



KI- Kontextanalysen

als notwendiges
Entscheidungswerkzeug für den
verantwortungsvollen Einsatz von
Machine Learning Verfahren

Warum ist das nötig?

Die Situation:

„KI-Einsatz in Europa nimmt zu

Eine neue Studie von Juniper Networks kommt zu dem Ergebnis, dass der Einsatz von KI in den vergangenen zwölf Monaten in Europa stark zugenommen hat.

63 Prozent der befragten Führungskräfte gaben an, dass KI-Einführungsziele bereits „größtenteils“ erreicht wurden. Vor einem Jahr lag der Anteil noch bei 6 Prozent.

Governance-Richtlinien sind für die meisten Befragten hingegen eine große Schwachstelle. **Nur 9 Prozent der befragten IT-Führungskräfte halten KI-Richtlinien in ihrem Unternehmen für >>vollständig ausgereift<<.**“

Quelle: bigdata-insider.de und „Startup-Insider“ vom 19.7.2022

sagte Kolbe. "Diese Technologie wird vieles in Wirtschaft und Gesellschaft verändern. Wir halten diese Veränderungen aber für gestaltbar."

Was ist eine menschenzentrierte KI?

Mehr als Usability und User Experience?

Das Problem:

Beispiel: „Gemeinsam wird es KI“

Civic Innovation Plattform,
Workshop auf der re:publica 22

Berlin Juni 22



gefördert vom Bundesministerium für Arbeit und Soziales

[Start - Civic Innovation Platform \(civic-innovation.de\)](https://civic-innovation.de), Abruf am 19.7.2022



Challenge Auswahl

Wähle eine Challenge aus, mit der du dich kreativ beschäftigen möchtest.

Wie kann KI genutzt werden, um die Arbeit zu verbessern?

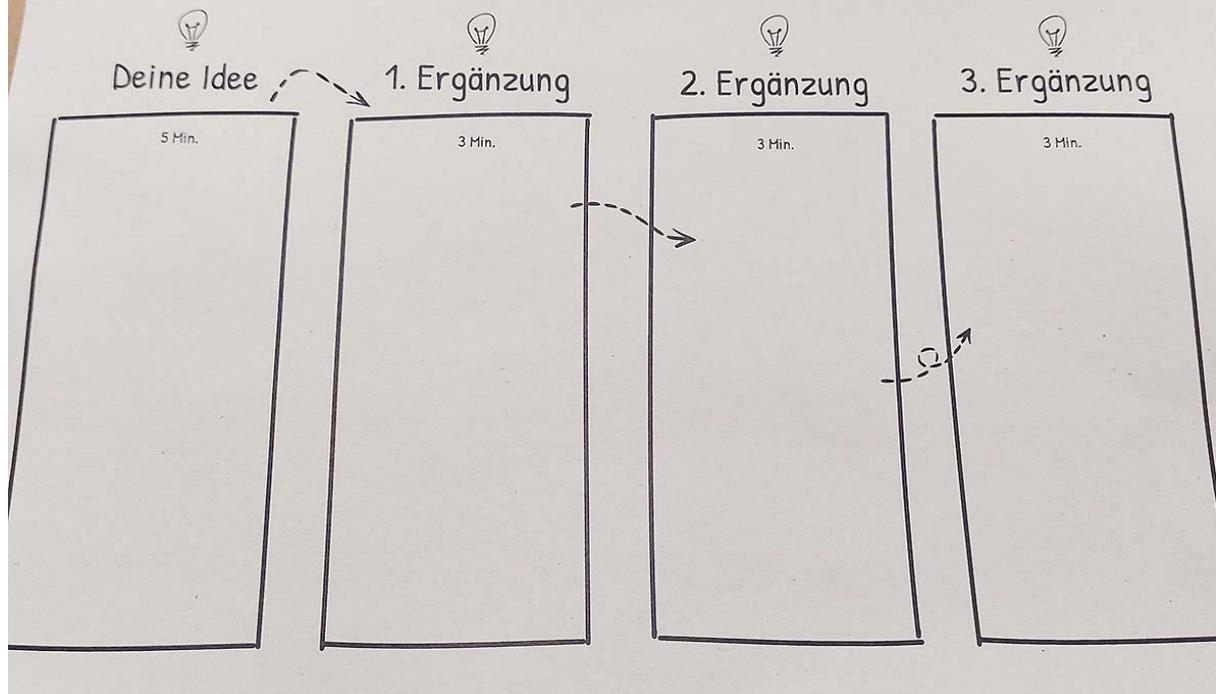
Wie kann KI genutzt werden, um das Leben zu verbessern?

Wie kann KI genutzt werden, um unsere Gesellschaft zu verbessern?

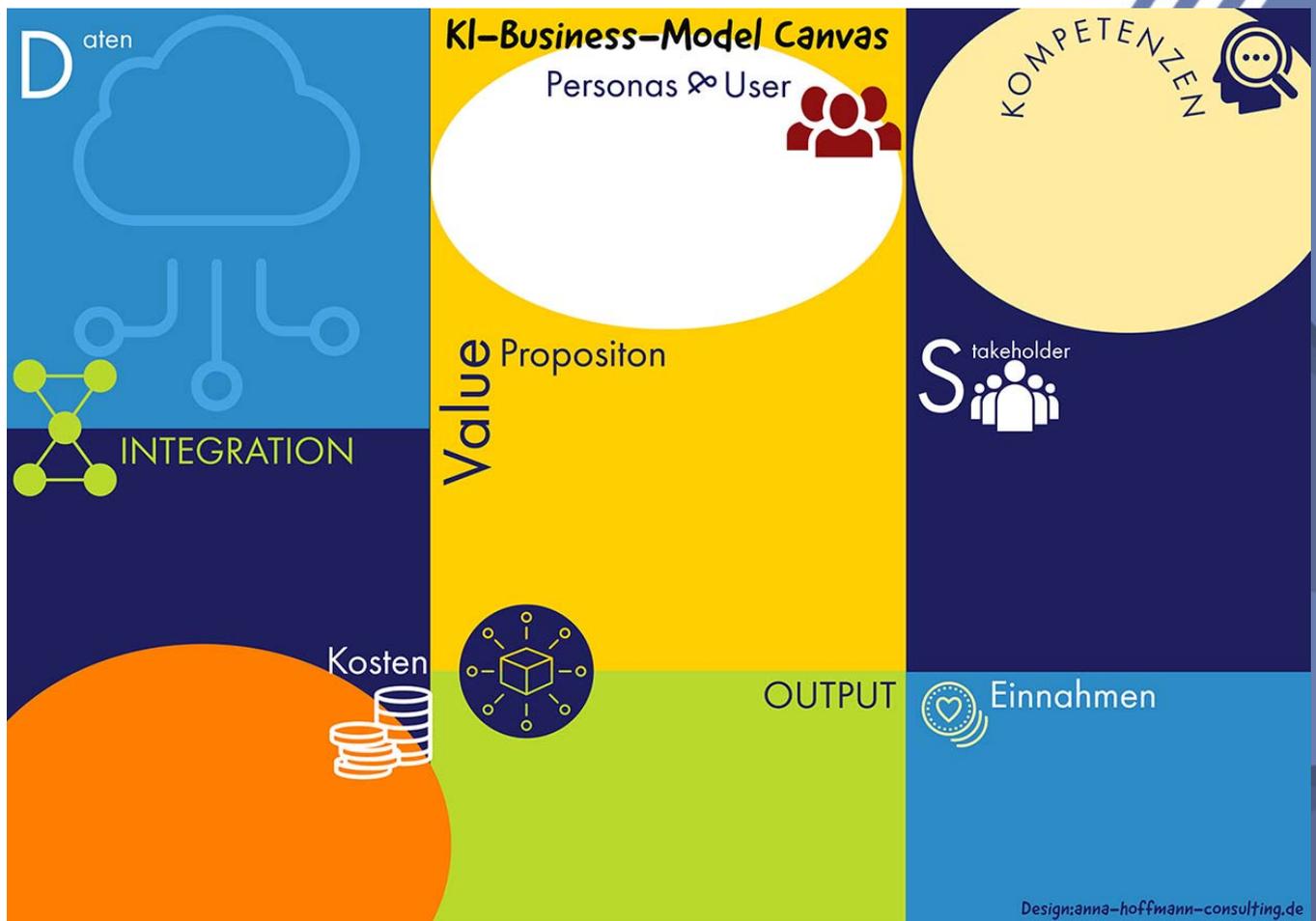
Wie kann KI genutzt werden, um die Umwelt zu verbessern?

