfahren zum Lernen von Interaktionsmustern basierend auf den gelieferten Daten der Minimalsensorik, sowie der Prädiktion zukünftiger Interaktion und der Anomalieerkennung.

Aufgabenstellung

Der/die Studierende soll in diesem Projekt ein ITS mitentwickeln, das mittels maschinellen Lernverfahren das Nutzungsverhalten der Bewohner*innen eines Smart Homes erlernt, um anschließend, anhand der Echtzeitdaten, geeignete Handlungsempfehlungen bereitzustellen oder, bei der Erkennung von Anomalien, geeignete Maßnahmen einzuleiten. Die Ausarbeitung des konkreten Anwendungsszenarios, zusammen mit dem Team des KogniHome e.V., ist Teil des Projekts.

Zur Veranschaulichung soll folgendes Szenario beschrieben werden: „Sabine (74) steht jeden Morgen zwischen 7 und 8 Uhr auf, geht auf die Toilette und macht sich danach einen Kaffee“. In diesem Handlungsablauf sind drei Aktivitäten zu erkennen: Aufstehen, auf Toilette gehen, Kaffee kochen. In jeder dieser Aktivitäten kann es jetzt zu Handlungsempfehlungen bzw. Anomalien kommen. Beispielsweise könnte das ITS den Gang zur Toilette durch geeignete Beleuchtung vorbereiten, oder bereits die Kaffeemaschine anschalten. Gleichzeitig überwacht das ITS die Aktivitäten und erkennt Abweichungen vom gewohnten Verhalten. Stürzt Sabine z.B. auf dem Weg zur Toilette und steht nicht wieder eigenständig auf, so soll das ITS dies erkennen und z.B. den Notruf oder die Pflegestelle verständigen.

Bezug zum Thema Data Science

Die Evaluation und Anwendung von KI-/ML-Verfahren zur Zustandsüberwachung und Prädiktion sind ein Kernthema der Data Science und werden beispielsweise in den Modulen „Data Mining & Machine Learning“ sowie „Künstliche Intelligenz“ behandelt. Die Aufnahme hochaufgelöster Messdaten (z.B. Verbrauchsdaten) von einer Vielzahl an Sensoren in komplexen Wohnumgebungen stellt hohe Anforderungen an Organisation und Verarbeitung der Daten. Dieses ist Kern des Moduls „Big Data Architekturen“.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Szenarios benötigt werden (Systembeschreibung, logistische Abläufe, relevante Kennzahlen) werden vom KogniHome e.V. bereitgestellt
- Es besteht Zugriff auf die Forschungswohnung des KogniHome e.V.
- Über die Forschungswohnung des KogniHome e.V. stehen umfangreiche Testdatensätze zur Verfügung.
- Der Ansprechpartner im KogniHome e.V. wird über die Projektlaufzeit zur Verfügung stehen
- Benötigte Materialien werden vom KogniHome e.V. bereitgestellt.
- Hardware für das komplexere Maschine Learning ist über das Data Science Lab, das CfADS, sowie dem KI-Rechencluster yourAI der HSBI verfügbar.

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die Konzepte und Strukturen (IoT-Software, Smart Home Protokolle, Schnittstellen etc.) der Forschungswohnung des KogniHome e.V.

Zweites Semester: Erstellung des Systemkonzepts zur Anomalieerkennung zur prädiktiven Steuerung von Assistenzsystemen. Recherche zu relevanten Arbeiten im Themenfeld des Einsatzes von KI-Verfahren für das Lernen von Ereignissequenzen und für die Anomalieerkennung. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Praktische Umsetzung verschiedenster maschinellen Lernverfahren im Bereich des unüberwachten und des (semi-)überwachten Lernens zur Optimierung einer automatisierten Interaktion von Mensch und Maschine, sowie deren Evaluation.

Viertes Semester: Masterarbeit und Kolloquium. Finale Evaluierung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit ersten quantitativen Ergebnissen als Prüfungsleistung.

Eignungskriterien

Zwingend:

- Programmierkenntnisse (vzw. Python)
- Erfahrung mit dem Versions-Kontroll-System „git“

Optional:

- Erfahrung in der Elektronikentwicklung
- Erfahrung im Bereich Smart Home Technologien/IoT-Geräten
- Programmierung von Mikrocontrollern

Erwerbbare Kompetenzen

- Verfahren der künstlichen Intelligenz zur Problemlösung
- Sensornähe Informationsverarbeitung
- Prädiktive Assistenz und Anomalieerkennung (Übertragbarkeit auf industrielle Prozesse)
- Kompetenzen in der Arbeit im Team

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 6 Effiziente Ausführung eines Algorithmus zur Objekterkennung auf Edge-Geräten

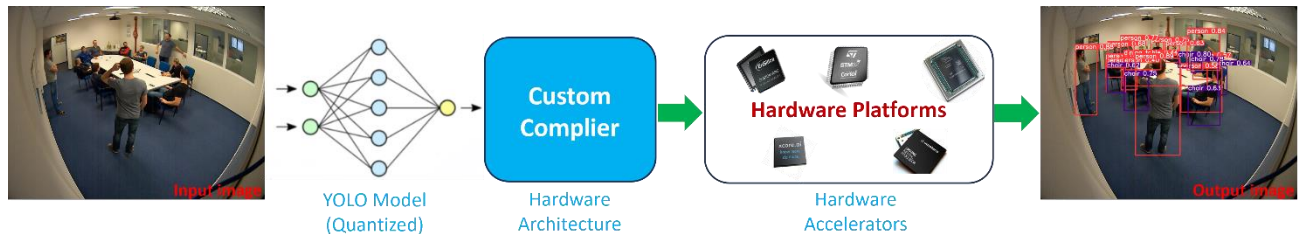


Bild 1: Effiziente Abbildung des YOLO-Vx-Modells auf Randgeräte

Projektübersichtss

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. habil. Thorsten Jungeblut Dr.-Ing. Qazi Arbab Ahmed
Projektkontext	Transferprojekt CareTech OWL in Zusammenarbeit mit einem externen Partner, der Steinel GmbH , im Rahmen des Forschungsprojekts CareTech OWL. Eine parallele Beschäftigung als studentische Hilfskraft (WHK) ist möglich.

Abstrakt

Das Paradigma der künstlichen Intelligenz (KI) hat eine Handvoll traditioneller Computer-Vision-Techniken ersetzt, um die korrekte Ausgabe eines komplexen Systems unter engen Einschränkungen und unterschiedlichen Bedingungen intelligent vorherzusagen. Viele KI-Algorithmen benötigen jedoch aufgrund ihrer massiven Rechenanforderungen enorme Mengen an Hardwareressourcen. Die Hauptaufgabe dieses Projekts ist die Entwicklung eines Frameworks für die effiziente Implementierung von Objekterkennungsalgorithmen auf stark ressourcenbeschränkten Hardwarearchitekturen (Low-End-FPGAs, eingebettete Mikrocontroller). Basierend auf dem Framework werden automatisierte Methoden zur Erkundung des Entwurfsraums geeigneter Kombinationen von Objekterkennungsalgorithmen und Hardware im Sinne eines HW/KI-Co-Designs untersucht. Insbesondere industrielle Anwendungen mit hohen Latenzanforderungen werden angesprochen, wie z. B. die Echtzeit-Anwesenheitserkennung von Menschen.

Kurzbeschreibung

Für die effiziente Ausführung von KI-Algorithmen haben Techniken wie Föderales Lernen (FL) und Cognitive Edge Computing (CEC) bereits die Last des Trainings und der Inferenz neuronaler Netze von der Cloud an den Ort der Datenentstehung verlagert. Die Partitionierung der Ausführung der Anwendung hat einen erheblichen Einfluss auf die Performanz und Ressourceneffizienz des Gesamtsystems. So kann hier zwischen grundlegend unterschiedlichen Ansätzen der dezentralen Merkmalsextraktion möglichst nah am Sensor mit anschließender Fusion gegenüber der zentralen

Verarbeitung in der Cloud unterschieden werden. Der erste Ansatz erfordert leistungsstarke Edge-Hardware für die Vorverarbeitung und bietet potenziell Vorteile für hohe Echtzeitanforderungen. Der zweite Ansatz stellt höhere Anforderungen an die Kommunikationsinfrastruktur, ermöglicht aber möglicherweise die Ausführung komplexerer Netzwerke. Das Hauptziel dieses Projekts ist es daher, automatisierte Methoden zur Erforschung des Entwurfsraums geeigneter Kombinationen von KI-basierten Objekterkennungsmethoden und ressourcenbeschränkter Hardware im Hinblick auf HW/KI-Codesign zu erforschen.

Im Bereich der effizienten Ausführung von KI-Verfahren auf eingebetteten Systemen (Cognitive Edge Computing) wurden in der Vergangenheit große Fortschritte erzielt. Auf allen Ebenen der unterschiedlichen Verarbeitungskonzepte (Cloud, Fog, Edge, Very Edge) in der Produktionskette findet sich eine Vielzahl potenzieller Hardwarearchitekturen und KI-Beschleuniger, die sich in den verfügbaren Systemressourcen (z.B. Leistung oder Stromverbrauch) unterscheiden. Beispiele für relevante Hardwarearchitekturen sind eingebettete Mikrocontroller mit integrierter KI-Beschleunigung, eingebettete GPUs/FPGAs, dedizierte KI-Hardwarebeschleuniger oder High-End-GPUs/FPGAs aus dem HPC-Bereich.

