Vertrauen in KI-Anbieter: Risikonarrative und das „Privacy-Paradox“ im Datenschutz

Esther von der Weiden

Zusammenfassung

Der Beitrag untersucht das Narrativ, dass Anbietern von künstlicher Intelligenz (KI) nicht zu trauen sei. Zunächst wird Vertrauen theoretisch auf zwei Ebenen differenziert: das organisationale Vertrauen in die Anbieter und das Vertrauen in die technischen Systeme selbst. Studien zeigen, dass beide Ebenen unterschiedlich bewertet werden und sich gegenseitig beeinflussen.

Mehrere Studien belegen, dass erhebliche Teile der Bevölkerung – besonders in westlichen Ländern wie Deutschland – gegenüber KI-Anbietern skeptisch eingestellt sind. Gleichzeitig wächst die tatsächliche Nutzung von KI-Systemen wie ChatGPT rasant, wobei häufig sensible Daten geteilt werden.

Diese Diskrepanz zwischen geäußerten Bedenken und tatsächlichem Verhalten erklärt der Beitrag mithilfe des „Privacy-Paradox“: Menschen handeln nicht allein nach ihren Einstellungen, sondern wägen Nutzen und Risiken situativ ab, unterliegen kognitiven Verzerrungen und sozialem Einfluss. Das „Paradox“ ist damit weniger ein echter Widerspruch als eine erwartbare Lücke zwischen Einstellung und Verhalten.

Vertrauen ist kein binäres Phänomen, sondern kontextabhängig ausgehandelt. Die Hauptverantwortung für die Vertrauensbildung liegt nicht beim Einzelnen, sondern bei den Anbietern und staatlichen Institutionen, die durch Transparenz, unabhängige Prüfungen, klare Governance und wirksame Regulierung verlässliche Rahmenbedingungen schaffen müssen.

Einleitung

Spätestens seit der Veröffentlichung von ChatGPT durch OpenAI im November 2022 sind künstliche Intelligenz (KI-) Systeme in den Fokus der breiten Öffentlichkeit gerückt und haben eine weite Verbreitung erfahren. Mittlerweile existiert eine Vielzahl an Anbietern von KI-Systemen – etwa

OpenAI, Google, Replit oder Anthropic – und KI wird zunehmend in die unterschiedlichsten Bereiche wie z. B. Softwareentwicklung, Medizin, Kundenbetreuung oder kreative Arbeit integriert. Unternehmen setzen KI-Systeme ein, um Prozesse zu optimieren, Ärzte nutzen sie zur Unterstützung bei Diagnosen, und Privatpersonen greifen auf sie zurück, um alltägliche Aufgaben zu erleichtern.

Gleichzeitig wächst jedoch auch die Zahl kritischer Stimmen. Es entwickelt sich das Narrativ, dass Anbietern von KI-Systemen nicht zu trauen sei. Insbesondere wird befürchtet, dass übertragene Daten entgegen gemachter Zusagen oder Zusicherungen zum Training der KI-Systeme verwendet und anschließend so auch in den Ausgaben für Dritte offengelegt werden. Diese Befürchtung einer übergriffigen Aneignung speist sich auch aus Narrativen über andere Dienste oder Erfahrungen mit diesen. Bedenken hinsichtlich mangelnder Transparenz, fehlender Kontrolle über die eigenen Daten und unklarer Geschäftsmodelle nähren dieses Narrativ zusätzlich.

Der Datenschutz ist in der Europäischen Union (EU) als Grundrecht verankert, wobei die Datenschutz-Grundverordnung (DSGVO) den konkreten Umgang mit personenbezogenen Daten regelt. Die EU-KI-Verordnung erweitert die DSGVO im Hinblick auf den Einsatz von KI-Systemen. Zum jetzigen Zeitpunkt ist sie allerdings noch nicht vollständig in Kraft getreten. Trotz dieser hohen Regulierung entwickelt sich dieses Narrativ und spiegelt damit unter Umständen die tatsächliche Einstellung von Teilen der Bevölkerung wider. Interessant ist daher die Frage, welche Unterschiede sich zwischen verschiedenen Regionen und Ländern feststellen lassen, die über unterschiedliche Regulierungsrahmen verfügen. Während dieser Beitrag nicht spezifisch auf die Regulierungen anderer Länder eingehen wird, werden Unterschiede in der Einstellung der Bevölkerung gegenüber KI-Systemen herausgestellt.

Dieser Beitrag ist wie folgt gegliedert. Zuerst wird beleuchtet, was Vertrauen überhaupt ist und in welchen Rahmen es definiert ist. Anschließend werden Studien analysiert, um zu verstehen, welche Faktoren das Vertrauen in KI-Systeme und KI-Anbieter beeinflussen. Danach wird das „Privacy-Paradox“ erklärt, um eine Grundlage zu haben, um das Nutzungsverhalten solcher KI-Systeme, welches in dem folgenden Abschnitt betrachtet wird, einzuordnen. Der Beitrag schließt mit einem Fazit zum Narrativ und der Verortung von Verantwortlichkeiten bezüglich der Vertrauensbildung.

1. Was ist Vertrauen?

Um das Risikonarrativ „KI-Anbietern kann man nicht trauen" angemessen zu analysieren, ist zunächst zu klären, was unter Vertrauen verstanden wird. In der Literatur wird zwischen interpersonalen Formen des Vertrauens gegenüber konkreten Personen und institutionellen Formen des Vertrauens gegenüber abstrakten sozialen Systemen und Organisationen unterschieden. Luhmann (1973) betont, dass Vertrauen soziale Komplexität reduziert und Handlungsfähigkeit unter Unsicherheit ermöglicht. Im Kontext von KI-Anbietern ist vor allem das organisationale bzw. institutionelle Vertrauen relevant, da Nutzende typischerweise mit einem komplexen sozio-technischen Arrangement interagieren, dessen Verlässlichkeit durch organisatorische Praktiken und Regulierung abgesichert wird. Aktuelle Forschung zu Vertrauen in KI betont, dass Vertrauen hier überwiegend institutionell und systemisch vermittelt ist und weniger auf interpersonalen Beziehungen beruht (Afroogh u. a. 2024; Stilgoe 2023).