Speed date your idea

Nehme deine Idee jetzt in Angriff und erhalte durch die Kreativmethode 4-1-3 weitere Ergänzungen von deinem Team.



Fotos: Anna Hoffmann, re:publica 22



Eine mögliche Lösung:

KI-Kontextanalyse

(Version Anna Hoffmann 2022)

In Punkt 4.4 „KI-spezifisches Risikomanagement“

des Berichts der **Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“ des Deutschen Bundestages** vom 28.10.2020 heißt es:

„Erst die Betrachtung des individuellen Anwendungskontextes und der individuellen Einsatzumgebung erlauben eine umfassende Bewertung der mit dem Gebrauch von Algorithmen und KI-Systemen einhergehenden Kritikalität.“¹

¹ [Drucksache 19/23700 \(bundestag.de\)](https://www.bundestag.de/Drucksache_19_23700), S. 66; Abruf 8.7.2022

Genau auf diesem Gesichtspunkt baut die KI-Kontextanalyse auf, die den spezifischen Anwendungskontext sowie die spätere Nutzerumgebung vertieft in den Blick nimmt.

Es handelt sich dabei um eine Weiterentwicklung der „Gestaffelten Kontextsitzungen“ nach Thomas Bevan („Theory of Machines“) ergänzt um Fragestellungen aus der „Human Centered AI“ (Mittelstand 4.0-Kompetenzzentrum Usability, Workshops und Inputs von Manuel Kulzer)².

Der Fragenkatalog der KI-Kontextanalyse bietet eine Lösungsidee, wie ML-basierte KI-Systeme schon ab dem ersten Entwurf des Use-Cases darauf durchleuchtet werden können, ob es sich um

- ein menschenzentriertes und verantwortungsvolles Business-Design handelt,
- welche Risiken sich zeigen und
- wo nachgebessert werden kann oder auch muss.

Die vorgestellte Form der KI-Kontextanalyse wird zurzeit am Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik (IPK) in Berlin für einen weitreichenden KI-Use Case erprobt.

1. Grundlegende Fragen zum Design der KI:

- Welchen Use-Case fördert die KI?
In welchem Sinnzusammenhang soll die KI einen Mehrwert generieren?
- Wie relevant ist der Einsatz der KI für den Erfolg des Prozesses/ Use Cases?
- Welcher konkrete Outcome/Nutzen soll durch die KI erzeugt werden?
- Welchen (Daten-)Output soll/muss die KI haben?

² [Mittelstand 4.0-Kompetenzzentrum Usability Kompetenzzentrum Usability \(kompetenzzentrum-usability.digital\)](https://www.kompetenzzentrum-usability.digital)

- Welcher Grad an Genauigkeit wird dafür benötigt?
- Wer soll später die KI anwenden/damit umgehen können?
Wer sind die Anwender/User der KI?
- Welche Stakeholder sollen durch die KI profitieren?

2. Grundlegende Fragen zum Aufbau einer nachhaltigen Lernumgebung für die KI:

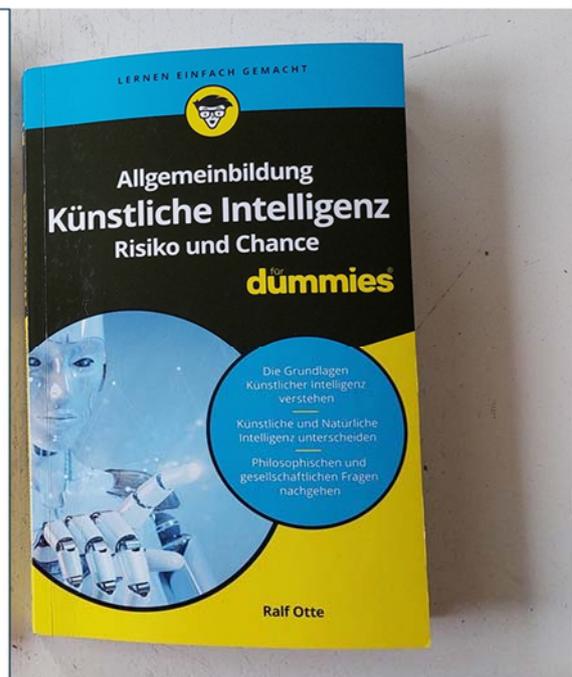
- Worüber soll/wird das KI-System konkret entscheiden?
- Welches Wissen wird für diese Entscheidung benötigt?
- Wie kann dieses Wissen zustande kommen:
 - Welche Daten werden dafür benötigt?
 - Welches menschliche Fachwissen wird für eine sichere Entscheidung benötigt?
- Welche Folgen könnte eine Fehlentscheidung der KI haben?

S.133:

„Eine Korrelation beschreibt eine datenbasierte Beziehung zwischen Merkmalen.

Eine Korrelation ist ein statistischer Zusammenhang zwischen bestimmten Erscheinungen beziehungsweise Merkmalen.“

Eine gefundene Beziehung zwischen den Merkmalen A und B muss keine kausale Beziehung sein. **Es muss kein Ursache-Wirkung Zusammenhang zwischen A und B bestehen** (z.B. wenn A, dann ergibt sich B).

