

Pressemitteilung

Universität Bielefeld

Jörg Heeren

10.06.2022

<http://idw-online.de/de/news795379>

Forschungsprojekte, Kooperationen
Gesellschaft, Informationstechnik, Philosophie / Ethik, Psychologie, Sprache / Literatur
überregional



Menschen erklären Maschinen, die sich selbst erklären

Mit Forschenden auf erster Konferenz des Transregios 318 diskutieren Die Prozesse des maschinellen Lernens sind für viele Menschen unverständlich. Im Sonderforschungsbereich/Transregio (SFB/TRR) 318 „Constructing Explainability“ (Erklärbarkeit konstruieren) der Universitäten Bielefeld und Paderborn erarbeiten Wissenschaftler*innen Wege, Erklärprozesse zu gestalten und Nutzer*innen die Erklärungen einer Künstlichen Intelligenz (KI) steuern zu lassen. Auf der ersten vom TRR 318 organisierten internationalen Konferenz liegt der Schwerpunkt auf der sozialwissenschaftlichen Forschung zu erklärender KI.

Die Tagung trägt den Titel „Explaining Machines“ und wird am 23. und 24. Juni im CITEC-Gebäude der Universität Bielefeld ausgerichtet. Die Konferenzsprache ist Englisch.

Eine der zentralen Fragen in der KI-Forschung ist derzeit, wie verständliche Erklärungen der dahinterliegenden maschinellen Prozesse erreicht werden können: Sollen Menschen in der Lage sein, zu erklären, wie Maschinen funktionieren, oder sollten die Maschinen selbst lernen, sich zu erklären?

„Die Zweideutigkeit im Titel unserer Konferenz ‚Explaining Machines‘ drückt verschiedene Möglichkeiten aus: Maschinen erklären sich selbst oder Menschen erklären Maschinen – oder vielleicht beides gleichzeitig“, sagt Professorin Dr. Elena Esposito. Die Bielefelder Soziologin leitet ein Teilprojekt des TRR 318 und organisiert die Konferenz gemeinsam mit ihrem Paderborner Kollegen Professor Dr. Tobias Matzner. Der Medienforscher leitet ebenfalls ein Transregio-Teilprojekt. „Wenn das Erklären von Maschinen eine soziale und politische Wirkung haben soll, reicht es nicht aus, dass die Erklärungen für Informatiker*innen verständlich sind“, so Matzner. „In die Erklärprozesse müssen unterschiedlich sozial situierte Menschen einbezogen werden – von Mediziner*innen über Rentner*innen zu Schüler*innen.“

Technische und soziale Herausforderungen von KI-Projekten

Das Organisationsteam hebt den interdisziplinären Fokus der Konferenz hervor. „Es reicht nicht aus, KI-Systeme nur mit der Expertise einer Fachrichtung zu entwickeln. Informatiker*innen, die die Maschinen entwerfen, müssen mit Sozialwissenschaftler*innen zusammenarbeiten, die untersuchen, wie Menschen und KI interagieren und unter welchen Bedingungen das geschieht“, sagt Esposito. „Die Herausforderung von KI-Projekten ist heute mehr denn je sowohl technischer als auch sozialer Natur. Mit der Konferenz möchten wir anregen, die Perspektiven und Erkenntnisse der Sozialwissenschaften mit in die Debatte um erklärende Maschinen einzubeziehen.“

Bislang arbeiten an der Erklärbarkeit von Künstlicher Intelligenz vor allem Informatiker*innen: „Während in diesem Forschungsansatz vor allem die Sichtweise besteht, dass Erklärbarkeit und Verstehen durch Transparenz entsteht – also dann, wenn möglichst alle Informationen verfügbar sind – ist eine alternative Sicht die der Ko-Konstruktion“, sagt die Linguistin und Transregio-Sprecherin Professorin Dr. Katharina Rohlfing von der Universität Paderborn. „In unserer Forschung betrachten wir Menschen nicht als passive Partner, die Erklärungen empfangen. Erklärungen entstehen

vielmehr an der Schnittstelle zwischen Erklärer*in und Erklärungsempfänger*in. Beide gestalten den Erklärprozess aktiv und arbeiten darauf hin, sich auf gemeinsame Ideen und Vorstellungen zu verständigen. Für die Erforschung von Erklärbarkeit ist eine fächerübergreifende Zusammenarbeit essentiell.“

Vortragende aus Medienforschung, Philosophie, Rechtswissenschaft und Soziologie

Zum Veranstaltungsprogramm gehören zehn Kurzvorträge. Unter den internationalen und renommierten Gästen sind der Experte für rechtliche Aspekte der Künstlichen Intelligenz Professor Dr. Frank Pasquale, die Juristin und Philosophin Professorin Dr. Mireille Hildebrandt, der Philosoph zum Thema Internet Dr. David Weinberger, die Philosophin zum Thema KI-Ethik Dr. Cansu Cansa, der Soziologe Dr. Dominique Cardon, die Rechtswissenschaftlerin Dr. Antoinette Rouvroy und der Medienforscher Dr. Bernhard Rieder. Im Anschluss an die Impulsvorträge haben die Teilnehmenden die Gelegenheit, die Forschungstexte, die ihnen nach ihrer Anmeldung per E-Mail zugehen, mit den Vortragenden zu diskutieren.