Wir betrachten die gesamte Verarbeitungskette, beginnend mit der sensornahen Vorverarbeitung (very edge), über die Verbindung der Sensorgruppen und ihrer Informationen in der Edge, über lokale dezentrale Cloud-Instanzen (Fog), bis hin zur zentralen Station aller relevanten Informationen (Cloud). Auf jeder Ebene der Verarbeitungskette existieren individuelle Ansätze, um die Ressourceneffizienz von KI-Methoden lokal zu optimieren, z. B. durch Reduzierung der numerischen Präzision (z. B. von 32 auf 16 oder 8 Bit), um eine effizientere Ausführung auf spezialisierter Hardware zu ermöglichen oder den lokalen Speicherplatz zu minimieren. Ziel dieses Projekts ist es, eine optimale Kombination von Hardware und Objekterkennungsalgorithmen (z.B. YOLO) im Rahmen einer ganzheitlichen Entwurfsraumexploration in der Nähe des Sensors zu ermitteln.

Aufgabenstellung

In diesem Projekt entwickeln die Studierenden ein Framework für die effiziente Ausführung von KI-basierten Objekterkennungsalgorithmen auf ressourcenbeschränkten Edge-Geräten. Um die nachhaltige Nutzung der entwickelten Methoden und Frameworks durch den Industriepartner zu unterstützen, sollen in diesem Projekt universell einsetzbare Modelle und automatisierte Entwurfswerkzeuge in Form eines Standard-Entwicklungsbaukastens zur Verfügung gestellt werden. Die erste große Herausforderung ist die Komprimierung des neuesten Algorithmus zur Erkennung von Objekten (z. B. YOLO), um die Modellgröße mit Hilfe von KI-Approximationstechniken mit akzeptabler Genauigkeit zu reduzieren. Die nächste Herausforderung besteht darin, ressourceneffiziente Techniken (Inferenz-Compiler) zu entwickeln, um das komprimierte Modell auf eine geeignete Hardwareplattform abzubilden, z. B. Low-End-FPGAs oder eingebettete (KI-) Mikrocontroller, unter Berücksichtigung der Ressourcennutzung in Bezug auf Fläche, Leistung/Energie und Latenz/Durchsatz, wie in Abbildung 1 dargestellt.

Bezug zum Thema Data Science

Die Bewertung und Anwendung von KI/ML-Methoden im Bereich "Machine Vision", z.B. der Einsatz von CNNs in der Objektklassifizierung, sind ein Kernthema der Data Science und werden beispielsweise in den Modulen "Data Mining & Machine Learning" und "Artificial Intelligence" behandelt. KI-gestützte Bildverarbeitung stellt hohe Anforderungen an die Organisation und Verarbeitung von Daten auf allen Ebenen von IoT-Verarbeitungskonzepten (Edge/Fog/Cloud). Dies ist der Kern des Moduls "Big Data Architekturen". Die Betrachtung des gesamten Systemprozesses vom bildgebenden Sensor bis in die

Cloud erfordert eine ganzheitliche Betrachtung des gesamten Data-Science-Prozesses, die im Modul "Data Science" behandelt wird.

Verfügbare Ressourcen

Sicherstellung der Verfügbarkeit von Daten, Rechenressourcen, Hardware, Anwendungsexperten

1. Die für die Erstellung des Szenarios erforderlichen Informationen (Systembeschreibung, Schnittstellen, Dokumentation, relevante Kennzahlen etc.) werden bereitgestellt
2. Steinel GmbH stellt umfangreiche Testdatensätze aus realen Produktionsumgebungen zur Verfügung
3. Der Ansprechpartner der Steinel GmbH steht Ihnen für die Dauer des Projekts zur Verfügung
4. Die für die Prototypenentwicklung benötigten Komponenten sowie weitere benötigte Materialien werden von der Steinel GmbH zur Verfügung gestellt
5. Hardware für das komplexere maschinelle Lernen steht über das Data Science Lab, das CfADS und den KI-Rechencluster yourAI an der Hochschule Bielefeld zur Verfügung

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfung. Vertrautheit mit dem Konzept KI-basierter Objekterkennungsalgorithmen, insbesondere YOLO, R-CNN, Bilderkennung, Approximationstechniken für neuronale Netze, Hardwareplattformen und Design-Flow-Tools.

Zweites Semester: Erstellung des Systemkonzepts zur Entwurfsraumexploration von kognitiven Edge-Computing-Architekturen. Recherche zu relevanten Arbeiten auf dem Gebiet der KI-basierten Objekterkennungsalgorithmen und der Modellkomprimierung im oben genannten Kontext. Erstellung einer Arbeit, die einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Entwicklung eines ersten Demonstrators und Proof-of-Concept zur Hardwarebeschleunigung für Human Presence Detector (HPD) mit z.B. YOLO-V7. Vergleich des entwickelten Frameworks mit einer klassischen Implementierung.

Viertes Semester: Masterarbeit und Kolloquium. Implementierung und Vergleich verschiedener Kombinationen von KI-basierten Objekterkennungsalgorithmen und Hardwarebeschleunigern. Systematische Bewertung und Untersuchung der Effizienz des entwickelten Rahmens. Abschließende Bewertung durch Vergleich der implementierten Strategien mit den modernsten Methoden. Erstellung einer Arbeit mit den ersten quantitativen Ergebnissen als Prüfung.

Eignungskriterien

Zwingend:

- Gute Kenntnisse in Python, Pytorch
- Gute Kenntnisse von C++

Optional:

- Erfahrung mit Hardware-Design-Flow-Tools
- Programmierung von Mikrocontrollern/FPGAs
- Grundkenntnisse in HDL (Verilog, VHDL)
- Erfahrung mit IoT-Geräten
- Erfahrung mit dem Versionskontrollsystem "git".

Erwerbbarer Kompetenzen

- Ressourceneffiziente Informationsverarbeitung am Edge (eingebettete Mikrocontroller, FPGAs) im Einklang mit dem IoT-Verarbeitungskonzept
- Sensorbezogene Informationsverarbeitung
- AI/ML-basierte Objekterkennungsmethoden
- Nutzung von eingebetteter Hardware zur Beschleunigung von AI/ML-Prozessen

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 7 Hardware-KI-Codesign für die effiziente Ausführung neuronaler Netze

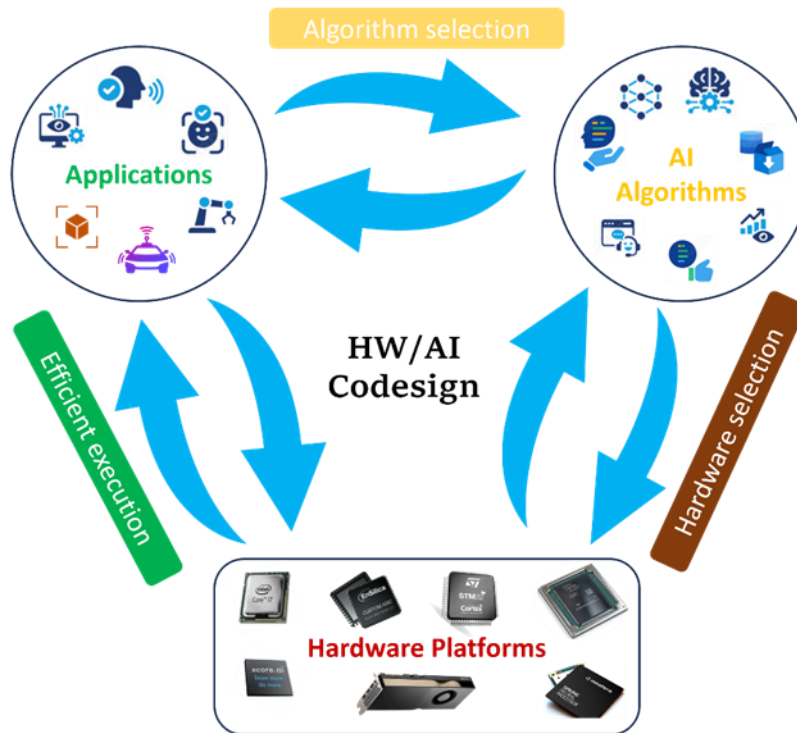


Bild 1: Design-Space-Exploration für Hardware-KI-Codesign

Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. habil. Thorsten Jungeblut, Dr.-Ing. Qazi Arbab Ahmed
Projektkontext	CareTech OWL Transferprojekt in Zusammenarbeit mit dem externen Partner, der Steinel GmbH , im Rahmen des Forschungsprojekts CareTech OWL. Eine parallele Beschäftigung als studentische Hilfskraft (WHK) ist möglich.

Abstrakt

Das Paradigma der künstlichen Intelligenz (KI) hat eine Handvoll traditioneller Computer-Vision-Techniken ersetzt, um die korrekte Ausgabe eines komplexen Systems unter engen Einschränkungen und unterschiedlichen Bedingungen intelligent vorherzusagen. Viele KI-Algorithmen benötigen jedoch aufgrund ihrer massiven Rechenanforderungen enorme Mengen an Hardwareressourcen. Ziel dieses Projekts ist die Entwicklung eines Frameworks für die effiziente Implementierung von KI- und ML-Algorithmen auf stark ressourcenbeschränkten Hardwarearchitekturen. Basierend auf dem Framework werden automatisierte Methoden zur Exploration des Entwurfsraums geeigneter Kombinationen von KI-Algorithmen und Hardware im Sinne von HW/KI-Co-Design erforscht.

