Themenkartenstapel:

Künstliche Intelligenz in der Arbeitswelt – Die technische Seite von hKI

Themenkartenstapel Überblick

Themenkarten

Die technische Seite von hKI: Was machen KI-Expert:innen mit hKI?

Diese Themenkarte behandelt die Grundlagen der technischen Umsetzung von humanzentrierter KI. Die Arbeit eines KI-Experten in humanzentrierten KI-Projekten (hKI) ist facettenreich und hochgradig anspruchsvoll. Sie erfordert technisches Know-how, interdisziplinäre Zusammenarbeit und ein tiefes Verständnis für menschliche Bedürfnisse und gesellschaftliche Verantwortung. Dieser Abschnitt beleuchtet die vielschichtigen Aufgaben, die KI-Experten im Rahmen der Entwicklung, Implementierung und Wartung von hKI-Systemen übernehmen.

Die folgende Abboldung symbolisiert diese komplexe Aufgabe: Der Kontext einer KI-Anwendung ist nur dem Menschen bekannt, welcher dieses Wissen in Form von Anweisungen an die KI transferiert. Die KI wiederum gibt Erklärungen über seine Verarbeitung der Daten an den Menschen zurück. Dieser Transfer ist das Hauptaugenmerk der folgenden Themenkarten.

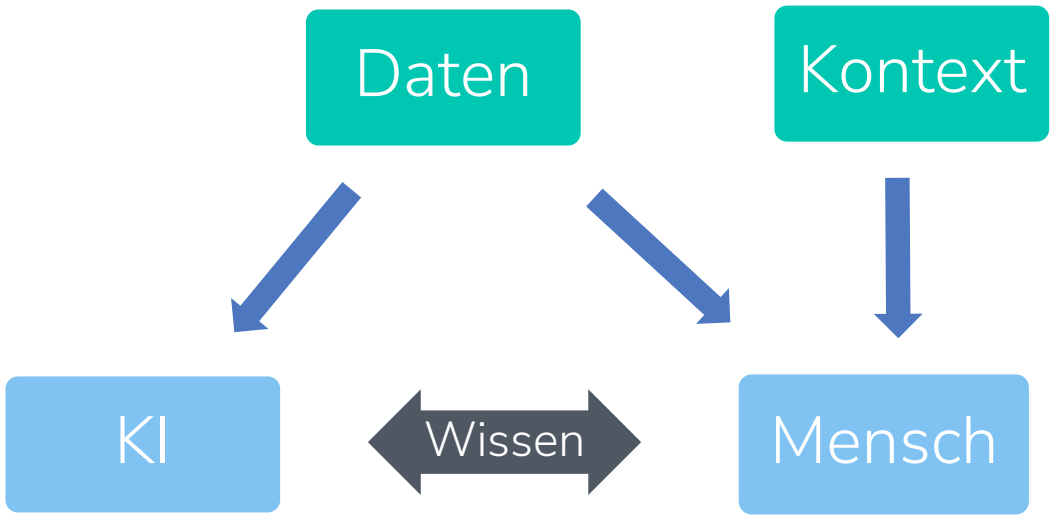

Abbildung 1: eigene Darstellung, adaptiert von Lamarr Institute (2024)

