«KÜNSTLICHE INTELLIGENZ UND ETHIK: WAS BEDEUTET ES, WENN MASCHINEN ANSTELLE VON MENSCHEN ENTSCHEIDEN?»

Impulsreferat von Dr. Dorothea Baur, baur consulting gmbh

Künstliche Intelligenz verbreitet sich immer mehr in Kontexten, wo direkt Einfluss auf Menschen genommen wird. Dr. Dorothea Baur bewertet diese neue «unsichtbare Hand» aus ethischer Perspektive und zeigt an eindrucksvollen Fallbeispielen auf, welchen Einfluss Künstliche Intelligenz bereits heute auf den Menschen hat. Dabei stellt sie in ihrem Referat als Ethikerin eine ganz andere Frage ins Zentrum als wir sie aus der Psychologie kennen. Es ist nicht die Frage nach dem sein, sondern es ist die Frage nach dem sollen. Und diese Frage hat es in sich – denn die Antwort darauf definiert, wie wir das Verhältnis zwischen Mensch und Künstlicher Intelligenz in Zukunft ausgestalten wollen.

Dr. Dorothea Baur, baur consulting gmbh

Aber ganz von vorne. Als Unterbereich der Philosophie beschäftigt sich die Ethik mit der philosophischen Reflexion von begründbaren moralischen Geltungsansprüchen. Oder einfacher gesagt, die Ethik fragt im Unterschied zur Psychologie nicht nach dem sein, also wie das Zusammenleben von Menschen und Künstlicher Intelligenz ist, sondern sie fragt nach dem sollen, also wie das Zusammenleben mit der künstlichen Intelligenz aussehen soll. Im Zentrum stehen somit normative Fragen, also an welchen moralischen Werten wir uns orientieren wollen, wenn es um die Ausgestaltung des Verhältnisses zwischen Mensch und Künstlicher Intelligenz geht. Und daraus entstehen wiederum verschiedene Handlungsorientierungen, die uns in der Ausgestaltung der KI leiten. So wird bei einem Einsatz von KI in der Medizin beispielsweise nicht die Fairness im Sinne von Zugangsmöglichkeiten zu medizinischer Versorgung für alle zu jeder Zeit priorisiert, sondern die Sicherheit, dass die KI möglichst sicher und zuverlässig arbeitet.

Der Geltungsbereich der Ethik ist eigentlich all das, wo der Mensch Entscheidungen trifft. So ist auch die künstliche Intelligenz ein menschgemachter Kontext, den wir selbst gestalten und in welchem wir definieren, was ein legitimes Produkt oder legitime Vorgehensweise ist. Und gerade im Bereich der Künstlichen Intelligenz zeigen aktuelle Fallbeispiele die Wichtigkeit einer grundlegenden Diskussion über die obengenannten Fragestellungen auf.

DISKRIMINIERUNG, BIAS UND ETHISCHE HERAUSFORDERUNGEN

Im August 2020 gingen die Menschen in England erstmals auf die Strasse, um gegen einen Algorithmus zu protestieren. Doch wie kam es dazu? Da wegen der Corona-Pandemie die Maturaprüfungen nicht abgelegt werden konnten, plante das Bildungsministerium zunächst, dass die Lehrpersonen die Noten auf Grundlage früherer Bewertungen vergaben (sogenannte predicted grades). Das führte jedoch dazu, dass die Noten im Durchschnitt besser waren als in den Vorjahren. Das Bildungsministerium entschied sich deshalb dafür, die Noten von einem Algorithmus korrigieren zu lassen. Die Folge war, dass mehr als 1/3 der Schülerinnen und Schüler ein downgrade erlebte. Besonders beunruhigend war jedoch das Verhältnis der Schülerinnen und Schüler aus Privatschulen und öffentlichen Schulen. Denn der Algorithmus hat nebst den historischen

Noten der Schülerinnen und Schüler weitere Daten zur Berechnung der Abschlussnote beigezogen, wie etwa die Schule, Klassengrösse und die Postleizahl. Dies hatte für Schülerinnen und Schüler aus Wohngegenden mit historisch geringeren Bildungsabschlüssen zur Folge, dass sie trotz guter Noten fast doppelt so häufig eine schlechtere Abschlussnote bekamen als Schülerinnen und Schüler aus besseren Wohngegenden und kleineren Klassen. Das Bildungsministerium entschuldigte sich nach den Protesten für den «mutanten Algorithmus» und machte die Noten rückgängig. Was bleibt ist aber die Frage nach den Rechten und Pflichten bei Entwicklung und Nutzung von KI. Denn der «mutante» Algorithmus ist nicht einfach so entstanden, er ist menschgemacht.

Doch dies ist längst nicht das einzige Beispiel, in dem KI Ungleichheiten in der Gesellschaft fortschriebt und zur Benachteiligung bestimmter Gruppen beiträgt. So zeigt ein Fall aus Österreich, dass ein bestimmtes soziodemografisches Merkmal bestimmte, wer Unterstützungsleistungen bei der Arbeitssuche erhält. Der Algorithmus teilte Personen in drei Gruppen ein, in welchem Masse sie Anspruch auf Unterstützungsleistungen erhalten. Der ersten Gruppe wurden Personen zugeteilt, die fähig waren, alleine ein neues Beschäftigungsverhältnis zu finden. Der zweiten Gruppe wird Unterstützung bei der Arbeitssuche zugesprochen und der dritten Gruppen wiederum keine, da sich bei diesen Personen eine Unterstützung kaum lohnt. Die Einteilung war das Ergebnis einer Datentabelle aus 81‘000 Datensätzen, die sich aus zahlreichen Merkmalen wie z.B Geschlecht, Altersgruppe, Staatengruppe etc. zusammensetzt. Eines dieser Merkmale war die Betreuungsverpflichtung, die jedoch nur bei Frauen einberechnet wurde. Dies hatte zur Folge, dass Frauen mit Betreuungspflichten durch den Algorithmus systematisch schlechtere Aussichten auf Unterstützungsleistungen erhielten.

