Forschungsbericht zur *Anwender-Studie* TrustKI

„Vertrauen als Wertmaß
für Vertrauenswürdigkeit"

gefördert durch

Bundesministerium
für Digitales
und Verkehr

if(is)
internet-sicherheit.

Informationen zur *Anwender-Studie TrustKI*

Studien-Design und -Autorin

Ulla Coester

Ergebnisauswertung

Yasin Zerria

Studien-Design und -Ausführung

Dominik Adler

Marcel Brauer

Prof. Dr. Norbert Pohlmann

Kontakt

Institut für Internet-Sicherheit - if(is), Westfälische Hochschule, Gelsenkirchen

Ulla Coester, Projektleiterin *TrustKI*

(Mobile) 0179 46 020 30

(E-Mail) coester@internet-sicherheit.de

(Webseite) www.vertrauenswürdigkeit.com

Genderhinweis:
Aus Gründen der besseren Lesbarkeit wird im Text auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für beide Geschlechter. Dies impliziert keine Benachteiligung des weiblichen Geschlechts, sondern ist im Sinne der sprachlichen Vereinfachung als geschlechtsneutral zu verstehen.

Präambel

Die konkrete Fragestellung, die im Auftrag des „Bundesministerium für Digitalisierung und Verkehr“ im Forschungsprojekt TrustKI untersucht wird lautet, ob eine „Vertrauenswürdigkeits-Plattform für KI-Lösungen“ geeignet ist den Entscheidungsprozess der Anwender zu vereinfachen. Unter den herrschenden Gegebenheiten ist das Erfordernis von Vertrauenswürdigkeit evident: Mit der zunehmenden Digitalisierung geht eine Komplexität einher, sodass es für Anwender immer schwieriger wird, die Wirkweise und Ziele von KI-Lösungen umfassend verstehen sowie einordnen zu können. Dies schränkt sie potenziell in ihrem Entscheidungsprozess ein. Um diesem Umstand konstruktiv entgegenzuwirken, ist es seitens der KI-Anbieter notwendig, entsprechende Maßnahmen zu ergreifen, damit es Anwendern möglich wird Vertrauen aufzubauen. Denn die durch Vertrauen evozierte Gewissheit – also die Annahme, dass es möglich ist, sich auf etwas Bestimmtes zu verlassen – ermöglicht generell, Komplexität zu reduzieren, weil dadurch die subjektive Überzeugung der Richtigkeit von Handlungen entsteht.

Basierend auf der Annahme, dass für diesen Prozess dezidierte Kriterien zum Aufbau eines Vertrauensverhältnisses erfüllt sein müssen, steht hierbei im Fokus zu untersuchen, was – im Kontext des Vertrauenswürdigkeits-Modells[1] – zur Beurteilung der Vertrauenswürdigkeit von KI-Anbietern geeignet ist.

Obwohl es grundsätzlich Normen und Gesetze gibt, die unter anderem dem Schutz der Privatsphäre dienen, sollten unserer Ansicht nach KI-Anbieter aus eigener Motivation auch legitim handeln und ihre diesbezügliche Intention auch offenlegen.

Unser erklärtes Ziel ist daher, nicht zuletzt auf Grundlage der Ergebnisse von *Anwender-Studien*, die Rahmenbedingungen dafür zu schaffen, indem wir eine Vertrauenswürdigkeits-Plattform etablieren, auf der KI-Anbietern unter anderem die Möglichkeit geboten wird, umfassend alle obligaten Informationen über ihr Unternehmen sowie ihre KI-Lösung darzulegen. Um hier eine Vergleichspräzision durchzusetzen, basiert der dazu notwendige Prozess auf einzelnen Vertrauenswürdigkeits-Aspekten[2], zu denen Fragen konzipiert werden, die aus Sicht der Anwender seitens der KI-Anbieter zwingend zu beantworten sind.

Die Intention dahinter ist, dass die KI-Anbieter den Anwendern ein Commitment – also ein wahrhaftiges Versprechen – bezüglich der Vertrauenswürdigkeit ihres Unternehmens und ihrer KI-Lösung geben.

Den Ausgangspunkt für unser Vorhaben stellt die *Anwender-Studie TrustKI* dar. In der entsprechenden Publikation werden die Ergebnisse der repräsentativen online erhobenen Befragung von 263 Führungskräften aus deutschen Anwenderunternehmen – die von Ende Juni bis Ende August durchgeführt wurde – vorgestellt und kurz kommentiert.

Der vorliegende *Forschungsbericht TrustKI* bietet eine ausführliche Interpretation aller relevanten Faktoren im Kontext der Analyse von Vertrauen und Vertrauenswürdigkeit gemäß spezifischer Sachverhalte.

[1] Die Vertrauenswürdigkeits-Aspekte im Kontext des Vertrauenswürdigkeits-Modells werden unter „Die Interdependenz zwischen Vertrauen und Vertrauenswürdigkeit“ (Kapitel 3) behandelt.

[2] Detaillierte Informationen zu den Vertrauenswürdigkeits-Aspekten sind unter „Die Interdependenz zwischen Vertrauen und Vertrauenswürdigkeit“ (Kapitel 3) sowie unter „Spezielle Aspekte im Kontext der Definition von Vertrauen und Vertrauenswürdigkeit“ (Kapitel 5.4) dargelegt.

Vertrauenswürdigkeit unter dem Aspekt der Ethik als offenkundiges Differenzierungsmerkmal

Im Folgenden sind die prägnanten Takeaways aus der *Anwender-Studie TrustKI*[3] aufgeführt:

- Generell lässt sich feststellen, dass die Teilnehmer seitens der KI-Anbieter ein wahrhaftiges *Commitment* bezüglich ihrer Vertrauenswürdigkeit erwarten – 75,1 Prozent (Top 2) der Teilnehmer sprechen sich dafür aus, dass ein KI-Anbieter die Dimension seines vertrauenswürdigen sowie werteorientierten Handelns in einer Kernbotschaft transparent und prägnant zusammengefasst darstellt.

- Der hohe Zuspruch zum *Standort Deutschland/Europa* legt nahe, dass die verlässlichen gesetzlichen Regelungen den Teilnehmern die Sicherheit geben, dass wenn der KI-Anbieter das in ihn gesetzte Vertrauen missbraucht, sie trotzdem in geregelter Form zu ihrem Recht kommen können.

- Die *Umsetzung der (ethischen) Sorgfaltspflicht*, die sich in erster Linie im *Nicht-Schädigungsprinzip* sowie dem Umgang mit Trainingsdaten manifestiert, scheint allgemein für die Anwender von hoher Relevanz zu sein. So ist es unter anderem für 91,8 Prozent (Top 2) der Teilnehmer substanziell, dass sie darüber informiert werden, auf welche Funktionalitäten ein KI-Anbieter zum Wohle seines Kunden verzichtet. 94,2 Prozent (Top 2) wollen nicht nur die Information erhalten, ob eine *Folgenabschätzung* vorgenommen wird, sondern auch weitere Details hinsichtlich der entsprechenden Umsetzung erfahren. Für 96,4 Prozent (Top 2) ist es wichtig, dass ein Störfall unmittelbar kommuniziert wird.

- Die *(ethische) Sorgfaltspflicht* spiegelt sich nach Ansicht der Teilnehmer im Umgang mit den Trainingsdaten wider. Das bezieht sich jedoch nicht nur auf deren Qualität, sondern auch auf den verantwortungsvollen Umgang mit diesen – so wollen 57,8 Prozent (Top 1) wissen, ob das Handling der Trainingsdaten gemäß ethischen Grundsätzen verläuft und ob die Anforderungen bezüglich Fairness sowie Gerechtigkeit erfüllt werden.

- Die detaillierten Ergebnisse in der Gesamtauswertung bezüglich des Vertrauenswürdigkeits-Aspekts *IT-Sicherheit* dokumentieren durchweg eine hohe Zustimmung seitens der Teilnehmer. Die Top 2-Werte liegen hier häufig – über alle segmentierten Gruppen hinweg – bei über 90,0 Prozent.

[3] Die relevanten Informationen zur Methodik der Auswertung sind unter „Studiendesign“ (Kapitel 4) aufgeführt.

- Prinzipiell hat sich im Rahmen der Studie gezeigt, dass bei der Beurteilung der *Vertrauenswürdigkeit* von KI-Anbietern sowie KI-Lösungen *bestimmte Merkmale* bei der Entscheidung eine Rolle spielen. Hier stechen in erster Linie Unterschiede zwischen den beiden *Altersgruppen* (jünger und älter als 49 Jahre) sowie zwischen den *KI-Kenntnisständen* und der bereits gemachten *Erfahrung mit KI* heraus. Dieser Sachverhalt könnte sich unserer Meinung nach auf die Urteilsfindung auswirken:

    - So ist es im Kontext des Vertrauenswürdigkeits-Aspekts *Zutrauen* für **Teilnehmer, die umfangreiche Kenntnisse** haben, wichtig zu erfahren, wie sichergestellt wird, dass bei *Personalfluktuation* das Wissen und die Kompetenz im Unternehmen erhalten bleibt. Dies lässt sich eventuell auf eine bereits gemachte Erfahrung, etwa dass die Qualität einer KI-Lösung maßgeblich von der Qualifikation des jeweiligen Entwicklers abhängt, zurückführen.

    - Bezüglich der Einhaltung *ethischer Grundprinzipien* haben vor allem die **Unternehmen, die KI-Lösungen bereits einsetzen oder sich mit einem möglichen Einsatz** beschäftigen sowie **Teilnehmer, die über umfangreiche Kenntnisse** verfügen, einen höheren Informationsbedarf den Umgang mit Trainingsdaten betreffend – letztere wollen zum Beispiel dezidiert wissen, ob und wie das Handling der Trainingsdaten gemäß ethischen Grundsätzen verläuft. Oder auch, in welchem Maße Daten abfließen können.

    - **Allen Teilnehmern, die über umfangreiche Kenntnisse** verfügen, ist es sehr wichtig oder wichtig (Top 2), über das Potenzial *unerwünschter Diskriminierung* in Kenntnis gesetzt zu werden.

    - Tendenziell ist es für **Teilnehmer mit umfangreichen Kenntnissen** relevanter Kenntnis darüber zu erhalten, wie ein KI-Anbieter *ethische Anforderungen* umsetzt. Hierzu fordern sie detaillierte Informationen – etwa dahingehend, ob ein *Ethik-Gremium* vorhanden ist und *Workshops mit Stakeholdern* durchgeführt werden. Das lässt darauf schließen, dass ihnen die Implikationen beim Einsatz von KI-Lösungen durchaus bewusster sind als solchen, die nicht über eine entsprechende Expertise verfügen.

Inhaltsangabe

1. Zielsetzung der Studie im Rahmen des Forschungsprojekts

Die steigende Komplexität im Zuge der digitalen Transformation allgemein sowie der Technologie – und hier im speziellen der Künstlichen Intelligenz (KI) – führt zunehmend dazu, dass Anwenderunternehmen vermehrt die notwendige Fachkompetenz zur Beurteilung von Technologien und deren Einsatz fehlt. Somit wird es für diese immer häufiger problematisch, Entscheidungsprozesse effizient absolvieren zu können. Ein weiterer Aspekt, der in diesem Kontext eine Rolle spielt, ist die Flut an (zum Teil auch widersprüchlichen) Informationen, die mitunter weder auf den tatsächlichen Bedarf der Anwender ausgerichtet sind noch diesen im notwendigen Maße gerecht werden.

Begründet durch den Fakt, dass dieser Dissens potenziell die Entscheidungsfindung erschwert, lässt sich deduzieren, dass Anwenderunternehmen Anbietern vertrauen müssen, da sie ansonsten in ihrer Handlungsfähigkeit eingeschränkt sind.

De facto wollen Anwender auch vertrauen können. Folgerichtig war das Ziel der *Anwender-Studie TrustKI* zu untersuchen, ob es grundsätzliche Anforderungen beziehungsweise Maßgaben für einen Vertrauensaufbau gibt – und falls dem so ist welche diese genau sind – damit die Anwenderunternehmen KI-Anbieter als vertrauenswürdig beurteilen.

Basierend auf der Annahme, dass für den Entscheidungsprozess spezifische Kriterien zur Beurteilung der Vertrauenswürdigkeit der KI-Anbieter relevant sind, stand somit im Fokus der Anwender-Studie TrustKI zu eruieren, ob die transparente Zurverfügungstellung aller, im Kontext der Vertrauenswürdigkeits-Aspekte relevanten, Informationen hierfür geeignet ist.

2. Definition wesentlicher Begriffe

Um ein einheitliches Verständnis der wesentlichen Begriffe zu gewährleisten, werden diese hier definiert.

Vertrauen

Der Fokus der Definition ist jeweils abhängig von der Perspektive – das bedeutet, ob hier das Hauptaugenmerk auf dem Anwender als Vertrauensgeber oder dem Anbieter als Vertrauensnehmer liegt – die für die Begriffsbestimmung von Vertrauen eingenommen wurde und somit für diese maßgeblich ist. Des Weiteren ist auch der Forschungsschwerpunkt von Bedeutung – Vertrauen wird unter anderem von Soziologen oder Psychologen untersucht, wodurch sich unterschiedliche Aspekte in der Ausprägung der Definition sowie der Hinzunahme von erklärenden Parametern ergeben.

