Die Achillesferse der Künstlichen Intelligenz: Warum Diskriminierung ein ungelöstes Problem bleibt

Bericht: Duale Hochschule Baden-Württemberg

Eine aktuelle Studie der DHBW Stuttgart am Studienzentrum Dienstleistungsmanagement hat die Fähigkeit der Künstlichen Intelligenz (KI) zur Erkennung diskriminierender Inhalte in Bildern und Werbeanzeigen untersucht und dabei sowohl beeindruckende Fortschritte als auch bestehende Grenzen aufgezeigt.

In der umfassenden Untersuchung wurde die KI mit einer Vielzahl von Bildern und Werbeanzeigen konfrontiert und um eine Bewertung gebeten. Darunter befanden sich unter anderem 60 Werbeanzeigen, die der Deutsche Werberat in der jüngeren Vergangenheit gerügt hatte. Die erzielten Ergebnisse zeigen, dass die KI eine erstaunlich hohe Fähigkeit besitzt, Diskriminierungen in Werbeanzeigen mit einer beeindruckenden Treffsicherheit zu identifizieren (F1 Score: 0.949). So wurden die vom Deutschen Werberat gerügten Werbeanzeigen in aller Regel von der KI als potenziell diskriminierend eingeschätzt. Gleichzeitig erfolgte bei den nichtdiskriminierenden Werbeanzeigen ein solcher Hinweis in den meisten Fällen nicht.

Dies ist insbesondere vor dem Hintergrund beeindruckend, da die KI noch vor zehn Jahren erhebliche Schwierigkeiten hatte, die in einem Bild dargestellten Objekte korrekt zu klassifizieren. In der Zwischenzeit und dank Millionen von Bildern hat die KI gelernt, weitgehend treffsicher zwischen einem Hund und einer Katze zu unterscheiden. Die rasanten Fortschritte der KI werfen die Frage auf, wo derzeit die Grenzen der KI verlaufen. „Wir wollten wissen, inwiefern die KI ein diskriminierendes Verhalten erkennt, wenn man dieser nur eine Werbeanzeige zur Verfügung stellt und diese auffordert, die Anzeige zu bewerten“, erläutert Studentin Helen Beckers das Vorgehen bei der Studie.

Algorithmische Diskriminierung

Dies ist von besonderer Relevanz angesichts des massiven Anwachsens von Diskriminierung durch Algorithmen, die Menschen aufgrund von Geschlecht, Religion, Weltanschauung, Rassismus oder Herkunft benachteiligen können. Dieser Aspekt gewinnt an Bedeutung, da Diskriminierung durch Algorithmen zunehmend zu einem ernsten Problem wird, das verschiedene Bereiche wie Bewerbungsverfahren, Kreditvergabe, Medizin und die Berechnung der Rückfallwahrscheinlichkeit von Straftäter*innen betrifft.

ChatGPTs Fähigkeit zur Aufdeckung von Sexualisierung und Stereotypen

Besonders aufschlussreich ist die Erkenntnis, dass ChatGPT Sexualisierung und stereotypisches Denken aufdecken kann. Eine modifizierte Anzeige mit vertauschten Geschlechterrollen zeigte, dass die KI auch in umgekehrten Konstellationen Diskriminierung identifizieren kann. „Diese unterschiedliche Bewertung durch ChatGPT in den zwei Szenarien unterstreicht die Fähigkeit, Diskriminierung auch in umgekehrten Konstellationen zu identifizieren“, nennt Student Sven Peter als weitere Erkenntnis aus der Studie.

Grenzen und Herausforderungen bei der Diskriminierungserkennung durch KI

An die Grenzen stieß die KI jedoch bei der Identifizierung weiterer Formen der Diskriminierung, wie der Objektivierung, der Respektlosigkeit sowie des Machtmissbrauchs. „Die Ergebnisse der Studie unterstreichen die Notwendigkeit, KI-Systeme weiterzuentwickeln, um Diskriminierung effektiver zu erkennen und zu verhindern“, berichtet Student Marius Funk. Die Studie wirft wichtige Fragen auf, wie KI-Technologien in der Zukunft eingesetzt werden können, um Diskriminierung in verschiedenen Bereichen zu bekämpfen und die Gleichberechtigung zu fördern.

Studienzentrum Dienstleistungsmanagement

Die Studie wurde unter Leitung von Prof. Dr. U. Bucher am Studienzentrum Dienstleistungsmanagement an der DHBW Stuttgart durchgeführt. Das Studienzentrum Dienstleistungsmanagement bereitet die Studierenden auf verantwortungsvolle betriebswirtschaftliche Tätigkeitsfelder in innovativen und kundenorientierten Projekten sowie in vielfältigen Funktions- und Servicebereichen von Dienstleistungsunternehmen vor.

Die DHBW Stuttgart:

Die Duale Hochschule Baden-Württemberg Stuttgart gehört mit rund 8.000 Studierenden in Bachelor und Masterstudiengängen zu den größten Hochschuleinrichtungen in den Regionen Stuttgart und Oberer Neckar. In Kooperation mit rund 2.000 ausgewählten Unternehmen und sozialen Einrichtungen bietet sie mehr als 60 national und international anerkannte, berufsintegrierte Studienrichtungen in den Bereichen Wirtschaft, Technik, Sozialwesen und Gesundheit an. Die Bachelor-Studierenden wechseln im dreimonatigen Rhythmus zwischen der Hochschule und dem Ausbildungsbetrieb, dem Dualen Partner. Auf diese Weise sammeln sie bereits während der Studienzeit Berufserfahrung. Finanzielle Unabhängigkeit, sichere Chancen auf dem Arbeitsmarkt durch außergewöhnlich hohe Übernahmequoten, kleine Kursgruppen und die internationale Ausrichtung bieten den Studierenden enorme Vorteile. Über den Bachelorabschluss hinaus bietet die DHBW ein duales Masterprogramm an. Weitere Informationen unter: www.dhbw-stuttgart.de

8.1.2024
Carolin Höll
Hochschulkommunikation
Duale Hochschule Baden-Württemberg
www.dhbw.de