

KI und Bias: Vortrag und Podiumsdiskussion zum Internationalen Frauentag vom 8. März 2024

Vom Umgang mit KI-generierter Diskriminierung

Am Internationalen Women's Day 2024 organisierte der DLH eine Podiumsdiskussion zum aktuellen Thema KI und Bias: Das Input-Referat hielt Sara Dolf-Metzler, MA-Management and Economics UZH, Managing Partner bei Diversidad, Fachspezialistin Diversität und Inklusion bei der SRG. Auf dem Podium diskutierten Marko Kovic, Soziologe und Dozent u.a. an der Hochschule Luzern, Werner Winkelmann, Informatik-Lehrperson an der Kantonsschule Küsnacht, sowie Anna Mätzener, Datenspezialistin und Co-Leiterin Strategische Initiative Daten an der Staatskanzlei des Kantons Zürich. Moderiert wurde der Anlass von Michael Grossenbacher, Projektleiter Digitalisierung und Bildung des Kantons Bern.

In ihrem dichten, forschungsbasierten Referat «KI und Bias – Diversität intelligent nutzen und Bias reduzieren» zeigte Dolf-Metzler zuerst anhand von Beispielen den Nutzen von Diversität auf: Diverse Teams arbeiteten produktiver, innovativer und risikobalancierter. Dass es jedoch mit der Gleichstellung in der Schweiz, auch im Bildungsbereich, noch nicht wirklich weit her ist, erläuterte sie anhand des Global Gender Gap Report des WEF von 2023¹: Die Schweiz ist demnach von Platz 13 auf den Rang 21 zurückgefallen. Der sog. Glass Ceiling Effekt (gläserne Decke, an die qualifizierte Frauen in Unternehmen und Organisationen stossen) zeige sich sowohl im Topmanagement von Unternehmen als auch bei der Verteilung von Professuren an den Schweizer Hochschulen (26% Professorinnen).

Unterschiedliche Formen von Bias

Aus der Forschung ist eine Vielzahl von Bias-Formen bekannt. Der sogenannte unconscious Bias als unbewusste Verzerrung der Wahrnehmungen bringt nicht nur Nachteile. Z.B. können damit beim Thinking fast Wissenslücken kompensiert oder schnelle Entscheidungen gefällt werden. Allerdings ist er auch schwierig zu kontrollieren und verstärkt sich bei Stress. Gerade in homogenen Gruppen kann es dadurch zu Fehleinschätzungen und Diskriminierungen kommen. Für den Bildungsbereich interessant ist der Performance Bias: Hierbei geht es um inkorrekte Einschätzungen von Fähigkeiten. Dolf-Metzler zitierte aus der Studie «Statesmen or Cheerleaders?» von 2019², wonach Leistungen von Frauen tendenziell unterschätzt und jene von Männern eher überschätzt werden. Nur 10% der Menschen schätzen demnach ihre Kompetenzen realistisch ein.

KI und Bias

Anschliessend gab die Referentin den Zuhörenden eine Aufgabe auf: Man sollte sich eine Person vorstellen, die an einer Hochschule studiert. Anschliessend zeigte sie ein KI-generiertes Bild als Antwort auf diese Frage. Der «highly gifted student» wird, wen wunderts, als männliche Person gezeigt. Die Referentin schlussfolgerte, dass KI als maschinelles Lernen Ungleichheiten und Diskriminierung verstärken könne, in dem sie menschliche Bias reproduziere. Dabei basiert die Verzerrung auf einer systemischen (Data und Algorithmic Bias) oder einer menschlichen Ebene (Interpretation Bias), was sich gegenseitig beeinflussen kann, aber zugleich auch auf Möglichkeiten verweist, Bias zu vermeiden.

Bias in KI-Systemen reduzieren

Zum Schluss ihres Referats zeigte Dolf-Metzler in Form von fünf Empfehlungen auf, wie Bias im Zusammenhang mit KI reduziert werden könne und schlug den Bogen zur Bildung: KI-Systeme müssten mit einer möglichst vielfältigen Datenbasis gefüttert werden. Algorithmen sollten möglichst fair und unverzerrt designed werden. Auch gehe es darum, Programmierende und Nutzende zu sensibilisieren und auf einer rechtlich-politischen Ebene die Entwicklung und Nutzung von KI-Systemen zu regulieren. Für die Schule interessant könne der Ansatz sein, mit Künstlicher Intelligenz Verzerrungen und Diskriminierungen sichtbar zu machen, z.B. bei der Bildergenerierung (vgl. das Beispiel des «highly gifted students»), da gerade Bilder länger in Erinnerung blieben und assoziativer als Texte wirkten. In der anschliessenden Diskussion thematisierten die Podiumsteilnehmenden u.a. verzerrte KI-Anwendungen im Alltag (z.B. Gesichtserkennung am Handy), den Versuch von Tech-Unternehmen, krasse Bias zu korrigieren und die Möglichkeit mittels Software Texte hinsichtlich Inklusion zu überprüfen. Einig war man sich darin, dass Lehrpersonen hinsichtlich KI vor grossen Herausforderungen stehen. Sensibilisierung, Aufklärung, aber auch ein reflektierter spielerischer Umgang mit KI wurden als

¹ <http://reports.weforum.org/global-gender-gap-report-2023>, abgerufen 8.3.2024.

² Dolder, Wyatt, Silvester: Statesmen or cheerleaders, 2019.

wichtige Anwendungsmöglichkeiten im Unterricht angesprochen. Kinder und Jugendliche sollten lernen, Resilienz aufzubauen, und Werkzeuge in die Hand bekommen, die das kritische Denken fördern, um Bias-Effekte, seien sie KI generiert oder auch nicht, erkennen zu können.

Bei der anschließenden Öffnung der Diskussionsrunde wurde u.a. die Wichtigkeit des Nachdenkens über KI-Produkte in der Schule betont, aber auch wie Trainingsdaten für KI-Systeme zustande kommen und welche ökonomischen Interessen damit verbunden sind. Einig war man sich darin, dass auch trotz Datenschutz und einem hohen Regulierungsgrad nicht davon ausgegangen werden kann, dass Bias-Effekte in KI-Systemen mit 100% Garantie eliminiert werden könnten. Was die Wichtigkeit der kritischen Auseinandersetzung mit der Thematik in der Schule unterstreicht.

Der Moderator Michael Grossenbacher schloss die Veranstaltung mit der Bemerkung, dass die letzten eineinhalb Stunden wie im Flug vergangen seien, und bedankte sich bei der Referentin, den Podiumsteilnehmenden und beim Publikum für die angeregte Diskussion.

Literaturangaben

Bildungsbericht 2023

Dolder, E., Wyatt, M., Silvester, J.: Statesmen or cheerleaders? Using topic modeling to examine gendered messages in narrative development feedback for leaders, The Leadership Quarterly, volume 30, Issue 2019.

2022 Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten

2023 Guidance for generative AI in education and research

2023 Women in workplace reports

2023 Global gender gap report.

<https://science.apa.at/mehrzumthema/man-kann-nicht-nicht-diskriminieren/>

<https://ulb-dok.uibk.ac.at/download/pdf/8090092.pdf>

<https://deutsches-schulportal.de/schulkultur/lehrkraefte-merken-oft-nicht-dass-sie-stereotype-reproduzieren/>

Ressourcen für den Unterricht

www.codedbias.com

Der KI-Chatbot von Google antwortet voreingenommen und falsch (nzz.ch)

<https://lehrerweb.wien/praxis/praxis-ideen/idea/130?cHash=ebd430aab088c67a96f6aa53b24d613c>