

Die KI hat entschieden, dass...Aber hat die KI wirklich entschieden?

Erarbeitet von

Prof. Susanne Hahn, Marie von Lobenstein M.A.

Disclaimer: Das Thema KI-„Entscheidungen“ ist äußerst komplex und wird aus verschiedenen Fachdisziplinen und Perspektiven betrachtet. In diesem Themenblock haben wir daher eine Auswahl getroffen, um einen ersten Überblick zu geben. Eine vollständige Behandlung des Themas würde den Rahmen der Lehrveranstaltung sprengen und ist im Rahmen dieses Kurses nicht möglich.

Lernziele	1
Inhalt	2
Einstieg.....	2
Was sind Entscheidungen?	2
„Entscheiden“ KI-Systeme denn?	3
Die ethische Perspektive.....	5
Take-Home Message	5
Weiterführendes Material.....	6
Disclaimer	6

Lernziele

- Du kannst definieren, was eine Entscheidung ist
- Du kannst einordnen, ob KI-Systeme entscheiden
- Du kannst erklären, wieso der Begriff „Entscheidung“ kritisch betrachtet werden sollte

Inhalt

Einstieg

Herzlich willkommen zu diesem Video über Entscheidungen und die Frage, ob KI-Systeme entscheiden. Die Nutzung von Ausdrücken wie „Algorithmen entscheiden“ oder „Künstliche Intelligenz trifft Entscheidungen“ deutet auf eine faszinierende Verschiebung hin: Begriffe, die einst untrennbar mit menschlichem Handeln verbunden waren, werden nun auf den Bereich der Maschinen und Algorithmen übertragen. Wir haben in einem anderen Video bereits über anthropomorphe Sprache in der KI-Wissenschaft und Berichterstattung gesprochen. In diesem Video gehen wir jedoch über die rein sprachliche Ebene hinaus und werfen grundlegende Fragen auf: Welche impliziten Annahmen, Assoziationen und Bedeutungszusammenhänge werden transportiert, wenn menschliche Handlungsphänomene auf Künstliche Intelligenz übertragen werden? Und welche potenziellen Missverständnisse oder Irreführungen könnten daraus resultieren? In diesem Video werden wir uns mit diesen Fragen auseinandersetzen und die komplexen Dynamiken erkunden, die entstehen, wenn die Sprache und Konzepte des menschlichen Handelns auf die Welt der künstlichen Intelligenz angewendet werden.

Hierfür habe ich Prof. Dr. Hahn, Professorin der HHU und Expertin im Bereich Digitalisierung und Künstliche Intelligenz, interviewt. Sie wird uns in diesem Video eine handlungsphilosophische Perspektive auf dieses Thema eröffnen.

Bei der Diskussion darüber, ob KI Entscheidungen trifft, ist es wichtig, zu definieren, was wir unter „Entscheidungen“ verstehen, und KI-Entscheidungen mit menschlichen Entscheidungen zu vergleichen.

Was sind Entscheidungen?

Frage: Frau Prof. Hahn, was ist das Verständnis von „Entscheidung“, mit dem Sie auf KI blicken?

Ich versuche zu ermitteln, welches Verständnis von Entscheidung unterstellt wird, wenn in einer bestimmten Weise über KI gesprochen wird. Z.B. wenn gesagt wird, wie problematisch die sogenannte Autonomie der Algorithmen ist oder welche Macht Algorithmen bei Entscheidungen zukommt, und stelle demgegenüber, wie KI im Sinne von Algorithmen maschinellen Lernens funktioniert.

Nehmen wir die Rede von autonom entscheidenden Algorithmen. Hier wird ein Ausdruck, den wir üblicherweise für menschliches Handeln verwenden, auf Algorithmen angewendet, und zwar in einer Weise, die ich als ein **emphatisches** Verständnis bezeichnen möchte.

Diese Auffassung von Entscheidung, wie sie beispielsweise im historischen Wörterbuch der Philosophie aufgeführt wird zeichnet sich dadurch aus, „dass etwas Ungewisses, Zweifelhafte zur Klärung kommt“; Entscheidungen finden in einer Situation der Unsicherheit statt, die entscheidende Person steht unter Handlungsdruck, hat Handlungsspielräume und es fehlen ausreichende Gründe so und nicht anders zu handeln. Passende Beispielszenarien: Eine Person entscheidet, ob eine lebensverlängernde Maßnahme durchgeführt wird oder eine Person entscheidet, ob am nächsten Tag ein Angriff gegen eine feindliche Stellung gestartet wird.

