An den Grenzen der KI

Die Rolle des Menschen in Bezug auf die Nutzung von KI

Sowohl in der Wirtschaft als auch in unserem täglichen Privatleben interagieren wir mit zunehmender Häufigkeit mit KI-Systemen und richten unser Verhalten nach algorithmischen Entscheidungen aus. Die Möglichkeiten zum Einsatz scheinen hierbei unbegrenzt. Jedoch sollten wir kritisch hinterfragen, für welche Zwecke und basierend auf welchen Daten und Modellen wir KI-Systeme einsetzen. Eine erste Orientierungshilfe bietet hier die vorgeschlagene KI-Verordnung der Europäischen Union (Artificial Intelligence Act, im Folgenden kurz AI-Act).

Text — Dr. Barbara Bushart, Dr. Christina Strobel

Mit dem Fortschreiten der digitalen Revolution werden digitale, automatisiert agierende Systeme sowohl in der Wirtschaft als auch in unserem täglichen Leben allgegenwärtig. Dabei übernehmen diese Systeme zunehmend Aufgaben von Menschen und passen sich ohne größeren Programmierungsaufwand flexibel an sich ändernde Rahmenbedingungen an. Heutzutage steuern solche Systeme bspw. Maschinen in Fertigungsstraßen, analysieren große Datenmengen, um Muster und Trends zu erkennen, oder machen Vorhersagen und leiten daraus die beste Strategie für Entscheidungen ab. Teilweise ist der Einsatz der Systeme den Nutzenden bewusst, teilweise verschwimmen jedoch die Grenzen oder die Nutzenden sind im Unklaren darüber, dass eine Künstliche Intelligenz (KI) anstelle eines Menschen eine Entscheidung trifft oder mit ihnen interagiert. Speziell im Steuer- oder im Gesundheitswesen werden KIs im Rahmen von Nudges dabei auch immer mehr zur Verhaltenssteuerung eingesetzt.

Was ist ein Nudge?

Ein Nudge ist jeder Aspekt der Entscheidungsarchitektur, der das Verhalten der Menschen auf vorhersehbare Weise verändert, ohne ihnen Optionen zu verbieten oder ihre wirtschaftlichen Anreize wesentlich zu verändern.

Die Grundlage für diese Systeme bildet die Analyse einer Vielzahl von Datenpunkten. KIs können dabei auf verschiedenste Arten aufgebaut sein, bspw. durch festgelegte Entscheidungsbäume bzw. Algorithmen, welche dazu führen, dass ein System bei Vorliegen bestimmter Daten bestimmte Aktionen ausführt. Als zielführender hat sich jedoch in den letzten Jahren der Einsatz von Machine Learning (oder in der Weiterentwicklung auch Deep

Learning) gezeigt, bei dem die KI selbstständig lernt und ihre Algorithmen eigenständig modifiziert.

Für einen Menschen ist es meist nicht nachvollziehbar, wie genau eine KI am Ende zu einer Entscheidung kommt, auch wenn es Bemühungen gibt, den Entscheidungsweg erklärbar zu machen (Explainable AI). Daher werden KIs zunehmend als autonome Einheiten angesehen, die ihre Aufgaben eigenständig ausführen. Wie eine Reihe von Studien ergeben hat, schreiben Menschen KIs dabei vermehrt menschliche Eigenschaften zu (Anthropomorphismus). Dies passiert nicht nur bei humanoiden Robotern, sondern auch bei rein digitalen KIs. So werden automatisierte Systeme bspw. als gleichwertige Teammitglieder wahrgenommen oder die Verantwortung für eine Entscheidung nahezu gleichwertig mit diesen geteilt.

Die steuernde Wirkung der Daten, Modelle und Methoden

Unabhängig vom Ansatz, mit welchem eine KI zu einem Ergebnis gelangt (meistens sind dies in der Praxis Mischformen), sind die Eingabedaten von entscheidender Bedeutung. Viel ausschlaggebender als die Entscheidungsfindung der KIs selbst ist somit das, was dieser vorausgeht: Die Fähigkeit, eine Fragestellung als mathematisch abbildbares Problem zu erkennen, dies zu formulieren, die richtigen Lern-Ansätze zu wählen und brauchbare Eingangsdaten zu nutzen. Erst diese Mathematisierung des Problems ermöglicht den Einsatz der KI. Dafür gilt es zunächst der Frage nachzugehen, welche Einflussfaktoren berücksichtigt werden müssen und inwieweit Daten zu diesen Faktoren vorliegen.

