

Erklärungen gemeinsam entwickeln | 03.2026

Im Newsletter des Transregio (TRR) 318 „Constructing Explainability“ präsentieren wir Forschungsprojekte, Workshops zu Künstlicher Intelligenz sowie aktuelle Nachrichten und Vorträge. Sie sind herzlich eingeladen, mit uns auf [LinkedIn](#) und [Instagram](#) 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 ▼](#)



„Sieker spricht und wir hören zu“

Schon am Dienstag, 13. Oktober, von 14 bis 18:30 Uhr findet in der Stadtteilküche Sieker ein offener Mitmachnachmittag rund um Künstliche Intelligenz statt. Der TRR 318 und der SFB 1646 „Sprachliche Kreativität in der Kommunikation“ der Universität Bielefeld laden dazu ein, verschiedene KI-Anwendungen selbst auszuprobieren und mit Forschenden ins Gespräch zu kommen. An verschiedenen Stationen können Besucher*innen unter anderem Roboter, KI-Games und Chatbots ausprobieren. Der Mitmachnachmittag ist Teil des gemeinsamen Programms „Sieker spricht und wir hören zu“.

[Weiterlesen: Sieker Programm](#)

[Weiterlesen: Mitmachnachmittag](#)



Wie eine KI die Welt sieht

Wie trifft Künstliche Intelligenz eigentlich Entscheidungen und warum können dabei bestehende Ungleichheiten fortgeschrieben werden? In einem neuen Erklärvideo zeigt Techniksoziologe Dr. Nils Klowait, wie KI Muster in Daten erkennt und auf neue Situationen überträgt. Außerdem erklärt er, warum nachvollziehbare und anpassungsfähige KI-Systeme wichtig sind und wie der TRR 318 dazu forscht.

[Video anschauen](#)



Soziologin Elena Esposito erhält renommierten Balzan-Preis

Die hochdotierte Ehrung gehört weltweit zu den Spitzenauszeichnungen der Wissenschaft. Mit dem diesjährigen Balzan-Preis wird Prof. Dr. Elena Esposito für ihre wissenschaftliche Arbeit zur Soziologie von KI und Algorithmen gewürdigt. Im TRR 318 leitet sie das Projekt (B01): Modelle des maschinellen Lernens mit einem dialogbasierten Ansatz erklären.

[Weiterlesen](#)



Erfolgreich promoviert im TRR 318

Weitere Promovierende des TRR 318 haben ihre Dissertationen erfolgreich abgeschlossen. In ihren Arbeiten beschäftigten sie sich unter anderem mit interaktiven und adaptiven KI-Erklärungen, Mensch-Roboter-Interaktion, Negationen und Biases in Sprachmodellen. Die Ergebnisse zeigen, wie vielfältig die Forschung zu verständlicher, adaptiver und sozialer KI im TRR 318 ist.

[Weiterlesen](#)



Save the Dates

Der TRR 318 ist am **28. Oktober bei der GENIALE** in der Wissenswerkstadt Bielefeld vertreten und gibt dort kurze Einblicke in die Forschung zu erklärbarer KI in der Medizin.

[Mehr erfahren](#)

Die **4. TRR 318-Konferenz „Scaffolding Social XAI“** findet am 16. und 17. März 2027 in Paderborn statt. Forschende verschiedener Disziplinen sind eingeladen, Beiträge zu sozialer und erklärbarer KI einzureichen. Die Einreichungsfrist ist der 1. November. Für die Special Session on XAI Literacy, die sich mit einem kritischen und kompetenten Umgang mit KI-Erklärungen beschäftigt, können noch bis zum 14. November Beiträge eingereicht werden.

[Mehr erfahren](#)

Im Kontext

„Wer entscheidet eigentlich, wie ein KI-System eingesetzt wird? Und was können Menschen tun, wenn sie mit seinen Entscheidungen oder Erklärungen nicht einverstanden sind? Mit solchen Fragen beschäftigen sich die Co-Construction Workshops (CCWS), die von uns im Projekt WIKO des TRR 318 entwickelt werden.