Als eine grundsätzliche Definition des Vertrauens wird die von Karl Girgensohn (1921) gesehen, der gemäß Prof. Carsten Gennerich, EFH Darmstadt, als Urheber der empirischen Vertrauensforschung bezeichnet werden kann[4]:

> „Vertrauen ist ein Sichöffnen und Sicherschließen gegenüber dem Objekt des Vertrauens, und Vertrauen ist zweitens stets ein ‚Sichanvertrauen' an das Objekt des Vertrauens, gestützt auf der Zuversicht, dass das Objekt des Vertrauens richtig handeln könne und werde."

In weiteren Definitionen wird das Objekt des Vertrauens auch als Vertrauensnehmer bezeichnet.

Vertrauenswürdigkeit[5]

Der *Vertrauenswürdigkeit* werden viele Facetten attribuiert. Zudem hat sie – ebenso wie die Reputation – eine *relationale Eigenschaft*: der Beobachter, insbesondere der mögliche Vertrauensgeber, entscheidet darüber, ob er Vertrauenswürdigkeit zuschreibt oder nicht. In diesem Rahmen gilt es drei Aspekte der Vertrauenswürdigkeit näher zu betrachten:

- **Kompetenz**: ist gerade im Rahmen von Wertschöpfungsprozessen ein notwendiger Bestandteil für eine erfolgreiche Kooperationsbeziehung, aber *kein hinreichender Aspekt* für das Vertrauensphänomen.
- **Nicht-Opportunismus**: wird als *wichtigster Aspekt* der Vertrauenswürdigkeit eingeschätzt, weil sich darin die *Bereitschaft und Fähigkeit* des Vertrauensgebers ausdrückt, *situativen Versuchungen des Missbrauchs von Vertrauen zu widerstehen* und sich keine Vorteile zu Lasten des Vertrauensnehmers zu verschaffen.
- **Folgenabschätzung**: das *Einbeziehen der Auswirkungen des Handelns* auf Dritte, zum Beispiel weitere Vertrauensnehmer, ist zusätzlich ein wesentlicher Bestandteil der Vertrauenswürdigkeit. Das bedeutet, die Handlungen des Vertrauensnehmers sind auch unter dem Aspekt zu beurteilen, dass diese *nicht zu Lasten Dritter* gehen.

[4] Zitat gemäß Ulf Bernd Kassebaum, aus „Interpersonelles Vertrauen Entwicklung eines Inventars zur Erfassung spezifischer Aspekte des Konstrukts", Dissertation

[5] gemäß Suchanek, A. „Vertrauen als Grundlage nachhaltiger unternehmerischer Wertschöpfung"

Digitale Transformation

Unter dem Begriff ‚Digitale Transformation' versteht man erhebliche Veränderungen des Alltagslebens, der Wirtschaft und der Gesellschaft durch die Verwendung digitaler Technologien und Techniken sowie **deren Auswirkungen**.(https://wi-lex.de/index.php/lexikon/technologische-und-methodische-grundlagen/informatik-grundlagen/digitalisierung/digitale-transformation/)

Künstliche Intelligenz (KI)

Für den Begriff Künstliche Intelligenz (KI) wird in diesem Dokument die Definition der OECD verwendet: "Ein KI-System besteht aus drei Hauptelementen, aus Sensoren, einer operativen Logik und Aktoren. Die Sensoren sammeln Rohdaten über die Umgebung und die Aktoren beeinflussen den Zustand der Umgebung. Die operative Logik ist das wichtigste Element eines KI-Systems. Sie liefert für bestimmte Ziele ausgehend vom Daten-Input der Sensoren Output für die Aktoren. Dies können Empfehlungen, Vorhersagen oder Entscheidungen zur Beeinflussung der Umgebung sein." Diese Definition ist passend, weil sie sowohl logisch operierende KI-Lösungen (zum Beispiel Bilderkennung) als auch physisch operierende KI-Lösungen (zum Beispiel autonomes Fahren) umfasst."

Trainingsdaten

Die Definition von Trainingsdaten ist an die des EU AI Act angelehnt: Trainingsdaten werden zum Trainieren eines KI-Systems verwendet, um dessen Parameter anzupassen.

Nachvollziehbarkeit

Die Nachvollziehbarkeit von Entscheidungen und Ergebnissen einer KI-Lösung sind auf zwei Ebenen zu betrachten: Die Erklärbarkeit bezeichnet die Darstellung der Mechanismen, die der Funktionsweise eines Algorithmus zugrunde liegen, während sich die Interpretierbarkeit auf die Bedeutung der Ergebnisse von KI-Systemen im Zusammenhang mit dem vorgesehenen funktionalen Zweck bezieht.

Autonomiegrad

Der Autonomiegrad einer KI-Lösung bestimmt, inwieweit eine KI-Lösung eigenständig operieren kann und wann ein Mensch eingreifen kann und muss. Die Norm SAE J3016 beispielsweise klassifiziert 6 Autonomiestufen für Kraftfahrzeuge.

3. Die Interdependenz zwischen Vertrauen und Vertrauenswürdigkeit

Basierend auf der – eingangs gewählten – Definition zum Vertrauen und unter Berücksichtigung der Tatsache, dass es keine eineindeutige Definition gibt, beziehen wir uns im Weiteren auf universelle Annahmen, um den Vertrauensbegriff im Kontext der Vertrauenswürdigkeit zu operationalisieren: Die am weitesten verbreitete Perspektive spezifiziert Vertrauen als eine Frage der Vernunft und stellt den Vertrauensgeber – also einen Anwender – als (begrenzt) rationalen Entscheider in den Vordergrund. Gemäß dieser Definition ist der Anwender vermehrt bereit, vertrauenswürdige Interaktionspartner anhand bestimmter Kriterien als solche anzusehen. Zu diesen Kriterien zählen nach einem anerkannten Modell *Kompetenz*, *Integrität* und *Wohlwollen* des Vertrauensnehmers.

Daraus lässt sich ableiten, dass die Vertrauenswürdigkeit von KI-Anbietern als Vertrauensnehmer seitens der Anwender anerkannt werden muss. Somit besteht eine Interdependenz zwischen Vertrauen und Vertrauenswürdigkeit. Daher ist es für KI-Anbieter notwendig, die konkreten Anforderungen in Bezug auf die Kriterien beziehungsweise Aspekte zu kennen, anhand derer es Anwendern möglich ist Vertrauen aufzubauen. Das Kardinalproblem, um Vertrauenswürdigkeit zu erschließen, liegt nach Ansicht von Niklas Luhmann darin, die Perspektive der Vertrauensgeber zu erschließen.
Er formuliert es folgendermaßen: „Wer sich Vertrauen erwerben will, muss [..] in der Lage sein, fremde Erwartungen in die eigene Selbstdarstellung mit einzubauen.“ [6]

Relevant in diesem Kontext ist, dass die fremden Erwartungen – gemeint sind hier jene der Vertrauensgeber – systematisch erhoben werden. Diese Maßgabe fand im Rahmen der *Anwender-Studie TrustKI* dahingehend Berücksichtigung, dass die sieben *Vertrauenswürdigkeits-Aspekte* aus dem bereits *etablierten Vertrauenswürdigkeits-Modell* jeweils als Basis für die dedizierten Fragestellungen gewählt wurden.

Folglich kann im Weiteren das *Vertrauenswürdigkeits-Modell* dann zur *strukturierten (Selbst-)Darstellung* der Vertrauensnehmer Anwendung finden.

Konkretisierung des Vertrauenswürdigkeits-Modells

Wie bereits begründet, stellt aufgrund der Tatsache, dass Anwender kein vollständiges Wissen über die Funktionsweise einer KI-Lösung haben können, deren Nutzung für diese theoretisch eine Risikohandlung dar. Basierend auf dieser Erkenntnis resultiert konsequenterweise die Frage, was sichergestellt sein muss, um Anwender in die Lage zu versetzen, eine KI-Lösung zu nutzen. Hierfür gilt schlüssig die Prämisse der Vertrauenswürdigkeit, die seitens der KI-Anbieter – als Vertrauensnehmer – nachzuweisen ist.

Insgesamt wird somit einerseits die Relevanz von Vertrauen beim Einsatz innovativer Technologien dokumentiert und gleichzeitig die Notwendigkeit unterstrichen, dass KI-Anbieter vertrauenswürdig agieren müssen, damit dieses Vertrauen auch gerechtfertigt ist. Aus dieser Interdependenz lässt sich das Erfordernis eines Vertrauenswürdigkeits-Modells begründen, denn der Aufbau der Vertrauenswürdigkeit erfordert eine

[6] Niklas Luhmann, Dirk Baecker (Hrsg.): *Einführung in die Systemtheorie.* 5. Auflage. Carl Auer, 2009*, (Seite 80 f.)*

gezielte Vorgehensweise, bei der es gilt, die verschiedenen Aspekte gleichwertig unternehmensspezifisch zu analysieren und anschließend in einer Strategie umzusetzen.

Die umfassenden Zusammenhänge zum Aufbau von Vertrauen inklusive der relevanten Vertrauenswürdigkeits-Aspekte – die im Rahmen der *Anwender-Studie TrustKI* näher untersucht wurden – sind in dem Vertrauenswürdigkeits-Modell dargestellt (siehe Abbildung 1).

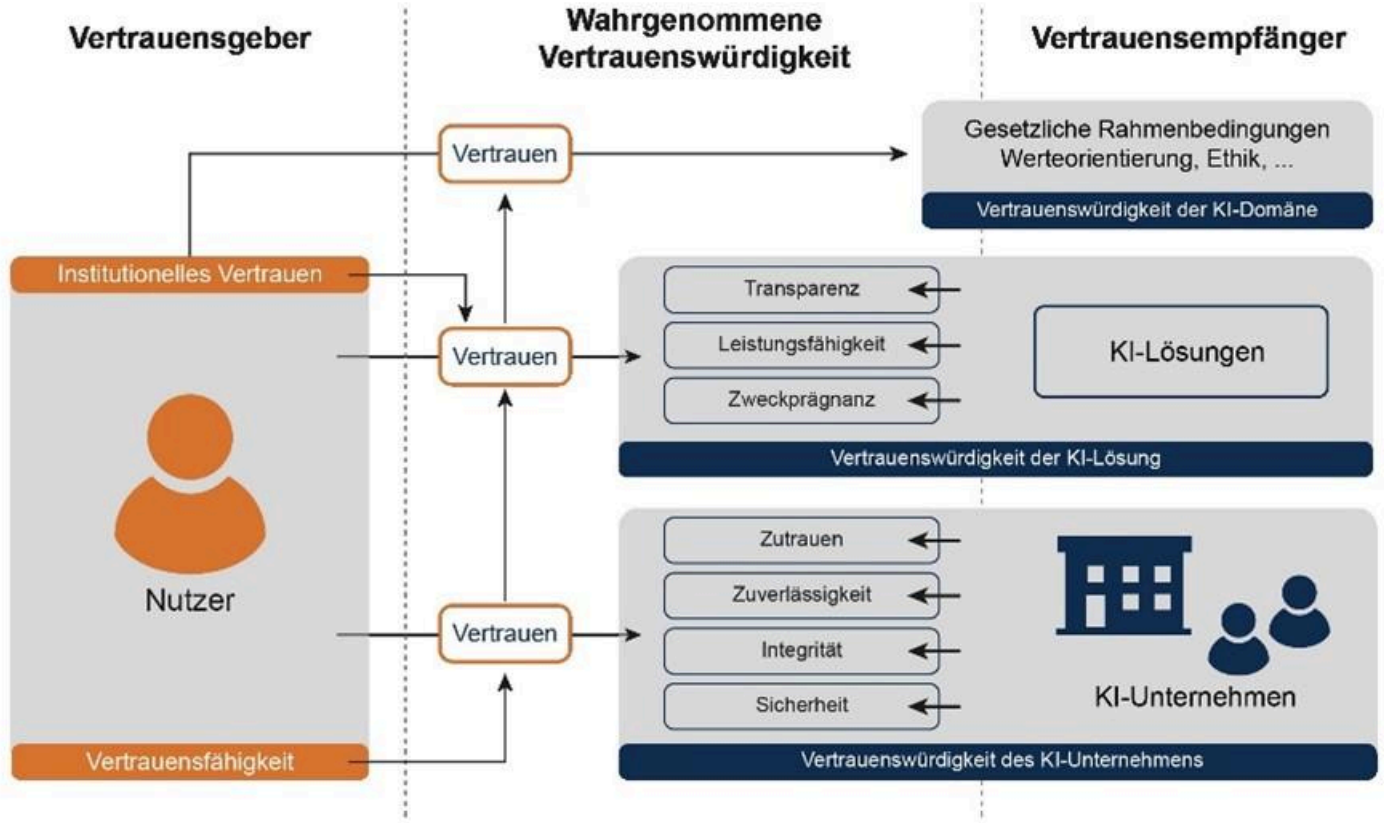


Abbildung 1: Vertrauenswürdigkeits-Modell

4. Studien-Design

Die *Anwender-Studie TrustKI* wurde vom 20. Juni bis zum 28. August 2023 online durchgeführt. Die Rekrutierung der überwiegend deutschen Teilnehmer erfolgte seitens der Projektmitarbeiter über die direkte Ansprache – unter anderem in beruflichen sozialen Netzwerken – oder in Kooperation mit einer Vielzahl an Verbänden. Zusätzlich wurde ein externes Marktforschungsinstitut mit der dedizierten Akquise von 100 weiteren Teilnehmern beauftragt, um die Stichprobengröße zu erhöhen.

