

KI und Entscheidungen

Erarbeitet von
Dr. Maike Mayer

Lernziele	1
Inhalt	2
Einstieg	2
Warum mit Maschinen entscheiden?	2
Exkurs: Menschliche Entscheidungsfindung	3
Ansätze für eine gute Mensch-KI-Interaktion	4
Wo KI bei Entscheidungen helfen kann und wo (besser) nicht	5
Fazit	6
Quellen	7
Disclaimer	8

Lernziele

- Du kannst KI-gestützte Entscheidungen differenziert einordnen:
 - Du kannst exemplarisch Ansätze für die Gestaltung einer guten Mensch-KI-Interaktion aufzeigen
 - Du kannst mögliche Vorteile des Einsatzes von KI-Systemen bei Entscheidungen benennen
 - Du kannst einschätzen, wo KI-Systeme in der Entscheidungsfindung eingesetzt werden können und wo eher nicht
- Du kannst exemplarisch Biases und Heuristiken der menschlichen Entscheidungsfindung benennen

Inhalt

Einstieg

Immer mal wieder geistern Schlagzeilen durch die Medien, dass der Einsatz von Automation oder KI-Systemen schief gegangen ist, teilweise sogar katastrophal: Der fehlerhafte Einsatz von ChatGPT eines Rechtsanwalts [1], ein Bewerbungsalgorithmus, der Frauen benachteiligt [2], Gesichtserkennungssoftware, die Gesichter mit dunkler Hautfarbe nicht erkennt [3], oder der fatale Unfall eines autonom fahrenden Fahrzeugs mit einer Passantin [4].

Quelle [1] [2] [3] [4]

Einblendung: Icons (Zeitung, Smartphone, Fernseher, Waage, Papierstapel, Gesicht, Auto)

Nicht zuletzt an solchen – teils drastischen – Beispielen wird deutlich, dass unsere Interaktion mit KI-Systemen nicht immer reibungslos funktioniert. Da kann man sich vielleicht auch schon mal die Frage stellen, warum wir überhaupt mit Hilfe von KI-Systemen entscheiden ...

Einblendung: Icon (Figur mit Fragezeichen)

Warum mit Maschinen entscheiden?

Auch wenn der KI-Einsatz schief gehen kann, kann es durchaus gute Gründe geben, mit der Unterstützung von KI-Systemen zu entscheiden. Schauen wir uns dafür doch mal ein paar exemplarische Punkte an [5].

Quelle [5]

Einblendung: Icon (Figur mit Idee, Lupe)

Gerade was große Datenmengen angeht, sind KI-Systeme leistungsfähiger als wir. Sie können sehr viel größere Datenmengen in kürzerer Zeit verarbeiten als wir es je könnten. Und genau das ermöglicht KI-Systemen, komplizierte Muster zu entdecken, die uns vielleicht entgangen wären. Eine wichtige Voraussetzung dafür ist natürlich, dass das KI-System ausreichend Daten zur Verfügung hat. Aber KI-Systeme können nicht nur kompliziertere Muster entdecken als wir, sondern auch subtilere Muster, die uns vielleicht entgangen wären. Wir brauchen nämlich recht starke Zusammenhänge, um Verbindungen zu erkennen. Hier müssen wir allerdings aufpassen, dass entdeckte Muster wirklich bedeutsam sind und nicht nur zufällige Datenvariationen.

Einblendung: Icon (Papierstapel, Pfeil, Uhr, Netzwerk, Glühbirne, Detektiv*in, Ausrufezeichen)

Die großen Datenmengen, die zum Trainieren von KI-Systemen verwendet werden, haben einen weiteren potentiellen Vorteil: mehr Feedback. Die Daten, mit denen trainiert wird, enthalten jede Menge Rückmeldungen darüber, wie bestimmte Fälle ausgegangen sind: Wurde ein Studium abgebrochen oder nicht? Oder stimmt die gegebene Diagnose? KI-Systeme haben dadurch deutlich mehr Vergleichsfälle zur Verfügung als wir vor einer Entscheidung. Zumal wir nach unserer Entscheidung oft auch keine Rückmeldung bekommen, ob sie richtig war oder wie ähnliche Fälle ausgegangen sind, um daraus zu lernen.

