



CONSTRUCTING EXPLAINABILITY

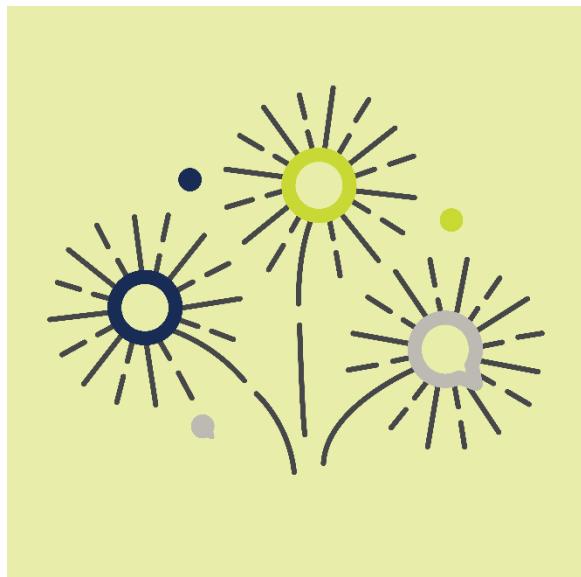
Erklärungen gemeinsam entwickeln | 04.2025

Im Newsletter des Transregio (TRR) 318 „Constructing Explainability“ präsentieren wir Forschungsprojekte, Workshops zu Künstlicher Intelligenz (KI) 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 ▼](#)

Der TRR 318 hat etwas zu feiern!

Die Deutsche Forschungsgemeinschaft (DFG) verlängert unseren Transregio 318 „Erklärbarkeit konstruieren“ um weitere dreieinhalb Jahre. Seit 2021 erforscht das TRR-Team, wie KI-Systeme verständlicher gestaltet werden können und sich an die Bedürfnisse der Nutzer*innen anpassen. In der kommenden Förderphase liegt der Fokus auf dem Kontext von Erklärungen: Wie verändern Situationen, Personen oder gemeinsames Wissen den Erklärbedarf und wie können KI-Systeme flexibel darauf reagieren? Mit über 60 Forschenden aus sieben Disziplinen setzt der TRR 318 seine erfolgreiche interdisziplinäre Arbeit fort.





Nachwuchswissenschaftler*innen aus verschiedenen Disziplinen kommen in der RTG zusammen

Das Graduiertenkolleg des TRR 318 strukturiert die Weiterbildung von Doktorand*innen und Postdoktorand*innen. Ricarda Kurock ist Mitarbeiterin der **Research Training Group (RTG)**. Im Interview spricht sie über Angebote, Ziele und Erfolge des Qualifizierungsprogramms. Ein Schlüsselement für sie ist die interdisziplinäre Zusammenarbeit.

[Weiterlesen](#)



Interview mit Promovierenden

Alle Nachwuchswissenschaftler*innen des TRR 318 sind automatisch Mitglieder des Graduiertenkollegs (RTG). Im Interview sprechen Psycholinguistin Dr. Josephine Fisher und Computerwissenschaftler Roel Visser über ihre Erfahrungen mit der Promotion in interdisziplinären Projekten des Transregio sowie über die Unterstützung und Betreuung der RTG.

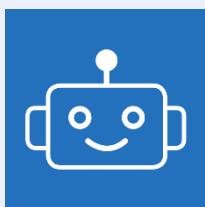
[Weiterlesen](#)



Science Café

Am 11.12.2025 findet in der Wissenswerkstadt Bielefeld das TRR 318-Gespräch „Erklärt und (nicht) verstanden? Wie Menschen und KI gemeinsam Bedeutung schaffen“ statt. Informatikerin Prof. Dr.-Ing. Britta Wrede und Linguistin Prof. Dr. Katharina Rohlfing geben Einblicke in ihre Forschung und laden zum offenen Gespräch über Erklärungen, KI und das Miteinander von Mensch und Maschine ein.

[Weiterlesen](#)



Danke für den Besuch!

Am 22.11.2025 fand der TRR 318-Tag auf der GENIALE in der Wissenswerkstadt Bielefeld statt. Wir bedanken uns herzlich bei allen Besucher*innen für ihr Interesse und die spannenden Gespräche!

