Vertrauen als ein Verhältnis zwischen Menschen und Maschinen?

Zur Akzeptanz von IT-Systemen durch Institutionenvertrauen und Vertrauensersatz

Georg Rainer Hofmann

information management institut

mainproject

EUROPÄISCHE UNION
EUROPÄISCHER SOZIALFONDS

ESF IN BAYERN
WIR INVESTIEREN IN MENSCHEN

TH Aschaffenburg
university of applied sciences

Autor
Prof. Dr. Georg Rainer Hofmann
Technische Hochschule Aschaffenburg, Information Management Institut (IMI)
E-Mail: hofmann@th-ab.de
Tel.: +49 6021 4206-700

Herausgegeber
Information Management Institut (IMI), Technische Hochschule Aschaffenburg

Lektorat und Layout
Meike Schumacher
Technische Hochschule Aschaffenburg, Information Management Institut (IMI)

Titelbild
Bundesregierung / Steffen Kugler

Die Deutsche Bibliothek - CIP Einheitsaufnahme
Vertrauen als ein Verhältnis zwischen Menschen und Maschinen?
Zur Akzeptanz von IT-Systemen durch Institutionenvertrauen und Vertrauensersatz

Aschaffenburg, April 2023

ISBN 978-3-9823413-7-8

TECHNISCHE HOCHSCHULE ASCHAFFENBURG
INFORMATION MANAGEMENT INSTITUT (IMI)
Würzburger Straße 45
D-63743 Aschaffenburg
https://www.th-ab.de/
https://www.imi.bayern/

In dieser Ausarbeitung wird für einige Berufs- und Personenbezeichnungen sowie personenbezogene Hauptwörter das Generische Maskulinum verwendet, anstelle der gleichzeitigen Verwendung der Sprachformen männlich, weiblich und divers. Dies erfolgt aus Gründen der besseren Lesbarkeit. Die Bezeichnungen gelten grundsätzlich für alle Geschlechter. Die verkürzte Sprachform ist redaktioneller Natur und beinhaltet keinerlei Wertung. Vergleiche Urteil des BGH vom März 2018 (VI ZR 143/17): Die formale Verwendung des Generischen Maskulinums verstößt nicht gegen das AGG.

Sollten wir einer Maschine vertrauen?
Braucht sie das? Vertraut sie uns dann auch?
– Vertrauen sich Maschinen untereinander?

Sicherheit allein schafft kein Vertrauen.
Aber allein Vertrauen schafft Sicherheit.

Motivation

Der technische Fortschritt von IT-Systemen, wie zum Beispiel der sogenannten „Künstlichen Intelligenz" (KI) oder den „Common Data Spaces" (CDSs), bringt die Frage nach der Akzeptanz dieser Systeme mit sich. Sowohl die aktive (ökonomische) Akzeptanz in Form des Erwerbs und der Nutzung, als auch die passive (soziale) Akzeptanz in Form der Duldung der aktiven Nutzung durch Dritte, sind für den Erfolg von IT-Systemen ausschlaggebend [ScHo16]. Eine geringe Akzeptanz der Systeme und ein mangelndes Vertrauen in dieselben wird in einen Zusammenhang gebracht: Man fragt nach der Gestaltung „zuverlässiger und vertrauenswürdiger Systeme", da diese besonders akzeptiert werden [CoPo20].

Es gibt kein einheitliches Verständnis des Begriffs „Vertrauen" – dazu ist das Phänomen zu komplex, und die akademischen Fachrichtungen, die es behandeln, sind zu unterschiedlich [RoWo15]. Vertrauen ist zwar im sozialen Kontext und der Lebenswirklichkeit allgegenwärtig – und doch ist es analytisch nicht so präzise erfasst, wie dies wünschenswert wäre.

In diesem Beitrag werden „Vertrauen" und dessen Zusammenhang mit „Zuverlässigkeit" und „Verstehen" im Kontext von IT-Systemen erörtert. Unterschieden werden ein „aktives Vertrauen" (des Vertrauensgebers) und ein „passives Vertrauen" (des Vertrauensnehmers). Es zeigt sich, dass die Begriffe „zuverlässig" und „vertrauenswürdig" keinesfalls das Gleiche sind, gar als Pleonasmus bezeichnet werden können. Hinzu kommt, dass „Vertrauenswürdigkeit" nur eine Eigenschaft(!) ist, aus der ein Vertrauen als eine Beziehung(!) nicht unbedingt folgen muss – ebenso wenig wie alle liebenswürdigen Personen geliebt und alle preiswürdigen Werke prämiert werden. Ein quasi „persönliches" Vertrauensverhältnis eines Menschen zu einem IT-System ist lediglich eine anthropomorphe Projektion und stellt keine Grundlage vertretbarer Akzeptanz dar. Ebenso absurd ist die umgekehrte Annahme, dass ein IT-System seinen Nutzern tatsächlich „persönlich" vertrauen könne.

Brauchbar sind hingegen zum einen Analogieschlüsse aus dem bekannten Markenvertrauen – als eine Form des Institutionenvertrauens – auf die (holistischen) Anwendungsszenarien von IT-Systemen. Nützlich ist zum anderen die analoge Übertragung der bekannten „Vertrauensbildenden Maßnahmen – Confidence-Building Measures (CBMs)" aus dem politischen Kontext. Es werden in der Folge Elemente zur Gestaltung von Markenvertrauen und CBMs für IT-Anwendungen identifiziert.

Auf dieser Basis lassen sich gestaltungsorientierte Konzepte „Vertrauen in Gesamt-IT-Anwendungsszenarien" und damit der Akzeptanzverbesserung von IT-Systemen entwickeln. Integrierte vertrauenswürdige Szenarien von IT-Anwendungen mit ihrem sozio-ökonomischen Kontext werden so greifbar und darstellbar.

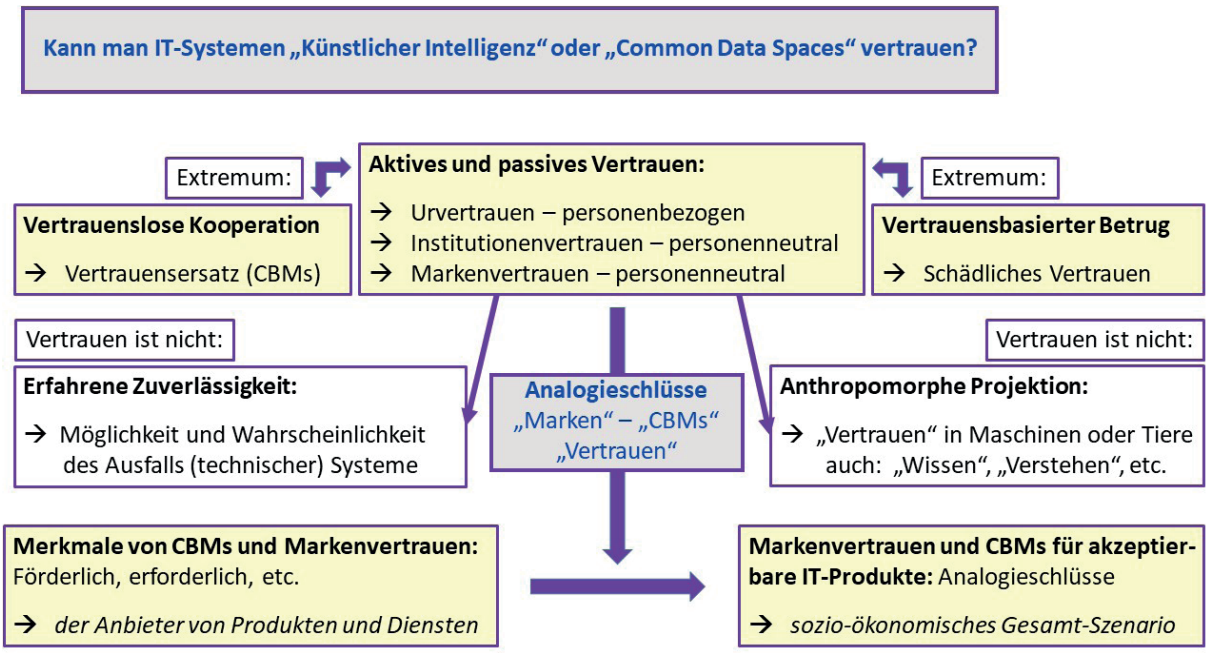


Diagramm zum Gang der Untersuchung. Ausgehend von der Frage nach dem möglichen Vertrauen in IT-Systeme, wie der KI und den CDSs, werden die Basisbegriffe des aktiven und passiven Vertrauens erörtert. Die vertrauenslose Kooperation, die Vertrauensersatz durch CBMs erfordert, und der vertrauensbasierte Betrug sind Extremfälle von Vertrauensverhältnissen. Von Vertrauen abzugrenzen sind Zuverlässigkeit und anthropomorphe Projektionen. Es wird gezeigt, dass sich für IT-Systeme ein Markenvertrauen aufbauen lässt, beziehungsweise CBMs konstruieren lassen, indem bekannte Faktoren per Analogieschluss übernommen werden.

Inhalt

Menschliches Vertrauen und Verstehen – technische Zuverlässigkeit

Archetypus Vertrauen – aktives und passives Vertrauen

Die Eingangsfrage soll sein, in welchem Alter und welchen Situationen ein Mensch die Fähigkeit erwirbt, anderen Menschen „persönlich" zu vertrauen. Seit etwa den 1960er Jahren geht man in der Anthropologie davon aus, dass es ein evolutionär-biologisch entwickeltes und angeborenes „Urvertrauen" gibt. Der Soziologe und Anthropologe Dieter Claessens (1921 – 1997) hat das „Urvertrauen" in die erste Lebensphase des Menschen verortet [Clae93]. Für einen Säugling ist das Vertrauen zu seinen ersten Kontaktpersonen (in der Regel sind das die Eltern) nicht-rational und alternativlos – aufgrund seines noch nicht entwickelten rationalen Urteilsvermögens und seiner eigenen Hilfslosigkeit. Dabei kann ein „aktives Vertrauen" (der Säugling vertraut als Vertrauensgeber – und ist dahingehend aktiv) von einem „passiven Vertrauen" (den Eltern als Vertrauensnehmern wird vertraut – sie sind mithin passiv) unterschieden werden.

Das aktive „Vertrauen-Können" ist ein Archetypus, der dem Menschen angeboren ist; es muss nicht erst rational erlernt oder begründet werden, um es ausüben zu können. Vertrauen muss daher nicht quasi „erarbeitet" werden. Aktives Vertrauen ist in Situationen von Hilfslosigkeit und Unsicherheit unabdingbar. Aktives Vertrauen schafft subjektive Sicherheit in (unsicheren) Lebenssituationen.

Mit dem aktiven Vertrauen des Säuglings korrespondiert – in einer reflexiven Symmetrie – das passive Vertrauen der ersten Kontaktpersonen. Diese Vertrauensnehmer sind in der Regel die Eltern, es können aber auch andere Personen sein, die sich der Pflege des Säuglings annehmen. Zur Anbahnung des Verhältnisses zwischen aktivem vertrauendem Säugling und passiven Pflege-Vertrauenspersonen hat der Mensch – ebenfalls angeborene – Vertrauen-Anbahnungssignale, insbesondere das Duchenne-Lächeln. Es wird interkulturell, universell und personen-neutral verstanden – nicht nur vom eigenen soziokulturellen Kontext oder der Verwandtschaft.

Ein ziemlich junger Mensch lächelt („Duchenne-Lächeln"), als ein Vertrauen-Anbahnungssignal. Eine Zurückweisung des Signals oder eine Enttäuschung des angebahnten Vertrauens löst beim Adressaten, dem Vertrauensnehmer, in aller Regel ein Störgefühl aus [Bild: jcphoto]

Das nach dem Physiologen Guillaume-Benjamin Armand Duchenne de Boulogne (1806 – 1875) benannte echte oder Duchenne-Lächeln ist dem Menschen angeboren. Das echte

Lächeln ist eine Aktion des gesamten Gesichts und ein Ausdruck des guten Willens und Wohlbefindens; es dient insbesondere der Aufnahme von Kommunikation und als Vertrauenssignal. Die natürlichen Entgegnungen sind dann haptische Signale, wie das In-den-Arm-Nehmen durch die passiv Vertrauenden; sie manifestieren das Urvertrauen und ein „einander Verstehen“ [Duch23].

Das echte Lächeln ist zu unterscheiden von quasi „falschen“ Arten des Lächelns, wie dem aufgesetzt-künstlichen, dem bösartig-zynischen, maliziösen oder höhnischen Lächeln, auch vom müden Lächeln oder vom Verlegenheits- und Ängstlichkeitslächeln. Das sichere Erkennen des Duchenne-Lächelns ist dem Menschen ebenfalls angeboren.

Das echte Lächeln ist ein angeborenes universelles Vertrauen-Anbahnungssignal. Es ist auch ein Schutz vor Vertrauensmissbrauch, weil die passiven Vertrauenspersonen ein – wohl ebenfalls angeborenes – Störgefühl (vulgo „schlechtes Gewissen“) entwickeln, wenn sie die Vertrauen-Anbahnung zurückweisen oder das angebahnte Vertrauen bösartig missbrauchen. Dieses Störgefühl bei den passiven Vertrauensnehmern nach einem Vertrauensbruch ist quasi die „soziale Versicherung“ für die aktiven Vertrauensgeber. Im Gegenzug wird der Vertrauensgeber nach einem Vertrauensbruch das Vertrauensverhältnis sehr schnell beenden. Auch dies kann als eine „soziale Versicherung“ verstanden werden, um Schaden vom Vertrauensgeber abzuwenden.

So wie ein Vogelküken per natürlicher Fertigkeit die Eierschale zur Umwelt durchbricht und den Pflegetrieb der Vogeleltern auslöst, so gelangt ein neugeborener Mensch per echtem Lächeln in seine neue soziale Umwelt, der er vertraut. Die Bindungsforschung (Attachment Research) des britischen Psychoanalytikers John Bowlby (1907 – 1990) konnte belegen, dass Menschen ein angeborenes Bedürfnis haben, enge und von intensiven Gefühlen geprägte Beziehungen zu Mitmenschen aufzubauen [Bowl01].