Neben diesem emphatischen Verständnis von Entscheidung gibt es aber auch eine andere Sicht auf Entscheidungen, wie sie in der Entscheidungstheorie entwickelt wird. Dort werden drei Arten von Entscheidung unterschieden: Entscheidungen unter Sicherheit, unter Risiko und unter Unsicherheit. Während man bei Entscheidungen unter Risiko Wahrscheinlichkeiten für das Eintreten einer Konsequenz angeben kann, geht dies bei Entscheidungen unter Unsicherheit eben nicht. Man könnte sagen, dass das eben geschilderte emphatische Verständnis von Entscheidung in dieser Klassifikation als eine Entscheidung unter Unsicherheit betrachtet werden kann.

Entscheidungen unter Unsicherheit sind unangenehm. Wir streben daher danach, für möglichst viele wichtige Entscheidungen allgemeine Zusammenhänge zwischen einer Handlung und einer Handlungsfolge zu bestimmen. Das Betreiben von Wissenschaft ist wesentlich auf dieses Ziel ausgerichtet. Doch auch in der Lebenswelt wollen wir wissen, wie wahrscheinlich es ist, dass beispielsweise jemand einen gewährten Kredit zurückzahlt. Lassen sich Merkmale bestimmen, die die Rückzahlung wahrscheinlicher machen, z. B. Berufszugehörigkeit, Immobilienbesitz etc.? Wenn man einen solchen allgemeinen Zusammenhang hat, kann man diesen zur Erstellung von Prognosen nutzen: Wie wahrscheinlich ist es, dass **diese** Person mit den und den Merkmalen den Kredit zurückzahlt?

Wenn man dann noch über eine Bewertung der möglichen Konsequenzen verfügt, d. h. als wie gravierend man einen Kreditausfall betrachtet oder wie sehr man an einem Abschluss des Darlehens wegen der Zinszahlungen interessiert ist, dann hat man Gründe für oder gegen eine Gewährung eines Kredits. Mit der Entscheidungsregel, dass der Erwartungswert zu maximieren ist, ist klar, was zu tun ist.

Solche Handlungssituationen fallen somit nicht unter das geschilderte emphatische Verständnis von Entscheidung. Beim emphatischen Verständnis wird gerade der Umstand als wesentlich hervorgehoben, dass es keinen Grund gibt, so und nicht anders zu handeln. Die entscheidende Person wählt vielmehr unter Wahrnehmung ihres Ermessensspielraums und ihrer Autonomie eine Handlung.

„Entscheiden“ KI-Systeme denn?

Frage: Nach diesem Verständnis von Entscheidung: „entscheiden“ Algorithmen/KI?

Dazu muss man sich anschauen, wie Algorithmen eingesetzt werden. Der Prozess der Entwicklung und Anwendung von Algorithmen, die als „Entscheidungsalgorithmen“ angesprochen werden, lässt sich in fünf Schritten erfassen: Generierung des Daten-Inputs, Datenaufbereitung, Erstellung des Voraussagemodells, Erstellung von Wenn-Dann-Beziehungen und Generierung des Outputs.

Die ersten beiden Schritte sind nicht spezifisch für die Algorithmen maschinellen Lernens, sondern gehören standardmäßig zu statistischen Verfahren. Im dritten Schritt wird ein Voraussagemodell erstellt, z. B. durch künstliche neuronale Netze oder durch logistische Regression. Mit einem solchen Voraussagemodell können neue Fälle, z. B. eine neue Anfrage für einen Kredit, eingeschätzt werden. Resultat der Einschätzung ist ein Wahrscheinlichkeitswert für eine bestimmte Konsequenz, z. B. die vollständige Rückzahlung eines Kredits. Bis zu diesem Schritt ist „noch nichts passiert“, d. h. es wurden lediglich Prognosen erstellt.

Erst mit der Aufstellung besonderer Wenn-Dann-Verknüpfungen werden Maßnahmen eingeleitet: **Wenn** eine Person einen Score von so-und-so-viel hat, d. h. wenn die Wahrscheinlichkeit für die Kreditrückzahlung wenigstens ein bestimmtes Maß hat, **dann** wird eine Maßnahme durchgeführt. Im Beispiel wäre der Output eine Finanzierungszusage. Mit dem Einsatz des Algorithmus wird eine **Sortierung von Anfragen** vorgenommen: Unterhalb des festgelegten Wahrscheinlichkeitswertes bekommt die anfragende Person keinen Kredit, ab dem festgelegten Wert bekommt sie ihn. Dieses Vorgehen unterscheidet sich von einer Erwägung, wie sie beispielsweise bei ermächtigten Personen vorkommt. Hier sind die entscheidenden Personen nicht im Vorhinein vollständig festgelegt, sondern können von Leitlinien abweichen.