[Weitere News](#)



„Dass man mit KI Stimmen generieren kann, also einen Text in gesprochene Sprache transformieren kann, ist allseits bekannt, und jede*r von uns hat solche Stimmen schon einmal gehört, beispielsweise bei Durchsagen an Bahnhöfen oder in Zügen. Auch als Lernhilfe, um die Aussprache beim Erlernen einer Fremdsprache zu verbessern, ist KI schon seit langem im Einsatz.“



Aber Stimmenerzeugung für die Sprachtherapie ist nochmal etwas ganz anderes: Wir entwickeln eine KI, die einzelne Stimmmerkmale, wie etwa die Rauigkeit oder Behauchtheit der Stimme, gezielt verändern kann. Dies hilft zum einen angehenden Sprach- und Stimmtherapeut*innen, menschliche Stimmeigenschaften besser zu verstehen. Andererseits kann es auch direkt die Sprachtherapie oder ein Stimmentraining unterstützen, wenn die Patientinn*en oder die Übenden versuchen, die von der KI vorgegebene Aussprache nachzuahmen. Mit den von uns entwickelten KI-basierten Stimmengeneratoren bieten sich viele Möglichkeiten, ein Stimmentraining noch effektiver zu gestalten.“

Prof. Dr. Reinhold Häb-Umbach

Projektleiter C06

Veröffentlicht

Forschungsartikel: Enhancing AI Interaction through Co-Construction: A Multi-Faceted Workshop Framework

In dem Artikel wird ein Workshop-Format des TRR 318 vorgestellt, das Menschen dabei unterstützt, besser mit Large Language Models wie ChatGPT umgehen zu können. Durch Beobachtung, Analyse und praktische Übungen sollen wissenschaftliche Forschung, Lernprozesse und Wissenschaftskommunikation miteinander verbunden werden. Erste Daten zeigen, dass Teilnehmende nach dem Workshop kritischer, reflektierter und dialogorientierter mit dem System umgehen.

[Weiterlesen](#)

Forschungsartikel: Trust, distrust, and appropriate reliance in (X)AI: A conceptual clarification of user trust and survey of its empirical evaluation

In dieser Studie werden die Konzepte von Vertrauen, Misstrauen und angemessenem Vertrauen in KI-Systeme theoretisch geklärt und empirisch untersucht. Es wird analysiert, wie Explainable AI (XAI) mit dem Vertrauen von Nutzer*innen zusammenhängt, und gezeigt, dass die bisherigen empirischen Befunde hierzu uneinheitlich sind. Durch die begriffliche Präzisierung und die Auswertung bestehender Studien entsteht eine Grundlage, um zukünftige Untersuchungen zu Nutzervertrauen und Reliance auf KI-Systeme fundierter zu gestalten.

[Weiterlesen](#)

Forschungsartikel: The power of combined modalities in interactive robot learning

Die Studie untersucht, wie verschiedene Eingabemodalitäten und darüber hinausgehende „Meta-Modalitäten“ das interaktive Lernen von Robotern verbessern können. Die Ergebnisse zeigen, dass zwar einzelne Modalitäten unterschiedlich wahrgenommen werden, ihre Kombination jedoch das Lernverhalten des Roboters und die Nutzbarkeit deutlich steigert. Damit liefert die Arbeit wichtige Impulse für die Gestaltung flexiblerer und nutzerfreundlicherer Mensch-Roboter-Lernumgebungen.

[Weiterlesen](#)

Forschungsartikel: Forms of Understanding for XAI-Explanations

Der Artikel entwickelt ein interdisziplinäres Modell dazu, was „Verstehen“ im Kontext von XAI überhaupt bedeutet und welche Formen es annehmen kann. Dabei wird unterschieden zwischen *enabledness* (dem „Wissen wie“) und *comprehension* (dem „Wissen dass“). Der Artikel beschreibt, wie sich Verstehen im Erklärprozess entwickelt und welche besonderen Herausforderungen dabei speziell in XAI auftreten.

[Weiterlesen](#)

Medientipps

TRR 318-Podcast: „Explaining Explainability“

Der TRR 318-Podcast zeigt, wie Menschen erklären und was KI-Systeme daraus lernen können, in spannenden Folgen wie „Explainability“, „Explaining“, „Understanding“, „Co-Construction“, „Scaffolding“, „Looking Back on the First Funding Phase“ und „Monitoring“. Moderiert von Prof. Dr.-Ing. Britta Wrede diskutieren in jeder Episode zwei Forschende aus unterschiedlichen Disziplinen ein zentrales Konzept der erklärbaren KI und bringen ihre Perspektiven zusammen.

[Podcast hören](#)

Podcast Folge: Wann macht KI Fehler-und wann wir?

In der neuen Folge von „Der KI-Podcast“ der ARD geht es darum, warum KI-Modelle wie Chatbots manchmal falsche Informationen erfinden und welche Arten von Fehlern dabei auftreten können. Es wird erklärt, wie man diese Fehler versteht, was Reasoning-Modelle verändern und warum kritisches Denken im Umgang mit KI entscheidend bleibt.