Damit identifiziert sich das psychologische Phänomen „Vertrauen“ als allgegenwärtig.

Bei [RoWo15] wird in zutreffender Weise ausgeführt, dass aktives Vertrauensgeben als soziale Praxis ohne Vorleistung oder Erfahrung von Zuverlässigkeit möglich ist. Vertrauen kann von daher ein präreflexives Konstrukt auf hermeneutischer Basis sein:

Das ist direkt evident, wenn beispielsweise ein Mensch in einer ihm fremden Stadt nach dem Weg zum Bahnhof fragen muss. Das bedingt, dass dieser Mensch präreflexiv festlegt, welchen Passanten er nach dem richtigen Weg fragt – wem er auf der Basis von welcher Hermeneutik vertrauen will. Vertrauen erscheint hier als eine Erfahrungs-unabhängige Vorleistung, allerdings auf der Basis einer sozialen Einschätzung, welcher Person eine kompetente und gutwillige Antwort zuzutrauen ist: Wer könnte den Weg kennen und ist auch willens, diesen mitzuteilen? Die vermutete Kompetenz und der Goodwill sind ebenfalls – wie weiter unten dargelegt werden wird – in Verbindung mit der Bewertung von Marken und auch IT-Systemen relevant.

Überträgt man das Erörterte auf das Vertrauen in ein IT-System, so würde das bedeuten, dass beispielsweise ein Fahrer dem KI-basierten autonomen Fahrsystem eines PKW beim Einsteigen freundlich und ehrlich zulächeln könnte, zur Signalisierung einer präreflexiven Vertrauensbereitschaft. Fahrer und KI fangen an, „sich zu verstehen" – woraufhin das KI-System sich besonders um diesen Fahrer kümmern würde, ihn per Handschlag oder Umarmung willkommen hieß, und sein Vertrauen keinesfalls enttäuscht. Eine eventuelle Fehlfunktion der KI (ein „Vertrauensbruch" der Maschine) mit einer Schädigung des Fahrers hätte bei der KI ein Störgefühl oder ein schlechtes Gewissen zur Folge. Das erscheint als einigermaßen absurd: Eine KI wird das Lächeln des Fahrers bestenfalls ikonisch erkennen können, aber keinesfalls wirklich „verstehen" – von einem vertrauensbildenden haptischen Körperkontakt zwischen IT-System und Benutzer ganz zu schweigen.

Das Urvertrauen weist allerdings auch Merkmale einer Institution (lateinisch institutum – Einrichtung) auf. In den Wirtschaftswissenschaften wird als „Institution" generell ein Ordnungs- oder Regelsystem bezeichnet, das Handlungen von Menschen so beeinflusst, dass sie für andere Menschen „berechenbar" werden. Eine Institution ist zwar nicht Personenunabhängig, aber sehr wohl Personen-neutral. In der Institution können beteiligte Personen durch andere ersetzt werden, ohne dass – im angestrebten Idealfall – die Institution darunter leidet oder gar zugrunde geht.

Das biologisch-evolutionär entstandene Regelsystem des Urvertrauens erlaubt – in gewissem Maße – den Personen-neutralen Austausch der Eltern durch andere pflegende Vertrauenspersonen. Und umgekehrt ist es zu beobachten, dass es in vielen Kulturen einen „Elternersatz" gibt, wie die Patenschaft. Die Paten sorgen als biologische Institution dafür, dass ein Säugling nicht verloren ist, wenn seine biologischen Eltern ausfallen sollten. Insofern ist das dem Menschen angeborene Urvertrauen auch ein Institutionenvertrauen – als Vertrauen in eine biologische Institution. Es mag an dieser Stelle dahingestellt bleiben, inwieweit der globale Erfolg des homo sapiens auch auf diesem Institutionenvertrauen und der damit möglichen Austauschbarkeit der Hilfspersonen für den Säugling beruht [Bowl01].

Archetypus Verstehen – verstehen und verstanden werden

Die Hermeneutik (griechisch ερμηνευειν – hermeneuein – auslegen, erklären) erörtert das Verstehen von Texten, Symbolen und Artefakten. Für das Verstehen bedarf es einer gemeinsamen sozialen Basis. In einem Festvortrag vor akademischem Publikum wurde im Herbst des Jahres 2022 eine „maschinelle Hermeneutik" erörtert und die Frage gestellt „Wissen Sie, wie lange Krokodile leben?" – worauf die einfache komödiantische Antwort „Ganz genauso wie die Kurzen!" gegeben wurde [Hof22b]. Das Publikum ließ durch ein Lächeln erkennen, dass man den Scherz „verstanden" hatte. Das Lächeln war ein hermeneutisch relevantes „Signal des Verstehens" sowohl für den Redner als auch für die anderen Anwesenden untereinander.

Das komödiantische Verstehen dieses Krokodil-Scherzes führt zu zwei grundlegenden hermeneutischen Fragen:

a) Könnte sich das Publikum dem Verstehen als solchem entziehen, speziell wenn man den Scherz nicht allzu lustig findet? Oder muss(!) man „es" (früher oder später) verstehen?

b) Was das Wesen dieses „es“ ist, dass da verstanden worden ist. Was genau wurde da – vom Redner zum Publikum – transferiert, als man „es“ (hier einen Scherz) verstanden hatte?

Zur Frage a) kann man feststellen, dass hier offenbar ein Archetypus wirkt, der den Menschen das Verstehen quasi aufzwingt, selbst wenn man das eigentlich gar nicht will – etwa, weil man den Scherz zu schlicht findet.

Das Verstehen der Welt scheint mithin für den Menschen unvermeidbar, kann er sich gegen das Verstehen gar nicht wehren.

Beauftragt man zur Scherzfrage „Wissen Sie, wie lange Krokodile leben?“ – und der Antwort „Ganz genauso wie die Kurzen!“ eine englische Version beim Google Übersetzer, so erhält man als Ergebnis: “Do you know how long crocodiles live?” – “Just like the short ones!” Man stellt aber fest, dass Google nicht lacht – es wird zur Übersetzung kein Smiley mit zurückgeschickt. Google lacht nicht – ganz offenbar kann Google nicht verstehen (nicht früher und auch nicht später!), um was „es“ hier eigentlich geht.

Wird eine Google-KI jemals diesen einfachen Krokodile-Scherz (früher oder später) als „lustig“ klassifizieren (können)? Unterstellt man, dass Google um eine umfangreiche Witzedatenbank erweitert wird, und Google sendet tatsächlich – zu dem in dieser Datenbank gefundenen Krokodile-Scherz – „blitzschnell“ einen Smiley. Aber, lacht Google wirklich? Oder wäre der Smiley nur ein bloßer Indikator für einen Treffer in dieser Witzedatenbank? Wenn aber Google den Smiley nun schneller schickt, als ein Mensch lächeln kann – hieße das, dass dieser „Lach-Algorithmus“ im „Verständnis der Sache“ dem Menschen geistig überlegen ist? Lässt sich das „Verstehen formal verstehen“ und darauf basierend das „Verstehen“ in einer KI programmieren? Diese erkenntnistheoretische Frage kann dahingestellt bleiben [Hof22b] – so auch für diesen Beitrag.

Zur Antwort auf Frage b) ist festzustellen, dass Menschen – wahrscheinlich exklusiv – das Verstehen ihrer Mitmenschen quasi „spüren“ können. Leider mussten die Dozierenden in den Corona-Videokonferenz-Vorlesungen mitunter erfahren, dass sie wegen der abgeschalteten Video-Rückkanäle nicht erkennen konnten, ob das „es“ von den Studierenden verstanden wurde, das die Vorlesungen vermitteln sollten. Auch im Kreis der Studierenden war nicht klar, ob „man“ das „es“ verstanden hatte. Das Fehlen einer Common-Sense-Atmosphäre (die Wahrnehmung allgemeiner Unsicherheit) führte zu Hemmungen, was das Stellen von Verständnisfragen anging.

Folgt man Werner Heisenberg, so bedeutet das „Verstehen“ in den Naturwissenschaften, speziell der Physik, dass nach Maßgabe nachvollziehbarer Experimente Daten gewonnen wurden, die in eine (mathematische, statistische) Modellbildung einfließen. Das verstehende Subjekt steht der objektiven Welt gegenüber. Die Modelle wiederum müssen nachvollziehbar sein und Prognosen zu einem Systemverhalten erlauben [Heis69]. Bei Erwin Schrödinger hingegen verstehen die Menschen die Welt als gemeinsame Intentionalität, als eine „kollektive Erfahrung“. Damit steht der Mensch nicht solo der Welt gegenüber. Die Subjekte (im Plural!) verstehen Sachverhalte holistisch als Teil der ihnen eigenen „gesamten“ Natur und Kultur [Schr85].

Luigi Nono und Karlheinz Stockhausen an den Internationalen Ferienkursen für Neue Musik, Darmstadt, im Jahr 1957. Das Bild zeigt zwei Personen in offenbar vertrauensvoller Unterhaltung; man will und man kann sich verstehen. [Bild: Seppo Heikinheimo]

Von diesen beiden Ansätzen des Verstehens ist Google mit seiner Übersetzung des Krokodile-Scherzes weit entfernt. Und der Google-Übersetzer hat offenbar kein allgemeines Modell für das, was ein Scherz sein könnte. Google könnte zwar per einfacher Rückmeldung behaupten, den Scherz verstanden zu haben – glaubhaft ist das nicht. Es wäre doch nur ein Finden des Scherzes in einer Datenbank. Man sollte mithin unterscheiden zwischen „einen Scherz erkennen“ und „einen Scherz verstehen“. Vielleicht kommt ersteres per Datenbank maschinell zustande. Das kann man fortgeschrittenen KI-Systemen schon zutrauen. Letzteres wird vermutlich nie maschinell gelingen.

Menschen unterliegen manchmal der Illusion der sogenannten anthropomorphen Projektion – Menschen vermuten quasi „Menschenartiges“ auch bei Tieren oder gar in Maschinen. Da gibt es den Hund oder das Pferd, denen man unterstellt, sie „verstünden“ ihre Besitzer. Und da gibt es Kraftfahrer, die ihrem Auto des Morgens gut zureden, der PKW möge bitte – trotz fast leerer Starterbatterie – „Verständnis zeigen“ und noch einmal anspringen und zum Arbeitsplatz fahren. Oder User, die die Enter-Taste des Computers extra fest drücken, um den Rechner zu mehr „Verständnis“ der Aufgaben und schnellerer Rechenarbeit zu motivieren [Hof22b].

Vertrauen in Institutionen und Organisationen

Der Ökonom (und Nobelpreisträger von 1993) Douglass Cecil North (1920 – 2015) definiert eine Institution über die (formalen wie informalen) „Spielregeln“, die das politische, wirtschaftliche und gesellschaftliche Zusammenspiel in einer Gesellschaft festlegen [Nort92]. Eine Institution ist zwar nicht Personen-unabhängig, sehr wohl aber Personen-neutral. In der Institution können beteiligte Personen durch andere ersetzt werden, ohne dass die Institution Schaden nimmt, indem die Spielregeln auch von diesen neuen Personen entsprechend beachtet werden.

Bekannte Beispiele für Institutionen sind etwa Nationen und Regierungen, die UNO, Kommunen, Behörden, Gerichte, Universitäten, Kirchen, auch DIN, ISO und IEC, das Rote Kreuz und die Malteser, und dergleichen mehr. Menschen vertrauen Institutionen, im Rahmen des Institutionenvertrauens. Eine Universität, beispielsweise, kann über Jahrhunderte hinweg als Institution bestehen und es kann ihr auch institutionell vertraut werden. Nichtsdestoweniger werden die Führungspersonen der Universität über die Jahrhunderte nicht dieselben bleiben können. Eine Institution gibt sich eine konkrete zeitgemäße Organisation für ihren

Betrieb und ihre Verwaltung. Ähnliches lässt sich für Nationen mit wechselnder Organisations- und Staatsform oder die Kirchen feststellen: Für ein Erreichen der gesteckten Ziele brauchen Institutionen eine Organisation.

Akzeptanzuntersuchungen im Metier des Cloud-Computing haben bereits gezeigt, dass Akzeptanz von IT-Anwendungen vom Vertrauen in deren institutionelles Umfeld abhängig ist. Nach [HoSc14] und [HoSc16] muss sich das Vertrauen allerdings auf die (menschlichen) Anbieter und Betreiber der IT-Systeme, weniger auf die (maschinellen) IT-Systeme als solche beziehen.

Mit dem Begriff „Organisation" (griechisch οργανον – lateinisch organum – Werkzeug) werden einerseits Rechts- und Wirtschaftssubjekte als die Struktur einer Tätigkeit oder Funktion bezeichnet. In diesem Beitrag interessiert der zweite Aspekt: Institutionen wie eine Partei oder ein Unternehmen verfügen jeweils über eine (interne) Organisation. Eine entsprechend schlecht gestaltete Organisation vermag das Vertrauen in eine Institution zu schädigen, wie etwa Beispiele korrupter Regierungen (Einfluss Dritter) im Bereich von Nationen, oder unzulängliches Qualitätsmanagement im Bereich von produzierenden Unternehmen zeigen.

Institutionenvertrauen als Grundlage von Kooperationen – Kontrolle von Organisationen

Das sogenannte „Institutionenvertrauen" bezeichnet das Vertrauen von Menschen in Institutionen. Das Institutionenvertrauen fördert die Kooperation von Personen (oder Personengruppen) mit Institutionen. Der Umgang mit Institutionen ist mit Unsicherheiten und unvollständigen Information verbunden. So dürfte beispielsweise ein Kunde kaum die gesamte Organisation der Produktions- und Handelsprozesse im Bereich von Lebensmitteln kennen und verstehen. Dennoch kauft und verzehrt der Kunde Lebensmittel nach Maßgabe seines Institutionenvertrauens in die Lebensmittelwirtschaft.