Der bereinigte Datensatz umfasst 263 Teilnehmer.

Die Auswertung beruht vorrangig auf Grundlage der deskriptiven Statistik. Im Kern enthält die Umfrage drei Fragetypen:

- Skalenfragen mit der Ausprägung: „unwichtig“, „weniger wichtig“, „wichtig“ und „sehr wichtig“.
- Rating-/Rangfragen: zwischen 3 bis 8 Rängen, wobei Rang 1 jeweils die beste Platzierung darstellt. Bei Rating-/Rangfragen wird die Anzahl der Befragten sowie die Häufigkeit derer betrachtet, die entweder den ersten Rang (Top 1) oder den ersten und zweiten Rang (Top 2) gewählt haben.
- Zustimmungsfragen, die mit „Ja“ oder „Nein“ zu beantworten waren.

Zu Beginn der weiteren Analysen wurden für eine differenzierte Betrachtung die folgenden Untergruppen gebildet:

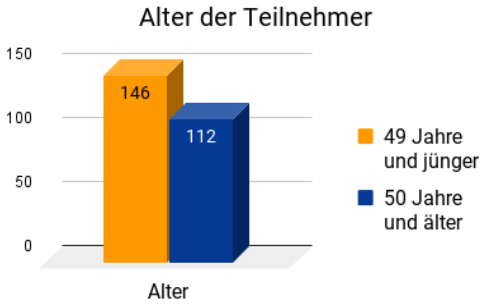
- Alter der Teilnehmer,
- Unternehmensart,
- KI-Kenntnisse,
- Compliance-Vorgaben,
- KI-Einsatz im Unternehmen sowie
- Vertrauenslevel.

Legende

Top 1: Prozentangaben mit "Top 1" geben bei Skalenfragen Auskunft über diejenigen, die die Ausprägung "sehr wichtig" gewählt haben im Verhältnis zu der Anzahl der Teilnehmer. Top 1 bei Rating-/Rangfragen gibt Auskunft über das Verhältnis der Teilnehmer, die den ersten Rang gewählt haben, und die Anzahl der Befragten.

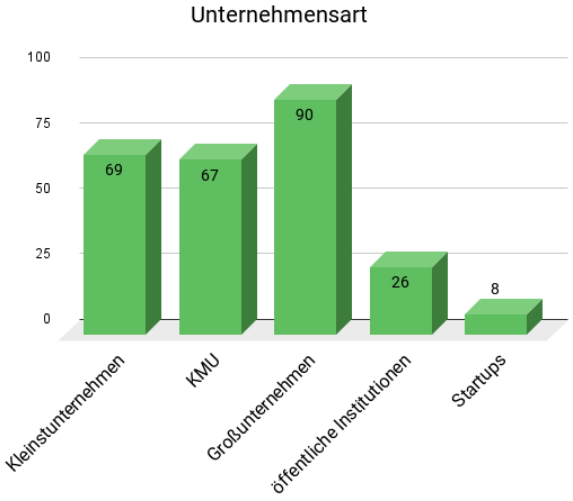
Top 2: Prozentangaben mit „Top 2“ geben bei Skalenfragen Auskunft über diejenigen, die die Ausprägung “sehr wichtig” oder “wichtig” gewählt haben im Verhältnis zu der Anzahl der Befragten. Top 2 bei Rating-/Rangfragen gibt Auskunft über das Verhältnis der Teilnehmer, die den ersten oder zweiten teilweise dritten Rang gewählt haben, und die Anzahl der Befragten.

Zustimmung: Der Begriff zeigt an, dass die Teilnehmer bei Zustimmungsfragen mit „Ja“ geantwortet haben.



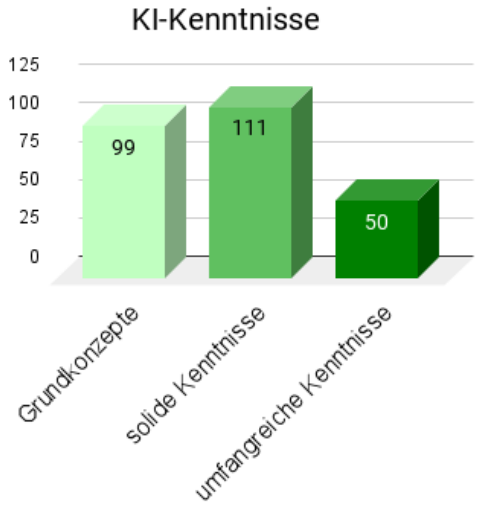
Alter der Teilnehmer

49 Jahre und jünger: N=146
50 Jahre und älter: N=112



Unternehmensart

Die drei Unternehmensarten Kleinstunternehmen, KMUs und Großunternehmen sind in der Stichprobe in ausreichender Größe vertreten, so dass sie separat betrachtet werden können. Öffentliche Institutionen und Startups konnten im Rahmen der Anwender-Studie nicht einbezogen werden, da in diesen Unternehmensarten der Schwellenwert N<30 lag.

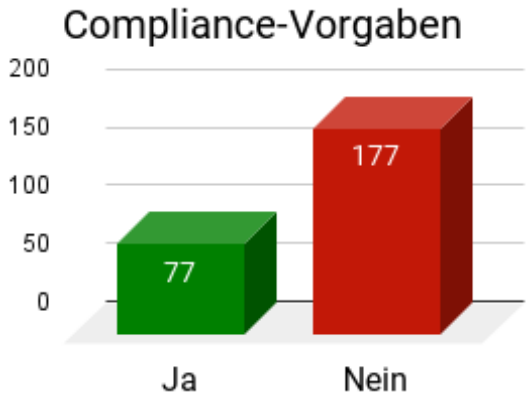


KI-Kenntnisse

Das Verhältnis zwischen Teilnehmern, die über Grundkenntnisse im Bereich KI verfügen, sowie jene mit soliden Kenntnissen ist relativ ausgeglichen. Ungefähr ein Fünftel der Teilnehmer zeichnet sich durch ein fundiertes Expertenwissen aus.

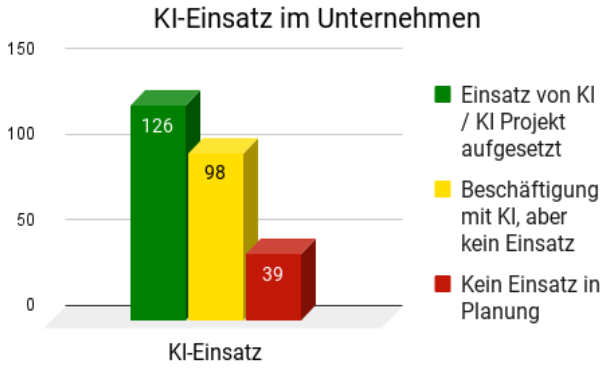
Compliance-Vorgaben

Etwa ein Drittel der Befragten gab an, dass sie in ihren Compliance-Vorgaben dezidiert Vorschriften für KI-Anbieter fixiert haben.



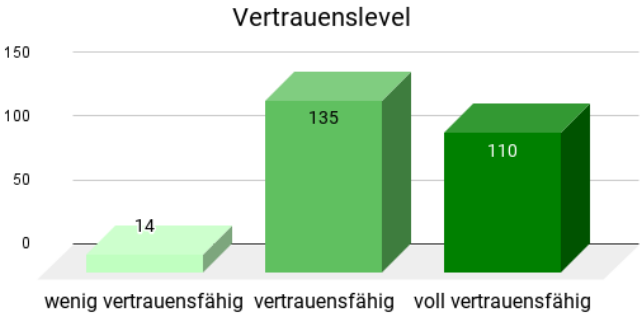
KI-Einsatz im Unternehmen

Ein Großteil der Teilnehmer setzt KI bereits im Unternehmen ein oder plant dies in naher Zukunft zu tun. Für circa 15 Prozent der Teilnehmer kommt dies perspektivisch nicht in Frage.



Vertrauenslevel

Da im Kontext von Studien zum Vertrauen das Vertrauenslevel relevant ist, wurde dies auf Grundlage der allgemein anerkannten, drei institutionellen Vertrauensfragen am Ende des Fragebogens ermittelt. Diese haben insgesamt fünf Ausprägungen, wodurch Teilnehmer zwischen einem und fünf Punkten vergeben können. Mittels Addition dieser Punkte wird das Vertrauenslevel ermittelt. Insgesamt verfügen 95 Prozent der Teilnehmer der Anwender-Studie über ein Maß an Vertrauensfähigkeit, das ihnen ermöglicht, Vertrauen aufzubauen.



5. Ergebnisse

5.1 Vertrauensfähigkeit und institutionelles Vertrauen – Auswertung grundlegender Parameter

Geringe Vertrauensfähigkeit & Misstrauen gegenüber KI-Anbietern

Der Aufbau von Vertrauen allgemein wird durch die Neigung zu vertrauen (Vertrauensfähigkeit) und die daraus resultierende Wahrnehmung über die Vertrauenswürdigkeit (einer Person) bestimmt.

Anhand der Ergebnisse unserer Anwender-Studie TrustKI lässt sich nachweisen, dass die allgemeine Vertrauensfähigkeit in Abhängigkeit zum Aufbau von Vertrauen in die KI steht. Menschen, die wenig vertrauensfähig sind, drücken insgesamt ein extrem starkes Misstrauen dem KI-Anbieter gegenüber aus und bekunden eine äußerst geringe Zustimmung bei der Frage, ob der KI-Lösung vertraut werden kann.

Hohe Vertrauensfähigkeit & hoher Informationsbedarf

Zur Operationalisierung von Vertrauen im Rahmen von KI wird im Weiteren das Modell des institutionellen Vertrauens als geeignete Grundlage angesehen.

Institutionelles Vertrauen lässt sich in diesem Kontext folgendermaßen ableiten: Dieses basiert auf der Annahme, dass das bereits gut erforschte Vertrauen in Institutionen sich in einen institutionsökonomischen Ansatz integrieren lässt. Vertrauen in Institutionen manifestiert sich zum einen über Regeln und Vorschriften, die einzuhalten sind. Zum anderen über die Vertrauenswürdigkeit der agierenden Personen im Bezugsrahmen von Institutionen und deren Motivation so zu handeln.[7]

Somit entwickelt sich institutionelles Vertrauen vorrangig aufgrund kognitiver Prozesse. Durch die Manifestierung von Regeln und Verfahren aller Art werden gleichermaßen Verständnis sowie Erwartungen geschaffen, die es Vertrauensgebern ermöglichen, Vertrauen aufzubauen.

Die Auswertung der Studie zeigt, dass ein hoher Bedarf seitens vertrauensfähiger/voll vertrauensfähiger[8] Teilnehmer[9] besteht, genau die Informationen seitens der KI-Hersteller zu erhalten, die aus ihrer Sicht in Bezug auf den Vertrauensaufbau relevant sind. Dies lässt sich daraus ableiten, dass vielfach auf keinerlei praktische

[7] Gemäß Maurer, A., Schmid, M. (Hrsg.), Neuer Institutionalismus, Campus Verlag Frankfurt/New York 2002

[8] Die Ergebnisse wurden auf Basis der allgemein anerkannten Trust Scale Questions der Vertrauensfähigkeit ermittelt. Die entsprechenden fünf Fragen wurden zum Abschluss des Fragebogens den Teilnehmern gestellt. Die Bewertung kann zwischen „stimme gar nicht zu“ bis „stimme voll und ganz zu“ vorgenommen werden. Zur besseren Vergleichbarkeit wurden die einzelnen Antworten addiert und in einer Skala wie folgt zusammengefasst: „wenig vertrauensfähig“, „vertrauensfähig“ und „voll vertrauensfähig“. In der weiteren Analyse dienen diese drei Abstufungen als eine Untergruppe. Für jede Frage wurde anhand der relativen Häufigkeit ein 2-Stichprobentest für die Anteilswerte durchgeführt.

[9] Da sich das Antwortverhalten beider Gruppen nur marginal voneinander unterscheidet, wurden diese zusammen ausgewertet.

Erfahrung – und daraus resultierend nicht die Möglichkeit zur eigenständigen Verifizierung von KI-Lösungen gegeben ist – zurückgegriffen werden kann.

