Center for AI

Das gemeinsame Forschungszentrum für Künstliche Intelligenz von IBM und fortiss

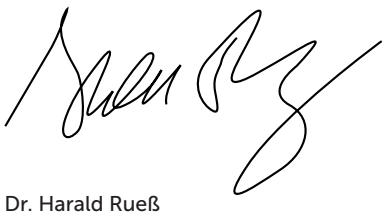
IBM — C4Ai — fortiss

Vorwort

Ökosysteme sind heute essenziell für Innovation und Wachstum. Deshalb sind wir sehr stolz, dass im April 2019 ein gemeinsames Forschungszentrum für Künstliche Intelligenz (KI) vom IBM Watson Center München und fortiss, dem Forschungsinstitut des Freistaats Bayern für softwareintensive Systeme, gegründet wurde. Die Partnerschaft wurde angelegt, um zuverlässige und sichere KI-Technologien für Wirtschaft und Gesellschaft zu entwickeln.

„Ein starkes Partner-Ökosystem aus Industrie und Forschung ist entscheidend, um das Potenzial von KI nachhaltig zu erschließen", erklärte Dr. Kareem Yusuf, IBM General Manager AI Applications & Blockchain, bereits 2019 bei der Gründung des gemeinsamen Forschungszentrums. Die Kooperation zwischen IBM und fortiss ermöglicht es, das Potenzial von KI weiter zu erforschen und zu realisieren. Gemeinsam wird geforscht und Wissen geteilt. IBM und fortiss entwickeln KI-Technologien für Wirtschaft und Gesellschaft, die in unsicheren, komplexen Umgebungen und Situationen zeitnahe und zuverlässige Entscheidungen treffen können.

Mit diesem Ziel vor Augen hat die Partnerschaft – die im IBM Watson Center München angesiedelt ist – bereits mehrere innovative Anwendungsfälle, Methoden und Proof of Concepts hervorgebracht, wie z.B. die Stresserkennung in Echtzeit zur Verbesserung der Sicherheit von Feuerwehrleuten, ein intelligentes System zur Verhinderung von Wassereinbrüchen auf der Basis von maschinellem Lernen und ontologischer Inferenz sowie Methoden für eine sichere und rechenschaftspflichtige KI zwischen Abteilungen im öffentlichen Sektor.



Dr. Harald Rueß

Andrea Martin

Center for AI

Unternehmen und Verwaltung in Bayern stehen derzeit vor der Herausforderung, Potenziale von KI-Technologien herauszuarbeiten, um neue Produkte zu entwickeln und neue Geschäftsfelder und Dienstleistungen zu erschließen. Dabei müssen die Forscher auch die Frage nach der Vertrauenswürdigkeit dieser Technologien beantworten.

Wirtschaft und Gesellschaft benötigen KI-Technologien, die in unsicheren, komplexen Umgebungen und Situationen zeitnah sichere Entscheidungen treffen. Die von Maschinen, Automaten oder Fahrzeugen getroffenen Entscheidungen sollten nicht nur nachvollziehbar und erklärbar, sondern auch robust gegen fehlerhafte Eingaben und gezielte Angriffe sein. Moderne KI-Systeme müssen zudem immer größere Datenmengen verarbeiten, aber ebenso auch aus kleinen Datenmengen nützliche Erkenntnisse gewinnen können – und das ohne unnötige Kompromisse bei Vertraulichkeit und Privatsphäre eingehen zu müssen.

Das gemeinsame Center for AI ist global vernetzt und arbeitet mit fortiss Wissenschaftlern aus München und Experten aus dem Netzwerk von IBM Forschungsinstituten zusammen, unter anderem aus Zürich (Schweiz), Dublin (Irland), Almaden und Yorktown (USA) sowie Hursley (Großbritannien). In der Einrichtung forschen und entwickeln insgesamt rund 50 Wissenschaftler neue KI-gestützte Softwarelösungen für unternehmens- und geschäftskritische Anwendungen für die Industrie als auch für den öffentlichen Sektor.

Gemeinsam identifizieren die Forscher von fortiss und IBM kontinuierlich wichtige neue Forschungsfragen, neue Geschäftsbedürfnisse oder -chancen, und entwickeln daraus sukzessive ein Portfolio lösungsorientierter KI-Projekte, um das Potenzial von KI nachhaltig zu erschließen. Ein gemeinsamer Lenkungsausschuss (Steering Committee) von IBM und fortiss begleitet die Aktivitäten und evaluiert regelmäßig den Fortschritt, unterstützt wichtige Herausforderungen zu identifizieren, und gibt entsprechende Handlungsempfehlungen.