[Podcast hören](#)

Was habe ich gelernt?

„Als ich vor knapp viereinhalb Jahren im TRR angefangen habe, war ich gespannt darauf zu sehen, wie die Zusammenarbeit in einem Sonderforschungsbereich gestaltet wird. Heute weiß ich, wie wertvoll der Austausch über Disziplingrenzen hinweg ist und dass Wissenschaft insbesondere durch Dialog und gemeinsames Nachdenken entsteht.
In den letzten Jahren durfte ich durch verschiedene Formate im TRR viel lernen und neue Perspektiven gewinnen: etwa durch unsere monatlichen Activity Afternoons, die allen Beteiligten Raum boten, Ideen einzubringen und Feedback aus anderen Teilprojekten zu erhalten, oder durch die Synthesis Groups, in denen Kolleg*innen mit unterschiedlichen Perspektiven gemeinsam an einer Frage arbeiteten. Der TRR schafft insgesamt ein Umfeld, das Austausch und Zusammenarbeit über Disziplingrenzen hinweg fördert. Mit der RTG verfügt er zudem über eine Struktur, die Doktorand*innen gezielt begleitet und unterstützt. Rückblickend habe ich gelernt, wie wichtig es ist, offen für unterschiedliche Perspektiven zu bleiben und sich regelmäßig Zeit für Austausch zu nehmen, auch dann, wenn der eigene Forschungsalltag gerade viel Raum einnimmt.
Mein Rat an zukünftige Doktorand*innen ist, die Vielfalt im TRR zu nutzen, um sich mit anderen auszutauschen und voneinander zu lernen. Besonders bereichernd ist es, dabei auch über die eigene Disziplin hinauszuschauen.“



*Dr. Olesja Lammert
Wissenschaftliche Mitarbeiterin im Projekt A03*



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

[Newsletter abonnieren.](#)

[Top: german version ▲](#)

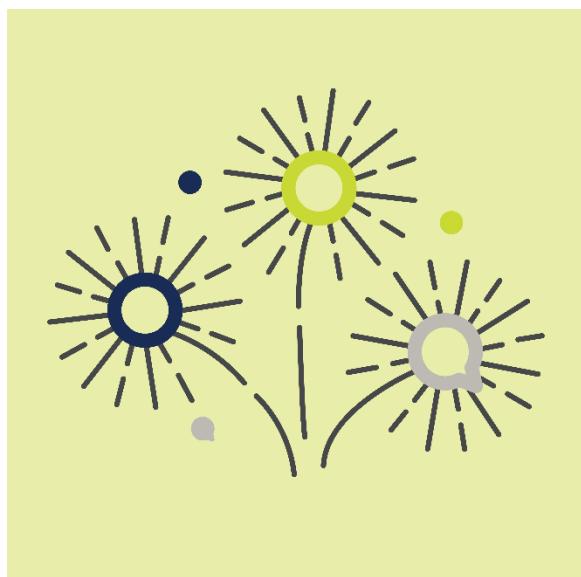
[Footer: Impressum ▼](#)

Developing explanations together | 04.2025

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 [LinkedIn](#) and [Instagram](#) and email us with your questions about AI. Let's develop explanations together!

TRR 318 Has Something to Celebrate!

The German Research Foundation (DFG) has decided to extend our Transregio for an additional three and a half years. Since 2021, the TRR team has researched how to make AI systems more understandable and adaptable to users' needs. In the upcoming funding phase, the focus will be on the context of explanations. How do situations, people, or shared knowledge affect the need for explanations, and how can AI systems respond flexibly to these changes? With over 60 researchers from seven disciplines, TRR 318 will continue its successful interdisciplinary work.





RTG Brings Together Early Career Researcher from Many Disciplines

TRR 318's graduate school provides a structured training program for doctoral researchers and post-doctoral researchers. Ricarda Kurock is the coordinator of the **Research Training Group (RTG)**. In the following interview, she speaks about the offerings, goals, and success of the program. A key element for her is the cooperative, interdisciplinary approach.

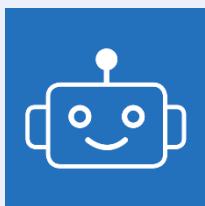
[Read more](#)



Interview with Doctoral Candidates

All early-career researchers of TRR 318 are automatically members of the Research Training Group (RTG). In this interview, psycholinguist Dr. Josephine Fisher and computer scientist Roel Visser talk about their experiences pursuing a PhD within the interdisciplinary projects of the Transregio, as well as about the mentoring provided by the RTG.

