

Jahrestagung der Fachgruppe Frauen und Informatik

„Künstliche Intelligenz - vertrauenswürdig, erklärbar, fair?“

Ursula Köhler, Gudrun Schiedermeier, Margita Zallmann

Die Jahrestagung 2021 der Fachgruppe Frauen und Informatik fand als virtuelle Tagung am 24. April 2021 zum Thema „Künstliche Intelligenz – vertrauenswürdig, erklärbar, fair?“ mit über 70 Teilnehmer*innen statt.

Zunächst wird kurz dargestellt, warum wir das Thema gewählt haben und anschließend ein Überblick über die Tagung gegeben.

Der Begriff Künstliche Intelligenz (KI) ist momentan omnipräsent in allen Medien, Diskussionsrunden oder Tagungen verschiedenster Fachrichtungen. Dabei gibt es keine allgemeingültige bzw. von allen Akteur*innen konsistent genutzte Definition von KI.

Alle Beiträge unserer Tagung bewegten sich im Bereich des maschinellen Lernens (ML), einem Teilgebiet von KI. KI hat schon eine längere Vergangenheit, aktuell befinden wir uns in der dritten Hype-Welle. Nach dem Beschreiben und Kategorisieren steht nun das Erklären im Fokus. Je mehr KI-Systeme unser Leben beeinflussen oder sogar bestimmen, umso wichtiger wird es, dass die Menschen und ihre Bedürfnisse bei den Entwicklungen beachtet werden (human-centered AI).

Die Plattform „Lernende Systeme“ beschreibt die Situation so: „Künstliche Intelligenz verändert unser Leben! Lernende Systeme durchdringen sämtliche Bereiche in Arbeitswelt und Alltag. Sie lösen vorgegebene Aufgaben selbstständig und reagieren auf ihre Umwelt. Das Verhältnis von Mensch und Maschine ändert sich damit grundlegend – und muss in allen Einsatzgebieten im Sinne des Menschen gestaltet werden.“ [1]

Dazu gehört auch, dass Verhalten und Ergebnisse der sogenannten intelligenten Systeme erklärbar, transparent und kontrollierbar sind. Die Schwerpunktsetzung unserer Tagung spiegelt unser Interesse an diesen Fragestellungen. Wir finden es beeindruckend und sind stolz darauf, dass alle Referentinnen mit ihren profunden Forschungen aus den Reihen unserer Fachgruppe kommen.

Ute Schmid (Universität Bamberg) hielt in ihrem Vortrag **Gemeinsam klüger - Erklärbares und interaktives maschinelles Lernen** ein Plädoyer für die Nutzung der KI zur kognitiven Unterstützung und Förderung der Menschen. Einleitend erklärte sie, dass aktuell KI oft synonym zu maschinellem Lernen, speziell zu tiefen neuronalen Netzen verwendet wird. Entsprechende Architekturen wie Convolutional Neural Networks ermöglichen es, dass Modelle, zum Beispiel für die Klassifikation von Bildern, direkt aus den Rohdaten gelernt werden. Ute Schmid führte weiter aus, dass die Ansprüche von tiefen Netzen an Menge und Qualität von Daten in vielen Bereichen kaum oder nur mit sehr hohen Kosten erfüllbar sind. Zudem ist es für viele Anwendungsbereiche unverzichtbar, dass transparent und nachvollziehbar ist, aufgrund welcher Information ein gelerntes Modell zu einer bestimmten Entscheidung gelangt ist. Entsprechend wird aktuell die dritte Welle der KI – explainable AI (XAI) ausgerufen. Es werden Methoden für erklärendes und interaktives maschinelles Lernen entwickelt, die eine Kooperation von Mensch und KI-Systemen ermöglichen. Dadurch kann menschliches Wissen in den Lernprozess einbezogen werden und menschliche Kompetenzen werden nicht beschnitten sondern ergänzt und erweitert.

In ihrem Vortrag **Vertrauenswürdigen maschinelles Lernen durch „Waschzettel“** berichtet **Katharina Morik** (Universität Dortmund) aus ihrer aktuellen Forschungsarbeit.

