CONSTRUCTING EXPLAINABILITY

TRR 318

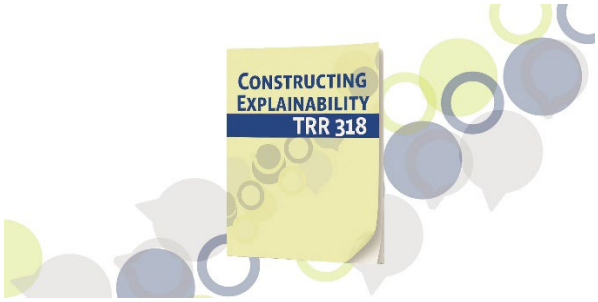
Erklärungen gemeinsam entwickeln | 03.2025

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 ▼

Erste TRR 318-Broschüre ist erschienen

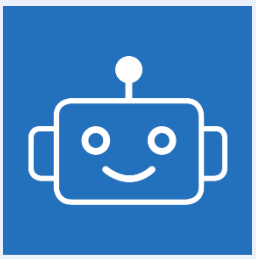
Unsere englischsprachige Broschüre gibt Einblicke in den Forschungsverbund und zeigt, was alles in der ersten Förderphase passiert ist. Ein gedrucktes Exemplar können Sie gerne **hier** anfordern.
Zur Broschüre



News

TRR 318 auf der GENIALE 2025

Am 22. November lädt der TRR 318 zum Science Festival GENIALE in die Wissenswerkstadt Bielefeld ein. Besucher*innen können an interaktiven Stationen spielerisch entdecken, wie Künstliche Intelligenz funktioniert und warum Erklärbarkeit wichtig ist – ob beim Minigolf mit einem Roboterarm, im „Katzen-Hotdog-Spiel" mit ChatGPT oder beim Workshop „Wie eine KI die Welt sieht". Im Experiment „Sag was!" erfahren Besucher*innen, wie KI Unterschiede zwischen Stimmen sichtbar und hörbar machen kann.
Für Schulklassen gibt es außerdem den Workshop „Hausaufgabenhelfer KI?", in dem Schüler*innen lernen, Sprachmodelle wie ChatGPT kritisch zu reflektieren und sinnvoll im Unterricht zu nutzen.
Weiterlesen



Roboter Nao als aufmerksamer Erklärer

André Groß, Dr. Birte Richter und Prof. Dr.-Ing. Britta Wrede aus Projekt **A05** haben ein Modell entwickelt, mit dem der Roboter Nao seine Erklärungen an die Aufmerksamkeit und den kognitiven Zustand von Nutzer*innen anpasst. Anhand von Blickbewegungen und Leistung entscheidet Nao, ob er direkte Anweisungen („schüttle die Flasche“) oder Negationen („nicht schwenken, sondern schütteln“) gibt.
Die Ergebnisse zeigen, dass Versuchspersonen beim Lösen von Aufgaben mit dem Modell 23% weniger Fehler machen.
Weiterlesen



Best Demo Award

Meisam Booshehri hat auf der 21. International Conference on Semantic Systems in Wien den Best Demo Award erhalten. Gemeinsam mit Christian Kullik und Prof. Dr. Philipp Cimiano entwickelte er die Suchmaschine „A Semantic Search Engine for the Exploration of Explainability Phenomena“, die Daten aus verschiedenen TRR 318-Projekten bündelt. Forschende können diese Daten so einfach durchsuchen, vergleichen und für ihre eigenen Projekte nutzen.
Weiterlesen



Muster in Erklärungen sichtbar machen

TRR-Mitglied Vivien Lohmer präsentierte auf der 19. International Pragmatics Association in Brisbane (Australien) Ergebnisse aus dem Projekt **A04**. Die Linguistin sowie Lutz Terfloth aus der Didaktik der Informatik und Linguistik-Professorin Dr. Friederike Kern untersuchen Erklärungen zu technischen Artefakten. Mit ihrer interdisziplinären methodischen Herangehensweise hat das Team Muster zwischen diskursiver und inhaltlich-semantischer Ebene sichtbar machen können. Das Fazit: Die Art und Weise, wie Erklärungen hervorgebracht werden, ist eng mit dem Inhalt verknüpft. Auf Basis der Visualisierung dieser Zusammenhänge können Erklärungen gezielter und über mehrere Datensets hinweg untersucht werden.



Disputationen

Viele der Doktorand*innen im TRR 318 haben ihre Promotion bereits erfolgreich abgeschlossen oder stehen kurz davor. An dieser Stelle herzliche Glückwünsche an: Jaroslaw Kornowicz, Olesja Lammert, Josephine Beryl Fisher, Fabian Fumagalli, Nils Klowait und Stefan Lazarov!