Im Zug einer praktikablen Lebensgestaltung übertragen Menschen Verantwortung und Kontrolle auf Institutionen, weil sie „solo" eben dieser Verantwortung nicht gerecht werden können, beziehungsweise die eigentlich erforderlichen Kontrollen aufwandsmäßig nicht bewältigen können. Dafür erwarten Menschen, dass die Institution ihre Aufgaben erfüllt – die jeweilig subjektiven Bewertungskriterien entsprechen. Erfüllt die Institution die übertragenen Aufgaben nicht, leidet das Institutionenvertrauen. Je höher das Institutionenvertrauen ist, desto mehr Akzeptanz erfährt die Institution, und desto höhere subjektive „Sicherheit" erfahren die aktiv Vertrauenden. Das Fehlen von Institutionenvertrauen führt in der Regel zu einer Verweigerung von Akzeptanz der (Angebote der) Institution.

Eine spezielle Rolle bei der Akzeptanz von Systemen spielt das kollektive Institutionenvertrauen, das aktiv von großen (Kunden-) Gruppen erbracht wird. Landläufig ausgedrückt besteht der Effekt darin, dass man annimmt, dass

- sich eine – hinreichend große – Kundengruppe quasi „schon nicht irren" kann, beziehungsweise
- eine große Nutzgruppe einen derart hohen sozialen Druck auf die Anbieter ausübt, so dass diese sich entsprechend intensiv bemühen, die Erwartungen der großen Kunden-

gruppe zu erfüllen, oder
- sich unter der großen Nutzergruppe sicher jemand findet, der ein eventuelles Problem bereinigt, wie man dies von den Bug-fixes bei Wikipedia oder Mozilla her kennt.

Die erfolgreiche Marktdurchdringung im Bereich der IT basiert nicht selten auf dem Finden einer hinreichend großen kritischen Zahl von Kunden, die die oben genannten positiven Akzeptanzeffekte auslöst. Dem steht die Möglichkeit eines fatalen kollektiven Irrtums gegenüber, wenn große Nutzergruppen von einen IT-Systemausfall kollektiv getroffen werden. Ein derart „falsches" kollektives Institutionenvertrauen beobachtet man etwa bei kollektiv akzeptierten und dann platzenden Spekulationsblasen.

In der Konsequenz stellt sich die Frage, ob – und wie – das Anwendungsszenario eines IT-Systems ein (passives) Institutionenvertrauen von (aktiv vertrauenden) Menschen erlangen kann.

Vertrauen versus Zuverlässigkeit – Zero-Trust-Transaction – Vertrauensbasierter Betrug

In der Ökonomie kann bei Transaktionen (wie Kaufvorgängen, etc.) unterschieden werden zwischen:
1) dem Zuverlässigkeits-Paradigma, der Annahme der beteiligten Parteien, dass die Transaktion zuverlässig einen günstigen Verlauf nehmen wird, und
2) dem Vertrauens-Paradigma, der notwendigen Voraussetzung von Vertrauen der beteiligten Parteien untereinander in einen günstigen Verlauf der Transaktion.

Als Abgrenzung zwischen 1) und 2) kann das Kriterium dienen, ob im Fall eines jeweils ungünstigen Verlaufs der Transaktion ein dahingehender Schaden kompensiert werden kann, beziehungsweise ob er ersetzbar ist.

Zu 1) Eine Annahme einer wahrscheinlichen Zuverlässigkeit bei Transaktionen zwischen zwei Parteien unterstellt, dass die jeweils andere die Erwartungen in Leistung und Gegenleistung zuverlässig erfüllt. So nimmt beispielsweise ein Verkäufer an, dass der Käufer den Kaufpreis entrichten wird. Im Gegenzug geht der Käufer davon aus, dass der Verkäufer die zu erwerbende Ware liefert. Wird die Transaktion korrekt durchgeführt, dann entsteht ein beiderseitiger Nutzwert. Nur wenn dieser Nutzwert höher liegt als der Erwartungswert des möglichen ungünstigen Verlaufs wird die Transaktion zu Stande kommen.

Wenn der ungünstige Verlauf – Käufer zahlt nicht, Verkäufer liefert nicht – bezüglich der Schadenssumme und der Eintrittswahrscheinlichkeit abgeschätzt werden kann, dann kann eine entsprechende Versicherung abgeschlossen werden, die einen Schaden kompensiert, beziehungsweise den Verlust ersetzt. In diesem Fall ist für die Transaktion keinerlei Vertrauen zwischen den Parteien der Transaktion erforderlich. Die abgeschlossene Versicherung ermöglicht die Transaktion auch unter sich misstrauenden Parteien.

Wie 1) zeigt, kommen ökonomische Transaktionen auch bei explizitem Misstrauen der Parteien zueinander zustande – nicht jede Kooperation setzt also Vertrauen voraus. Dies wird auch als „vertrauenslose Kooperation" oder „Zero-Trust-Transaction" bezeichnet: Im politischen Raum werden Kooperationen unter sich misstrauenden Parteien manchmal mit sogenannten „Vertrauensbildenden Maßnahmen" abgesichert – wie den Confiden-

ce-Building Measures (CBMs) der Organization for Security and Co-operation in Europe (OSCE). Die CBMs sind in der Regel eine Reihe von gegenseitigen Informations- und Offenlegungs-Pflichten [CBMs23]. Anzumerken ist, dass die CBMs keinerlei Vertrauen zu bilden in der Lage sind – sie sollten daher besser als „Vertrauensersatz-Maßnahmen" bezeichnet werden.

Eine Zero-Trust-Transaction bringt die Notwendigkeit umfassender Vorsichtsmaßnahmen – den CBMs, wie Versicherungen, Überwachungen, Kontrollen, etc. – mit sich. Diese vertrauensersetzenden Vorsichtsmaßnahmen können sehr aufwändig und teuer sein.

Besonders dramatische Auswirkungen kann ein übertriebenes Misstrauen haben. Die Neigung, dem sozialen Kontext immerzu feindselige Aktivitäten gegenüber der eigenen Person zuzuordnen, kann in der Lebenspraxis sehr lästig sein. Nicht nur für die Betroffenen ist ein (krankhaftes?) Misstrauen sehr aufwändig, das Ausräumen von unberechtigten Verdächtigungen ist für die Mitmenschen ebenfalls nicht immer einfach.

Korrespondierend zur „Zero-Trust-Transaction" existiert das Phänomen des „Vertrauensbasierten Betrugs". So gelingt es beim Betrug des sogenannten „Enkeltricks" dem Trickbetrüger meist über Telefon oder Messenger-Diensten gegenüber den Opfern ad-hoc ein – trügerisch falsches – Vertrauensverhältnis aufzubauen, so dass die Betrüger, sich als nahe Verwandte ausgebend, beispielsweise die Herausgabe von Bargeld erwirken können.

Zu 2) Ein Vertrauen bei Transaktionen ist speziell dann absolut erforderlich, wenn der ungünstige Verlauf nicht bezüglich der Schadenssumme und der Eintrittswahrscheinlichkeit abgeschätzt werden kann. Deshalb kann keine entsprechende Versicherung abgeschlossen werden, die einen Schaden vollständig kompensieren könnte.

Ein erster typischer Fall, der notwendigerweise auf Vertrauen basieren muss, ist die Übergabe unersetzlichen Vermögens an Dritte, wie beispielsweise die private Leihgabe eines Kunstwerks an ein Museum. Der Verlust des Originals ist finanziell nicht wirklich zu kompensieren. Das Vertrauen des Leihgebers ist vor allem ein Institutionenvertrauen in das Museum, das über ein sicheres Gebäude, seriöses und geschultes Personal und dergleichen verfügt. Dem vergleichbar ist die Übergabe unersetzlicher Daten zur Daten-Speicherung in einem CDS.

Ein zweiter Fall notwendigen Vertrauens ist die Auslieferung der eigenen Person an normative Handlungen Dritter, wie beispielsweise einem Chauffeur in einem Taxi oder einem Reisebus. Ein Verkehrsunfall mit Körperverletzungen, gar Invalidität, ist nicht finanziell zu kompensieren. Von daher ist ein Vertrauen auf zwischenmenschlicher Ebene in die Qualifikation, Gesundheit und den guten Willen des Fahrers ohne Alternative. Dem vergleichbar ist die Auslieferung der eigenen Person an normative Maschinen, wie einem KI-Autopiloten in einem Fahrzeug.

Ein dritter Fall ist die Annahme, dass eine Person bezüglich eines – für das Gelingen einer Transaktion – entscheidenden Faktums nicht die Unwahrheit sagt. Die Kooperationspartner im Fall 2) zeigen offenbar ein gegenseitiges „spürbares Verstehen" der Situation des jeweils anderen, aus dem ein Vertrauen erwächst, das wiederum die Transaktion zustande kommen lässt.

In beiden Fällen zu 2) stellt sich zudem die Frage, wie ein „Vertrauen“ in IT-Systeme dargestellt werden kann. Keinesfalls haltbar ist eine Synonymität der Begriffe „vertrauenswürdig“ und „zuverlässig“.

Zuverlässigkeit von (IT-) Systemen – externer Einfluss und Sabotage

Die Zuverlässigkeit eines technischen (IT-) Systems bezeichnet den Umstand, ob das System in einem gegebenen Zeitintervall wie beabsichtigt funktioniert. Zuverlässigkeit kann stochastisch und quantitativ beschrieben werden.

Die Zuverlässigkeit großer (IT-) Systeme kann in der Regel nicht allein empirisch durch die Beobachtung der Ausfallhäufigkeit ermittelt werden. Mit analytischen Zuverlässigkeitsmodellen wird die Ausfallstruktur des Gesamtsystems modelliert. Zuverlässigkeitsanalysen korrespondieren mit dem Zuverlässigkeitsmanagement, welches speziell zuverlässigkeitserhöhende Maßnahmen umfasst, dazu zählen unter anderem:

- Bewährte und qualifizierte Komponenten,
- Einsatz redundanter Komponenten,
- Ex-post-Auswertung der Zuverlässigkeitsdatenbasis.

Für ein technisches System besteht prinzipiell die Möglichkeit des Ausfalls. Offenbar haben komplizierte und komplexe Systeme eine höhere Ausfallneigung als einfache Systeme.

Zwischen Zuverlässigkeit und Vertrauen besteht ein grundlegender struktureller Unterschied: Zuverlässigkeit ist eine Eigenschaft eines (IT-) Systems, während Vertrauen eine sozio-psychologische Relation (Beziehung zwischen Parteien) ist.

Für die Zuverlässigkeit von technischen (IT-) Systems nicht unwichtig sind externe Einflüsse in Form von Sabotage oder als Effekt krimineller Handlungen. In der Konsequenz sind bei Zuverlässigkeitsanalysen und dem Zuverlässigkeitsmanagement entsprechende Bedrohungsanalysen unabdingbar. Die Zuverlässigkeit von (IT-) Systemen hängt ab von einer gelingenden Abwehr von Sabotage und Cyberkriminalität.

Markenvertrauen als spezielle Form des Institutionenvertrauens

Das sogenannte „Markenvertrauen“ als spezielle Form des Institutionenvertrauens verdient eine besondere Aufmerksamkeit. Für IT-Systeme kann durchaus eine „Markenposition“ erreicht werden, und somit einem IT-System ein Markenvertrauen entgegengebracht werden. Im Kontext dieser Betrachtung soll der Begriff „Marke“ nicht (nur) juristisch – etwa als Schutzmarke – verstanden werden, sondern vielmehr im Sinne

a) eines Herkunfts- oder Markenzeichens für eine Ware. Die Nutzer oder Käufer einer Ware wollen den Hersteller sicher identifizieren können. Der Aspekt wird typischerweise durch eine – juristisch abgesicherte – Kennzeichnung erfüllt, meist ein Signet („(geschütztes) Markenzeichen“) des Herstellers.
b) einer Abkürzung von Lieferbedingungen. Die Nutzer oder Käufer einer Ware wollen sicher sein, dass die Ware oder Dienstleistung gewisse Eigenschaften hat, und damit spezielle Qualitätsmerkmale erfüllt.

Zu a) stellt die Marke für die Verwender eines Markenzeichens ein Instrument der Marktzugangs-, Preis- und damit der Produktpolitik dar. Die Marke ist ein Instrument, um die Eigenschaften der eigenen Produkte deutlich zu signalisieren und sie von Konkurrenzprodukten einigermaßen sicher zu unterscheiden.

Zu b) kann man etwa die nationale und internationale Standardisierung sehen. So ist eine ISO- oder DIN-Norm ein Standard, der Objekte beschreibt, die insbesondere Gegenstand einer Handelstransaktion sein können. So kann sich ein Kauf eines Schlosserhammers auf eine Spezifikation nach DIN 1041 beziehen. Der Käufer erspart sich so die aufwändige Beschreibung des zu erwerbenden Schlosserhammers in Form sicherheitstechnischer Festlegungen, Abmessungen und Gewichte, etc.

Beim Markenvertrauen zeigen sich zwei Phänomene [RoWo15]:

c) Zuverlässigkeitsaspekt: Die Käufer gehen davon aus, dass ein Marken-Produkt die mit der Marke assoziierten Eigenschaften und Qualitätsmerkmale zuverlässig erfüllt.
d) Vertrauensaspekt: Die Käufer vertrauen dem Anbieter beziehungsweise Hersteller des Produkts, dass diese die zugesicherten Qualitätsmerkmale der Marke tatsächlich einhalten.

Damit ein Markenvertrauen wirksam werden kann, müssen sich die Produkte der Marke von alternativen Angeboten unterscheiden lassen. Dies wird ermöglicht durch einerseits eine entsprechende „objektive" Gestaltung der Produkte selbst, andererseits durch die Etablierung einer sozio-ökonomischen „Markenumgebung", wie sie etwa die sogenannten „Standard-Marken" realisiert haben [Lang03].

Der Zuverlässigkeitsaspekt zu c) kann insbesondere durch Fehler in der Produktion, aber auch durch Dritte als Markenfälschung („Produktpiraterie") gestört werden. Dem Markenvertrauen in eine qualitativ hochwertige Original-Ware ist es abträglich, wenn diese wegen der Ähnlichkeit mit einem minderwertigen Plagiat verwechselt wird.