Im organisationsbezogenen Verständnis ist Vertrauen die Bereitschaft, sich aufgrund positiver Erwartungen hinsichtlich der Absichten oder des Verhaltens einer anderen Partei verwundbar zu machen (Mayer u. a. 1995). Dieses Verständnis umfasst üblicherweise die Dimensionen Kompetenz, Wohlwollen und Integrität (Mayer u. a. 1995). Diese Triade ist auf Personen und Organisationen zugeschnitten. Beziehen sich Urteile auf technische Artefakte, stehen dagegen Merkmale wie Zuverlässigkeit, Robustheit, Vorhersagbarkeit, Sicherheit und Erklärbarkeit im Vordergrund. Normativ lässt sich argumentieren, dass Technologie ohne Intentionalität kein „Wohlwollen" zeigen kann und Technologie damit nicht vertraut werden kann (Stilgoe 2023; Prognos AG und Z_punkt); daher sprechen wir bei Artefakten im strengen Sinne eher von Verlässlichkeit oder Konfidenzerwartungen. Deskriptiv sprechen Nutzende dennoch von „Vertrauen in Systeme"; gemeint ist damit meist die Erwartung verlässlicher Leistung unter Unsicherheit, vermittelt durch Standards, Zertifizierungen und die verantwortlichen Organisationen (Lee und See 2004; Hoff und Bashir 2015).

Überträgt man diese Unterscheidung auf KI und KI-Anbieter, lassen sich drei Ebenen differenzieren: die allgemeine Idee moderner KI, konkrete KI-Systeme mit ihren wahrgenommenen Leistungs- und Zuverlässigkeitseigenschaften sowie die Organisationen, die diese Systeme entwickeln, implementieren und betreiben. Neuere Forschung differenziert explizit zwischen Vertrauen in technische Systeme und Vertrauen in die Organisationen, die diese Systeme entwickeln und einsetzen (Lankton u. a. 2015). So zeigen

empirische Studien, dass Nutzende eigenständige Urteile gegenüber technischen Systemen, insbesondere KI-Systemen, bilden (Glikson und Woolley 2020; Bach u. a. 2024). Vertrauen in Technologie bezieht sich dabei auf wahrgenommene Zuverlässigkeit, Vorhersagbarkeit und Leistungsfähigkeit eines Systems, während Vertrauen in Organisationen stärker an institutionelle Strukturen, Verantwortlichkeit und Regulierungen gebunden ist.

Empirische Arbeiten wie die von Yao u. a. (Yao u. a. 2025) beschreiben in einem mehrstufigen Vertrauensübertragungsmodell, dass sich Vertrauen zwischen diesen Ebenen übertragen kann. Im Gesundheitswesen, wo sich die Vertrauensebenen auf das KI-System, das interpersonale Vertrauen in die behandelnden Ärztinnen und Ärzte sowie das institutionelle Vertrauen in das Krankenhaus verteilen, zeigt sich, dass Vertrauen in ein KI-System über die behandelnden Fachpersonen als Zwischenschicht die Einstellung gegenüber der Institution beeinflussen kann. Solche Transfers sind kontextabhängig und nicht zwangsläufig kausal in nur eine Richtung.

Für die Untersuchung des Narrativs „KI-Anbietern kann man nicht trauen" ist daher zentral, die Ebenen und Dimensionen zu trennen. Misstrauen kann auf artefaktbezogene Ursachen (mangelnde Zuverlässigkeit, fehlende Erklärbarkeit, Sicherheitsprobleme) oder auf organisationale Ursachen (wahrgenommene fehlende Integrität, unklare Verantwortlichkeit, unzureichende Transparenz, mangelnde externe Auditierung) zurückgehen. Eine differenzierte Betrachtung macht sichtbar, an welchen Punkten Misstrauen entsteht und wie es begründet wird. Sie zeigt zugleich, welche Gegenmaßnahmen Vertrauen begründen können, etwa durch klare Governance, Haftungs- und Beschwerdemechanismen, unabhängige Prüfungen, Regulierungs-Compliance und evidenzbasierte Leistungsnachweise.

2. Faktoren für das Vertrauen

Ziel dieses Beitrags ist es, die verschiedenen Ebenen des Vertrauens gegenüber KI-Anbietern empirisch zu beleuchten: einerseits das organisationale Vertrauen in die anbietenden Unternehmen und Institutionen (Dimensionen u. a. Kompetenz, Integrität), andererseits das Vertrauen in bzw. angemessene Verlassen auf die technischen Systeme selbst (Merkmale u. a. Zuverlässigkeit, Vorhersagbarkeit, Sicherheit, Erklärbarkeit). Diese Einordnung folgt etablierten Modellen des Organisationsvertrauens und der Forschung zum Vertrauen in Automatisierung (Mayer, Davis, und Schoorman, 1995; Lee und See, 2004; Hoff und Bashir, 2015). Da die Studien

unterschiedliche Frageformulierungen und Messansätze nutzen, werden Prozentwerte kontextuell interpretiert und nicht direkt verglichen; mehrere Erhebungen sind industrienah und werden daher mit anderen Quellen ergänzt.

2.1 Organisationales Vertrauen in Anbieter und den Umgang mit Daten

Die Bitkom-Studie (Bitkom 2023) zeigt, dass deutsche Konsumierende inländischen Anbietern beim Umgang mit ihren Daten vergleichsweise stärker vertrauen: 65 % geben an, deutschen Anbietern ihre Daten anzuvertrauen; 19 % halten ihre Daten dort für sicher. Das im Vergleich niedrigere Vertrauen in Anbieter aus anderen europäischen Ländern (50 %) und den USA (32 %) sowie sehr geringe Werte für China (3 %) und Russland (1 %) deuten auf den Einfluss institutioneller Kontexte, wahrgenommener Integrität und Kompetenz sowie geopolitischer Zuordnungen für Organisationsvertrauen hin. Kaspersky (Kaspersky 2021) berichtet für Deutschland und andere europäische Länder deutliche Institutionen- und Branchenunterschiede: So trauen 52 % der befragten Deutschen Amazon mit ihren Daten, 47 % dem deutschen Staat; Facebook liegt mit 27 % am unteren Ende. Diese Befunde stützen, dass Integrität und Datensicherheit zentrale Treiber des organisationsbezogenen Vertrauens sind.

