

Erklärungen gemeinsam entwickeln | 01.2026

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 ▼](#)



TRR 318 eröffnet zweite Förderphase

Mit einem Kick-off-Event am 3. März ist der TRR 318 offiziell in seine zweite Förderphase gestartet. Mitglieder und Assoziierte kamen zusammen, um sich kennenzulernen, auszutauschen und Ziele zu schärfen. In der zweiten Förderphase rückt insbesondere der Kontext von Erklärungen in den Fokus: Wie können KI-Systeme angemessen auf unterschiedliche Situationen, Wissensstände

und Dialogverläufe reagieren? Zwei neue Projekte legen zudem den Fokus auf große Sprachmodelle (Large Language Models, LLMs).

News



Neues Projekt B07

Im Projekt B07 „Kommunikative Praktiken des Erfragens von Informationen und Erklärungen von LLM-basierten Agenten“ untersuchen die Forschenden, wie Nutzer*innen mithilfe von Prompts Informationen und Erklärungen von KI-Systemen auf Basis großer Sprachmodelle (LLMs) wie ChatGPT erfragen. Anhand von Gesprächsanalysen untersucht das Team, wie sich Verständnis im Austausch mit KI schrittweise entwickelt und wie Nutzer*innen lernen, die Antworten der Systeme besser einzuordnen. Ziel ist es, einen kontextsensitiven Rahmen für die Mensch-Maschine-Interaktion zu entwickeln.



Neues Projekt C07

Im Projekt C07 „Ko-Konstruktions-optimierte große Sprachmodelle für Erklärungen“ wird untersucht, wie große Sprachmodelle so trainiert werden können, dass sie Erklärungen gemeinsam mit Nutzer*innen entwickeln. KI-Systeme sollen im Dialog erkennen, was bereits verstanden wurde, wo noch Unsicherheiten bestehen und wie Erklärungen schrittweise weiter aufgebaut werden können. Dazu werden Methoden entwickelt, mit denen Sprachmodelle den Gesprächsverlauf erfassen und relevantes Fachwissen sowie Kontextinformationen gezielt einbinden. Ziel ist es, kontextsensitive und ko-konstruktive Erklärungen zu ermöglichen.



Vortrag „Eingeloggt! Bielefeld“-Woche

Zum Auftakt der „Eingeloggt! Bielefeld“-Woche halten Prof. Dr.-Ing. Britta Wrede und Prof. Dr. Katharina Rohlfing einen Vortrag „Zum Erklären gehören mindestens Zwei – Warum die KI Sie zum Erklären braucht!“. Darin zeigen sie, wie verständliche KI nur im Zusammenspiel mit den Nutzer*innen entsteht und warum Nachfragen, Vorwissen und Dialog entscheidend für gelungene Erklärungen sind. Die Veranstaltung findet am Montag, 16. März 2026 von 17 bis 19 Uhr in der Wissenswerkstadt Bielefeld statt. Die Initiative „Eingeloggt!“, gefördert von der Körber-Stiftung, stärkt digitale Kompetenzen und richtet sich insbesondere an Menschen ab 50 Jahren.

[Weiterlesen](#)



Blick hinter die TRR-Kulissen: unsere Activity Afternoons

Wie schaffen es rund 70 TRR-Mitglieder und -Assoziierte, in Kontakt zu bleiben? Einmal im Monat lädt der TRR 318 zu den Activity Afternoons ein – sowohl digital als auch in Präsenz. Bei diesen Treffen haben die Teilnehmenden die Möglichkeit, aktuelle Forschungsthemen zu diskutieren. Gelegentlich werden auch Gäste eingeladen, wie beispielsweise der Linguist Prof. Dr. Michael Franke von der Universität Tübingen, der am 29. April 2026 einen Einblick in seine Forschung geben wird.