TRR 318-Quiz

Wie gut kennen Sie den TRR 318? Testen Sie Ihr Wissen in unserem TRR 318-Quiz.
Zum Quiz

Weitere News





„Viele der aktuell bekanntesten auf KI – genauer: auf Large Language Models – basierenden Anwendungen werden uns durch ihr Design als Schreibassistenten angeboten. Chatbots wie ChatGPT unterstützen uns bei der Ideenfindung oder beim Erstellen ganzer Texte. Andere Anwendungen übersetzen unsere Entwürfe, verbessern Stil oder korrigieren Rechtschreibung und Grammatik. Solange wir einschätzen können, ob das, was die KI vorschlägt, inhaltlich richtig, dem Kontext und der Leserschaft angemessen und unserer Intention entsprechend ist, mag KI vielleicht tatsächlich manche Prozesse abkürzen oder Sicherheit geben. Problematisch wird es, wenn wir glauben, die KI könne besser sagen, was wir sagen wollen – und ihr große Teile der Autorschaft überlassen. Autorschaft abzugeben heißt aber nicht, Verantwortung abzugeben. Besonders für Lernende ist entscheidend, die Ziele und Zwecke von Texten in sozialen Kontexten zu verstehen. Die Fragen, wer dabei auf welche Weise zu wem spricht und was als guter, richtiger, interessanter, kontroverser oder auf andere Weise relevanter Beitrag zum gemeinsamen Diskurs angesehen wird, sind dabei zentral. Die besten Schreibassistent*innen im Hinblick auf diese Fragen sind Menschen, die Teil der jeweiligen sozialen Kontexte sind. Erst an ihren expliziten und

impliziten Reaktionen auf unsere Texte – und nicht durch die Vorschläge einer KI – lernen wir, so zu schreiben, dass wir wirklich etwas zu sagen haben."

Dr. Andrea Karsten
*Mitarbeiterin des **Graduiertenkollegs** im TRR 318*

Veröffentlicht

Forschungsartikel: Interacting with fallible AI: Is distrust helpful when receiving AI misclassifications
Die Studie untersucht, wie Menschen auf fehlerhafte Empfehlungen von KI-Systemen reagieren und ob gezielte Warnhinweise ihre Urteilsfähigkeit verbessern. Das Ergebnis: Ein Aufruf zur Skepsis führt nicht zu besseren Entscheidungen – im Gegenteil, er kann die Leistung sogar verschlechtern.
Weiterlesen

Forschungsartikel: Speech Synthesis along Perceptual Voice Quality Dimensions
Die Studie betrachtet, wie Sprachsynthese-Systeme gezielt wahrnehmbare Stimmqualitäten wie Rauheit, Hauchen, Resonanz und Stimmfülle verändern können. Dafür wird ein neues Modell eingesetzt, das aus Beispielen lernt, statt direkt akustische Merkmale zu manipulieren, und dessen Ergebnisse von phonetischen Expert*innen bewertet wurden.
Weiterlesen

Ratgeber: Metaphern zum Erklären
Der englischsprachige Ratgeber erforscht, wie Metaphern das Erklären beeinflussen, sowohl im Alltag als auch in der Wissenschaft. Es wird anhand konkreter Beispiele gezeigt, welche Chancen und Risiken sprachliche Bilder bergen und welche Metaphern besonders gut zum Verständnis von Erklärprozessen im TRR 318 passen.
Weiterlesen

Medientipps

Tagesschau-Artikel: Strengere Regeln für ChatGPT und Co. treten in Kraft
Seit dem 2. August 2025 gelten in der EU strengere Vorschriften für große KI-Modelle wie ChatGPT: Anbieter müssen offenlegen, mit welchen Daten ihre Systeme trainiert wurden – insbesondere im Hinblick auf urheberrechtlich geschützte Inhalte – und darlegen, welche technischen Schutzmaßnahmen sie gegen Missbrauch ergreifen. Ziel ist es, Urheberrechte wirksamer zu schützen, mögliche Klagen zu vermeiden und klare Standards für den verantwortungsvollen Umgang mit KI zu setzen.
Artikel lesen

Podcast Folge: Was kommt nach ChatGPT?
In der neuen Folge von „Der KI-Podcast" der ARD geht es um die Frage, wie die nächste Generation von KI-Sprachmodellen aussehen könnte. Diskutiert werden neue Ansätze jenseits immer größerer Modelle, beispielsweise KI mit echtem Weltverständnis, lernfähig wie

der Mensch, oder sogar auf Basis von Gehirnsignalen.
Podcast hören

Was habe ich gelernt?



