

Also Available
in English



Womit Menschen mit KI wirklich vertrauen

Immer noch nichts Neues!?

Volume 4, Nr. 6

Februar 2023

Michael Leyer Universität Marburg

Lilian Do Khac Universität Marburg & Adesso SE

Layout & Design: Oliver Behn



White Paper Serie des Lehrstuhls ABWL:
Digitalisierung und Prozessmanagement
Volume 4

Vertrauen in KI ist eines der großen Themen dieser Tage und möglicherweise eines der führenden Konzepte, das in Zukunft zwischen Mensch und KI vermittelt. Es werden neue empirische Erkenntnisse sowie neue Modelle und Faktoren geschaffen, um das Vorhandensein oder Fehlen von Vertrauen in KI zu ermitteln. Es gibt zwar viele Studien über die Wahrnehmung von KI, aber die Frage ist, was in dieser Hinsicht wirklich neu ist. Die zugrundeliegenden Theorien sind bekannt, und die meisten Menschen verstehen nicht, was KI ist, wenn sie entscheiden, ob sie die Nutzung von Diensten oder die Zusammenarbeit ablehnen oder akzeptieren. Und doch gibt es Unterschiede zu herkömmlicher Software, die vielleicht nicht bewusst wahrgenommen werden.

Wem wir mit KI vertrauen

Der erste Aspekt ist die Bezugsperson/das Objekt des Vertrauens. Bei KI besteht der Hauptunterschied zu herkömmlicher Software darin, dass sie innerhalb der gesetzten Grenzen eigene Strategien entwickelt, um die gestellten Aufgaben zu lösen. KI lernt aus den bereitgestellten Daten und es ist oft unklar, wie die erzielten Ergebnisse berechnet werden bzw. ob sie angemessen sind. KI ist jedoch nicht objektiv, sondern wird von normativen Ansichten gesteuert, die entweder in den bereitgestellten Daten, in der Art und Weise, wie der Algorithmus trainiert oder ursprünglich programmiert wurde, oder in der Art und Weise, wie er angewendet wird, enthalten sind. Es gibt also mehr Akteure als die KI, denen man vertrauen muss. Neben der KI selbst können dies der Eigentümer der KI, das Programmierunternehmen (falls nicht identisch), das Schulungsunternehmen (falls nicht identisch mit dem Eigentümer), das Unternehmen, das die KI nutzt, um den

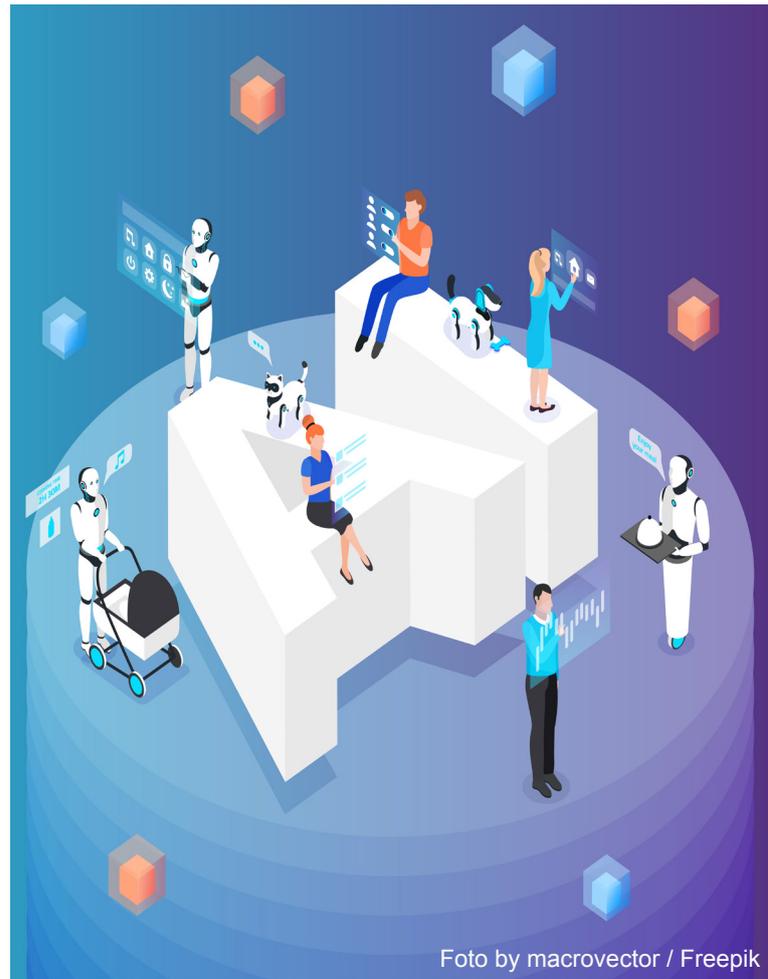


Foto by macrovector / Freepik

Kunden Dienstleistungen anzubieten (falls nicht identisch mit dem Eigentümer) und die normative Haltung zur KI (Regulierung & Normen) sein.