Der Vertrauensaspekt zu d) schafft zwei Sicherheiten. Zum einen, dass die Anbieter die Kunden zu den Produkteigenschaften nicht belügen. Und zweitens, dass die Anbieter der Markenprodukte alle Situationen berechtigter subjektiver Unzufriedenheit des Käufers zu bereinigen in der Lage sind. Dies bedeutet, dass die Anbieter in der Lage sind, eventuelle begründete Beschwerden im Interesse der Kunden auszuräumen. Diese Problemlösungsabsicht „Guten Willens" trägt zur Kundenloyalität bei.

Markenvertrauen schafft nicht zuletzt die Sicherheit, dass der Markenanbieter seine Kunden dahingehend versteht, dass er den Gebrauch und Anwendung des Produkts auf den Kunden abzustimmen in der Lage ist. Kunden vertrauen – symmetrisch – darauf, dass der Anbieter das gewohnte Qualitätsniveau der Produkte einhalten wird.

Die Anbieter wiederum haben „einen Ruf zu verlieren" – einen quasi „Bruch des Markenvertrauens" werden sie nach aller Möglichkeit zu vermeiden versuchen, da dies ihre Marktposition ruinieren könnte. Durch entsprechende organisatorische Mechanismen (Feedback, Vorschlagswesen, Kulanz, etc.) des offenen Diskurses mit den Kunden etablieren und bewahren Anbieter eine entsprechende vertrauenswürdige Markenkultur. Diese Markenkultur verhindert wiederum, dass das Vertrauen der Kunden enttäuscht wird.

Anthropomorphe Projektionen als Grundlage irrigen Vertrauens in Maschinen

Es ist dem Menschen seit jeher eigentümlich, dass er zu anthropomorphen (griechisch ανθρωπος – anthropos – Mensch, sowie μορφη – morphe – Gestalt) Projektionen neigt [Hof22a – pp. 213ff.]. Menschen erkennen auch in Tieren oder Dingen etwas Menschenähnliches. Man meint etwa, ein Haushund würde den Menschen wirklich „verstehen" und daher lernen [Zime92]. Auch in Computer und Roboter wird ein humanes Wesen hineinprojiziert und sie werden damit quasi „sozial und intellektuell überschätzt" [Jime11] [Hof22a – pp. 216ff.].

Die algorithmische Datenverarbeitung ist mit dem biologischen menschlichen Denken keinesfalls gleichzusetzen. Menschliches Wissen beinhaltet Elemente gemeinsamer Intentionalität, die die „Wahrheit der Wissensinhalte" markiert [Toma10]. Umgekehrt speichert das menschliche Gehirn sehr wohl – quasi-maschinell – gewisse Dateninhalte, es erfolgt im menschlichen Gehirn auch eine gewisse algorithmische Datenverarbeitung. Der Mensch erkennt nun die partielle Ähnlichkeit seines Denkens mit dem Wirken der Datenverarbeitungsmaschinen – und es kommt zu der psychischen Projektion, dass diese Maschinen menschenähnlich, anthropomorph, etwas „wissen" könnten. Man kann im Folgenden einige Aspekte anthropomorpher Projektion identifizieren.

Erkennen von Gesichtern als Anthropomorphe Projektion

Die Menschen erkennen menschliche Gesichter und ihre emotionale Botschaft quasi sofort. Der Mechanismus funktioniert auch am Rande des Gesichtsfeldes und übersteuert durchaus, wenn auch Nicht-Menschen und Sachen als Menschen-ähnlich erkannt werden und ihnen ein Gesichtsausdruck zugebilligt wird. Die Fähigkeit des Menschen, eine Identität zu erkennen, kann quasi „übersteuern" [Hofm21].

Das „freundliche Gesicht" eines Musikschranks aus den 1960er Jahren. [Bild: Hofmann]

Die sofortige empathische Einschätzung von Mitmenschen mag überlebensnotwendig und ein Ergebnis der erfolgreichen Evolution des Menschen sein. Freundliche und feindliche Gesichtsausdrücke müssen schnell und sicher erkannt werden. Dieser ererbte biologische Mechanismus lässt Menschen aber auch in Maschinen „Gesichtsausdrücke" hineinprojizieren und erkennen.

Nicht-humanes Lernen als Anthropomorphe Projektion

Die Menschen lernen und lehren per genetischem Programm. Das Erklären von Sachverhalten gegenüber anderen Mitmenschen und das Erkennen, dass die lernenden Gegenüber den Inhalt verstanden haben – das kann einem Menschen durchaus Freude machen [Hof22b].

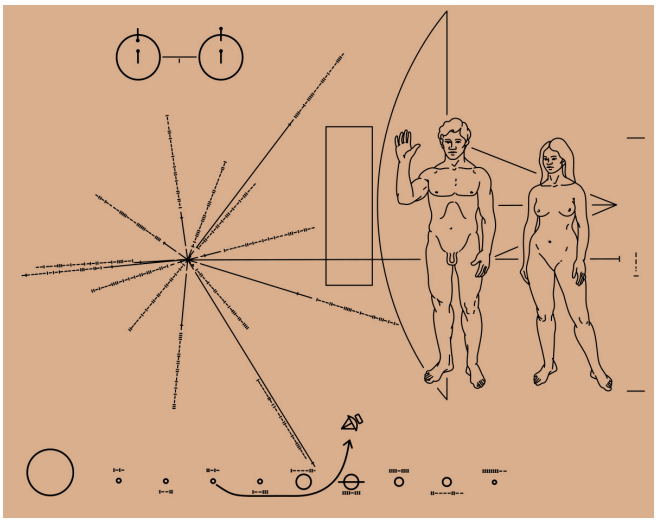
Es ist ein kleines Erfolgserlebnis, wenn man einer Maschine „etwas beibringen“ konnte, und wenn es nur darin bestand, einem Wecker per „Programmierung“ eine Weckzeit beizubringen. Es gibt Menschen, die mit Vergnügen Computer programmieren – weil es eben einfach Spaß machen kann, zu sehen, dass die Maschine das macht, was man ihr beigebracht hat.

Ein einfacher Automat wie ein Wecker kann „lernen“, wann er eine Person zu wecken hat. Der Wecker – wie alle Automaten – ist natürlich von einem menschlichen Lernen weit entfernt. Er weiß nicht, was er tut, er versteht die Bedeutung des Weckvorgangs nicht. [Bild: Deutsches Uhrenmuseum, Glashütte].

Auch einem Hund kann per Dressur vielerlei beigebracht werden. Auch hier stellt sich eine gewisse Befriedigung ein, wenn der Hund das Gelernte korrekt ausführt. Maschinen oder Tiere assistieren dem Menschen bei dessen Unzulänglichkeiten. Die Projektion besteht darin, (fälschlicherweise) anzunehmen, dass Automaten oder Tiere Handlungsabläufe tatsächlich „verständnisvoll erlernen“ könnten. Wobei über diesen speziellen Aspekt hinaus Automaten und Tiere keinesfalls gleichgesetzt werden sollen, erst recht nicht, wenn es um den „besten Freund des Menschen“ geht.

Nicht-humanes Wissen als Anthropomorphe Projektion

Die Menschen können Faktenwissen erwerben und im Gedächtnis speichern [Hofm21 – pp. 220]. Die maschinelle Speicherung von Daten wird als „Speicherung von Wissen“ fehlinterpretiert. Ohne eine menschliche Erklärungskomponente für die Interpretation der Texte und Symbole bleiben diese ohne hermeneutische Funktion.



Anders als Daten existiert ein „Wissen“ nicht von Menschen unabhängig. Die Plakette an den Raumsonden Pionier II sollte Außerirdischen irdisches „Wissen“ vermitteln – sie ist aber schon hier auf der Erde reichlich nutzlos. Das liegt an den unerklärten formalen Symbolen: Die Plakette ist als Wissensspeicher untauglich, weil die für die Wissensübermittlung nötige Hermeneutik fehlt. Die Symbole kann niemand ohne menschliche Erklärung interpretieren, obwohl die eigentlichen Symbole und Datenlage doch deutlich lesbar sind. Die adressierten Außerirdischen werden diese Plakette niemals entziffern können, denn eine Erklärung durch einen Menschen fehlt draußen in den Tiefen des Alls.
[Bild: NASA]]

Die Projektion besteht darin, (fälschlich) anzunehmen, dass nicht-humanes „Wissen“ in der Form selbsterklärender Dokumente oder als wissende Computer existiert: „Ein Buch weiß nicht, was in ihm steht.“ – das Zitat wird Martin Heidegger zugeschrieben.

Nicht-humanes Verstehen als Anthropomorphe Projektion

Wenn Menschen anderen Menschen etwas erklären, dann „spüren“ sie es, wenn diese anderen Menschen das Erklärte tatsächlich verstanden haben [Hof22b] [Hofm21 – pp. 223ff.]. Auch Computer oder Tiere „hören zu“ und geben verständnisvolle Antworten und zeigen ein scheinbares Verständnis von Sachverhalten.

Ein Vertreter der Art Canis Lupus (Wolf) macht einen recht verständigen Eindruck. Dieser wird durch die Zuchtform mit der Ausprägung großer Augen und einem nach vorn gerichteten menschenähnlichen Blick noch verstärkt.
[Bild: Elles Rijsdijk]

Die Projektion besteht darin, (fälschlich) anzunehmen, dass „Der Hund die Sprache des Menschen versteht, obwohl er selbst nicht reden kann“ – so Kaiser Friedrich II. in seinem Buch über die Falknerei. Friedrich II. schloss daraus, dass dies der Grund für die leichte Dressierbarkeit des Hundes sein müsse. Ähnlich ist eine Projektion, anzunehmen, dass der Google-Übersetzer den notorischen Krokodil-Scherz wirklich „verstehen“ und darum übersetzen kann.

Nicht-humanes Vertrauen als Anthropomorphe Projektion

Die Menschen vertrauen Menschen – dies hat vielfältige Gründe. Misstrauen unter Menschen ist sehr aufwändig und teuer. Umgekehrt löst ein Vertrauensbruch psychische Störgefühle aus. Menschen vertrauen auch Tieren und Maschinen – weil ein Misstrauen in der Haustierhaltung keine brauchbare Alternative ist.

Ein junger Mensch vertraut einem Hund. Es ist schwierig, aus der bisherigen Verlässlichkeit des Tieres bedingungslos auf sein künftiges Verhalten schließen zu wollen [Zime92]. Die meisten Angriffe durch Hunde auf Menschen dürften durch tierische „Ersttäter“ erfolgen, die sich bislang einwandfrei verhalten haben.
[Bild: Vertrauen – de.wikipedia.org].

Die Passagiere und Piloten vertrauen dem komplexen und komplizierten Autopiloten eines Verkehrsflugzeuges. Das Flugzeug vom Typ Boeing 737 max wurde ausführlich getestet und hatte sich als sehr zuverlässig erwiesen. Bekanntermaßen kam es dennoch zu fatalen Unglücksfällen.
[Bild: Acefitt – en.wikipedia.org].

Die Projektion besteht in einem (falschen) Vertrauen in Tiere oder Maschinen nach Maßgabe der bislang gezeigten Zuverlässigkeit.

Kritische Wertung

Zu unterscheiden sind die Begriffe „Daten“ und „Wissen“. Man kann sich zu einer reinen Faktenfrage zusätzliche erkenntnistheoretische Fragen stellen. Sieht man als Beispiel die Frage nach einem Faktum (als Datenlage), etwa wie die Hauptstadt des US-Bundestaates Kalifornien heißt. Die Faktenfrage nach der kalifornischen Hauptstadt kann per Wikipedia geklärt werden. Interessanter ist die Antwort auf die Frage, woher das System das erfragte Faktum bezogen hat, und noch spannender, woher man denn wissen kann, dass die Systemantwort wirklich wahr ist.

Für einen Menschen kann das Faktum „Sacramento ist die Hauptstadt von Kalifornien“ die Auskunft einer Person gewesen sein, wobei man sich an diese Person und den Vorgang selbst gar nicht mehr erinnern muss. Entscheidend war die Glaubwürdigkeit der Person, und darum hat man sich das Faktum als „wahr“ gemerkt. Das glaubwürdige und damit in einem sozialen Kontext akzeptierte Wissen einer Person ist ein durch Computer nicht simulierbarer „Social Common Sense“.

> Die Unterscheidung der Begriffe „Daten“ und „Wissen“ führt zur Frage nach der dem Vertrauen in eine Informationsquelle. Das wahre und verlässliche Wissen ist ein in der natürlichen Intelligenz biologisch fundiertes, intersubjektiv gebildetes, soziales Phänomen.

Ein Faktum gilt dann als „wahr“, wenn es aus einer verlässlichen Quelle stammt, wenn also sozio-psychologisch vertraute Personen als Autoren fungieren oder fungierten. In der entwickelten Informationsgesellschaft etablierte sich konsequenterweise die Erkenntnis, dass das sogenannte „Wissen“ eine anthropozentrisch verankerte gemeinsam-intersubjektive Intentionalität ist.

Zu Beginn des Jahres 2023 wurde das KI-System „ChatGPT“ in den Nachrichten diskutiert. Einerseits wurde Chat-GPT als großer Durchbruch in der KI gesehen. Andererseits wurde auf Risiken hingewiesen. Es ist nicht hinreichend bekannt, wie ChatGPT die Texte auswählt, die in seine Antworten einfließen. Die KI orientiert sich an Texten, die im Internet vorhanden ist. Wenn oft im Netz behauptet wird, dass die Erde eine Scheibe ist, so könnte ChatGPT diese Falschinformation übernehmen und weitergeben. [https://www.tagesschau.de/inland/regional/nordrheinwestfalen/wdr-story-53399.html]. Konsequenterweise wird an der Entwicklung von Systemen gearbeitet, die entscheiden können, ob ein Text von einem Menschen oder einem Computer geschrieben wurde. Damit soll der Sorge begegnet werden, dass etwa bei Studienarbeiten geschummelt wird, oder auch Desinformationskampagnen generiert werden. [https://www.zeit.de/digital/internet/2023-02/chatgtp-erfinder-programm-schummeln-schularbeiten-kuenstliche-intelligenz]

Das erkenntnisbasierte Wissen beim Menschen besteht aus audio-visuellen, aber auch haptischen und anderen unmittelbaren biologisch-physischen Erkenntniselementen. Damit wäre klar, dass das, was der Mensch holistisch unter „Verstehen“ versteht, einem quasi „körperlosen“ Computer nicht zugänglich sein kann – völlig unabhängig von der Rechenleistung der Maschine. Das menschliche Verständnis der Welt hat seine Ursache nicht in einer reinen Rechenleistung des Gehirns, sondern auch in der biologischen Physis. Da den Maschinen das holistische Verstehen und damit die Bedeutung der Modelle fehlen, kann man sagen, dass die Computer nicht wissen, was sie tun.