Technische Erfordernisse bei der Entwicklung von hKI-Systemen

Zu Beginn der Arbeit von KI-Experten steht die Vorbereitung eines KI-Systems. Dies betrifft die Datengrundlage sowie die Entwicklung und Optimierung von Algorithmen, die die Grundlage aller KI-Systeme bilden. Während klassische Algorithmen häufig rein leistungsorientiert sind, müssen hKI-Algorithmen zusätzlichen Anforderungen gerecht werden. Sie sollen nicht nur effizient, sondern auch gerecht und transparent sein. Diese Ziele sind herausfordernd, da sie teils gegenläufig sind.
Um Fairness zu gewährleisten, analysieren Experten zunächst die Datengrundlage. Verzerrte Daten können dazu führen, dass bestimmte Gruppen benachteiligt werden. Dieses Problem ist besonders in sensiblen Anwendungsbereichen wie der Strafjustiz, der Personalrekrutierung oder der medizinischen Diagnostik relevant. Mit Methoden wie „Fairness-Constraints“ wird sichergestellt, dass Algorithmen keine diskriminierenden Muster verstärken. Eine weitere Technik, die hier eingesetzt wird, sind adversarielle Netzwerke. Diese trainieren Modelle so, dass sie selbstständig problematische Muster erkennen und ausgleichen können.
Ein weiteres Schlüsselmerkmal von hKI ist die Erklärbarkeit der Modelle, die unter dem Begriff Explainable AI (XAI) bekannt ist. Modelle wie neuronale Netzwerke, insbesondere Deep-Learning-Systeme, gelten häufig als "Black Boxes", deren innere Funktionsweise nur schwer nachzuvollziehen ist. Für hKI sind solche "Black Boxes" nicht akzeptabel. Durch XAI-Methoden wie SHAP (Shapley Additive Explanations) oder LIME (Local Interpretable Model-agnostic Explanations) können Experten aufzeigen, welche Faktoren zu einer Entscheidung beigetragen haben. Diese Erklärungen stärken das Vertrauen in die Technologie und ermöglichen es Anwendern, fundierte Entscheidungen zu treffen.

Technische Erfordernisse im laufenden hKI-System

Schnittstellen: Interaktion zwischen Mensch und Maschine optimieren

Die Schnittstelle zwischen Mensch und KI ist ein zentraler Bestandteil humanzentrierter KI. Während klassische KI-Systeme oft auf statischen, technischen Interfaces beruhen, müssen hKI-Schnittstellen nutzerzentriert und dynamisch gestaltet sein. KI-Experten arbeiten eng mit UX-Designern, Psychologen und anderen Fachleuten zusammen, um Interfaces zu entwickeln, die intuitiv, flexibel und inklusiv sind. Es stellt sich die Frage: Wie kann ein Mensch bestmöglich befähigt werden, die Konzepte und innere Logik einer KI zu verstehen? Dazu existieren eine Vielzahl von Methoden und Werkzeugen.

Eine besondere Rolle spielt hierbei die Anpassungsfähigkeit der Schnittstellen. Adaptive Systeme erkennen, welche Informationen der Nutzer benötigt, und passen die Darstellung entsprechend an. Beispielsweise kann ein medizinisches Diagnose-System Ärzten detaillierte Daten zu Krankheitsbildern präsentieren, während es Patienten eine vereinfachte Erklärung bietet. Solche Flexibilität ist entscheidend, um die breite Akzeptanz von hKI-Systemen zu fördern.

Ein weiteres Schlüsselelement sind Feedbackmechanismen. Experten implementieren Rückmeldeschleifen, die es Nutzern ermöglichen, direkt Einfluss auf die Funktionsweise der KI zu nehmen. Solche Systeme lernen kontinuierlich aus den Interaktionen und verbessern sich entsprechend. Darüber hinaus berücksichtigen KI-Experten emotionale Aspekte, etwa durch die Integration von Technologien zur Stimmungs- und Gefühlsanalyse. In Bereichen wie der Pflege oder im Bildungswesen kann eine solche emotionale Sensibilität die Qualität der Interaktion erheblich steigern.

Integration in bestehende Infrastrukturen: Technische Herausforderungen

Die Integration von hKI-Systemen in bestehende technische Infrastrukturen ist eine der anspruchsvollsten Aufgaben für KI-Experten. Viele Organisationen verfügen über komplexe IT-Ökosysteme, die durch die Einführung von KI ergänzt werden sollen, ohne dass es zu Ausfällen oder Inkompatibilitäten kommt. Hierbei ist die Interoperabilität, also das problemfreie Zusammenarbeiten verschiedener technischer Systeme, ein zentrales Kriterium.

