

OBJEKTIVITÄT VON ALGORITHMEN

FORSCHUNGSINTERESSE:

Welche Faktoren beeinflussen den Entwicklungsprozess von Algorithmen?

VORGEHEN:

Die Wissenschaftler*innen interviewten Data Scientists und Entwickler*innen von Social Media Plattformen hinsichtlich sogenannter personalisierter Services (z. B. beim Produktkauf, wo ein Algorithmus Kaufempfehlungen ausspricht). Dabei fokussierten sie sich auf drei Aspekte: das Algorithmic Management (datengetriebene Formen, Arbeitsprozesse zu organisieren wie z. B. die Planung von Schichtdiensten), Anonymität (durch die Verbindungen verschiedener Daten kann Anonymität aufgehoben werden) und Datennachbarschaften (algorithmische Empfehlungen auf Basis von Entscheidungen anderer Personen mit ähnlichem Verhalten).

ERGEBNISSE/ERKENNTNISSE:

Die Untersuchung ergab, dass bestimmte Annahmen der Entwickler*innen in die Gestaltung der Social Media Plattformen einfließen (z. B. dass „Datennachbarn“ als Gleichgesinnte mit ähnlichen Interessen gelten). Das bedeutet, dass bei der Entwicklung von Algorithmen stets betrachtet werden sollte, wer diese entwickelt und für wen.

DISKRIMINIERUNG DURCH DATENNACHBARSCHAFTEN

FORSCHUNGSINTERESSE:

- Auf welchem Prinzip basiert die Anonymisierung persönlicher Nutzungsdaten?
- Welche Folge hat die Anonymisierung auf personalisierte Services (z. B. Kaufempfehlungen)?

VORGEHEN:

- Hintergrund: Immer mehr Unternehmen nutzen Nutzerdaten, um personalisierte Services anzubieten. Dazu werden Verhaltensdaten anonymisiert und in sogenannten Datennachbarschaften gruppiert, wobei Personen mit ähnlichem Kauf- und Klickverhalten zusammengefasst werden, um ihre Privatsphäre zu schützen.
- Die Wissenschaftler*innen analysierten eine bisher unveröffentlichte Studie, in der es um die Entstehung des Nachbarschaftsprinzips „Homophilie“ geht. Dieses Prinzip wird heutzutage für Empfehlungsalgorithmen verwendet. In der Studie geht es um Freundschaftsnetzwerke in zwei Wohnprojekten. Die Studie beschreibt zwei soziale Konzepte, die zur Bildung von freundschaftlichen Beziehungen zwischen Nachbar*innen beitragen: „Gleiches und Gleiches gesellt sich gern“ (Homophilie) und „Gegensätze ziehen sich an“.

ERGEBNISSE/ERKENNTNISSE:

Von diesen beiden Konzepten hat sich nur das Prinzip der Homophilie in der Entwicklung digitaler Netzwerke und ihren Personalisierungsalgorithmen durchgesetzt. Das ist deshalb problematisch, weil personalisierte Dienste (wie z. B. Kaufempfehlungen) Nutzer*innen zu Gruppen mit ähnlichen Interessen zusammenfassen - sogenannte Datennachbarschaften. Dadurch wird Vielfalt/Andersartigkeit ausgeschlossen, z. B. indem bestimmte Produkte gar nicht erst angeboten werden. In diesem Zusammenhang werden Phänomene der systematischen Datendiskriminierung deutlich. Daher sollte das Prinzip der Anonymität und der homophilen Datennachbarschaften neu gedacht werden, um Diskriminierung in digitalen Technologien entgegenzuwirken.

ML-BASIERTES AUKTIONSSYSTEM

FORSCHUNGSINTERESSE:

Inwiefern kann Machine Learning ein Auktionssystem verbessern?

VORGEHEN:

- Gemeinsam mit einem Praxispartner entwickelten die Wissenschaftler*innen ein automatisiertes Auktionssystem, das gebrauchte Waren anbietet und mithilfe von Machine Learning (ML) Produktvorschläge für Nutzer*innen ausgibt.
- Der Praxispartner hatte u. a. die Anforderung, dass die Entscheidungen des Systems erklärbar sein müssen, damit es auch im direkten Kundenkontakt eingesetzt werden kann.
- Die Wissenschaftler*innen entwickelten einen ML-Algorithmus, den sie mit Produktdaten, Informationen über die Käufer*innen sowie den Auktionsverlauf trainierten. Als Grundlage verwendeten sie einen Datensatz, der Informationen über 3.200 Nutzer*innen, 270.000 Artikel und 375.000 Gebote enthielt.

ERGEBNISSE/ERKENNTNISSE:

Mithilfe dieser Daten entwickelten die Wissenschaftler*innen ein Modell, das bis zu elf Prozent effektiver als derzeit existierende State-of-the-art-Systeme arbeitet.