Branchenunterschiede bestätigt die eos-Studie (eos 2020): In der EU genießen Banken beim Datenumgang mit 54 % das höchste Vertrauen, während Social-Media- und Messenger-Dienste mit 12 % deutlich niedriger bewertet werden. Die Kaspersky-Erhebung (Kaspersky 2021) zeigt zugleich, welche Aspekte Konsumierende für vertrauensrelevant halten: 92 % nennen Datenschutz als wichtigen Faktor; nur 52 % haben jedoch das Gefühl, Kontrolle über ihre Daten zu haben. Sorgen über Datenmissbrauch (77 %) und Diebstahl (60 %) sind verbreitet, gleichzeitig geben viele ihre Daten für Gegenleistungen frei (z. B. Zugang zu Online-Diensten: 83 %; Rabatte: 41 %). Ähnlich berichtet eos (2020), dass 34 % ihre Daten für eine konkrete Gegenleistung verkaufen würden, während nur 17 % einen konkreten Wert benennen können; 66 % fühlen sich in der EU dazu gezwungen, Daten preiszugeben, um digitale Dienste nutzen zu können. Eine Studie des Centre for International Governance Innovation (CIGI) unterstützt dieses Muster: Global wird Social-Media beim Schutz persönlicher Daten systematisch misstraut (CIGI und Ipsos 2019).

Länderspezifische Unterschiede im Organisationsvertrauen gegenüber Technologieanbietern verdeutlicht YouGov (YouGov 2023): Global trauen etwa 40 % Tech-Unternehmen, während ihnen 48 % misstrauen; die Werte variieren stark zwischen den Ländern (z. B. Indien/Indonesien: 66 % Vertrauen; USA/Großbritannien: 25 %; Deutschland: 27 % Vertrauen bei 57 % Misstrauen). Diese Variation spiegelt neben Unternehmenspraktiken auch die allgemeine institutionelle Vertrauenslage und kulturelle Einstellungen gegenüber Technologie wider. Direkt bezogen auf das Narrativ „KI-Anbietern kann man nicht trauen" zeigt KPMG (2023), dass 61 % der Befragten in 17 Ländern KI-Anbietern gar nicht oder nur ambivalent vertrauen; höhere Vertrauenswerte finden sich in den BICS-Ländern (Brasilien, Indien, China, Südafrika), während in vielen westlichen Ländern geringere Werte berichtet werden. Die von KPMG vermuteten Gründe (zunehmende wirtschaftliche Rolle neuer Technologien) sind plausibel, sollten aber als Hypothesen verstanden werden. Ergänzend liefert Cisco (2024) die Anbieterperspektive: 94 % der befragten Datenschutz- und Sicherheitsverantwortlichen glauben, dass Kundinnen und Kunden ohne hinreichenden Datenschutz nicht weiter kaufen würden; 80 % sehen positive Effekte von Datenschutzinvestitionen auf Loyalität und Vertrauen; 91 % sehen zusätzlichen Bedarf, um die Nutzung von KI mit Kundendaten zu erklären. Diese Selbsteinschätzungen zeigen angebotsseitige Handlungsbereitschaft, ersetzen aber nicht die Nutzerperspektive.

2.2 Vertrauen in KI-Systeme

Systembezogene Einstellungen werden in der Deloitte-Studie (Deloitte 2024) sichtbar: 34 % der Befragten sind unsicher oder ahnungslos in Bezug auf KI-Systeme. Unter Personen mit Nutzungserfahrung liegt das Vertrauen in KI-Systeme deutlich höher (57 %) als unter Nichtnutzenden (33 %). Die wahrgenommene Vorteilhaftigkeit der Nutzung und die Unterschiede zwischen privaten Anwendungsfällen (z. B. Steuererklärung: 64 % Vertrauen) und staatlichen Anwendungen (z. B. Entscheidungen über Sozialleistungen: 50 % Vertrauen) zeigen, dass Systemvertrauen kontextabhängig ist und von wahrgenommener Zuverlässigkeit, Vorhersagbarkeit und Sicherheit abhängt. 66 % nennen die Vertraulichkeit und Sicherheit der Daten als zentralen Faktor für Vertrauen; 50 % glauben, dass der staatliche KI-Einsatz regulierbar ist; 51 % trauen Unternehmen zu, KI verantwortungsvoll einzusetzen. Ergänzend berichtet das Eurobarometer (Europäische Kom-

mission 2025), dass 73 % der Europäerinnen und Europäer glauben, dass Digitalisierung ihr Leben künftig erleichtern wird. 89 % sehen Forschung und Innovation als notwendig, um Systeme sicherer zu machen, und 83 % halten es für wichtig, dass KI-Systeme so entwickelt werden, dass Rechte und Werte respektiert werden. Diese normativen Erwartungen sind Teil der Vertrauensbasis in technische Systeme und die dahinterstehenden Organisationen.

Mehrere Studien deuten darauf hin, dass Regulierung als Voraussetzung für Vertrauen gesehen wird (Deloitte 2024; KPMG 2023). Das Edelman Trust Barometer (Edelman 2024) liefert Makrokontext: In „Entwicklungsländern" liegt das allgemeine Vertrauen in Institutionen höher (63 %) als in „entwickelten Ländern" (49 %). Regierungen werden im Mittel als weniger kompetent und ethisch eingeschätzt als Unternehmen, und 39 % bewerten Innovationen als schlecht gemanagt (USA: 56 %; Nigeria: 24 %). Ablehnung gegenüber KI ist bei jenen höher, die Technologiemanagement generell als schlecht bewerten (43 %), und niedriger bei jenen mit positiver Einschätzung (26 %). Datenschutzsorgen sind besonders ausgeprägt unter Skeptikerinnen und Skeptikern. Der gesellschaftliche Nutzen und ein besseres Verständnis von KI-Systemen werden als Schlüssel zur Vertrauensbildung genannt. Diese Ergebnisse stützen, dass Vertrauen in KI-Anbieter nicht isoliert entsteht, sondern an breiteres institutionelles Vertrauen und wahrgenommenes Innovationsmanagement gekoppelt ist. Vor diesem Hintergrund gewinnen Governance-Rahmen an Bedeutung: Prinzipien und Standards wie die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) Grundsätze für KI, die EU-KI-Verordnung sowie ISO/IEC-Normen (z. B. 23894 für KI-Risikomanagement, 42001 für KI-Managementsysteme) adressieren Verantwortlichkeit, Sicherheit und Transparenz und könnten institutionelles Vertrauen stützen.