Zudem wird einerseits generell nach guten Gründen für das Vertrauen aus Vernunft gesucht, andererseits nach Indizien der Verlässlichkeit. Diese Annahme lässt sich anhand der Anwender-Studie TrustKI belegen, denn es besteht ein hoher Informationsbedarf bezüglich der – im Kontext von KI – relevanten Kriterien wie Kompetenz, Wohlwollen und Integrität bei der KI-Entwicklung, auch wenn die Teilnehmer bei der Suche nach guten Gründen für ihr Vertrauen nicht notwendigerweise einen Bedarf an weiteren Einzelheiten bezüglich der konkreten Umsetzung der jeweiligen Kriterien haben.

Fokussierung des Informationsbedarfs

Teilweise besteht die Anforderung, dass Informationen komprimiert zur Verfügung gestellt werden sollen. Das zeigt sich beispielsweise bei dem Vertrauenswürdigkeits-Aspekt Integrität an einigen Fragen – exemplarisch an folgender: „Möchten Sie wissen, wie der KI-Hersteller die Anforderungen der Gesellschaft bezüglich Ethik konkret umsetzt?“. Hier gaben 77,6 Prozent der Teilnehmer an, dass dies für sie von Interesse ist. Bei der dezidierten Nachfrage bezüglich der Realisierung tendiert die Mehrzahl jedoch vorrangig dazu, keine ausführliche Auskunft diesbezüglich erhalten zu wollen.

Doch insgesamt lässt sich daraus keinesfalls induzieren, dass die Teilnehmer über wesentliche Fakten im Sinne des institutionellen Vertrauens prinzipiell keine Details erfahren wollen. So gibt es auch bei dem Vertrauenswürdigkeits-Aspekt Integrität im Rahmen der Fragestellungen erkennbar Aspekte, denen Teilnehmer eine höhere Bedeutung beimessen: Zum Beispiel konkret bezüglich der Umsetzung von ethischen Anforderungen der Gesellschaft (77,5 Prozent – Zustimmung), ob „ein Ethik-Gremium im Unternehmen etabliert wurde und welche Verantwortung sowie Befugnisse dieses reell hat“ – hier ist das Interesse an mehr Informationen tendenziell ausgeglichen. Gleiches gilt für die Frage, ob die Teilnehmer „etwas über die Prozesse erfahren möchten, die der KI-Hersteller etabliert hat, um zu gewährleisten, dass sich kein Mitarbeiter gegen die festgelegten (ethischen) Werte und Regeln hinwegsetzen kann“. Hier ist der Wert der Zustimmung mit 78,3 Prozent ebenfalls hoch – und auch bezüglich der korrespondieren Frage „ja, in Form einer transparenten Dokumentation bezüglich der jeweiligen Umsetzung seiner ethischen Werte über den gesamten Entwicklungsprozess“ kann die Interessenlage als relativ ausgewogen bezeichnet werden.

Zudem lässt sich an verschiedenen Stellen der Anwender-Studie TrustKI ablesen, dass spezifische Rahmenbedingungen, die im Kontext des institutionellen Vertrauens ausschlaggebend sind – wie etwa die Vertrauenswürdigkeit der Mitarbeiter – die Teilnehmer sehr interessieren: Zum Beispiel daran, dass im Rahmen des Vertrauenswürdigkeits-Aspekts Zutrauen den agierenden Personen im Bezugsrahmen der Institution und deren Motivation zu handeln eine hohe Bedeutung beigemessen wird. Einerseits wird die Kompetenz der Mitarbeiter als relevanter Faktor angesehen – so erachten beispielsweise 62,4 Prozent (Top 1) der Teilnehmer es als sehr wichtig, dass die Mitarbeiter über umfassende Branchenerfahrung und Kenntnisse im Anwendungsbereich der KI-Lösung verfügen. Andererseits messen knapp 60,0 Prozent (Top 2) der Teilnehmer mit der Forderung nach Soft Skills auch der damit assoziierten Handlungsweise, geprägt durch die kollektiven Wertvorstellungen der Mitarbeiter eine ausschlaggebende Bedeutung bei.

5.2 Holistische Perspektive der Transparenz – Auswertung von relevanten Parametern

Grundsätzlich wird im Kontext von KI der Begriff der Vertrauenswürdigkeit in hohem Maße mit der Forderung nach Transparenz assoziiert. Wobei sich diese momentan vorrangig auf die KI-Lösung bezieht und die Anforderungen nach Erklärbarkeit, Interpretierbarkeit und Nachvollziehbarkeit umfasst – sodass die, von der Anwendung getroffenen Entscheidungen auch klar zu deuten und darzulegen sind – oder Reproduzierbarkeit der Ergebnisse sowie allgemein auf deren Funktionalität.

Wie bereits dargestellt (vgl. Kapitel 3), liegt das Kernproblem bei dem Prozess zum Aufbau von Vertrauenswürdigkeit darin, die Perspektive der Vertrauensgeber zu erfassen. Analog zum Vertrauensaufbau können grundsätzliche Forderungen formuliert werden: Zum einen müssen die KI-Anbieter dem Faktor „Konsistenz“ eine hohe Priorität einräumen und zum anderen ist es essenziell, den Erwartungen der Anwender kontinuierlich gerecht zu werden.

Damit es Anwenderunternehmen möglich ist KI-Lösungen zu vertrauen gilt es somit, den bereits etablierten Maßstab zu erweitern – das heißt, die Vertrauenswürdigkeit sollte (oder kann) nicht einzig aus der Transparenz der KI-Lösung resultieren, sondern muss seitens der Anwender auf den KI-Anbieter extendiert werden können.

Hinsichtlich möglicher Prüfkriterien für die Vertrauenswürdigkeit des KI-Anbieters ist Transparenz in einem übergeordneten beziehungsweise holistischen Sinne von hoher Relevanz.

Denn unter dem Aspekt, dass der Anwender auf die Verlässlichkeit der jeweiligen KI-Lösung beziehungsweise auf die Kompetenz und Integrität sowie das wohlwollende Verhalten des KI-Anbieter vertraut, nimmt er – qua Definition – billigend ein gewisses Risiko in Kauf. Unter anderem gegeben durch den Sachverhalt, dass er grundsätzlich kein ausreichendes Expertenwissen im Kontext eines individuellen Anwendungsfalls haben kann, um eine spezifische KI-Lösung vollumfänglich zu evaluieren und somit stets auf – durch externe Quellen – zur Verfügung gestelltes Know-how vertraut werden muss. Infolgedessen erwächst aus der Nutzung von KI-Lösungen eine gewisse Vulnerabilität für Anwender, da es ihnen nicht möglich ist, bestimmte Gegebenheiten – etwa im Bereich der Privatsphäre – als Risikofaktor zu identifizieren.

Um daraus entstehende Konsequenzen für die Anwenderunternehmen erkennbar, bewertbar und möglicherweise reduzierbar machen zu können, ist die Forderung nach einer übergeordneten Transparenz, die auf Ethik referenziert – also mehr umfasst als den momentan allseitig erhobenen Transparenz-Anspruch an die KI-Lösung – nicht nur plausibel, sondern definitiv geboten. Denn die Implikationen resultierend aus dem Einsatz von KI allgemein, betreffen sowohl die Anwenderunternehmen als auch potenziell deren Kunden unmittelbar sowie mittelbar die Gesellschaft.

Diese Achillesferse wird mittlerweile auch seitens der Anwender so gesehen, wodurch der Wunsch nach einer holistischen Transparenz – die sich in allen sieben Vertrauenswürdigkeits-Aspekten manifestiert – besteht. Auf Grundlage dieses Leitgedankens haben wir 28 primäre Transparenzkriterien eruiert. Aus diesen wurden im Rahmen unserer Anwender-Studie nachfolgend jene Kriterien selektiert, die aus Sicht der Teilnehmer ihren Informationsbedarf decken und somit ausschlaggebend im Sinne einer insgesamt transparenten Darstellung sind.

Angaben zum Unternehmen

Angaben zum Standort sowie zur Unternehmensstruktur sind für alle Teilnehmer im Sinne der Vertrauenswürdigkeit relevant.

Für den Großteil der Befragten ist der Standort Deutschland sehr wichtig oder wichtig. Der hohe Zuspruch von 79,5 Prozent (Top 2) legt nahe, dass die verlässlichen gesetzlichen Regelungen – unter anderem bezüglich des Datenschutzes und der EU-Produktsicherheitsvorschriften – dem Anwender die Gewissheit geben, dass er in geregelter Form zu seinem Recht kommen kann, wenn der KI-Anbieter das in ihn gesetzte Vertrauen missbraucht. Zudem ist dies ein potenzieller Indikator dafür, dass dem gemeinsamen Wertverständnis eine große Bedeutung beigemessen wird. Dies lässt darauf schließen, dass der übergeordnete Begriff der Transparenz und die damit inhärent assoziierte Zuverlässigkeit im Rahmen der Vertrauenswürdigkeit der KI-Anbieter von hoher Relevanz ist.

Auch über die Unternehmensstruktur will ein Großteil der Befragten – präzise 82,3 Prozent (Top 2) – Informationen erhalten. Im Kontext der Bedeutung des Standorts könnte diese Frage einen weiteren Hinweis darauf geben, dass es Anwender als opportun erachten sicherzustellen, dass das Wertverständnis des KI-Anbieters mit ihrem eigenen kongruiert. Durch die transparente Darstellung der Unternehmensstruktur lässt sich erschließen, ob und wenn ja welche Personen, Organisationen beziehungsweise kulturellen Gegebenheiten konkret Einfluss auf den KI-Anbieter haben (könnten).

Wertekodex

Im Kontext der KI gibt es etliche Auslegungen dahingehend, an welchen Grundwerten sich Anwenderunternehmen beim Einsatz prinzipiell orientieren sollten. Im Rahmen der Anwender-Studie TrustKI wurden diesbezüglich generelle Werte identifiziert.

Wertekodex – Absichtserklärung des KI-Anbieters

Generell lässt sich feststellen, dass die Teilnehmer seitens der KI-Anbieter ein wahrhaftiges Commitment bezüglich ihrer Position hinsichtlich des vertrauenswürdigen und ethischen Handelns erwarten – 75,1 Prozent (Top 2) der Teilnehmer sprechen sich dafür aus, dass ein KI-Anbieter seine Werte sowie Handlungsweise in einer Kernbotschaft transparent und prägnant zusammengefasst darstellt.

Wertekodex – Grundlage zur Einhaltung ethischer Anforderungen

Auf die Frage, ob der KI-Anbieter erläutern sollte, welche Prozesse er etabliert hat, um zu gewährleisten, dass sich kein Mitarbeiter gegen die festgelegten (ethischen) Werte und Regeln hinwegsetzen kann, sagen 78,3 Prozent (Top 2) der Befragten, dass sie dies wissen möchten. Allerdings besteht kein Interesse daran zu erfahren, wie der KI-Anbieter dies gedenkt umzusetzen – die Teilnehmer wollen lediglich die Gewissheit haben, dass diese Prozesse bestehen und der KI-Anbieter somit den grundsätzlichen Anforderungen gerecht werden kann. Des Weiteren hat die Frage, ob der KI-Anbieter darüber Auskunft geben soll, wie er sicherstellt, dass seine Mitarbeiter die vorgegebenen ethischen Werte umsetzen (können) eine hohe Bedeutung für die

Anwender, wie der Wert von 74,9 Prozent (Top 2) zeigt. Aber auch bei diesem Aspekt ist es für die Teilnehmer nicht entscheidend, weitere Details zu erfahren – die höchste Nennung diesbezüglich liegt mit 42,6 Prozent (Zustimmung) bei der Forderung nach einem Code-of-Conduct, in dem die Handlungsvorgaben beschrieben werden.

Wertekodex – Einhaltung essenzieller Werte

Privatheit

Für die Teilnehmer ist es substanziell, dass sie bei der Nutzung von KI die Kontrolle über die Abgabe ihrer Daten behalten – das sagen immerhin 72,6 Prozent (Zustimmung). Der Wunsch nach Kontrolle über die Daten auf allen Ebenen ist nachvollziehbar, da so die potenzielle Vulnerabilität der Anwender minimiert wird – denn ohne Kontrollmöglichkeiten entsteht eine Macht-Asymmetrie zwischen KI-Anbietern (und hier vor allem den großen Technologie-Unternehmen) und Anwendern. Diese Macht-Asymmetrie wird möglicherweise verstärkt, wenn die Privatheit durch AGB ausgehebelt werden kann – dieser Fakt ist auch den Anwendern bewusst, von daher legen 76,8 Prozent (Zustimmung) der Teilnehmer Wert darauf zu erfahren, ob dies möglich ist und der KI-Anbieter so ermächtigt würde, Daten an Dritte weiterzuverkaufen. Dabei ist es insgesamt für die Teilnehmer elementar, dass der KI-Anbieter genau offenlegt, wie er mit den Daten umgeht – ob er diese für eigene (Werbe-)Zwecke verwendet (89,4 Prozent – Top 2), diese alternativ an Dritte weitergibt (92,3 Prozent – Top 2) oder eventuell Regierungsbehörden überlässt (94,9 Prozent – Top 2).

