CONSTRUCTING EXPLAINABILITY

TRR 318

Erklärungen gemeinsam entwickeln | 02.2023

Im Newsletter des Transregios (TRR) 318 „Constructing Explainability" präsentieren wir Forschungsprojekte, Workshops zu künstlicher Intelligenz (KI) sowie aktuelle Nachrichten und Vorträge. Sie sind herzlich eingeladen, mit uns auf Twitter zu interagieren und uns eine E-Mail mit Ihrer Frage zu KI zu schreiben. Lassen Sie uns gemeinsam Erklärungen entwickeln!

english version below: click here ▼

Halbzeit der ersten Förderphase

Im Juli 2021 fiel der Startschuss für den TRR 318 Constructing Explainability. Die Förderung von der Deutschen Forschungsgemeinschaft (DFG) beläuft sich zunächst auf vier Jahre, danach kann das Projekt um zwei weitere Phasen verlängert werden. Für die erste Förderphase heißt das aber erst einmal: Halbzeit! Viele Projekte haben erste Ergebnisse erzielt und Forschungsartikel veröffentlicht. Aber es laufen auch neue Projekte an: So begrüßen wir im Juli Suzanna Alpsancar und das Teilprojekt B06 im TRR. Das Forschungsprojekt zur erklärbaren KI und Ethik bringt die wissenschaftliche Disziplin Philosophie in den Transregio. Wir schauen gespannt auf die kommenden zwei Jahre!

Neues Teilprojekt über Ethik von erklärbarer KI

Die Deutsche Forschungsgemeinschaft (DFG) fördert ab Juli das neue Teilprojekt B06 im Forschungsbereich „B - Soziale Praxis“. Ziel des Projekts ist es, einen ethischen Rahmen für das Erklären von Künstlicher Intelligenz zu schaffen. Juniorprofessorin Dr. Suzana Alpsancar leitet das Teilprojekt gemeinsam mit Professor Dr. Tobias Matzner (bereits Leiter des Teilprojekts B03).

Weiterlesen

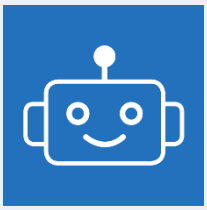
News

Gastwissenschaftler Marco am TRR 318

Marco Matarese ist Doktorand an der Universität Genua und unterstützt als Gastforscher aktuell das TRR-Teilprojekt A01„Adaptives Erklären“. Im Interview erzählt er von seiner bisherigen Zeit am Transregio, was er mit seinen Kolleg*innen erforscht und worin für ihn das Potenzial und die Gefahren von KI liegen.
Weiterlesen

Handgesten, Kopfnicken und Co.

Wenn wir die Regeln eines Brettspiels erklären, nutzen wir oft multimodale Signale; das sind nonverbale Formen der Kommunikation wie Mimik, Kopfnicken oder zeigende Handgesten. Die Forschenden des Projekts A02 haben diese Signale in Brettspiel-Erklärungen untersucht und halten die Ergebnisse in einem Daten-Korpus fest.
Weiterlesen



Der zögernde Roboter

Oft geraten Menschen beim Erklären ins Stocken, weil sie über ihre nächsten Sätze nachdenken müssen. Sollten Roboter bei Erklärungen diese natürlichen Verzögerungen imitieren? In einer neuen Studie der Universitäten Bielefeld und Bremen prüfen Wissenschaftler*innen, wie Menschen den Angaben eines Roboters besser folgen können.
Weiterlesen

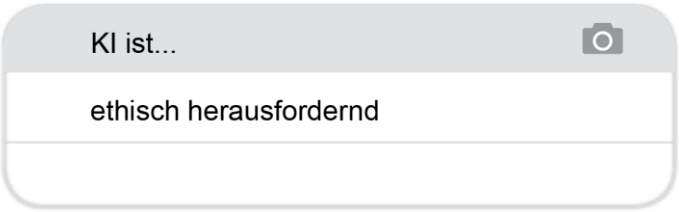
Neue Medien-Webseite

Das Projekt Öffentlichkeitsarbeit dokumentiert die Forschung des Transregios inzwischen auf einigen digitalen Kanälen. Eine neue Medien-Seite auf der Webseite stellt nun eine Übersicht aller verfügbaren Plattformen bereit. So sind dort auch die vergangenen Newsletter-Ausgaben zum Nachlesen zu finden. **Mehr entdecken**