Einblendung: Icons (Figur mit Idee, Daumen runter, Daumen hoch, Fragezeichen); Schlagwort („Feedback“)

Wird außerdem klar, woran der Fehler eines KI-Systems gelegen hat, haben KI-Systeme gegenüber uns Menschen einen weiteren Vorteil: Der identifizierte Fehler lässt sich schnell und flächendeckend beheben, indem man das KI-System anpasst. Automobilhersteller könnten einen Fehler beispielsweise in allen betroffenen Modellen durch ein Softwareupdate beheben oder Banken direkt in allen Filialen ihres Unternehmens. Das Gleiche bei Menschen zu erreichen würde sehr viel länger dauern. Voraussetzung für diesen KI-Vorteil ist natürlich, dass wir herausfinden, woran das Problem ursprünglich lag und wie wir es beheben können.

Einblendung: Icons (Code mit ‚Bugs‘, Pfeile, Kreuze, Person in Auto, Gebäude, Uhr, Lupe)

Des Weiteren muss man fairerweise natürlich auch sagen, dass wir Menschen (mit oder ohne KI-Unterstützung) ohnehin – ähnlich wie KI-Systeme – nicht immer perfekte Entscheidungen treffen ...

Exkurs: Menschliche Entscheidungsfindung

Werfen wir doch mal einen Blick auf die menschliche Entscheidungsfindung [6]. Zunächst suchen wir in unserer Umwelt nach Informationen und Hinweisen, die für unsere Entscheidung relevant sind. Dabei spielt unter anderem unsere Aufmerksamkeit eine Rolle (die manchmal wirklich sehr selektiv sein kann), unsere Wahrnehmung und unser Gedächtnis. Wusstest du übrigens, dass die Genauigkeit unserer Entscheidung abnehmen kann, wenn wir viele Informationen zur Verfügung haben? Zwar tendieren wir dazu, nach vielen Informationen zu suchen, um ein möglichst vollständiges Bild zu bekommen, aber zu viele Informationen überfordern uns. Außerdem tun wir uns schwer damit, Informationen für unsere Entscheidung zu gewichten. Wir tendieren dazu, allen Informationen mehr oder weniger den gleichen Wert beizumessen, unabhängig davon, wie hilfreich sie tatsächlich sind. Das bezeichnet man auch als „**as if**“-Heuristik.

Quelle [6]

Einblendung: Icons (Kopf mit Zahnrädern & Glühbirne, Fernglas, Natur, Fragezeichen, Papiermappe, Ausrufezeichen, Waage); Schlagworte („Aufmerksamkeit“, „Wahrnehmung“, „Gedächtnis“, „as if“-Heuristik“)

Als Nächstes stellen wir Annahmen oder Hypothesen auf, wie die aktuelle Situation aussieht und wie sie sich zukünftig entwickeln wird. Im Rückgriff auf das, was wir schon wissen, entwickeln wir dann verschiedene Handlungsmöglichkeiten. Wir überlegen, was Ergebnisse dieser Optionen sein könnten, wie wahrscheinlich sie eintreten etc. Unsere Entscheidungsfindung ist dabei oft iterativ. Wir suchen also im Entscheidungsprozess auch immer wieder nach neuen Informationen. Allerdings neigen wir unter Umständen dabei dazu, Informationen zu suchen, die zu unserer Annahme oder Vermutung passen. Das bezeichnet man auch als **Confirmation Bias**. Generell neigen wir dazu, unsere erste Hypothese zu favorisieren und schwenken zum Teil nur schwer auf Basis von neuen Informationen auf weitere Hypothesen um. Das bezeichnet man als **Anker-Heuristik**. Und wenn wir wiederholte Entscheidungen treffen müssen, kann es zu einer Art Entscheidungsmüdigkeit (oder auch **Decision Fatigue**) kommen: Wir geben uns weniger Mühe, genau hinzusehen und Informationen zu analysieren.