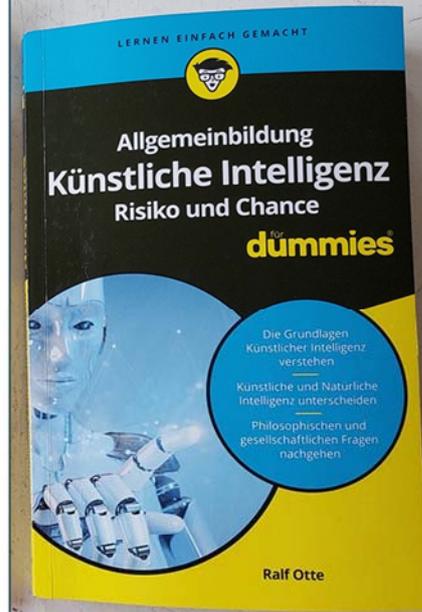


S.134:

**„Korrelation und Kausalität
(Ursächlichkeit) sind nicht
dasselbe.**

**Es handelt sich um vollkommen
unterschiedliche Konzepte zur
Beschreibung der Welt!“**

Beide Konzepte haben ihren Platz
in dem wissenschaftlichen
Fachgebiet der Künstlichen
Intelligenz, dürfen aber nicht
verwechselt werden.



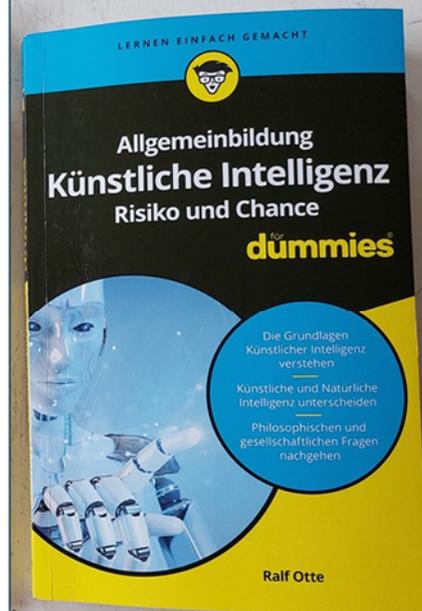
S.134: **BEISPIEL**

Korrelationsaussage:

Schuhgröße und Einkommen
korrelieren in den Daten positiv:

>Je größer die Schuhgröße, desto
höher das Einkommen.<

**Es gibt aber keinen kausalen
Zusammenhang zwischen beiden
Daten. Die Schuhgröße sagt nichts
über das Einkommen aus oder
umgekehrt.**

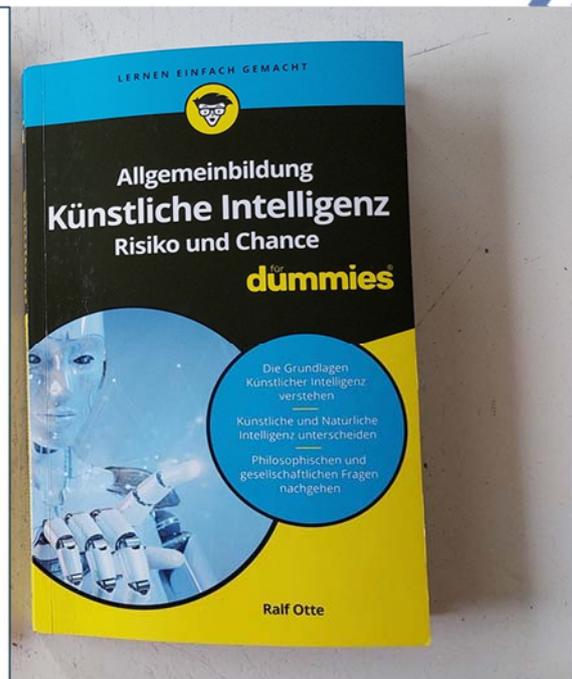


S.135:

„Wenn Sie die Korrelation von einer Aussage nicht bewerten können, so ist es meist vernünftiger, anzunehmen, dass hinter der Korrelation kein kausaler Zusammenhang steckt.

Lieber vermuten Sie keinen Zusammenhang als einen falschen.“

Die Kosten für einen falschen Zusammenhang/Irrtum können hoch sein.



3. Grundlegende Fragen zur späteren Einbindung der KI in wirtschaftliche Strukturen:

- An welcher Stelle, in welchem Prozess soll die KI eingebunden werden?
Wie sieht die Integrationsschnittstelle genau aus?
- Welche Produkte und Systeme werden für den Prozess genutzt, in die die KI eingebunden werden soll?
- Welche Aufgaben soll die KI in dem Gesamtprozess exakt übernehmen?
- Welches technische/digitale Vorwissen ist bei den Anwendern der KI anzutreffen/vorauszusetzen?

4. Grundlegende Fragen zur langfristigen Qualitätssicherung der KI:

- Wie viele Ausgangsdaten werden für einen möglichst umfassenden Lern- und Evaluierungsprozess benötigt?
- Wo können diese Daten vorliegen oder gewonnen werden?
- Welche Fachexpertise wird zum Labeln der Daten benötigt?
- Wie kann eine „Human-in-the-loop“-Schnittstelle eingebunden werden?
- Wie können langfristige Fehler/Ungenauigkeiten der KI sicher erkannt werden?
- Sind die Testdaten für den Lernprozess aus einem anderen Datenpool als die späteren Evaluierungsdaten?
[Testdaten für die die Messung des Lernerfolg vs. Real-Life-Daten für eine echte Praxiserprobung]
- Wie kann sichergestellt werden, dass relevante/fehlende Daten regelmäßig ergänzt werden?
- Wie kann ein nachhaltiges Design der KI aussehen, durch das notwendige Aktualisierungen leicht implementiert werden können?
- Wie häufig soll die KI in der Praxis zum Einsatz kommen?

5. Grundlegende Fragen zum Risikomanagement:

- In welchen Anwendungsfällen könnte die KI durch Übertragung in einen anderen Kontext ungeplant Schaden verursachen?
- Wo und wie könnte die KI missbraucht werden?

Anwendungsbeispiel KI-Kontextanalyse

Use-Case: Eifel barrierefrei erleben

[Quelle: [Eifel barrierefrei erleben - Civic Innovation Platform \(civic-innovation.de\)](https://civic-innovation.de), Zugriff 27.6.2022]

„Das Ziel dieser Idee ist die Entwicklung einer digitalen Anwendung, die passgenaue Empfehlungen für barrierefreie Reiseerlebnisse in Nordrhein-Westfalen anbietet. Dabei sollen die Informationen nicht nur auffindbar sein, sondern an die Betroffenen „herangetragen“ werden, indem die Daten auf die jeweiligen Nutzer*innen zugeschnitten sind.

Unter Beachtung der Bedürfnisse und Anforderungen der Nutzer*innen sollen verlässliche und konkrete Informationen zu Angeboten für Menschen mit Beeinträchtigungen zur Verfügung gestellt werden. Der verfolgte Ansatz zielt aber auch auf andere Gruppen ab, wie etwa Familien mit Kinderwagen oder Personen mit schwerem Gepäck. An der Idee beteiligt sind der *Tourismus NRW e.V.*, das *Fraunhofer-Institut für Software- und Systemtechnik ISST* sowie die *Eifel Tourismus (ET) GmbH*.“

„Eure Idee in drei Sätzen erklärt.

