

Also Available
in English



Mehr als nur KI-Fähigkeiten

Operationalisierung von vertrauensvoller KI

Volume 5, Nr. 6

Dezember 2024

Michael Leyer Universität Marburg

Lilian Do Khac Universität Marburg

Sarah Engel Universität Marburg

Layout & Design: Oliver Behn



White Paper Serie des Lehrstuhls ABWL:
Digitalisierung und Prozessmanagement

Volume 5

Einleitung

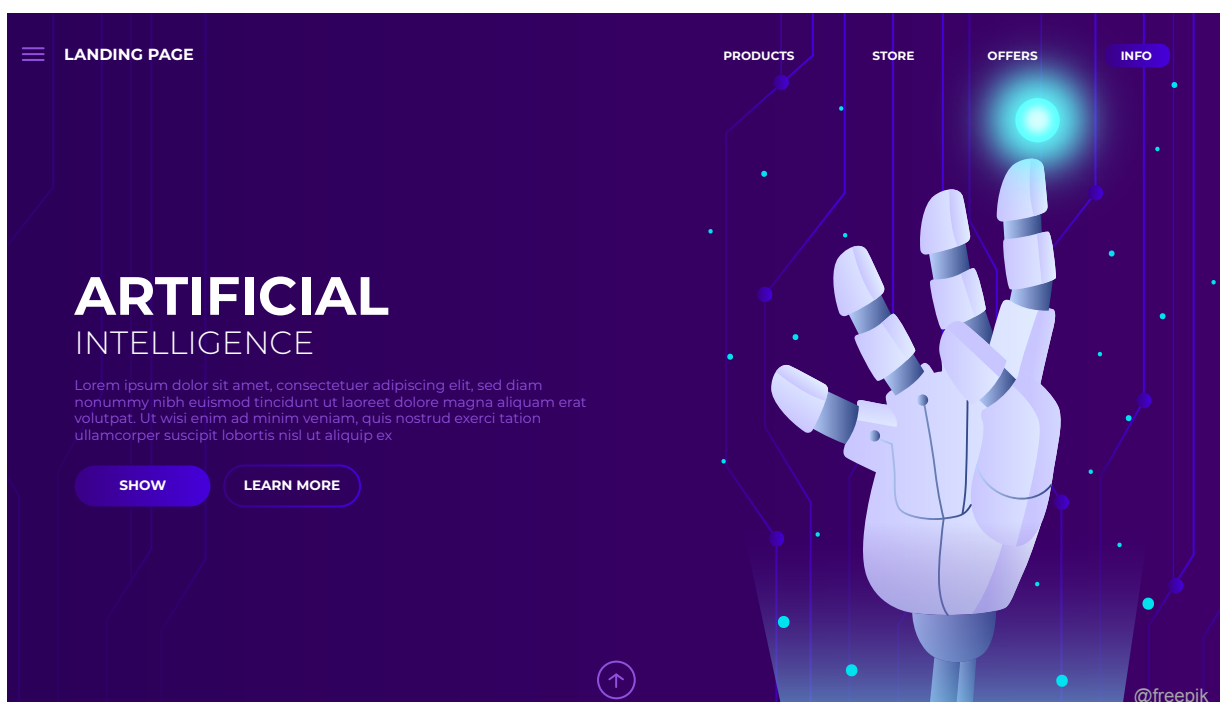
Die Durchsetzung von Qualitätsstandards in der Entwicklung und Bereitstellung von KI-Systemen, die sich gewinnbringend in das Mensch- und Naturgefüge integrieren, wird wohl eine der prägendsten Fragestellungen des digitalen Zeitalters sein. Schon jetzt sind wesentliche regulatorische Instrumente auf die Wege gebracht worden, die grundlegende Interessen wie Gesundheit, Sicherheit und fundamentale Rechte sichern sollen und damit auch eine funktionierende soziale Marktwirtschaft. Dabei geht es um mehr als nur der Identifizierung von Verantwortlichkeiten (im englischen „Responsible“), also aufgabenfokussiert, sondern um die Verpflichtung und Rechenschaft gegenüber dem Ergebnis (im englischen „Accountable“), also ergebnisfokussiert.

Letzteres wird, vor allem im europäischen Raum, unter dem Oberbegriff „Trustworthy AI“ adressiert. KI-Systeme können in nahezu jedem Industriezweig in nahezu jeder Aufgabe eingesetzt werden und bringen somit umfas-

send skalierte Potentiale wie auch Risiken mit sich. Um diese zu beherrschen und den Sorgfaltspflichten nachzukommen ist, zum einen das Wissen über die jeweiligen konkreten Stellhebel direkt am KI-System wie zum anderen auch über einer geeigneten Governance um solche Systeme, wichtig. Ein Zeugnis über die Erfüllung von Qualitätsstandard bei KI-Produkten oder KI-Dienstleistungen resultiert in Vertrauenswürdigkeit und drückt damit Wettbewerbsfähigkeit im Zeitalter der Digitalisierung aus.