„Ich habe mich einige Wochen lang ganz auf das Schreiben meiner Dissertation im Rahmen des Projekts A05 des TRR 318 konzentriert. Zuvor hatte ich viele Monate mit Lesen, Schreiben, Programmieren und Strukturieren verbracht – eine Phase, die sich im Nachhinein als zentrales Fundament der Arbeit erwiesen hat. Dank dieser Vorbereitung verlief das eigentliche Schreiben deutlich schneller und zielgerichteter, als ich erwartet hatte. Die zuvor investierte Zeit in die Strukturierung hat mir beim Schreiben Orientierung und Sicherheit gegeben.
Auch wenn das Schreiben streckenweise mühsam und zeitintensiv war und es mir an manchen Tagen schwerfiel, überhaupt einen Satz zu Papier zu bringen, empfand ich es zugleich als befreiend. Es bot mir den Raum, Gedanken, die über Monate hinweg in meinem Kopf kreisten, endlich klar zu formulieren und in die Dissertation einfließen zu lassen.
Im Verlauf dieses Prozesses wurde mir zunehmend bewusst, dass ein Dissertationsthema nie wirklich abgeschlossen ist. Es gibt immer etwas, das sich überarbeiten oder erweitern ließe. Diese Erkenntnis hat meinen Blick auf die Dissertation selbst verändert: Ich betrachte sie heute weniger als abgeschlossenes Produkt, sondern vielmehr als Zwischenstand eines fortlaufenden Forschungsprozesses.
Rückblickend war die Dissertation für mich nicht nur ein wissenschaftliches Projekt, sondern auch eine intensive Phase der persönlichen Auseinandersetzung, in der ich viel über mein eigenes Denken und Arbeiten gelernt habe.“

André Groß
*Doktorand im Projekt **A05***

TRR digital

Oder direkt per Mail mit Fragen oder Feedback an uns.

Newsletter abonnieren.

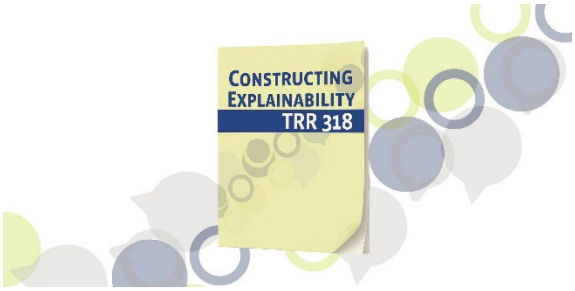
Top: german version ▲
Footer: Impressum ▼

Developing explanations together | 03.2025

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!

The first TRR 318 Brochure is now available!

This brochure offers insights into the research network and highlights the accomplishments of the first funding period.
Request a printed copy **here**.
To the brochure

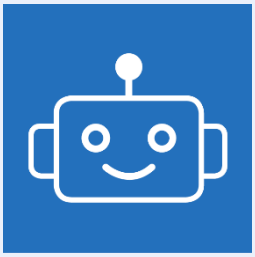


News

TRR 318 at GENIALE 2025

On November 22, TRR 318 is inviting visitors to the GENIALE Science Festival at the Wissenswerkstadt Bielefeld. Visitors can discover how artificial intelligence works and why explainability is important through interactive activities, such as playing mini golf with a robot arm, playing the “cat hot dog game“ with ChatGPT, and attending the “How AI sees the world“ workshop. At the “Say Something!“ experiment, visitors will learn how AI can distinguish

between voices and make them visible and audible.
School classes can also attend the workshop “AI as a Homework Helper,“ where students will learn to critically reflect on language models such as ChatGPT and use them sensibly in class.
Read more



Robot Nao as an Attentive Explainer

André Groß, Dr. Birte Richter, and Professor Britta Wrede of project **A05** developed a model that enables the Nao robot to adapt its explanations to users' attention and cognitive states. Based on eye movements and performance, Nao decides whether to give direct instructions ("shake") or negations ("do not shake"). Results show that participants make 23% fewer errors when using the model to solve tasks.
Read more



Best Demo Award

Meisam Booshehri received the Best Demo Award at the 21st International Conference on Semantic Systems in Vienna. Together with Christian Kullik and Prof. Dr. Philipp Cimiano, he developed the search engine “A Semantic Search Engine for the Exploration of Explainability Phenomena”, which bundles data from various TRR 318 projects. Researchers can easily search, compare, and use these datasets for their own projects.
Read more



Making Patterns Visible in Explanations

TRR member Vivien Lohmer presented findings from project **A04** at the 19th International Pragmatics Association Conference in Brisbane, Australia. The linguist, together with Lutz Terfloth from the Department of Computer Science Education and linguistics professor Friederike Kern, is investigating explanations of technical artefacts. Using their interdisciplinary methodological approach, the team has visualized patterns between the discursive and content-semantic levels. Their conclusion: the way explanations are produced is closely linked to their content. By visualizing these relationships, explanations can be examined more precisely and across multiple data sets.