Ein frühes Beispiel einer anthropomorphen Projektion einer „verstehenden Maschine“ war das bereits im Jahr 1966 von Joseph Weizenbaum vorgestellte Computerprogramm namens „ELIZA“. Es war ein interaktives Programm, es konnte per schriftlichem Wortwechsel verschiedene Gesprächspartner simulieren. Berühmt wurde ELIZA für die – freilich ziemlich oberflächliche – Simulation einer Gesprächspsychotherapie. Sarkastische Zeitgenossen schlossen daraus, dass die Gesprächspsychotherapie per se eine triviale Angelegenheit sein müsse. Einige euphorisierte, aber ansonsten intellektuell tieffliegende Psychotherapeuten glaubten gar, mit ELIZA die Psychotherapie weitgehend automatisieren zu können. Joseph Weizenbaum war in der Tat darüber mehr als befremdet, in der Folge fand er zunehmend zu einer systemkritischen Position [Hof22a – pp. 226f.].

Es ist der geradezu selbstverständliche Anspruch der KI, einige der physio-biologischen Phänomene des menschlichen Gehirns elektromechanisch nachzuahmen.

Ein Begriff, der ab dem Jahr 2009 verstärkt durch die Welt geisterte, war der der „Singularität“. Damit meinten einige der KI-Protagonisten den Punkt, an dem die KI die „Rechenleistung“ des menschlichen Hirns übertroffen hätte – und damit die KI dem Menschen an „Verstand“ überlegen sei. Diese Art der Betrachtung ist aber wenig zielführend, weil das menschliche „Verstehen“ von Sachverhalten mehr ist als nur das Anwenden einer – wenn auch enormen – formalen Rechenleistung.

Im Jahr 1997 gab es ein interessantes Schach-Experiment. Der damalige Schach-Weltmeister Garri Kasparow spielte gegen den IBM-Computer „Deep Blue“. Man war damals sehr gespannt, ob es einer Maschine gelingen könnte, den besten Schachspieler der Welt zu besiegen. Das würde bedeuten, dass Schachcomputern, wie „Deep Blue“ jeden menschlichen Gegner schlagen könnten. Dies gelang der Maschine tatsächlich, aber es war für Informatiker auch nicht weiter überraschend. Sie wussten schon immer, dass genau das eines Tages passieren würde, eben weil Schach „nur“ ein formales Spiel ist. In der Folge wurde der Fortschritt speziell in der KI zunehmend als ein Konkurrenzkampf zwischen Mensch und Maschine betrachtet [Hof22a – pp. 310].

Deep Blue schlägt Kasparow, Briefmarke zu € 1,10, Deutschland, 2021. Eine Gedenkmarke zur Würdigung einer Niederlage: Im Februar 1996 traten Schachweltmeister Garri Kasparow und Deep Blue von IBM gegeneinander an, und der Computer besiegte den amtierenden Weltmeister. Etwa 25 Jahre später gewinnen auch preiswerte Schachprogramme gegen die meisten menschlichen Gegner.

Es gab sogar eine deutsche Gedenkmarke zur Würdigung des 25. Jahrestags der Niederlage(!) des damaligen Schachweltmeisters Garri Kasparow gegen den IBM-Schachcomputer „Deep Blue“. Dass maschinelle Rechenkraft – punktuell – stärker als die menschliche sein kann, war ein logisches und vorhersehbares Ereignis. Man kann durchaus darüber sinnieren, warum es damals keine Sondermarke gab, als die erste Eisenbahn schneller als eine Pferdekutsche gefahren ist.

Die völlig übertriebenen Erwartungen in die Möglichkeiten (neuer) Computer, die übertreibend auch „Elektronengehirne“ genannt wurden, haben eine gewisse jahrzehntealte Tradition. Dies ist jeweils auf diverse anthropomorphe Projektionen zurückzuführen.

Ansätze und Kritik zum „Vertrauensaufbau“ in Maschinen

Eine einfache Suche im Netz mit dem Stichwort „Vertrauen in KI“ fördert per Jahresbeginn 2023 eine gewisse allgemeine Skepsis – eine quasi „Vertrauenskrise“ – zutage. Ein belastbares Vertrauen in Maschinen wird nicht unkritisch gesehen. Als exemplarisch seien angeführt:

I KPMG AG (2021): Geringes Vertrauen in KI: Nur etwa ein Viertel (28 Prozent) der Befragten in den fünf Ländern ist bereit, künstlicher Intelligenz im Allgemeinen zu vertrauen. (https://home.kpmg/de/de/home/themen/2021/05/studie-buergerinnen-und-buerger-haben-wenig-vertrauen-in-ki.html)

II. Martina Mara (2019): Der Umgang mit KI erfordert informiertes Vertrauen. Man weiß aus Forschung und Praxis, dass viele Menschen KI nur sehr wenig Vertrauen entgegenbringen und sich zum Beispiel gar nicht vorstellen können, relevante Entscheidungen an ein neuronales Netz auszulagern. Gleichzeitig gibt es aber das Phänomen des Overtrust, des völligen Überschätzens maschineller Fähigkeiten. (https://www.zukunftsinstitut.de/artikel/der-umgang-mit-ki-erfordert-informiertes-vertrauen/)

III Studie „AI-Ambitions 2022“ von Fivetran (2022?): (…) vertrauen 90 Prozent der Unternehmen künstlicher Intelligenz nicht genug, um auf menschliche Entscheidungsfindung zu verzichten. Die Umfrage unter CIOs, IT-Verantwortlichen und Data Scientists zeigt: Trotz großer Ambitionen und Investitionsbereitschaft gelingt es Unternehmen nur bedingt, KI im Rahmen von Entscheidungsprozessen zu nutzen. (https://www.it-finanzmagazin.de/90-prozent-der-unternehmen-vertrauen-kuenstlicher-intelligenz-ki-nicht-genug-145480/)

IV Rainer Hoffmann, EnBW AG (2021): Kann eine Technologie wirklich vertrauenswürdig oder sogar verantwortlich sein? Ganz klar nein! Trotzdem gewinnen mit zunehmendem Einsatz von Künstlicher Intelligenz (KI) Begriffe wie „Trustworthy AI“, „Responsible AI“ oder „Ethical AI“ an Bedeutung. Was steckt also dahinter? Technologien menschliche Eigenschaften zuzuschreiben (Anthropomorphismus) ist leider irreführend. (https://www.energie-klimaschutz.de/vertrauen-in-kuenstliche-intelligenz/)

In der Konsequenz wird gefragt, wie sich Vertrauen in Maschinen aufbauen ließe. Es sollen in der Folge drei Ansätze vorgestellt werden.

Ansatz von Hoffmann

Bei Hoffmann [Hoff21] wird – richtigerweise – bezweifelt, dass man Verantwortung, moralische Abwägungen und Vertrauen von einer Technologie erwarten kann. Eine Technologie werde von Menschen für bestimmte Zwecke eingesetzt. Daher müsse man in die Menschen

vertrauen, die eine Technologie nutzen und betreiben. Menschen seien Teilhaber der Organisation, welche Werte wie Verantwortung, moralische Abwägungen und Vertrauen glaubhaft vertritt. Will man also einer KI-Lösung vertrauen, muss das Vertrauen darin bestehen, dass die involvierten Organisationen KI verantwortungsvoll betreiben und anbieten. Zwei Werte seien von den KI-betreibenden Organisationen zu verfolgen, nämlich:

1) Transparenz, und
2) Gerechtigkeit.

Danach bedeutet nach 1) die Transparenz einer KI, dass die Herkunft und die Eigenschaften der verwendeten (Trainings-) Daten bekannt sind, ebenso wie das Lernverfahren des Algorithmus. Es muss nachvollziehbar sein, wie das System zu seinen Ergebnissen kam. Dieser Aspekt ist auch unter dem Begriff der „Explainable AI" bekannt. Transparenz nach 1) bedeutet weiter, dass die Entwicklung des gesamten Systems dokumentiert wurde und Entscheidungen des Systems festgehalten werden. Transparenz kann auch bei Black-Box Verfahren erreicht werden, bei denen das Zustandekommen des Ergebnisses vollumfänglich durch den Menschen nachvollziehbar ist.

Eine Kritik einer „vertrauensbildenden Transparenz" wird anführen, dass es sehr wohl Markenprodukte mit aktivem und passivem Vertrauen gibt, bei denen die Hersteller – etwa im Bereich von Lebensmitteln – sowohl Details der Zutaten als auch der Herstellung gegenüber den Kunden verschweigen – im Sinn von „nach altem Geheimrezept mit natürlichen Zutaten". Offenbar ist eine mangelnde Transparenz dem Markenvertrauen nicht unbedingt abträglich.

Zu 2) der vertrauensbildenden Eigenschaft der Gerechtigkeit von KI-Systemen wird ausgeführt, dass KI-Systeme keine Werkzeuge zur Verletzung gesellschaftlich akzeptierter ethischer Werte sein dürfen. Dies bedeutet etwa den Ausschluss von Diskriminierungen; keine Vorteilserlangung für eine bestimmte Gruppe, Vertretbarkeit der sozialen Auswirkungen der KI-Systeme.

Eine „vertrauensbildende Gerechtigkeit" existiert bei Markenprodukten kaum. Hier kann man sehen, dass es Markenprodukte mit großem passivem Vertrauen gibt, die aber wegen ihrer ethisch-relevanten Auswirkungen keine Akzeptanz (mehr) erfahren oder gar verboten werden. So dürfte das Markenvertrauen in Glühbirnen, Plastiktragetaschen, Pflanzenschutzmittel, Feuerwerkskörper durchaus hoch sein – gleichwohl wurden die Produkte verboten, beziehungsweise sie sind ethisch umstritten. Ethische Werte der Anwendung sind orthogonal zum eigentlichen Markenvertrauen positioniert; sie sind allerdings relevant für die Gesamtakzeptanz der Produkte.

Bei [Hoff21] wird eine Abwägung zwischen Folgen und Vertrauen empfohlen. Nicht für jede KI-Lösung muss Vertrauenswürdigkeit umgesetzt werden. Die Abwägung zwischen den Folgen und dem Vertrauen einer KI-Lösung ist zu beachten. Eine KI-basierte Fernsehprogrammempfehlung ist eher harmlos, verglichen mit einer Personalauswahl oder gar einem KI-basierten autonomen Fahrzeug.

Ansatz von Coester und Pohlmann

In [CoPo20] stellen Coester und Pohlmann die Frage „Wie können wir der KI vertrauen?“ und schlagen als Antwort einen „Mechanismus für gute Ergebnisse“ vor.

Zunächst stellen [CoPo20] fest, dass innovative Technologien und die IT-Infrastruktur zunehmend vielschichtig und auch intransparent werden. Daraus resultiert das Dilemma, dass zu vermehrtem Einsatz das Wissen über Hintergründe der IT-Systeme zurückbleibt. Dies könnte entweder zu unverhältnismäßiger Ablehnung oder blindem Vertrauen führen. Beides verhindert eine sinnvolle Nutzung. Als eine Definition von „Vertrauen“ wird bei [CoPo20] die subjektive Überzeugung von der Richtigkeit einer Aussage oder von Handlungen angegeben. Ein KI-System kann generell als vertrauenswürdig eingestuft werden, wenn es sich für den vorgesehenen Zweck immer wie erwartet verhält. Daraus lässt sich folgern, dass Vertrauenswürdigkeit nachweisbar ist.

Eine Kritik muss darauf hinweisen, dass hier wohl „Zuverlässigkeit“ mit „Vertrauen“ verwechselt wird. Die angesprochene Nachweisbarkeit ist ein deskriptiver ex-post Mechanismus, der aus einer beobachteten Reihe von Verhaltensweise auf die Zukunft schließt. Analog könnte man meinen: Da der FC Bayern München in den Jahren von 2012/2013 bis 2021/2022 zuverlässig Deutscher Fußballmeister wurde, wird er das sicher auch in den Folgejahren werden. Eine solche einfache Fortsetzung einer Zeitreihe ist offenbar nicht vertretbar.

In Bezug auf KI sind somit grundlegend folgende Faktoren relevant, die im Weiteren erläutert werden:

1. Die Eingangsdaten der KI müssen eine hohe Qualität für den Anwendungsfall aufweisen.
2. Die IT-Anwendung und das KI-System sind von KI und Anwendungsexperten konzipiert sowie manipulationssicher und vertrauenswürdig umgesetzt.
3. Ergebnisse nachzuvollziehen wird ermöglicht.
4. Bei der Entwicklung und Anwendung werden jeweils ethische Grundsätze eingehalten.

Hier muss kritisch angemerkt werden, dass der Punkt 1) das „Zutaten-Argument“ anspricht, welches für ein Markenvertrauen offenbar sekundär ist. Der Punkt 2) adressiert das soziale Umfeld der System-Ontogenese. Der Aspekt bei Punkt 3) wird sich bei KI-Blackbox-Systemen kaum einhalten lassen.

Bevor innovative Technologien, im Sinne von Anwendungen und Diensten, eine breite Akzeptanz in der Gesellschaft allgemein sowie beim Nutzer im Speziellen erfahren, müssen bestimmte Rahmenbedingungen erfüllt sein. Ein Kriterium und ebenso ein entscheidender Erfolgsfaktor, über den ein hoher Grad an Zustimmung zu erreichen ist, wird hierbei zukünftig die Vertrauenswürdigkeit der Hersteller und Anbieter sein [CoPo20].