Insbesondere industrielle Anwendungen mit hohen Latenzanforderungen werden angesprochen. Dabei wird nicht nur die gesamte Kette als linearer Prozess vom Modelltraining bis zur Inferenz berücksichtigt, sondern auch der Einfluss der Wahl möglicher Hardwarekonfigurationen auf die ursprüngliche Modellentwicklung.

Kurzbeschreibung

Für die effiziente Ausführung von KI-Algorithmen haben Techniken wie Föderales Lernen (FL) und Cognitive Edge Computing (CEC) bereits die Last des Trainings und der Inferenz neuronaler Netze von der Cloud an den Rand der Datenentstehung verlagert. Die Partitionierung der Ausführung der Anwendung hat einen erheblichen Einfluss auf die Performance und Ressourceneffizienz des Gesamtsystems. So kann hier zwischen grundlegend unterschiedlichen Ansätzen der dezentralen Merkmalsextraktion möglichst nah am Sensor mit anschließender Fusion versus zentraler Verarbeitung in der Cloud unterschieden werden. Der erste Ansatz erfordert leistungsstarke Edge-Hardware für die Vorverarbeitung und bietet potenziell Vorteile für hohe Echtzeitanforderungen. Der zweite Ansatz stellt höhere Anforderungen an die Kommunikationsinfrastruktur, ermöglicht aber möglicherweise die Ausführung komplexerer Netzwerke. Das Hauptziel dieses Projekts ist es daher, automatisierte Methoden zur Exploration des Entwurfsraums geeigneter Kombinationen von KI-Methoden und ressourcenbeschränkter Hardware im Sinne von HW/KI-Codesign zu erforschen.

Im Bereich der effizienten Ausführung von KI-Verfahren auf eingebetteten Systemen (Cognitive Edge Computing) wurden in der Vergangenheit große Fortschritte erzielt. Auf allen Ebenen der unterschiedlichen Verarbeitungskonzepte (Cloud, Fog, Edge, Very Edge) in der Produktionskette findet sich eine Vielzahl potenzieller Hardwarearchitekturen und KI-Beschleuniger, die sich in den verfügbaren Systemressourcen (z.B. Leistung oder Stromverbrauch) unterscheiden. Beispiele für relevante Hardwarearchitekturen sind eingebettete Mikrocontroller mit integrierter KI-Beschleunigung, eingebettete GPUs/FPGAs, dedizierte KI-Hardwarebeschleuniger oder High-End-GPUs/FPGAs aus dem HPC-Bereich.

Wir betrachten die gesamte Verarbeitungskette, beginnend mit der sensornahen Vorverarbeitung (very edge), über die Verbindung der Sensorgruppen und ihrer Informationen in der Edge, über lokale dezentrale Cloud-Instanzen (Fog), bis hin zur zentralen Station aller relevanten Informationen (Cloud). Auf jeder Ebene der Verarbeitungskette existieren individuelle Ansätze, um die Ressourceneffizienz von KI-Prozessen lokal zu optimieren, z. B. durch Reduzierung der numerischen Präzision (z. B. von 32 auf 16 oder 8 Bit), um eine effizientere Ausführung auf spezialisierter Hardware zu ermöglichen oder den lokalen Speicherplatz zu minimieren. Ziel ist es, eine optimale Kombination von Hard- und Software (d.h. KI-Algorithmus) im Rahmen einer ganzheitlichen Entwurfsraumexploration auf allen Verarbeitungsebenen zu ermitteln.

Aufgabenstellung

In diesem Projekt werden die Studierenden ein automatisiertes Toolkit für die Entwurfsraumexploration entwickeln, das den Benutzer beim HW/KI-Codesign unterstützen kann. Um den nachhaltigen Einsatz der entwickelten Methoden und Frameworks durch den Industriepartner zu unterstützen, sollen in diesem Projekt universell einsetzbare Modelle und automatisierte Entwurfswerkzeuge in Form eines Entwicklungsbaukastens zur Verfügung gestellt werden, der von den konkreten Anwendungsfällen abstrahiert und es ermöglicht, anhand eines zukünftigen Anwendungsszenarios Empfehlungen für eine optimale Kombination von KI-Methoden und geeigneter (Edge-)Hardwarearchitektur zu erhalten. Ziel ist es, dem Anwender vorläufige Modelle und automatisierte Entwurfswerkzeuge zur Verfügung zu stellen, die Empfehlungen für eine Kombination

aus KI-Methoden und geeigneter (Edge-) Hardwarearchitektur auf Basis eines gegebenen Anwendungsszenarios und Bewertungsmaßnahmen ermöglichen.

Bild 1 veranschaulicht das allgemeine Konzept der Exploration des Entwurfsraums für das Hardware-AI-Codesign. Die erste Hauptherausforderung ist die Komprimierung/Optimierung des neuronalen Netzes, um die Größe zu reduzieren und dennoch eine akzeptable Genauigkeit zu erreichen, indem KI-Approximationstechniken verwendet werden. Die nächste Herausforderung besteht in der Entwicklung ressourceneffizienter Techniken zur Abbildung des komprimierten (KI-)Modells auf eine geeignete Hardwareplattform wie GPUs, CPUs, ASICs und FPGAs unter Berücksichtigung der Ressourcennutzung in Bezug auf Fläche, Leistung/Energie und Latenz/Durchsatz.

Bezug zum Thema Data Science

Die Evaluierung und Anwendung von KI/ML-Methoden im Bereich "Maschinelle Sehens", z.B. der Einsatz von CNNs in der Objektklassifikation, sind ein Kernthema der Data Science und werden z.B. in den Modulen "Data Mining & Machine Learning" und "Künstliche Intelligenz" behandelt. KI-gestützte Bildverarbeitung stellt hohe Anforderungen an die Organisation und Verarbeitung von Daten auf allen Ebenen von IoT-Verarbeitungskonzepten (Edge/Fog/Cloud). Dies ist der Kern des Moduls "Big Data Architekturen". Die Betrachtung des gesamten Systemprozesses vom bildgebenden Sensor bis zur Cloud erfordert eine ganzheitliche Betrachtung des gesamten Data Science Prozesses, die im Modul "Data Science" behandelt wird.

Verfügbare Ressourcen

Sicherstellung der Verfügbarkeit von Daten, Rechenressourcen, Hardware, Anwendungsexperten

- Sicherstellung der Verfügbarkeit von Daten, Rechenressourcen, Hardware, Anwendungsexperten
- Bereitstellung der für die Erstellung des Szenarios erforderlichen Informationen (Systembeschreibung, Schnittstellen, Dokumentation, relevante Kennzahlen etc.)
- Steinel GmbH stellt umfangreiche Testdatensätze aus realen Produktionsumgebungen zur Verfügung
- Der Ansprechpartner der Steinel GmbH steht während der gesamten Projektlaufzeit zur Verfügung
- Die Steinel GmbH stellt die für die Prototyp-Entwicklung benötigten Komponenten sowie weitere benötigte Materialien zur Verfügung
- Hardware für das komplexere maschinelle Lernen ist über das Data Science Lab, den CfADS und den KI-Computing-Cluster yourAI an der Hochschule Bielefeld verfügbar

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in das Konzept der KI-Algorithmen (CNN, DNN), insbesondere der Objekterkennung und Bilderkennung, der verwendeten Hardware-Plattformen und Design-Flow-Tools.

Zweites Semester: Erstellung des Systemkonzepts für die Design Space Exploration von Cognitive Edge Computing Architekturen. Recherche zu relevanten Arbeiten im Bereich der Nutzung von KI/ML-Methoden zur Datenverarbeitung und Modellkompression im oben genannten Kontext. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Entwicklung eines ersten Demonstrators und Proof-of-Concept für die Hardware-Beschleunigung einer relevanten Anwendung. Vergleich eines KI-Prozesses mit einer klassischen Implementierung.

Viertes Semester: Masterarbeit und Kolloquium. Implementierung und Vergleich verschiedener Kombinationen von KI/ML-Verfahren und Hardwarebeschleunigern. Systematische Evaluierung und Erforschung der Effizienz der Kombinationen. Vergleich verschiedener Verarbeitungskonzepte (Embedded AI, Edge, Cloud). Abschließende Bewertung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit den ersten quantitativen Ergebnissen als Prüfung.

Eignungskriterien

Mandatory:

- Gute Kenntnisse in Python, Pytorch
- Gute Kenntnisse von C++

Optional:

- Erfahrung mit Hardware-Design-Flow-Tools
- Programmierung von Mikrocontrollern/FPGAs
- Grundkenntnisse in HDL (Verilog, VHDL)
- Erfahrung mit IoT-Geräten
- Erfahrung mit dem Versionskontrollsystem "git".

