



CONSTRUCTING EXPLAINABILITY

Erklärungen gemeinsam entwickeln | 03.2023

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 ▼](#)

2. TRR Konferenz „Measuring Understanding“

Am 6. und 7. November findet die diesjährige 2. TRR-Konferenz „Measuring Understanding“ statt. An der Universität Paderborn werden an diesen beiden Tagen internationale Wissenschaftler*innen über die aktuelle XAI-Forschung zum Thema Verstehen diskutieren. [Weiterlesen](#)



Science Slam als Rahmenprogramm

Teil des Konferenz-Programms vom 6. November ist ein Science Slam ab 18 Uhr. Hierbei werden Forschende aus verschiedenen Disziplinen ihre Eindrücke aus Forschung und wissenschaftlichen Alltag humorvoll und lyrisch präsentieren.

[Weitere Infos & Anmeldung](#)



Best Paper Award

Das Team des Teilprojekts C03 hat auf der World Conference on eXplainable Artificial Intelligence in Lissabon den „Best Paper Award“ gewonnen. Die Auszeichnung bekam der Beitrag „iPDP: On Partial Dependence Plots in Dynamic Modeling Scenarios“.

[Weiterlesen](#)



Podcast

Unser neuer englischer Forschungspodcast „Explaining Explainability“ ist veröffentlicht. In der ersten Episode spricht Moderatorin Prof. Britta Wrede mit den TRR-Sprecher*innen Prof. Philipp Cimiano und Prof. Katharina Rohlfing über das Thema 'Explainability'.

[Mehr entdecken](#)



YouTube

Der TRR 318 hat nun einen eigenen YouTube-Kanal. Das Video über unsere Co-Construction Workshops ist hier auch abrufbar, weitere Videos folgen in Kürze. [Mehr erfahren](#)



Menschliche Stimme

Wie kann man die menschliche Stimme zerlegen, verstehen und manipulieren? Die Forschenden des Teilprojekts C06 erstellten dafür ein Modell aus neuronalen Netzen, das die verschiedenen klanglichen Aspekte der Stimme voneinander trennt. [Weiterlesen](#)

Weitere News

KI ist...



oft unverständlich

„Wir alle haben täglich mit „Künstlicher Intelligenz“ zu tun. Sie empfiehlt mir zum Beispiel „Bücher, die mir gefallen könnten“, entscheidet, dass mein Smartphone erst einmal nur zu 80 Prozent aufgeladen wird, zeigt mir Werbung für Dinge, über die ich mich gestern im Café unterhalten habe, und beantwortet mir auf Zuruf Fragen. Manche Handlungen der KI sind nachvollziehbar (ich habe bereits ein eBook dieser Autorin gekauft), andere nicht (hört mein Smartphone meine Gespräche im Café etwa mit?). Die Verständlichkeit von KI ist dabei nicht immer wichtig (Hauptsache, das Smartphone ist morgen früh voll aufgeladen), für eine verantwortungsvolle Nutzung oft aber schon (kann ich den Antworten eines Large Language Models vertrauen?). Erklärbare KI hat das Ziel, Künstliche Intelligenz für Nutzer*innen verständlich zu machen und durch Erklärungen den Verstehensprozess zu ermöglichen. Im Idealfall wird dabei sowohl Wissen darüber erworben, wie die KI funktioniert (comprehension) als auch Wissen darüber, wie die KI genutzt werden kann (enabledness). Beide Formen des Verstehens greifen ineinander, bedingen sich gegenseitig und ermöglichen erst dadurch eine bewusste und verantwortliche Nutzung von KI und führt zu Handlungsfähigkeit (agency).“