Ein Grundgerüst für KI-Systeme

In der Beziehung von Menschen zu KI fragen wir uns: Welche Eigenschaften muss eine KI für optimale Chancen zur Vertrauensgewinnung eines Menschen mitbringen? In diesem Zusammenhang wird die Vertrauensbildung einer empfangenden Partei durch die drei vertrauensbildenden Faktoren Fähigkeiten, Wohlwollen und Integrität der entsendenden Partei hergestellt. Die eigene Grundeinstellung von der Empfängerpartei spielt auch eine Rolle inwieweit solche Faktoren erfolgreich sein können.

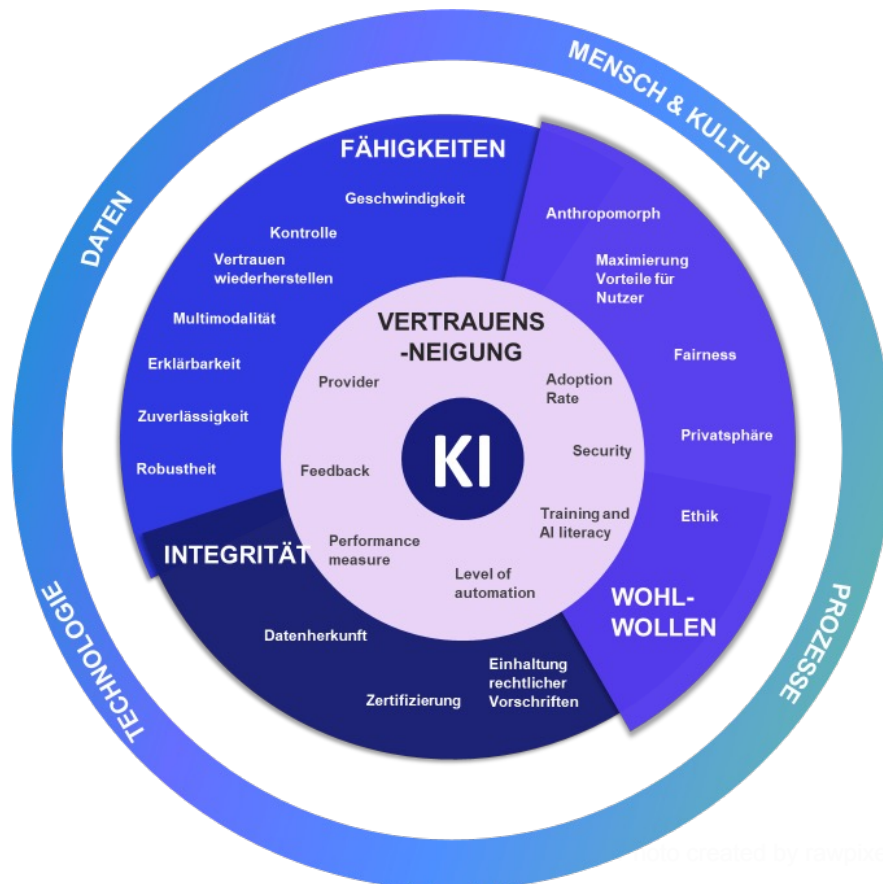


Ein Rückblick über den akademischen Diskurs zu KI oder autonomen Systemen im Zusammenhang mit Vertrauen legt nahe, dass bereits viele Stellhebel für die Optimierung von (teil-) autonomen Systemen bekannt sind.

Dabei tragen aus der Sicht des Menschen konzeptionelle Elemente wie Robustheit, Zuverlässigkeit oder Multimodalität zu einer positiven Wahrnehmung der Fähigkeiten einer KI bei. Robustheit und Zuverlässigkeit sind bereits breit und tief diskutierte Elemente, welche auch in bisherigen (Forschungs-) Arbeiten fest verankert sind. Multimodalität hingegen ist ein weniger durchdrungenes Konzept, welches jedoch schon mit den neueren KI-Generationen, die verschiedene sensomotorische Eingangskanäle verarbeiten können (Text, Bild, Audio), Erfolge vorweisen kann. Konzeptionelle Elemente wie Anthropomorphismus, Privatsphäre oder Fairness tragen zu einer positiven Wahr-

nehmung des Menschen hinsichtlich des Wohllollens zu einer KI bei. Anthropomorphe Eigenschaften sind solche hinsichtlich menschlicher oder menschenähnlicher Gestalt und Verhalten. Eine positive Wahrnehmung hinsichtlich der Integrität einer KI können über Datenherkunft oder Zertifizierungen erzeugt werden. Datenherkunft soll dabei Transparenz über die eingesetzten Trainingsdaten für das KI-System wie auch ein Zeugnis der Vertrauenswürdigkeit über diese Trainingsdaten erbringen. Solche Kennzeichnungen gibt es beispielsweise in der Lebensmittelindustrie. Die interne Wahrscheinlichkeit einer KI zu vertrauen wird durch konzeptionelle Elemente wie Training und Ausbildung über die KI-Systeme oder aufgrund des Herstellers hinter der KI beeinflusst.