Erwerbbare Kompetenzen

- Ressourceneffiziente Informationsverarbeitung am Edge (eingebettete Mikrocontroller, FPGAs) im Einklang mit dem IoT-Verarbeitungskonzept
- Sensorbezogene Informationsverarbeitung
- AI/ML-Methoden
- Nutzung von eingebetteter Hardware zur Beschleunigung von AI/ML-Prozessen

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 8 Cognitive Edge Computing für KI-/ML-basierte Oberflächeninspektion

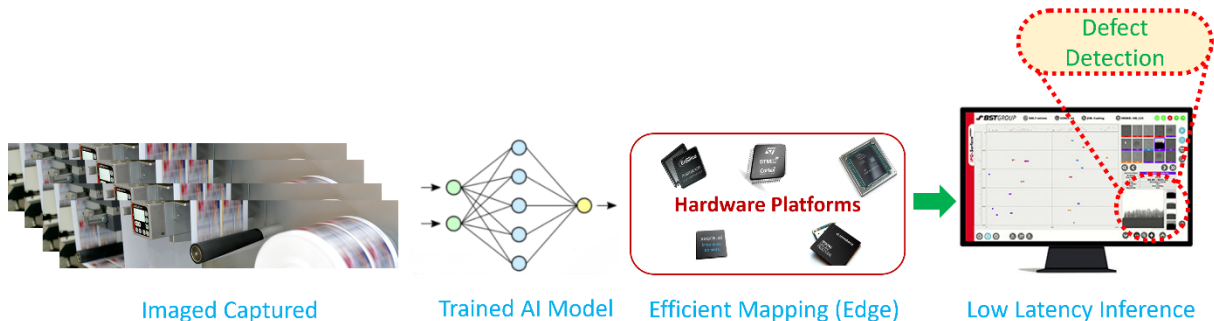


Bild 1: Effiziente Oberflächeninspektion mit AI on edge

Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externem Partner
Projektverantwortung	Prof. Dr.-Ing. Thorsten Jungeblut Dr. Ing. Qazi Arbab Ahmed
Projektkontext	<p>CareTech OWL Transferprojekt in Zusammenarbeit mit einem externen Partner, der BST Group, als Teil des CareTech OWL Forschungsprojekts.</p> <p>Das Partnerunternehmen BST Group stellt umfangreiche Testdatensätze sowie einen prototypischen Hardwareaufbau zur Verfügung. Eine parallele Beschäftigung als wissenschaftliche Hilfskraft (WHK) ist möglich.</p> <p>Die BST Group ist sehr an der Rekrutierung und langfristigen Beschäftigung von Nachwuchskräften interessiert.</p>

Abstrakt

Das Paradigma der künstlichen Intelligenz (KI) hat eine Handvoll traditioneller Computer-Vision-Techniken ersetzt, um auf intelligente Weise die korrekte Ausgabe eines komplexen Systems unter engen Beschränkungen und wechselnden Bedingungen vorherzusagen. Ziel des Projektes ist die Entwurfsraumexploration von KI-/ML-Hardwarebeschleunigern in der Oberflächeninspektion. Im Zentrum steht nicht nur das Lernen der Modelle auf HPC-Systemen, sondern auch die effiziente Ausführung (Inferenz) auf eingebetteter Hardware. Ergebnis der Entwurfsraumexploration ist die Partitionierung der Anwendung, d.h. welche KI-Verfahren können auf dem Sensor ausgewertet werden, welche Verfahren können über Edge-Hardware (z.B. eingebettete GPU/FPGA) beschleunigt werden und welche erfordern leistungsfähige HPC-Hardware in der Cloud.

Kurzbeschreibung

Die Entwicklungen im Bereich der intelligenten technischen Systeme (ITS) führen derzeit zu einem Umbruch in der gesamten Wertschöpfungskette der industriellen Produktion. Die zunehmende

Leistungsfähigkeit der Informationsverarbeitung bietet für das Anwendungsgebiet der industriellen Bildverarbeitung viele neue Möglichkeiten, an denen Mensch und Maschinen bisher an ihre Grenzen stoßen. Insbesondere die Nutzung Maschinellem Lernverfahren und Methoden der künstlichen Intelligenz versprechen bisher nicht gekannte Möglichkeiten beispielsweise in der Objektklassifikation oder der visuellen Qualitätskontrolle. Etwa seit dem Jahr 2010 werden mit Deep Neuronal Networks (DNN) bzw. Convolutional Neuronal Networks (CNN) bedeutende Fortschritte erzielt.

Die Leistungsfähigkeit etablierter KI-/ML-Verfahren basiert aber bisher meist auf der Nutzung leistungsfähiger dezentraler Rechenressourcen (High-Performance Computing) in der Cloud. Nicht nur für das Lernen der Modelle, sondern auch für deren Ausführung (Inferenz) ist der Anwender auf diese leistungsfähigen Ressourcen angewiesen. Die im Bereich in der industriellen Bildverarbeitung auftretenden Anforderungen weichen jedoch aufgrund hoher Ansprüche an niedrige Latenz, Echtzeitfähigkeit oder Datenlokalität von den durch große Anbieter von KI-Know-How adressierten Fragestellungen ab. Auch Fragestellungen wie Wartbarkeit, Zertifizierbarkeit oder Privacy erschweren die Nutzung populärer Modelle wie DNNs oder CNNs in der Cloud.

Aber auch im Bereich der effizienten Ausführung von KI-/ML-Verfahren auf eingebetteten Systemen (**Cognitive Edge Computing**) wurden in der Vergangenheit große Fortschritte gemacht. Auf allen Ebenen der unterschiedlichen Verarbeitungskonzepte in der vernetzten Produktion (Edge/Fog/Cloud-Computing) finden sich geeignete Hardwarebeschleuniger, die auf einen geeigneten Kompromiss zwischen den Systemressourcen wie benötigter Leistungsfähigkeit (z.B. Klassifikationsgenauigkeit), Leistungsaufnahme/Energiebedarf oder Datendurchsatz/Latenz abzielen. Beispiele für relevante Hardwarearchitekturen sind eingebettete Mikrocontroller, Embedded GPUs, eingebettete FPGAs oder dedizierte KI-Hardwarebeschleuniger.

Die BST Group entwickelt Systeme für die Oberflächeninspektion (z.B. für die Batteriezellproduktion), die unter anderem typische Fehler Fertigungsprozess (z. B. Beschichtungsaussetzer) verlässlich und umgehend erkennen können. Durch ihren modularen Aufbau können BST Systeme an unterschiedlichste Anwendungen perfekt angepasst werden. Bilderfassung und Fehlererkennung erfolgen in Echtzeit. Die Systeme sind anwendbar auf uniforme, texturierte sowie bedruckte Oberflächen. Die umgehende und automatische Erkennung und Anzeige auch kleinster Fehler und Abweichungen ermöglichen eine schnelle und sichere Anpassung des Prozesses zur Ausschussvermeidung. Das System besteht aus mehreren optischen Sensoren, deren Sensordaten geeignet fusioniert werden. Die Anwendung maschineller Lernverfahren verspricht hier eine höhere Leistungsfähigkeit in der Fehlererkennung und Optimierung des gesamten Prozesses. Mit Bahnlaufgeschwindigkeiten von mehreren hundert Metern pro Sekunde stellen sich allerdings sehr hohe Anforderungen an die Geschwindigkeit bzw. Latenz der eingesetzten KI-/ML-Verfahren. Bild 1 zeigt einen Überblick über die vorgeschlagene Methodik.

Aufgabenstellung

Ziel des Projektes ist die Entwurfsraumexploration von KI-/ML-Hardwarebeschleunigern für den Einsatz in der Oberflächeninspektion. Im Zentrum steht nicht nur das Lernen der Modelle auf HPC-Systemen, sondern auch die effiziente Ausführung (Inferenz). Die Betrachtung der gesamten Systemarchitektur vom intelligenten Sensor über Edge-Gateways zur lokalen Datenvorverarbeitung bis hin zur Cloud-Infrastruktur stellt hohe Anforderungen an die Abbildung der KI-Verfahren auf die Hardware. Die Wahl der Zielarchitektur hat wiederum Einfluss auf die Wahl und das Training der Modelle. Hieraus ergibt sich ein iterativer Zyklus (Model-to-Inference-to-Model), der sowohl die Auswahl geeigneter KI-Verfahren, als auch die Bestimmung der (Hyper-)Parameter des Modells beinhaltet. Diesen Ansatz einer ganzheitlichen Entwurfsraumexploration kann man auch in Anlehnung an den etablierten Begriff HW/SW-Co-Design als HW-/KI-Co-Design bezeichnen. Die Entwurfsraumexploration soll dabei unterschiedlichen Entwurfsziele, wie beispielsweise

Klassifikationsgenauigkeit, Latenz oder Ressourcenbedarf der Hardware berücksichtigen. Ergebnis der Entwurfsraumexploration ist die Partitionierung der Anwendung, d.h. welche KI-Verfahren können direkt auf dem Sensor ausgewertet werden (z.B. durch Dimensionsreduktion oder Feature-Extraktion), welche Verfahren können über Edge-Hardware (beispielsweise eingebettete GPU/FPGA oder dedizierte TPU-Beschleuniger) beschleunigt werden und welche erfordern leistungsfähige HPC-Hardware in der Cloud.