> **„IBM und fortiss entwickeln KI-Technologien für Wirtschaft und Gesellschaft, die in unsicheren, komplexen Umgebungen und Situationen zeitnahe und zuverlässige Entscheidungen treffen können".**
>
> Dr. Harald Rueß

Steering Committee

fortiss

Dr. Harald Rueß
Wissenschaftlicher Geschäftsführer

Prof. Dr. Helmut Krcmer
Senior Research Fellow

Prof. Dr. Birte Glimm
Research Fellow

Prof. Dr. Ute Schmid
Research Fellow

IBM

Andrea Martin
Leiterin IBM Watson Center Munich

Alessandro Curioni
Vice President Europe &Africa, Director IBM Research

Kareem Yusuf, Ph.D
General Manager AI Applications & Blockchain

Dirk Wittkopp
Managing Director IBM Germany R&D

„Künstliche Intelligenz ist einer der wesentlichen Motoren der zukünftigen Wirtschaftsentwicklung. Wir sind stolz, dass fortiss und IBM diese zukunftsweisenden Technologien gemeinsam in Bayern entwickeln werden. Sie sind für die Wettbewerbsfähigkeit des Hochtechnologiestandorts Bayern von herausragender Bedeutung“

Hubert Aiwanger

ACRA4DT

Automatisierte Konfiguration von Robotern und Analytics in Industrie4.0 mit digitalen Zwillingen

Im Projekt Automated Configuration of Robots and Analytics in I4.0 with Digital Twins (ACRA4DT) wird durch das Hinzufügen von semantischem Wissen zur roboterbasierten Fertigung ein zusätzlicher Kontext für Anomalieerkennungsverfahren geschaffen, die auf maschinellem Lernen basieren. Dies ermöglicht die automatisierte Integration von Data Analytics für die Kleinserienfertigung.

Projektbeschreibung
Es sind viele manuelle Konfigurations- und Programmierschritte erforderlich, um hochgradig digitalisierte industrielle Fertigungsprozesse einzurichten. Dies beginnt mit der Programmierung der Steuerungslogik für einzelne Roboter und setzt sich fort bis zur Konfiguration von prozessüberwachenden Analytics-Verfahren.
Diese manuellen Schritte sind sehr zeitaufwändig und erfordern ein hohes Maß an Fachwissen. Folglich sind sie nur bei überwiegend statischen Produktionsplänen, bei denen die Betriebszeit die Programmierzeit bei weitem übersteigt, oder bei hochwertigen Gütern wirtschaftlich durchführbar. Auf der anderen Seite steigt die Nachfrage nach flexibleren Produktionslinien und der Einzellosfertigung von mittelwertigen Gütern ständig. Zwar versprechen Roboter und 3D-Drucker die erforderliche Flexibilität bei der Bearbeitung, doch in Wirklichkeit ist der Aufwand für die Anpassung von Programmen und Analytics-Verfahren noch zu hoch, um sie auf breiter Basis einzusetzen.

Forschungsbeitrag
Traditionelle Ansätze für Analytics-Verfahren in der Industrieautomation und Robotik verwenden Rohdaten von mehreren Sensoren und erfordern für jeden neuen Anwendungsbereich oder Anwendungsfall einen hohen manuellen Aufwand.
Bei einem wissensbasierten Ansatz zur Entwicklung von Produktionssystemen wird das Wissen über Produkte, Prozesse und Ressourcen formal in semantischen Repräsentationssprachen dargestellt. Auf der Grundlage eines solchen semantischen Digital Twin-Modells können Sensor-Rohdaten automatisch mit semantischen Beschreibungen sowie Kontextinformationen, z.B. der aktuell ausgeführten Montageaufgabe und den involvierten Bauteilen oder assoziierter Parameter, angereichert werden. fortiss untersucht in diesem Projekt, wie insbesondere für die Kleinserienfertigung Ansätze des maschinellen Lernens zur Anomalieerkennung durch die Integration dieser Art von Informationen automatisiert und optimiert werden können.

Alexander Perzylo
fortiss

Dr.-Ing. Bashar Altakrouri
IBM

Stresserfassungssystem für Feuerwehrleute

Stresslevel einer Einsatzkraft in Echtzeit messen und einschätzen

Das Projekt zielt darauf ab, den Stressstatus von Feuerwehrleuten rechtzeitig zu erkennen, um Unfälle und falsche Reaktionen im Notfalleinsatz zu vermeiden. Auf Basis von Human-centered Machine Learning entwickeln fortiss und IBM Ansätze für die Überwachung und das Management von physischem und psychischem Stress in sicherheitskritischen Anwendungen und Missionen.

Projektbeschreibung
Durch unsachgemäße Reaktionen in gefährlichen Einsätzen verletzt sich jedes Jahr eine große Anzahl von Feuerwehrleuten. Extreme Hitze, schlechte Sicht durch Rauchentwicklung, Zeitdruck, sind nur ein Teil der externen Faktoren, unter denen sie reaktionsschnell agieren müssen. Der Stress, der in einer solchen Situation entsteht, beeinflusst die körperliche und mentale Reaktionsfähigkeit und kann zu einer potenziell schweren Beeinträchtigung der kognitiven Fähigkeiten führen.