Jun.-Prof. Dr. Hendrik Buschmeier
Projektleiter im **Teilprojekt A02**

Veröffentlicht

Forschungsartikel: Adding Why to What? Analysis of an Everyday Explanation

Diese Publikation beleuchtet die Herausforderungen bei der Erklärung von Technologien für Laien und analysierte dafür 20 Spieleerklärungen auf die Art der Priorisierung in der Erklärung. Die Ergebnisse zeigen, dass Erklärer*innen zunächst die Logik des Spiels (Architektur) betonen, bevor sie sich auf Empfehlungen oder die Bedeutung von Entscheidungen (Relevanz) konzentrieren. Die Erkenntnisse inspirieren zukünftige Forschungsfragen zur Entwicklung synthetischer Erklärungen. [Weiterlesen](#)

Forschungsartikel: The Role of Response Time for Algorithm Aversion in Fast and Slow Thinking Tasks

In dieser Publikation untersuchen die Forschenden den Einfluss der KI-Reaktionszeit auf die Algorithmus-Aversion in verschiedenen Denkaufgaben. Es wird deutlich, dass längere Reaktionszeiten zu einer geringeren Abneigung führen, unabhängig vom Denkt tempo. Zu dem gibt es signifikante Unterschiede in der Abneigung bei schnellen Denkaufgaben je nach Aufgabenbereich. [Weiterlesen](#)

TRR 318 bei der „4th Summer School on Social Human-Robot Interaction“

Vom 18. bis 23. September waren Josephine Beryl Fisher (A01), Vivien Lohmer (A04) und Amit Singh (A05) bei der 4th Summer School on Social Human-Robot Interaction am European Centre of Geological Education in Chęciny, Polen.

Forscher*innen der ganzen Welt und verschiedenen Disziplinen kamen zusammen und verbrachten fünf Tage mit interessanten Vorträgen (u.a. von Kerstin Fischer, Judith Holler und Roger K. Moore), spannenden Workshops (Programming NAO und Programming Furhat) und Methodenworkshops (z.B. Recurrence Quantification Analysis und Conversational Analysis in HRI). Darüber hinaus hatten alle Doktorand*innen die Möglichkeit, ihre Forschung in einer eigenen Sitzung als Poster zu präsentieren.

Josephine Beryl Fisher (A01) präsentierte ihre Arbeit zu einschlägigen Beiträgen (substantive contribution) durch den Explainee. Die Erfassung von „einschlägigen Beiträgen“ stand dabei im Vordergrund.

Vivien Lohmer (A04) präsentierte ihre aktuelle Arbeit zu interaktiven Strukturen von Erklärung und darüber, wie die Teilnehmer*innen das Ende einer Erklärung wechselseitig anzeigen.

Amit Singh (A05) präsentierte seine aktuelle Forschung zu 'interaktiven Kontrasten', in denen er den Effekt von sprachlichen Reparaturen - z.B. Negationen - auf die Online-Aufmerksamkeit des Erklärenden in einem interaktiven HRI-Kontext untersucht.

Forschungsposter ansehen:

- [Josephine Beryl Fisher](#)
- [Vivien Lohmer](#)
- [Amit Singh](#)

YouTube Video „Die WAHRE Gefahr hinter künstlicher Intelligenz“

Im Video des Youtube Kanals „SPACE FROGS“ erklären die YouTuber, was die Gefahr hinter der künstlichen Intelligenz ist. Denn das Problem ist nicht das KI-System selbst...

[Video sehen](#)

Artikel „Wenn Künstliche Intelligenz in den Dialog geht“

KI unterstützt die Entscheidungsfindung in vielen Bereichen, beispielsweise in Banken, Polizei und Medizin. Doch wie kommt die KI zu Ihren Ergebnissen und Antworten? Mit diesen Fragen beschäftigen sich die Forscher*innen des TRR. [Artikel lesen](#)

Podcast von ARD „Der KI-Podcast“

Von den Auswirkungen auf die Arbeitswelt bis hin zur Nutzung im Alltag - Der Podcast beschäftigt sich mit den Fragen der KI-Revolution. [Podcast hören](#)

Neues KI-Magazin: Human Magazin

Das Magazin „human“ richtet sich an alle Menschen mit Interesse am Thema KI und bietet Orientierung und Wissen. Der Fokus liegt auf den Auswirkungen von KI auf verschiedene Bereiche – Wirtschaft, Politik, Gesellschaft und Kultur. [Artikel lesen](#)

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?

