Wege zum verantwortungsvollen Umgang mit KI

# Kritisch nachdenken, erklären, abstimmen

Bild: Lee – adobe.istock.com

Beim Einsatz von Algorithmen geht es oft um Effizienz, selten um Verständlichkeit und Werte; sei es im Sozialbereich, oder in der Wirtschaft. Forschungsprojekte von Wissenschaft und NGOs zeigen, worüber Verantwortliche, Betroffene und Interessierte reden sollten. **Von Michaela Ortis**

Die Wissenschaftscommunity rund um Maschinelles Lernen diskutiert schon lange, dass ML-Algorithmen verständlicher sein müssen, damit die Menschen ableiten können, wieso es zu bestimmten Vorhersagen gekommen ist und was diese bedeuten. Einer darunter ist Sebastian Tschiatschek, er hat in der Schweiz direkte Demokratie erlebt: „Das hat mich motiviert, dieses Abstimmen und Mitreden auch in meinem Forschungsbereich zu ermöglichen."

## Algorithmen für Laien

Der Assistenzprofessor für Maschinelles Lernen am Informatikinstitut der Uni Wien forscht im Projekt „Interpretability and Explainability as Drivers to Democracy" (gefördert vom WWTF), wie Algorithmen und Entscheidungsfindungsprozesse für Laien verständlich gemacht werden können. Fokussiert wird auf alle Gruppen, die im Lebenszyklus eines Algorithmus mitspielen bzw. betroffen sind: Beamte und Politik, die entscheiden, ob ein Algorithmus eingesetzt und wie er entwickelt wird, Firmen mit ihren wirtschaftlichen Interessen und die Bevölkerung. Je nach Wissensstand brauchen sie spezifische Erklärungen. Anschaulich sind Diagramme um zu zeigen, welche Trends der Algorithmus darstellt. Um ein Gefühl zu bekommen, was ein Algorithmus tut und ob er für bestimmte Bevölkerungsgruppen fair ist, empfiehlt sich Algorithmic Recourse, wo man Varianten direkt überprüfen kann, z.B.: Wenn ich andere Eigenschaften hätte, welche Voraussagen würde das System dann über mich machen. Bzw. umgekehrt: Welche Eigenschaften müsste ich verändern, um eine andere Vorhersage zu erhalten.

» **Sebastian Tschiatschek, Universität Wien:** „Man sollte idealerweise allen die Möglichkeit geben, Vorhersagen zu verstehen."

Foto: Julia Glück

In die Forschung fließen interdisziplinäre Gebiete ein, wie die Demokratietheorie mit Partizipationsmodellen für ausgewählte repräsentative Personen aus der Bevölkerung; dazu Tschiatschek: „Mein Zugang ist etwas anders, man sollte idealerweise allen die Möglichkeit geben, Vorhersagen über sie zu verstehen. Ob sie es dann nützen oder nicht, ist eine andere Sache. Aber hier muss man aufpassen, dass das nicht zur Pseudo-Mitsprache wird: Wenn Entwickler von Algorithmen etwa sagen, wir drucken auf 1.000 Seiten die Parameter eines neuronalen Netzwerks aus und Sie können alles nachschauen, hilft das de facto nichts."

## Was ist das Ziel?

Die Gesetzeswerke der EU sowie vieler einzelner Staaten propagieren, dass man bei automatisierten Entscheidungen in high-risk AI Systemen einen „Human oversight" braucht, also eine Person, die die Ergebnisse des Algorithmus kontrolliert und prinzipiell entscheiden kann, das AI-System nicht zu nutzen oder die Entscheidungen des Systems zu verändern oder zu verwerfen. Tschiatschek untersucht, was ein menschlicher Einfluss bedeutet, um abzuleiten, welche Informationen die Verantwortlichen bzw. die Ausführenden benötigen: „Wenn ein Human in the Loop Entscheidungen des Algorithmus verändern darf, wird dessen implizites Wertesystem verändert und das erfordert eine Abstimmung. Man erwartet sich durch den Algorithmeneinsatz Objektivierung, aber das passiert bei so einem Setting nur zu einem gewissen Grad. Werte festzulegen

ist ein großes und nicht einfaches Thema. So ein Prozess wäre jedoch interessant, denn Verantwortliche müssten explizit sagen, was sie mit dem Algorithmus erreichen wollen – aber das tun sie oft nicht so gerne."