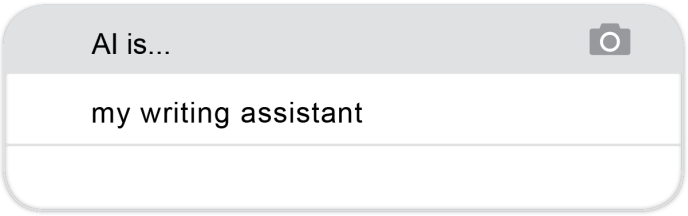


Disputations Many of the doctoral students in TRR 318 have successfully defended their theses or are about to do so. Congratulations to Jaroslaw Kornowicz, Olesja Lammert, Josephine Beryl Fisher, Fabian Fumagalli, Nils Klowait and Stefan Lazarov!



TRR 318 Quiz How well do you know the TRR 318? Test your knowledge in our TRR 318 quiz.
To the quiz

More News





“Many of the best-known AI applications – or, more precisely, large language models – are presented to us by design as writing assistants. Chatbots, such as ChatGPT, help us find ideas and create entire texts. Other applications translate our drafts, improve style, and correct spelling and grammar. As long as we can assess whether the AI's suggestions are factually correct, appropriate to the context and audience, and in line with our intentions, AI may indeed help shorten certain processes or provide a sense of security. However, it becomes problematic when we believe the AI is better at expressing what we want to say, and leave a large part of the authorship to it. Nevertheless, relinquishing authorship does not mean relinquishing responsibility. Understanding the aims and purposes of texts in social contexts is crucial, especially for learners. Questions about who speaks to whom, how, and what constitutes a good, correct, interesting, controversial, or otherwise relevant contribution to the everyday discourse are central to this. The best writing assistants for these questions are people who are part of the respective social contexts. It is only through their explicit and implicit reactions to our texts – and not through the suggestions of an AI – that we learn to write in a way that truly means we have something to say.”

Dr. Andrea Karsten
*Research associate of the **Research Training Group** within TRR 318*

Published

Research Paper: Interacting with fallible AI: Is distrust helpful when receiving AI misclassifications?
This study investigates how people react to incorrect recommendations from AI systems, and whether targeted warnings improve judgment. The findings reveal that encouraging skepticism does not lead to better decisions; in fact, it can even worsen performance.
Read more

Research Paper: Speech Synthesis along Perceptual Voice Quality Dimensions
This study examines how speech synthesis systems can alter perceptual voice qualities, such as roughness, breathiness, resonance, and weight. A new model that learns from examples instead of directly manipulating acoustic features was used to achieve this, and its results were evaluated by phonetic experts.
Read more

Guidebook: Metaphors for Explaining
The guidebook explores how metaphors influence explanations in everyday life and science. Using concrete examples, it demonstrates the opportunities and risks associated with linguistic images, as well as which metaphors are particularly well-suited to understanding explanatory processes in TRR 318.
Read more

What have I learned?



“For several weeks, I focused entirely on writing my dissertation for project A05 of TRR 318. Before that, I spent many months reading, writing, programming, and structuring—a phase that, in hindsight, proved to be the foundation of my work. Thanks to this preparation, the actual writing process was much faster and more purposeful than I had expected. The time I invested in structuring gave me both direction and confidence during the writing phase.
Although writing was at times difficult and demanding, and there were days when even a single sentence felt out of reach, it was also a liberating experience. It gave me the opportunity to finally bring long-circling thoughts onto the page and express them clearly within the dissertation.
Throughout this process, I realized that a dissertation topic is never truly finished. There is always something that could be revised or expanded upon. This

realization changed the way I view the dissertation itself. I've come to see it as a milestone in a larger, ongoing process rather than a final product.
The dissertation turned out to be more than just a scientific project for me; it also became an intensive phase of personal reflection, during which I gained valuable insights into my own thinking and working habits."

André Groß
*PhD student in project **A05***

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

Wenn Sie diese E-Mail (an: trr318_news@lists.uni-paderborn.de) nicht mehr empfangen möchten, können Sie diese **hier** kostenlos abbestellen.

If you no longer wish to receive this email (to: trr318_news@lists.uni-paderborn.de), you can unsubscribe free of charge **here**.