„Das Ziel vieler wissenschaftlicher Projekte ist in vielen Fachbereichen die Publikation und Präsentation auf Fachkonferenzen. Leider existieren Konferenzen nicht auf magische Weise, sondern sie müssen organisiert werden. Da stehen Menschen dahinter, die wie die Teilnehmer*innen auch „nur“ Wissenschaftler*innen sind und meist keine Erfahrungen mit Großprojekten haben. Da kann man sich als erstmaliger Organisator schnell ins kalte Wasser geworfen fühlen. Man lernt aber schnell zu schwimmen. Das Wichtigste dabei sind Co-Organisatoren*innen und eine gute Arbeitsteilung. Denn Aufgaben gibt es überraschend viele. So wollen Aufrufe für Beiträge geschrieben, Websites gebaut, Keynote Speaker akquiriert, und Präsentationen vorbereitet werden. All das muss zu einem thematischen Konzept passen, was natürlich auch entwickelt werden muss. Der wichtigste und zeitaufwändigste Teil der Organisation ist eine gute Öffentlichkeitsarbeit. Denn die letzten Marketing-Meter funktionieren meist doch über Mund-zu-Mund-Werbung. Für das Überspringen all dieser zeitraubenden und anstrengenden Hürden wird man allerdings auch belohnt. Man erhascht einem wertvollen Einblick hinter die Kulissen, führt Gespräche mit Experten:innen auf Augenhöhe, und erweitert sein persönliches Netzwerk. All das ist in der Wissenschaft von immenser Bedeutung.“



Maximilian Muschalik
Doktorand im [Teilprojekt C03](#)

TRR digital



Oder [direkt per Mail](#) mit Fragen oder Feedback an uns.

Newsletter [abonnieren](#).

[Top: german version](#) ▲

[Footer: Impressum](#) ▼

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!



2nd TRR Conference "Measuring Understanding"

This year's 2nd TRR Conference "Measuring Understanding" will take place on November 6 and 7. On these two days at Paderborn University, international scientists will discuss current XAI research on the topic of understanding. [Read more](#)



Science Slam as additional programme

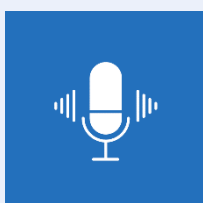
Part of the conference program on November 6 is the Science Slam. Researchers of psychology and linguistics will present their impressions from research and their scientific everyday life in a lyrical and humorous way. [Register here](#)

News



Best Paper Award

The team of subproject C03 has won the "Best Paper Award" at the World Conference on eXplainable Artificial Intelligence in Lisbon. The award was given to the paper "iPDP: On Partial Dependence Plots in Dynamic Modeling Scenarios". [Read more](#)



Podcast

The new research podcast "Explaining Explainability" is out. In the first episode, host Prof. Britta Wrede talks with TRR speakers Prof. Philipp Cimiano and Prof. Katharina Rohlfing about the topic of explainability. [Discover more](#)



YouTube

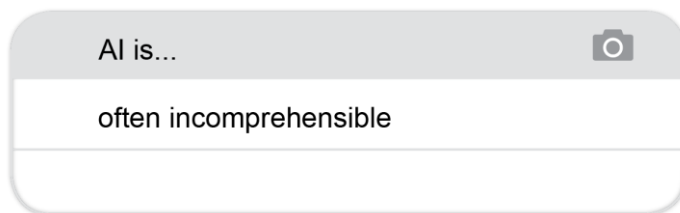
The TRR 318 now has its own YouTube channel. The first video about our co-construction workshops is already uploaded, more videos will follow soon. [Learn more](#)



Human voice

How can the human voice be dissected, understood and manipulated? The researchers of subproject C06 created a neural network model that separates the different sonic aspects of the voice. [Read more](#)

[More News](#)