Wer KI vertraut

Die Frage, wer einer KI (oder den Unternehmen, die sie einsetzen) Vertrauen entgegenbringt, ist ebenfalls von großer Bedeutung, da sie über die Fähigkeit entscheidet, eine KI zu beurteilen. Laien, die sich nicht mit KI auskennen, können den Aufwand und die Infrastruktur dahinter nicht nachvollziehen. Daher können sie z. B. das Training und die Programmierung nicht berücksichtigen. Sie werden eher nach den Eigenschaften der KI selbst oder des Anbieters, der sie einsetzt, urteilen. Die Ergebnisse einer empirischen Studie zeigen, dass tatsächlich in

bestimmten Situationen das Vertrauen in den KI-Besitzer das fehlende Vertrauen in eine KI überwiegt, um die KI zu nutzen. Darüber hinaus sind es für Laien oft andere Merkmale als die Beschaffenheit der KI selbst, die in Betracht gezogen werden. Dies kann z. B. die Stimme sein, die von einer KI verwendet wird, die ein beliebter Schauspieler ist, mit dem Vertrauen verbunden ist. Ein anderes Beispiel ist die Verwendung einer beliebten Figur, wenn ein Avatar vorhanden ist. Solche Faktoren können andere Aspekte wie die Qualität der Ergebnisse einer KI überlagern. Die Art des Algorithmus ist oft nicht relevant, da Laien die Details nicht verstehen, selbst wenn sie wissen, welche Art von Algorithmus angewendet wird. Wir haben in einer empirischen Studie gezeigt, dass Laien unabhängig von der Art eines Algorithmus ein ähnliches Vertrauen, eine ähnliche Bereitschaft und Akzeptanz entwickeln. Selbst wenn diese Algorithmen durch grundlegende Merkmale gekennzeichnet sind, die gegensätzlich sind und unterschiedliche Haltungen herausfordern sollten.

Im Gegensatz dazu werden Experten, z. B. von einer Regulierungsbehörde, jedoch die gesamte Wertschöpfungskette einschließlich aller Akteure beurteilen und im Detail analysieren. Sie werden daran interessiert sein zu verstehen, welche Daten verwendet werden, wie sie verarbeitet werden und warum eine KI ihre Ergebnisse ableitet. In ähnlicher Weise werden Unternehmen, die KI von anderen Anbietern einsetzen, an solchen Details interessiert sein, um sicherzustellen, dass eine KI entsprechend arbeitet, um minimale Risikobedingungen zu erreichen. Wenn Faktoren, die über die funktionale Natur der KI hinausgehen, berücksichtigt werden, so beziehen sich diese in der Regel darauf, wie eine KI die Arbeit eines Managers bedroht oder unterstützt.

Verständnis von Vertrauen mit KI

Um zu verstehen, warum Menschen in diesen unterschiedlichen Kontexten Vertrauen zu einer KI oder zu verwandten Akteuren haben oder entwickeln, gibt es eine ganze Reihe von The-



Picture by Freepik

orien. Eines der Standardmodelle beschreibt, dass Fähigkeiten, Wohlwollen und Integrität zu Vertrauen führen. Vertrauen, das durch diese Antezedenzen genährt wird, kann dispositionaal, situationsbedingt und erlernt sein. Weitere Modelle, die an die KI angepasst wurden, beschreiben eher Nuancen der konzeptionellen Unterschiede, aber die zugrunde liegende Logik bleibt ähnlich. Der Mensch versucht, das Verhalten einer KI zu verstehen, ähnlich wie er andere Menschen versteht. Auch diese weisen bestimmte Merkmale auf und bleiben oft Black Boxes. Bei der Beurteilung von KI ist es schwieriger, auf Vorerfahrungen mit Menschen zurückzugreifen, und die Situation ist oft mehrdeutig. Und genau hier liegt die Hauptneuheit bei der KI - die spezifischen Parameter, die zwischen Menschen, Maschinen und lernenden Maschinen unterschiedlich bewertet werden. Menschen versuchen oft, Vorhersagen über das KI-Verhalten zu treffen, die sie aus früheren Erfahrungen mit Menschen ableiten, sie haben Annahmen aus Filmen, wenn deutlich hervorgehoben wird, dass eine KI handelt, oder sie vergleichen die KI-Leistung mit der menschlichen (eigenen) Leistung. Daher ist es notwendig, die spezifischen Parameter zu identifizieren, die sich aus der funktionalen Natur der KI, ihrem Aussehen oder Verhalten ergeben. Die Kategorien und Beziehungen der grundlegenden Vertrauensmodelle geben Aufschluss über die zugrunde liegende Logik der Vertrauensbildung, die denselben Mustern folgt wie bei Menschen.

Experten führen eine solche Bewertung auf analytischere Weise durch und konzentrieren sich stärker auf die verschiedenen Aspekte der Daten, Algorithmen und beteiligten Unternehmen. Letztendlich wird eine zentrale Bewertung für eine Anwendung vorgenommen, während

die Bewertung für ein Netz von menschlichen Akteuren, die die Aufgabe ausführen, eher verstreut wäre. Im letzteren Fall müsste auch das menschliche Verhalten gegenüber den verschiedenen Akteuren ermittelt werden, um das Vertrauen in die erbrachten Dienste zu bewerten, während das Ergebnis eines solchen Verhaltens im KI-Design verankert ist und als solches getestet werden kann. Dies erfordert andere Verfahren, bietet aber auch ein größeres Potenzial an Beweisen, die analytisch geprüft werden können.

Fazit

Während der grundlegende Mechanismus, wie Vertrauen entsteht, derselbe bleibt, können die Parameter der Bewertung eine unterschiedliche Relevanz und Priorität für Menschen und KI haben. Der KI fehlen einige Merkmale, die der Mensch in Bezug auf das Erscheinungsbild hat, aber dies stellt unterschiedliche Merkmale in den Details oder Schwellenwerten dar, die innerhalb der bestehenden Mechanismen zu berücksichtigen sind. Hier kann die Neuartigkeit erforscht werden, um zu verstehen, warum und wann Menschen mit unterschiedlichem Hintergrund KI vertrauen.

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