Was bedeutet das für die Frage, ob Algorithmen entscheiden? Wenn man ein **emphatisches** Verständnis von Entscheidung unterstellt, dann sind die einzelnen Sortierungsfälle, die durch Algorithmen vorgenommen werden, **keine Entscheidungen in diesem Sinne**. „Der Algorithmus“ ist kein autonomer Akteur, der unter Ausnutzung von Ermessensspielräumen Entscheidungen trifft und dem man auch prima facie „unsinnige“ Entscheidungen nachsehen würde, wenn er sie situativ erklärt.

Die Rede von autonom entscheidenden Algorithmen verdeckt, dass mit der **Herstellung** einer Wenn-Dann-Verknüpfung eine **Entscheidung** gefällt wird, die die Sortierung der zukünftigen Fälle festlegt. *Ex ante* wird über die Annahme eines Voraussagemodells und – insofern ein Output automatisiert verknüpft wird – über eine Verknüpfung entschieden. Hier handelt es sich um eine Entscheidung im emphatischen Sinn: Wenn man einen Wahrscheinlichkeitswert festlegt, unterhalb dessen keine Kreditvergabe erfolgt, dann weiß man nicht, welche Einzelfälle, die man bei einer Einzelfallbetrachtung vielleicht noch gerade zu den positiv beschiedenen Fällen gezählt hätte, der Grenzziehung zum Opfer fallen; bzw. wer noch zu den positiv beschiedenen Fällen zählt, aber besser in die andere Gruppe gefallen wäre.

Der Einsatz eines Algorithmus mit einer Maßnahmenverknüpfung stellt somit eine Entscheidung dar, für die die **entscheidende Person auch verantwortlich ist**.

Die ethische Perspektive

Frage: Wieso ist es wichtig, dass wir das Konzept der „Entscheidung“ aus dieser ethisch/geisteswissenschaftlichen Perspektive betrachten?

Die Analyseskizze zeigt, dass eine unreflektierte Übertragung auf Algorithmen ein emphatisches Verständnis von Entscheidung mitbringt, das nicht der tatsächlichen Arbeitsweise von Algorithmen als nachgeordneten Sortierungsmechanismen entspricht. Hingegen ist die **Entscheidung über den Einsatz eines Algorithmus** eine folgenreiche Entscheidung, die mit Verantwortlichkeiten einhergeht. Diese Verantwortungsebene wird aufgedeckt.

Konstruktiv gewendet heißt das: Wenn man eine Perspektive einnimmt – bestenfalls von informatischer Expertise begleitet – die die Arbeitsweise von Algorithmen und ihre Leistungsfähigkeit und Limitationen einerseits und die Analyse unterschiedlicher Entscheidungstypen andererseits umfasst, dann wird damit der Raum eröffnet für die Frage: Welche Abläufe und Handlungssituationen eignen sich für algorithmische Sortierungen?

Take-Home Message

Die Rede von „Algorithmen, die entscheiden“ stellt eine unreflektierte Übertragung der Entscheidungsrede auf Algorithmen dar. Wenn wir davon sprechen, dass „Algorithmen Entscheidungen treffen“, übertragen wir damit ein Verständnis von Entscheidungen, das nicht der Funktionsweise von KI entspricht. Algorithmen sind eher wie Werkzeuge, die Informationen sortieren, anstatt wie Menschen, die Entscheidungen treffen. Wenn wir Algorithmen als Entscheider betrachten, vergessen wir oft, dass Menschen dafür verantwortlich sind, wie und wann diese Algorithmen verwendet werden. Es ist wichtig, dass wir die Arbeitsweise von Algorithmen verstehen und ihre Stärken und Schwächen realistisch einschätzen.

Weiterführendes Material

Claus von Bormann (1972): «Entscheidung», in: J. Ritter (Hg.): Historisches Wörterbuch der Philosophie, Basel: Schwabe Verlag. DOI: 10.24894/HWPh.876

Hahn, S. (2024, January). Algorithmische „Entscheidungen“ in der Medizin? Eine Reflexion zu einem handlungsbezogenen Ausdruck. In *Der Einsatz von KI & Robotik in der Medizin* (pp. 13-26). Nomos Verlagsgesellschaft mbH & Co. KG.

Ruschmeier, H., & Steinrötter, B. Der Einsatz von KI & Robotik in der Medizin. DOI: doi.org/10.5771/9783748939726-13

Disclaimer

Transkript zu dem Video „KI und Ich: KI-„Entscheidungen““, Marie von Lobenstein. Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.