"We all interact with "artificial intelligence" on a regular basis. For example, it recommends "books I might like", decides that my smartphone should stop charging at 80 per cent and continue later, shows me advertisements for things I talked about in the café yesterday, and answers my questions on demand. Some of the actions of AI-driven technology are understandable (I've already bought an ebook by this author), some are not (is my smartphone listening to my conversations in the café?). Understanding AI is not always important (I don't care as long as my smartphone is fully charged tomorrow morning), but it is often important for appropriate and responsible use (can I trust the answers of a large language model?). Explainable AI aims to make artificial intelligence understandable to users, and to facilitate the process of understanding by providing explanations. Ideally, both knowledge of how the AI works (comprehension) and knowledge of how the AI can be used (enabledness) are gained. These two forms of understanding are intertwined and interdependent, and are the only way to enable a



conscientious and responsible use of AI in one's own actions, and to give us agency."

Jun.-Prof. Dr. Hendrik Buschmeier
Project Leader of **subproject A02**

Published

Research Paper: Adding Why to What? Analysis of an Everyday Explanation

This publication highlights the challenges of explaining technology to laypeople and analysed 20 game explanations for the type of prioritisation in the explanation. The results show that explainers first emphasize the logic of the game (architecture) before focusing on recommendations or the importance of decisions (relevance). The findings inspire future research questions to develop synthetic explanations. [Read more](#)

Research Paper: The Role of Response Time for Algorithm Aversion in Fast and Slow Thinking Tasks

In this publication, the researchers investigate the impact of AI response time on algorithm aversion in various thinking tasks. It becomes clear that longer reaction times lead to less dislike, independent of the speed of thinking. There are significant differences in the dislike in fast thinking tasks depending on the task. [Read more](#)

TRR on a research trip

TRR 318 at the „4th Summer School on Social Human-Robot Interaction“

From September 18 – 23 Josephine Beryl Fisher (A01), Vivien Lohmer (A04) and Amit Singh (A05) were at the 4th edition of the Summer School in Social Human-Robot-Interaction at the European Centre for Geological Education in Chęciny, Poland.

Researchers from all over the world and from various disciplines came together and spend five days with interesting talks (e.g. Kerstin Fischer, Judith Holler and Roger K. Moore among others), exciting workshops (programming NAO and Furhat) and learning interesting methods to analyse interaction (e.g., Recurrence Quantification Analysis and conversational analysis in HRI). In addition to this, all PhDs had the chance to present their research as a poster in a dedicated session.

Josephine Beryl Fisher (A01) presented her work on substantive contributions through the Explainee. The focus was on the recording of "substantive contributions".

Vivien Lohmer (A04) presented her ongoing work concerning the interactive structure of an explanation and how participants are making the end of an explanation accountable.

Amit Singh (A05) presented his ongoing work on 'interactive contrasts' investigating the effect of linguistic repairs – e.g., negations - on explainee's online attention in an interactive HRI context.

View posters:

- [Josephine Beryl Fisher](#)
- [Vivien Lohmer](#)
- [Amit Singh](#)

What have I learned?

"The goal of scientific projects in many disciplines is publishing and presenting at scientific conferences. Unfortunately, these conferences do not magically exist; they have to be organized. Organized by people who, like the participants, are "only" scientists and usually have no experience of large-scale projects. As a first-time organizer, you can easily feel thrown into the deep end. But you learn to swim very quickly. The most important things are co-organizers and a good division of tasks, and there are surprisingly many of these. Calls for papers have to be written, websites set up, keynote speakers recruited and presentations prepared. All of this has to fit into a thematic concept, which of course also has to be developed in the process. The most important and time-consuming part of the organization is good outreach effort. After all, word of mouth is usually the ultimate marketing tool in the final stages. However, there are rewards for jumping over all these time-consuming and exhausting hurdles. You get a behind-the-scenes look, you talk to experts on an equal footing and you expand your personal network. All of this is invaluable in science."



Maximilian Muschalik
PhD student in [subproject C03](#)

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