Die Betroffenen sind sich ihrer eingeschränkten Urteilsfähigkeit jedoch oft nicht bewusst. Daher ist es von entscheidender Bedeutung, Feuerwehrpersonal über die potenzielle Gefahr in Bezug auf ihren aktuellen körperlichen und emotionalen Zustand genauer zu informieren. fortiss arbeitet an einer Möglichkeit, das Stresslevel einer Einsatzkraft der Feuerwehr in Echtzeit zu messen und einzuschätzen, um sie auf dieser Basis direkt im Einsatz bei ihren Entscheidungen zu unterstützen

Forschungsbeitrag
In diesem Projekt untersuchen fortiss und IBM den mentalen Stress und entwickelt benutzerzentrierte Machine-Learning-Algorithmen zur Stressüberwachung. Sie dienen der verbesserten systeminternen Repräsentation des Menschen und basieren auf Data-Mining sowie der Erfassung von kognitiven Eigenschaften. Untersucht werden u.a. generelle Indikatoren für Stress, wie Herzfrequenz, Gehirnaktivität, Muskelspannung, Hautfeuchtigkeit oder Cortisonausschuss. Berücksichtigt werden weiterhin individuelle Reaktionen auf Stress, die nicht nur von der aktuellen mentalen Belastbarkeit des Einzelnen abhängt, sondern auch von der Situation in der er sich während der Messung befindet. Das datenbasierte System, das dabei entsteht, soll die Anforderungen und Bedürfnisse des Feuerwehrpersonals besser erfüllen.
fortiss und IBM entwickeln anhand unterschiedlicher Einsatzszenarien der Feuerwehr und den Erfahrungen aus solchen Einsätzen, neue personalisierte Stresserkennungsmodelle, um einen nachvollziehbaren Benutzer-Stresszustand zu liefern. Ziel des Projektes ist es, die Leistungsfähigkeit eines intelligenten lernenden Systems zu steigern.

Dr. Yuanting Liu
fortiss

Dr.-Ing. Bashar Altakrouri
IBM

DR&P

Nutzerorientierte Verwaltung durch proaktive und interaktionslose Leistungen

Die Zukunft der Verwaltung ist proaktiv: Verwaltungsleistungen werden ohne Antrag und Zutun der Nutzer*innen automatisch erbracht. Das Projekt Digital Readiness Assessment and Piloting for German Public Services (DR&P)" erforscht Konzepte und Methoden, die eine solche Zukunft möglich machen und probiert sie prototypisch aus.

Projektbeschreibung
Verwaltungsleistungen sind bürokratisch und langwierig. Oft müssen Bürger*innen und Unternehmen persönlich beim Amt erscheinen und Daten wie die Adresse mehrfach einreichen. Das antragslose Kindergeld in Österreich gilt als Vorbild auch für Verwaltungsleistungen in Deutschland. Aber wie können solche nutzerzentrierten Leistungen technisch und organisatorisch umgesetzt werden? fortiss und IBM denken nutzerzentrierte Verwaltungsleistungen über das Optimieren von Online-Anträgen hinaus und zeigen, wie diese komplett ohne Anträge auskommen können.

Eine proaktive Verwaltung wird von alleine aktiv und hilft ihren Nutzer*innen, damit sie möglichst wenig Aufwand haben. Das wird erreicht durch einheitliche Schnittstellen zwischen den beteiligten IT-Systemen, um automatisierte Datenabfragen und Anträge zu ermöglichen.

Im Rahmen des Projektes wird eine Readiness Assessment-Methode für Verwaltungsleistungen entwickelt und pilotiert. Im Fokus stehen dabei Effizienz von Verwaltungsprozessen, Nutzerzentrierung der Leistungen und die Attraktivität für Verwaltungsmitarbeitende. Dabei kommen Technologien wie KI und Digital Ledger Technologies (DLT) zum Einsatz.

Forschungsbeitrag
Die Digitalisierung ermöglicht die Transformation der öffentlichen Verwaltung von bürokratischen Genehmigungsstellen hin zum nutzerzentrierten Dienstleister. Die Nutzer*innen der Verwaltung werden aktiv unterstützt und mit individuellen Verwaltungsservices versorgt. Die Ergebnisse des Projekts helfen abstrakte Konzepte der Nutzerfreundlichkeit wie Once-Only und Proaktivität auf praktische Verwaltungsarbeit anzuwenden und zu implementieren.

In Zusammenarbeit entwickeln IBM und fortiss eine offene Schnittstellen-Spezifikation nach dem OpenAPI Standard, der die typischen Interaktionen zwischen Nutzer*innen und Verwaltung sowie innerhalb der Verwaltung abdeckt. Diese Spezifikation ist für alle Verwaltungsleistungen einheitlich und kann deshalb leicht skaliert werden.

Peter Kuhn
fortiss

Felizitas Müller
IBM

AFML

Accountable Federated Machine Learning für die deutsche öffentliche Verwaltung

Wissen teilen ohne Daten offenzulegen? Das Ziel des Projekts Accountable Federated Machine Learning (AFML) ist die Entwicklung eines Prototyps eines städteübergreifenden Ideenklassifikators im Kontext von Bürgerbeteiligung. Dies geschieht unter Einhaltung der gesetzlichen Vorschriften, besonders in Bezug auf Datenschutz, und basierend auf nachweisbarem, föderiertem maschinellem Lernen. Eine Föderierung ist wichtig, denn Städte verfügen häufig nicht über genügend Daten, um ein sinnvolles Modell zu trainieren. Dabei teilen die Städte nur das maschinell erlernte Wissen, aber keine Daten. Die Umsetzung dieser Anwendung soll das Potenzial von AFML demonstrieren und einen Rahmen für weitere Anwendungen entwickeln.