Einblendung: Icons (denkende Figur, Gedankenblase mit A/B/C, Pfeile im Kreis, Häkchen, A, schlafendes Gesicht, pfeifende Figur); Schlagworte („Confirmation Bias“, „Anker-Heuristik“)

Bei dem, was ich euch gerade erzählt habe, habe ich die menschliche Entscheidungsfindung natürlich nur angerissen und habe das Ganze stark vereinfacht dargestellt. Auch die präsentierten Heuristiken und Biases sind nur ein kleiner Auszug. Sie sollen euch aber dafür sensibilisieren, dass wir nicht perfekt entscheiden. Die Frage wäre nun: Wie können wir eine gute Mensch-KI-Interaktion gestalten? Vor allem vor dem Hintergrund, dass wir uns – wie wir wissen – manchmal schwer damit tun, angemessen mit KI-Systemen umzugehen?

Einblendung: Icons (Figur mit Idee, Ausrufezeichen, Fragezeichen, Mensch, Pfeile, Smartphone, Laptop, Glühbirne, Blitz)

Ansätze für eine gute Mensch-KI-Interaktion

Neben gezielten Schulungen für die Nutzung von KI-Systemen, Aufklärung über ihre Funktionsweisen oder dem Ansatz der menschenzentrierten Automation, gibt es noch weitere Ansätze, um unsere Interaktion mit KI-Systemen für uns möglichst gut zu gestalten. Beispielsweise über die Gestaltung der KI-Systeme selbst. Ein interessantes Konzept in diesem Zusammenhang ist die sogenannte **Usability** oder Gebrauchstauglichkeit. Diese beschreibt, inwieweit wir mit einem System effektiv, effizient und zufriedenstellend unser Ziel erreichen können [7]. Wird bei der Gestaltung des Systems und bei der Schnittstelle, über die wir mit dem System interagieren, beispielsweise berücksichtigt, wie wir Dinge wahrnehmen, oder wie wir bestimmte Aufgaben üblicherweise bearbeiten, macht uns das den Umgang mit dem System zusätzlich leichter. Mit der Frage, wie sich Maschinen und unsere Interaktion mit ihnen bestmöglich gestalten lassen, beschäftigt sich beispielsweise die Ergonomie und die Ingenieurpsychologie [7]. Dort finden sich viele interessante

Hinweise für eine gute Gestaltung von Systemen. Man kann Nutzende übrigens auch wählen lassen, ob sie ein System einsetzen möchten und welchen Grad der Automatisierung das System haben soll. Das bezeichnet man auch als **adaptable automation** [8].

Quelle [7] [8]

Einblendung: Icons (Figuren an Plakat, Glühbirne, stolze Figur, Laptop, Smartphone, Figur am Computer, Glühbirne, Pfeil, Laptop, sehende Figur, Arbeitsblatt, Fragezeichen, Glühbirne, denkende Figur); Schlagworte („Usability“, „effektiv“, „effizient“, „zufriedenstellend“, „Ergonomie“, „Ingenieurpsychologie“, „adaptable automation“)

Auch Transparenz, beispielsweise in Form von Explainable AI, kann unseren Umgang mit KI-Systemen unterstützen. Erklärungen für die Systemempfehlung helfen uns dabei, einzuschätzen, ob wir die Empfehlungen eines KI-Systems annehmen wollen. Entscheidungen und Annahmen, die in die Entwicklung eines Systems geflossen sind, sollten ebenfalls transparent gemacht werden [5]. Dabei geht es nicht darum, dass jemand böswillig Systeme entwickelt, die später beispielsweise diskriminierende Empfehlungen aussprechen. Manchmal wurde etwas schlicht nicht bedacht, eine Annahme traf nicht zu oder das System hat Schlussfolgerungen gezogen, die nicht geplant oder nicht gewünscht waren. Aber nur wenn wir nachvollziehen können, welche Annahmen in einem System stecken, können wir fehlerhafte Annahmen beheben und einschätzen, wofür ein System geeignet ist [5].