Ansatz von Troncoso

Im Beitrag von Troncoso [Tron22] wird der Aufbau von Markenvertrauen als ein wesentlicher Bestandteil jeder Marketingstrategie angesehen, da das Verbraucherverhalten vom Grad des Vertrauens beeinflusst wird. Die Positionierung eines Unternehmens als vertrauenswürdig wird daher als wichtig angesehen. Weiter wird das „Markenvertrauen“ als die Bereit-

schaft des Kunden angesehen, sich auf die Fähigkeit der Marke zu verlassen, die erklärte Funktion zu erfüllen. Kunden zögern, Marken zu akzeptieren, von denen sie nichts wissen.

Damit kann man bei [Tron22] mindestens drei Komponenten eines Systems sehen: Das sind das anbietende Unternehmer, die Kunden und das eigentliche Produkt. Das Unternehmen vermittelt Produkteigenschaften – und der Kunde glaubt, dass das Produkt die kommunizierten Eigenschaften erfüllt.

Als Grundlage für den Aufbau von Vertrauen sieht [Tron22] nicht nur das Vermeiden falscher Werbung, sondern eine Reihe von Maßnahmen, darunter:

- **Fokus auf Kundenerfahrung und Customer Experience.**
  Investitionen in positive Kundenerlebnisse zahlen sich aus. Kunden erwähnen positive Erfahrungen gegenüber Dritten. Leider gilt auch das Gegenteil: Kunden sind eher bereit, ihre negativen Erfahrungen weiterzugeben. Vor diesem Hintergrund müssen ihre Vertriebs- und Kundenserviceteams die Auswirkungen positiver Erfahrungen verstehen.

- **Klares Ziel und Leitbild.**
  Der Aufbau von Markenvertrauen besteht auch darin, den Zweck, die Werte und das Leitbild des Unternehmens genau zu definieren und dem Zielpublikum zu kommunizieren. Kunden akzeptieren Marken, die einen Wert repräsentieren, der mit den eigenen Werten übereinstimmt.

- **Verstehen, was Kunden brauchen.**
  Das bedeutet, per Marktforschung herauszufinden, was die Zielgruppe von der Marke erwartet, beziehungsweise was ihre Kunden in Zukunft erwarten werden. Das Markenvertrauen bezieht sich nicht nur auf einen (physischen oder digitalen) Artikel oder eine einmalige Dienstleistung. In der Zukunft werden möglicherweise Support, Reparaturen oder zusätzliche Artikel benötigt. Auch diese Lösungen müssen bedarfsgerecht mit angeboten werden können, um eine Marke als vertrauenswürdig und zuverlässig zu positionieren.

Der Aufbau eines Markenvertrauens wird bei Troncoso [Tron22] vor allem als ein Marketing-Problem der „ehrlichen“ Werbemaßnahmen und für den Kunden mit den „richtigen“ Inhalten verstanden. Die Maßnahmen berücksichtigen nicht im wünschenswerten Maß, dass Vertrauensnehmer (passive Rolle) immer „holistisch“ als Personen oder Pseudo-Personen (als Institutionen) positioniert sind.

Ansatz von LEAM

In der Machbarkeitsstudie „LEAM – Large European AI Models“ der Akademie für Künstliche Intelligenz (AKI) im KI Bundesverband werden vertrauenswürdige „Große KI-Modelle für Deutschland“ adressiert [LEAM23, pp. 45ff.]. Um KI-spezifische Risiken systematisch zu reduzieren werden in [LEAM23] sechs Dimensionen der Vertrauenswürdigkeit identifiziert:

- **Fairness**
  Zu vermeiden ist eine ungerechtfertigte Diskriminierung von Personen durch unaus-

gewogene, mit Bias behaftete Trainingsdaten oder die statistische Unterrepräsentation von Personengruppen. Es darf nicht zu einer Bevorzugung oder Benachteiligung von geschlechtsspezifischen oder ethnischen Gruppen kommen.

- **Autonomie und Kontrolle**
  Die Unterstützung von Menschen durch eine KI-Anwendung muss ausreichenden Handlungsspielraum in der Interaktion mit der KI-Anwendung bereithalten. Die spezifische Herausforderung liegt darin, dass die Möglichkeiten der Interaktion mit den Menschen meist erst im Design der konkreten nachgelagerten KI-Anwendung festgelegt werden können.

- **Transparenz**
  Unter Transparenz werden Aspekte der Nachvollziehbarkeit, Reproduzierbarkeit und Erklärbarkeit subsumiert. Die grundlegende Funktionsweise der KI-Anwendung muss angemessen nachvollziehbar sein und Ergebnisse der KI-Anwendung müssen reproduziert und begründet werden können.

- **Verlässlichkeit**
  Ein wesentlicher Aspekt ist hier die Faktentreue, Plausibilität und der Wahrheitsgehalt von Aussagen, die sich im Lauf der Zeit ändert. Ein weiteres Problem ist die absichtliche Generierung von „Fake News“ oder ähnlichen Artefakten. Dieser Bereich muss noch deutlich weiterentwickelt werden.

- **Sicherheit**
  Diese Dimension adressiert sowohl Eigenschaften der funktionalen Sicherheit als auch die Absicherung gegenüber Angriffen und Manipulationen der KI-Anwendung.

- **Datenschutz**
  Diese Dimension bezieht sich auf den Schutz sensibler Daten im Kontext von Entwicklung und Betrieb einer KI-Anwendung. Dabei wird sowohl der Schutz personenbezogener Daten als auch von Geschäftsgeheimnissen adressiert. Hier gibt es einen hohen Forschungsbedarf.

Eine Kritik an [LEAM23] muss benennen, dass hier „Vertrauen“ in IT-Systeme – hier KI-Systeme – auf die Eigenschaft(!) der Vertrauenswürdigkeit reduziert wird. Nach Maßgabe der obigen Ausführungen zum Begriff des „wahren Wissens“ als eine sozio-philosophische Kategorie dürfte die Darstellung von Verlässlichkeit von KI-Systemen wohl falsch eingeschätzt werden.

Richtig erkannt wird bei [LEAM23], dass es weiterer intensiver Forschung und Entwicklung bedarf, um die Anforderungen aus allen beschriebenen sechs Dimensionen der vertrauenswürdigen KI systematisch abzudecken. Neben der Komplexität stellt die Vielfalt möglicher Anwendungen eine besondere Herausforderung dar. Grundsätzlich sind geeignete organisatorische Maßnahmen zu ergreifen, um in Situationen, in denen z. B. ein mögliches Fehlverhalten eines Modells auftritt, reagieren zu können. Hierbei ist das Wechselspiel zwischen den beteiligten Organisationen zu berücksichtigen. Zu beachten ist, dass zur Behebung gefundene Fehler auch technische Maßnahmen, wie Modellverbesserungen adressiert werden.

Markenvertrauen und CBMs für IT-Systeme

Es sollen nun konstruktiv zwei Möglichkeiten für vertrauensähnliche Beziehungen zwischen Menschen und Maschinen erörtert werden. Es wird
- einerseits für den Aufbau von Markenvertrauen und
- andererseits für Confidence-Building Measures (CBMs)

jeweils dargelegt, wie diese für IT-Systeme realisiert werden könnten.

Markenvertrauen für IT-Systeme

Das bekannte Phänomen des Markenvertrauens klassischer Produkte lässt sich durchaus auf IT-Systeme übertragen. Aus einem bekannten Katalog von Standard-Markenprodukten [Lang03] lässt sich ableiten, dass unterschieden werden kann zwischen

- **Statischen Marken:**
  Bei Produkten dieser Marken erwartet der Käufer eine verlässliche Konstanz der Produkteigenschaften. Die Anbieter sollen das Produkt nicht ändern. Marken dieser Art finden sich insbesondere im Lebensmittelbereich – wie etwa bei der Würzsauce Maggi, oder beim Magenbitter Underberg, dessen Motto sogar „semper idem“ (lateinisch – Immer das Gleiche) ist

und

- **Dynamischen Marken:**
  Bei Produkten dieser Marken erwartet der Käufer – unbeschadet der Qualität – keine Konstanz der Produkteigenschaften, sondern eine fortlaufende Innovation, die dem Stand der Technik entspricht. Die Anbieter sollen die Produkte sehr wohl ändern, indem sie sie fortlaufend weiterentwickeln und verbessern. Marken dieser Art finden sich insbesondere im Technikbereich – wie bei der Firma 3M („Improving life through innovation“), die für die Marken „Post-it“ und „Scotch“ bekannt ist, oder auch bei der Firma Stihl („Pioniergeist seit mehr als 95 Jahren – immer wieder revolutionäre Innovationen entwickeln“), die motorbetriebene Geräte für die Forstwirtschaft fertigt.

Ein Markenvertrauen für IT-Systeme wird sich eher dem dynamischen Typ zuordnen lassen; der Kunde erwartet eine Fortentwicklung der Produkte, gleichwohl sollen die Innovationen den gewohnten und erwarteten Qualitätsstandard nicht verletzen.

Die Unterscheidung bei [RoWo15] zwischen
- Reflexiven Vertrauens – zwei Parteien vertrauen sich gegenseitig und
- Präreflexiven Vertrauens – eine Partei vertraut der anderen, ohne ein Gegenvertrauen

lässt sich beim IT-Markenvertrauen in ein präreflexives Vertrauen eingrenzen, da es kaum ein Vertrauensverhältnis zwischen Anbietern hin zum Kunden geben muss, beziehungsweise geben wird. Möglicherweise vertrauen gerade die beiden oben genannten Anbieter Underberg und Stihl darauf, dass die Kunden keinen groben Unfug mit den Produkten anstellen, was einerseits – mit jedem auf seine Art – durchaus möglich ist und andererseits den Ruf der Produkte schädigen würde. Essentiell ist dieser Aspekt für IT-Systeme kaum.

Bei [Hart11] und [RoWo15] werden weiter unterschieden:

- **Das Markenvertrauen des homo oeconomicus**
  als eine ökonomische Austauschbeziehung (Exchange Relationship) und rationales Kalkül. Dieses Markenvertrauen ist geprägt von pragmatisch-rationalen Kosten-Nutzwert-Abwägungen. Für jede Leistung der Beziehungspartner – Geld beziehungsweise Produkt – wird eine äquivalente Gegenleistung – Produkt beziehungsweise Geld – erwartet. Ein Eigeninteresse und das Streben nach materiell messbarem Gewinn stehen im Vordergrund, und

- **das Markenvertrauen des homo sociologicus**
  als eine Partner-Beziehung (Communal Relationship). Hier orientieren sich die Vertrauenspartner verstärkt am sozio-ökonomischen Wohlergehen des jeweils anderen. Eine Norm der gegenseitigen Verantwortung und Unterstützung steht im Vordergrund, ohne dass immer eine konkrete Gegenleistung erwartet wird.

Für die Akzeptanz von IT-Systemen dürfte anzunehmen sein, dass die Exchange Relationship im Vordergrund steht. Speziell wenn die aktive Akzeptanz eines IT-Produkts durch eine Firma oder Organisation erfolgen soll, dürften rational-ökonomische Argumentationen in der Belegschaft und Unternehmensführung im Vordergrund stehen. Die Communal Relationships sind sicher für die Akzeptanz förderlich, aber kaum die alleinige Basis einer ökonomischen Kaufentscheidung.

Oben wurde dargelegt, dass ein „Vertrauen" in maschinelle Systeme einer anthropomorphen Projektion entstammt: Die Systeme werden zunächst „vermenschlicht", um sie dann als passive Vertrauensnehmer einzuordnen. Die Existenz von „Vertrauen in (IT-) Systeme" hat in einem aufgeklärten Diskurs keinen Bestand.

Nichtsdestoweniger kann davon gesprochen werden, dass maschinelle (IT-) Systeme Gegenstand von Markenvertrauen sein können. In der Folge sollen daher (Marken-) vertrauensbildende Maßnahmen für IT-Systeme, wie CDS- und KI-Anwendungen dargelegt werden. Dabei können vier funktionale Kategorien für ein Markenvertrauen identifiziert werden:

I. Schädlich
II. Neutral
III. Förderlich
IV. Nötig

Konsequenterweise soll für diese vier Kategorien nun jeweils eine Unterliste angegeben werden; dabei ist versucht worden, die Unterkategorien nach ihrer relativen Wichtigkeit zu sortieren:

I Schädlich für ein Markenvertrauen

a Schädlich: Eine Anonymität der Anbieter, so dass ein sozialer Kontext des Produkts nicht fassbar ist. Dies tritt etwa auf, wenn ein Produkt irgendwo (in einem Drittstaat) hergestellt und semi-anonym per Fullfillment-Center importiert wurde. In der Konsequenz sind keine menschlichen Ansprechpartner für den Fall von Problemen verfügbar; niemand kann für eine Fehlfunktion oder Qualitätsmangel verantwortlich gemacht werden.

b Schädlich: Ein direkter Einfluss Dritter auf die Anbieter, so dass die bilaterale Bezie-

hung zwischen Anbieter und Kunden beeinflusst oder gestört wird. Die liegt etwa vor, wenn Anbieter eines IT-Systems auf Weisung (auswärtiger) staatlicher Stellen oder Einflussnehmer einen vom Kunden erwarteten Datenschutz unterlaufen.

c Schädlich: Das Verbreiten expliziter Falschinformation von Seiten des Anbieters über ein Produkt ist ebenso schädlich, wie das Verschweigen für die Kaufentscheidung und den Betrieb relevanter Informationen.

d Schädlich: Ein erratisches unkalkulierbares Verhalten des Anbieters, so dass Kunden nicht wissen können, warum welche Entscheidungen getroffen worden sind.

e Schädlich: Eine fehlende Disziplin, mangelnde Termintreue, offenbare Unordnung in der Datenhaltung, auch quasi „innovative" und eher sinnlose neue Vokabeln sind für ein Markenvertrauen schädlich.