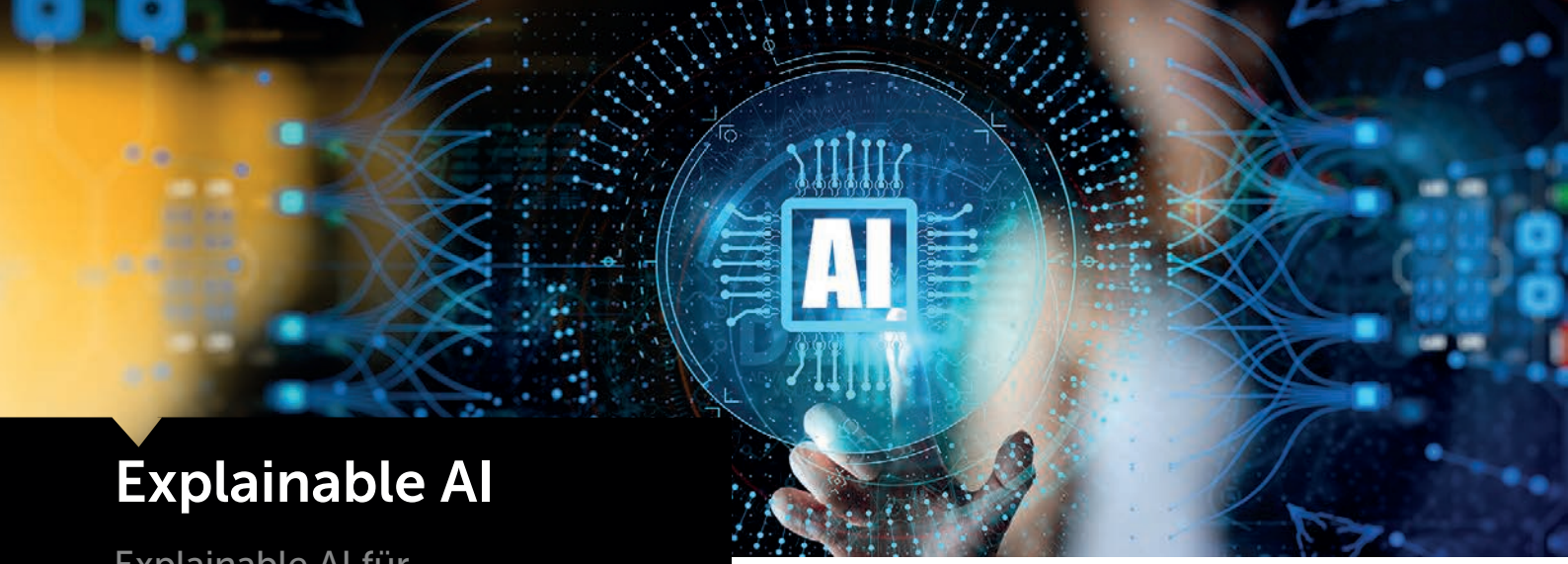
Projektbeschreibung
Im Projekt AFML wird ein Prototyp entwickelt, welcher das Trainieren von Modellen zur Klassifizierung von Bürgerideen anhand von föderiertem maschinellem Lernen ermöglicht. Dabei wird trotz der Dezentralität eine Nachweisbarkeit und Verifizierbarkeit des Prozesses und der Einhaltung von Kriterien (u.a. Datenschutz, Sicherheit, Datenverzerrung) der entstehenden Ergebnisse sicherstellt. Als Grundlage dienen Daten und Modelle im Kontext von Bürgerbeteiligung. Die entstandenen Modelle sollen über Stadtgrenzen hinweg genutzt und weiter trainiert werden. Dies soll jedoch ohne einen direkten Austausch von Daten zwischen den Beteiligten geschehen, indem das Modell jeweils lokal trainiert wird, und nur die entstehenden Änderungen in ein aggregiertes Modell überführt werden. Das Prinzip hierbei lautet: "Share Knowledge not Data". Mit AFML werden einzelne Schritte und lokale Trainingsiterationen nachweisbar protokolliert, sodass die Ergebnisse verifiziert und Manipulationen oder Fehler erkannt werden können.

Forschungsbeitrag
Federated Machine Learning (FML) ist ein vielversprechendes und spannendes Forschungsfeld im Bereich des maschinellen Lernens. Es ermöglicht Modelle dezentralisiert und lokal zu trainieren, und sie gleichzeitig übergreifend für die verschiedenen Beteiligten zu nutzen. Modelle auf der Basis großer Datenmengen zu trainieren soll auch in Bereichen möglich sein, in welchen kein Datenaustausch möglich ist.

