AUS DER FORSCHUNG

Vertrauen in Algorithmen verändert sich mit der Aufgabe

Algorithmusvertrauen hängt von der Aufgabe ab: Prüfe Objektivität, beobachtete Fehler, Beratungsquelle, Expertise und Nutzung vor dem Urteil.

Sinan Isoglu, MBA (Quantic)
Führungskraft für Umsatzwachstum, Dozent und Doktorand

VERÖFFENTLICHT
1. September 2026

LESEZEIT
8 Min.

WÖRTER
1.711

https://isoglu.com/de/insights/journal/algorithm-vertrauen-aendert-sich-mit-der-aufgabe/

INHALT

MANAGEMENT SUMMARY

Menschen können algorithmische Hinweise in einem Kontext schätzen, sich nach einem sichtbaren Fehler von einem Algorithmus abwenden und Hinweise je nach objektiv oder subjektiv erlebter Aufgabe unterschiedlich nutzen. Castelo, Bos und Lehmann zeigen, dass die wahrgenommene Objektivität einer Aufgabe die Algorithmusnutzung in Labor- und Feldstudien verändert. Dietvorst, Simmons und Massey zeigen, dass ein beobachteter Fehler Vertrauen und spätere Nutzung eines Algorithmus auch dann senken kann, wenn der Algorithmus zuvor besser als ein Mensch war. Logg, Minson und Moore dokumentieren Algorithmuspräferenz, zeigen aber Grenzen bei einem Vergleich mit der eigenen Schätzung und bei Prognoseexpertise. Der Beitrag führt Aversion und Präferenz als Randbedingungen zusammen und entwickelt eine Prüfkarte für aufgabenbezogenes Vertrauen. Er bewertet kein aktuelles Produkt oder Modell.

Schlagwörter: Vertrauen in Algorithmen · Algorithmusaversion · Algorithmuspräferenz · Entscheidungsunterstützung

Algorithmusvertrauen ist keine stabile Haltung. Es verändert sich mit der Aufgabe, dem beobachteten Fehler und der Person, die entscheidet.

Die kurze Antwort lautet: Ein System sollte nicht als vertrauenswürdig oder nicht vertrauenswürdig bezeichnet werden, ohne die Entscheidungsgrenze zu nennen. Menschen können algorithmische Hinweise eher bei einer objektiv erlebten Aufgabe nutzen, sich nach einem sichtbaren Fehler von einem Algorithmus abwenden und in einer anderen Situation algorithmische Hinweise menschlichen Hinweisen vorziehen. Diese Befunde widersprechen sich nicht, wenn Aufgabenobjektivität, Erfahrung, Vergleichspunkt und Expertise als Bedingungen behandelt werden und nicht in einer einzigen Vertrauenszahl verschwinden.

Castelo, Bos und Lehmann zeigen, dass die wahrgenommene Objektivität einer Aufgabe die Algorithmusnutzung verändert. Dietvorst, Simmons und Massey zeigen eine spezifische Abwendung, nachdem Personen einen Fehler des Algorithmus beobachtet haben. Logg, Minson und Moore zeigen in mehreren Schätz- und Prognosesituationen eine Präferenz für algorithmische Hinweise, mit schwächerer Präferenz beim Vergleich mit der eigenen Schätzung oder bei Prognoseexpertise. Zusammen begründen die Quellen eine Karte für aufgaben- und erfahrungsabhängige Nutzung, keine universelle Regel.

Warum muss die Systembewertung bei der Aufgabenstruktur und nicht beim KI-Label beginnen?

Stell dir dieselbe Entscheidungshilfe für zwei Fragen vor. Die erste verlangt eine numerische Prognose mit definiertem Ergebnis und historischem Vergleich. Die zweite verlangt eine Einschätzung von Geschmack, Passung oder emotionaler Reaktion. Die Oberfläche ist gleich geblieben. Der Entscheidungskontext nicht.

Vor der Frage, ob Nutzende dem Algorithmus vertrauen, sollten diese Felder benannt werden:

Tabelle 1 Warum muss die Systembewertung bei der Aufgabenstruktur und nicht beim KI-Label beginnen?