II Neutral für ein Markenvertrauen

a Neutral: Obwohl im Bereich von IT-Systemen eine „Transparenz" als förderlich diskutiert wird, können sehr viele Produkte mit hohem Markenvertrauen identifiziert werden, bei denen eine Transparenz der Zutaten und der Herstellungsprozesse nicht weiter von Belang ist, oder gar geheim gehalten wird. Eine explizite Offenlegung der Algorithmen und der Trainingsdaten für ein KI-System scheint von daher überbewertet zu werden.

b Neutral: Der Preis für den Erwerb für ein Produkt ist eher belanglos, soweit keine explizit prohibitive Preisbildung zum Ansatz kommt.

III Förderlich für ein Markenvertrauen

a Förderlich: Der Anbieter zeigt durch eine entsprechende Kommunikation, dass das Produkt eine technische – auch innovative – Exzellenz repräsentiert. Der Grad der Innovation darf aber nicht vermitteln, dass das Produkt experimenteller Natur und unerprobt ist, so dass der Kunde ein unangemessenes Risiko eingeht.

b Förderlich: Einzelne Personen in der Anbieter-Organisation sollten austauschbar sein. Der Anbieter sollte nicht ausfallen, wenn einzelne Individuen in der Organisation nicht mehr verfügbar sind. Dies ist zum Beispiel durch entsprechende Nachfolge- und Vertretungsregelungen erreichbar.

c Förderlich: Eine geschlossene Kundengruppe ist nützlich. Dies ist beispielsweise der Fall, wenn nur berufsständische oder qualifizierte Kunden zugelassen werden.

d Förderlich: Für ein präreflexives Vertrauen ist die Darstellung der Historie der Zuverlässigkeit und der Akzeptanz der Produkte nützlich.

e Förderlich: Eine große Kundengruppe – eine sogenannte „kritische Masse" – und damit verbundene qualifizierte Peer-2-Peer-Akzeptanzsignale fördern Markenvertrauen und vermindern den Grenzaufwand bei der Kundengewinnung.

IV Nötig für ein Markenvertrauen

a Nötig: Die Anbieter zeigen, dass sie etwas zu verlieren haben, wenn sie als Vertrauensnehmer den vertrauensgebenden Kunden enttäuschen. Der Anbieter hat „einen Ruf zu verlieren", im Sinne eines verlorenen sozialen Kredits, der nur aufwändig wieder hergestellt werden kann.

b Nötig: Der Anbieter kann glaubhaft darstellen, dass er die Kundensituation verstanden hat. Der Anbieter weiß, wofür das Produkt beim Kunden verwendet wird und kann daher dessen Erwartungen bedienen.

c Nötig: Das Verstehen der Kundensituation wird durch eine Kette menschlicher Vertrauenspersonen vom Anbieter weitergegeben, bis es den Endkunden erreicht hat. Der Anbieter unterhält dafür einen Fachhandel und Kundenberater, die beim Kunden

eine hohe Produktkompetenz vertreten können.

d Nötig: Das Wohlergehen der Kunden ist im zentralen Interesse („Good Will“) des Anbieters. Wenn der Kunde zeigt, dass das Produkt – warum auch immer – versagt hat, wird sich der Anbieter um eine Lösung oder Reparatur im Interesse des Kunden bemühen.

e Nötig: Das Produkt hat eine definierte Qualität, auf die sich der Kunde verlassen kann.

f Nötig: Es gibt einen Einfluss der Kunden auf die Anbieter. Der Anbieter organisiert einen Offenen Diskurs mit den Kunden, so dass eine Verbesserung und Selbstkorrektur der Produkte systematisch erzielt werden kann.

g Nötig: Die Transaktionen des Kunden mit dem Anbieter gehen über das Bedienen eines Automaten hinaus. Der Anbieter ist zu Nicht-normative Korrekturen, wie dem Gewähren einer Kulanz, bereit.

Diese Komponenten sind nicht als eine abschließende Aufzählung zu verstehen, vielmehr sind sie eine synoptische Darstellung auf einer phänomenologischen Basis.

Confidence-Building Measures (CBMs) für Maschinen

Dieser Abschnitt folgt weitgehend den Ausführungen von [Hofm19 – pp. 73ff.]. Die Programmierung der Mensch-Maschine-Interaktion definiert formal, was ein Mensch quasi „zu tun hat“, wenn er mit der Maschine interagiert: Maschinen, Automaten und formale Prozesse bewegen Menschen zu einem bestimmten Handeln.

Der Begriff „Automat“ bedeutet „sich selbst bewegend“, als selbsttätiges Handeln einer Maschine. Die griechische Antike kannte die Αυτοματια (Automatia) als die Göttin der ohne menschliches Zutun eintretenden glücklichen Ereignisse. Aber schon früh wurde erkannt, dass Automaten auch eine Kehrseite haben und gefährlich werden können. Die ersten Automaten waren wohl die prähistorischen Beute-Fallgruben. Bereits in den biblischen Sprüchen (Kapitel 26, Vers 27) findet sich das Sprichwort „Wer eine Grube gräbt, fällt selbst hinein“. Selbst technisch primitive Automaten können gefährlich sein, wenn ihr Agieren nicht dem intendierten Modell der Nutzung entspricht. Seit jeher sind Vorkehrungen erforderlich, um Unfälle mit Automaten zu vermeiden.

Der Fortschritt in der KI wird als zunehmender Konkurrenzkampf zwischen Mensch und Maschine betrachtet. Dies gilt speziell für die künstlichen neuronalen Netze (Artificial Neural Networks – ANN); sie haben ein biologisches Vorbild in den natürlichen neuronalen Netzen, den Nervenzellvernetzungen der Gehirne diverser Lebewesen. Ein ANN-System kann die Struktur von Lern-Beispieldaten adaptieren und diese nach Maßgabe der – von Menschen organisierten – Programmierung verallgemeinern. Das ANN-System kann, von Beispielen abstrahierend, gewisse Gesetzmäßigkeiten quasi „erkennen“. Den verwendeten ANN-Systemen wird ein Verhalten antrainiert; es ist jedoch kein Einblick in den erlernten algorithmischen Lösungsweg möglich. Das Verhalten dieser sogenannten „impliziten“ ANN-Algorithmen ist nicht unbedingt einwandfrei vorhersehbar. Die Lernbeispiele für das Training hat man nicht immer unter voller Kontrolle. Bekannt wurde im Jahr 2015 ein ANN-Programm, das dunkelhäutige Menschen als Gorillas bezeichnete – was schon arg bedenklich ist [https://www.spiegel.de/netzwelt/web/google-fotos-bezeichnet-schwarze-als-gorillas-a-1041693.html]. Man könnte meinen, das Programm habe eine „schlechte Erziehung“ genossen.

Die Normen der KI-Systeme verlangen nach anthropozentrischer Ergänzung. Wenn Entscheidungen von Computern nicht sinnhaft sind, dann müssen diese erkannt werden können. Es müssen Mechanismen – Confidence-Building Measures (CBMs) – für Maschinen verfügbar sein.

Zu diesen Aspekt wird ein Beispiel betrachtet: Personenkraftwagen erfahren seit Jahren einen zunehmenden Grad an Automatisierung. Eine automatische Überwachung von Werkstatt-Servicezyklen entlastet den Fahrer, hat aber Nachteile, wenn der Bordcomputer verhindert, dass sich der Wagen „zu Ihrer eigenen Sicherheit" oder „zu Ihrem eigenen Vorteil" in Betrieb nehmen lässt. Dies kann sogar gefährlich werden, wenn es um eine Fahrt in einem echten Notfall geht. Das Beispiel zeigt aber auch, dass sich die Nicht-Sinnhaftigkeit einer Automatisierung jeweils kaum vorhersagen lässt. Diese Nicht-Sinnhaftigkeit – oder Fehlerhaftigkeit – hat sich erst später ergeben, typischerweise als Defizite in der Modellierung, die der Anwendungssituation zugrunde liegt. Es dürfte kaum einen Ansatz geben, mit dem sich diese Defizite sicher und im Voraus erkennen lassen.

Einen Ansatz bietet hier der Kritische Rationalismus [Popp96]: Es kann nicht entscheidend sein, wie man im Voraus eine sinnvolle und fehlerfreie Automatisierung konstruiert. In den Vordergrund ist stattdessen die Frage zu rücken, wie schlechte und fehlerhafte Automaten erkannt („falsifiziert") und quasi „gebessert" werden können. Confidence-Building Measures (CBMs), mittels derer man nicht-sinnhafter Automatisierung ausweicht und entkommen kann, sind:

- **OFF** – das kontrollierte Abschalten von Automaten und dadurch die Möglichkeit, einen Prozess manuell anhalten und abschalten zu können. Diese CBM ist beispielsweise als Notausschalter an konventionellen Maschinen und Anlagen bekannt. Ein derartiger „Kill Switch" dient dazu, im Gefahrfall die Anlage rasch in einen sicheren Zustand zu bringen, und

- **ESC** – das Ausweichen vor den normativen Vorgaben eines Automaten und das Wiedererlangen der Kontrolle durch manuell und sinnhaft – die Situation verstehende – handelnde Menschen. Die CBM des ESC greift bereits bei harmlosen manuellen Korrekturen, die etwa nötig sein können, weil die Autokorrektur eines Textverarbeitungsprogramms sinnlose „Korrekturen" im Text vornimmt, die manuell revidiert werden müssen. Die CBM ist entscheidend bei der manuellen Übersteuerung von Autopiloten in Fahrzeugen und Flugzeugen. Eine sinnvoll gestaltete ESC-Funktion hätte die notorischen Unglücke der Boeing-Flugzeuge vom Typ 737 durchaus verhindern können.

Vor dem Hintergrund einer Digitalen Ethik sollte eine Inbetriebnahme von (neuen) Automaten ohne eine systematische Planung dieser beiden CBM-Optionen „OFF" und „ESC" eigentlich nicht möglich sein dürfen. Eine nicht-sinnhafte Automatisierung muss korrigierbar sein, den Menschen muss eine letztliche Handlungsautonomie bleiben. Das ist der Sinn von Vertrauensersatz-Maßnahmen (CBMs) für KI-, CDS- und andere IT-Systeme.

Fazit und offene Fragen

Es gibt kein einheitliches Verständnis des Phänomens Vertrauen. Vertrauen ist zwar im sozialen Kontext und der Lebenswirklichkeit allgegenwärtig, aber doch analytisch nicht so erfasst, wie dies wünschenswert sein könnte. Es kommt im Diskurs zu Vermischungen des Vertrauens (als einer Beziehung) mit Zuverlässigkeit und Vertrauenswürdigkeit (als Eigenschaften). Dieser Umstand hat eine gewisse Ähnlichkeit mit der Alltäglichkeit und dem analytischen Verständnis der Gravitation vor der Moderne, als die „Schwere" eines Körpers nicht als Auswirkung eines (Kraft-) Feldes, sondern als eine Eigenschaft(!) der massereichen Körper selbst galt.

Vertrauen erscheint als eine Beziehung(!) zwischen Vertrauensgebern und Vertrauensnehmern – wenn diese Rollen von beiden Parteien simultan eingenommen werden, dann ist die Vertrauensbeziehung symmetrisch. Vertrauensgeber können nur Menschen sein, während als Vertrauensnehmer auch Institutionen auftreten können. Eine Vertrauenswürdigkeit ist eine Eigenschaft von Menschen oder Institutionen, die einer Vertrauensbeziehung förderlich sein kann, aber dafür weder hinreichend noch notwendig ist.

Das institutionelle Markenvertrauen und Vertrauensersatz-Maßnahmen (CBMs) aufgreifend können Analogieschlüsse angestellt werden, um Komponenten von prospektivem Markenvertrauen und von Zero-Trust-Transactions zu identifizieren. Gesamt-Szenarien vertrauenswürdiger IT-Anwendungen mit ihrem sozio-ökonomischen Kontext erscheinen so darstellbar.

Als Gegenstand offener Fragen lassen sich einige Verständniskontexte des Vertrauens identifizieren. Als Gegenstände künftiger Arbeiten können identifiziert werden:

1. Eine sprachphilosophische Betrachtung des Begriffs „Vertrauen", speziell im englischen Sprachraum. Die Analogien zwischen deutsch-englischen Vokabelmengen sollten adressiert werden, wie:
   a) {Vertrauen, Zutrauen, Zuverlässigkeit, Sicherheit, Erwartung, Berechenbarkeit} einerseits, und
   b) {Trust, Confidence, Hope, Security, Safety, Reliability} andererseits.

2. Eine systematische Analyse von Vertrauensmarken. Die hier bereits gefundenen Faktoren des Markenvertrauens – von I. bis IV. – sollten einer präziseren Gewichtung unterzogen werden.

3. Eine systematische Analyse von CBMs für Automaten. Ansätze aus anderen Metiers, wie etwa dem Maschinen- und Anlagenbau sollten auf eine Transferierbarkeit untersucht werden.

4. Die Frage nach Vertrauenstransfer von Herstellern zum Kunden über Instrumente wie der Kundenbetreuung oder den qualifizierten Fachhandel.

Es sollte gelingen, künftig genauere „Vertrauensmodelle" und die Strukturen von „Vertrauensplattformen" daraus abzuleiten.

Im Rahmen eines Erwartungsmanagements muss allerdings relativierend festgehalten werden: Die Versuche, die vertrauensvolle Kommunikation von Unternehmen mit der Kund-

schaft komplett auf „vollautomatische Systeme“ zu übertragen und zu delegieren, scheiterten stets. Am Ende musste man wieder auf einen humanen persönlichen und verständnisvollen Ansprechpartner als „Back-up“ und Quelle des Vertrauens zurückkehren. Vertrauen ist nicht technisch darstellbar, es basiert immer auf einer menschlichen und inter-subjektiven Kommunikation. Diese Erkenntnis wurde in der Entwicklung der Informationsgesellschaft etwa um die Jahrtausendwende direkt evident [Hof22a – pp. 213ff.].

Abschließend sei betont, dass Formale Systeme und Prozesse – wie IT-Systeme – immer revidierbar sein müssen. Es ist nicht vertretbar, dass ein formaler Prozess quasi „super-sicher“ ist und nicht mehr durch humane Intervention korrigiert werden kann. Das wäre das Ende jeder humanen Verantwortungsethik – sie würde durch das reine Feststellen von Systemversagen abgelöst. Der bewährte anthropozentrische Orientierungspunkt der individuellen Freiheit darf nicht gegen in Aussicht gestellte Nutzwerte Digitaler Systeme eingetauscht werden.