2.3 Synthese und Implikationen für das Narrativ

Die zusammengeführten Befunde legen nahe, dass Misstrauen gegenüber KI-Anbietern auf mehreren Ebenen entsteht und unterschiedlich begründet wird. Auf der organisationalen Ebene stehen Integrität, Transparenz, Verantwortlichkeit und nachweisliche Kompetenz im Umgang mit Daten im Vordergrund; Defizite hier nähren das Narrativ (Bitkom 2023; Kaspersky 2021; eos 2020; YouGov 2023; KPMG 2023). Auf der Systemebene geht es um wahrgenommene Zuverlässigkeit, Sicherheit, Erklärbarkeit und die

Kalibrierung von Vertrauen durch Erfahrung und Leistungsnachweise (Deloitte 2024; Europäische Kommission 2025). Zwischen den Ebenen könnte es Vertrauensübertragungen geben: Interpersonales Vertrauen in fachlich verantwortliche Personen oder Institutionen kann die Akzeptanz eines Systems erhöhen; umgekehrt kann hohe Systemzuverlässigkeit das Vertrauen in die anbietende Organisation stützen.

Aus diesen Einsichten folgen praktische Hebel zur Vertrauensbildung und zur differenzierten Bewertung des Narrativs: Anbieter sollten Evidenz zur Leistungsfähigkeit, Sicherheit und Robustheit ihrer Systeme bereitstellen, Erklärbarkeit und sinnvolle menschliche Kontrolle gewährleisten, klare Datenschutz- und Governance-Praktiken etablieren und unabhängige Prüfungen sowie Compliance mit anerkannten Standards und Regulierung nachweisen. Gleichzeitig ist die allgemeine institutionelle Vertrauenslage und die Qualität staatlicher Regulierung mitentscheidend dafür, ob diese Maßnahmen gesellschaftlich als glaubwürdig wahrgenommen werden.

3. Das „Privacy-Paradox"

Die Forschung zeigt, dass Offenlegungsentscheidungen selten rein attitudengesteuert sind, sondern etwa von Nutzenkalkülen, sozialem Einfluss und situativen Frames geprägt werden. Einige der betrachteten Studien zeigen ausgeprägte Sorgen um persönliche Daten und begrenztes Vertrauen in Unternehmen und KI-Anbieter. Gleichzeitig weisen Nutzungsdaten auf eine schnelle Verbreitung generativer KI hin. Diese Einstellungs-Verhaltens-Diskrepanz wird als „Privacy-Paradox" bezeichnet: Menschen äußern Datenschutzbedenken, passen ihr Verhalten jedoch nicht immer entsprechend an (Gerber u. a. 2018).

In ihrer systematischen Literaturübersicht identifizierten Gerber u. a. (2018) acht theoretische Erklärungsansätze für das Privacy-Paradox, also das Phänomen, dass Nutzende zwar Bedenken hinsichtlich ihres Datenschutzes äußern, ihr Verhalten jedoch oft nicht entsprechend anpassen.

Privacy Calculus: Nutzende führen eine rationale Abwägung zwischen den erwarteten Vorteilen (z. B. Rabatte, Bequemlichkeit) und den wahrgenommenen Kosten oder Risiken (z. B. Datenmissbrauch) durch. Wenn der Nutzen die Kosten überwiegt, sind Nutzende eher bereit, ihre Daten weiterzugeben. Dieser Ansatz findet starke empirische Unterstützung und erklärt das „Paradox" dadurch, dass Nutzende gleichzeitig besorgt sind und Daten teilen können, wenn der Nutzen ihre Sorgen überwiegt.

Begrenzte Rationalität und kognitive Verzerrungen: Menschen treffen Entscheidungen nicht ausschließlich rational, sondern unterliegen kognitiven Verzerrungen wie dem *Availability Bias* (Überbewertung leicht erinnerbarer Ereignisse) oder dem *Optimism Bias* ("Mir passiert schon nichts"). Diese Verzerrungen führen dazu, dass das Verhalten nicht immer den geäußerten Einstellungen entspricht.

Mangel an persönlicher Erfahrung und Schutzwissen: Da die meisten Nutzenden keine persönlichen Erfahrungen mit Datenschutzverletzungen gemacht haben, basieren ihre Einstellungen oft auf Heuristiken oder Erfahrungen anderer. Fehlendes technisches Wissen verhindert zudem den effektiven Einsatz von Schutzmaßnahmen. Dieser Ansatz findet jedoch nur schwache empirische Unterstützung.

Sozialer Einfluss: Das soziale Umfeld beeinflusst Datenschutzentscheidungen stark. Sozialer Druck durch Peergroups oder Familie sowie das Gefühl der Reziprozität ("Wenn andere teilen, muss ich auch") können dazu führen, dass Nutzende Daten teilen, obwohl sie Bedenken haben.

Risiko-Vertrauens-Modell: Dieses Modell unterscheidet zwischen Risiko und Vertrauen. Während Vertrauen direkt das Verhalten beeinflusst, wirkt sich Risiko eher auf Einstellungen und Intentionen aus. Die empirischen Befunde zu diesem Modell sind jedoch gemischt.

Quantentheorie-Ansatz: Dieser Ansatz vergleicht Entscheidungsprozesse mit der Quantenmechanik. Präferenzen sind nicht fix, sondern ändern sich im Entscheidungsmoment. Das „Paradox" wird dadurch erklärt, dass geäußerte Intentionen nicht immer der tatsächlichen Entscheidung entsprechen.