Die Konferenz „Explaining Machines“ ist die erste große wissenschaftliche Veranstaltung des Transregios 318. In den kommenden drei Förderjahren sind weitere Fachkonferenzen geplant, die das Konzept der Erklärbarkeit aus unterschiedlichen wissenschaftlichen Disziplinen zum Thema haben.

Journalist*innen sind eingeladen, über die Veranstaltung zu berichten. Eine Anmeldung ist erforderlich bei: communication@trr318.uni-paderborn.de. Die Tagungsleiter*innen und Transregio-Sprecher*innen stehen während der Konferenz vor Ort für Medienanfragen zur Verfügung.

Sonderforschungsbereich/Transregio (TRR) 318

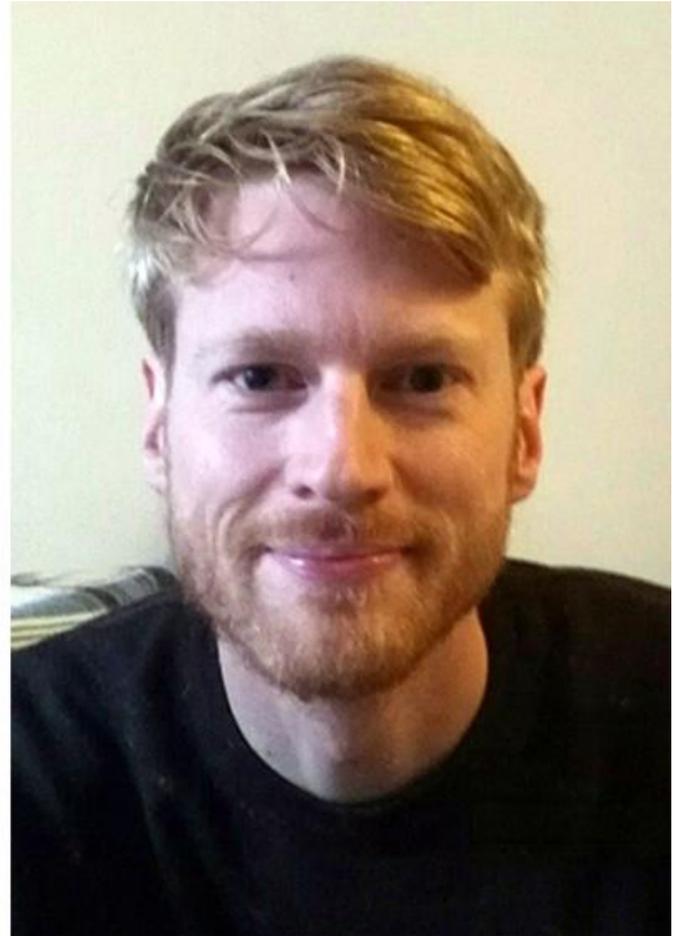
Das stark interdisziplinär ausgerichtete Forschungsprogramm mit dem Titel „Constructing Explainability“ („Erklärbarkeit konstruieren“) geht über die Frage nach Erklärbarkeit von KI als Grundlage algorithmischer Entscheidungsfindungen hinaus. Der Ansatz fördert die aktive Teilnahme der Menschen an soziotechnischen Systemen. Ziel ist es, die Mensch-Maschine-Interaktion zu verbessern, das Verständnis von Algorithmen in den Mittelpunkt zu stellen und dieses als Produkt eines multimodalen Erklärprozesses zu untersuchen. Die Deutsche Forschungsgemeinschaft (DFG) stellt dafür bis Juli 2025 rund 14 Millionen Euro Fördergelder zur Verfügung.

wissenschaftliche Ansprechpartner:

Linda Thomßen, Universität Bielefeld
Referentin für Wissenschaftskommunikation, TRR 318
Telefon: 0521 106-86695
E-Mail: communication@trr318.uni-paderborn.de

URL zur Pressemitteilung: <https://trr318.uni-paderborn.de/explaining-machines> Website zur Konferenz „Explaining Machines“

URL zur Pressemitteilung: <https://trr318.uni-paderborn.de/forschungsprofil> Forschungsprofil des Transregios 318



Sie übernehmen die Konzeption der Konferenz: Prof'in Dr. Elena Esposito (li.) von der Universität Bielefeld und Prof. Dr. Tobias Matzner (re.) von der Universität Paderborn.
Foto li.: Universität Bielefeld/ Michael Adamski, Foto re.: Universität Paderborn



Wie können Erkenntnisse aus den Sozialwissenschaften die Debatte um erklärende Maschinen bereichern? Das ist ein Aspekt, um den es auf der ersten Konferenz des Transregios 318 geht.
Foto: TRR 318