Katharina Morik setzt ähnlich wie die anderen Referentinnen den Schwerpunkt auf menschenorientiertes maschinelles Lernen. Aus dem zunehmenden Einsatz in unterschiedlichen gesellschaftlichen Bereichen resultiert ein großer Bedarf an Vertrauenswürdigkeit. Für dessen Gewährleistung seien die praktischen Herausforderungen in der Entwicklung von Anwendungen enorm, trotz der seit Jahren bekannten und eingesetzten ML-Methoden. Ihr Vortrag führte dies zur Dokumentation von Lernprozessen und Interpretation von Lernergebnissen genauer aus. Deutlich wurde, dass die besondere Problematik heutiger KI-Entwicklungen durch die riesigen Datenmengen und Modelle entsteht. Zudem gibt es ein Dickicht an Tools, deren Anwendung sehr viel Wissen über maschinelles Lernen erfordert.

Katharina Morik stellte vor diesem Hintergrund einen eigenen und neuen Ansatz der KI-Forschung vor, an dem sie gerade arbeitet. In Anlehnung an sogenannte Waschzettel, auf Waschmaschinen oder auch auf Kleidungsstücken verortbar, schlägt sie **Waschzettel für Methoden** vor. Indem strukturierte Symbole die Güte, den Energie- und Speicherverbrauch der Anwendung anzeigen, soll eine nachvollziehbare Dokumentation auf einen Blick geliefert werden.

Eine ganz andere Perspektive nahm **Gudrun Schiedermeier** (HaW Landshut) in ihrem Vortrag **Diskriminierende KI-Systeme** ein. Sie schilderte mögliche Formen und Ursachen von Diskriminierung und beleuchtete die Fairness-Thematik, auch im Hinblick auf Frauen. Nach einer Definition des Begriffs Diskriminierung wurden Beispiele für diskriminierende KI-Systeme vorgestellt. Zunächst beschrieb Gudrun Schiedermeier einen Algorithmus zur Klassifizierung von Arbeitslosen in Österreich, danach ein US-amerikanisches Automatisches Entscheidungssystem im Justizvollzug. Genauer wurde auf die Ungleichbehandlung aufgrund der Hautfarbe bei Gesichtserkennungs-Software eingegangen, sowie auf die Ungleichbehandlung aufgrund neuer Lernverfahren mit großen Datenmengen.

Im zweiten Teil behandelte Gudrun Schiedermeier die Frauenfeindlichkeit in KI-Systemen, dabei wurde die Frauenfeindlichkeit von Sprach-Assistent*innen und deren Auswirkungen auf die Gesellschaft diskutiert.

Im dritten Teil beschäftigte sich der Beitrag mit Lösungsansätzen zur Erkennung und Behebung von Diskriminierung. Vorschläge dazu umfassen einerseits die Verwendung einer gender-neutralen Stimme, andererseits den Bau einer feministischen Alexa oder auch eine diversere Besetzung der Entwicklungsteams. Nach Ansicht der Autor*innen des Whitepapers „Künstliche Intelligenz und Diskriminierung“ der AG3 Plattform lernende Systeme lassen sich die Diskriminierungsrisiken u.a. durch Erklärbarkeit und Überprüfung sowie effektiven Rechtsschutz und Rechtsdurchsetzung vermeiden. [2] Die EU-Kommission hat im April 2021 einen Gesetzesvorschlag zur Regulierung von KI-Systemen vorgestellt. Je nachdem, welche Risiken mit den KI-Systemen verbunden sind, unterliegen sie unterschiedlichen Regulatorien, angefangen bei KI-Systemen mit minimalem Risiko mit keinen regulatorischen Vorgaben, bis hin zum Verbot für gefährliche KI-Systeme wie z.B. Social Scoring. [3]

Es gilt zu hinterfragen, wer entscheidet und wer verdient an KI-Anwendungen, wer wird verletzt, wer ausgegrenzt oder belästigt. Diese Fragen betreffen nicht nur die Technik-Industrie und Regierungen, sondern jede Einzelne, jeden Einzelnen, deren Leben KI-Tools und -Dienste beeinflussen.