Die Workshops bieten einen spielerischen und niedrigschwelligen Zugang zu kritischer KI- und Datenkompetenz. In Rollenspielen betrachten die Teilnehmenden KI-Systeme aus unterschiedlichen Perspektiven und untersuchen, wie sich Erklärbedarfe und Handlungsmöglichkeiten je nach beteiligter Person oder Interessengruppe unterscheiden.

Dabei reflektieren sie auch soziale und politische Dimensionen von KI, etwa Machtverhältnisse und Ungleichheiten. Sie setzen sich mit ihren individuellen und gemeinsamen Möglichkeiten auseinander, KI-Systeme mitzugestalten, zu hinterfragen, zu verändern oder auch abzulehnen.”

*Ula Przybylska für **Projekt WIKO**: Fragen zu erklärbaren Technologien*

Veröffentlicht

Studie: Partner models of explainers

Wie gut eine Erklärung funktioniert, hängt stark davon ab, wen wir vor uns haben. In einer neuen Studie haben Viviane Gladow und Prof. Dr. Heike Buhl aus dem Projekt [A01](#) „Adaptives Erklären“ untersucht, wie Lehrkräfte, Hochschullehrende und Ärzt*innen ihr Gegenüber beim Erklären einschätzen. Im Mittelpunkt stehen sogenannte Partnermodelle, also Vorstellungen von Wissen, Verständnis, Interesse und weiteren Voraussetzungen einer Person. Die Ergebnisse zeigen, warum Erklärungen flexibel angepasst werden müssen und was das für künftig erklärende KI-Systeme bedeutet.

[Weiterlesen](#)

Forschungsartikel: Making sense of AI: how lay users negotiate the shifting human-machine boundaries in talking about AI

Die Studie untersucht, wie Laien KI in alltäglichen Situationen einordnen, beschreiben und bewerten. Im Mittelpunkt steht die Frage, wie Menschen mithilfe vertrauter Kategorien

Bedeutung herstellen, auch ohne technisch präzise Definitionen. Die Autor*innen zeigen damit, dass Erklärbarkeit nicht nur eine Eigenschaft von KI-Systemen ist, sondern auch durch menschliche Interpretation und situatives Verstehen entsteht.

[Weiterlesen](#)

Methodenentwicklung: A Methodological Framework for Studying Online Language–Vision Interplay in Human–Robot Interaction

Wie lassen sich Sprache und Blickverhalten im Kontext bereits durchgeführter Handlungen während der Mensch-Roboter-Interaktion untersuchen? Amit Singh und Prof. Dr. Katharina Rohlfing aus dem Projekt A05 des TRR 318 haben dafür eine neue experimentelle Methode entwickelt. Sie verbindet kontrollierte Untersuchungsbedingungen mit einer fortlaufenden Interaktion zwischen Mensch und Roboter. So lässt sich in Echtzeit untersuchen, wie Menschen Sprache verstehen und eigene Äußerungen im Hinblick auf die vorangegangenen Handlungen abstimmen.

[Weiterlesen](#)

Medientipps

Diskussionsbeitrag: Gesundes Misstrauen gegenüber agentischen KI-Systemen in der Hochschullehre

Ein neuer Beitrag im Hochschulforum Digitalisierung beschäftigt sich mit dem Einsatz agentischer KI-Systeme in der Hochschullehre. Im Mittelpunkt steht die Frage, wie solche Systeme verantwortungsvoll genutzt werden können und warum ein kritischer Umgang mit ihren Empfehlungen wichtig ist. Die Autor*innen aus dem TRR 318 plädieren dafür, „gesundes Misstrauen“ als Teil von KI-Kompetenz und Hochschulbildung zu verstehen.