Quelle [5]

Einblendung: Icons (Figur mit Handy und Ausrufezeichen, Sprechblase, Glühbirne, Teufel, durchgestrichener Teufel, Glühbirne); Schlagwort („Transparenz“)

Soweit, so gut. Aber für welche Entscheidungsprozesse können wir KI-Systeme nun einsetzen? Gibt es vielleicht auch Entscheidungen, bei denen wir KI-Systeme besser nicht einsetzen sollten oder sie gar nicht eingesetzt werden können?

Einblendung: Icons (nachdenkende Figur mit Fragezeichen)

Wo KI bei Entscheidungen helfen kann und wo (besser) nicht

Vielleicht ein wenig trivial mag es erscheinen, dass der Einsatz von KI-Systemen bei der Entscheidungsfindung mindestens zu genauso guten Entscheidungen führen sollte wie wenn der Mensch alleine entschieden hätte. Idealerweise sollten die Entscheidungen sogar besser werden. Wichtig ist daher, dass die Qualität der Entscheidungen oder Empfehlungen der KI-Systeme geprüft werden können [9]. Vor allem, wenn wir nicht genau nachvollziehen können, wie sie zustande gekommen sind. Können wir die Qualität beurteilen, können wir KI-Systeme prinzipiell einsetzen, um unsere Entscheidungsfindung zu unterstützen. Vorausgesetzt natürlich, sie sind gut genug.

Quelle [9]

Einblendung: Icons (Figur mit Smartphone, Gleichzeichen, denkende Figur, Daumen hoch, Ausrufezeichen, Detektiv*in, Figur vor Computer erst mit Fragezeichen, dann mit Glühbirne, Daumen hoch)

Es gibt aber durchaus auch Entscheidungen, bei denen der Einsatz von KI-Systemen schwierig bis unmöglich wird. Ein ziemlich offensichtliches Beispiel sind sogenannte **singuläre Entscheidungen** [5, 9]. Dabei handelt es sich um Entscheidungen, die entweder einmalig getroffen werden oder sehr selten. Das bedeutet in der Konsequenz natürlich auch, dass es kaum ähnliche Fälle gibt oder sogar gar keine, die man als Trainingsdaten für ein System heranziehen könnte. Häufig fallen politische Entscheidungen beispielsweise in diese Kategorie. Hinzu kommt bei diesem Entscheidungstyp oft, dass das Ergebnis sehr stark von der jeweiligen Ausgangslage abhängt. Was vielleicht in einem Land funktioniert hat, muss nicht unbedingt in einem anderen Land funktionieren. Dadurch wird es auch schwierig, die Qualität der entsprechenden Entscheidung dann zu beurteilen.

Quelle [5] [9]

Einblendung: Icons (Warnschild, durchgestrichener Papierstapel, Urkunde mit Siegel, Globus); Schlagworte („singuläre Entscheidungen“)

Eine weitere Kategorie sind **Werturteile** [5, 9]. Im Prinzip fallen darunter all diejenigen Entscheidungen, bei denen sich Expertinnen und Experten in einem angemessenen Rahmen einig sein müssen. Große Abweichungen sind hier nicht vorgesehen oder müssen zumindest erklärbar sein. Das klingt zu abstrakt? Noten sind ein gutes Beispiel dafür. Wir erwarten, dass sich die Dozierenden hier in gewissem Maße einig sind. Es könnten aber auch z. B. Strafmaße sein. Bei Werturteilen haben wir oft das Problem, dass wir keine spezifischen und konkreten Regeln vorgeben können, wie zu entscheiden ist. Wir können sie der Maschine also nicht vorgeben und sie kann sie auch nicht in den Daten finden. Bei Werturteilen kommt es aber auch auf die Begründung einer Entscheidung an. Menschen können sich austauschen, diskutieren oder ihre Entscheidung begründen. Das kann ein KI-System oft schwer leisten.