Feld	Frage	Warum es zählt
Aufgabencharakter	Wird die Aufgabe als objektiv, subjektiv oder gemischt erlebt?	Die wahrgenommene Objektivität verändert die Bereitschaft, algorithmische Hinweise zu nutzen.
Beratungsquelle	Ist der Hinweis als algorithmisch, menschlich oder unbekannt gekennzeichnet?	Die Quellenbezeichnung kann die Befolgung unabhängig vom Zahlenwert verändern.
Beobachtete Leistung	Wurde ein Fehler des Systems gesehen?	Ein sichtbarer Fehler kann die Nutzung verändern, auch wenn die relative Leistung besser bleibt.
Vergleichspunkt	Wird der Hinweis mit einer eigenen Schätzung oder mit keiner Alternative verglichen?	Präferenz kann schwächer werden, wenn das eigene Urteil der direkte Konkurrent ist.
Expertise	Verfügt die entscheidende Person über Prognose- oder Fachexpertise?	Expertise kann die Stärke der Befolgung verändern.
Konsequenz	Was geschieht, wenn der Hinweis befolgt oder ignoriert wird?	Nutzung ist eine Entscheidung und nicht nur eine Einstellung.

Tabelle aus diesem Essay. Quellen und Einordnung stehen im Beitrag.

Diese Felder verhindern einen Kategorienfehler. Eine Person kann einem Algorithmus bei einer numerischen Prognose vertrauen und ihn bei einer subjektiven Empfehlung ablehnen. Sie kann einen gekennzeichneten Algorithmus im Experiment schätzen und nach einem auffälligen Fehler weniger nutzen. Keine dieser Beobachtungen beweist eine allgemeine Haltungsänderung über alle Aufgaben hinweg.

Abbildung 1 Die Karte für aufgabenbezogenes Algorithmusvertrauen

Prüffeld	Beispielgrenze	Zulässige Interpretation	Stoppsignal
Aufgabencharakter	Objektiv, subjektiv oder gemischt	„Nutzung wird für diese Aufgabe geprüft.“	Von einer Aufgabe wird auf alle Aufgaben geschlossen.
Beobachtete Leistung	Kein beobachteter Fehler oder sichtbarer Fehler	„Diese Erfahrung kann die Nutzung beeinflussen.“	Ein Fehler gilt als Beweis allgemeiner Unterlegenheit.
Beratungsquelle	Algorithmisch, menschlich oder unbekannt bezeichnet	„Die Quellenbezeichnung gehört zum Entscheidungskontext.“	Quellenbezeichnungen fehlen, während Einstellungen verglichen werden.
Expertise und Vergleich	Eigene Schätzung, menschlicher Vergleich oder kein direkter Konkurrent	„Die Reaktion ist durch Vergleich und Expertise begrenzt.“	Novizen und Experten werden gleich behandelt.
Freigabefrage	Welche Nutzungsentscheidung steht an?	„Nutzen, prüfen, überstimmen oder testen.“	„Nutzende vertrauen dem Algorithmus“ ist das Endergebnis.

Eigene Entscheidungslogik auf Grundlage von Castelo, Bos und Lehmann (2019), Dietvorst, Simmons und Massey (2015) sowie Logg, Minson und Moore (2019). Die Fragen sind synthetisch und bewerten kein aktuelles System.

Wie verändert die wahrgenommene Aufgabenobjektivität das menschliche Vertrauen in Algorithmen?

Castelo, Bos und Lehmann berichten, dass die Algorithmusnutzung in vier Online-Laborstudien mit mehr als 1.400 Teilnehmenden und zwei Online-Feldstudien mit mehr als 56.000 Teilnehmenden mit der wahrgenommenen Objektivität variierte. Konsumentinnen und Konsumenten vertrauten Algorithmen bei als subjektiv erlebten Aufgaben weniger und nutzten sie weniger als bei als objektiv erlebten Aufgaben. Das ist eine Randbedingung der Nutzung, keine Aussage, dass alle subjektiven Aufgaben Algorithmen ablehnen oder alle objektiven Aufgaben sie akzeptieren.

Die Autoren berichten außerdem, dass eine stärkere wahrgenommene Objektivität oder eine stärkere affektive Menschenähnlichkeit des Algorithmus Vertrauen und Nutzung bei ansonsten subjektiven Aufgaben erhöhte. Das ist relevant, weil Teams eine Aufgabe anhand ihrer Datenform als objektiv beschreiben, während Nutzende sie als interpretativ erleben. Eine Bewertung, ein Ranking oder eine Prognose kann numerisch aussehen und doch Annahmen über Geschmack, Passung, Fairness oder Identität enthalten.

Die praktische Folgerung ist, den erlebten Aufgabencharakter vor der Interpretation einer Nutzungsquote zu besprechen oder zu messen. Wenn Erklärung, Oberfläche oder soziale Rahmung verändert werden, kann sich auch die Wahrnehmung der Aufgabe verändern. Das bedeutet nicht, dass der Algorithmus genauer geworden ist. Es bedeutet, dass sich die Nutzungsbedingungen verändert haben.

Wie lösen beobachtbare Systemfehler eine überproportionale Algorithmen-Aversion aus?