Illusion der Kontrolle: Nutzende verwechseln die Kontrolle über die Veröffentlichung ihrer Daten mit der Kontrolle über die Nutzung dieser Daten durch Dritte. Datenschutzeinstellungen können ein trügerisches Sicherheitsgefühl vermitteln, was zu einer erhöhten Datenweitergabe führt, obwohl Bedenken bestehen.

Methodisches Artefakt: Dieser Ansatz argumentiert, dass das „Privacy-Paradox" auf methodische Unzulänglichkeiten zurückzuführen ist. So werden Verhalten und Einstellungen oft unterschiedlich operationalisiert, oder es werden indirekte Effekte übersehen. Differenzierte Modelle können das Paradox oft auflösen.

Die von Gerber u. a. (2018) identifizierten Erklärungsansätze zeigen, dass das „Privacy-Paradox" ein vielschichtiges Phänomen ist, das durch eine Kombination aus rationalen Abwägungen, kognitiven Verzerrungen, sozialem Einfluss und methodischen Herausforderungen erklärt werden

kann. Während einige Ansätze wie der *Privacy Calculus* und die Illusion der Kontrolle starke empirische Unterstützung finden, bieten andere wie der Quantentheorie-Ansatz oder das Risiko-Vertrauens-Modell interessante theoretische Perspektiven, die jedoch noch weiter untersucht werden müssen.

Kokolakis (2017) widmet den methodischen Herausforderungen im Bereich des „Privacy-Paradox" mehr Raum als Gerber u. a. (2018) und identifiziert mehrere Faktoren, die zu widersprüchlichen Befunden führen können.

Interpretation der Befunde: Ein zentrales Problem ist die ambivalente Interpretierbarkeit von Forschungsergebnissen. Identische Daten können unter Umständen zu gegensätzlichen Schlussfolgerungen führen. Beispielsweise kann ein Aufpreis von 1,50 Dollar für einen Dienst entweder als "gering absolut" oder als "Verdreifachung des Preises" interpretiert werden. Diese unterschiedliche Bewertung führt dazu, dass identische Daten zu gegensätzlichen Schlussfolgerungen führen können.

Kontextabhängigkeit: Datenschutz-Verhalten ist hochgradig kontextsensitiv. Beispielsweise verhalten sich Nutzende in der vertrauten Umgebung ihres Universitätsseminars anders als beim Online-Shopping in einer anonymen, riskanteren Umgebung. Studien, die verschiedene Kontexte vergleichen, ohne diese Unterschiede zu berücksichtigen, führen zu nicht vergleichbaren Ergebnissen. Generalisierungen über Kontexte hinweg sind daher gefährlich.

Diversität persönlicher Informationen: Persönliche Informationen sind kein kohärentes Objekt. Verschiedene Datentypen wie Standort, Gesundheit, Browsing-Historie, Alter oder Gewicht werden unterschiedlich bewertet. Die Sensitivität dieser Informationen wird oft vernachlässigt, obwohl sie zentral ist. Studien, die verschiedene Datentypen direkt vergleichen, ignorieren diese Sensitivität und könnten das „Privacy-Paradox" als Artefakt dieser Vernachlässigung darstellen.

Diversität der Datenschutzbedenken: Datenschutzbedenken sind kein einheitliches Konstrukt. Es gibt verschiedene Arten von Bedenken, wie soziale Bedrohungen (Mobbing, Stalking), organisationale Bedrohungen (Sekundärnutzung, Drittparteien, Marketing) und unangemessener Zugriff (Arbeitgeber, Öffentlichkeit). Diese Vielfalt wird oft nicht berücksichtigt, was zu einer ungenauen Darstellung der tatsächlichen Datenschutzbedenken führt.

Methodische Vielfalt und ihre Probleme: Kokolakis identifiziert mehrere methodische Herausforderungen, die zu widersprüchlichen Ergebnissen führen können:

a. Umfragen eignen sich gut zur Erfassung von Überzeugungen und Einstellungen, aber nicht zur Erfassung des tatsächlichen Verhaltens. Selbsteinschätzungen, insbesondere zu irregulärem oder infrequentem Verhalten, sind oft ungenau. Beispielsweise weicht das tatsächliche Verhalten von Google-Nutzenden stark von ihrem berichteten Verhalten ab.
b. Experimente können aufgrund unterschiedlicher Rahmenbedingungen, mangelndem Realismus und dem Bewusstsein der Teilnehmer, dass sie an einem Experiment teilnehmen, zu unterschiedlichen Ergebnissen führen. Beispielsweise ergaben die Untersuchungen von Huberman u. a. (2005) und Carrascal u. a. (2013) trotz gleicher Auktionsmethode unterschiedliche Ergebnisse aufgrund unterschiedlicher Rahmenbedingungen.
c. Kokolakis verweist auf die Studie von Dienlin und Trepte (2014), die zeigt, dass unterschiedliche Modelle und Analysemethoden zu widersprüchlichen Ergebnissen führen können. Während eine einfache Regression oft keine Zusammenhänge zwischen Datenschutzbedenken und spezifischem Datenschutz-Verhalten findet, kann ein differenziertes Strukturgleichungsmodell indirekte Zusammenhänge nachweisen und das Paradox so auflösen.

Die von Kokolakis identifizierten methodischen Herausforderungen verdeutlichen, dass das „Privacy-Paradox" ein komplexes Phänomen ist, das durch eine Vielzahl von Faktoren beeinflusst wird. Die unterschiedlichen methodischen Ansätze und die Vernachlässigung von Kontexten, Informationssensitivität und der Vielfalt von Datenschutzbedenken führen zu widersprüchlichen Befunden in der Forschung. Ein interdisziplinärer und differenzierter Ansatz ist daher notwendig, um das Phänomen umfassend zu verstehen und gezielte Maßnahmen zum Schutz der Privatsphäre zu entwickeln.