Der Forschungsschwerpunkt liegt dabei derzeit auf dem Thema Secure & Privacy Preserving (Sicherheit & Wahrung der Privatsphäre). Im Projekt AFML soll dieses Konzept um die Komponente Nachweisbarkeit erweitert werden, sodass trotz des dezentralen Charakters von FML Verantwortlichkeiten definiert, Ergebnisse verifiziert und Manipulationen oder Fehler in einzelnen Prozessschritten erkannt werden können. Diese Erweiterung der aktuellen Forschung schafft Vertrauen in dezentral trainierte ML-Modelle.

Dian Balta
fortiss

Dr.-Ing. Bashar Altakrouri
IBM



Explainable AI

Explainable AI für Fahrassistenzsysteme

Im Projekt Explainable Artificial Intelligence (AI) für Fahrassistenzsysteme wird untersucht, wie das Fahrspurwechselverhalten von Fahrzeugen anhand des digitalen Zwillings von Providentia++ in Echtzeit vorhergesagt und erklärt werden kann.

Projektbeschreibung
Das Forschungsprojekt zielt darauf ab, das Potenzial und den Wert von Explainable AI im sicherheitskritischen Bereich der Vorhersage von Fahrspurwechseln zu demonstrieren. Dies eröffnet die Möglichkeit, den Stand der Forschung im Bereich des maschinellen Lernens zu erklären.
Das Ziel ist die Entwicklung eines maschinellen Lernmodells, das Spurwechsel vorhersagen und begründen kann, indem es die Daten des Providentia++ Digital Twin nutzt. Die Argumentation des maschinellen Lernmodells soll dargelegt werden, um Vertrauen aufzubauen und die Vorhersage des Modells in einem sicherheitskritischen Bereich zu rechtfertigen.
Schließlich werden die Vorhersagen und Erklärungen des maschinellen Lernmodells in einer Live-Webanwendung visualisiert, um das Potenzial von Explainable AI in sicherheitskritischen Bereichen für Kunden zu demonstrieren.

Forschungsbeitrag
Schichtnormalisierte Long Short-term Memory Modelle (LSTM) haben sich als robuste, dem neuesten Stand der Technik entsprechende maschinelle Lernmodelle für die Vorhersage von Fahrspurwechseln in Echtzeit erwiesen. Ihre innere Funktionsweise ist jedoch zu komplex und kompliziert, um von einem Beobachter verstanden zu werden. Daher sind schichtnormalisierte LSTMs sogenannte Blackboxes. Für sicherheitskritische Anwendungen wie die Vorhersage von Fahrspurwechseln muss ihre Funktionsweise explizit gemacht werden.
Um das Verhalten des Modells zu erklären, werden verschiedene Attributionsmethoden verglichen. Die schichtnormalisierte Relevanzausbreitung (Layer-wise Relevance Propagation, LRP) wird aufgrund ihrer Robustheit und ihres geringen Rechenaufwands als besonders geeignet für eine Erklärung in Echtzeit angesehen. LRP wurde jedoch nicht auf die schichtnormalisierte LSTM-Architektur ausgeweitet. Der bedeutende Forschungsbeitrag des Projektes ist, dass LRP auf normalisierte LSTMS erweitert wurde.

Dr. Yuanting Liu
fortiss

Dr.-Ing. Bashar Altakrouri
IBM

Interview

Das Center for AI – eine erfolgreiche Kooperation für eine verlässliche KI

Das Landesforschungsinstitut des Freistaats Bayern für softwareintensive Systeme fortiss und das Technologieunternehmen IBM haben vor genau zwei Jahren das Center for AI ins Leben gerufen, um sichere KI-Methoden für Wirtschaft und Gesellschaft zu entwickeln. Die Expert*innen von IBM und die Wissenschafter*innen von fortiss haben inzwischen mehrere KI-Forschungsprojekte erfolgreich umgesetzt. Im Fokus der Kooperation stand dabei immer die Frage: „Wann ist eine KI vertrauenswürdig und wie können Nutzer verstehen, wie oder warum eine KI Entscheidungen trifft?“ Dr. Holger Pfeifer, Kompetenzfeldleiter Software Dependability bei fortiss, koordiniert das Center for AI und ist Spezialist für Robuste KI. Im folgenden Interview berichtet er über die gewinnbringende Zusammenarbeit mit einem internationalen Technologieführer und über die durchgreifenden Fortschritte der vergangenen zwei Jahre.

Herr Pfeifer, das Center for AI wurde vor zwei Jahren ins Leben gerufen, um verlässliche und sichere KI-Technologien für Wirtschaft und Gesellschaft zu entwickeln. Welche Ergebnisse sind bisher zu verzeichnen?

Wir haben eine Reihe von Forschungslinien aufgesetzt und darin mehrere Projekte gestartet, z. B. in den Bereichen Mensch-zentriertes Maschinelles Lernen, Robotik, Behördendienstleistungen und verteiltes Lernen sowie Anomaliedetektion im Gebäudemanagement.