Bezug zum Thema Data Science

Die Evaluation und Anwendung von KI-/ML-Verfahren im Bereich des „maschinellen Sehens“ z.B. den Einsatz von CNNs in der Objektklassifikation sind ein Kernthema der Data Science und werden beispielsweise in den Modulen „Data Mining & Machine Learning“ sowie „Künstliche Intelligenz“ behandelt. Die KI-gestützte Bildverarbeitung stellt hohe Anforderungen an Organisation und Verarbeitung der Daten auf allen Ebenen der IoT-Verarbeitungskonzepte (Edge/Fog/Cloud). Dieses ist Kern des Moduls „Big Data Architekturen“. Die Betrachtung des gesamten Systemprozesses vom bildgebenden Sensor bis zur Cloud erfordert eine ganzheitliche Betrachtung des vollständigen Data Science Prozesses, welches im Modul „Data Science“ behandelt wird.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Szenarios benötigt werden (Systembeschreibung, Schnittstellen, Dokumentationen, relevante Kennzahlen etc.) werden von der BST Group bereitgestellt
- Die BST Group stellt umfangreiche Testdatensätze aus realen Produktionsumgebungen zur Verfügung
- Der Ansprechpartner der BST Group wird über die Projektlaufzeit zur Verfügung stehen
- Die für den prototypischen Aufbau benötigten Komponenten sowie sonstiges benötigtes Material wird von der BST Group bereitgestellt
- Hardware für das komplexere Maschine Learning ist über das Data Science, das CfADS sowiedem KI-Rechencluster yourAI der FH Bielefeld verfügbar

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die das Konzept der Oberflächeninspektionssysteme von BST, die Schnittstellen der zur Verfügung gestellten intelligenten Sensoren.

Zweites Semester: Erstellung des Systemkonzepts zur Entwurfsraumexploration von Cognitive-Edge-Computing-Architekturen. Recherche zu relevanten Arbeiten im Themenfeld des Einsatzes von KI-/ML-Verfahren für die Sensordatenverarbeitung im o.g. Kontext. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Entwicklung eines ersten Demonstrators und Proof-of-Concept zur Hardwarebeschleunigung einer Anwendung zur Oberflächeninspektion von BST. Vergleich eines KI-Verfahrens mit einer klassischen Realisierung.

Viertes Semester: Masterarbeit und Kolloquium. Implementierung und Vergleich unterschiedlicher Kombinationen aus KI-/ML-Verfahren und Hardwarebeschleunigern. Systematische Evaluation und Exploration der Effizienz der Kombinationen. Vergleich verschiedener Verarbeitungskonzepte (Embedded AI, Edge, Cloud). Finale Evaluierung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit ersten quantitativen Ergebnissen als Prüfungsleistung.

Eignungskriterien

Zwingend:

- Gute Erfahrung mit Python, Pytorch
- Gute Kenntnisse in C++

Optional:

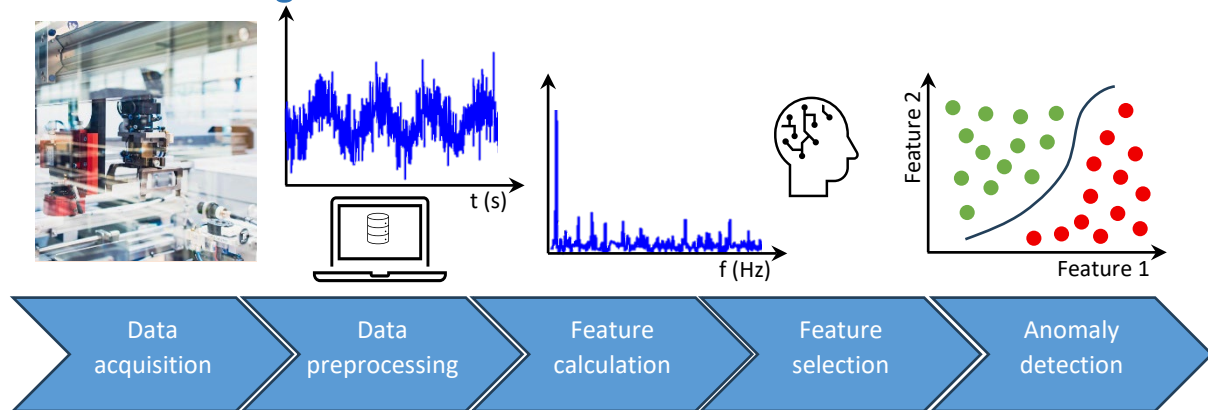
- Erfahrung mit IoT-Geräten
- Grundkenntnisse in HDL (Verilog, VHDL)
- Erfahrung mit Hardware-Design-Flow-Tools
- Programmierung von Mikrocontrollern/FPGAs

Erfahrung mit IoT-Geräten Erwerbbarer Kompetenzen

- Ressourceneffiziente Informationsverarbeitung auf den verschiedenen Ebenen (Edge, Fog, Cloud) im Sinne des IoT-Verarbeitungskonzepts
- Sensornaher Informationsverarbeitung
- KI-/ML-Verfahren
- Einsatz eingebetteter Hardware zur Beschleunigung von KI-/ML-Verfahren

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 9 Systematisches Daten- und Feature-Engineering für die KI-basierte Zustandsüberwachung in der industriellen Automatisierung



Project overview

Number of Students	1
Project Type	Master student project
Project Owner	Dr.-Ing. Annika Junker Prof. Dr.-Ing Wolfram Schenck Dr.-Ing. Heiko Stichweh (Lenze)
Project Context	Project within the Center for Applied Data Science Gütersloh (CfADS) with the external partner Lenze in Groß Berkel

Abstract

This Master student project focuses on developing AI-driven methods for machine condition monitoring using data from frequency inverters, eliminating the need for additional sensors. Conducted in close collaboration with Lenze and embedded within the Center for Applied Data Science Gütersloh (CfADS), the project aims to develop innovative, physics-informed data processing and AI techniques for predictive maintenance.

Short description

The increasing digitalization in industrial environments and the rise of Industry 4.0 have created new opportunities for automated machine condition monitoring. Instead of adding costly sensors, modern drive systems already provide rich operational data. Lenze's advanced frequency inverters deliver high-quality signals that can be leveraged for sensorless monitoring, reducing hardware costs and simplifying integration. Combined with AI-based anomaly detection, this approach can significantly improve machine availability, reduce downtime, and optimize maintenance strategies. This research project aims to develop a physics-informed AI system for anomaly detection in storage and retrieval machines, using only inverter data. Key objectives include:

- Extracting time- and frequency-domain features with drivetrain knowledge.
- Preprocessing and segmenting data for comparable motion patterns.
- Building robust ML models for reliable anomaly detection.
- Development of robust machine learning models for reliable anomaly detection based on representative training datasets.

The expected outcome is a scalable solution for predictive maintenance in industrial automation.

Task definition

The objective of this research project is to design, implement, and critically evaluate a robust data and feature processing pipeline for AI-supported condition monitoring systems. The methods are designed to work with various machines and applications. The following work packages are planned:

Data Preprocessing and Feature Engineering: The student will systematically prepare operational data collected from frequency inverters, ensuring reliable performance across diverse operational motions, such as acceleration, constant speed, and deceleration. Preprocessing steps should guarantee that the resulting dataset allows for meaningful comparison and supports transfer to similar systems.

Feature Selection and Construction: Both data-driven features (e.g., statistical measures, time-series characteristics, and patterns learned directly from data) and physics-informed features (e.g., metrics derived from drivetrain models, domain knowledge, or physical laws) should be considered. Domain knowledge is to be explicitly integrated in the construction of physics-motivated features, leveraging expertise in drive technology and system dynamics.

Model Development and Robustness: Develop and train machine learning models for anomaly detection using the engineered feature sets. Robustness must be validated by applying the trained models to data from an additional, comparable machine. The student should ensure that both feature processing and model performance are transferable, not narrowly tailored to a single device or set of operating conditions. To avoid overfitting and improve model generalization, feature reduction techniques such as principal component analysis should be implemented, efficiently selecting the most relevant features for anomaly detection.

Comparative Analysis: The student will conduct a critical comparison between data-driven and physics-informed features, discussing their respective strengths and limitations. This includes:

- Assessing performance and robustness of models built on each approach
- Evaluating transferability to new machines or operational scenarios
- Considering factors such as interpretability, computational requirements, and resilience to changing system conditions

Documentation and Critical Reflection: The methodology, selection of algorithms, and all major decisions must be thoroughly documented. The student is expected to provide a reasoned discussion of the results, limitations, and practical implications, particularly with respect to the transferability and reliability of the developed system.

Reference to the topic of data science

This project addresses key aspects of data science, including data processing, particularly the preprocessing of raw data, and the targeted computation and selection of relevant features for subsequent AI-based condition monitoring. In addition, approaches from physics-informed machine learning will be explored, as the integration of physical domain knowledge into the machine learning pipeline represents an exciting and forward-looking research field.

Available Resources

- **Data access:** Availability of realistic operational data from modern frequency inverters across different machine types and applications. This data forms the basis for developing and training anomaly detection methods.
- **Expert support:** Guidance from specialists in automation and drive technology and data science available to advise on data interpretation, modeling, and implementation questions.

Project plan

First Semester:

- Conduct a literature review on condition monitoring and physics-informed machine learning.
- Familiarize with the operation of automation systems using frequency inverters.
- Perform data exploration and preprocessing for an exemplary system.
- Write a research exposé summarizing the findings.

Second Semester:

- Implement the calculation and selection of relevant features from both data-driven and physically motivated domains for the exemplary system of the first semester.
- Develop and evaluate a machine learning model to identify abnormal machine states.
- Write a paper of a literature review as the mandatory artefact.

Third Semester:

- Test the transferability of the developed model to at least two other systems: assess to what extent the model can be applied to comparable, yet previously unseen installations.
- Identify features critical for model quality reducing complexity and improving generalizability.
- Write a scientific paper with first results as a mandatory artifact.
- Define specific research questions to be addressed within the scope of the Master thesis.

Fourth Semester:

- Master thesis and colloquium.