Die Ergebnisse von Kokolakis (2017) und Gerber u. a. (2018) verdeutlichen diese Ansätze, dass die Diskrepanz zwischen Datenschutzeinstellung und -verhalten nicht mehr als Paradox bezeichnet werden sollte. Die verschiedenen Erklärungsansätze zeigen, dass das Verhalten von Nutzenden im Umgang mit ihren Daten durch die Interaktion verschiedener Faktoren entsteht. Aus diesen Befunden folgt: Das „Paradox" ist weniger ein Widerspruch als eine erwartbare Attitüden-Verhaltens-Lücke, die aus Nutzen-Risiko-Abwägungen, kognitiven und sozialen Prozessen sowie Mess-

und Kontexteffekten entsteht. Für die Analyse bedeutet das: Hohe Nutzung von KI-Systemen kann mit niedrigerem Organisationsvertrauen koexistieren. Damit liefert das „Privacy-Paradox" die Brücke, um die im nächsten Abschnitt dargestellten Nutzungsdaten zu interpretieren, ohne sie als Widerlegung von Vertrauenssorgen zu missverstehen.

Vor dem Hintergrund des „Privacy-Paradox" steht zu erwarten, dass tatsächliche Nutzungsentscheidungen von generativen KI-Systemen stark vom konkreten Nutzen, Kontext und Design abhängen – und daher auch bei bestehender Skepsis positiv ausfallen können. Die folgenden Daten zu KI-Nutzung sind entsprechend als Ausdruck kontextueller Nutzen-Risiko-Abwägungen zu lesen, nicht als unmittelbares Maß für Organisationsvertrauen.

4. Nutzung von KI-Systemen

Nachdem wir festgestellt haben, dass das Vertrauen in KI-Anbieter von verschiedenen Faktoren beeinflusst wird, lohnt sich ein genauerer Blick darauf, wie Menschen KI-Systeme tatsächlich nutzen. Die Art und Weise der Nutzung kann wertvolle Hinweise darauf geben, welche Erwartungen und Bedürfnisse die Nutzenden haben.

OpenAI, einer der weltweit führenden Anbieter von KI-Systemen, hat in einer Studie (Chatterji u. a. 2025) das Nutzungsverhalten des eigenen KI-Systems ChatGPT analysiert. ChatGPT ist ein KI-basierter Chatbot, der auf der Verarbeitung natürlicher Sprache basiert und Nutzenden ermöglicht, Fragen zu stellen, Texte zu verfassen oder Informationen zu erhalten. In der Studie wurden die Länder nach ihrem Bruttoinlandsprodukt pro Kopf in Niedrig- bis Mittel-Einkommensländer (10.000–40.000 US-Dollar) und Hoch-Einkommensländer (über 40.000 US-Dollar) eingeteilt. Die Ergebnisse zeigen, dass die Nutzung von ChatGPT in den wirtschaftlich schwächeren Ländern seit der Einführung im Jahr 2022 deutlich schneller gewachsen ist als in den wohlhabenderen Ländern. Dieser Befund deckt sich mit den in Abschnitt 2 betrachteten Studien, in denen für „Entwicklungsländer" eine höhere Akzeptanz von KI-Systemen festgestellt wurde als in westlichen Ländern.

Obwohl die Studie von OpenAI keine detaillierte Aufschlüsselung nach einzelnen Ländern bietet, liefert sie dennoch wertvolle Einblicke in das allgemeine Nutzungsverhalten. Für die Auswertung wurden die geführten Chats automatisiert in berufliche und nicht berufliche Unterhaltungen un-

terteilt. Dabei machen nicht berufliche Chats mit rund 70 % den größten Anteil aus, während berufliche Anwendungen etwa 30 % ausmachen. Beide Bereiche zeigen ein kontinuierliches Wachstum.

Die thematische Analyse der Chats ergab, dass die meisten Unterhaltungen in drei Hauptkategorien fallen: praktische Anleitungen, Informationssuche und Schreiben. Im beruflichen Kontext dominiert das Schreiben als häufigstes Nutzungsszenario. Im privaten Bereich stehen Themen wie Nachhilfe, Ratgeber zu Gesundheit, Fitness und Schönheit, die Bearbeitung oder Bewertung von Texten sowie Unterstützung bei persönlicher Kommunikation im Vordergrund. Gerade die private Nutzung umfasst demnach auch sensible und schützenswerte Daten, insbesondere Gesundheitsinformationen, die beispielsweise in der EU durch die DSGVO besonders geschützt sind.

Um das Verhältnis zwischen Vertrauen und tatsächlicher Nutzung besser zu verstehen, lohnt sich ein Blick auf eine Studie der Wochenzeitung „Zeit Online" (Killian 2025), die die Verwendung von ChatGPT in Deutschland untersucht hat. Laut der Studie nutzen die Deutschen ChatGPT im europäischen Vergleich am häufigsten. Weltweit zählt Deutschland neben den USA und Südkorea zu den drei Ländern mit der höchsten Nutzung. Die Nutzungszahlen haben sich in Deutschland im Jahr 2025 im Vergleich zu 2024 verdreifacht.

Die Studie unterscheidet zwar nicht explizit zwischen privater und beruflicher Nutzung, listet jedoch die zehn häufigsten von ihnen identifizierten Anwendungsszenarien auf. An erster Stelle steht das Schreiben von Texten und die Unterstützung bei persönlicher Kommunikation. Es folgen Lernen, Nachhilfe und verschiedene Ratgeberfunktionen. Auf Platz sechs rangiert die Suche nach Beziehungsratschlägen, wobei viele Nutzende Hilfe beim Sortieren ihrer Gefühle oder beim Formulieren komplexer Themen suchen. Auch die Nachfrage nach Ratschlägen zu Gesundheit, Fitness und Schönheit ist hoch (Platz acht). Hier geht es vor allem um Unterstützung bei der Diagnose von Krankheiten, Tipps zu Ernährung und Sport sowie Empfehlungen für das eigene Aussehen.