VERHÄLTNIS VON MENSCH UND KÜNSTLICHER INTELLIGENZ

FORSCHUNGSINTERESSE:

- Wie wirkt sich Künstliche Intelligenz auf das Verhalten von Menschen in ökonomischen Verhandlungen aus?
- Beeinflussen ökonomische Erwartungen die Präferenzen für oder gegen Künstliche Intelligenz?
- Inwiefern unterscheiden Menschen zwischen lernenden und nicht-lernenden Algorithmen?
- Wächst das Vertrauen der Menschen in lernende oder nicht-lernende Algorithmen mit der Zeit und wenn ja, warum?

VORGEHEN:

- Im ersten Schritt führten die Wissenschaftler*innen eine Literaturrecherche zu den Auswirkungen von KI auf das Verhalten von Beschäftigten durch. Im zweiten Schritt untersuchten sie in zwei Studien das Verhältnis von Mensch und KI.
- In der ersten Studie „Ökonomische Verhandlung zwischen zwei Akteuren“ sollten zwei Personen Geld aufteilen. Zunächst verhandelten sie ohne Technologie, dann mit einer automatisierten Entscheidungshilfe für eine Person und einem Recommender System für die andere. Ziel war es, herauszufinden, wie sich das Verhalten durch den Einsatz der Technologien verändert.
- Die Studie wurde mit dem Fokus auf ökonomische Erwartungen wiederholt, an der 480 Personen teilnahmen. Die Wissenschaftlerinnen untersuchten, ob der ökonomische Eigennutz die Präferenzen für oder gegen eine Verhandlung mit KI beeinflusst. Die Probandinnen wurden in „Arbeitgeberinnen“ und „Arbeitnehmerinnen“ unterteilt und sollten vor der Verhandlung die Höhe des Geldes einschätzen, das sie aushandeln würden.
- Die zweite Untersuchung „Algorithmus Aversion“ erforschte, wie Menschen die Entscheidungen von Algorithmen ablehnen und ob sie zwischen lernenden und nicht lernenden Algorithmen unterscheiden. Proband*innen sollten eine Zahl ermitteln und die Empfehlung eines Algorithmus berücksichtigen, um zu prüfen, wie nah sie dem Ergebnis kamen und wie viel Geld sie verdienten.

ERGEBNISSE/ERKENNTNISSE:

- Die Literaturrecherche ergab, dass algorithmische Systeme nur dann effektiv sind, wenn sie menschliches Verhalten verstehen und Prognosen zu Verhaltenstrends abgeben können, ohne Verhaltensmuster radikal zu verändern. Zudem wurden Forschungslücken aufgezeigt, insbesondere bezüglich der Auswirkungen von Transparenz und Interpretierbarkeit von KI auf menschliches Verhalten.
- In der Studie „Ökonomische Verhandlung zwischen zwei Akteuren“ fühlten sich nur wenige Personen bei Verhandlungen zwischen Menschen benachteiligt. Mit einer automatisierten Entscheidungshilfe fühlte sich die Person ohne technische Unterstützung benachteiligt und forderte mehr Geld, was das Risiko eines Misserfolgs erhöhte. Eine Erklärung zur Algorithmus-Empfehlung verbesserte das Gefühl, aber die Verhandlung blieb als unfair empfunden. Dies zeigt, dass Algorithmen das Verhalten der Menschen beeinflussen.
- In der zweiten Studie erwarteten 49 % der Arbeitnehmer*innen ein höheres Gehalt bei Verhandlungen mit einer KI. Dennoch entschieden sich 52 % für Verhandlungen mit einem Menschen. Dies deutet darauf hin, dass soziale Faktoren wichtiger sind als ökonomische bei der Ablehnung von KI. Die Verhandlung mit einer KI führte zu höheren Geldforderungen, was zeigt, dass KI die Entscheidungsfindung verändert, ohne monetäre Anreize zu nutzen.
- Die Studie „Algorithmus Aversion“ zeigte, dass Proband*innen einem Algorithmus nur dann vertrauen, wenn er keine Fehlentscheidungen trifft. Auch nach dem Lernen aus Fehlern blieb das Vertrauen gering.

KÜNSTLICHE INTELLIGENZ

VERTRAUENSWÜRDIGKEIT VON KÜNSTLICHER INTELLIGENZ

FORSCHUNGSINTERESSE:

Wie kann die Vertrauenswürdigkeit von Künstlicher Intelligenz gesteigert werden?

VORGEHEN:

Die Wissenschaftler*innen erstellten ein Survey zum Thema „Kausalität für vertrauenswürdige KI“. Ein Survey gibt einen Überblick über aktuelle Forschungsinhalte innerhalb eines Forschungsgebietes. Dies dient einem besseren Verständnis für das Forschungsgebiet und soll andere Wissenschaftler*innen zum Forschen animieren.