KI aus künstlerischer Sicht betrachtet

Im Kunstmuseum Marta Herford ist ab dem 17. Juni die Ausstellung „SHIFT. KI und eine zukünftige Gemeinschaft“ zu sehen. Besucher*innen haben hier auch die Möglichkeit, an Workshops und Vorträgen teilzunehmen. **Mehr erfahren**

Weitere News



„Aktuell werden KI-Systeme in vielen Bereichen in atemberaubendem Tempo top-down eingeführt. Das bedeutet, dass eine zentrale Instanz entscheidet und alle untergeordneten Bereiche folgen. KI-Systeme haben aber oft eine erhebliche politische und gesellschaftliche Tragweite. Die Auswirkungen können wir noch gar nicht richtig einschätzen und sie werden im aktuellen Hype kaum berücksichtigt. Eine der größten ethischen Herausforderungen besteht derzeit in der rasanten Verbreitung bei (noch) fehlender Regulierung und Evaluation dieser Systeme. Unser Ziel ist es daher, KI-Systeme interdisziplinär und ko-konstruktiv zu entwickeln. Das heißt: Wir möchten eng mit den zukünftigen Nutzer*innen zusammenarbeiten, die Systeme auf sie abstimmen und gemeinsam auswerten. Damit bauen wir zum einen das Potenzial für eine demokratisch legitimierte gesellschaftliche Nutzung von KI-Systemen auf. Zum anderen können wir so die Expertise, Praktiken und gesellschaftlich relevanten Anforderungen in die technische Gestaltung der Systeme

einbringen. Letztlich können wir damit auch das technische Entwicklungspotenzial von KI-Systemen besser erschließen.“

Jun.-Prof. Dr. Ilona Horwath
*Projektleiterin der Teilprojekte **B03** und **Ö***

Medientipps

Artikel vom Spektrum „Wie eine KI lernt, sich selbst zu erklären“
Große Sprachmodelle wie ChatGPT und Co. haben das Portenzial für einen erheblichen Nutzen, bergen aber auch einige Risiken in ihrer Nutzung. Mit dem Ansatz der erklärbaren KI (XAI) sollen die Systeme ihre Antworten nun erklären können – zumindest teilweise. **Artikel lesen**

Artikel vom Lamarr-Blog „Wie eine Maschine Objekte erkennt“
Erkennung und Identifizierung von Autos bei Fahrassistenten oder Prozesskontrolle in einem Wirtschaftsunternehmen – diese und mehr Fälle erfordern das Erkennen und Verorten von Objekten. Im Feld des „maschinellen Lernens“ gibt es für Objekterkennung verschiedene Methoden und Verfahren. **Artikel lesen**

Veröffentlicht

Forschungsartikel: DAGA-Paper
Gängige Sprachtrennungssysteme zerlegen ein Sprachsignal in zwei verschiedene Einbettungen: eine Inhaltseinbettung und eine Sprechereinbettung. Um zu verstehen, wie die Informationen in der Sprecher-Einbettung kodiert sind, wird in dieser Arbeit ein solches System untersucht. Die Ergebnisse könnten eine Manipulation dieser Einbettungen in eine gewünschte Richtung mit veränderten akustischen Eigenschaften ermöglichen. **Weiterlesen**

Forschungsartikel: Halting the Decay of Talk
In dieser Publikation wird untersucht, wie Menschen mit atypischen körperlichen Fähigkeiten in der virtuellen Realität (VR) interagieren und wie sie die interaktionellen Herausforderungen in diesen neuen sozialen Umgebungen bewältigen. Die Ergebnisse stehen im Zusammenhang mit einer erneuten Diskussion über die Beziehung zwischen Handlung und Umgebung und der ko-konstruierten Natur situierten Handelns. **Weiterlesen**

Gesucht

Teilnehmende für Studien

Eine Auswahl:
- Gesellschaftsspiele erklären und erklärt bekommen,
- sich von virtuellen Agenten anleiten lassen,
- Alltagsbegegnungen mit künstlicher Intelligenz festhalten,
- Robotern Bewegungen beibringen.

