

Erklärungen gemeinsam entwickeln | 02.2025

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 ▼](#)

3. TRR 318-Konferenz: Contextualizing Explanations

Am 17. und 18. Juni findet an der Universität Bielefeld die dritte Konferenz des TRR 318 „Contextualizing Explanations“ (ContEx25) statt. Im Mittelpunkt steht die Frage, wie Erklärungen so an unterschiedliche Kontexte angepasst werden können, dass sie für Nutzer*innen verständlich und hilfreich sind – besonders im Einsatz von Künstlicher Intelligenz in entscheidungskritischen Bereichen. Internationale und interdisziplinäre Forschende beleuchten dabei die Rolle des Kontexts in Erklärprozessen aus verschiedenen Perspektiven. Interessierte, die an der Konferenz teilnehmen möchten, können sich [hier](#) bis zum 1. Juni registrieren.

[Zur Konferenz-Webseite](#)





Neue Podcast-Folge „Co-Construction“

Die vierte Folge des TRR 318-Podcasts „Explaining Explainability“ ist veröffentlicht (in englischer Sprache). Darin beschäftigen sich Moderatorin Prof. Dr.-Ing. Britta Wrede und ihre Gäste Prof. Dr. Henning Wachsmuth und Prof. Dr. Axel Ngonga Ngomo mit Ko-Konstruktion in Sprachmodellen wie ChatGPT und DeepSeek.

[Podcast anhören](#)



Intelligentes Prompting

Im Projekt [INF](#) wurde ein Modell entwickelt, das die optimale Kombination verschiedener Prompt-Techniken vorhersagt, um die Ausgabe von Sprachmodellen zu verbessern. Professor Dr. Henning Wachsmuth und Maximilian Spliethöver haben ihr Modell auf drei Sprachmodelle angewendet und dabei in vielen Fällen bessere Ergebnisse erzielt als bisherige Ansätze. Spliethöver hat die Studie des TRR 318 auf der NAACL 2025 in New Mexico vorgestellt.

[Weiterlesen](#)



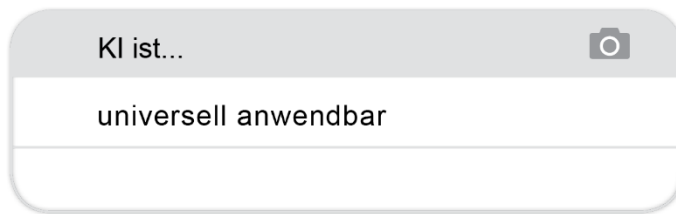
Gastwissenschaftler aus Warschau zu Besuch

Im März war Hubert Baniecki, Doktorand an der Universität Warschau, als Gastwissenschaftler im Projekt C03 an der LMU München zu Besuch. Während seines Aufenthalts forschte er gemeinsam mit dem Team an Methoden zur Erklärung komplexer KI-Systeme. Der Austausch war nicht nur fachlich intensiv, sondern auch persönlich bereichernd – ein Beispiel für die internationale Vernetzung und Zusammenarbeit im TRR 318.



Instagram

Der TRR 318 ist auf Instagram! Folgen Sie unserem Account [sfb_trr318](#), um keine spannenden Einblicke und Updates zu verpassen.



„Das Chat-Fenster eines Sprachmodells suggeriert uns schier unbegrenzte Möglichkeiten. Wir können jede beliebige Frage stellen zu jedem beliebigen Thema und bekommen vom Sprachmodell eine Antwort. Diese generelle Einsetzbarkeit erscheint im Vergleich mit früheren KI-Methoden und -systemen als dramatischer Fortschritt, bei denen man sich häufig mit Fehlermeldungen oder eindeutig absurden Antworten konfrontiert sah, sobald man den Bereich der Trainingsdaten verließ. Tatsächlich aber führt die oberflächliche Plausibilität der Antworten von Sprachmodellen uns häufig in die Irre. Wir schreiben den Modellen Wissen, Verständnis, sogar Empathie zu, die sie gar nicht haben. Für wirklich angemessene Antworten in einem bestimmten Kontext - sei es Hilfe bei den Hausaufgaben, eine wissenschaftliche Frage oder gar eine medizinisch-diagnostische Entscheidung - braucht es Wissen über und Verständnis für diesen Kontext, sowohl bei den Menschen als auch beim KI-System. Sich die relevanten Teile eines Kontexts gemeinsam zu erarbeiten ist ein wesentlicher Teil unserer Vision für ko-konstruktive Erklärungen im TRR 318 und auf der 3. TRR 318-Konferenz: Contextualizing Explanations.“