ERGEBNISSE/ERKENNTNISSE:

- Das Survey beleuchtet, wie man KI mit dem Konzept des kausalen Denkens vertrauenswürdiger machen kann. Es wird dargestellt, wie Wissenschaftler*innen es geschafft haben, mittels Kausalität KI-Systeme interpretierbarer, fairer, robuster, die Privatsphäre besser schützend und sicherer zu gestalten.
- Es werden verfügbare Datensätze, Werkzeuge und Pakete aufgelistet, die für die kausalitätsbasierte Forschung und Entwicklung vertrauenswürdiger KI relevant sind. Um einen Einblick zu geben, was Kausalität in diesem Zusammenhang bedeutet, wird ein Beispiel aus der Robustheit vorgestellt: Hier besteht das Problem der KI darin, falsche Zusammenhänge bei der Auswertung von Daten zu nutzen, wenn die Datengrundlage verändert wird.

KÜNSTLICHE INTELLIGENZ

REPRÄSENTATIVITÄT UND ROBUSTHEIT VON KÜNSTLICHER INTELLIGENZ

FORSCHUNGSINTERESSE:

- Wie repräsentativ sind die Entscheidungen einer Künstliche Intelligenz (KI) zur Bilderkennung?
- Wie robust sind die Entscheidungen einer Künstliche Intelligenz (KI) zur Bilderkennung?

VORGEHEN:

- Bei der Repräsentativität Künstlicher Intelligenz geht es um den Zusammenhang zwischen Trainingsdaten und den Entscheidungen der KI, während die Robustheit beschreibt, wie die KI auf kleine Änderungen in den Daten reagiert.
- Die Wissenschaftler*innen trainierten ein KI-Modell zur Bilderkennung und testeten es mit neuen Daten, um sicherzustellen, dass die Entscheidungen der KI nicht auf irrelevanten Merkmalen basieren. Die Auswahl der Trainingsdaten ist entscheidend für die Leistung der KI. Die KI sollte nicht nur grobe Unterschiede zwischen Objekten erkennen, sondern auch kleine Veränderungen in der Datengrundlage korrekt einordnen. Wenn eine Änderung in den Bilddaten dazu führt, dass die KI ihre Entscheidung erheblich ändert, ist sie nicht robust. Das Modell sollte daher in der Lage sein, relevante Merkmale korrekt zu identifizieren.
- Um zu verstehen, was das KI-Modell gelernt hat, manipulierten die Wissenschaftler*innen verschiedene Modellkomponenten und änderten die Trainingsdaten. Sie testeten das Modell, indem sie Pixel im Bild veränderten und überprüften die Ergebnisse. Nachdem die KI eine Entscheidung getroffen hatte, bewertete ein Optimierungsalgorithmus, ob diese korrekt war. Das Modell

musste dann lernen, warum es falsch entschieden hatte, indem es die relevanten Merkmale identifizierte und daraus Rückschlüsse für zukünftige Entscheidungen zog.

ERGEBNISSE/ERKENNTNISSE:

Stück für Stück verbesserte sich das KI-Modell. Die KI wurde also im Laufe der Versuche immer robuster. Die Wissenschaftler*innen konnten in diesem Zusammenhang auch zeigen, dass es eine Verbindung gibt zwischen der Robustheit von Daten (Wie stark reagieren die Label auf minimale Veränderungen der Datengrundlage?) und der Robustheit des KI-Modells (Wie stark reagiert die KI auf minimale Veränderungen der Datengrundlage?).

KÜNSTLICHE INTELLIGENZ

GENERALISIERBARKEIT VON KÜNSTLICHER INTELLIGENZ

FORSCHUNGSINTERESSE:

Wirkt sich die Robustheit von Daten positiv auf die Generalisierbarkeit von KI aus?

VORGEHEN:

- Hintergrund: Neben der Repräsentativität und Robustheit ist auch die Generalisierbarkeit eine wichtige und erforderliche Eigenschaft Künstlicher Intelligenz. KI muss in der Lage sein zu verallgemeinern. Dafür muss die KI lernen, von einzelnen Merkmalen auf die Gesamtheit zu schließen. Wenn sie z. B. mit Bildern von einfachen Küchenstühlen trainiert wird, soll sie auch Bürodrehstühle als Stühle erkennen, obwohl diese anders aussehen. Erkennt das Modell, welche funktionalen Merkmale mit dem Objekt übereinstimmen (sitzen, anlehnen), und überträgt dies auf anders gestaltete Objekte mit derselben Funktion, kann die KI generalisieren. Die Generalisierbarkeit erhöht das Einsatzspektrum der KI, da sie nicht nur für den einen trainierten Fall eingesetzt werden kann, sondern für weitere Anwendungsfälle mit ähnlichem Kontext.
- Die Wissenschaftler*innen des Zukunftslabors führten Versuche durch, um zu testen, ob sich die Robustheit von Daten auch positiv auf die Generalisierbarkeit der KI-Modelle auswirkt.

ERGEBNISSE/ERKENNTNISSE:

Die Wissenschaftler*innen identifizierten Verfahren, die Robustheitsinformationen nutzen.