Somit wird von 51,7 Prozent (Zustimmung) der Anwender darauf Wert gelegt, dass sie mehr Informationen zum verantwortungsvollen Umgang mit der Technologie explizit im Kontext der Privatheit erhalten, auch weil dies wesentlich ist im Hinblick auf mögliche Manipulation.

Autonomie

Anhand der Ergebnisse bei der konkreten Frage nach Autonomie und Selbstbestimmung der Anwender mit 43,7 Prozent (Zustimmung) diesbezüglich lässt sich keine übermäßige hohe Relevanz nachweisen. Jedoch zeigt sich an anderer Stelle die Bedeutung, die der Autonomie eingeräumt wird, daran, dass 95,3 Prozent (Top 2) die Aufklärung darüber, was bei den einzelnen Autonomiegraden zu beachten ist, als wichtig einschätzen.

Wertekodex – Umsetzung der (ethischen) Sorgfaltspflicht

Nicht-Schädigung

Dem Anerkennen beziehungsweise der Umsetzung des Nicht-Schädigungsprinzips im Kontext der Integrität schreiben die Teilnehmer allgemein eine hohe Bedeutung zu. So ist es für 91,8 Prozent (Top 2) substanziell, dass sie darüber informiert werden, auf welche Funktionalitäten der KI-Anbieter zum Wohle des Kunden verzichtet. 94,2 Prozent (Top 2) wollen nicht nur mehr darüber erfahren, ob eine Folgenabschätzung vorgenommen, sondern auch wie diese umgesetzt wird – unter anderem interessieren sich 90,7 Prozent (Top 2) dafür, ob in diesem Kontext Assessments mit Stakeholdern stattfinden. Dies zeigt, dass ein wohlwollendes Verhalten der KI-Anbieter im Sinne der Anwender erwartet wird und dass es möglich ist, den KI-Anbieter an seinem Verhalten diesbezüglich zu messen. Diese Annahme erschließt sich des Weiteren auch daraus, dass 53,2 Prozent (Top 1) der Teilnehmer von KI-Anbietern eine zielgruppengerechte Aufklärung über potenzielle

Folgen erwarten. Da sich die Teilnehmer offenbar auch darüber bewusst sind, dass es Konflikte sowohl bei der Entwicklung als auch bei dem Inverkehrbringen der Lösung geben kann, wollen 79,1 Prozent (Zustimmung) darüber informiert werden, welche Implikationen genau entstehen können – etwa dahingehend, ob ökonomische Sachzwänge über ethische Werte gestellt werden. Zur Rechenschaftspflicht zählt für 68 Prozent (Zustimmung) unter anderem, dass der KI-Anbieter Auskunft darüber gibt, was er unternimmt, um physische, psychische oder finanzielle Schädigungen – die bei der Nutzung der KI-Lösung potenziell auftreten – abzuwenden. Die Dimension des Nicht-Schädigungsprinzips wird durch weitere Aussagen belegt, so sagen 95,7 Prozent (Top 2) aus, dass sie über Risiken bezüglich der Verletzung der physischen, psychischen und finanziellen Unversehrtheit im Vorfeld informiert werden möchten.

Im Sinne des Nicht-Schädigungsprinzips spielt nicht zuletzt der adäquate Umgang mit Daten in zweierlei Hinsicht eine wichtige Rolle: Zum einen im Kontext der IT-Sicherheit, also dass keine wertvollen Daten aus dem Unternehmen abfließen können – was für 95,7 Prozent (Top 2) der Teilnehmer von elementarer Bedeutung ist. Zum anderen die inadäquate Nutzung von Daten, da hier nach Meinung von 87 Prozent (Top 2) unter anderem ein Zusammenhang besteht bezüglich des Potenzials für unerwünschte Diskriminierung. Beide Sachverhalte implizieren, dass dadurch ein großer Schaden nicht nur subjektiv für ein Unternehmen, sondern ebenso für die jeweiligen Kunden sowie insgesamt die Gesellschaft entstehen kann.

Zusammenfassend lässt sich aus diesen Ergebnissen ableiten, dass mitnichten nur die Kompetenz der KI-Anbieter eine Rolle spielt, sondern auch deren Verantwortungsbewusstsein dahingehend, alles zu tun, um sicherzustellen, dass den Anwendern kein Schaden zugefügt wird – was einmal mehr die Interdependenz zwischen Vertrauen und Vertrauenswürdigkeit bestätigt.

Diese Einsicht lässt sich analog auf die KI-Lösung abbilden. Für 93,4 Prozent (Top 2) der Anwender ist es wichtig, dass kompetente Mitarbeiter mit Branchenerfahrung und Kenntnissen im Anwendungsbereich zur Verfügung stehen und für 95,3 Prozent (Top 2) gleichzeitig auch, dass ausreichend Mitarbeiter mit KI- und Datenkompetenz im Unternehmen angestellt sind. Das könnte daraus resultieren, dass diese als relevante Indikatoren für Fehlerfreiheit und Verlässlichkeit der KI-Lösungen gewertet und somit ebenfalls im Sinne der Vermeidung oder Einschränkung möglicher Schadenspotentiale als bedeutsam erachtet werden.

Da eine KI-Lösung, die weder akkurat noch zuverlässig arbeitet, einen Schaden verursachen könnte, ist es nachvollziehbar, dass es für 96,4 Prozent (Top 2) der Teilnehmer substanziell ist, unmittelbar über einen Störfall informiert zu werden.

Trainingsdaten

Aufgrund der Tatsache, dass die Bedeutung von Trainingsdaten mittlerweile sehr stark in der breiten Öffentlichkeit kommuniziert wurde, lässt sich erklären, dass 92,2 Prozent (Top 2) der Anwender die Qualität der Trainingsdaten als sehr wichtig oder wichtig erachten.

Doch nicht nur der Qualität auch dem verantwortungsvollen Umgang mit den Trainingsdaten wird seitens der Anwender eine hohe Bedeutung beigemessen – so wollen 57,8 Prozent (Zustimmung) der Anwender wissen, ob das Handling der Trainingsdaten gemäß ethischen Grundsätzen verläuft und die Anforderungen bezüglich Fairness sowie Gerechtigkeit erfüllt werden.

5.3 Holistische Transparenz – Selektion des relevanten Informationsbedarfs

Im Kontext der Transparenz sind sowohl deren Art als auch Umfang von Bedeutung. Gemäß dem funktionalistischen Ansatz von Niklas Luhmann[10] kommt Vertrauen durch den Rückgriff auf Informationen zustande und dient der Reduktion von Komplexität. Doch der Mensch – als informationsverarbeitendes Wesen – kann nur handlungsfähig werden, wenn es ihm gelingt, angemessene Formen der Informationsreduktion zu entwickeln. Basierend auf dieser Annahme wurden in einem weiteren Schritt jene Merkmale ermittelt, die innerhalb der Transparenz dazu beitragen, valide dem Wissensbedarf der Anwender gerecht zu werden.

Folgerichtig bestand der nächste Analyseschritt darin, aus den identifizierten Transparenzkriterien jene Merkmale zu eliminieren, die nicht weiter zur Differenzierung beitragen und somit keine zwingende Notwendigkeit besteht hier entsprechende Informationen aufzubereiten, da dies für den Prozess des Vertrauensaufbaus nicht unmittelbar nutzbringend wäre.

Zu diesem Zweck erfolgte für jede einzelne Frage zunächst die Berechnung der relativen Häufigkeiten (ohne „keine Angabe“), die auf den sogenannten Top 2-Werten basieren. Bei Skalenfragen lauten diese „sehr wichtig“ und „wichtig“, bei Ratingfragen „Rang 1“ und „Rang 2“ und bei Zustimmungsfragen wurden ausschließlich „Ja“-Antworten berücksichtigt.

Die Reduktion der Merkmale ließ sich mittels folgender Vorgehensweise durchführen: Im ersten Schritt fand der Vergleich der Einschätzung aller Teilnehmer auf Basis der Top 2-Werte mit den einzelnen Transparenzparametern[11] in einer neu erstellten Kreuztabelle statt. Im zweiten Schritt wurde festgelegt, dass nur diejenigen beibehalten werden, die eine Differenz von 5 Prozent oder mehr aufweisen.

Darauf basierend erfolgte eine Evaluierung dahingehend, welche Kombination von Transparenzkriterien und -merkmalen eine stochastische Abhängigkeit besitzt.[12] Nach Durchführung des Tests sowie der Überprüfung der Bedingung verbleiben die nachfolgend genannten fünf Transparenzparameter

- Absichtserklärung des KI-Anbieters
- Einhaltung ethischer Anforderungen
- Privatheit
- Autonomie
- Trainingsdaten (Handling nach ethischen Grundsätzen)

sowie die drei subsumierten Merkmale

- Angaben zum Code of Conduct [Einhaltung ethischer Anforderungen]
- Folgenabschätzung
- Rechenschaftspflicht [Umsetzung der (ethischen) Sorgfaltspflicht]

[10] Niklas Luhmann, Dirk Baecker (Hrsg.): Einführung in die Systemtheorie. 5. Auflage. Carl Auer, 2009, (Seite 80 f.)

[11] Beispiel: Der Aussage „Der KI-Hersteller legt Wert darauf, dass seine Belegschaft divers in Alter, Geschlecht, Herkunft zusammengesetzt ist.“ stimmen 51 Prozent aller Befragten zu. Die Teilnehmer, die im Transparenzparameter „Standort Deutschland“ angegeben haben, der Standort Deutschland sei ihnen (sehr) wichtig, stimmen dieser Aussage zu 58 Prozent zu. Die Differenz zum Totalwert beträgt in diesem Fall somit 7 Prozent.

[12] Zu diesem Zweck wird ein Chi-Quadrat-Unabhängigkeitstest für alle Paare durchgeführt. Der Fokus liegt hierbei auf dem zugehörigen P-Wert. Liegt der Signifikanzwert unter 5 Prozent, kann von einer stochastischen Abhängigkeit gesprochen werden.

Korrelierter Informationsbedarf

Für die vorhandenen Wertepaare aus dem vorangegangenen Schritt fand eine Evaluation hinsichtlich ihrer Korrelation statt.[13]

In einem entsprechenden Arbeitsvorgang wurden von den 229 vorselektierten Fragen jene eliminiert, die – obwohl eine Korrelation vorliegt – im Kontext der Kernfrage in keinem Sinnzusammenhang stehen.[14] Dies führte zu einer Reduktion auf 139 Fragen, die jedoch unterschiedlich stark korrelieren, woraus sich eine entsprechende Relevanz bezüglich deren Beantwortung bemisst.[15]

Dezidierte Angaben zur Dokumentation der Vertrauenswürdigkeit des KI-Anbieters

Mithilfe des vorab beschriebenen Prozesses konnten dediziert Fragen (im Folgenden „Korrespondierende Fragen“ genannt) extrahiert werden, deren Beantwortung zwingend erforderlich ist zur Dokumentation der Vertrauenswürdigkeit des KI-Anbieters.

Absichtserklärung des KI-Anbieters

Transparenzmerkmal: Wie wichtig ist es für Sie, dass der KI-Hersteller Ihnen seine Werte und Handlungsweise prägnant in einem Satz als Kernbotschaft mitteilt?

- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant sind:
  - *„Wie wichtig ist es für Sie, dass der KI-Hersteller sich mit den ethischen Werten der Gesellschaft konstant auseinandersetzt?“*
  - *„Der KI-Hersteller legt Wert darauf, dass seine Belegschaft divers bezüglich Alter, Geschlecht und Herkunft zusammengesetzt ist.“*

Einhaltung ethischer Werte

Transparenzmerkmal: Möchten Sie wissen, auf welchem Wege der KI-Hersteller sicherstellt, dass seine Mitarbeiter die vorgegebenen ethischen Werte umsetzen (können)?

- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant ist:
  - *„Möchten Sie etwas über die Prozesse erfahren, die der KI-Hersteller etabliert hat, um zu gewährleisten, dass sich kein Mitarbeiter über die festgelegten (ethischen) Werte und Regeln hinwegsetzen kann?“*

[13] Dazu werden die jeweiligen Korrelationskoeffizienten und entsprechende P-Werte berechnet, um zu den Transparenzparametern und -merkmalen die jeweils relevante(n) Kernfrage(n) zu ermitteln. Auch hier wurde die Signifikanzgrenze von unter 5 Prozent festgelegt.

[14] Dies basiert auf der Annahme, dass es sich um eine Scheinkorrelation und weitere darunter liegende Faktoren handelt.

[15] Die Fragen lassen sich gemäß Cohen aufgrund ihrer Stärke wie folgt unterteilen: Zwei Fragen haben mit einer Korrelation von über 0,67 den höchsten Effekt, 50 Fragen liegen zwischen 0,3 und 0,5 und haben demnach einen mittleren Effekt, während 87 Fragen einen schwachen Effekt haben, weil sie zwischen 0,1 und 0,3 liegen.