Ziel unserer Idee ist die Entwicklung einer digitalen Anwendung in Form einer Progressive Web App (PWA), welche, mit Hilfe eines Recommender Systems, Empfehlungen für barrierefreie Reiseerlebnisse anbietet. Hierbei sollen die Bedürfnisse und Anforderungen der Nutzer*innen beachtet werden. Der Fokus liegt dabei auf beeinträchtigten Menschen, die Umsetzung kommt aber allen Nutzer*innen gleichermaßen zu Gute.

Was macht eure Idee besonders?

Durch die Anwendung entsteht ein „supervised learning Ansatz“ für barrierefreie touristische Angebote. Die Informationen sollen nicht nur auffindbar sein, sondern an die Nutzer*innen „herangetragen“ werden, indem die von der PWA ausgespielten Daten passgenau zu den Nutzer*innen und deren Bedürfnissen ausgespielt werden und ihnen somit die Reiseplanung erleichtert wird. Beeinträchtigte Menschen benötigen verlässliche und konkrete Informationen zu den Angeboten, damit sie einschätzen können, ob diese für sie nutzbar sind.“

Beantworten Sie die grundlegenden Fragen zum Design einer KI (Machine Learning) für den Use-Case: „Eifel barrierefrei erleben“:

- In welchem Kontext/Zusammenhang soll die KI einen Mehrwert generieren?

- Welcher konkrete Nutzen soll durch die KI erzeugt werden?
- Welches Ergebnis soll/muss die KI haben?
- Welcher Grad an Genauigkeit wird dafür benötigt?
- Wer soll durch die KI profitieren?

- Worüber soll/wird das KI-System konkret entscheiden?
- Welches Wissen wird für diese Entscheidung benötigt?
- Wie kann dieses Wissen zustande kommen:
 - Welche Daten werden dafür benötigt?
 - Welches menschliche Fachwissen wird für eine sichere Entscheidung benötigt?
- Welche Folgen könnte eine Fehlentscheidung der KI haben?

Mögliche Antworten für den Use-Case: „Eifel barrierefrei erleben“.

- In welchem Kontext/Zusammenhang soll die KI einen Mehrwert generieren?
 Es sollen passgenaue Empfehlungen für barrierefreie Reiseerlebnisse in Nordrhein-Westfalen ermöglicht werden.
- Welcher konkrete Nutzen soll durch die KI erzeugt werden?
 Es sollen verlässliche und konkrete Informationen zu Angeboten für Menschen mit Beeinträchtigungen zur Verfügung gestellt werden.
- Welches Ergebnis soll/muss die KI haben?
 Die Daten sollen auf die Bedürfnisse und Anforderungen der jeweiligen Nutzer*innen zugeschnitten sein.
- Welcher Grad an Genauigkeit wird dafür benötigt?
 Ein hoher, da Menschen mit Beeinträchtigung sich nicht immer selber weiterhelfen können, wenn die Angaben der KI nicht stimmen.
- Wer soll durch die KI profitieren?
 - Beeinträchtigte Menschen, damit sie einschätzen können, ob bestimmte Angebote für sie nutzbar sind.
 - Familien mit Kinderwagen oder Personen mit schwerem Gepäck

- Tourismus NRW e.V, Eifel Tourismus (ET) GmbH
- Worüber soll/wird das KI-System konkret entscheiden?

Mögliche Antwort:

Über die Auswahl von möglichen Reiseerlebnissen für Menschen mit Beeinträchtigung, Familien mit Kinderwagen oder Personen mit schwerem Gepäck.

- Welches Wissen wird für diese Entscheidung benötigt?

Mögliche Antwort:

- Anzahl der möglichen Touristischen Attraktionen,
 - Wissen über die Umfeldbedingungen bei diesen Touristischen Attraktionen,
 - Wanderwege,
 - Veränderungen durch Wetterbedingungen,
 - Höhenverhältnisse,
 - mögliche Reisewege,
 - Reisemöglichkeiten im Öffentlichen Nahverkehr,
 - Kontexte der einzelnen Knotenpunkte im Öffentlichen Nahverkehr (Rampen, Treppen, Fahrstühle vorhanden etc.),
 - Aktuelle Behinderungen im Wegenetz (Baustellen, Umleitungen, Staus etc.),
 - Störungen im Betrieb (z.B. Ausfälle im Öffentlichen Nahverkehr, Polizei- oder Feuerwehreinsätze),
 - Kriterien von verschiedenen Beeinträchtigungen bei den Personen, die beraten werden sollen
- Wie kann dieses Wissen zustande kommen:
 - Welche Daten werden dafür benötigt?

Mögliche Antwort:

Daten über Attraktionen, Verkehrswege, den Öffentlichen Nahverkehr, Aktuelle Störungen, Wetterlage, Hilfsangebote, Wanderwege, Rettungsdienste etc.

- Welches menschliche Fachwissen wird für eine sichere Entscheidung benötigt?

Mögliche Antwort:

Fachwissen über das Reiseverhalten sowie die Anforderungen und

Bedürfnisse von Menschen mit Beeinträchtigungen und Familien mit Kinderwagen

- Welche Folgen könnte eine Fehlentscheidung der KI haben?

Mögliche Antwort:

Für beeinträchtigte Menschen und für Familien:

Sie haben keine schönen Ferien/Reiserlebnisse. Sie vertrauen ähnlichen digitalen Tools nicht mehr (auch wenn andere bessere Ergebnisse bringen). Im Ernstfall kann es auch zu Unfällen und schweren Verletzungen kommen, wenn die KI einen Risikobereich nicht angezeigt hat. Es liegt also eine potentielle Gefährdung der Gesundheit auf mehreren Ebenen vor: physisch und psychisch.

Für Tourismus NRW e.V und Eifel Tourismus (ET) GmbH:

Eine nachhaltige Schädigung des Rufs der Region als Reiseziel für beeinträchtigte Menschen und Familien sowie ein negatives Image der Tourismusbrand „Eifel Tourismus“ als Mit-Herausgeber der KI-Software.