Im Themenbereich Mensch-zentriertes Maschinelles Lernen (human-centered machine learning, HCML) gehen wir der Frage nach, wie Anwendungen, die Maschinelles Lernen einsetzen, besser den Anforderungen eines menschlichen Nutzers gerecht werden können. Das heißt auf der einen Seite wird es für den Menschen einfacher, Entscheidungen der KI zu verstehen und nachzuvollziehen, und auf der anderen Seite, wie die individuellen Unterschiede zwischen den Nutzern schon direkt in die Entwicklung der Lernalgorithmen einfließen können. Diese Fragestellungen werden am Anwendungsfall des Stress-Managements für Feuerwehrleute im Einsatz untersucht. Da Menschen unterschiedlich auf Stresssituationen reagieren, muss ein Lernalgorithmus personalisiert, also in der Lage sein, das Ausmaß des Stresses für Feuerwehrleute individuell zu erkennen. Umgekehrt müssen die Signale der KI z. B. dem/der Zugführer*in so präsentiert werden, dass diese/r leicht erkennen kann, welche seiner/ihrer Einsatzkräfte sich in einer besonderen Stresssituation befindet. Hier entwickeln wir neue Modelle für die Stresserkennung auf Basis von verschiedenen Biosignalen und sogenanntem selbstüberwachtem Lernen (self-supervised learning).

Für den Einsatz von Robotern in der Fertigung entwickelten wir handhabbare und wirtschaftlich durchführbare Konfigurationen und Analyselösungen, die auch für kleine und mittlere Unternehmen nutzbar sind. Dazu wurden semantische Modelle erarbeitet, die beschreiben, welche Arbeitsvorgänge der Roboter durchführen soll. Diese Informationen werden mit Daten aus den Roboteraktionen verknüpft. Mittels Maschinellen Lernens ist das System dann in der Lage, ständig zu lernen, Abweichungen zu erkennen und bei Bedarf eine Warnung zu generieren, idealerweise schon bevor ein Fehler entsteht.

In Bezug auf Behördendienstleistungen haben wir untersucht, wie diese so gestaltet werden können, dass der Bürger sie online in Anspruch nehmen kann, ohne mühsam Formulare ausfüllen zu müssen, und wie sie dem

Bürger proaktiv zur Verfügung gestellt werden können, z. B. bei bestimmten Ereignissen, ohne dass der Dienst dabei speziell beantragt werden muss. Wir haben dies an zwei beispielhaften Anwendungen demonstriert: der Beantragung von Kindergeld sowie der Erlaubnis zur Eröffnung einer Gaststätte.

Und schließlich arbeiten wir seit Kurzem an einer Anwendung zur Erkennung von Wasserschäden in Gebäuden. Hier nutzen wir Sensoren, die in dem Gebäude des Münchner IBM Watson Centers verbaut sind, um dynamisch Feuchtigkeitsdaten zu sammeln, und verknüpfen diese Ergebnisse mit statischen Gebäudeinformationen und Umweltdaten. Werden nun durch spezielle KI-Algorithmen Ereignisse in den dynamischen Sensordaten erkannt, die auf das Eindringen von Wasser hindeuten, so können diese mithilfe der ontologischen Daten nicht nur genauer lokalisiert werden, sondern es kann auch deren Ursache besser qualifiziert werden. So kann z. B. erkannt werden, ob das Ereignis durch ein offen stehendes Fenster, durch das Regen eindringt, oder etwa durch einen Rohrbruch verursacht wurde.

Mit IBM verbindet Sie bereits eine mehrjährige Zusammenarbeit. Warum ist gerade IBM ein guter Partner für diese Themen und was ist Gegenstand dieser gemeinsamen Unternehmung?

Ich denke, die zentralen Punkte sind, dass IBM weltweit führend auf dem Gebiet der KI ist und über ein großes, weltweites Netzwerk an renommierten Forschungszentren und Businesspartnern verfügt, die wissen, was in der Praxis benötigt wird. Neben der führenden Industrieexpertise von IBM ist auch der ethische Grundsatz sehr wichtig. IBM überlässt Partnern und Kunden die Entscheidung, welche Daten die KI nutzen soll (Transparenz) und wie sie mit ihren Daten eigene kognitive Lösungen bauen können.

Welche Herausforderungen sehen Sie beim jetzigen Stand der KI-Technologie?

KI-Systeme sind lernbasiert, d. h., ihre Funktion bzw. ihr Verhalten ergibt sich aus Daten, anhand derer sie trainiert werden. Solche Verfahren führen jedoch auch zu unvorhersehbarem Verhalten im Betrieb, was eine große Herausforderung für die Frage nach der Beherrschbarkeit solcher Systeme darstellt. Ferner sind KI-Lösungen auch immer nur so gut wie die Daten, mit denen sie trainiert worden sind. Dies stellt zunächst einmal die Frage nach der Verfügbarkeit geeigneter Daten und im zweiten Schritt nach deren Qualität – Stichwort repräsentative Daten und Einseitigkeit. KI-Systeme können sich durch fortwährendes Lernen auch weiterentwickeln und sich erfahrungsbasiert an neue Gegebenheiten anpassen oder optimieren. Dies macht die Entwicklung und vor allem das Testen und Absichern sehr schwierig. Etablierte Verfahren aus dem Software- und Systems-Engineering setzen voraus, dass genaue Spezifikationen des Systemverhaltens und der Einsatzumgebungen vorliegen und die Systeme vor der Inbetriebnahme umfassend verifiziert werden, was für KI-Systeme nicht ohne Weiteres machbar ist. Daher sind solche Verfahren auch nicht direkt übertragbar. Ferner operieren KI-Systeme oftmals in teilweise unbekannten oder unsicheren Umgebungen und müssen in robuster Weise mit ungenauen oder unsicheren Eingabedaten oder Situationen umgehen können, die vorab nicht oder nur unvollständig modelliert worden sind.