Es gibt also viele Stellhebel, jedoch bleibt noch unklar wie ein geeignetes Kalibrierungs-Set für



Created by [researchgate.net](https://www.researchgate.net/publication/351111111) / Freepik

die jeweiligen Stellhebel zu welchem Anlass, für welche Domäne etc. bestimmt werden kann oder grundlegend sein sollte.

Das richtige Maß für KI-Governance

Die Einführung von KI-Systemen hat einen besonders tiefgreifenden Einfluss in die Geschäftsprozesse und das Zusammenspiel von Organisationen im Vergleich zu vorherigen technischen Innovationen, da sie die Machtverteilung zwischen Technologie und menschlichem Bediener verstärken. KI-Systeme können keine Absichten haben, daher erfordert die Verantwortlichkeit für ihre Handlungen eine Verfolgung der Verantwortlichkeit über mehrere Stakeholder hinweg, von der Konzeption bis zur Nutzung. Wenn KI-Systeme sub-optimal in Geschäftsprozesse integriert sind, ist es unwahrscheinlich, dass die erwarteten Effizienzgewinne eintreten, und mögliche Risiken können sogar eskalieren, was den Nutzungsvorteil der KI untergraben kann. Um sicherzustellen,

dass verteilt verantwortliche Akteure reibungslos miteinander interagieren, ist eine Transparenz und ein Rahmen für die Zusammenarbeit erforderlich, um diesen komplexen Zusammenhang zu koordinieren. Dies ist das traditionelle Problem welches eine Governance-Struktur erfordert. Ein funktionierendes Rahmenwerk der KI-Governance stellt daher sicher, dass vertrauenswürdige KI-Systeme entwickelt und kontinuierlich verbessert werden. Als solches sollte die AI-Governance die organisatorische Nutzung von KI und die Bereitstellung von Organisationen zur Erreichung ihrer Ziele leiten. Um dies zu erreichen, ist es unerlässlich, KI-Systeme als vertrauenswürdig bewerten zu können.

Die vier wesentlichen organisatorischen Schlüsselfähigkeiten, um eine funktionsfähige KI-Governance zu betreiben, sind: Daten, Technologie, Mensch und Kultur sowie Prozesse. Daten bilden die Grundlage für KI-Anwendungen. Daher ist die Beherrschung der Datenqualität und das Management der Datenquellen eine



@freepik

Kernkompetenz für erfolgreiche KI-Anwendungen. Technologie bringt eine KI-Anwendung überhaupt erst zum Leben. Kompetenz über das richtige Grundgerüst eines KI-Systems mit dem entsprechenden Kalibrierungs-Set ist von Bedeutung. Anschließend muss die Organisation mit ihren Mitarbeitern und ihrer Kultur in der Lage sein, diese KI-Anwendungen gewinnbringend einzusetzen, was sich durch starke Prozesse anschließend skalieren lässt.

Nächste Schritte in das KI-Zeitalter

Dieser Rahmen für vertrauenswürdige KI kann als Leitfaden dienen, um Ihre unternehmenseigene KI-Reise zu navigieren und erfolgreiche Implementierungen zu gewährleisten. Durch die Berücksichtigung unserer vorgeschlagenen Stellhebel können gezielte Verbesserungsmaßnahmen in der Mensch-Maschine-Interaktion erzielt werden. Unternehmen sollten sich der verschiedenen Elemente und ihres Reifegrads in Bezug auf vertrauenswürdige KI bewusst sein. Eine ausgewogene Steuerung dieser Aktivitäten ist entscheidend, um die gewünschten Ergebnisse zu erreichen. Unser Rahmenwerk bietet eine umfassende Orientierung, die es ermöglicht, Vertrauen aufzubauen und eine positive Mensch-KI-Beziehung zu gestalten. In der Umsetzung sollten Unternehmen solche Aspekte berücksichtigen, um eine vertrauenswürdige und erfolgreiche KI-Integration zu erreichen, die zu einem Wettbewerbsvorteil und einer besseren Kundenzufriedenheit führt. Beachten Sie dabei die Kontrolle, Erklärbarkeit, Zuverlässigkeit und Robustheit, um ein solides Fundament für Ihre KI-Strategie zu schaffen, sowie das KI-Governance Gerüst bzgl. Daten, Technologie, Mensch & Kultur sowie den dazugehörigen Prozessen.

KONTAKTDATEN

Prof. Dr. Michael Leyer
Lehrstuhl ABWL:
Digitalisierung und Prozessmanagement

Fachbereich Wirtschaftswissenschaften

Adjunct Professor, School of Management,
Queensland University of Technology,
Brisbane, Australien

Email michael.leyer@wiwi.uni-marburg.de