Quelle [5] [9]

Einblendung: Icons (Waage mit Figürchen drin, 1-, 2, 3+, Polizist*in, durchgestrichene Checkliste, Kopf mit Sprechblasen); Schlagwort („Werturteile“)

Fazit

KI-gestützte Systeme haben also das Potential, uns in vielen, aber nicht in allen Bereichen unseres Alltags zu unterstützen und zu besseren Entscheidungen beizutragen. Aber das kann nur gelingen, wenn wir bei der Implementierung solcher Systeme auch die Nutzerinnen und Nutzer mit ihren Eigenschaften berücksichtigen. Der soziale Kontext, in dem ein System eingesetzt wird, wie es gestaltet und gehandhabt wird, beeinflusst maßgeblich, wie gut die

gemeinsame Entscheidungsfindung von Mensch und Maschine funktioniert. Dafür ist es wichtig zu verstehen, wie Menschen mit automatisierten Systemen umgehen und im Hinterkopf zu halten, wie KI-Systeme entstehen und funktionieren. Basierend darauf kann man dann überlegen, wie die Mensch-KI-Interaktion bestmöglich gestaltet werden kann.

Einblendung: Icon (Glühbirne, Figur vor Laptop mit Schrift, Pfeil, Daumen hoch, Daumen runter, Ausrufezeichen); Schlagwort („Kontext“)

Das Herausfordernde – aber auch das Spannende – daran ist, dass die richtige Gestaltung der Mensch-KI-Interaktion durchaus von System zu System, von Anwendungszweck zu Anwendungszweck und unter Umständen auch von Nutzenden zu Nutzenden unterschiedlich sein kann. Mögliche Ansätze müssen daher individuell ausgewählt und angepasst werden.

Einblendung: Icons (Figur mit Idee, Welt mit Figürchen drum herum)

Quellen

- Quelle [1] Weiser, B., & Schweber, N. (2023, June 8). The ChatGPT lawyer explains himself. The New York Times. Verfügbar unter: <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html> (zuletzt abgerufen am 12.12.2023)
- Quelle [2] Dastin, J. (2018, October 18). *Insight - Amazon scraps secret AI recruiting tools that showed bias against women*. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/> (zuletzt abgerufen am 2.01.2024)
- Quelle [3] Buolamwini, J. (2016, November). *How I'm fighting bias in algorithms* [Video]. TEDxBeaconStreet. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms (zuletzt abgerufen am 2.01.2024)
- Quelle [4] National Transportation Safety Board (2019). *Collision between vehicle controlled by developmental automated driving system and pedestrian, Tempe, Arizona, March 18, 2018* (NTSB/HAR-19/03, Product No. PB2019-101402). <https://www.nts.gov/investigations/accidentreports/reports/har1903.pdf> (zuletzt abgerufen am 2.01.2024)
- Quelle [5] Zweig, K. (2023). *Die KI war's! Von absurd bis tödlich: Die Tücken der künstlichen Intelligenz*. Heyne.
- Quelle [6] Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering Psychology and human performance* (4th ed.). Routledge. [Chapter 8: Decision Making, p. 245-283].

- Quelle [7] Vollrath, M. (2015). *Ingenieurpsychologie. Psychologische Grundlagen und Anwendungsgebiete*. W. Kohlhammer GmbH. [Kapitel 8: Automation, S. 181-205].
- Quelle [8] Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2022). *Engineering psychology and human performance* (5th ed.). Routledge. [Chapter 13: Human-Automation Interaction, p. 516-551]. <https://doi.org/10.4324/9781003177616>
- Quelle [9] Zweig, K. (2023). Droht KI den Menschen zu ersetzen? *Aus Politik und Zeitgeschichte*, 73(42), 4-8. Verfügbar unter <https://www.bpb.de/shop/zeitschriften/apuz/kuenstliche-intelligenz-2023/> (zuletzt abgerufen am 9.01.2024)

Disclaimer

Transkript zu dem Video „Mensch-KI-Interaktion: KI und Entscheidungen“, Dr. Maike Mayer. 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.