Dietvorst, Simmons und Massey berichten in fünf Studien eine Abwendung von Algorithmen, nachdem Teilnehmende die Leistung algorithmischer Prognostiker beobachtet hatten. Teilnehmende verloren nach dem gleichen Fehler schneller das Vertrauen in einen Algorithmus als in einen Menschen, selbst wenn der Algorithmus zuvor besser abgeschnitten hatte. Die Beobachtung, wie ein Algorithmus arbeitet, senkte auch die Wahrscheinlichkeit, dass Teilnehmende zukünftige Anreize an den Algorithmus statt an einen schlechteren menschlichen Prognostiker knüpften.

Dieser Befund beschreibt eine bestimmte Erfahrung: Die entscheidende Person sieht einen Fehler des Systems. Er zeigt nicht, dass Algorithmen niemals genutzt werden sollten, und nicht, dass jeder Fehler dieselbe Wirkung hat. Sichtbarkeit, Größe des Fehlers, Aufgabenrisiko und der verfügbare Vergleich gehören in den Prüfdatensatz.

Für die Gestaltung bedeutet das: Ein Team kann ein System gegen einen menschlichen Vergleich bewerten, die Kalibrierung beobachten und nach Fehlern eine Prüfung vorsehen. Diese Kontrollen sind kein Beleg dafür, dass das System vertraut wird. Sie machen Nutzung und Korrektur beobachtbar. Ein Freigabetext sollte sagen, ob das System genutzt, geprüft, überstimmt oder nach einem Fehler getestet wird.

Der stärkste nicht gedeckte Satz wäre: „Nutzende verlassen Algorithmen, wenn diese Fehler machen.“ Der zulässige Satz ist enger: In den berichteten Studien senkte die Beobachtung algorithmischer Fehler Vertrauen und Nutzung stärker als die Beobachtung vergleichbarer menschlicher Fehler, obwohl der Algorithmus besser geleistet hatte. So bleibt die Grenze der Quelle erhalten.

Unter welchen Bedingungen zeigen Anwender eine ausgeprägte Algorithmen-Präferenz?

Logg, Minson und Moore berichten in sechs Experimenten eine Präferenz für Algorithmen. Laien folgten Hinweisen eher, wenn sie glaubten, diese stammten von einem Algorithmus und nicht von einer Person. Die berichteten Kontexte umfassen numerische Schätzungen und Prognosen, darunter Urteile über die Popularität von Liedern und romantische Anziehung.

Das kann wie das Gegenteil von Algorithmusaversion klingen, bis der Vergleich sichtbar wird. In den Präferenzstudien geht es häufig darum, ob Personen dem Hinweis einer gekennzeichneten Quelle folgen. In den Aversionstudien geht es darum, wie Personen reagieren, nachdem sie einen Fehler des Algorithmus gesehen haben. Quellenbezeichnung, Leistungserfahrung und Aufgabenrahmung sind unterschiedliche Bedingungen.

Logg und Kollegen berichten auch Grenzen. Die Präferenz für Algorithmen wurde schwächer, wenn Personen die Schätzung des Algorithmus mit ihrer eigenen Schätzung verglichen oder wenn sie Prognoseexpertise hatten. Der direkte Konkurrent kann daher das eigene Urteil und nicht ein allgemeiner menschlicher Prognostiker sein. Eine Person kann einem Algorithmus mehr vertrauen als einem unbekannten menschlichen Durchschnitt und ihn trotzdem überstimmen, wenn er einer gut begründeten Expertenschätzung widerspricht.

Die Prüfung muss daher den Vergleichspunkt festhalten. „Algorithmische Hinweise wurden bevorzugt“ ist unvollständig, wenn nicht gesagt wird, was bevorzugt wurde, bei welcher Aufgabe, mit welcher Expertise und nach welcher Erfahrung.

Warum sind Aversion und Wertschätzung kontextabhängige Reaktionen statt universeller Gesetze?

Die drei Quellen lassen sich als einfache Sequenz lesen:

1. Die Aufgabe wird als objektiv oder subjektiv gerahmt.
2. Die Beratungsquelle wird bezeichnet oder bleibt unklar.
3. Die entscheidende Person beobachtet die Leistung, einschließlich sichtbarer Fehler.
4. Der Hinweis wird mit einem Menschen, der eigenen Schätzung oder keiner direkten Alternative verglichen.
5. Expertise und Risiko beeinflussen, ob die Person folgt, prüft oder überstimmt.

Diese Sequenz erklärt, warum eine einzelne Vertrauensbefragung ein schwaches Freigabekriterium ist. Sie verdichtet Aufgabe, Vorgeschichte, Vergleich und Expertise in eine Zahl. Ein hoher Wert kann bei einer risikoreichen Aufgabe mit geringer Nutzung einhergehen. Ein niedriger Wert kann mit einem nützlichen algorithmischen Beitrag vereinbar sein, wenn Nutzende ihn prüfen und korrigieren müssen.