Anna Hoffmann (Consultant und Prozessberaterin für die digitale Transformation, Agilität und Zukunftsfähigkeit; Hochschuldozentin und KI-Trainerin)

Autorisierte Prozessberaterin im Programm „UnternehmensWERT: Mensch PLUS“ des Bundesministeriums für Arbeit und Soziales zur Etablierung von agilen Prozessen
Consultant im Programm „Innovationen brauchen Mut“ der Wirtschaftsförderung Brandenburg
Agile Coach, Scrum Trainerin, Scrum Master, Scrum Product Owner
Senior Consultant Machwüth Team International

Agiles Business Coaching:

www.anna-hoffmann-coaching.de
info@anna-hoffmann-coaching.de

Zukunftsfähigkeit und Innovationen:

www.intense-impact.de
info@intense-impact.de

Handy: 0173 8914 579
Tel.: 03329 6981537

Büroanschrift:

Anna Hoffmann Business Consulting
Hermann-Scheidemann-Weg 7
14532 Stahnsdorf

Anhang

Bewertungsliste für vertrauenswürdige KI (Pilotversion)

Quelle: ETHIK-LEITLINIEN FÜR EINE VERTRAUENSWÜRDIGE KI, S. 34 - 43

[ethics_guidelines_for_trustworthy_ai-de_87FCEDA3-0D4B-1829-E9376DFoCB2DFFFE_60425.pdf](https://www.ethik-leitlinien.de/ethics_guidelines_for_trustworthy_ai-de_87FCEDA3-0D4B-1829-E9376DFoCB2DFFFE_60425.pdf), Abruf am 19.7.2022

„Verwenden Sie bei der Entwicklung, Einführung oder Verwendung eines KI-Systems die Bewertungsliste für vertrauenswürdige KI und passen Sie sie an den spezifischen Anwendungsfall an, auf den das System ausgerichtet ist.

Denken Sie daran, dass eine solche Bewertungsliste niemals vollständig sein wird. Bei der Gewährleistung einer vertrauenswürdigen KI geht es nicht darum, Kästchen anzukreuzen, sondern darum, kontinuierlich Anforderungen zu erkennen, Lösungen zu bewerten, die Ergebnisse während des gesamten Lebenszyklus des KI-Systems laufend zu verbessern und die Akteure in diesen Prozess einzubeziehen.“

1. Vorrang menschlichen Handelns und menschliche Aufsicht

Grundrechte:

Haben Sie für die Anwendungsfälle, bei denen die Möglichkeit einer Beeinträchtigung der Grundrechte besteht, eine Folgenabschätzung in Bezug auf die Grundrechte durchgeführt? Haben Sie mögliche Kompromisse zwischen den verschiedenen Grundsätzen und Rechten erkannt und dokumentiert?

Beeinflusst das KI-System die Entscheidungsfindung menschlicher Endnutzer?

(z. B. Empfehlungen für Handlungen oder Entscheidungen, Vorschlag verschiedener Möglichkeiten)

Besteht in diesen Fällen die Gefahr, dass das KI-System die menschliche Autonomie beeinträchtigt, indem es unbeabsichtigt in den Entscheidungsprozess des Endverbrauchers eingreift?

Haben Sie in Betracht gezogen, ob das KI-System den Nutzern mitteilen sollte, dass eine Entscheidung, ein Inhalt, eine Empfehlung oder ein Ergebnis die Folge einer auf Algorithmen basierenden Entscheidung ist?

Falls das KI-System über ein Chatbot oder ein Konversationssystem verfügt, werden die menschlichen Endnutzer auf die Tatsache aufmerksam gemacht, dass sie mit einem nicht-menschlichen Wesen, also einer Maschine, interagieren?

Vorrang menschlichen Handelns:

Falls das KI-System Teil des Arbeitsprozesses ist, haben Sie bei der Aufgabenverteilung zwischen dem KI-System und menschlichen Beschäftigten eine sinnvolle Interaktion und **eine angemessene menschliche Aufsicht und Kontrolle berücksichtigt?**

Verbessert oder erweitert das KI-System die menschlichen Fähigkeiten?

Haben Sie Vorkehrungen getroffen, um im Rahmen der Arbeitsprozesse einem übermäßigen Vertrauen in das KI-System vorzubeugen?

Menschliche Aufsicht:

Haben Sie das angemessene Maß an menschlicher Kontrolle für das jeweilige KI-System und den spezifischen Anwendungsfall überprüft?

Können Sie gegebenenfalls das Ausmaß an menschlicher Kontrolle bzw. menschlicher Teilnahme beschreiben? Welche Person ist mit der Kontrolle beauftragt und welche Werkzeuge stehen für den menschlichen Eingriff zur Verfügung?

Haben Sie Mechanismen und Maßnahmen eingeführt, um eine solche potenzielle menschliche Kontrolle oder Aufsicht zu gewährleisten oder **sicherzustellen, dass Entscheidungen unter der Gesamtverantwortung eines Menschen getroffen werden?**

Haben Sie Maßnahmen zur Nachprüfbarkeit und Verhinderung von Problemen im Zusammenhang mit der Steuerung und Kontrolle der Autonomie der KI ergriffen?

Haben Sie für den Fall eines selbstlernenden oder autonomen KI-Systems oder Anwendungsfalls spezifischere Vorkehrungen zur Kontrolle und Aufsicht eingeführt?

Welche Art von Erkennungs- und Reaktionsmechanismen haben Sie etabliert, um zu beurteilen, ob etwas schief gehen könnte?

Haben Sie eine „Stoptaste“ oder gegebenenfalls ein Verfahren zum sicheren Abbrechen eines Vorgangs sichergestellt? Wird dabei der Prozess ganz oder teilweise abgebrochen oder die Kontrolle an einen Menschen übergeben?

2. Technische Robustheit und Sicherheit

Schutz gegen Angriffe und Sicherheit:

Haben Sie mögliche Angriffsformen eingeschätzt, für die das KI-System anfällig sein könnte?

Haben Sie insbesondere verschiedene Arten möglicher Schwachstellen wie Datenverunreinigung, physische Infrastruktur, Cyberangriffe berücksichtigt?

Haben Sie Maßnahmen oder Systeme eingeführt, um die Integrität und Belastbarkeit des KI-Systems gegenüber möglichen Angriffen zu gewährleisten?

Haben Sie das Verhalten Ihres Systems in unerwarteten Situationen und Umgebungen bewertet?

Haben Sie in Erwägung gezogen, ob und inwieweit Ihr System einen doppelten Verwendungszweck (Dual-Use) haben könnte? Wenn ja, haben Sie geeignete vorbeugende Maßnahmen für diesen Fall ergriffen (**z. B. Verzicht auf die Veröffentlichung der Forschung oder Verzicht auf den Einsatz des Systems**)?

Auffangplan und allgemeine Sicherheit:

Haben Sie sichergestellt, dass Ihr System über einen befriedigenden Auffangplan verfügt, falls es feindlichen Angriffen oder anderen unerwarteten Situationen ausgesetzt ist (z. B. technische Umschaltverfahren oder das Heranziehen eines menschlichen Bedieners, bevor die Arbeit fortgesetzt wird)?

Haben Sie das vom KI-System verursachte Risiko in diesem speziellen Anwendungsfall berücksichtigt?

Haben Sie ein Verfahren zur Messung und Bewertung von Risiken und Sicherheit eingeführt?

Haben Sie die notwendigen Informationen für den Fall der Bedrohung der körperlichen Unversehrtheit von Menschen bereitgestellt?

Haben Sie eine Versicherung in Betracht gezogen, die mögliche, vom KI-System verursachte Schäden deckt?