Zum wichtigen Thema Werte forscht Lupita Svensson von der Universität Lund, bezogen auf Automatisierung im Sozialwesen, wo Entscheidungen oft vulnerable Personen betreffen und umso heikler sind. Sie sprach mit Sozialarbeiter:innen in schwedischen Kommunen, welche eng in die Entwicklung der Algorithmen eingebunden waren, das wurde von den Behörden so verlangt. „Das ist auf der einen Seite wichtig und gut, aber man muss auch etwas anderes bedenken: Diese Algorithmen sind nicht neutral, denn sie werden stark von den Menschen, die sie konstruieren, beeinflusst", sagt Svensson. Es brauche mehr Studien, um die daraus resultierenden Konsequenzen zu sehen. Wenn gewünscht wird, die Handhabung von Sozialhilfe zu ändern, können sich Sozialbehörden weder auf alte Daten noch auf Sozialarbeitskräfte verlassen. Vielmehr müssen sie definieren, was die neuen Werte sein sollen.

» **Lupita Svensson, Universität Lund:** „Entscheidungen über heikle Situationen , wo schutzbedürftige Menschen involviert sind, sollten nicht Algorithmen überlassen werden." Foto: Peter Frodin, Lund University

Svensson plädiert, Algorithmen dort einzusetzen, wo sie Effizienz und Neutralität bringen, wie beim Beschaffen und Klassifizieren von Informationen nach klaren Regeln: „Aber Entscheidungen über heikle Situationen, wo schutzbedürftige Menschen involviert sind, sollten nicht Algorithmen überlassen werden. Die technologische Entwicklung schreitet schnell voran und es ist leicht, sich darauf zu stürzen, ohne nachzudenken, was der Auftrag ist – deshalb ist die Diskussion darüber so wichtig."

## Mensch oder Maschine

Die Rolle des Human in the Loop wurde an Hand des AMS-Algorithmus untersucht, berichtet Rainer Stummer von der Bürgerrechts-NGO epicenter.works: „Jeder kann sich das gut vorstellen: Der AMS-Algorithmus zeigt mir an, dass die Person, die vor mir sitzt, schlechte Chancen am Arbeitsmarkt hat. Das beeinflusst, wie ich über sie denke – obwohl ich bei dem, was ich über sie weiß, zu einer anderen Entscheidung kommen würde." Derzeit ist der AMS-Algorithmus durch die Datenschutzbehörde gestoppt, weil die rechtliche Grundlage fehlt.

Aus Sicht von AMS-Mitarbeiter:innen und Klient:innen habe der Algorithmus nicht mehr Zeit für Gespräche ermöglicht. Er habe vorgegriffen und Menschen in Kategorien eingeteilt, mit starkem Bias. Oft hätten die Fachkräfte anders entschieden, sie fühlten sich bevormundet. Dazu Stummer: „Ein wichtiger Punkt kam in den Befragungen immer wieder: Menschen können bessere Entscheidungen treffen, die zur individuellen Situation einer Person passen."

Die erste Frage für Verantwortliche sei daher: Ist Technik die richtige Lösung für unser Problem? Kann Technik helfen, stehen weitere Fragen an: Bei welchen Entscheidungen sind Menschen betroffen und welche Tragweite haben diese?

» **Hannes Stummer, epicenter. works:** „Menschen können bessere Entscheidungen treffen, die zur individuellen Situation einer Person passen." Foto: privat

Kann die Technologie Qualitätsstandards wie Zuverlässigkeit oder Nichtdiskriminierung erfüllen? Der AI-Act der EU möchte künftig KI-Systeme in Kategorien einteilen. Je nach Risiko benötigen Systeme andere Qualitätsstandards und Sorgfältigkeit bei der Herstellung; gewisse Anwendungsfälle sollen verboten werden. „Eine Folgenabschätzung betrifft auch die personenbezogenen Daten, die solche Systeme vielleicht sammeln. Denn wir können oft gar nicht ermessen, welche Rückschlüsse Daten, die jetzt gesammelt werden, später einmal zulassen werden", betont Stummer.

Für den Einsatz von Algorithmen sprächen laut Tschiatschek Argumente wie Effizienzsteigerung, Kostenreduktion und Objektivität. Ein Mensch könne jedoch Einzelfälle differenzierter betrachten und Aspekte berücksichtigen, die bei der Entwicklung nicht eingeplant waren. „Algorithmen werden kommen und ich halte es für wichtig, wie bei allen Technologien kritisch darüber nachzudenken", resümiert Tschiatschek. ■

*Die Recherche wurde im Rahmen des Stipendiums Forschung & Journalismus der Österreichischen Akademie der Wissenschaften gefördert.*