Experten entwickeln hKI-Systeme, die modular aufgebaut sind, sodass sie schrittweise in bestehende Prozesse integriert werden können. Dies ist besonders in kritischen Bereichen wie der industriellen Fertigung oder dem Gesundheitswesen wichtig, wo Systeme in Echtzeit zuverlässig funktionieren müssen. Eine weitere Herausforderung besteht in der Skalierbarkeit der Systeme. Mit zunehmender Nutzerzahl und wachsender Datenmenge müssen hKI-Systeme ihre Leistung aufrechterhalten können.

Ein praktisches Beispiel ist die Einführung von KI-Systemen in Produktionslinien, die eine Vielzahl von Sensoren und Maschinen verknüpfen. Hierbei müssen hKI-Systeme nicht nur Daten in Echtzeit verarbeiten, sondern auch Fehlermuster frühzeitig erkennen und Handlungsempfehlungen geben. Experten arbeiten daran, diese Systeme so zu gestalten, dass sie sowohl robust als auch flexibel sind, um zukünftigen Anforderungen gerecht zu werden.

Interdisziplinäre Zusammenarbeit: Synergien schaffen

Die Entwicklung von hKI ist ein interdisziplinäres Unterfangen, das die Zusammenarbeit von Experten aus verschiedenen Bereichen erfordert. Ethiker, Soziologen, Designer und Domänenexperten bringen ihre spezifischen Perspektiven und Fachkenntnisse ein. KI-Experten agieren dabei oft als Vermittler, die technische Möglichkeiten mit den Anforderungen anderer Fachbereiche in Einklang bringen.

Ethiker tragen dazu bei, dass hKI-Systeme den Prinzipien der Gerechtigkeit und Verantwortung entsprechen. Soziologen analysieren die gesellschaftlichen Auswirkungen der Technologie und identifizieren potenzielle Akzeptanzprobleme. Designer sorgen für intuitive und ansprechende Interfaces, während Domänenexperten sicherstellen, dass die Systeme auf die Anforderungen ihrer Branche zugeschnitten sind. Diese Zusammenarbeit ist nicht nur notwendig, um hKI-Systeme erfolgreich zu entwickeln, sondern auch, um deren Akzeptanz und langfristige Nutzung zu sichern.
Überwachung und kontinuierliche Verbesserung
Nach der Einführung eines hKI-Systems endet die Arbeit der KI-Experten nicht. Vielmehr beginnt eine Phase der kontinuierlichen Überwachung und Weiterentwicklung. Systeme müssen regelmäßig überprüft werden, um sicherzustellen, dass sie zuverlässig und sicher bleiben. KI-Experten nutzen dafür fortschrittliche Monitoring-Tools, die Echtzeitdaten analysieren und potenzielle Probleme frühzeitig erkennen.
Zusätzlich fließen Rückmeldungen der Nutzer direkt in die Optimierung der Systeme ein. Dies ist besonders in dynamischen Anwendungsbereichen wie der autonomen Mobilität oder der medizinischen Diagnostik von Bedeutung, wo Anforderungen und Bedingungen sich stetig ändern. Ein gutes Beispiel ist die Aktualisierung von Algorithmen zur Bildanalyse in der Radiologie, die mit immer neuen Datensätzen trainiert werden, um ihre Genauigkeit zu verbessern.