Zur Webseite für mehr Informationen zu den aktuellen Studien. Oder: Mailingliste mit Einladungen zu neuen Studien **abonnieren.**

Schulklassen für KI-Workshops an der Universität Paderborn

Ein TRR-Forschungsteam aus dem Fachbereich Didaktik der Informatik bietet für Schulklassen Workshops zu KI an. Vorwissen ist nicht erforderlich, die Termine werden individuell vergeben. Interessierte Lehrkräfte sind eingeladen, sich per E-Mail anzumelden: **communication@trr318.uni-paderborn.de**

Was habe ich gelernt?

„Ergebnisse und Datenquellen zu dokumentieren, ist Teil des Datenmanagements und ein Qualitätsmaß für wissenschaftliches Arbeiten. Das wurde mir während des für den TRR 318 angebotenen Workshops ‚Data Sharing & Open Science' bewusst. Es ist unerlässlich, dass wir Forschenden unsere Ergebnisse auch einige Jahre später noch rekonstruieren und validieren können. Und wenn wir unsere Forschungsdaten teilen und offen zugänglich machen, können auch andere Wissenschaftler*innen darauf aufbauen und ihre eigenen Forschungsprojekte verbessern. Um die eigenen Veröffentlichungen entsprechend zu schützen, waren Lizenzen ein Thema des Workshops, die im normalen Wissenschaftsalltag leider viel zu kurz kommen. Wir alle wissen, dass diese Methoden wichtig sind, wenden sie aber in der Praxis oft nicht konsequent an. Der Workshop hat mir noch einmal vor Augen geführt, wie wichtig es ist, immer wieder darauf zu achten."

André Groß
*Doktorand im **Teilprojekt A05***

TRR digital

Oder direkt per Mail mit Fragen oder Feedback an uns.

Newsletter abonnieren.

Top: german version ▲
Footer: Impressum ▼

Developing explanations together | 02.2023

In the newsletter of Transregio (TRR) 318 "Constructing Explainability" we present our research projects, workshops on artificial intelligence (AI) as well as new publications and upcoming talks. You are invited to interact with us on Twitter and email us with your questions about AI. Let's develop explanations together!

Half-time of the first funding phase

In July 2021, the starting signal was given for TRR 318 Constructing Explainability. The funding from the German Research Foundation (DFG) is initially for four years, after which the project can be extended by two further phases. For the first funding phase, however, this means half-time for now! Many projects have achieved initial results and published research articles. But new projects are also starting up: For example, in July we welcome Suzanna Alpsancar and subproject B06 to the TRR. The research project on explainable AI and ethics brings the scientific discipline of philosophy to the Transregio. We are looking forward to the next two years!

New subproject on ethics of explainable AI

The German Research Foundation (DFG) will fund the new subproject B06 in the research area "B - Social Practice" starting in July. The aim of the project is to create an ethical framework for explaining artificial intelligence. Junior professor Dr. Suzana Alpsancar and professor Dr. Tobias Matzner (already leader of subproject B03) is leading the subproject.

Read more

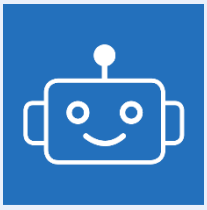
News

Guest scientist Marco at TRR 318

Marco Matarese is a PhD student at the University of Genoa and currently supports the TRR subproject A01 "Adaptive Explaining" as a visiting researcher. In this interview, he talks about his time at the Transregio, what he is researching with his colleagues, and what he sees as the potential and the dangers of AI.
Read more

Hand gestures, nods and co.