Grundsätzlich kann eine KI zwar gewisse Eingangsdaten schwächer bis gar nicht bei der Entscheidungsfindung berücksichtigen, sodass man unter Umständen dazu neigen würde, sämtliche verfügbaren Daten einer KI zur Verfügung zu stellen. Dies führt jedoch oftmals dazu, dass bspw. korrelierende Daten, welche jedoch keinen relevanten Einfluss auf eine Entscheidung haben, durch eine KI hoch gewichtet werden und es somit unter Umständen zu fehlerhaften Aussagen, Bias oder Diskriminierungen durch die KI kommt.

Beispiele hierfür sind etwa KIs im Personalwesen, die basierend auf Vergangenheitsdaten bestimmte ethnische Gruppen schlechter bewerten, da diese bislang keine Führungspositionen im Unternehmen einnahmen, oder Algorithmen zum Predictive Policing, durch welche Polizeikräfte präventiv in Stadtteile geschickt werden, in welchen in der Vergangenheit hohe Kriminalitätsraten auftraten und dadurch dort wiederum vermehrt Verstöße festgestellt wurden. In beiden Fällen bestätigt sich das System selbst.

Daher liegt die Entscheidung zum Einsatz der KI und die Auswahl der Daten, Modelle und Methoden weiterhin bei denjenigen, die die Fragestellung als mathematisches Problem formulieren und den Einsatz der KI gestalten. Ihnen obliegt es, den Algorithmus hinter einer Fragestellung zu modellieren, relevante Einflussfaktoren zu definieren, vorhandene Daten zu evaluieren und sich um die Beschaffung noch fehlender, aber benötigter Daten zu kümmern. Das Problem ausgelassener oder nicht verfügbarer Daten kann nur durch menschliches Urteilsvermögen erkannt und überwunden werden. Hier zwingt die Nutzung von KI uns, strukturiert und genau über die Datenstrukturen nachzudenken.

Ein nicht außer Acht zu lassendes Problem ist jedoch, dass Menschen tendenziell nicht besonders gut darin sind, statistische Informationen richtig zu deuten. Frühere Forschungsarbeiten bspw. von Kahnemann & Tversky haben gezeigt, dass allein die Wortwahl bei der Präsentation einer Vorhersagewahrscheinlichkeit Menschen in ihrer Entscheidung beeinflusst. Zusammen mit Forscher*innen der Harvard Medical School stellten sie bspw. Ärzt*innen zwei Behandlungsmöglichkeiten für Lungenkrebs vor: Bestrahlung oder Operation. Die Teilnehmenden wurden in zwei Gruppen geteilt und erhielten unterschiedliche Informationen über die kurzfristige Überlebensrate bei einer Operation. Wenn man ihnen sagte, dass „die Ein-Monats-Überlebensrate 90 % beträgt“, entschieden sich 84 % der Ärzt*innen für eine Operation. Jedoch sank diese Rate auf 50 %, wenn man ihnen sagte, dass „die Sterblichkeitsrate im ersten Monat bei 10 % liegt“. Beide Sätze sagen das Gleiche aus, aber die Formulierung der Information führt zu erheblichen Veränderungen bei der Entscheidung.

Aus Theorie wird Praxis

Wenn im Rahmen der Digitalisierung über den Einsatz von KI nachgedacht wird, sollte zunächst nicht die KI-Methode im Vordergrund stehen, sondern vielmehr die Formulierung und Modellierung der Fragestellung. Zudem gilt es

Dr. Barbara Bushart
Parlamentarische Beraterin & Juristin

Dr. Barbara Bushart wurde an der FSU Jena mit einer Monographie zu Hannah Arendts Rechtsverständnis promoviert und forscht seitdem vor allem im Bereich Grundlagen des Rechts. Sie ist parlamentarische Beraterin für Innenpolitik im Sächsischen Landtag.

Dr. Christina Strobel
Wissenschaftliche Mitarbeiterin & Gründerin

Dr. Christina Strobel forscht an der Technischen Universität Hamburg zum Thema Mensch-KI-Interaktion und ist Gründerin der KI-Ökosystem-Beratung AlgoTrust, die Unternehmen bei der ethischen und nachhaltigen Entwicklung und Nutzung von KI berät.

zu überprüfen, welche Daten für einen sinnvollen Einsatz der KI benötigt werden. Damit ergeben sich zukunftsgerichtet für die Praxis der Arbeitswelt insbesondere zwei Folgen:

1. Die KI verlagert den Schwerpunkt auf spezifische Fähigkeiten. Während bislang Routineaufgaben einen großen Teil der Arbeitszeit in Anspruch genommen haben, sind es künftig vermehrt evaluierende und strategische Aufgaben. Mit der zunehmenden Nutzung von KI steigt der Wert des Urteilsvermögens der Anwender*innen und Entwickler*innen der KI.