> **Ein wichtiger Aspekt der Vertrauenswürdigkeit ist die Frage nach der Nachvollziehbarkeit der KI-Entscheidungen: Kann ich als Nutzer verstehen, wie oder warum die KI eine bestimmte Entscheidung trifft?**
>
> Dr. Holger Pfeifer

Momentan ist KI für viele Menschen noch eine Blackbox mit vielen Unbekannten. Wie wollen Sie an diesem Punkt das Vertrauen der Menschen gewinnen?

Ein wichtiger Aspekt der Vertrauenswürdigkeit ist die Frage nach der Nachvollziehbarkeit der KI-Entscheidungen: Kann ich als Nutzer verstehen, wie oder warum die KI eine bestimmte Entscheidung trifft? Hier geht es darum, KI-Methoden so zu entwickeln, dass sie dem Nutzer auch Erklärungen für die Ergebnisse liefern kann. Am IBM fortiss Center for AI haben wir dazu erst kürzlich ein Projekt gestartet, bei dem ein intelligenter Fahrassistent entwickelt werden soll, der bei einer Autobahnfahrt Empfehlungen gibt, welche Fahrspur benutzt werden bzw. wann die Spur gewechselt werden soll. Zugleich soll das System in der Lage sein zu erklären, warum eine bestimmte Empfehlung ausgesprochen wird. Da die Daten, auf deren Grundlage eine Empfehlung zum Spurwechsel getroffen wird, möglicherweise aus entfernteren Straßenabschnitten stammen, die der Fahrer noch gar nicht einsehen kann, wie zum Beispiel ein Unfall oder ein Stauende hinter einer Kurve, ist es für das Vertrauen des Fahrers in die Entscheidungen wichtig, diese nachvollziehen zu können.


10.200.2.68
10.200.2.69

Ab wann wird eine KI als robust und vertrauenswürdig definiert?

Als robust sehen wir eine KI-Lösung an, wenn sie auch in unbekannten Umgebungen oder unvorhergesehenen Situationen noch sicher funktioniert und gute Ergebnisse liefert. Hier können wir es mit einer Reihe von Unsicherheiten oder besser: Ungewissheiten zu tun haben. Das kann zum Beispiel daran liegen, dass bei der Entwicklung bzw. dem Training der KI gar nicht alle möglichen Situationen oder Umgebungseinflüsse betrachtet worden oder überhaupt bekannt sind. Außerdem kann es vorkommen, dass eine KI-Komponente schon eine falsche Entscheidung trifft, wenn die Eingabe nur leicht von den bekannten Situationen abweicht. Man kennt solche Probleme zum Beispiel aus der Bilderkennung, wo die Änderung nur weniger Bildpunkte etwa dazu führen kann, dass eine rote Verkehrsampel als „grün" wahrgenommen oder ein Fußgänger gar nicht erkannt wird. Dies bietet natürlich auch die Gefahr von böswilligen Manipulationen, bei denen ein Angreifer versucht, die KI durch gezielte Eingaben zu stören. Insoweit ist das also auch eine Frage der Sicherheit der KI-Anwendung und damit ihrer Vertrauenswürdigkeit. Um KI-basierte Anwendungen in sicherheitskritischen Bereichen wie zum Beispiel dem autonomen Fahren oder der Medizinrobotik einsetzen zu können, müssen wir darauf vertrauen können, dass solches Fehlverhalten nahezu ausgeschlossen ist.

Welche Lösungen entwickelt das Center for AI um gezielt Behördendienste serviceorientierter und direkter zu gestalten?

Die Zukunft der öffentlichen Verwaltung ist proaktiv und interaktionslos. Behördliche Dienste sollen automatisch bereitgestellt werden, ohne dass Anwendungen benötigt werden und ohne dass der Bürger mit einer Anwendung interagieren muss. Mit unserem Projekt DR&P (Digital Readiness Assessment and Piloting for German Public Services) am Center for AI wollen wir das Konzept der proaktiven und interaktionslosen Behördendienste in die Praxis umsetzen. Daher haben wir eine Analysemethode für die Einsatzbereitschaft bestimmter Dienste entwickelt und angewendet, bestehende Software-Frameworks erweitert und zwei Demonstratoren (für das Beantragen von Kindergeld sowie für die Anmeldung einer Gaststätte) entwickelt. Durch unsere Forschung bieten wir Behördenpraktikern einen strukturierten Engineering-Ansatz zur Verknüpfung von visionärem Service-Design mit fortschrittlichen Technologien, um eine höhere Servicequalität für Bürger und Unternehmen zu erreichen. Eine Behörde, die proaktiv Dienstleistungen erbringt, gilt als benutzerfreundlich und verbessert die Servicequalität, da sie dem Benutzer eine Dienstleistung liefert (benutzerzentriert), anstatt sie nur zu genehmigen (regierungszentriert). Für die Bereitstellung solcher proaktiver und interaktionsloser Dienste werden intelligente Datenverarbeitung mittels Maschinellen Lernens und rechenschaftspflichtiger Datenaustausch mittels Distributed Ledger Technologie (DLT) die technologische Basis bilden.