When explaining the rules of a board game, we often use multimodal signals; these are nonverbal forms of communication such as facial expressions, head nods, or pointing hand gestures. The researchers of project A02 have investigated these signals in board game explanations and record the results in a data corpus.
Read more



The hesitant robot

People often get bogged down when explaining because they have to think about their next sentences. Should robots imitate these natural delays when explaining? In a new study by the Universities of Bielefeld and Bremen, scientists are examining how people can better follow the statements of a robot. **Read more**

New media page

The Public Relations project documents Transregio's research on a number of digital channels. A new media page on the TRR website now presents an overview with all available platforms. Thus, past newsletter issues can also be found there for reference.

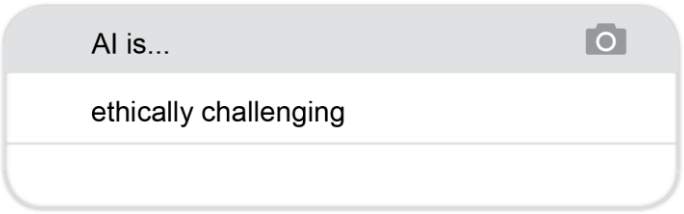
Discover more

AI viewed from an artistic perspective

Starting June 17, the Marta Herford Art Museum will host the exhibition "SHIFT. KI and a future community". Visitors will also have the opportunity to participate in workshops and lectures there.

Learn more

More News



"Currently, AI systems are being introduced top-down in many areas at a breathtaking pace. This means that a central authority decides and all subordinate areas follow. However, AI systems often have a considerable political and social impact. We can't even properly assess the implications yet, and they are hardly considered in the current hype. One of the biggest ethical challenges at the moment is the rapid proliferation in the absence (yet) of regulation and evaluation of these systems. Our goal is therefore to develop AI systems in an interdisciplinary and co-constructive way. This means that we want to work closely with future users, adapt the systems to them and evaluate them together. In this way, we are building the potential for a democratically legitimized social use of AI systems. On the other hand, it allows us to contribute expertise, practices and socially relevant requirements to the technical design of the systems. Ultimately, we can also use this to better tap the technical development potential of AI systems."

Jun.-Prof. Dr. Ilona Horwath
*Project Manager of the subprojects **B03** and **Ö***

Published

Research Paper: Speech Disentanglement for Analysis and Modification of Acoustic and Perceptual Speaker Characteristics
Popular speech disentanglement systems decompose a speech signal into two different embeddings: A content and a speaker embedding. To understand, which information is encoded in the speaker embeddings, in this work, such a system is investigated. The results could enable manipulating these embeddings in a wanted direction with modified acoustic properties. **Read more**

Research Paper: Halting the Decay of Talk
This paper investigates how people with atypical bodily capabilities interact within virtual reality (VR) and the way they overcome interactional challenges in these emerging social environments. The findings are subsequently related to renewed discussions of the relationship between agency and environment, and the co-constructed nature of situated action. **Read more**

What have I learned?

"Documenting results and data sources is part of data management and a quality measure for scientific work. I realized this during the 'Data Sharing & Open Science' workshop offered for TRR 318. It is essential that we researchers can still reconstruct and validate our results several years later. And if we share our research data and make it openly available, other scientists can also build on it and improve their own research projects. In order to protect one's own publications appropriately, licenses were a topic of the workshop, which unfortunately get far too little attention in the normal everyday life of science. We all know that these methods are important, but often do not apply them consistently in practice. The workshop made me realize once again how important it is to keep paying attention to this."

André Groß
*PhD student in **subproject A05***

TRR digital

Or message us directly for questions or feedback.

Subscribe to the newsletter for free.

Gefördert durch DFG Deutsche Forschungsgemeinschaft

TRR 318 „Constructing Explainability“

Teilprojekt Ö „Fragen zu erklärbaren Technologien“
Universität Bielefeld
Universitätsstraße 25
33615 Bielefeld

communication@trr318.uni-paderborn.de