Necessary Competencies

Mandatory:

- Bachelor's degree in an engineering-related field.
- Interest in both data-driven and physically motivated modeling.
- Experience with the implementation of machine learning methods with Python.
- Strong passion for exploring complex technical challenges in depth.

Acquirable Competencies

- Application of machine learning methods to real industrially relevant problems.
- Application of physics-informed machine learning, which is a particularly exciting and forward-looking research field.

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 10 AI-Based Network Optimization and Strategic Internationalization for CTL



Project overview

Number of Students	1
Project Type	Project with external partner (Softward GmbH, 100% subsidiary of CTL AG)
Project Owner	Prof. Dr. Christian Schwede (Softward GmbH and CTL AG, Contact: Malte Konersmann, Softward GmbH, Tobias Jüngling, CTL AG)
Project Context	This project is conducted in close collaboration with Softward GmbH, a wholly-owned subsidiary of Cargo Trans Logistik (CTL) AG and the group's digital innovation specialist. Softward is responsible for the ongoing development of the CTL network's digital infrastructure and service optimization. The project aims to leverage AI-based methods to identify untapped optimization potential in CTL's network management and develop actionable strategies for international expansion. Students Assistant Job is included.

Abstract

CTL AG operates a complex hub-and-spoke groupage network with six main hubs across Germany. Transport routes and planning are currently based on rules derived from postal codes and static tables. With internationalization moving to the forefront of CTL's strategy, Softward GmbH seeks to harness

AI-driven optimization techniques to flexibly plan routes and depot utilization, simulate various expansion scenarios, and support strategic decision-making. The objective is to create a dynamic planning system that increases efficiency, scalability, and service quality—providing measurable value as the CTL network enters international markets.

Short description

Currently, CTL's route planning relies on a deterministic system based on postal code zoning and fixed product/service definitions (e.g., Express, Fixed Date, Classic). With growing cross-border business, an adaptive AI-driven routing framework is vital. This project, led by Softward GmbH, investigates the use of machine learning, reinforcement learning, and advanced optimization algorithms to dynamically analyze, predict and optimize network operations. Key questions to be addressed:

- Which AI methods unlock the most impactful network optimizations (e.g., lead time reduction, better resource allocation, increased depot utilization) in the CTL setting?
- How do various internationalization strategies (new international hubs/multihubs/national hubs as gateway) affect network performance, and how can AI-driven simulation steer this process?
- How can a dynamic planning platform be designed to integrate with CTL/Softward systems and scale with international growth requirements?

Task definition

- Analyze current routing and network planning (based on postal codes and fixed service rules)
- Curate and model relevant network datasets (shipment flows, depot capacities, service times, cross-border factors) e.g. in a digital twin.
- Design and implement AI-based optimization algorithms for flexible, scenario-driven network planning
- Simulate and benchmark several international expansion scenarios (adding new hubs, cross-border routing, dynamic time slot optimization)
- Evaluate efficiency gains versus the rule-based baseline; quantify benefits for internationalization
- Develop a prototype or proof-of-concept; propose roadmap for integration into CTL's operational landscape

Reference to the topic of data science

- Data exploration & preprocessing
- Graph/network modeling, combinatorial & stochastic optimization
- Predictive modeling, simulation (e.g., reinforcement learning, scenario analysis)
- Integration of AI models into operational logistics IT (APIs, microservices)

Available resources

- Access to comprehensive CTL routing, shipment, and depot data via Softward

- Mentorship and subject matter support from Softward and CTL AG (operational, technical, strategic perspectives)
- Modern data analysis and AI/ML toolkits, simulation capabilities
- API and system access for realistic prototyping
- Opportunities for workshops with logistics and IT experts at Softward and CTL

Project plan

- **First semester:** Network data exploration, evaluation of current system, identification of key gaps and AI opportunities. Research Expose.
- **Second semester:** Literature review, AI method benchmarking, initial model and simulation development, building prototype. Paper on literature review results.
- **Third semester:** Validation of optimization techniques and prototype, comparison of different internationalization strategies, interim results and recommendations. Paper on quantitative results.
- **Fourth semester:** Prototype refinement, documentation, integration planning, delivery and defense of Master thesis

Necessary competencies

Mandatory: Solid knowledge of data science, statistics, machine learning, React, MySQL, MSSQL, Azure, DOTnet

Optional: Experience in network/graph modeling, optimization algorithms, logistics or transport sector, software/API integration, Python

Acquirable competencies

- Applying and evaluating modern AI/optimization in a real logistics context
- Building and validating scalable simulation and planning models
- Cross-disciplinary collaboration (data science, logistics, digital transformation)
- Business communication and presentation of data-driven strategic impact

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitel: 11 Multi-Agenten-Steuerung der IoT-Factory: Large Language Models und Reinforcement Learning am realen Usecase



Projektübersicht

Anzahl Studierende	1
Art	Studierendenprojekt
Projektverantwortung	Prof. Dr. rer. oec. Pascal Reusch
Projektkontext	Das Projekt findet in der IoT-Factory des CfADS (https://www.hsbi.de/iium/cfads/projekte/iot-factory) statt und wird durch den Laboringenieur Roman Sliwinski und weiteren Mitarbeitenden des CfADS betreut. Außerdem steht das Data-Analytics-Cluster des CfADS zur Verfügung.

Abstrakt

Für die IoT-Factory am CfADS soll eine Alternative zu der bisherigen zentralen MES-Steuerung in Form eines Multi-Agenten-Systems entwickelt werden. Dazu soll OptiFlow, eine haus eigene Python-basierte Steuerungsplattform, die unter anderem bereits AGV-Flottenkoordination und Materialflusssteuerung in der IoT-Factory realisiert, zu einem agentischen System weiterentwickelt werden, in dem jede physische Komponente als autonomer Agent modelliert wird, der kooperativ über die optimale Auftragsabwicklung verhandelt. Je nach Schwerpunkt sollen Large Language Model- (LLM-) basierte Agentenkommunikation, lernende Agenten mittels Reinforcement Learning oder Ansätze der Agentenökonomie erforscht und an der realen Anlage evaluiert werden.

Kurzbeschreibung

Mit der IoT-Factory bietet das Center for Applied Data Science (CfADS) der Hochschule Bielefeld eine vollständige modulare Fertigungslinie für die praxisnahe Untersuchung aktueller Forschungsfragen. Durch eine große Bandbreite an Sensoren innerhalb der einzelnen Module und die Datenverarbeitungskapazitäten des Data-Analytics Clusters des CfADS eröffnet die IoT-Factory eine Fülle neuer Methoden zur Beantwortung aktueller Fragestellungen.

Die aktuelle Steuerung der Anlage basiert auf einem zentralen MES (Festo MES4.0) in Verbindung mit OptiFlow, einer am CfADS entwickelten Python-basierten Steuerungsschicht, die Materialfluss, AGV-Flottenlogik, Stationskoordination und Anlagenverfahren in einer einheitlichen Plattform bündelt. Beide Systeme arbeiten deterministisch: Auftragspriorisierung, Routing und Ressourcenzuweisung folgen statischen und ineffizienten Regeln und können nicht eigenständig auf veränderte Bedingungen reagieren. Ziel des Projekts ist es, OptiFlow von einer deterministischen Steuerungsplattform zu einem

agentischen System weiterzuentwickeln. Die derzeit statische Logik für Materialfluss, AGV-Koordination und Stationssteuerung soll schrittweise durch kooperative Agenten ersetzt werden, die ihre Entscheidungen autonom und datengetrieben treffen. Die physischen Komponenten der IoT-Factory, Roboterzellen, Transportmodule, Lager, Werkstückträger und die Werkstücke selbst, sollen dabei als autonome, kommunikationsfähige Agenten innerhalb von OptiFlow repräsentiert werden und kooperativ über die optimale Auftragsabwicklung verhandeln.

Aufgabenstellung

Dieses Forschungsprojekt verfolgt einen grundlegend neuen Ansatz: Anstelle der zentralen MES-Steuerung soll ein Multi-Agenten-System (z.B. mit OFacT) angewendet werden, in dem jede physische Komponente der IoT-Factory als autonomer, kommunikationsfähiger Agent agieren soll. Diese Agenten sollen kooperativ aushandeln, wie eingehende Fertigungsaufträge optimal in der IoT-Factory abgearbeitet werden.

Im Rahmen des Forschungsprojekts können verschiedene Schwerpunkte gesetzt werden. Möglich ist der Einsatz von LLMs in der Kommunikation der Agenten. Dabei kann eine Variante mit natürlichsprachlicher LLM-to-LLM-Kommunikation als auch eine Variante mit strukturierter Kommunikation, die per LLM für menschliche Beobachter „übersetzt“, implementiert werden. Ein zweiter Schwerpunkt adressiert Reinforcement-Learning Agenten mit Fokus auf Optimierung von Fertigungsabläufen, prädiktive Ressourcenplanung und KI-gesteuerte AGV-Routingstrategien. Auch weitere Themen wie Agentenökonomie können vereinbart werden.

Bezug zum Thema Data Science

Das Projekt adressiert mehrere Kernbereiche des Forschungsmasters Data Science: Entwicklung autonomer Softwareagenten auf Basis von Künstlicher Intelligenz (Studienziel des Masters), Anwendung von LLMs für Reasoning und Entscheidungsfindung, Data Engineering beim Aufbau der Kommunikationsinfrastruktur und der Verknüpfung heterogener Datenquellen (MES, SPS, Sensorik), sowie die Anwendung verschiedener KI-Ansätze (z.B. LLMs oder Reinforcement Learning). Darüber hinaus berührt das Projekt Themen der verteilten Systeme, der Echtzeitkommunikation und der Mensch-Maschine-Interaktion.