2. Die Vorbereitung des Einsatzes von KIs ist für die Qualität der KI von fundamentaler Bedeutung. Daher gilt es sicherzustellen, dass die Prinzipien, nach denen die KI funktioniert, bekannt sind. Hierzu sind Schulungen der Anwender*innen und Entwickler*innen von KI im Umgang mit sowie bei der Interpretation und Formulierung von Wahrscheinlichkeiten, statistischen Aussagen und Ergebnissen nötig.

Nur durch ein umfassendes Verständnis davon, was eine KI macht, kann gewährleistet werden, dass eine KI richtig und zielführend eingesetzt wird. Zudem muss sichergestellt sein, dass eine Überwachung von sich selbst weiterentwickelnden KIs gewährleistet ist, um ungewolltes Verhalten frühzeitig zu identifizieren und diesem entgegenwirken zu können.

Ein Weg in die Zukunft? – Der Artificial Intelligence Act

Neben einem funktionalen Verständnis wird jedoch auch ein Verständnis der grundlegenden rechtlichen Rahmenbedingungen immer wichtiger. Der Entwurf des AI-Acts, der im Jahre 2021 von der Europäischen Kommission veröffentlicht wurde, stellt die Grundlage für die künftige Entwicklung und Nutzung von KI im europäischen Raum dar. Während die europaweit geltende Datenschutz-Grundverordnung bereits seit Inkrafttreten im Jahr 2016 den Zugriff auf und den Umgang mit Daten regelt – sei es aus dem virtuellen oder dem realen Raum –, nimmt der AI-Act vor allem auf die Risiken zunehmender Automatisierung, die Einsatzgebiete der KIs und die zugrunde liegenden Daten Bezug:

- Abhängig von einer Risikoeinstufung der KIs müssen verschiedene Anforderungen an diese sowie ihren Einsatz erfüllt werden. Dabei wird der Bereich der Hochrisiko-KIs, also derjenigen, die grundrechtsrelevant sind, den strengsten Vorschriften unterworfen. Dies bedeutet, das ein Risikomanagementsystem implementiert werden muss, dass die Hochrisiko-KI und ihren Einsatz permanent evaluiert. Daneben müssen die verwendeten Daten engmaschig kontrolliert und ausgewertet werden (bspw. auf Biases), und auch die grundlegenden Design-Entscheidungen müssen dokumentiert und begründet werden. In weniger kritischen und/oder risikobehafteten Bereichen werden geringere Anforderungen gestellt. Es werden jedoch auch einige Bereiche benannt, in welchen der Einsatz von KIs grundsätzlich verboten ist (wie bspw. im social scoring).

- Auch an die technische Dokumentation werden hohe Ansprüche gestellt. Diese muss bereits vor der Markteinführung der KI zur Verfügung stehen. Darüber hinaus müssen die Funktionsweisen der KI ständig durch ein Überwachungssystem dokumentiert werden. Insbesondere sollen dabei die Anwendungszeit, die Datengrundlage und diejenigen Menschen, die Ergebnisse bestätigt haben, aufgezeichnet werden. Diese drei Aspekte sind eine wichtige Ergänzung zu den heute bereits bestehenden Regelungen. Denn es gibt zwar bereits Gesetze wie bspw. in Deutschland das Allgemeine Gleichbehandlungsgesetz, die einen diskriminierungsfreien Zugang zum Arbeitsmarkt garantieren sollen, allerdings regulieren diese nicht den Prozess selbst, sondern nur das gewünschte Ergebnis. Damit können zwar Entwickler*innen bzw. Entscheider*innen im Nachhinein für eine rechtswidrige Entscheidung in Haftung genommen werden, sie sind jedoch dadurch nicht automatisch verpflichtet, Sicherungen in ihr System zu integrieren, die das Risiko einer automatisierten Fehlausgabe minimieren. Daher haben diese Regelungen heute meistens keinen direkten Einfluss auf die Gestaltung von KIs.