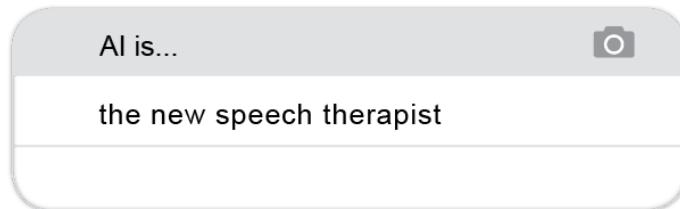
[Read more](#)



Thanks For Visiting!

On November 22, 2025, the TRR 318 Day took place at GENIALE in the Wissenswerkstatt in Bielefeld. Thank you to all the visitors for their interest and the exciting discussions!

[More News](#)



"It is well-known that AI can generate voices by transforming text into spoken language. We have all heard these voices before, for example, in announcements at train stations or on trains. AI has also long been used as a learning aid to improve pronunciation when learning a foreign language.

However, voice generation for speech therapy is a different matter entirely. We are developing AI that can alter specific voice characteristics, such as roughness or breathiness. On the one hand, this helps aspiring speech and voice therapists better understand human voice characteristics. On the other hand, it can directly support speech therapy or voice training by helping patients or practitioners imitate the AI's specified pronunciation. The AI-based voice generators we have developed offer many opportunities to make voice training more effective."



*Professor Reinhold Häb-Umbach
Project Leader C06*

Published

Research Paper: Enhancing AI Interaction through Co-Construction: A Multi-Faceted Workshop Framework

The article describes a workshop format developed by TRR 318 to help people learn how to better use large language models such as ChatGPT. Through observation, analysis, and practical exercises, the workshop links scientific research, learning processes, and science communication. Initial data shows that, after the workshop, participants interact with the system in a more critical, reflective, and dialogue-oriented manner.

[Read more](#)

Research Paper: Trust, distrust, and appropriate reliance in (X)AI: A conceptual clarification of user trust and survey of its empirical evaluation

This study clarifies the concepts of trust, mistrust, and appropriate trust in AI systems from theoretical and empirical perspectives. The study analyzes the relationship between explainable AI (XAI) and user trust, revealing that the existing empirical findings are inconsistent. This conceptual clarification and evaluation of existing studies provides a foundation for designing future studies on user trust and reliance on AI systems more informatively.

[Read more](#)

Research Paper: The power of combined modalities in interactive robot learning

This study examines how combining different input modalities and additional “meta-modalities” can enhance interactive learning in robots. Although individual modalities are perceived differently, the results show that combining them significantly improves the robot's learning behavior and usability. Thus, the study provides important insights for designing more flexible and user-friendly human-robot learning environments.

[Read more](#)

Research Paper: Forms of Understanding for XAI-Explanations

The article develops an interdisciplinary model of what “understanding” means in the context of XAI and which forms it can take. It distinguishes between enabledness (“knowing how”) and comprehension (“knowing that”). The article also describes how understanding develops throughout the explanation process and which specific challenges arise in XAI.

[Read more](#)

Media tips

TRR 318 Podcast: “Explaining Explainability”

The TRR 318 podcast explores how people explain things and how AI systems can learn from this. Exciting episodes include “Explainability,” “Explaining,” “Understanding,” “Co-Construction,” “Scaffolding,” “Looking Back on the First Funding Phase,” and “Monitoring.” Hosted by Professor Britta Wrede, each episode features two researchers from different disciplines discussing a central explainable AI concept and combining their perspectives.

[Listen here](#)

What have I learned?

“When I started at TRR just under four and a half years ago, I was excited to experience collaboration in a research center firsthand. Today, I know how valuable interdisciplinary exchange is, and I understand that science is created through dialogue and joint reflection. In recent years, I have learned a lot and gained new perspectives through various TRR programs. For example, our monthly Activity Afternoons gave everyone the opportunity to contribute ideas and receive feedback



from other subprojects. The Synthesis Groups allowed colleagues with different perspectives to work together on a question. Overall, TRR creates an environment that promotes collaboration and exchange across disciplinary boundaries. The RTG also has a structure that provides targeted guidance and support to doctoral students. In retrospect, I have learned how important it is to remain open to different perspectives and to regularly set aside time for exchange, even when one's own research routine is very demanding.

My advice to future doctoral students is to take advantage of the diversity within the TRR to exchange ideas with others and learn from each other. Looking beyond your own discipline is particularly enriching."

*Dr. Olesja Lammert
Research Associate in project A03*

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