Die Nutzung von KI-Systemen wie ChatGPT wächst weltweit rasant. Dabei zeigt sich, dass die Anwendungsmöglichkeiten weit über den beruflichen Kontext hinausgehen und vor allem im privaten Bereich eine zentrale Rolle spielen. Sensible und schützenswerte Daten werden dabei häufig geteilt, was den Bedarf an staatlichen Regulierungen zum Schutz der Privatsphäre unterstreicht. Die Analyse der Nutzung in Deutschland bestätigt diese globalen Trends und verdeutlicht, wie tief KI-Anwendungen bereits in

den Alltag und die verschiedenen Lebensbereiche der Menschen integriert sind. Vertrauen in KI-Anbieter ist daher auch eine Frage der gesellschaftlichen und rechtlichen Rahmenbedingungen, die den verantwortungsvollen Umgang mit persönlichen Daten sicherstellen müssen.

Abschließend lässt sich festhalten, dass sich das Narrativ, KI-Anbietern könne nicht vertraut werden, zumindest in Deutschland nicht in der tatsächlichen Nutzung widerspiegelt. Die hohe und stetig wachsende Nutzung von ChatGPT zeigt, dass viele Menschen die Angebote von KI-Anbietern aktiv in ihren Alltag integrieren und zum Teil auch sensible Daten teilen. Das Misstrauen gegenüber KI scheint daher weniger die tatsächliche Nutzung zu beeinflussen.

5. Fazit

Die Untersuchung des Vertrauens in KI-Anbieter zeigt, dass das Narrativ vom grundsätzlichen Misstrauen gegenüber diesen Unternehmen differenzierter betrachtet werden muss als zunächst angenommen. Vertrauen erweist sich als vielschichtiges, kontextabhängiges Phänomen, das sich je nach Land, Branche, Anwendungskontext und Governance-Rahmen sehr unterschiedlich ausprägt. Auffällig sind länderspezifische Differenzen, doch sie folgen keinem einfachen Nord-gegen-Süd-Muster: In vielen wohlhabenden Ländern finden sich ausgeprägtere Vorbehalte, während in einigen Schwellenländern höhere Zustimmungswerte berichtet werden. Solche Unterschiede scheinen weniger eine Frage der Geografie zu sein als Resultat variierender institutioneller Vertrauenslagen, wahrgenommener Chancen und Risiken sowie konkreter Anbieter- und Regulierungspraxis.

Zentral ist die Unterscheidung zwischen organisationsbezogenem Vertrauen in Anbieter und dem angemessenen Verlassen auf technische Systeme. Ersteres bezieht sich auf Integrität, Verantwortlichkeit und Kompetenz von Unternehmen und Institutionen; letzteres auf Zuverlässigkeit, Sicherheit, Vorhersagbarkeit und Erklärbarkeit konkreter Systeme. Beide Ebenen sind miteinander verknüpft, determinieren sich aber nicht gegenseitig. Hohe Systemnutzung kann mit Skepsis gegenüber Anbietern koexistieren, und gute organisatorische Governance kann Akzeptanz für Systeme stärken, ohne alle Vorbehalte zu beseitigen.

Die wiederkehrende Diskrepanz zwischen geäußerten Datenschutzbedenken und tatsächlichem Nutzungsverhalten ist vor diesem Hintergrund nicht überraschend. Sie wird in der Forschung als „Privacy-Paradox“ disku-

tiert. Welche Mechanismen diese Lücke im Einzelfall erklären, bleibt offen: Denkbar sind Nutzen-Risiko-Abwägungen, kognitive und soziale Einflüsse, Gestaltungseffekte von Systemen sowie methodische und kontextuelle Faktoren. Wichtig ist, die Nutzung von KI-Anwendungen nicht vorschnell als Beleg für hohes Organisationsvertrauen zu deuten, sondern als Ausdruck konkreter, situativer Abwägungen und wahrgenommenen Nutzens – häufig unter flankierenden Schutzstrategien der Nutzenden.

Die Studienergebnisse unterstreichen die Bedeutung verlässlicher Datenschutz- und Sicherheitspraktiken sowie klarer, wirksam durchgesetzter Regeln. Regulierung kann Vertrauen stützen, setzt aber wiederum Vertrauen in regulierende Institutionen voraus, das je nach Land unterschiedlich ausgeprägt ist.

Verantwortung für die Nutzung von KI-Systemen darf nicht beim Einzelnen abgeladen werden. Die Verantwortung des Einzelnen scheitert in der Praxis an strukturellen Grenzen: Informationsasymmetrien, intransparente Datenflüsse, komplexe Modelle und gestalterische Einflussnahmen wie z. B. bestimmte Standardeinstellungen oder aufdringliche Dialoge erschweren es den Nutzenden, informierte Entscheidungen zu treffen. Es ist daher legitim und notwendig, die Hauptverantwortung dort zu verankern, wo Risiken entstehen und wo Nutzen realisiert wird: bei den Organisationen, die Systeme konzipieren, betreiben und monetarisieren, und bei den Institutionen, die Regeln definieren und durchsetzen.

Für die weitere Entwicklung heißt das: Die Vertrauensfrage bei KI ist keine binäre Ja/Nein-Entscheidung, sondern ein dynamisches Aushandeln zwischen individuellem Nutzen, wahrgenommenen Risiken, technischen Eigenschaften der Systeme und der Glaubwürdigkeit der dahinterstehenden Organisationen und Aufsichtsstrukturen. Die Nutzungsdaten deuten darauf hin, dass viele Menschen KI-Systeme in ihren Alltag integrieren, ohne dass damit pauschal hohes Vertrauen in alle Anbieter einherginge. Die Verantwortung liegt nicht allein bei den Nutzenden, sondern vor allem bei Anbietern und staatlichen Institutionen, die verlässliche Rahmenbedingungen schaffen und glaubwürdig durchsetzen müssen.

Literatur

Afroogh, Saleh, Ali Akbari, Emmie Malone, Mohammadali Kargar, und Hananeh Alambeigi (2024). „Trust in AI: Progress, Challenges, and Future Directions". *Humanities and Social Sciences Communications* 11(1): 1568. doi:10.1057/s41599-024-04044-8.

Bach, Tita Alissa, Amna Khan, Harry Hallock, Gabriela Beltrão, und Sonia Sousa (2024). „A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective". *International Journal of Human–Computer Interaction* 40(5): 1251–66. doi:10.1080/10447318.2022.2138826.