Wissensvermittlung und Nutzerakzeptanz

Ein oft unterschätzter Aspekt der Arbeit eines KI-Experten ist die Vermittlung von Wissen an Nutzer und die Förderung der Akzeptanz von hKI-Systemen. Viele Menschen stehen KI skeptisch gegenüber, insbesondere wenn sie deren Funktionsweise nicht verstehen. KI-Experten sind daher auch als Vermittler gefragt. Sie organisieren Schulungen, erstellen technische Dokumentationen und fördern ein allgemeines Verständnis für die Technologie.
Die Wissensvermittlung geht dabei über rein technische Aspekte hinaus. Ziel ist es, Nutzer zu befähigen, die Möglichkeiten und Grenzen der Systeme realistisch einzuschätzen. Diese Maßnahmen tragen nicht nur zur besseren Nutzung der Systeme bei, sondern bauen auch Ängste und Vorurteile ab. So wird hKI als Unterstützung wahrgenommen, die den Menschen stärkt, anstatt ihn zu ersetzen.
Weitere Informationen zu Gestaltungsansätzen bei der KI-Einführung in Unternehmen finden sich in diesem Change-Management-Whitepaper der Plattform Lernende Systeme.

Wie setzt man die Kriterien von hKI um?

Im vierten Themenkartenstapel „Kriterien einer auf den Mensch ausgerichteten Künstlichen Intelligenz“ sind die Kriterien erläutert, die für ein erfolgreiches hKI-System erreicht werden sollten. Diese müssen jeweils technisch umgesetzt werden, dies wird im Folgenden skizziert. Die technische Umsetzung stellt dabei nur eine Möglichkeit dar, dieses Kriterium umzusetzen.

1. Erklärbarkeit und Transparenz von KI
- Definition: Die Prozesse und Entscheidungen der KI sollten für Anwender:innen nachvollziehbar und verständlich sein. Erklärbarkeit bedeutet, dass das System seine Entscheidungen oder Vorhersagen plausibel begründen und darstellen kann.
- Technische Umsetzung: LIME (Lokale Interpretierbare Modell-Agnostische Erklärungen) ermöglicht es, Vorhersagen eines interpretierbaren Modells zu erklären, indem ein simpleres Modell gebaut wird, das sich in dem speziellen Fall einer Vorhersage genau so verhält wie das eigentliche, komplexere Modell.

2. Entscheidung durch den Menschen
- Definition: In sensiblen Entscheidungen bleibt der Mensch in der Kontrollposition und kann die KI-Entscheidungen überstimmen oder anpassen. Dies wird als „Human-in-the-Loop“ bezeichnet.
- Technische Umsetzung: Visuelle Schnittstellen im Programm, die grafisch ansprechend und übersichtlich entwickelt sind und die natürlichen Stärken der menschlichen Wahrnehmung unterstützen. Logische Zusammenhänge visualisieren zu können, vereinfacht dem „human-in-the-loop“ das Überwachen der KI.

3. Mitbestimmung und Akzeptanz
- Definition: Beschäftigte und Stakeholder sollten die Möglichkeit haben, an der Entwicklung und Einführung von KI-Systemen teilzuhaben und ihre Bedenken oder Anforderungen einzubringen, um die Akzeptanz und Nutzen der Technologie zu erhöhen.
- Technische Umsetzung: Technisch einfach verständliche Wege, Prototypen von KI zu bauen sowie interaktive Dashboards erlauben es Anwender:innen, schnell ein Grundverständnis für die Funktionsweisen zu entwickeln und früh an der iterativen Weiterentwicklung der KI mitzuwirken. Mit der Wahrnehmung der Beschäftigten beschäftigt sich diese Fallstudie des Forschungsprojektes ai:conomics.

4. Technische Robustheit
- Definition: KI-Systeme müssen zuverlässig und fehlerresistent sein, um ihre Funktion unter verschiedenen Bedingungen zu erfüllen. Technische Robustheit reduziert das Risiko, dass Systeme durch unvorhergesehene Eingaben oder Fehler falsch reagieren.
- Technische Umsetzung: Ein robustes „Safety-Management“ benötigt eine erweiterte Software-Architektur, die Sicherheitsprüfungen implementiert. Wenn ein System zu „unsicher“ wird, also sich auf statistisch dünnem Eis bewegt, indem eine vorher definierte Sicherheitsgrenze überschritten wird, können automatische Entscheidungen des Systems blockiert werden und eine Bewertung durch menschliche Anwender:innen erfordert werden.