*Prof. Dr. Benjamin Paaßen
Mitorganisator*in der TRR-Konferenz und assoziiertes
TRR-Mitglied*

Veröffentlicht

Forschungsartikel: Algorithm, Expert, or Both? Evaluating the Role of Feature Selection Methods on User Preferences and Reliance

Die Studie untersucht, wie verschiedene Methoden der Merkmalsauswahl – algorithmusbasiert, expertenbasiert und kombiniert – die Präferenzen der Nutzer*innen sowie ihr Vertrauen in KI-gestützte Entscheidungssysteme beeinflussen. Die Ergebnisse zeigen, dass Nutzer*innen die kombinierte Methode bevorzugen, während sie in der tatsächlichen Nutzung gleich viel Vertrauen in alle Methoden setzen. Laut den TRR-Forschenden deutet dies auf eine Diskrepanz zwischen den Einstellungen der Nutzer*innen und ihrem

tatsächlichen Verhalten hin.

[Weiterlesen](#)

Medientipps

Podcast-Folge: Dr. ChatGPT - Was taugt KI in der Medizin?

In der neuen Folge von „Der KI-Podcast“ der ARD geht es darum, ob Künstliche Intelligenz die Lösung für die Herausforderungen des deutschen Gesundheitssystems sein kann. Dabei wird der Einsatz von KI in der Diagnostik und ihre Anwendung in der Augenheilkunde betrachtet.

[Podcast anhören](#)

Was habe ich gelernt?

„Bei der Planung der 2. TRR 318-Konferenz vor zwei Jahren habe ich gelernt, wie viele verschiedene Aspekte in die Konferenzgestaltung einfließen. Von der inhaltlichen Gestaltung der verschiedenen Slots bis zu sozialen Events gibt es viel zu berücksichtigen. Zuerst gilt es, ein passendes Datum zu finden und Konferenzräume für den Zeitraum anzufragen. Dann muss das Thema der Konferenz festgelegt werden, das wir als TRR 318 gemeinsam bestimmt haben. Die erste öffentlich sichtbare Aktion ist der Call for Papers auf der speziell dafür erstellten Konferenzhomepage.

Besonders an unserem Konferenzorganisationsteam war, dass wir ausschließlich aus Doktorand*innen bestanden. Es war eine tolle Erfahrung, im Team mit Vivien Lohmer (Universität Bielefeld) und Maximilian Muschalik (LMU München) zu arbeiten, da wir alle unsere individuellen Stärken einbringen konnten. Ich habe meine Kreativität in die Gestaltung eines Poetry Slams einfließen lassen, der als soziales Event am Abend in der Studiobühne stattfand. Dort konnten wir uns kreativ mit unseren Forschungsthemen oder allgemein mit dem Thema Wissenschaft auseinandersetzen. Die anderen Organisator*innen brachten ihre Stärken beispielsweise bei der Suche nach Keynote-Speakern aus den jeweiligen Fachrichtungen ein. So konnten wir unsere interdisziplinären und persönlichen Stärken optimal nutzen. Außerdem wurden wir tatkräftig von den zentralen Projekten Ö, Z, RTG und den Sprecher*innen des TRR 318 unterstützt.“



Josephine Beryl Fisher

*Doktorandin im Projekt **A01***



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

Newsletter [abonnieren](#).

[Top: german version](#) ▲

[Footer: Impressum](#) ▼

Developing explanations together | 02.2025

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!

3rd TRR 318 Conference: Contextualizing Explanations

On 17 and 18 June, the third conference of the TRR 318 "Contextualizing Explanations" (ContEx25) will take place at Bielefeld University. The conference will focus on the question of how explanations can be adapted to different contexts in such a way that they are understandable and helpful for users - especially in the use of artificial intelligence in decision-critical areas.

International and interdisciplinary researchers will shed light on the role of context in explanation processes from different perspectives.

Those interested in attending the conference can register [here](#) until June 1.