- *„Möchten Sie wissen, wie der KI-Hersteller die Anforderungen der Gesellschaft bezüglich Ethik konkret umsetzt?"*

Fakten bezüglich Handlungen im Code-of-Conduct

Transparenzmerkmal: Fakten über die Handlungsvorgaben im Code-of-Conduct des KI-Herstellers.

- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant ist:
  - *„Fakten – in Form einer transparenten Dokumentation bezüglich der jeweiligen Umsetzung seiner ethischen Werte über den gesamten Entwicklungsprozess.“*
  - *„Möchten Sie wissen, auf welchem Wege der KI-Hersteller sicherstellt, dass seine Mitarbeiter die vorgegebenen ethischen Werte umsetzen (können)?*

Informationen zum verantwortungsvollen Umgang

Transparenzmerkmal: Mehr Informationen zum verantwortungsvollen Umgang mit der Technologie unter dem Aspekt, dass die Privatheit sichergestellt wird, auch um eine Manipulation der Anwender zu verhindern.

- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant ist:
  - *Fakten zum Umgang bei einem Konflikt zwischen ethischen Werten und ökonomisch sinnvollem Handeln, zum Beispiel wenn die Etablierung eines hohen Sicherheitsgrades zu viele potenzielle Anwender abschrecken würde.*
  - *Fakten dazu, welche Maßnahmen ergriffen werden, um zyklisch die Einhaltung vorgegebener Regeln zu evaluieren.*

Rechenschaftspflicht

- Transparenzmerkmal: „Der KI-Hersteller ist bereit Auskunft darüber zu geben, was er unternimmt, um physische, psychische oder finanzielle Schädigungen der Anwender (zum Beispiel „Abkapseln von der Umwelt durch übermäßigen Gebrauch von Chatbots“), die bei der Nutzung seiner KI- Lösung auftreten könnten, abzuwenden.“
- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant ist:
  - *„Möchten Sie etwas über die Prozesse erfahren, die der KI-Hersteller etabliert hat, um zu gewährleisten, dass sich kein Mitarbeiter über die festgelegten (ethischen) Werte und Regeln hinwegsetzen kann?”*
  - *„Möchten Sie wissen, auf welchem Wege der KI-Hersteller sicherstellt, dass seine Mitarbeiter die vorgegebenen ethischen Werte umsetzen (können)?“*
  - *„Möchten Sie wissen, wie der KI-Hersteller die Anforderungen der Gesellschaft bezüglich Ethik konkret umsetzt?“*

Handling der Trainingsdaten nach ethischen Grundsätzen

Transparenzmerkmal: Ob und wie das Handling der Trainingsdaten gemäß ethischen Grundsätzen verläuft, zum Beispiel wie die Anforderungen bezüglich Fairness und Gerechtigkeit erfüllt werden.

- Korrespondierende Fragen, deren Beantwortung im Kontext der Vertrauenswürdigkeit relevant ist:
  - *„Wie wichtig ist es für Sie, dass der KI-Hersteller sich mit den ethischen Werten der Gesellschaft konstant auseinandersetzt."*
  - *„Möchten Sie etwas über die Prozesse erfahren, die der KI-Hersteller etabliert hat, um zu gewährleisten, dass sich kein Mitarbeiter über die festgelegten (ethischen) Werte und Regeln hinwegsetzen kann?"*

Im Rahmen der Reduktion der Ergebnisse hat sich gezeigt, dass – hinsichtlich der holistischen Transparenz – die Teilnehmer einen hohen Informationsbedarf bezüglich der Integrität von KI-Anbietern haben. Ein Erklärungsansatz für dieses starke Interesse – etwa nach der konstanten Auseinandersetzung mit ethischen Werten der Gesellschaft oder ob seitens des KI-Anbieters bestimmte Prozesse implementiert wurden, zum Beispiel um die Einhaltung ethischer Werte zu garantieren – könnte sein, dass allgemein die Teilnehmer im Rahmen des technologischen Fortschritts zunehmend auch die KI-Anbieter für den Zustand der Gesellschaft beziehungsweise das Wohlergehen der Menschen in der Verantwortung sehen.

Die Ergebnisse verschiedener Studien dokumentieren, dass unter anderem Wohlwollen und Fairness wichtige Komponenten zum Aufbau von Vertrauen in der Mensch-/KI-Interaktion sind. In diesem Kontext ist es beispielsweise essenziell, dass das Verhalten eines KI-Anbieters als Vertrauensnehmer konsistent, vorhersehbar und aufrichtig ist – dieser Anforderung der Anwender lässt sich nicht zuletzt durch etablierte sowie gut dokumentierte Prozesse nachkommen.

5.4 Spezielle Aspekte im Kontext der Definition von Vertrauen und Vertrauenswürdigkeit

Im Kontext der Definition von Vertrauen beziehungsweise Vertrauenswürdigkeit werden spezielle Aspekte bezüglich Kompetenz, Integrität und Wohlwollen noch einmal dediziert betrachtet.[16] Den Rahmen für diese Analyse bilden die Kriterien Unternehmensgröße, Alter der Teilnehmer, deren Kenntnisse bezüglich KI sowie in welchem Umfang KI im Einsatz ist. Eine selektive Behandlung der einzelnen Aspekte erscheint von daher sinnvoll, um unter anderem zu eruieren, ob der jeweilige Kenntnisstand oder der Einsatz von KI im Unternehmen einen Einfluss auf den Informationsbedarf hat, beziehungsweise den Blick für bestimmte Fragestellungen schärft. Nachfolgend werden einige prägnante Erkenntnisse der Analyse vorgestellt.

Kompetenz

Unter Kompetenz werden die Vertrauenswürdigkeits-Aspekte Zutrauen und Zuverlässigkeit subsumiert.

Allgemein wird ersichtlich, dass sich in Abhängigkeit von den Kenntnissen und der Erfahrung, die bei dem Einsatz von KI generiert wurde, sowohl die Präferenzen hinsichtlich des Informationsbedarfs differenzieren lassen als auch bestimmte Prioritäten.

Im Kontext des Vertrauenswürdigkeits-Aspekts Zuverlässigkeit ist für Teilnehmer, die umfangreiche Kenntnisse haben, wichtig zu erfahren, wie sichergestellt wird, dass bei Personalfluktuation das Wissen und die Kompetenz im Unternehmen erhalten bleibt. Dies lässt sich eventuell zurückführen auf die bereits gemachte Erfahrung, dass die Qualität einer KI-Lösung maßgeblich von der Qualifikation des jeweiligen Entwicklers abhängt.

Bezüglich dem Vertrauenswürdigkeits-Aspekt Zutrauen wollen Teilnehmer, die nur mit den Grundkonzepten der KI vertraut sind, mehr Belege – zum Beispiel in Form von formulierten Grundsätzen – dafür, dass der KI-Anbieter kooperativ handelt.

Bemerkenswert ist, dass Informationen etwa dahingehend ob sich der Firmensitz eines KI-Anbieters in der EU/Deutschland befindet oder detaillierte Auskünfte dazu, seit wann der Anbieter im Bereich KI tätig ist aber auch Fragestellungen im Hinblick auf die Diversifikation, für Großunternehmen eine klar höhere Relevanz haben.

In den selektierten Altersgruppen zeigen sich ebenfalls Unterschiede in den Präferenzen. Während nach Ansicht der Altersgruppe ab 50 – im vollkommenen Gegensatz zu den jüngeren Teilnehmern – die IT-Infrastruktur für KI einen entscheidenden Beitrag zum Wettbewerbserfolg leistet, zeigt sich im Weiteren, dass letztere hierfür der Forschung einen erheblich höheren Stellenwert beimessen.

[16] Zur Auswertungsmethodik: Im Zuge einer detaillierteren Betrachtung der jeweiligen Zielgruppen fand ein Vergleich der Untergruppen untereinander statt. Hierzu wurden die Teilnehmer der segmentierten Untergruppe, die mit „Top 2“ geantwortet haben, in Relation gesetzt zu der Anzahl derjenigen, die die Frage generell beantwortet haben. Diese Berechnung der relativen Häufigkeit wurde mit den anderen vorhandenen Untergruppen wiederholt. Liegt im direkten Vergleich eine Differenz von 10 Prozentpunkten oder mehr vor, kann davon ausgegangen werden, dass die jeweilige Frage für die entsprechende Gruppe von Bedeutung ist.

Integrität

Bei der Betrachtung des Vertrauenswürdigkeits-Aspekts Integrität wird offensichtlich, dass sich die Einstellung der Teilnehmer teilweise deutlich unterscheidet – hier bekunden jene mit umfangreichen Kenntnissen im Bereich KI sowie jene, die KI-Lösungen bereits im Einsatz haben deutlich mehr Interesse an Nachweisen bezüglich ethischem Verhalten.

Tendenziell ist es für Teilnehmer mit umfangreichen Kenntnissen weitaus essenzieller Kenntnis darüber zu erhalten, wie ein KI-Anbieter ethische Anforderungen umsetzt. Hierzu fordern sie detaillierte Informationen – etwa dahingehend, ob ein Ethik-Gremium vorhanden ist und Workshops mit Stakeholdern durchgeführt werden. Das lässt darauf schließen, dass ihnen die Implikationen beim Einsatz von KI-Lösungen durchaus bewusster sind als solchen, die nicht über eine entsprechende Expertise verfügen. Für diese These gibt es zwei weitere Indikatoren: Zum einen legen die Teilnehmer mit umfangreichen Kenntnissen Wert darauf, dass KI-Anbieter sich mit den ethischen Werten der Gesellschaft konstant auseinandersetzen – zum anderen wollen diese eher erfahren, wie gewährleistet wird, dass sich die Mitarbeiter nicht über definierte ethische Werte hinwegsetzen können und möchten dies auch gerne in einer Dokumentation festgehalten wissen. Zusätzlich ist es für Unternehmen, die bereits eine KI-Lösung im Einsatz haben, zudem wichtig, dass eine Dokumentation bezüglich der Einhaltung ethischer Grundsätze während des gesamten Entwicklungsprozesses zur Verfügung steht. Relativierend könnte hier allerdings zum Tragen kommen, dass diese Teilnehmer möglicherweise vorrangig Großunternehmen zuzurechnen sind und sich diese Fragestellungen in den jeweiligen Compliance-Vorgaben wiederfinden. Für die Vermutung spricht, dass diese vergleichsweise mehr Informationen über die Handlungsvorgaben im Code-of-Conduct des KI-Anbieters wissen wollen.

Im Gegensatz dazu ist es für Teilnehmer, die sich mit KI beschäftigen, aber noch keine KI-Lösung im Einsatz haben, vorrangig von Interesse, ob durch die Einhaltung ethischer Werte ein Konfliktpotential im Hinblick auf bestimmte Kundenwünsche entstehen könnte. Diese Annahme basiert möglicherweise auf mangelnder positiver Erfahrung – ebenso wie die Vermutung, dass es potenziell zu Konflikten zwischen der Durchsetzung ethischer Werte und ökonomischem Handeln kommt. Diese Fragestellung ist überwiegend für Unternehmen, die noch keine KI-Lösung einsetzen, relevant.

Die Auswertung der selektierten Altersgruppen zeigt, dass sich die älteren Teilnehmer umfassender mit ethischen Aspekten auseinandersetzen und dementsprechend auch mehr Wert auf die Einhaltung bestimmter Grundsätze legen, mittels derer sich eine ethische Handlungsweise sicherstellen lässt.

Wohlwollen

Unter Wohlwollen wird der Vertrauenswürdigkeits-Aspekt IT-Sicherheit subsumiert sowie die Umsetzung relevanter Anforderungen der Gesellschaft im Umgang mit Daten. Ebenso wie bei der Integrität zeigt sich auch im Kontext der Beurteilung des wohlwollenden Verhaltens von KI-Anbietern, dass Teilnehmer mit Expertise im Bereich KI oder jene, die bereits KI-Lösungen im Unternehmen einsetzen, sich diesbezüglich tiefergehender mit relevanten Fragestellungen beschäftigt haben.

Bei der Beurteilung des wohlwollenden Verhaltens stehen – neben der IT-Sicherheit – in erster Linie die Frage nach der Gewährleistung der Privatheit sowie der Umgang mit den Trainingsdaten und möglichen Implikationen, die daraus resultieren könnten, im Fokus.

Beachtenswert ist hier jedoch, dass es für Teilnehmer aus Großunternehmen – im Gegensatz zu jenen aus KMUs und Kleinstunternehmen – weniger relevant ist, Informationen zum verantwortungsvollen Umgang der Technologie unter dem Aspekt der Privatheit zu erhalten. Eine annehmbare Erklärung dafür könnte sein, dass sich diese per se mit solchen Themenstellungen, auch unter den rechtlichen Aspekten, bereits deutlich differenzierter auseinandergesetzt haben.