Haben Sie die potenziellen Sicherheitsrisiken infolge der vorhersehbaren (und anderer) Anwendungen der Technologie, einschließlich eines versehentlichen oder böswilligen Missbrauchs, ermittelt? Gibt es einen Plan zur Begrenzung oder Minderung dieser Gefahren?

Haben Sie beurteilt, ob die Möglichkeit besteht, dass das KI-System den Anwendern oder Dritten Schaden zufügt? Wenn ja, haben Sie die Wahrscheinlichkeit, den potenziellen Schaden, den betroffenen Personenkreis und die Schwere des Schadens eingeschätzt?

Falls die Gefahr besteht, dass das KI-System Schäden verursacht, haben Sie die Haftungs- und Verbraucherschutzbestimmungen berücksichtigt? In welcher Art und Weise haben Sie dies getan?

Haben Sie die möglichen Auswirkungen oder Sicherheitsrisiken für Umwelt oder Tiere in Betracht gezogen?

Haben Sie bei Ihrer Risikoanalyse berücksichtigt, ob Sicherheits- oder Netzwerkprobleme (z. B. Bedrohungen der Cybersicherheit) ein Sicherheitsrisiko oder Schäden durch ein unbeabsichtigtes Verhalten des KI-Systems Schäden verursachen könnten?

Haben Sie die möglichen Auswirkungen eines Ausfalls Ihres KI-Systems eingeschätzt und ob dieser zu falschen Ergebnissen, zur Nichtverfügbarkeit Ihres Systems oder zu gesellschaftlich inakzeptablen Ergebnissen (z. B. diskriminierende Praktiken) führen könnte?

Haben Sie Schwellenwerte sowie Maßnahmen zur Lenkung und Kontrolle für die oben genannten Szenarien festgelegt, damit alternative oder Auffangpläne aktiviert werden?

Haben Sie Auffangpläne festgelegt und erprobt?

Präzision

Haben Sie beurteilt, welcher Grad an Präzision und welche Definition von Genauigkeit im Zusammenhang mit dem KI-System und dem jeweiligen Anwendungsfall erforderlich sind?

Haben Sie beurteilt, wie die Präzision gemessen und gewährleistet wird?

Haben Sie Maßnahmen ergriffen, um sicherzustellen, dass die verwendeten Daten umfassend und aktuell sind?

Haben Sie Maßnahmen ergriffen, um zu beurteilen, ob zusätzliche Daten erforderlich sind, z. B. um die Präzision zu verbessern oder Verzerrungen zu vermeiden?

Haben Sie den Schaden bewertet, der entstehen würde, wenn das KI-System ungenaue Vorhersagen machen würde?

Haben Sie Methoden eingeführt, um zu messen, ob Ihr System eine inakzeptable Anzahl nicht präziser Vorhersagen ausgibt?

Wenn ungenaue Vorhersagen gemacht werden, haben Sie die nötigen Schritte zur Lösung des Problems veranlasst?

Zuverlässigkeit und Wiederholbarkeit:

Haben Sie eine Strategie zur Überwachung und Erprobung eingeführt, die Ihnen zu erkennen hilft, ob Ihr KI-System die Ziele, Zwecke und vorgesehenen Anwendungen erfüllt?

Haben Sie getestet, ob bestimmte Kontexte oder Bedingungen berücksichtigt werden müssen, damit die Wiederholbarkeit gewährleistet ist?

Haben Sie Prüfverfahren oder -methoden zur Messung und Sicherstellung verschiedener Aspekte der Zuverlässigkeit und Wiederholbarkeit eingeführt?

Haben Sie Verfahren eingeführt, die beschreiben, wann ein KI-System bei bestimmten Einstellungen ausfällt?

Haben Sie diese Prozesse zur Erprobung und Prüfung der Zuverlässigkeit von KI-Systemen eindeutig dokumentiert und operativ umgesetzt?

Haben Sie Vorkehrungen getroffen oder ein Kommunikationssystem eingerichtet, mit dem Sie den (End-)Nutzern die Zuverlässigkeit des KI-Systems garantieren können?

3. Schutz der Privatsphäre und Datenqualitätsmanagement

Achtung der Privatsphäre und Gewährleistung des Datenschutzes:

Haben Sie je nach Anwendungsfall einen Mechanismus eingerichtet, der es anderen ermöglicht, Fragen der Privatsphäre oder des Datenschutzes im Zusammenhang mit den Prozessen der Datenerfassung (für Ausbildung und Einsatz) und Datenverarbeitung des KI-Systems zu kennzeichnen?

Haben Sie die Art und den Umfang der in Ihren Datensätzen enthaltenen Daten beurteilt (z. B. ob personenbezogene Daten enthalten sind)?

Haben Sie Methoden in Betracht gezogen, wie Sie das KI-System entwickeln oder das Modell ausbilden können, wobei keine oder eine minimale Anzahl von potenziell sensiblen oder personenbezogenen Daten verwendet werden?

Haben Sie je nach Anwendungsfall Vorkehrungen zur Kenntnisnahme und Kontrolle personenbezogener Daten getroffen (z. B. eine gültige Einwilligungserklärung und ggf. eine Widerrufsmöglichkeit)?

Haben Sie Maßnahmen zur Verstärkung der Privatsphäre ergriffen, z. B. durch Verschlüsselung, Anonymisierung und Aggregation?

Falls es einen Datenschutzbeauftragten (DSB) gibt, haben Sie diese Person frühzeitig in den Prozess einbezogen?

Qualität und Integrität der Daten:

Haben Sie Ihr System auf potenziell relevante Normen (z. B. ISO, IEEE) oder weitverbreitete Protokolle für Ihr tägliches Datenqualitätsmanagement abgestimmt?

Haben Sie Kontrollmechanismen für die Erhebung, Speicherung, Verarbeitung und Nutzung der Daten eingerichtet?

Haben Sie beurteilt, inwieweit Sie die Qualität der verwendeten externen Datenquellen unter Kontrolle haben?

Haben Sie Prozesse zur Gewährleistung der Qualität und Integrität Ihrer Daten eingeführt?

Haben Sie weitere Prozesse in Betracht gezogen? Wie stellen Sie sicher, dass Ihre Datensätze weder beeinträchtigt noch gehackt wurden?

Datenzugriff:

An welche Protokolle, Prozesse und Verfahren haben Sie sich gehalten, um ein ordnungsgemäßes Datenqualitätsmanagement sicherzustellen?

Haben Sie festgesetzt, wer unter welchen Umständen auf die Benutzerdaten zugreifen darf?

Haben Sie sichergestellt, dass diese Personen dazu berechtigt sind und es erforderlich ist, dass sie auf die Daten zugreifen, und dass sie über die notwendigen Kompetenzen verfügen, um die Einzelheiten der Datenschutzvorschriften zu verstehen?

Haben Sie einen Überwachungsmechanismus eingerichtet, mit dem protokolliert wird, wann, wo, wie, von wem und zu welchem Zweck ein Datenzugriff erfolgte?