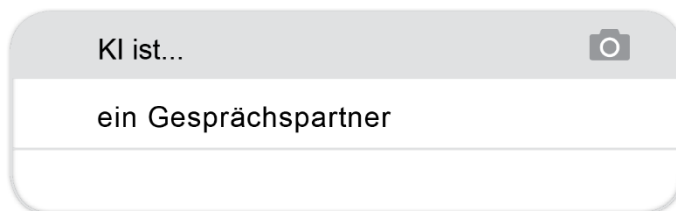


Disputationen

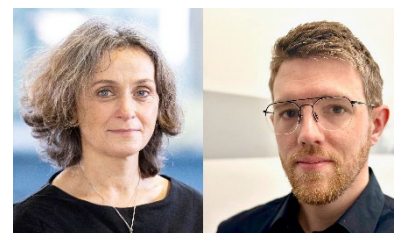
Weitere Doktorand*innen im TRR 318 haben ihre Promotion erfolgreich abgeschlossen. An dieser Stelle herzliche Glückwünsche an: André Groß, Tobias Peters, Amelie Robrecht, Dimitry Mindlin und Helen Beierling!

[Weiterlesen](#)

Weitere News



„KI auf der Grundlage von LLMs ist heutzutage so gut, dass man denken könnte, sie handelt wie ein echter Gesprächspartner. Aber der Schein trügt: LLMs können Gesprächspartnerschaft lediglich simulieren, ebenso wie sie Verstehen simulieren; daran kann auch gutes Prompting nichts ändern. So verfügen LLMs nicht über die Fähigkeit, materielle und situative Kontexte – wie beispielsweise soziale Rollen und damit verbundene Aufgaben – dynamisch im laufenden Gespräch zu aktualisieren und anzupassen sowie die Gesprächshistorie auch über einen längeren Zeitraum mit einzubeziehen. Insofern kann man nur in einem rudimentären Sinn davon reden, dass LLMs ko-konstruktiv bzw. 'Gesprächspartner' im eigentlichen Sinne sind. Gerade diese Einschränkungen im ko-konstruktiven Gesprächshandeln von LLMs machen es



notwendig, empirisch zu untersuchen, wie Menschen ihre Interaktionen mit solchen Systemen tatsächlich gestalten. Dies ist ein zentrales Forschungsvorhaben unseres neuen Projekts B07. Zudem betrachten wir, wie Menschen zu einer kritischeren Haltung gegenüber KI befähigt werden können.“

Prof. Dr. Friederike Kern & Dr. Björn Stövesand
Projektleiterin **A04**, **B07** & Assoziierter Postdoktorand
B07

Veröffentlicht

Forschungsartikel: Generation and evaluation of adaptive explanations based on dynamic partner-modeling an non-stationary decision making

Die Studie untersucht, wie Dialogsysteme Erklärungen dynamisch an ihre Nutzer*innen anpassen können. Dafür wird ein Partnermodell entwickelt, das mithilfe von Nutzerfeedback kontinuierlich aktualisiert und zur Auswahl der jeweils passenden Erklärstrategie genutzt wird. Die Ergebnisse einer Online-Studie zeigen, dass angepasste Erklärungen das Verständnis der Nutzer*innen deutlich verbessern.

[Weiterlesen](#)

Medientipps

TRR 318 YouTube Kanal

Auf dem YouTube-Kanal erscheinen regelmäßig spannende Shorts mit Fragen und Antworten rund um die TRR-Forschung sowie bald neue Videos aus der Reihe „KI ist...“. Außerdem folgt in Kürze ein neues Erklärvideo darüber, wie Künstliche Intelligenz die Welt sieht.

[Kanal abonnieren](#)

Podcast Folge: Prompten wir alle falsch?

In der neuen Folge von „Der KI-Podcast“ der ARD geht es um die Frage, ob wir KI eigentlich noch richtig prompten und welche neuen Strategien im Jahr 2026 zu besseren Ergebnissen führen. Dafür werden verschiedene Techniken getestet, über aktuelle Entwicklungen im Umgang mit KI diskutiert, sowie ein eigens entwickeltes Videospiel vorgestellt.

