

## Erklärungen gemeinsam entwickeln | 01.2025

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](#) ▼

### Fragen zu KI?

Der TRR 318 lädt dazu ein, Fragen rund um KI zu stellen. Im letzten Jahr wurden in Bielefelder und Paderborner Kneipen Postkarten verteilt, über die Interessierte ihre Fragen direkt an Forschende richten konnten, online oder per Post. Zahlreiche Fragen wurden eingereicht, etwa zu den Auswirkungen von KI auf Arbeit und Gesellschaft oder zur Zukunft von Bildern und Texten. Expert\*innen des TRR 318 haben diese Fragen beantwortet – die Antworten sind auf der Website abrufbar. Und die Aktion läuft weiter: Wer Fragen zu KI hat, kann diese auf der Webseite stellen. Das Forschungsteam des TRR 318 freut sich auf den Austausch!

[Fragen und Antworten](#)





### Neue Podcast Folge „Understanding“

Die dritte Folge des TRR 318 Podcasts „Explaining Explainability“ ist veröffentlicht. Darin beschäftigen sich Moderatorin Prof. Dr.-Ing. Britta Wrede und ihre Interviewpartner\*innen Prof. Dr. Hendrik Buschmeier und Prof. Dr. Heike Buhl mit dem Prozess des Verstehens.

[Podcast anhören](#)



### Erklärvideo von TRR 318 Forscher

In einem neuen Erklärvideo widmet sich Nils Klowait der Frage: *Wie kann man KI demokratisch gestalten?* Anhand der Geschichte des Fahrrads zeigt er, dass technologische Entwicklungen aus Kompromissen verschiedener Interessengruppen entstehen. Da KI viele gesellschaftliche Bereiche beeinflusst, ist ihre Entwicklung besonders sensibel. Dies macht die demokratische Gestaltung von KI so entscheidend.

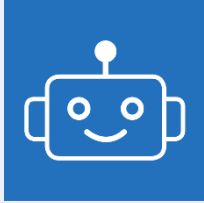
[Video ansehen](#)



### Fragen an Dr. David Johnson

TRR 318 Assoziierter Dr. David Johnson leitet die unabhängige Forschungsgruppe „Human-Centric Explainable AI“ und erforscht, wie Erklärungen in der Mensch-KI-Interaktion gemeinsam konstruiert werden. In einem Interview gibt er Einblicke in die Forschung und Ziele seiner Forschungsgruppe.

[Weiterlesen](#)



## Neue Open-Source-Software

Forschende des Projekts C03 haben *shapiq* entwickelt, ein Open-Source-Tool zur Verbesserung erklärbarer KI. Die Software ergänzt bestehende Methoden, indem es nicht nur einzelne Merkmale analysiert, sondern auch deren Wechselwirkungen berücksichtigt. Dadurch schafft *shapiq* mehr Transparenz für komplexe Modelle wie neuronale Netze und Entscheidungsbäume und stärkt Vertrauen, Fairness und Optimierung in der KI.

[Weiterlesen](#)

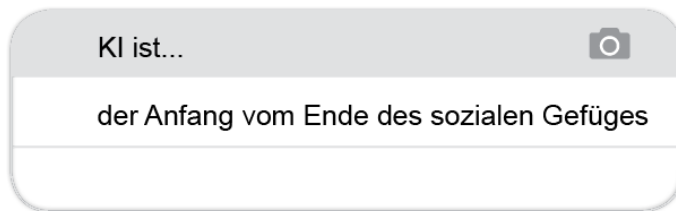


## Instagram

Der TRR 318 ist auf Instagram! Folgen Sie unserem Account [sfb\\_trr318](#), um keine spannenden Einblicke und Updates zu verpassen.