[To the conference website](#)





New Podcast Episode “Co-Construction”

The fourth episode of the TRR 318 podcast “Explaining Explainability” has been published. In this episode, host Professor Britta Wrede and guests Professor Henning Wachsmuth and Professor Axel Ngonga Ngomo discuss co-construction in large language models such as ChatGPT and DeepSeek.

[Listen in](#)



Intelligent Prompting

The [INF](#) project developed a model that predicts the optimal combination of different prompting techniques to improve the output of language models. Professor Henning Wachsmuth and Maximilian Spliethöver applied their model to three language models and in many cases achieved better results than previous approaches. Spliethöver presented the TRR 318 study at NAACL 2025 in New Mexico.

[Read more](#)



Guest Researcher from Warsaw

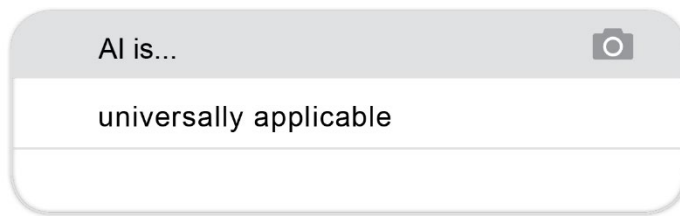
In March, Hubert Baniecki, a PhD student at the University of Warsaw, visited LMU Munich as a guest researcher in project C03. During his stay, he worked with the team on methods for explaining complex AI systems. The exchange was not only professionally intensive, but also personally enriching - an example of the international networking and cooperation in TRR 318.



Instagram

The TRR 318 is on Instagram! Follow our account [sfb_trr318](#) to make sure you don't miss any exciting insights and updates!

[More News](#)



“The chat interface of a language model suggests almost unlimited possibilities. We can ask any question on any topic and get an answer from the language model. This general applicability appears to be a dramatic advance over earlier AI methods and systems, where we were often confronted with error messages or clearly absurd answers as soon as we went beyond the training data. In reality, however, the superficial plausibility of the answers given by language models often leads us astray. We attribute knowledge, understanding, and even empathy to models that they do not have. To achieve appropriate answers in a given context - be it help with homework, a scientific question, or even a medical diagnosis - knowledge and understanding of that context is required, both by humans and by the AI system. Working out the relevant parts of a context together is an essential part of our vision for co-constructive explanations in TRR 318 and at the 3rd TRR 318 Conference: Contextualizing Explanations.”



Professor Benjamin Paaßen
Co-organizer of the TRR Conference and associated
TRR member

Published

Research Paper: Algorithm, Expert, or Both? Evaluating the Role of Feature Selection Methods on User Preferences and Reliance

The study investigates how different feature selection methods—algorithm-based, expert-based, and combined—affect user preferences and their trust in AI-supported decision-making systems. The results show that users prefer the combined method, while they place equal trust in all methods during actual use. According to the TRR researchers, this indicates a discrepancy between users' attitudes and their actual behavior.

[Read more](#)

What have I learned?

“In planning the 2nd TRR 318 conference two years ago, I learned how many different aspects go into conference design. From the content of the various slots to the social events, there is a lot to consider. The first step is to find a suitable date and request conference rooms for that time. Then the topic of the conference needs to be defined, which we as the TRR 318 have determined together. The first publicly visible action is the call for papers on the specially created conference homepage.

What was special about our conference organizing team was that we were exclusively PhD students. It was a great experience to work in a team with Vivien Lohmer (Bielefeld University) and Maximilian Muschalik (LMU Munich), as we were all able to contribute our individual strengths. I used my creativity to design a poetry slam, which took place as a social event in the evening on the studio stage. We were able to creatively explore our research topics or the topic of science in general. The other organizers contributed their strengths, for example in finding keynote speakers from their respective disciplines. This allowed us to make the most of our interdisciplinary and personal strengths. We were also actively supported by the central projects Ö, Z, RTG and the speakers of TRR 318.”



Josephine Beryl Fisher
*PhD student in project **A01***

TRR digital



Or **message us directly** for questions or feedback.

Subscribe to the newsletter for free.

TRR 318 „Constructing Explainability“

Teilprojekt Ö „Fragen zu erklärbaren Technologien“

Universität Bielefeld

Universitätsstraße 25

33615 Bielefeld

communication@trr318.uni-paderborn.de