Verfügbare Ressourcen

Sicherstellung der Verfügbarkeit von Daten, Rechenressourcen, Hardware, Anwendungsexperten

- IoT-Factory: 23-modulige Forschungs- und Produktionsanlage mit Fließband- und AGV-basiertem Transport, Montage-/Demontagerobotern, Kommissionierstation sowie assistierten Handarbeitsplätzen für manuelle und kombinierte Fertigung
- MES-Zugang: Vollständiger Zugang zum MES4-System von Festo inkl. Auftragsverwaltung, Workplan-Logik und Fehlerbehandlung
- Offene Datenbasis: Datenbanken, Maschinensensordaten und Teile des Softwarequellcodes stehen für Forschungszwecke offen zur Verfügung
- Rechenressourcen: Data-Analytics-Cluster des CfADS
- Betreuung: Laboringenieur M.Sc. Roman Sliwinski, Laborverantwortlicher Prof. Pascal Reusch sowie weitere Mitarbeitende des CfADS

Projektplan

Erstes Semester: Einarbeitung in Multi-Agenten-Systeme, LLM-basierte Agenten-Frameworks und die Architektur von OptiFlow. Analyse der bestehenden MQTT-Kommunikationsinfrastruktur und Konzeption einer agententauglichen Erweiterung. Erstellung eines ersten Proof-of-Concept mit zwei kommunizierenden Agenten an der realen Anlage. Erstellung eines Forschungsexposés ist Prüfungsleistung.

Zweites Semester: Implementierung des Multi-Agenten-Systems als Prototyp mit Anbindung an die realen Anlagenmodule über OptiFlow und MQTT. Je nach Schwerpunkt: Umsetzung und Vergleich verschiedener Kommunikationsvarianten (LLM-basiert vs. strukturiert), Entwicklung erster RL-Policies oder Entwurf ökonomischer Allokationsmechanismen. Erster Produktionslauf an der IoT-Factory. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, ist Prüfungsleistung.

Drittes Semester: Systematische Evaluation des agentischen Systems anhand verschiedener Szenarien (Einzelaufträge, parallele Aufträge, Ressourcenkonflikte, Störszenarien). Vergleich mit der zentralen MES-Steuerung anhand definierter KPIs wie Durchlaufzeit, Ressourcenauslastung und Entscheidungsqualität. Erstellung eines Papers mit ersten quantitativen Ergebnissen ist Prüfungsleistung.

Viertes Semester: Masterarbeit und Kolloquium

Eignungskriterien

Zwingend:

- Bereitschaft zur praktischen Arbeit an einer aktiven und komplexen Produktionsanlage
- Fortgeschrittene Programmierkenntnisse in Python
- Interesse an Künstlicher Intelligenz, insbesondere LLMs und Agentensystemen

Optional:

- Erfahrung mit Data-Science-Projekten, Digitalen Zwillingen, Intralogistik und Produktionssystemen
- Kenntnisse in Reinforcement Learning oder Natural Language Processing
- Erfahrung mit LLM-Frameworks (z.B. LangChain, CrewAI) oder RL-Bibliotheken (z.B. Stable Baselines3, RLlib)
- Grundkenntnisse in Softwarearchitektur (z.B. Django, modulare Python-Backends)

Erwerbbare Kompetenzen

- Praktische Erfahrung in der Entwicklung und Integration von Softwareagenten an realer Fertigungsinfrastruktur (SPS-Netzwerk, MES, MQTT, OPCUA, Fahrerlose Transportfahrzeuge)
- Fähigkeit zur vergleichenden Evaluation verschiedener KI-Kommunikationsansätze
- Umfassende Erfahrung im wissenschaftlichen Arbeiten, inklusive Publikation auf internationalen Konferenzen
- Mitarbeit an einem realen Open-Source-Softwareprojekt (OptiFlow) mit Django-Backend und modularer Python-Architektur
- Je nach Schwerpunkt: fundiertes Wissen in LLM-Engineering (Prompting, RAG, Agentendesign), Deep Reinforcement Learning (Policy-Optimierung, Reward-Design) oder Mechanismusdesign (Auktionen, Ressourcenallokation)
- Breites Skillset über Data Engineering, LLM-Engineering, IoT-Kommunikation und Produktionslogistik hinweg

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 12 Dialogorientierte Materialmodellierung und -simulation (Simulation Copilot)



Projektübersicht

Anzahl Studierende	1 bis 2
Art	Vorlaufforschung als Basis für einen Projektantrag
Projektverantwortung	Prof. Dr. Christian Schröder
Projektkontext	Projekt in der AG Computational Materials Science and Engineering im Kontext des Bielefelder Instituts für Angewandte Materialforschung (BifAM). Anstellung als HiWi an der HSBI ist möglich.

Abstrakt

Ziel ist ein dialogbasiertes Interface, das mit einem Large Language Model (LLM) Materialmodelle in natürlicher Sprache spezifiziert, fehlende Parameter gezielt nachfragt, Simulationen über Tool-Aufrufe startet und Ergebnisse nachvollziehbar zusammenfasst. Untersucht werden Prompting, RAG, Domänen-Ontologien und Evaluationsmetriken in Fallstudien. Ergebnis: Prototyp plus quantitative Evaluation.

Kurzbeschreibung

Ausgangssituation: Materialmodellierung und -simulation sind ein wesentlicher Teil moderner Produktentwicklung. Dabei verfügen Produktentwickler zwar über ein enormes Domänenwissen, Erfahrung im Umgang mit den üblicherweise sehr komplexen Softwaretools ist jedoch häufig nicht oder nur lückenhaft vorhanden. Jede Umgebung verlangt eine eigene Eingabesprache, Parameter, Konvergenzkriterien und Postprocessing. Viele Iterationen entstehen durch unvollständige Spezifikationen, falsche Einheiten, ungeeignete Randbedingungen oder unklare Zielgrößen.

Motivation: LLMs ermöglichen natürlichsprachliche Interaktion, können aber ohne domänenspezifische Absicherung zu Halluzinationen, falschen Einheiten oder nicht reproduzierbaren Workflows führen. Ein „Holodeck-ähnlicher“ Dialog (für alle, die Star Trek kennen 😊), der

systematisch nachfragt und Simulationstools korrekt bedient, könnte Zugänglichkeit, Geschwindigkeit und Reproduzierbarkeit deutlich verbessern.

Ziel: Forschung und Prototyping eines dialogorientierten „Simulation Copilot“, der (i) Aufgaben in natürliche Sprache annimmt, (ii) fehlende Informationen aktiv klärt, (iii) Simulationen über definierte Schnittstellen startet, (iv) Ergebnisse zusammenfasst/visualisiert und (v) Reproduzierbarkeit und Nachvollziehbarkeit sicherstellt.

Aufgabenstellung

Sie entwickeln und untersuchen einen Prototypen für ein Simulationstool (wird gemeinsam mit dem Betreuer festgelegt) mit folgenden Mindestanforderungen:

1. Domänenmodell & Dialogschema

- Definition eines strukturierten Schemas für: Materialparameter, Modelltyp, Geometrie, Randbedingungen, Numerik, Einheiten, Zielgrößen usw.
- Entwurf einer Nachfrage-Strategie (Clarification Policy): Wann muss das System nachfragen, wann darf es Standardwerte vorschlagen?

2. Tool-Integration

- Anbindung mindestens eines Simulators (z.B. LAMMPS für MD oder ein FEM-Workflow via COMSOL-Schnittstellen).
- Implementierung von „Tool Calling“: Generieren von Input-Dateien, Starten von Runs, Sammeln von Logs/Outputs.

3. Wissenszugriff & Absicherung

- Aufbau einer domänenspezifischen Wissensbasis (z.B. Handbuchauszüge, Materialdatenblätter, Best-Practice-Leitfäden) und RAG-Pipeline. RAG zur Quellenstützung (z.B. „Warum diese Cutoff-Wahl?“ „Welche Randbedingung passt?“).

4. Evaluation

- Erstellung eines Benchmark-Dialogs typischer Aufgaben (z.B. „Elastizitätsmodul schätzen → Simulation konfigurieren → Spannungs-Dehnungs-Kurve liefern“).
- „Assumption Ledger“: Liste aller Annahmen/Defaults, die im Dialog getroffen wurden.
- Quantitative Metriken: Erfolgsrate, Anzahl Dialogturns, Anteil notwendiger Korrekturen, Simulationsabbruchrate, Zeit bis Ergebnis, Reproduzierbarkeit (Run-Determinismus/Protokollqualität).