Literaturverzeichnis

[Baur10] Baurmann, Michael: „Kollektives Wissen und epistemisches Vertrauen“, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 50/2010, Köln, 2010

[Bowl01] Bowlby, John: „Frühe Bindung und kindliche Entwicklung (Child Care and the Growth of Love)". Ernst Reinhardt Verlag, München 2001

[CBMs23] Wikipedia-Artikel: „Confidence-building measures“, Januar 2023 https://en.wikipedia.org/wiki/Confidence-building_measures

[Clae93] Claessens, Dieter: „Das Konkrete und das Abstrakte: Soziologische Skizzen zur Anthropologie“ suhrkamp taschenbuch wissenschaft, Frankfurt am Main, 1993

[CoPo20] Coester, Ulla und Pohlmann, Norbert: „Wie können wir der KI vertrauen? – Mechanismus für gute Ergebnisse“, IT & Production – Zeitschrift für erfolgreiche Produktion, Technik-Dokumentations-Verlag, 2020/21 https://norbert-pohlmann.com/artikel/wie-koennen-wir-der-ki-vertrauen/

[Duch23] Wikipedia-Artikel: „Guillaume Duchenne de Boulogne“, Januar 2023 https://en.wikipedia.org/wiki/Guillaume_Duchenne_de_Boulogne

[eco21] eco netTALK „Potenzial von SSI & Blockchain“, 14. Juni 2021 https://www.eco.de/event/eco-nettalk-potenzial-von-ssi-blockchain/

[FiDa20] Fischer, Hans-Peter und Damm, Frank: „Cyber Security: Vertrauen durch Sicherheit für eine digitale Welt“, managerwissen, Manager Magazin, 2020 https://manager-wissen.com/cyber-security-vertrauen-durch-sicherheit-fuer-eine-digitale-welt

[Hart11] Hartmann, Martin: „Die Praxis des Vertrauens“, Suhrkamp, Frankfurt am Main, 2011

[Heis69] Heisenberg, Werner: „Der Teil und das Ganze“, Piper, München, 1969

[Hoff21] Hoffmann, Rainer: „Trustworthy Artificial Intelligence – Vertrauen in Künstliche Intelligenz“, Stiftung Energie & Klimaschutz, Karlsruhe, 2021 https://www.energie-klimaschutz.de/vertrauen-in-kuenstliche-intelligenz/

[Hofm19] Hofmann, Georg Rainer: Das Weinberg-Paradoxon. Warum sich das gute Handeln nicht vollständig durch Gesetze und Gebote regeln lässt – ein Essay über nicht-normative Ethik. wbg, Darmstadt, 2019

[Hofm21] Hofmann, Georg Rainer: „Was ist und zu welchem Zweck braucht man eine „Sichere Identität“? Grundlagen, Akzeptanzfragen und Gestaltungsaspekte des Identitätsmanagements“, IMI-Verlag, Aschaffenburg, 2021

[Hof22a] Hofmann, Georg Rainer: „Globale Provinz. Entdeckung und Besiedlung der digitalen Welt 1980 bis 2020“, Vergangenheitsverlag, Berlin, 2022

[Hof22b] Hofmann, Georg Rainer „Wissen wir, wie lange Krokodile leben? – Do we know how long crocodiles live?“, Festvortrag Corps Hannovera, Hannover, 18. November 2022
https://www.imi.bayern/publikationen/2022/

[HoSc14] Hofmann, Georg Rainer und Schumacher, Meike: „Studie zur Akzeptanz von Cloud Computing“, EuroCloud Deutschland_eco e.V., EuroCloud Austria, 2012, Köln, Wien, 2. Aufl. 2014

[HoSc16] Hofmann, Georg Rainer und Schumacher, Meike: „Akzeptanzfaktoren des E-Invoicing (‚Elektronische Rechnung‘)“, eco – Verband der Internetwirtschaft e.V., Köln, 2016

[Hube14] Hubert, Martin: „Menschliches Denken ist an Gemeinsamkeit orientiert“, Deutschlandfunk, 2014
https://www.deutschlandfunk.de/evolution-des-geistes-menschliches-denken-ist-an-100.html

[Jime11] Jiménez, Fanny: „Warum wir Gesichter blitzschnell erkennen können“, Welt digital, 10. Dez. 2011
https://www.welt.de/wissenschaft/article13759042/Warum-wir-Gesichter-blitzschnell-erkennen-koennen.html

[KPMG21] KPMG: „Bürgerinnen und Bürger haben wenig Vertrauen in künstliche Intelligenz (Trust in Artificial Intelligence. A five country study)“, KPMG, März 2021
https://home.kpmg/de/de/home/themen/2021/05/studie-buergerinnen-und-buerger-haben-wenig-vertrauen-in-ki.html

[Lang03] Langenscheidt, Florian (Hrsg.): „Deutsche Standards. Marken des Jahrhunderts“, 14. Auflage, Köln, 2003

[LEAM23] Akademie für Künstliche Intelligenz (AKI): „Große KI-Modelle für Deutschland – Machbarkeitsstudie LEAM – Large European AI Models“, KI Bundesverband, Berlin 2023

[Lisch13] Lischka, Konrad: „Dinge mit Gesicht: Die Welt steckt voller Lächeln“, Hoffmann und Campe, 2013
eine Variante auch online unter https://dingemitgesicht.de/

[Luhm68] Luhmann, Niklas: „Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität“ (1968), 4. Auflage, Lucius & Lucius, Stattgart, 2000

[Mara22] Mara, Martina: „Der Umgang mit KI erfordert informiertes Vertrauen“, zukunftsinstitut, Frankfurt 2022
https://www.zukunftsinstitut.de/artikel/der-umgang-mit-ki-erfordert-informiertes-vertrauen/

[Nort92] North, Douglass: "Institutionen, Institutioneller Wandel und Wirtschaftsleistung (Institutions, Institutional Change, and Economic Performance)", Mohr, Tübingen 1992

[Nune16] Nunez Mariano: „Vertrauensbildende Maßnahmen sind gefragt", Security-Insider, 2016
https://www.security-insider.de/vertrauensbildende-massnahmen-sind-gefragt-a-544156/

[Popp96] Popper, Karl Raimund: „Alles Leben ist Problemlösen: Über Erkenntnis, Geschichte und Politik", 8. Aufl., Piper, München, 1996

[RoWo15] Rommerskirchen, Jan und Woll, Anne-Kathrin: „Normative Erwartungen und internalisierte Werte-Marken als ethische Konstrukte",. Journal of Business and Media Psychology, 6(1), 10-25, 2015
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-61944-1

[ScHo16] Schumacher, Meike und Hofmann, Georg Rainer: „Case-based Evidence – Grundlagen und Anwendung: Prognose und Verbesserung der Akzeptanz von Produkten und Projekten", Springer Vieweg, 2016

[Schr85] Schrödinger, Erwin: „Mein Leben, meine Weltansicht", Verlag Zsolnay, Wien, 1985

[Toma10] Tomasello, Michael: „Warum wir kooperieren", 4. Auflage, edition unseld, Suhrkamp, Frankfurt am Main, 2010

[Tron22] Troncoso, Darya Jandossova: „Aufbau Von Markenvertrauen", marketsplash.com, 2022
https://marketsplash.com/de/markenvertrauen/

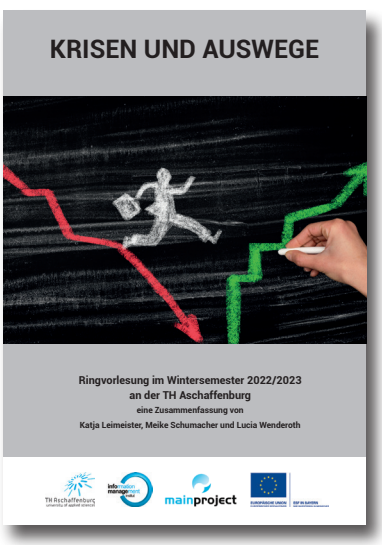
[Zime92] Zimen, Erik: „Der Hund: Abstammung – Verhalten – Mensch und Hund", Goldmann Verlag, 1992

[Ziea13] Ziske, Christine et al.: „Vertrauen und IT-Sicherheit. Vertrauensmodelle für die Informationsgesellschaft", TeleTrusT – Bundesverband IT-Sicherheit e.V., Berlin, 2013

Weitere Publikationen aus dem IMI-Verlag

Titel: Was ist und zu welchem Zweck braucht man eine „Sichere Identität"?
Autor: Georg Rainer Hofmann
Verlag: IMI Verlag, Aschaffenburg, 11/2021
ISBN: 978-3-9823413-0-9

Inhalt:
Bei (Digitalen) Geschäfts- und Verwaltungsprozessen können Betrug und anderes ungesetzliches Verhalten durch die „sichere Identität" der Beteiligten, Bürger, Kunden und Geschäftspartner erschwert werden. Bei der „Sicheren Identität" spielen die „Vertrauensvolle Identität" und die „Zuverlässige Identität" eine wichtige Rolle. Prof. Dr. Georg Rainer Hofmann geht in dieser Veröffentlichung unter anderem auf folgende Punkte ein: Symbole der Identität, Ontogenese und Fälschungssicherheit der Identitätsdokumente, Selbst-Souveräne Identitäten SSIs u.v.m.



Titel: Krisen und Auswege
Untertitel: Ringvorlesung im Wintersemester 2022/2023 an der TH Aschaffenburg
Autoren: Katja Leimeister, Meike Schumacher und Lucia Wenderoth
Verlag: IMI Verlag, Aschaffenburg, 02/2023
ISBN:978-3-9823413-2-3

Inhalt:
Krisen gab es schon immer. Aktuell hat sich jedoch ein ganzes Spektrum an Krisen entfaltet – es gibt politische und militärische Krisen, es brechen Energiemärke und Lieferketten weg, wir sehen ganze Branchen gefährdet, auch die Gesundheit macht uns Sorgen, vom Klimawandel ganz zu schweigen.
In den Veranstaltungen der Ringvorlesung „Krisen und Auswege", die im Wintersemester 2022/2023 stattfand, wurden Phänomene, Konzepte und Lösungen aus vielfältiger Perspektive erschlossen und damit die Komplexität des Gegenstands aufgezeigt. Herausforderungen für Unternehmen und Privatpersonen wurden identifiziert sowie der Einfluss auf technologische und organisatorische Entwicklungswege dargelegt.

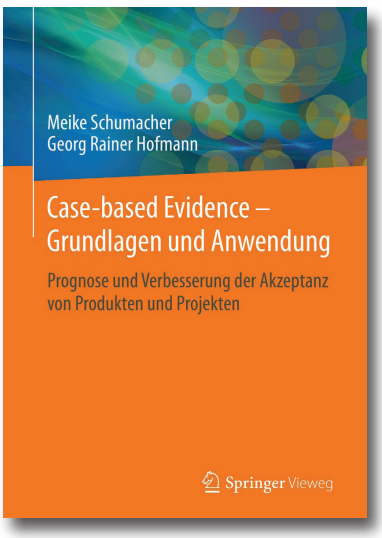
In der Publikation „Krisen und Auswege" wurden die einzelnen Vorträge der Ringvorlesung zusammengefasst.

Weitere Publikationen des Autors

Titel: Globale Provinz
Untertitel: Entdeckung und Besiedlung der digitalen Welt 1980 bis 2020
Autor: Georg Rainer Hofmann
Verlag: Vergangenheitsverlag, Berlin, 11/2021
ISBN: 978-3-86408-277-1

Inhalt:
Dieser Bericht basiert auf den Erinnerungen eines Vertreters der „Generation X", der erlebt hat, wie in circa 40 Jahren das private, berufliche und öffentliche Leben vom Gebrauch des sogenannten »Netzes« und der Mensch-Maschine-Symbiose immer mehr, sowohl positiv als auch negativ, geprägt wurde. Georg Rainer Hofmann, einer der Aktivisten dieser Entwicklung in Deutschland, zeichnet in einem komplexen Bild die technischen, ökomischen, sozio-politischen, und auch ethisch-philosophischen Hintergründe und Herausforderungen.

Mit einem Geleitwort von Alexander Rabe und mit Beiträgen von: Wolfgang Alm, Bernd Becker, Christof Blum, Ralf Cordes, Peter Egloff, José Luis Encarnação, Lucia Falkenberg, Andreas Hufgard, Andreas Kindt, Richard Knapp, Günter Koch, Detlef Krömker, Guerino Mazzola, Radu Popescu-Zeletin, Olaf Reubold, Gerd Rossbach, Gerd Schürmann, Hans-Georg Stark, Harald Summa, Rainer Thome, Thomas Wolf, Ruben Zimmermann.



Titel: Case-based Evidence – Grundlagen und Anwendung
Untertitel: Prognose und Verbesserung der Akzeptanz von Systemen und Projekten
Autoren: Meike Schumacher, Georg Rainer Hofmann
Verlag: Springer Vieweg, Wiesbaden, 03/2016
ISBN: 978-3-658-10612-6

Inhalt:
Das Praxisbuch erläutert anschaulich anhand konkreter Fälle, wie Analogien für die Prognose und Erhöhung der Akzeptanz neuer Produkte und Projekte genutzt werden können. Der Leser erhält eine „Schritt für Schritt Anleitung" zur Methodenanwendung mit zahlreichen Hinweisen und Best-Practice Beispielen.

Die Autoren erläutern allgemein verständlich die Teilschritte der Methode: Zunächst werden die kritischen Phänomene identifiziert. Analogieschlüsse ausgewählter Vergleichsfälle werden auf den aktuellen Fall übertragen. Diese isomorphen Mechanismen bilden die Grundlage eines synoptischen Modells, welches schließlich in einer Serie qualifizierter Experteninterviews evaluiert wird.

Information Management Institut (IMI)

Technische Hochschule Aschaffenburg
Würzburger Straße 45
63743 Aschaffenburg

www.imi.bayern
www.mainproject.eu

ISBN 978-3-9823413-7-8