[Weiterlesen](#)

Magazinartikel: Perspektiven auf KI in der Psychologie

Zwei Ausgaben der Psychologischen Rundschau widmen sich aktuellen Perspektiven auf Künstliche Intelligenz aus verschiedenen Fachgruppen der Deutschen Gesellschaft für Psychologie. Mit dabei sind auch Beiträge aus der Pädagogischen Psychologie und der Allgemeinen Psychologie, an denen TRR-Mitglieder Prof. Dr. Heike Buhl und Prof. Dr. Ingrid Scharlau beteiligt sind.

[Generative Künstliche Intelligenz im Kontext von Lehren, Lernen und Beraten](#)
[Vorhersage ist nicht Erklärung](#)

Was habe ich gelernt?

„Die wissen aber schon ganz schön viel über KI‘, dachte ich, als ich mich im Jugendhaus Sieker mit einer kleinen Gruppe 12-jähriger Mädchen über ChatGPT und andere KI-Anwendungen ausgetauscht habe. Wir haben über den hohen Wasserverbrauch und die Fehler von KI gesprochen und darüber, wie die Mädchen selbst mit ChatGPT kommunizieren.



Dabei ist schnell deutlich geworden: KI wird ganz selbstverständlich genutzt, aber ziemlich pragmatisch. ‚Ich nutze ChatGPT, wenn ich etwas wissen oder nachgucken will‘, war eine der Antworten. Auch ‚bitte‘ und ‚danke‘ seien nicht nötig: ‚ChatGPT ist doch nicht mein Freund.‘ Mit der Vorstellung, KI stärker zu vermenschlichen oder eine emotionale Beziehung zu ihr aufzubauen, konnten die Mädchen eher wenig anfangen.

Anschließend haben wir einige Gedanken, Wörter und Motive aus unserem Austausch in Graffitis verwandelt. Besonders beeindruckt hat mich, mit wie viel Motivation und Spaß sich die Mädchen auf diese ganz analoge Weise mit KI auseinandergesetzt haben und wie reflektiert sie bereits über ihren eigenen Umgang mit der Technologie gesprochen haben.

Und gelernt habe ich auch: Spraken sieht leichter aus, als es ist.“

Linda Thomßen
Referentin für Wissenschaftskommunikation
am **TRR 318**

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 as well as new publications and upcoming talks. You are invited to interact with us on [LinkedIn](#) and [Instagram](#) and email us with your questions about AI. Let's develop explanations together!



"Sieker Speaks and We Listen"

Coming up on Tuesday, October 13, from 2:00 to 6:30 p.m., an open, hands-on event exploring artificial intelligence will take place at Stadtteilküche Sieker. The TRR 318 and the Collaborative Research Centre 1646 "Linguistic Creativity in Communication" at Bielefeld University invite visitors to try out various AI applications and talk with researchers. At different interactive stations, visitors will have the opportunity to experiment with robots, AI games, chatbots, and more. The event is part of the joint programme "Sieker Speaks and We Listen."

[More: Sieker Programm](#)

[More: Mitmachnachmittag](#)

News



Explanatory Video: How AI Sees the World

How does artificial intelligence actually make decisions, and why can existing inequalities be perpetuated in the process? In a new explainer video, technology sociologist Dr. Nils Klowait shows how AI recognizes patterns in data and applies them to new situations. He also explains why transparent and adaptable AI systems are important and how TRR 318 is conducting research in this area.

[Watch the video](#)



Sociologist Elena Esposito Receives Prestigious Balzan Price

This highly endowed honor is one of the world's most prestigious awards in science. This year's Balzan Prize recognizes Prof. Dr. Elena Esposito for her academic work on the sociology of AI and algorithms. In TRR 318, she leads project B01: "Models of Machine Learning with a Dialogical Approach to Explanation."

[Read more \(in German\)](#)



Successful Doctorates at TRR 318

More doctoral researchers at TRR 318 have successfully completed their dissertations. Their research covered topics including interactive and adaptive AI explanations, human–robot interaction, negotiation, and biases in language models. The results highlight the diversity of research on understandable, adaptive, and social AI within TRR 318.