Bitkom (2023): Datenschutz: Größtes Vertrauen in deutsche Anbieter. Presseinformation. URL: https://www.bitkom.org/Presse/Presseinformation/Datenschutz-groesstes-Vertrauen-in-deutsche-Anbieter (besucht am 12.12.2024).

Carrascal, Juan Pablo; Riederer, Christopher; Erramilli, Vijay; Cherubini, Mauro und de Oliveira, Rodrigo (2013): Your browsing behavior for a big mac: Economics of personal information online. In: Proceedings of the 22nd International Conference on World Wide Web, S. 189-200.

Chatterji, Aaron; Cunningham, Thomas; Deming, David J.; Hitzig, Zoe; Ong, Christopher; Shan, Carl Yan und Wadman, Kevin (2025): How People Use ChatGPT (NBER Working Paper No. 34255). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w34255.

Cisco (2024): Privacy Benchmark Study 2024. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-privacy-benchmark-study-2024.pdf (besucht am 12.12.2025).

Deloitte (2024): Trust in Generative AI in Europe. https://www.deloitte.com/us/en/insights/topics/digital-transformation/trust-in-generative-ai-in-europe.html (besucht am 12.12.2025).

Dienlin, Tobias und Trepte, Sabine (2014): Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology, 45(3), S. 285-297.

Edelman (2024): Trust Barometer: Key Insights Around AI. URL: https://www.edelman.com/sites/g/files/aatuss191/files/2024-03/2024%20Edelman%20Trust%20Barometer%20Key%20Insights%20Around%20AI.pdf (besucht am 12.12.2025).

eos (2020): Was sind Daten wert? Studie. URL: https://de.eos-solutions.com/de/dam/jcr:055985a9-dda5-4718-90fd-73159a71c1dc/EOS%20Studie%20Was%20sind%20Daten%20wert.pdf (besucht am 12.12.2025).

Europäische Kommission (2025): Eurobarometer: Digitalisierung (Survey 3362). URL: https://europa.eu/eurobarometer/surveys/detail/3362 (besucht am 12.12.2025).

Gerber, Nina; Gerber, Paul und Volkamer, Melanie (2018): Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. Computers & Security, 77, S. 226-261.

Glikson, Ella, und Anita Williams Woolley (2020). „Human Trust in Artificial Intelligence: Review of Empirical Research". *Academy of Management Annals* 14(2): 627–60. doi:10.5465/annals.2018.0057.

Hoff, Kevin Anthony, und Masooda Bashir (2015). „Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust". *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57(3): 407–34. doi:10.1177/0018720814547570.

Huberman, Bernardo A.; Adar, Eytan und Fine, Leslie R. (2005): Valuating privacy. IEEE Security & Privacy, 3(5), S. 22-25.

Kaspersky (2021): Datenschutz-Passe: 83 Prozent der Deutschen würden persönliche Daten gegen Gratis-Services tauschen trotz Sicherheitsbedenken. Pressemitteilung. https://www.kaspersky.de/about/press-releases/datenschutz-passe-83-prozent-der-deutschen-wurden-personliche-daten-gegen-gratis-services-tauschen-trotz-sicherheitsbedenken (besucht am 12.12.2025).

Killian, Nicolas (2025): Was die Deutschen ChatGPT fragen. Zeit Online. https://www.zeit.de/digital/internet/2025-04/chatgpt-ki-europa-deutschland-anfragen-beziehung (besucht am 12.12.2025).

Kokolakis, Sokratis (2017): Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security, 64, S. 122-134. doi:10.1016/j.cose.2015.07.002.

KPMG (2023): Trust in AI: Global Insights 2023. https://assets.kpmg.com/content/dam/kpmgsites/au/pdf/2023/trust-in-ai-global-insights-2023.pdf (besucht am 12.12.2024).

Nancy Lankton, D. Harrison McKnight und John Tripp (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association for Information Systems* 16(10): 880–918. doi:10.17705/1jais.00411.

Lee, J. D., und K. A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46(1): 50–80. doi:10.1518/hfes.46.1.50_30392.

Luhmann, Niklas. 1973. *Vertrauen: ein Mechanismus der Reduktion sozialer Komplexität*. 2. erweiterte Auflage. Stuttgart: Ferdinand Enke Verlag.

Mayer, Roger C., James H. Davis, und F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20(3): 709. doi:10.2307/258792.

Prognos AG und Z_punkt (o.J.): Digitales Vertrauen. VORAUS:schau!-Prozess, Zukunftsbüro beim Bundesministerium für Bildung und Forschung. URL: https://www.bmftr.bund.de/SharedDocs/Downloads/DE/v/zukunft-des-vertrauens--digitale-welten.pdf?__blob=publicationFile&v=3 (besucht am 12.12.2024).

Stilgoe, Jack (2023). What Does It Mean to Trust a Technology? *Science* 382(6676): eadm9782. doi:10.1126/science.adm9782.

Yao, Jie, Zhibo Zhou, Huaqing Cui, Yujie Ouyang, und Wenhao Han (2025). „Trust Transfer from Medical AI to Doctors and Hospitals: Integrating Digital, AI, and Scientific Literacy in a Cross-Sectional Framework". *BMC medical ethics* 26(1): 144. doi:10.1186/s12910-025-01300-7.

YouGov (2023): 2022 vs. 2023: Do Global Consumers Trust Tech Companies With Their Personal Data? URL: https://yougov.com/articles/47075-2022-vs-2023-do-global-consumers-trust-tech-companies-with-their-personal-data (besucht am 12.12.2025).

Über die Autorin

Esther von der Weiden
arbeitet an der Stabsstelle Datenschutz und am Lehrstuhl für soziotechnisches Systemdesign und künstliche Intelligenz der Ruhr-Universität Bochum. Nachdem sie ihr Studium der angewandten Informatik an der Universität Duisburg-Essen 2024 mit dem M. Sc. abgeschlossen hat, beschäftigt sie sich als Doktorandin mit der Schnittstelle zwischen KI und Datenschutz.