Bezüglich der Einhaltung ethischer Grundprinzipien haben vor allem die Unternehmen, die KI-Lösungen bereits einsetzen oder sich mit einem möglichen Einsatz beschäftigen sowie Teilnehmer, die über umfangreiche Kenntnisse verfügen, einen höheren Informationsbedarf den Umgang mit Trainingsdaten betreffend – letztere wollen zum Beispiel dezidiert wissen, ob und wie das Handling der Trainingsdaten gemäß ethischen Grundsätzen verläuft. Oder auch, in welchem Maße Daten abfließen können.

Allen Teilnehmern, die über umfangreiche Kenntnisse verfügen, ist es sehr wichtig oder wichtig, über das Potenzial unerwünschter Diskriminierung in Kenntnis gesetzt zu werden. Im Gegensatz dazu haben Teilnehmer, die nur mit den Grundkonzepten der KI vertraut sind, weniger Interesse daran, Informationen über die Erfüllung von Anforderungen bei Datenschutz und Urheberrecht zu erhalten – eventuell, weil sie keine Vorstellung davon haben, wie Datensätze entstehen und dass es hierbei möglich ist gesetzliche Vorgaben zu umgehen. Daraus lässt sich insgesamt die Schlussfolgerung ziehen, wie essenziell ein adäquater – auf den Bedarf der jeweiligen Zielgruppe abgestimmter – Erkenntnisaustausch ist.

Im Hinblick auf die IT-Sicherheit gab es ebenfalls ein bemerkenswertes Ergebnis: Für Unternehmen, die sich mit dem Einsatz von KI beschäftigen, ist ein Entscheidungskriterium, dass ein KI-Anbieter mehr in IT-Sicherheit investiert als der Wettbewerb.

Bei der Analyse der selektierten Altersgruppen kristallisieren sich bedeutsame Unterschiede heraus. Während den Teilnehmern der Altersgruppe über 50 wichtig ist Informationen darüber zu erhalten, wie beziehungsweise, dass sie die Kontrolle über die Abgabe der eigenen Daten behalten können sowie, dass die Privatheit im Umgang mit der Technologie sichergestellt wird, bekunden die jüngeren Teilnehmer kein gesteigertes Interesse daran. Eine mögliche Erklärung wäre an dieser Stelle rein spekulativ, von daher gilt es die Gründe hierfür zu erforschen. Des Weiteren ist noch hervorzuheben, dass die Teilnehmer in der Altersgruppe über 50 eher wissen wollen, inwieweit die Möglichkeit besteht Schäden, die bei der Nutzung auftreten können, zu verhindern.

5.5 Der Wert der Vertrauenswürdigkeit

Wie bereits ausgeführt wird es – aufgrund der immens gestiegenen Komplexität der Technologie – für Anwender zunehmend diffizil, KI-Lösungen zu verstehen und hinsichtlich der für sie relevanten Kriterien bewerten zu können. Aufgrund dieser Erwägung lässt sich bereits prognostizieren, wie wertvoll das Instituieren einer konstanten Vertrauensbasis für KI-Anbieter ist: eine nachhaltige unternehmerische Wertschöpfung, die Raison d'être von Unternehmen, kann ohne Vertrauen nicht ermöglicht werden. Insofern lässt sich auch plausibilisieren, dass KI-Anbieter dem Erhalt ihrer Vertrauenswürdigkeit oberste Priorität beimessen müssen.[17]

Im Rahmen der Anwender-Studie TrustKI sollte auch nachgewiesen werden, dass durch die – aus Sicht des Anwenders – hinreichende Begründung hinsichtlich seiner relevanten Aspekte, er den Wert der Vertrauenswürdigkeit des KI-Anbieters entsprechend honorieren würde, weil diese für ihn einen Mehrwert darstellt. Zur Erbringung dieses Nachweises galt es im Weiteren, exakt die entsprechenden Fragen zu selektieren.[18] Basierend auf dieser Annahme wurden alle Fragestellungen der Anwender-Studie dahingehend überprüft, ob sie mit der Frage „Mehrausgabe" korrelieren. Nach Überprüfung der Ergebnisse anhand ihres P-Wertes[19] bleiben 46 Fragen übrig, die gemäß ihrer Logik geprüft werden.

Insgesamt ließen sich so 37 Fragen ermitteln, die mit der Frage nach Mehrausgaben korrelieren.[20]

Nachfolgend die Anzahl der korrelierenden Fragen bezüglich der einzelnen Vertrauenswürdigkeits-Aspekte: Zutrauen (5), Integrität (22), IT-Sicherheit (3), KI-Lösung (6).

Zusammengefasst lassen sich die Aspekte wie folgt abbilden:

Zutrauen

Hinsichtlich des Vertrauenswürdigkeits-Aspekt Zutrauen sind zwei Kriterien von hoher Relevanz: Im Kontext der Belegschaft erachten die Teilnehmer Diversität und Erfahrung als wichtig. Des Weiteren legen sie Wert auf die transparente Darstellung des Unternehmens.

Der Mehrwert lässt sich aus Sicht der Teilnehmer dadurch rechtfertigen, dass KI-Anbieter respektive die Verteilung der Unternehmensanteile offen darlegen und dass sich der Hauptsitz in der EU oder Deutschland befindet. Eine diverse Belegschaft mit entsprechender Erfahrung im Anwendungsbereich oder der Branche

[17] gemäß Suchanek, A. „Vertrauen als Grundlage nachhaltiger unternehmerischer Wertschöpfung"

[18] An dieser Stelle sei zu vermerken, dass Mehrausgaben nicht als explizite Ordnung gelten sollen, sondern diese repräsentativ im Sinne ihrer inhärenten Bedeutung gewertet werden. Personen, die bereit sind mehr für eine KI-Lösung zu zahlen, zeigen einen höheren Bedarf an bestimmten Informationen. Sollte dieser adäquat erfüllt werden, wäre hierfür konkludent eine höhere Bezahlung aus ihrer Sicht gerechtfertigt.

[19] P-Wert < 0,05

[20] Eine positive Korrelation bedeutet, dass Personen, die bereit sind Mehrausgaben zu tätigen, höher in den entsprechenden Fragen antworten und diese als bspw. wichtiger erachten. Bei Zustimmungsfragen bedeutet eine positive Korrelation Zustimmung, während eine negative Korrelation als Ablehnung zu interpretieren ist. Die stärkste Korrelation beträgt 0,23.

belegt zusätzlich die Kompetenz des KI-Anbieter. Aufgrund dessen wird ebenfalls Wert darauf gelegt, dass auch Fort- und Weiterbildungsmöglichkeiten angeboten werden.

Integrität

Mit Blick auf den Vertrauenswürdigkeits-Aspekt Integrität resultiert der Mehrwert für die Teilnehmer aus drei Kategorien: Das unternehmerische Handeln, dediziert Ethik-Anforderungen sowie die Prozesse zur Sicherstellung des ethischen Handelns.

Im Sinne des unternehmerischen Handelns sind Dimensionen wie Mitarbeiterführung und -motivation, sowie Auskünfte, was zur Vermeidung von Schäden getan wird, substanziell. Die Kategorie der dedizierten „Ethikanforderungen" umfasst unter anderem Anforderungen hinsichtlich der konstanten Auseinandersetzung mit den Werten der Gesellschaft sowie die Maßnahmen zur Umsetzung ebendieser.

Bezüglich der "Sicherstellung von Ethik" ist es erforderlich Informationen offenzulegen, wie ethische Werte umgesetzt werden und was KI-Anbieter tun, damit diese nicht ausgehebelt werden können.

IT-Sicherheit

Für eine Wertsteigerung im Sinne des Vertrauenswürdigkeits-Aspekts IT-Sicherheit sind zwei Kriterien maßgeblich: Zum einen ist es notwendig, dass ein gewisser Grad an IT-Grundschutz vorhanden ist, um die Sicherheit der Prozesse kontinuierlich zu verbessern. Zum anderen ist es für die Teilnehmer mit Blick auf die IT-Infrastruktur substanziell, dass diese durch entsprechende Maßnahmen wirksam geschützt wird.

Ein weiterer Mehrwert ergibt sich nach Ansicht der Teilnehmer daraus, dass ein KI-Anbieter mehr in IT-Sicherheit investiert als seine Wettbewerber.

KI-Lösung

Bezüglich der Vertrauenswürdigkeits-Aspekte der KI-Lösung ist allgemein erkennbar, dass sich aus der transparenten Darstellung der KI-Lösung definitiv Mehrwerte generieren lassen: Hierunter fallen nach Maßgabe der Teilnehmer unter anderem Informationen bezüglich potenzieller Probleme, wie etwa die Möglichkeit unerwünschter Diskriminierung oder ob eine Aushebelung der AGB stattfinden könnte.

Insgesamt ist es fundamental, dass der Verwendungszweck detailliert dargelegt wird und der KI-Anbieter generelle Informationen über die Trainingsdaten zur Verfügung stellt. Wichtig für eine Wertsteigerung ist zudem, dass sich keine ethischen Probleme durch das Inverkehrbringen der KI-Lösung ergeben.

5.6 Die Bedeutung der IT-Sicherheit im Kontext der Vertrauenswürdigkeit

Die detaillierten Ergebnisse in der Gesamtauswertung zeigen, dass bei dem Vertrauenswürdigkeits-Aspekt IT-Sicherheit durchweg eine hohe Zustimmung seitens der Teilnehmer zu verzeichnen ist. Die Top 2-Werte liegen hier häufig – über alle segmentierten Gruppen hinweg – bei über 90 Prozent. Aufgrund dessen wird die hohe Relevanz, die die Teilnehmer dem Vertrauenswürdigkeits-Aspekt IT-Sicherheit zuschreiben, an dieser Stelle noch einmal gesondert analysiert:

Grundsätzlich lässt sich feststellen, dass Sicherheit ein Megatrend ist, dem paradoxe Entwicklungsdynamiken innewohnen: Das Empfinden für Risiken und Gefahren nimmt zu, obwohl die Menschen aktuell de facto in der sichersten aller Zeiten leben. Doch eben diese Sicherheit führt dazu, dass Menschen die Gefühle von Unsicherheit intensiver wahrnehmen. Dies wird ein Stück weit dadurch erklärbar, dass unsere Gesellschaft sich in einem Daueralarm befindet – zudem haben Krisen, wie etwa die Corona-Pandemie, plötzlich transparent gemacht, dass das Leben fragil ist und sich Umstände von einem auf den anderen Tag ändern können.

Hinzu kommt die gestiegene Komplexität sowie Dynamik aufgrund der Digitalisierung. Dies bedeutet zum einen Änderungen der gewohnten Lebensumstände für jeden Einzelnen. Aber zum anderen auch – rein faktisch – eine höhere Verwundbarkeit, da sich mit jedem vernetzten Gerät die Angriffsfläche vergrößert, nicht nur im Unternehmen, sondern auch zuhause.

Somit wird im digitalisierten und globalisierten 21. Jahrhundert grundsätzlich die Frage danach, was Sicherheit bedeutet und wer sie verantwortet, neu verhandelt – des Weiteren rückt zudem das Thema Resilienz zunehmend in den Fokus.

Dass folglich insbesondere IT-/beziehungsweise Cyber-Sicherheit, inklusive der damit verbundenen Herausforderungen, in den Medien große Aufmerksamkeit zuteil wird, kann als weiterer Erklärungsansatz für die Ergebnisse gewertet werden.

Aufgrund der Vorgehensweise im Kapitel „Spezielle Aspekte im Kontext der Definition von Vertrauen und Vertrauenswürdigkeit – Wohlwollen" (siehe Fußnote 18) lässt sich erklären, dass der Vertrauenswürdigkeits-Aspekt IT-Sicherheit in den Auswertungen als relevanter Parameter keine Beachtung findet – in den untersuchten Untergruppen sind keinerlei Unterschiede aufgetreten, weil IT-Sicherheit durchgängig eine hohe Bedeutung beigemessen wird.

5.7 KI-Lösung – Auswertung relevanter Ergebnisse

Die Vertrauenswürdigkeit von KI-Lösungen kann unter zwei Aspekten bewertet werden: Zum einen ist es möglich eine Evaluierung bezüglich der eigentlichen Leistung, für die das spezielle System eingesetzt wird, vorzunehmen – im Sinne von Explainable AI (xAI). Die Realisierung solcher KI-Prüfungen findet anhand definierter Kriterien statt. Generell gibt es zwei prominente Ansätze, um diese vorzunehmen: Prozess- und Produktprüfungen. Erstere adressieren die Prozesse, nach denen KI-Anwendungen entwickelt sowie betrieben werden und beruhen auf der Annahme, dass gute Prozesse auch zu guten KI-Anwendungen führen. Letztere zielt darauf ab, die Funktionalitäten einer konkreten KI-Anwendung zu validieren beziehungsweise über eine strukturierte Risikoanalyse nachzuweisen, dass wesentliche Risiken im Hinblick auf die Vertrauenswürdigkeit hinreichend gut gemindert sind.[21]

Zum anderen kann der Bewertungsmaßstab im Hinblick auf Vertrauenswürdigkeit darauf basieren, welches Versprechen der KI-Anbieter bezüglich Entwicklung sowie Bereitstellung seiner KI-Lösung abgibt und ob dieses kongruent ist mit Bezugnahme auf die allgemeine Erwartungshaltung der Gesellschaft. Hier sind Art und Umfang der Informationen, die dem Anwender zur Verfügung gestellt werden, dahingehend zu beurteilen, ob diese ihn in die Lage versetzen, die Vertrauenswürdigkeit des KI-Anbieters und dessen Lösung einzuschätzen.