[Read more](#)



Safe the Dates

TRR 318 will be represented at **GENIALE at Wissenswerkstadt Bielefeld on October 28**, offering brief insights into research on explainable AI in medicine (event in German). [Read more](#)

The **4th TRR 318 Conference "Scaffolding Social XAI"** will take place on 16–17 March 2027 in Paderborn. Researchers from various disciplines are invited to submit contributions on social and explainable AI. The submission deadline is 1 November. Contributions to the Special Session on XAI Literacy, focusing on the critical understanding and use of AI explanations, can be submitted until 14 November.

[Read more](#)

[More News](#)

In Context

"Who decides how an AI system is used? And what can people do if they disagree with its decisions or explanations? These are some of the questions addressed in the Co-Construction Workshops (CCWS), which are being developed by us as part of the WIKO project within TRR 318.



The workshops offer a playful and accessible approach to critical AI and data literacy. Through role-play,

participants explore AI systems from different perspectives and examine how needs for explanations and possibilities for action vary across stakeholders.

They also reflect on the social and political dimensions of AI, including power relations and inequalities, and consider their individual and collective agency in shaping, questioning, modifying, or rejecting AI systems.”

*Ula Przybylska for **Project WIKO**: Questions about Explainable Technologies*

Published

Interview Study: Partner models of explainers

How well an explanation works depends greatly on who we are talking to. In a new study, Viviane Gladow and Prof. Dr. Heike Buhl from project A01 “Adaptive Explanation Generation” investigated how school teachers, university lecturers, and physicians assess the person they are explaining something to. The focus is on so-called partner models – ideas about a person’s knowledge, understanding, interests, and other prerequisites. The results show why explanations need to be adapted flexibly and what this means for future explainable AI systems

[Read more](#)

Research Paper: Making sense of AI: how lay users negotiate the shifting human–machine boundaries in talking about AI

The study examines how laypeople categorize, describe, and evaluate AI in everyday situations. It focuses on how people use familiar categories to make sense of AI, even without technically precise definitions. The authors show that explainability is not only a property of AI systems but also emerges through human interpretation and situated understanding.

[Read more](#)

Research Paper: A Methodological Framework for Studying Online Language–Vision Interplay in Human–Robot Interaction

How can language and gaze be studied in the context of previous actions during human–robot interaction? Amit Singh and Prof. Dr. Katharina Rohlfing from TRR 318 project A05 have developed a new experimental method for this purpose. It combines controlled experimental conditions with continuous interaction between humans and robots. This allows researchers to investigate in real time how people understand language and adapt their own utterances based on previous actions.

[Read more](#)

What Have I Learned?

“They really know quite a lot about AI,’ I thought, talking to a group of twelve-year-old girls at the Sieker Youth Centre about ChatGPT and other AI applications. We discussed AI’s high water consumption, its errors and how the girls themselves interact with ChatGPT.

It quickly became clear that they use AI as a matter of course, but in a fairly pragmatic way. ‘I use ChatGPT when I want to look something up,’ was one response. They also said that there’s no need to say ‘please’ or ‘thank you’ because ‘ChatGPT isn’t my friend, after all’. The girls weren’t interested in humanizing AI or building an emotional connection with it.

Afterwards, we turned some of the thoughts, words and motifs from our discussion into graffiti. I was particularly impressed by how motivated and enthusiastic the girls were about engaging with AI in this analogue way, and by how thoughtful they were about their own use of the technology.

I also learned: Spray-painting is harder than it looks.”

Linda Thomßen

Science Communication Manager at [TRR 318](#)



TRR digital



Or [message us directly](#) for questions or feedback.

[Subscribe](#) to the newsletter for free.

TRR 318 „Constructing Explainability“

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

communication@trr318.uni-paderborn.de