Aber nicht nur im öffentlichen Bereich zeigt sich Diskriminierung durch Algorithmen, sondern auch in der Privatwirtschaft: Amazon nutzt seit 2014 KI im Recruiting, um die besten Talente zu entdecken. Ein Jahr später kam heraus, dass das Modell für Stellen im Bereich Softwareentwicklung und Technik die Lebensläufe nicht geschlechterneutral bewertet hat, obwohl die Angaben zum Geschlecht absichtlich aus den Lebensläufen entfernt wurden. Doch wie konnte das passieren? Der selbstlernende Algorithmus hat gelernt, dass Personen mit dem Hobby «Baseball» historisch gesehen besser bewertet wurden als Personen mit typisch weiblichen Hobbys, zu welchen der Algorithmus kaum oder wenige Daten finden konnte. Der Algorithmus hat daher durch sogenannte Stellvertretervariablen angefangen, systematisch männliche Kandidaten zu bevorzugen, weil deren Profile stärker mit Profilen übereinstimmen, die historisch «positiv» bewertetet wurden. Und genau hier liegt eine zentrale Herausforderung, wenn Maschinen anstelle von Mensch Entscheidungen treffen: Der Algorithmus unterscheidet nicht zwischen Korrelation und Kausalität. Und auch wenn versucht wird, bestimmte Merkmale wie Geschlecht oder Einkommen zu entfernen, gibt beispielweise die Postleizahl stellvertretend Auskunft über das Einkommen.

Wie ein solcher Bias in der Künstlichen Intelligenz zustande kommt, verdeutlicht auch die Fehlerrate der automatischen Gesichtserkennung eindrucksvoll. Bei weissen Männern liegt diese nur gerade bei 1%, bei dunkelhäutigen Frauen hingegen bei 35%. Ein Bias entsteht in einem KI-Projekt demnach bereits im Rahmen der Konstruktion der Anwendung - repräsentieren die zugrundeliegenden Daten die betroffenen User beispielsweise nur zu einem bestimmten Teil, zieht sich dieser Bias durch die ganze KI-Anwendung. Dabei ist ein Bias in der künstlichen Intelligenz nicht per se schlecht, sondern muss immer im Verhältnis zur Aufgabe betrachtet werden. In einer medizinischen Diagnose ist ein Bias überlebenswichtig, in einem anderen Anwendungsfall kann er diskriminierend sein. So bedeutet beispielsweise der Bias in den aktuellen Fallbeispielen, dass sich nur gewisse Menschen auf die KI verlassen können. Ungleichbehandlung bedeutet im Kontext von KI demnach, dass bestimmte Gruppen damit rechnen müssen, Opfer von ungerechter Benachteiligung werden – und dies mit potentiell schwerwiegenden Folgen, stellt man sich den Anwendungsfall in der Medizin oder im Polizeibereich vor.

Und auch die zukünftige Entwicklung von KI lässt ethische Herausforderungen erwarten: Im Jahr 2022 werden ca. 10% aller persönlichen Geräte «Emotion AI Fähigkeiten» besitzen, die ihnen erlauben, menschliche Emotionen zu erkennen und darauf zu antworten. KI wird aber simplifizierend auf 6 Emotionen programmiert, die nicht das ganze Spektrum an menschlichen Emotionen abdecken. Wie zuverlässig kann ein solcher Algorithmus und seine Entscheidungen also sein und womit legitimeren wir seinen direkten Einfluss auf das Verhalten von Menschen und den Eingriff in die Privatsphäre?

DIE WICHTIGKEIT ETHISCHER FRAGEN IN KI-PROJEKTEN

Was lässt sich also aus Perspektive der Ethik für den Einsatz von KI festhalten? Trotz vielen erfolgsversprechenden Anwendungsfällen von KI z.B zur Reduktion von Foodwaste, Pestiziden und vereinfachtem Zugang zu Dienstleistungen, stellen sich in der Ausgestaltung von KI zahlreiche ethische Herausforderungen. Gerade wenn es darum geht, KI in Kontexten anzuwenden, die einen direkten Einfluss auf den Menschen haben, werden ethische Fragen zentral. Als erfolgsversprechend führt Dorothea Baur die Berücksichtigung von ethischen Fragen in allen Phasen der Entwicklung von KI-Projekten auf. So betreffen zentrale ethische Fragen in der Planung den Zweck der KI, die Repräsentativität der Daten und die Klärung von möglichen Abhängigkeiten und der Privatsphäre der Nutzerinnen und Nutzer. In der Entwicklung der KI gilt es die Gefährdung möglicher verletzlicher Gruppen zu berücksichtigen und die Verständlichkeit und Erklärbarkeit des Algorithmus jederzeit zu gewährleisten. Auch während des Einsatzes muss die KI Vorsicht betreffend Monitoring walten lassen und sicherstellen, dass Individuen Entscheidungen anfechten können.