[Podcast anhören](#)

Was habe ich gelernt?

„Im TRR 318 werden die Promovierenden dazu ermutigt, sich nicht auf disziplinäre Grenzen zu beschränken und über das hinauszugehen, was sie im Masterstudium gelernt haben. Dazu gehört auch das Publizieren von Forschungsergebnissen. Mein Weg von einem groben Manuskript zu einer druckfertigen Version war voller Lernprozesse:



- a) Flexibel bleiben: Die Veröffentlichung beginnt erst nach Erhalt der Gutachten. Eingestellt auf Überarbeitungen zu sein und flexibel zu bleiben, wenn Änderungen erforderlich sind, macht den Umgang mit den Kommentaren entspannter.
- b) Offene Wissenschaft und öffentliche Dokumentation der Daten stärken Vertrauen: Wer preregistriert und den Analyseworkflow transparent hält, hat schon viel geschafft. Dies gibt auch den Herausgeber*innen Vertrauen in die eigene Arbeit.
- c) Ein Manuskript ist mehr als nur separate Abschnitte: Wir Menschen bevorzugen Geschichten gegenüber unzusammenhängenden Informationsbrocken. Dies wird noch deutlicher, wenn wir gemeinsam mit Kolleg*innen aus anderen Disziplinen schreiben. Die Verbindungen zwischen den Informationsbrocken zu finden, verwandelt reiche Ideen in eine zusammenhängende Geschichte.
- d) Gutachter*innen schätzen Bemühungen: Kleinere Vorschläge sind oft genauso wichtig wie die größeren im endgültigen Entwurf.“

Amit Singh

Doktorand im Projekt [A05](#)

TRR digital



Oder **direkt per Mail** mit Fragen oder Feedback an uns.

Newsletter **abonnieren**.

[Top: german version](#) ▲

[Footer: Impressum](#) ▼

Developing explanations together | 01.2026

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!



TRR 318 Launches Second Funding Phase

With a kick-off event on March 3rd, TRR 318 officially launched its second funding phase. Members and associates came together to get to know each other, exchange ideas, and refine their goals. In this second funding phase, the focus is particularly on the context of explanations: How can AI systems appropriately respond to different situations, levels of knowledge, and interactions? Additionally, two new projects focus on large language models (LLMs).

News



New Project B07

In project B07 “Communicative practices of requesting information and explanation from LLM-based agents,” researchers are investigating how users formulate prompts to request information and explanations from AI systems based on large language models (LLMs), such as ChatGPT. Using conversation analysis, the team is examining how understanding gradually develops in interactions with AI and how users learn to better categorize the systems' responses. The goal is to develop a context-sensitive framework for human-machine interaction.



New Project C07

Project C07 “Co-construction-following large language models for explaining” explores how to train large language models to generate explanations alongside users. The focus is on AI systems recognizing in dialogue what has already been understood, where uncertainties still exist, and how explanations can be built up step by step. The researchers are developing new methods to enable language models

to capture the course of a conversation and incorporate specific knowledge and contextual information. The goal is to facilitate context-sensitive, co-constructive explanations.



Public Talk “Eingeloggt! Bielefeld” week

At the opening of the “Eingeloggt! Bielefeld” week, Professor Britta Wrede and Professor Katharina Rohlfing will give a public talk titled “Explaining Takes at Least Two – Why AI Needs You to Explain!”. They will explore how understandable AI depends on interaction, showing why prior knowledge, questions, and dialogue are essential for successful explanations. The event will take place on Monday, March 16, 2026, from 17–19 pm at the Wissenswerkstadt Bielefeld. The “Eingeloggt!” initiative, supported by the Körber-Stiftung, promotes digital literacy, particularly for people aged 50 and above.