Bezug zum Thema Data Science

Teilbereiche:

- Natural Language Processing / Conversational AI
- Information Retrieval (RAG), Wissensrepräsentation
- MLOps/LLMOps und Experimentmanagement
- Datengetriebene Modellierung & Surrogatmodelle (optional, aber empfohlen)

- Evaluation/Statistik (quantitative Studien, Hypothesentests)

Methoden (Auswahl, projektabhängig):

- Prompt Engineering, Structured Output (JSON-Schemata), Guardrails
- Retrieval-Augmented Generation: Embeddings, Vektordatenbank, Chunking/Ranking
- Dialog-Policy-Design: regelbasiert + datengetrieben (z.B. Bandit/aktive Nachfrageheuristiken)
- Validierung: Constraints, Unit-Checks, Konsistenzregeln, „self-check“ mit Begründungspflicht
- Experimentdesign & Auswertung: A/B-Vergleiche (mit/ohne RAG, mit/ohne Nachfragen), Konfidenzintervalle, Effektstärken
- Optional: Bayesian Optimization/Active Learning zur automatisierten Parametersuche im Simulationsraum; Surrogate (Gaussian Processes/NN) zur Beschleunigung

Verfügbare Ressourcen

Vollausgestatteter Arbeitsplatz, Zugriff auf Rechenressourcen (HPC Cluster mit > 5000 CPU-Cores und leistungsfähigen GPUs), Simulationssoftware, Domänen- und Anwendungsexperten, ein cooles Team von Leuten 😊

Projektplan

1. Semester – Forschungsexposé

- Stand der Forschung: LLMs in Engineering/Simulation, Tool-Calling, RAG, Validierung
- Präzise Forschungsfragen und Hypothesen (z.B. „Nachfrage-Strategie senkt Abbruchrate um x%“)
- Methodik, Datengrundlagen, geplante Metriken, Risikoanalyse, Ethik/Compliance (Lizenzierung von Tools/Daten)
- Auswahl der Use Cases gemeinsam mit Betreuer, Definition der Metriken und Baselines

2. Semester – Survey/Overview Paper

- Systematische Literaturrecherche: LLMs in Engineering, Tool-Calling, Validierung, Reproducibility-by-design
- Taxonomie: Eingabespezifikation, Nachfrage-Strategien, Wissensquellen, Dialogsysteme für Engineering, Simulatoranbindung, Verifikation/Validierung, Reproduzierbarkeit
- Identifikation von Forschungslücken und Ableitung der eigenen Forschungsfragen und eines Referenz-Workflows für „Dialog → Setup → Run → Analyse“

3. Semester – Paper mit ersten quantitativen Ergebnissen

- Implementierter Prototyp inkl. Simulatoranbindung
- Benchmark-Auswertung und Variantenvergleich (Baseline vs. +RAG vs. +Constraints vs. +Policy)
- Vergleich mehrerer Varianten (z.B. Baseline Prompting vs. RAG vs. RAG+Constraints)
- Diskussion von Fehlerfällen, Limitationen, Generalisierbarkeit

4. Semester – Masterarbeit & Kolloquium

- Ausbau zur robusten Forschungsplattform (Logging, Reproduzierbarkeit)
- Erweiterte Evaluation (mehr Tasks, ggf. Nutzerstudie mit Material-/Simulationsanwendern)
- Abschluss: Masterarbeit (Methodik, Ergebnisse, Guidelines) + Kolloquium (Demo + wissenschaftliche Einordnung)

Eignungskriterien

Zwingend: Gute allgemeine Informatikkenntnisse, Programmierkenntnisse (Python usw.), Interesse an interdisziplinärem Arbeiten, Neugierde, Enthusiasmus

Optional: Erfahrung in LLMs und RAG

Erwerbbarer Kompetenzen

- **Wissenschaftliche Kompetenz:** Hypothesenbildung, experimentelles Design, quantitative Evaluation, sauberes Reporting
- **Data-Science-Kompetenz:** RAG-Pipelines, NLP/LLM-Workflows, Metriken, statistische Auswertung, Benchmarking
- **Software-Engineering:** Tool-Integration, Schnittstellen/Orchestrierung, Testbarkeit, Logging, Reproduzierbarkeit, Deployment-Grundlagen
- **Domänenkompetenz Materialsimulation:** Parameterisierung, Randbedingungen, Einheiten, Interpretation von Simulationsergebnissen
- **Qualität & Sicherheit:** Umgang mit Halluzinationen, Validierungsstrategien, Dokumentation von Annahmen/Unsicherheiten
- **Kommunikation:** Strukturierte Dialogdesigns, verständliche Ergebniszusammenfassungen, wissenschaftliches Schreiben

Forschungsprojekt für den Forschungsmaster Data Science

Projekttitle: 13 KI-Produktionssteuerung von Galvanikanlagen auf Basis eines Digitalen Zwillings



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externem Partner
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt in Kooperation mit der UBB GmbH. Anstellung als studentische Hilfskraft geplant.

Abstrakt

Die UBB GmbH betreibt zwei Galvanikanlagen zur Beschichtung mit Zink von zuvor gefertigten Stahlbauteilen. Ziel des vorliegenden Projekts ist die Entwicklung eines KI-basierten Agentensystems, das eine automatisierte Ressourcenplanung und -steuerung der Galvanik ermöglicht. Durch die Einführung dieses Systems sollen die Ressourcenauslastung verbessert, die Durchlaufzeiten verkürzt und die Termintreue erhöht werden. Im Rahmen des Projekts werden zunächst die Prozesse, Kundenanforderungen und Ressourcen in einem digitalen Zwilling abgebildet. Darauf aufbauend wird ein KI-Agentensystem entwickelt, das die Abläufe und den Ressourceneinsatz unter Berücksichtigung priorisierter Kundenforderungen und technischer Randbedingungen steuert.

Kurzbeschreibung

Die UBB GmbH mit Sitz in Burgbernheim ist eine hundertprozentige Tochtergesellschaft der Müpro International GmbH. Die Müpro Gruppe umfasst 14 internationale Tochtergesellschaften, beschäftigt etwa 650 Mitarbeitende und hat ihren Hauptsitz in Wiesbaden. Die UBB produziert mit seinen ca. 90 Mitarbeitern im 3-Schichtbetrieb in verschiedenen Fertigungsbereichen Befestigungssysteme als Produkte des Kernsortiments der Müpro Gruppe.

Im Fertigungsbereich Galvanik werden Metallteile mit einer definierten Zinkschicht versehen, um einen effektiven Korrosionsschutz für die spätere Anwendung der Fertigprodukte sicherzustellen. Die beiden Galvanikanlagen – eine Trommel- und eine Gestellanlage – sind integraler Bestandteil der Produktionskette und werden über das eingesetzte ERP-System (SAP) grob geplant und gesteuert. Die detaillierte Planung und konkrete Einsteuerung der Fertigungsaufträge in den Galvanikprozess erfolgen derzeit manuell durch Fachpersonal. Dabei werden neben einer gegebenenfalls angepassten Auftragspriorisierung auch galvanotechnische Parameter wie der Füllgrad der Trommeln und Gestelle sowie die Bauteilgeometrien berücksichtigt. Diese manuelle Vorgehensweise führt zu Unschärfen in der Planung und Terminabfolge, was negative Auswirkungen auf die Termintreue und Kapazitätsauslastung hat.

Aufgabenstellung

Die Aufgabenstellung besteht darin, ein innovatives KI-Agentensystem zu konzipieren und zu implementieren, das die bestehenden manuellen und transaktionsbasierten Planungs- und Steuerungsprozesse im Galvanikbereich der UBB GmbH automatisiert und optimiert. Im Mittelpunkt steht die vollständige Digitalisierung aller relevanten Abläufe, einschließlich der Erfassung und Modellierung von Prozessparametern, Ressourcen und Kundenanforderungen im digitalen Zwilling. Das zu entwickelnde System soll auf Grundlage von Echtzeitdaten sowie priorisierten Kundenforderungen eigenständig Entscheidungen zur Ressourcenplanung und Steuerung treffen und flexibel auf Veränderungen im betrieblichen Umfeld reagieren können. Ziel ist es, die Effizienz und Qualität der Galvanikprozesse signifikant zu steigern, die Transparenz im Warenfluss zu erhöhen und die operative Steuerung zu entlasten beziehungsweise zu übernehmen. Darüber hinaus soll das Agentensystem eine kontinuierliche Anpassung und Verbesserung der Abläufe ermöglichen, um den steigenden Anforderungen internationaler und volatiler Kundenprojekte gerecht zu werden.

Bezug zum Thema Data Science

Entwicklung eines KI-Agenten (Reinforcement Learning, Optimierung, Agentic AI) und das Erstellen Datenbasierter Digitaler Zwillinge.

Verfügbare Ressourcen

Zugang zu den Daten, IT-Systeme und Anlagen des Unternehmens wird sichergestellt. Ein fachlicher Ansprechpartner steht außerdem im Unternehmen bereit. Rechenressourcen werden durch die HSBI mit dem Your-AI Cluster bereitgestellt.

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die Aufgabenstellung im Unternehmen, Sichtung der Daten und Definition der Optimierungskriterien.

Zweites Semester: Literaturrecherche zur Verfahren des KI-basierten Produktionsteuerung. Implementierung des digitalen Zwillings. Erstellung eines Papers, das einen Überblick über das Forschungsgebiet gibt ist Prüfungsleistung.

Drittes Semester: Implementierung eines ersten Verfahren und Auswertung der Ergebnisse. Veröffentlichung eines Papers mit den ersten Ergebnissen ist Prüfungsleistung.

Viertes Semester: Masterarbeit und Kolloquium.

Eignungskriterien

Zwingend:

- Programmierkenntnisse (vzw. Python)

Optional:

- Erfahrung mit Maschinellem Lernen
- Erfahrung mit Optimierungsverfahren
- Erfahrung in Produktion und Logistik

Erwerbbarer Kompetenzen

- KI-Verfahren zur Produktionssteuerung
- Digitale Zwilling in der Produktion
- Publikation von wissenschaftlichen Veröffentlichungen auf internationalen Konferenzen
- Kompetenzen in der Arbeit im Team