4. Transparenz

Rückverfolgbarkeit:

Haben Sie Maßnahmen zur Gewährleistung der Rückverfolgbarkeit ergriffen?

Diese könnten folgende Dokumentation umfassen:

Verwendete Methoden für den Entwurf und die Entwicklung des algorithmischen Systems:

o Im Falle eines regelbasierten KI-Systems sollte die Art der Programmierung oder die Art und Weise, wie das Modell erstellt wurde, dokumentiert werden.

o Im Falle eines lernbasierten KI-Systems sollte die Methode zur Ausbildung des Algorithmus, einschließlich der Angabe, welche Eingabedaten erhoben und ausgewählt wurden und wie dies geschah, dokumentiert werden.

Verwendete Methoden für die Erprobung und Validierung des algorithmischen Systems:

- o Im Falle eines regelbasierten KI-Systems sollten die Szenarien oder Fälle, die zur Erprobung und Validierung verwendet wurden, dokumentiert werden.
- o **Im Falle eines lernbasierten Modells sollten Angaben zu den zur Erprobung und Validierung verwendeten Daten dokumentiert werden.**

Ergebnisse des algorithmischen Systems:

- o Die Ergebnisse des Algorithmus oder die vom Algorithmus getroffenen Entscheidungen sowie mögliche andere Entscheidungen, die sich aus anderen Fällen (z. B. für andere Benutzeruntergruppen) ergeben könnten, sollten dokumentiert werden.

Erklärbarkeit:

Haben Sie beurteilt, inwieweit die Entscheidungen und damit das Ergebnis des KI-Systems nachvollziehbar sind?

Haben Sie sichergestellt, dass eine Erklärung der Gründe, warum ein System eine bestimmte Wahl getroffen hat, die zu einem bestimmten Ergebnis führt, für alle Benutzer, die eine Erklärung wünschen, verständlich gemacht werden kann?

Haben Sie beurteilt, inwieweit die Entscheidung des Systems die Entscheidungsprozesse der Organisation beeinflusst?

Haben Sie beurteilt, warum dieses spezielle System in diesen speziellen Bereich eingeführt wurde?

Haben Sie das Geschäftsmodell im Zusammenhang mit diesem System bewertet (z. B. auf welche Weise es Wert für die Organisation schafft)?

Haben Sie das KI-System von Anfang an auf die Möglichkeit der Interpretierbarkeit ausgelegt?

Haben Sie recherchiert und versucht, das einfachste und am besten interpretierbare Modell für die jeweilige Anwendung zu verwenden?

Haben Sie geprüft, ob Sie die Daten, die Sie für die Ausbildung und Erprobung verwenden, analysieren können? Können Sie das im Laufe der Zeit ändern und aktualisieren?

Haben Sie beurteilt, ob Sie nach der Ausbildung und Entwicklung des Modells über Möglichkeiten zur Überprüfung der Interpretierbarkeit verfügen oder ob Sie Zugang zum internen Workflow des Modells haben?

Kommunikation:

Haben Sie den (End-)Nutzern – mit einem Haftungsausschluss oder anderweitig – mitgeteilt, dass sie mit einem KI-System und nicht mit einem anderen Menschen interagieren?

Haben Sie Ihr KI-System als solches ausgewiesen?

Haben Sie Mechanismen eingerichtet, mit denen Sie die Nutzer über die Gründe und Kriterien, die den Ergebnissen des KI-Systems zugrunde liegen, informieren?

Wird dies den angesprochenen Nutzern klar und verständlich mitgeteilt?

Haben Sie Prozesse eingeleitet, die die Rückmeldungen der Benutzer berücksichtigen und auf deren Grundlage das System angepasst wird?

Haben Sie auch über potenzielle oder tatsächlich erkannte Risiken, wie z. B. Verzerrungen, informiert?

Haben Sie je nach Anwendungsfall auch den Informationsaustausch und die Transparenz gegenüber anderen Zielgruppen, Dritten oder der Öffentlichkeit erwogen?

Haben Sie klargestellt, worin der Zweck des KI-Systems besteht und wer oder was von dem Produkt/der Dienstleistung profitieren kann?

Wurden die Anwendungsszenarien für das Produkt angegeben und klar verständlich mitgeteilt?

Wurden dabei auch alternative Kommunikationsformen in Betracht gezogen, um sicherzustellen, dass sie für die angesprochenen Benutzergruppen verständlich und angemessen sind?

Haben Sie je nach Anwendungsfall die menschliche Psychologie und mögliche Einschränkungen wie Verwechslungsgefahr, Bestätigungsfehler oder geistige Ermüdung in Betracht gezogen?

Haben Sie eindeutig die Merkmale, Grenzen und potenziellen Mängel des KI-Systems folgenden Personen mitgeteilt?

Bei der Entwicklung: denjenigen Personen, die das KI-System in ein Produkt oder eine Dienstleistung einbauen?

Bei der Einführung: den Endnutzern oder Verbrauchern?

5. Vielfalt, Nichtdiskriminierung und Fairness

Vermeidung unfairer Verzerrungen:

Haben Sie eine Strategie oder eine Reihe von Verfahren vorgesehen, um die Entstehung oder Verstärkung unfairer Verzerrungen im KI-System hinsichtlich sowohl der Verwendung der Eingangsdaten als auch des Entwurfs der Algorithmen zu vermeiden?

Haben Sie mögliche Einschränkungen, die sich aus der Zusammensetzung der verwendeten Datensätze ergeben könnten, bewertet und zur Kenntnis genommen?

Haben Sie beachtet, dass die Daten die Vielfalt der Nutzer widerspiegeln und repräsentativ sind?

Haben Sie das KI-System auf bestimmte Anwendergruppen oder problematische Anwendungsfälle getestet?

Haben Sie die verfügbaren technischen Hilfsmittel recherchiert und eingesetzt, um ein besseres Verständnis von Daten, Modell und Leistung zu erzielen?

Haben Sie während der Phasen der Entwicklung, Einführung und Nutzung des Systems Prozesse eingerichtet, mit denen Sie das System auf mögliche Verzerrungen untersuchen und überwachen können?

Haben Sie abhängig vom Anwendungsfall einen Mechanismus vorgesehen, der es anderen ermöglicht, Probleme im Zusammenhang mit Verzerrungen, Diskriminierung oder schlechter Leistung des KI-Systems zu kennzeichnen?

Haben Sie klare Schritte und Kommunikationswege darüber in Betracht gezogen, wie und an wen solche Themen herangetragen werden sollten?

Haben Sie außer den (End-)Nutzern auch weitere, möglicherweise indirekt vom KI-System Betroffene berücksichtigt?

Haben Sie untersucht, ob unter den gleichen Bedingungen die Entscheidungen bestimmten Schwankungen unterliegen könnten?

Wenn ja, haben Sie über die möglichen Ursachen nachgedacht?