5. Datensicherheit und Datenschutz

- Definition: Die Verarbeitung personenbezogener Daten durch KI-Systeme muss den geltenden Datenschutzbestimmungen entsprechen und Datensicherheit gewährleisten. Nutzer:innen sollen darauf vertrauen können, dass ihre Daten sicher und verantwortungsvoll behandelt werden.
- Technische Umsetzung: Sicherheitsmechanismen wie Verschlüsselungsprotokolle, Anwendung im kontrollierten Raum (z.B. auf lokalen Servern oder einer privaten Cloud) für die Datenübertragung begleiten die technische Seite des Datenmanagements eines KI-Systems. Weiter können Daten pseudonymisiert oder anonymisiert werden.

6. Verantwortung und Haftung

- Definition: Es muss klar geregelt sein, wer für die Entscheidungen und Fehler der KI-Systeme verantwortlich ist. Dies betrifft sowohl ethische als auch rechtliche Aspekte.
- Technische Umsetzung: Mit juristischer Unterstützung, z.B. durch externer Gutachter:innen, wird frühzeitig im Dialog aller Stakeholder festgestellt, welche Szenarien rechtlich wie zu bewerten sind. Noch bevor das KI-System den Weg in die Nutzung im Unternehmen findet, besteht bei allen Beteiligten ein Konsens, wer bei Fehlern und technischen Ausfällen haftet.

7. Diskriminierungsfreiheit

- Definition: KI-Systeme sollten so gestaltet sein, dass sie keine Vorurteile oder Benachteiligungen gegenüber bestimmten sozialen Gruppen aufweisen.
- Technischen Umsetzung (Beispiel): Eine KI, die Stellenbewerbungen vorsortiert, wird so trainiert, dass sie keine Voreingenommenheiten bzgl. Geschlechter-, Alters- oder Herkunftsmerkmalen aufweist. Regelmäßige Überprüfungen und Anpassungen helfen sicherzustellen, dass das System neutral bleibt.Viele Ansätze für KI-Beschränkungen zugunsten der Fairness werden entwickelt, beispielsweise im Max-Planck-Institut für Software-Systeme.

8. Ökologische Verantwortung

- Definition: KI-Entwicklung und -anwendung sollten umweltbewusst gestaltet sein und möglichst wenig Ressourcen verbrauchen. Energieeffiziente Modelle und optimierte Datenverarbeitungsprozesse tragen zur ökologischen Verantwortung bei.
- Technischen Umsetzung (Beispiel): Ein Unternehmen verwendet energieeffiziente KI-Modelle und cloudbasierte Server, die mit erneuerbarer Energie betrieben werden, um den ökologischen Fußabdruck zu reduzieren. Auch bei der Datenverarbeitung und dem KI-Training wird darauf geachtet, den Ressourcenverbrauch zu minimieren.

Best-Practice-Beispiele

Best-Practice-Beispiele geben einen praxisnahen Einblick, wie humanzentrierte KI in verschiedenen Branchen sinnvoll eingesetzt wird.

Medizinische Diagnostik:
In der medizinischen Diagnostik beispielsweise unterstützen KI-Systeme Ärztinnen bei der Auswertung von Bilddaten wie Röntgenaufnahmen oder MRT-Scans. Diese Systeme können Anomalien in den Bildern erkennen und so Ärztinnen Hinweise auf mögliche gesundheitliche Probleme geben, was die Diagnose beschleunigt und präzisiert. Dabei bleibt die Entscheidung über die Diagnose und Behandlung jedoch stets in der Hand der Fachleute. Diese Art von KI-Einsatz zeigt, wie die Technologie als Unterstützung und nicht als Ersatz dient und das Wohl der Patienten in den Vordergrund stellt.