- Der AI-Act sieht ebenfalls vor, dass Transparenz geschaffen werden muss und die Nutzenden Zugriff auf zugrunde liegende Informationen haben müssen. So muss es möglich sein, dass für die Nutzenden einer KI erkenn-

bar ist, welche Daten in welcher Menge der Entscheidung zugrunde liegen. Damit kann der Gefahr einer Fehlentscheidung aufgrund der Verwechslung von Korrelation und Kausalität zumindest teilweise begegnet werden. Es ist jedoch fraglich, inwieweit Nutzende diese Informationen verstehen und analysieren können. Hier wird unabhängigen Kontrollstellen eine besondere Bedeutung zukommen.

Kausalität und Korrelation

Kausalität beschreibt einen Zusammenhang zwischen zwei Faktoren, bei denen der eine Faktor ursächlich für den anderen Faktor ist. Korrelation bedeutet, dass lediglich ein Zusammenhang zwischen zwei Faktoren besteht.

Die wohl wichtigste Anforderung ist die ebenfalls geregelte notwendige menschliche Überwachung. Die Systeme müssen so beschaffen sein, dass sie effektiv von Menschen beaufsichtigt werden können. Dies dient dazu, die Risiken für Gesundheit, Sicherheit oder die fundamentalen Rechte von Menschen zu minimieren. Die Qualifikation der Aufsichtspersonen ist hierbei von entscheidender Bedeutung.

Unverzichtbar – der menschliche Faktor

Wie dargelegt, ist der menschliche Faktor nicht nur aus technisch-gestalterischer, sondern auch aus rechtlicher Sicht beim Einsatz einer KI unverzichtbar. Dabei gilt es vor allem auf vier Themenfelder zu achten:

1. KIs sind ineffektiv in Bereichen, in denen relevante Daten spärlich sind. Eine der Hauptschwächen von KI stellen unbekannte sowie nicht datentechnisch erfasste Einflussfaktoren dar. Diese gilt es zu identifizieren und wenn möglich anhand von verfügbaren Stellvertreterdaten (Proxy-Daten) in die KI zu integrieren.

2. KIs geben manchmal falsche Antworten, von deren Richtigkeit sie überzeugt sind. Dies kann bspw. daran liegen, dass relevante Daten nicht vorlagen oder berücksichtigt wurden oder nicht zwischen Kausalität und Korrelation unterschieden werden kann. Speziell bei abstrakten Fragestellungen ist es häufig schwierig, erklärende Drittvariable zu finden, welche eine vermeintliche Kausalität als Korrelation erklären. Es besteht somit die Gefahr, dass ein Ereignis als ursächlich für ein anderes betrachtet wird, während es tatsächlich lediglich korreliert.

3. KIs treffen keine Entscheidungen, sie nehmen statistische Analysen vor. Um basierend auf den Analysen Entscheidungen zu automatisieren, müssen umfangreiche und umfassende Daten zu allen relevanten Einflussfaktoren vorliegen. Dies muss vor der Automatisierung überprüft werden.

4. Beim Einsatz von KIs sind rechtliche Anforderungen zu beachten. Dies gilt nicht nur für den Umgang mit Daten (Datenschutz, Anonymisierung, Nutzungsrechte, Verhinderung von Biases usw.), sondern bspw. auch für Prozesse zum Monitoring, zur Sicherung der Transparenzanforderung und der Regelung von Verantwortlichkeiten.

Fazit

Der AI-Act weist bereits in die richtige Richtung, wenn er detaillierte Dokumentation und hohe Transparenz fordert, die vielen der aufgezeigten Risiken entgegenwirken können. Gleichzeitig wird jedoch auch deutlich, dass der wichtigste Faktor für die Nutzung von KI immer noch der Mensch ist. Eine der größten Gefahren bei der Nutzung von KI liegt darin, dass Menschen bei Entscheidungen, die sie „gemeinsam“ mit einer KI treffen, dieser einen erheblichen Teil der Verantwortung zusprechen. Sie verstehen sich also nicht mehr, wie es das Rechtssystem bereits jetzt und der AI-Act auch zukünftig vorsieht, als Letztentscheider*innen. Dieses Selbstverständnis muss deswegen zukünftig bei allen Beteiligten enorm gestärkt werden. Denn es zeigt sich, dass entgegen der landläufigen Meinung der Mensch im Zuge der zunehmenden Automatisierung nicht überflüssig wird. Seine menschlichen Fähigkeiten wie Urteilsvermögen, Reflexion und systemübergreifendes Denken sind vielmehr unverzichtbarer denn je, nicht nur beim Design, sondern auch bei der Nutzung von KI – insbesondere, wenn diese zur Verhaltenssteuerung eingesetzt wird. ▪