Im Rahmen der Anwender-Studie TrustKI lag bezüglich der KI-Lösung das Hauptaugenmerk darauf zu eruieren, welche Informationen die Anwender hinsichtlich der einzelnen Vertrauenswürdigkeits-Aspekte für notwendig erachten, um die jeweilige KI-Lösung als vertrauenswürdig einzustufen.

Insgesamt dokumentieren die Ergebnisse der Studie, dass insbesondere die folgenden Angaben seitens der KI-Anbieter zur Verfügung gestellt werden sollten, um den Informationsbedarf der Teilnehmer zu decken.[22]

[21] „Wie kann die Vertrauenswürdigkeit und Transparenz von KI-Systemen gewährleistet werden?“, Dr. Poretschkin, M., Blogbeitrag, https://lamarr-institute.org/de/blog/ki-vertrauenswuerdigkeit/ (Abruf: 02. Januar 2024)

[22] Zur Auswertungsmethodik:
Da bei 50 Prozent ein ausgewogenes Verhältnis vorliegt, kann nicht von Relevanz gesprochen werden, daher liegt die Schwelle bei 60 Prozent für Relevanz.

- Bei Rating-/ Rangfragen werden die Befragten und deren Häufigkeit betrachtet, die den ersten Rang (Top 1) oder den ersten und zweiten Rang (Top 2) gewählt haben.
    - Top 1 erhält Relevanz, wenn 50 Prozent der Befragten diesen Punkt auf Rang 1 gewählt haben.
    - Top 2 erhält Relevanz, wenn 80 Prozent der Befragten diesen Punkt auf Rang 1 oder 2 gewählt haben.

- Bei Skalenfragen werden die Befragten und deren Häufigkeit betrachtet, die “sehr wichtig” (Top 1) oder “sehr wichtig” und “wichtig” (Top2) gewählt haben.
    - Top1 erhält Relevanz, wenn 50 Prozent der Befragten diesen Punkt als “sehr wichtig” gewählt haben.
    - Top 2 erhält Relevanz, wenn 80 Prozent der Befragten diesen Punkt als “sehr wichtig” oder “wichtig” gewählt haben.

Vertrauenswürdigkeits-Aspekte

Zweckprägnanz

Die Zweckprägnanz manifestiert sich im Verwendungszweck der KI-Lösung insgesamt. Dies bedeutet, dass bei der Entwicklung von Funktionen die Intention der KI-Lösung zielgenau definiert sein muss.

Frage: Welche Informationen zum Verwendungszweck einer KI-Lösung sind wichtig, damit Sie dieser vertrauen können?

- Für 59,1 Prozent (Top1) der Teilnehmer ist es sehr wichtig, dass der KI-Anbieter den Verwendungszweck der KI-Lösung detailliert darstellt.
- Angaben darüber, ob die – bei der Nutzung der KI-Lösung – generierten Daten für eigene Zwecke wie Werbung genutzt werden, halten 58,2 Prozent (Top1) für relevant.
- Bezüglich der Aufklärung dahingehend, ob die Daten der KI-Lösung an Dritte, zum Beispiel Werbeagenturen, weitergegeben werden, sind 65,9 Prozent (Top1) der Meinung, dass diese Information sehr wichtig ist.
- Eine Offenlegung darüber, unter welchen Umständen Daten an Regierungsbehörden weitergegeben werden, verlangen 67,2 Prozent (Top1).

Leistungsfähigkeit

Die Leistungsfähigkeit der KI-Lösung ist das, was der Anwender unmittelbar erfassen und auch kontrollieren kann.

Frage: Welche informationstechnischen Sicherheitsrisiken würden Sie am ehesten davon abhalten, eine KI-Lösung einzusetzen?

- Kenntnisse über einen möglichen Abfluss/Ausspähen von kritischen Daten erachten 86,4 Prozent der Teilnehmer als wichtig und sehr wichtig (Top 2).
- 87,9 Prozent (Top 2) der Teilnehmer sehen die Möglichkeit von Manipulationen der Ergebnisse einer KI-Lösung als ein großes Sicherheitsrisiko an.

Transparenz

Im Rahmen der Transparenz-Erklärung geben die KI-Anbieter Auskunft über alle relevanten Fakten, die erforderlich sind, um im gegebenen Kontext eine valide Entscheidung über die Vertrauenswürdigkeit der KI-Lösung treffen zu können.

Frage: Wie wichtig ist es für Sie, dass der KI-Anbieter Nachweise unabhängiger Dritter (zum Beispiel TÜV) über seine KI-Lösung zu folgenden Themen zur Verfügung stellt?

- Nachweise unabhängiger Dritter (zum Beispiel TÜV) über die KI-Lösung zu dem Thema "technische Robustheit und Sicherheit der KI-Lösung" halten 60,0 Prozent (Top 1) für sehr wichtig.

- Nachweise unabhängiger Dritter (zum Beispiel TÜV) über die Präzision und Zuverlässigkeit der Verarbeitungslogik der KI ist für 94,6 Prozent (Top 2) wichtig und sehr wichtig sowie für 60,9 Prozent (Top 1) sehr wichtig.
- Nachweise unabhängiger Dritter (zum Beispiel TÜV) über die KI-Lösung zu dem Thema "Qualität der Trainingsdaten" halten 92,2 Prozent für wichtig und sehr wichtig (Top 2) sowie 53,7 Prozent für sehr wichtig (Top 1).

Frage: Wie wichtig ist es Ihnen, dass Sie im Vorfeld über die folgenden Risiken durch den Einsatz der KI-Lösung informiert werden?

- Kenntnisse über einen möglichen Abfluss/Ausspähen von kritischen Daten erachten 95,7 Prozent als wichtig und sehr wichtig (Top 2) sowie 68,6 Prozent als sehr wichtig (Top 1).
- Aufklärung über Risiken durch den Einsatz der KI-Lösung in Bezug auf das "Potenzial der unerwünschten Diskriminierung" halten 87,0 Prozent für wichtig oder sehr wichtig (Top 2).
- Aufklärung über Risiken durch den Einsatz der KI-Lösung in Bezug auf den "Grad der Unzuverlässigkeit der Ergebnisse" wollen 58,5 Prozent erhalten (Top 1).

Für alle Befragten sind Informationen über die genannten Risiken im Vorfeld von Relevanz: Abfluss/Ausspähen von kritischen Daten / Verletzung der physischen, psychischen, finanziellen Unversehrtheit / Grad der Unzuverlässigkeit der Ergebnisse / Möglichkeiten der Fehlinterpretation der Ergebnisse / Potential der unerwünschten Diskriminierung relevant. Besonders erwähnenswert in diesem Kontext ist, dass die Ergebnisse in Top 1 bei allen Antworten außer bei „Potential der unerwünschten Diskriminierung“ über 50 Prozent liegen.

Weitere Kriterien

Funktionsweise

Frage: Möchten Sie im Einzelfall wissen, wie das Ergebnis zustande gekommen ist?

- 61,1 Prozent beantworteten diese Frage mit "Ja". Erwähnenswert an dieser Stelle ist, dass diesbezüglich der Informationsbedarf bei Unternehmen, die noch keine KI im Einsatz haben, mit 74,4 Prozent weitaus höher liegt.

Frage: Über welche Aspekte der Trainingsdaten möchten Sie informiert werden?

- Bezüglich der Erfüllung regulatorischer Anforderungen bei Urheberrecht und Datenschutz wollen 63,5 Prozent (Zustimmung) Informationen erhalten

Ethik

Frage: Können sich durch das Inverkehrbringen und den Einsatz einer KI-Lösung ethische Probleme ergeben?

- 72,6 Prozent der Befragten haben diese Frage mit „Ja“ beantwortet – das heißt, sie sind der Meinung, dass sich aus dem Inverkehrbringen und dem Einsatz einer KI-Lösung ethische Probleme ergeben können.

Frage: Ich möchte wissen, ob im Rahmen des Einsatzes einer KI-Lösung die Privatheit zum Beispiel durch AGB ausgehebelt werden kann, wodurch der KI-Hersteller das Recht zum Verkauf von Daten an Dritte erhalten könnte.

- Wie die Privatheit ausgehebelt werden kann, möchten 79,8 Prozent (Zustimmung) der Befragten wissen. Interessant hierbei ist unter anderem, ob es möglich ist, die Privatheit zum Beispiel mittels AGB zu umgehen und dadurch der KI-Anbieter zum Beispiel das Recht zum Verkauf von Daten an Dritte erhalten könnte.

Interpretation prägnanter Ergebnisse

Die Häufung von, als maßgeblich erachteten, Fragestellungen zum Vertrauenswürdigkeits-Aspekt Transparenz zeigen, dass diesem seitens der Anwender ein hoher Stellenwert im Rahmen des Vertrauensaufbaus eingeräumt wird. Hieraus lässt sich ableiten, dass ein Teil der Vertrauenswürdigkeit einer KI-Lösung erst über die Vertrauenswürdigkeit des Herstellerunternehmens zu entstehen scheint, da die jeweiligen AGB seitens des KI-Anbieters erstellt werden und dieser für die entsprechende Ausgestaltung in der KI-Lösung verantwortlich ist.

Trainingsdaten haben eine große Bedeutung für die Anwender. So wollen generell 89,0 Prozent (Zustimmung) der Befragten mehr über Trainingsdaten erfahren und 70,0 Prozent (Zustimmung) speziell bezüglich deren Qualität. Insgesamt 53,7 Prozent (Top 1) erachten es sogar als sehr wichtig, dass eine unabhängige (akkreditierte) dritte Partei diese nachweist. Allerdings sind die Details nicht von so hoher Relevanz: bei der direkten Nachfrage zu einer punktuellen Auskunft, die von den KI-Anbietern bezüglich der Trainingsdaten gefordert werden könnte, lagen die Ergebnisse nicht so hoch: lediglich 42,8 Prozent (Zustimmung) wollen mehr über die Herkunft der Trainingsdaten (Länder, Branchen) erfahren und 37,3 Prozent (Zustimmung) über die Personen oder Organisationen, die den Trainingsdatensatz erstellt haben.

Bezüglich der Fragestellung, bei welchen Anwendungsbereichen der KI die Anwender die größten ethischen Herausforderungen sehen, zeigen sich klare Tendenzen. Bei Bereichen, wo durch den Einsatz von KI für den Einzelnen potenziell ein Mehrwert entstehen kann und hierdurch ein nicht zu hohes persönliches Risiko eingegangen wird – wie etwa zur Unterstützung von Krebsdiagnosen – zeigt sich eine hohe Akzeptanz. Als moralisch nicht vertretbar werden hingegen Anwendungsbereiche gesehen, die aufgrund von KI pauschalierende Ergebnisse liefern mit wenig Eingriffs- beziehungsweise Korrekturmöglichkeiten durch den Menschen, wie etwa die Vorhersage von Straftaten.

Im Hinblick auf die Autonomiegrade zeigt sich, dass Repräsentanten von Anwenderunternehmen, die nur über KI-Grundkenntnisse verfügen sowie keinen Einsatz planen oder selbst keine KI anbieten, die Autonomiegrade oft nicht kennen. Über alle Branchen hinweg besteht jedoch der Bedarf, darüber informiert zu werden, was bei den Autonomiegraden einer KI-Lösung zu beachten ist. So erachten 95,3 Prozent (Top 2) der Befragten es als sehr wichtig oder wichtig, dass der KI-Anbieter über die verschiedenen Autonomiegrade (58,1 Prozent – Top 1) informiert, um seine Vertrauenswürdigkeit unter Beweis zu stellen.

6. Ausblick

Im Kontext der *Anwender-Studie TrustKI* wurden substanzielle Ergebnisse ermittelt, die Aufschluss darüber geben, was Führungskräfte in Anwenderunternehmen als notwendig erachten, um KI-Anbietern und deren KI-Lösung vertrauen zu können.

Im weiteren werden wir mit den Erkenntnissen aus der *Anwender-Studie TrustKI* in den Dialog mit den KI-Anbietern treten, um mit diesen im Rahmen von moderierten Workshops die Basis für ein gemeinsames Ökosystem (Vertrauenswürdigkeits-Plattform) zu konzipieren.

Wir laden alle KI-Anbieter ebenso wie Führungskräfte aus Anwenderunternehmen ein, daran aktiv mitzuwirken und gemeinsam ein Ökosystem für vertrauenswürdige KI-Lösungen zu etablieren.

Anhang

Die Anwender-Studie *TrustKI* kann unter folgendem Link abgerufen werden:

www.vertrauenswürdigkeit.com/studienergebnisse-anwender-studie-trustki/