Lena Wiese (Goethe-Universität Frankfurt und Fraunhofer ITEM) fügte mit ihrem Impulsvortrag **Beipackzettel für Machine Learning Modelle** weitere Gedanken in Richtung Transparenz und Verständlichkeit hinzu. Auf maschinellem Lernen beruhende Entscheidungen werden so allgegenwärtig, dass viele Menschen auch in ihrem Alltag von ML-basierten Entscheidungen betroffen sein werden. Eine Transparenz bezüglich der Eigenschaften des zugrunde liegenden ML-Modells sollte daher für alle

Betroffenen verfügbar und verständlich sein. Basierend auf üblichen Verbraucherkennzeichnungen (wie zum Beispiel der Lebensmittelampel oder des Energieeffizienzlabels) schlugen Lena Wiese und ihre Mitautorinnen (Christin Seifert, Universität Duisburg-Essen und Stefanie Scherzinger, Universität Passau) vor, sogenannte „Beipackzettel“ für vortrainierte und veröffentlichte ML-Modelle zu erstellen. Diese Beipackzettel sollten genauso intuitiv verständlich sein wie andere Verbraucherkennzeichnungen und die Merkmale des trainierten ML-Modells umfassend darstellen. Diese Beipackzettel sollten strukturiert so aufgebaut sein, dass detaillierte Informationen zu einzelnen Bereichen (wie etwa Fairness für einzelne Betroffenenengruppen) flexibel abgerufen werden können. Anhand dieser Detailinformationen ist dann jede(r) in der Lage, die Anwendbarkeit des ML-Modells in dem jeweiligen Einsatzkontext beurteilen zu können.

Abschließend diskutierten alle Vortragenden und zusätzlich Meike Klettke (Universität Rostock) unter der Moderation von Christine Regitz (SAP) über Themen im Bereich KI und Ethik in der Forschung sowie der universitären Lehre. Diese lebendige **Podiumsdiskussion** traf bis zum Schluss auf großes Interesse seitens des Publikums.

Einigkeit herrschte bei den Expertinnen, dass wir die Entwicklung und den Einsatz von KI-Systemen nicht allein den Herstellern überlassen dürfen. Essenziell für die Förderung von Vertrauen in KI-Systeme ist vielmehr, dass herstellerunabhängige Instanzen Prüfungen und Risikoabschätzungen durchführen. Diese kritische Begleitung ist wichtig, aber es bedarf auch der Durchsetzung der Regelungen. Auf der anderen Seite enthalten die Entwicklungen ein innovatives Potenzial, das wir nutzen sollten.

Großen Anklang fand die **Verleihung des Preises der Fachgruppe** an Jana Eisoldt für eine herausragende Bachelorarbeit und die Verleihung des Preises an Judith Hemp für eine ebenso ausgezeichnete Masterarbeit. Die Preisträgerinnen wurden aus 70 Einreichungen ausgewählt. Der Preis dient dem Ziel, Frauen mit hervorragenden Ergebnissen im Informatik-Studium sichtbar zu machen. Weitere Informationen zur Preisverleihung finden sich auf der Website: <https://fg-frauen-informatik.gi.de/mitteilung/preisverleihung-der-fachgruppe-frauen-und-informatik>

Die in der folgenden Abbildung zusammengestellten Kommentare geben einen Eindruck von einem spannenden Tag mit vielen Anregungen für weitere Diskussionen zum Thema „Künstliche Intelligenz“. Auch die von Edna Kropp parallel geposteten Twitter-Tweets erzeugen eine große und positive Resonanz.



Abbildung 1: Ausgewählte Kommentare der Teilnehmerinnen, zusammengestellt von Phöbe Günzler

Links

- [1] <https://www.plattform-lernende-systeme.de/arbeitsgebiete.html>; abgerufen am 01.06.2021
- [2] <https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaeetze.html> ; abgerufen am 29.06.2021
- [3] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788 ; abgerufen am 29.06.2021

Dr. Ursula Köhler

Sprecherin Fachgruppe Frauen und Informatik

ulla.koehler@frauen-informatik.de

Prof. Dr. Gudrun Schiedermeier

gschied@haw-landshut.de

Margita Zallmann, Dipl. Inf.

Margita.zallmann@frauen-informatik.de