Die Karte liefert damit einen besseren Test für einen neuen Entscheidungsprozess. Soll die vorgesehene Person dem Hinweis automatisch folgen, ihn vor der Handlung prüfen, ihn als einen von mehreren Inputs nutzen oder ihn als Vergleichswert behandeln? Das sind unterschiedliche Nutzungsregeln. Sie sollten nicht hinter dem Wort Vertrauen verschwinden.

Welche Governance-Sequenz steuert den Rollout automatisierter Entscheidungssysteme?

Nutze diese Sequenz, wenn geprüft wird, ob eine Aussage über Nutzung belegt ist:

1. Definiere die Aufgabe und halte fest, ob sie objektiv, subjektiv oder gemischt erlebt wird.
2. Benenne den Vergleich: menschliche Prognose, eigene Schätzung oder kein direkter Konkurrent.
3. Halte fest, ob ein Fehler beobachtet wurde und wie er gezeigt wurde.
4. Trenne Befolgung von Genauigkeit, Kalibrierung und Ergebnisqualität.
5. Beschreibe relevante Expertise, statt von einer einheitlichen Reaktion auszugehen.
6. Entscheide, ob die Regel automatische Nutzung, Prüfung, Übersteuerung oder einen kontrollierten Test vorsieht.
7. Schreibe den stärksten gedeckten Satz und den stärkeren universellen Satz, der außerhalb der Evidenz bleibt.

Das ist kein Rezept, um Vertrauen herzustellen. Es hält das Entscheidungsobjekt sichtbar. Ein Team kann eine Prüfung vor der Handlung wählen, auch wenn Algorithmuspräferenz hoch ist. Es kann einen nützlichen Algorithmus in einer subjektiven Situation testen, ohne zu behaupten, die Aufgabe sei objektiv geworden.

Wo liegen die empirischen Grenzen der Forschung zur Mensch-Algorithmus-Interaktion?

Die drei Quellen bewerten kein aktuelles Produkt, Modell, Teammitglied oder organisatorisches Einsatzfeld. Sie begründen keine universelle Eigenschaft des Algorithmusvertrauens, keine garantierte Reaktion auf jeden Fehler und keine allgemeine Regel, dass Algorithmen menschliches Urteil ersetzen sollten. Sie zeigen, dass wahrgenommene Aufgabenobjektivität, beobachtete Fehler, Quellenbezeichnungen, Vergleichspunkte und Expertise die Nutzung in den berichteten Situationen begrenzen können.

Die Stopregel ist konkret: Gib „Nutzende vertrauen dem Algorithmus“ erst frei, wenn Aufgabe, beobachtete Leistung, Vergleichspunkt, Expertise und Nutzungsregel benannt sind. Fehlt eines davon, kann das System trotzdem genau oder nützlich sein. Die Vertrauensaussage ist noch nicht ausreichend begrenzt.

Die Aufgabenabgrenzung verbindet sich mit hybrider Intelligenz und ihrer Fähigkeitsgrenze und dem Kundenmodell, das nicht jedes Ergebnis prognostizieren kann.

Castelo, N., M. W. Bos und D. R. Lehmann. (2019). Task-Dependent Algorithm Aversion. Journal of Marketing Research, 56(5), 809-825. DOI

Dietvorst, B. J., J. P. Simmons und C. Massey. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. Journal of Experimental Psychology: General, 144(1), 114-126. DOI

Logg, J. M., J. A. Minson und D. A. Moore. (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. Organizational Behavior and Human Decision Processes, 151, 90-103. DOI

ENDE DES BEITRAGS

ZITIERWEISE

Isoglu, S. (2026, 1. September). Vertrauen in Algorithmen verändert sich mit der Aufgabe. Sinan Isoglu. https://isoglu.com/de/insights/journal/algorithm-vertrauen-aendert-sich-mit-der-aufgabe/

WEITERLESEN

Aus der Forschung
Was ist Interferenz? Wenn die Behandlung einer Einheit ein anderes Ergebnis verändert
https://isoglu.com/de/insights/journal/was-ist-interferenz/

Aus der Forschung
Mehr Werbebeobachtungen garantieren keine besseren Entscheidungen
https://isoglu.com/de/insights/journal/mehr-werbebeobachtungen-garantieren-keine-besseren-entscheidungen/

Aus der Forschung
Conjoint-Analyse schätzt Präferenzen, nicht Nachfrage
https://isoglu.com/de/insights/journal/conjoint-analyse-schaetzt-praeferenzen-nicht-nachfrage/

Sinan Isoglu, MBA (Quantic)
Führungskraft für Umsatzwachstum, Dozent und Doktorand