Bildung:
Ein weiteres Beispiel ist die Anwendung von KI in der Bildung. Adaptive Lernsysteme analysieren den Fortschritt von Schüler*innen und passen die Lerninhalte individuell an deren Bedürfnisse an. Ein solches System kann erkennen, welche Themen die Lernenden bereits verstanden haben und bei welchen sie noch Unterstützung benötigen. Hierdurch wird der Lernprozess effizienter und zielgerichteter, was besonders in großen Schulklassen mit heterogenen Leistungsniveaus von Vorteil ist. Diese Systeme tragen dazu bei, die Chancengleichheit zu fördern, da sie eine gezielte Förderung ermöglichen.

Industrie 4.0 und Automatisierung:
In der Industrie und Fertigung wird KI genutzt, um Prozesse effizienter und sicherer zu gestalten. Ein Beispiel ist der Einsatz von KI zur Überwachung und Steuerung von Produktionsprozessen. Diese Systeme können Abweichungen erkennen, bevor sie zu größeren Problemen führen, und so die Produktion optimieren. Durch den Einsatz von KI können Beschäftigte von Routineaufgaben entlastet werden und sich auf kreative und strategische Aufgaben konzentrieren. Auch hier zeigt sich der Nutzen humanzentrierter KI, indem sie den Arbeitsalltag der Menschen erleichtert und zur Wertschöpfung beiträgt.

Literatur

Danave, Sameer (2024): Explainable AI: Seven Tools and Techniques for Model Interpretability. https://dzone.com/articles/explainable-ai-7-tools-and-techniques-for-model (Abruf am 28.11.2024).

Fleck, Lara/Graus, Evie/Klinger, Maximilian (2022): Verändert Künstliche Intelligenz die Zukunft unserer Arbeit? Wahrnehmungen von betroffenen Arbeitnehmer:innen. https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/Veraendert_Kuenstliche_Intelligenz_die_Zukunft_unserer_Arbeit.pdf (Abruf am 12.11.2024).

Fraunhofer IKS: SAFE AI: Absicherung von Künstlicher Intelligenz (KI). https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz/absicherung-ki.html (Abruf am 26.11.2024).

Kutzias, Damian (2021): Triple KI: Die drei Säulen erfolgreicher KI-Implementierung im Unternehmen. https://blog.iao.fraunhofer.de/triple-ki-die-drei-saeulen-erfolgreicher-ki-implementierung-im-unternehmen/ (Abruf am 29.11.2024).

Lamarr Institute for Machine Learning and Artificial Intelligence (2024): Research Area Human-centered AI Systems. https://lamarr-institute.org/research/human-centered-ai-systems/ (Abruf am 29.11.2024).

Pinkwart, Niels/ Beudt, Susan (2020): Künstliche Intelligenz als unterstützende Lerntechnologie. https://publica-rest.fraunhofer.de/server/api/core/bitstreams/2d4f9603-b95c-4548-912a-d4a3b596ebf1/content (Abruf am 29.11.2024).

Stowasser, Sascha/Suchy, Oliver/Huchler, Norbert/Müller, Nadine/Peissner, Matthias/Stich, Andrea/Vögel, Hans-Jörg/Werne, Jochen (2020): Einführung von KI-Systemen in Unternehmen. Gestaltungsansätze für das Change-Management, Whitepaper aus der Plattform Lernende Systeme. https://www.acatech.de/publikation/einfuehrung-von-ki-systemen-in-unternehmen-gestaltungsansaetze-fuer-das-change-management/ (Abruf am 25.09.2024).

Visani, Giorgio (2020); LIME: explain Machine learning predictions. Intuition and Geometrical Interpretation of LIME. https://towardsdatascience.com/lime-explain-machine-learning-predictions-af8f18189bfe. (Abruf am 29.11.2024):

Zafar, Muhammed Bilal/ Valera, Isabel/ Roduriguez, Manuel Gomez et. al. (2017): Fairness Constraints: Mechanisms for Fair Classification. https://doi.org/10.48550/arXiv.1507.05259 (Abruf am 29.11.2024).