[Read more](#)



Behind the Scenes at TRR: Our Activity Afternoons

How do approximately 70 TRR members and associates manage to stay in touch? TRR 318 invites them to Activity Afternoons once a month, both digitally and in person. At these meetings, participants can discuss current research topics. Occasionally, guests are invited as well. For example, on April 29, 2026, linguist Professor Michael Franke from the University of Tübingen will provide insight into his research.

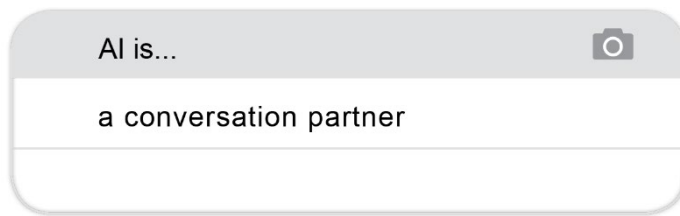


Disputations

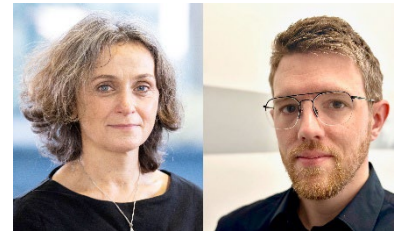
More doctoral students in TRR 318 have successfully defended their theses. Congratulations to André Groß, Tobias Peters, Amelie Robrecht, Dimitry Mindlin, and Helen Beierling!

[Read more](#)

[More News](#)



“AI based on LLMs has become so advanced that it can seem like a real conversation partner. However, appearances can be deceiving. LLMs can only simulate conversation and understanding. Even good prompting cannot change that. LLMs cannot dynamically update or adapt to material or situational contexts, such as social roles and associated tasks, during an ongoing conversation. Nor can they incorporate conversation history over a longer period of time. Therefore, it can only be said in a rudimentary sense that LLMs are co-constructive or “conversational partners” in the true sense of the word. These limitations in the co-constructive conversational behavior of LLMs make it necessary to empirically investigate how humans shape their interactions with such systems. This is a key research focus of our new B07 project. Additionally, we are exploring ways to empower individuals to adopt a more critical stance toward AI.”



*Professor Friederike Kern & Dr. Björn Stövesand
Project Leader [A04](#), [B07](#) & Associated Postdoctoral
Researcher [B07](#)*

Published

Research Paper: Generation and evaluation of adaptive explanations based on dynamic partner-modeling an non-stationary decision making

The study examines how dialogue systems can adapt their explanations dynamically to users. To this end, a partner model is being developed that uses user feedback to continuously update and select the most appropriate explanation strategy for each case. Results from an online study show that customized explanations significantly improve user comprehension.

[Read more](#)

Media tips

TRR 318 YouTube Channel

The YouTube channel regularly features short, exciting videos with questions and answers about TRR research. New videos from the "AI is..." series will be coming soon. Additionally, a

new video explaining how artificial intelligence sees the world will be released shortly.

[Subscribe the channel](#)

What have I learned?

“The doctoral candidates in TRR 318 are encouraged not to confine themselves within disciplinary boundaries and to go beyond what they learned in their master’s studies. This also includes publishing research results. My journey from a rough manuscript to a camera-ready version has been full of learning processes:

a) Staying flexible: Publishing only begins after receiving reviews. Being prepared for revisions and remaining flexible when changes are needed makes handling the comments much easier.

b) Open science and public documentation of data build trust: Those who preregister and keep their analysis workflow transparent have already accomplished a lot. This also gives editors confidence in your work.

c) A manuscript is more than just standalone sections: We humans tend to prefer stories over disjointed information chunks. This disjointedness becomes even more apparent when we write together with colleagues from other disciplines. Finding the connections between the information chunks transforms rich ideas into a coherent narrative.

d) Reviewers appreciate effort: Minor suggestions are often just as important as the major ones in the final draft.”



Amit Singh

PhD student [A05](#)

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