Wie können Behörden in einer föderierten Umgebung unter Einhaltung des Datenschutzes Wissen teilen, ohne Daten freizugeben?

Die Stichworte sind hier „Föderiertes Maschinelles Lernen (federated machine learning – FML)" und „Rechenschaftspflicht (accountability)". FML ist ein Ansatz, der es mehreren Parteien ermöglicht, kooperativ ein gemeinsames maschinelles Lernmodell aus ihren Daten zu erstellen, ohne diese Daten teilen zu müssen. Die Idee ist, dass alle Parteien maschinelle Lernaufgaben auf ihren privaten Datensätzen ausführen und die resultierenden Modellaktualisierungen austauschen, um ein kombiniertes Modell der gesamten Daten zu erstellen. Auf diese Weise bleiben die Daten privat und die Parteien tauschen nur Modell-Updates und Testdaten zur Beurteilung der Qualität der gelernten Modelle aus. IBM Research hat ein Framework für föderiertes Lernen entwickelt, welches wir in unserem gemeinsamen Projekt „AFML – accountable federated machine learning" nun mit dem von fortiss entwickelten System Evidentia verbinden. In Evidentia können verteilte Arbeitsabläufe wie das föderierte Lernen spezifiziert und die Durchführung von Arbeitsschritten fälschungssicher dokumentiert werden. So ist jederzeit klar nachvollziehbar und belegbar, welcher Akteur was genau zu welcher Zeit gemacht hat, welche Entscheidungen getroffen wurden und warum und ob die tatsächlichen Handlungen der Teilnehmer dem vereinbarten Lernprozess entsprechen. Dies erlaubt es einem Konsortium von Akteuren, Ansprüche auf überprüfbare Weise zu erfassen, auch wenn kein gegenseitiges Vertrauen besteht.

Welche Branchen profitieren besonders davon?

Am Center for AI forschen wir zunächst einmal branchenunabhängig bzw. -übergreifend. Für KI gibt es unzählige Anwendungsfelder. Überall dort, wo Daten anfallen, kann KI und vor allem Maschinelles Lernen potenziell gewinnbringend eingesetzt werden. Gerade auch im Bereich der Fertigungsindustrie mit ihrem hohen Anteil an wiederkehrenden und vorhersehbaren Tätigkeiten besteht besonderes Potenzial. Natürlich arbeiten wir auch an konkreten Anwendungsfällen für bestimmte Felder, wie zum Beispiel dem Einsatz von Techniken des Maschinellen Lernens zur Erkennung von Anomalien der roboterbasierten Fertigung.

Wie können Unternehmen Zugang zu diesen Technologien bekommen?

Im Center for AI als gemeinsamer Forschungsunternehmung zwischen IBM und fortiss erforschen und entwickeln wir KI-Lösungen für gezielte Herausforderungen. Wenn Unternehmen mit uns kooperieren wollen, ist unser Bereich fortiss Mittelstand die zentrale Anlaufstelle. Hier bündeln wir unserer Services für Unternehmen und informieren sie gerne zu den Kooperationsformen.

Impressum

Herausgeber
fortiss GmbH
Guerickestraße 25
80805 München
E-Mail: info@fortiss.org
www.fortiss.org

Autoren
Andrea Martin
Dr. Harald Rueß
Dr. Holger Pfeifer
Kathrin Kahle
Silvia Hervé

Gestaltung
Victoria Plewniok
Kathrin Kahle

Druck
Viaprinto

September 2021

Bildnachweise
Seite 1: Adobe Stock, ©Lukassek
Seite 3: ©Astrid Eckert
Seite 4: ©fortiss, ©IBM
Seite 5: ©Astrid Eckert
Seite 6: shutterstock, ©metamorworks
Seite 7: shutterstock, ©Rawpixel.com
Seite 8: shutterstock, ©Alexander Supertramp
Seite 9: shutterstock, ©Gorodenkoff
Seite 10: Adobe Stock, ©greenbutterfly
Seite 11: shutterstock, ©VAKS-Stock Agency
Seite 12: ©fortiss, Kathrin
Seite 13: ©Astrid Eckert
Seite 15: shutterstock, ©Photon photo

IBM Watson Center Munich
Highlight Towers
Mies-van-der-Rohe-Str. 8
80807 München
Deutschland

https://www.ibm.com/business-operations/resources/munich-center

fortiss GmbH
Guerickestraße 25
80805 München
Deutschland

www.fortiss.org

IBM C4Ai fortiss