Haben Sie in Bezug auf solche Schwankungen Vorkehrungen zur Messung oder Bewertung möglicher Auswirkungen auf die Grundrechte getroffen?

Haben Sie eine angemessene Arbeitsdefinition von „Fairness“ festgelegt, die Sie bei der Gestaltung von KI-Systemen anwenden?

Wird Ihre Definition üblicherweise verwendet? Haben Sie andere Definitionen in Betracht gezogen, bevor Sie sich für diese entschieden haben?

Haben Sie eine quantitative Analyse oder Metriken zur Messung und Prüfung der angewandten Definition von Fairness vorgesehen?

Haben Sie Mechanismen eingeführt, mit denen Sie gewährleisten können, dass Sie faire KI-Systeme einsetzen? Haben Sie andere mögliche Vorkehrungen in Betracht gezogen?

Barrierefreiheit und universeller Entwurf:

Haben Sie sichergestellt, dass das KI-System eine Vielzahl individueller Vorlieben und Fertigkeiten berücksichtigt?

Haben Sie geprüft, ob das KI-System für Menschen mit besonderen Bedürfnissen oder Behinderungen oder für von Ausgrenzung bedrohten Menschen geeignet ist? Wie wurden diese Gesichtspunkte in das System integriert und wie werden sie überprüft?

Haben Sie sichergestellt, dass die Informationen über das KI-System auch für Nutzer assistierender Technologien zugänglich sind?

Haben Sie diese Gruppe in die Entwicklungsphase des KI-Systems einbezogen oder konsultiert?

Haben Sie die Auswirkungen Ihres KI-Systems auf ihre potenzielle Zielgruppe berücksichtigt?

Ist das am Aufbau des KI-Systems beteiligte Team für Ihre Zielgruppe repräsentativ?

Repräsentiert das System unter Einbeziehung möglicherweise indirekt betroffener Gruppen die allgemeine Bevölkerung?

Haben Sie beurteilt, ob es Personen oder Gruppen geben kann, die überproportional von unerwünschten Auswirkungen betroffen sein könnten?

Haben Sie Rückmeldungen von anderen Teams oder Gruppen erhalten, die unterschiedliche Hintergründe und Erfahrungen repräsentieren?

Beteiligung der Interessenträger:

Haben Sie Vorkehrungen getroffen, um die Beteiligung verschiedener Interessenträger an der Entwicklung und Nutzung des KI-Systems sicherzustellen?

Haben Sie den Weg für die Einführung des KI-Systems in Ihre Organisation geebnet, indem Sie die betroffenen Arbeitnehmerinnen und Arbeitnehmer sowie deren Vertreter im Voraus informiert und in den Prozess einbezogen haben?

6. Gesellschaftliches und ökologisches Wohlergehen

Nachhaltige und umweltfreundliche KI:

Haben Sie während der Entwicklung, der Einführung und Nutzung des KI-Systems Maßnahmen zur Messung der Umweltauswirkungen eingeführt (z. B. Energieverbrauch des Rechenzentrums, Art der von den Rechenzentren verwendeten Energie usw.)?

Haben Sie Maßnahmen zur Reduzierung der durch Ihr KI-System während seines gesamten Lebenszyklus verursachten Umweltbelastung eingeführt?

Soziale Auswirkungen:

Falls das KI-System direkt mit Menschen interagiert:

Haben Sie beurteilt, ob das KI-System den Menschen ermutigt, sich mit dem System verbunden zu fühlen und Empathie zu entwickeln?

Haben Sie sichergestellt, dass das KI-System deutlich zu verstehen gibt, dass es soziale Interaktionen nur simuliert und dass es weder „verstehen“ noch „fühlen“ kann?

Haben Sie ein gutes Verständnis der sozialen Auswirkungen des KI-Systems sichergestellt?

Haben Sie beispielsweise das Risiko drohender Arbeitsplatzverluste und der Dequalifizierung von Arbeitnehmerinnen und Arbeitnehmer beurteilt? Welche Maßnahmen wurden ergriffen, um diesen Risiken entgegenzuwirken?

Gesellschaft und Demokratie:

Haben Sie die weiter reichenden gesellschaftlichen Auswirkungen der Nutzung des KI-Systems über die einzelnen (End-)Nutzer hinausgehend bewertet und dabei beispielsweise andere, möglicherweise indirekt betroffene Akteure berücksichtigt?

7. Rechenschaftspflicht

Nachprüfbarkeit:

Haben Sie Mechanismen eingeführt, die die Nachprüfbarkeit des Systems durch interne und/oder externe Prüfer erleichtern? Haben Sie z. B. die Rückverfolgbarkeit und Protokollierung der Prozesse und Ergebnisse des KI-Systems sichergestellt?

Minimierung und Meldung negativer Auswirkungen:

Haben Sie eine Risiko- oder Folgenabschätzung für das KI-System unter Berücksichtigung verschiedener, direkt oder indirekt betroffener Interessenträger durchgeführt?

Haben Sie zur Entwicklung von Verfahren zur Rechenschaftspflicht einen Rahmen für Aus- und Weiterbildungsmaßnahmen geschaffen?

Welche Mitarbeiterinnen und Mitarbeiter oder Abteilungen sind daran beteiligt? Gehen diese Maßnahmen über die Entwicklungsphase hinaus?

Werden im Rahmen dieser Schulungen auch die möglicherweise für das KI-System geltenden rechtlichen Rahmenbedingungen vermittelt?

Haben Sie die Einrichtung eines „Ethikausschusses“ oder etwas ähnliches in Betracht gezogen, um die umfassende Rechenschaftspflicht sowie ethische Verfahrensweisen einschließlich potenziell unklarer Grauzonen zu erörtern?

Gibt es neben internen Initiativen oder Rahmenbedingungen zur Überwachung von Ethik und Rechenschaftspflicht auch externe Leitlinien oder wurden zusätzliche Prüfverfahren eingeführt?

Sind für Drittpersonen (z. B. Lieferanten, Verbraucher, Händler/Anbieter) oder Mitarbeiterinnen und Mitarbeiter Verfahren zur Meldung potenzieller Schwachstellen, Risiken oder Verzerrungen des KI-Systems/der Anwendung vorgesehen?

Dokumentation von Kompromissen:

Haben Sie Vorkehrungen zum Erkennen relevanter Interessen und Werte im Zusammenhang mit dem KI-System und zum Feststellen möglicherweise erforderlicher Kompromisse getroffen?

Auf welche Prozesse stützen Sie sich, um Entscheidungen bezüglich solcher Kompromisse zu treffen?

Haben Sie die Dokumentation einer solchen Kompromisslösung sichergestellt?

Rechtsschutz:

Haben Sie angemessene Vorkehrungen zum Schadensausgleich im Falle eines Schadens oder nachteiliger Auswirkungen des KI-Systems getroffen?

Haben Sie Vorkehrungen getroffen, um die (End-)Nutzer/Dritte über ihre Möglichkeiten zum Einlegen von Rechtsmitteln zu informieren?