Weitere News

---



„Die KI ist eine radikale Umgestaltung der Gesellschaft, die tiefgreifende Auswirkungen auf die Art und Weise haben wird, wie Sozialität verstanden und praktiziert wird. Aber dies ist nicht das erste Mal, dass so etwas geschieht. In der Vergangenheit hat unsere Gesellschaft die Auswirkungen anderer radikaler Innovationen verarbeitet, wie vor allem die Schrift, die es uns seit Jahrtausenden ermöglicht, miteinander zu kommunizieren, ohne einen Menschen vor uns zu haben. In schriftlicher Form kommunizieren wir mit räumlich und zeitlich weit entfernten Menschen, die uns vielleicht unbekannt oder sogar anonym sind. Wie heute, wurde es damals befürchtet, dass dies zum Ende der Sozialität führen würde - während die Erfahrung gezeigt hat, dass sich komplexere und artikuliertere Formen der Sozialität entwickelt haben.“



*Prof. Dr. Elena Esposito*  
*Projektleiterin **B01***

## Veröffentlicht

### **Forschungsartikel: Explain with, rather than explain to: How explainees shape their own learning**

In diesem Artikel untersuchen Forschende des TRR 318 gemeinsam mit Kolleg\*innen des University College London, die Rolle der Personen, die eine Erklärung erhalten und ihren Einfluss auf den eigenen Lernerfolg. Dafür wurden Erklärdialoge von Erwachsenen zu (answesenden und abwesenden) unbekanntem Objekten analysiert. Wenn das unbekannte Objekt anwesend war, wurde häufiger nach den Funktionen gefragt.

[Weiterlesen](#)

## Medientipps

### Podcast Folge: ChatGPT als Co-Autor? KI beim wissenschaftlichen Schreiben.

In der neuen Folge des Podcasts „Autonomie & Algorithmen“ von Dr. Christiane Attig und dem TRR Assoziierten Prof. Dr. Benjamin Paaßen wird die Rolle von ChatGPT im wissenschaftlichen Schreiben diskutiert.

[Podcast anhören](#)

## Was habe ich gelernt?

„In unserer Forschung zu Explainable AI untersuchen wir, wie Menschen KI im Alltag erleben und welche Erklärmomente für sie wichtig sind. Besonders spannend finde ich, wie vielfältig die Perspektiven auf KI sind – sei es durch persönliche Erfahrungen oder alltägliche Begegnungen. Unsere Interviews und Gruppendiskussionen zeigen, dass es essenziell ist, Nutzer\*innen ohne technischen Hintergrund aktiv in die Entwicklung von KI einzubeziehen. Denn neben der Frage 'Was ist technisch möglich?' muss immer auch reflektiert werden: 'Wie und für wen wird KI gestaltet?' Wir haben in den Gesprächen erfahren, dass Erklärbarkeit im Alltag auf ganz unterschiedliche Weise verstanden wird und oft nur kontextsensitiv betrachtet werden kann. Bis hierhin habe ich gelernt, dass eine technische Erklärung von einer KI als hilfreich und sogar als partizipatives Moment wahrgenommen werden kann. Manchmal geht es aber auch einfach darum, dass ein KI-System ohne Begründung oder Hin und Her funktioniert und die Interaktion zwischen Nutzenden und KI im Rahmen des Alltags reibungslos gelingt. Diese Vielfalt an Eindrücken weiter zu erforschen, wird eine spannende Herausforderung sein.“

*Patrick Henschen*

*Doktorand im Projekt **B03***



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 (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!

### Questions about AI?

TRR 318 invites people to ask questions about AI. Last year, postcards were distributed in pubs in Bielefeld and Paderborn, giving people the opportunity to send their questions directly to the researchers, either online or by mail. Numerous questions were submitted, for example on the impact of AI on work and society or the future of images and texts. Experts from TRR 318 answered these questions - the answers are available on the website.

And the campaign continues: Anyone with questions about AI can ask them on the website. The TRR 318 research team is looking forward to the exchange!

[Questions and Answers](#)





## New Podcast Episode “Understanding”

The third episode of the TRR 318 podcast “Explaining Explainability” has been published. In this episode, host Professor Britta Wrede and interviewees Professor Hendrik Buschmeier and Professor Heike Buhl discuss the process of understanding.

[Listen in](#)



## Explanatory Video by TRR 318 Researcher

In a new explanatory video, Nils Kloweit explores the question: How can AI be designed democratically? Using the history of the bicycle as an example, he shows that technological developments are the result of compromises between different interest groups. As AI affects many areas of society, its development is particularly sensitive. This is what makes the democratic design of AI so crucial.

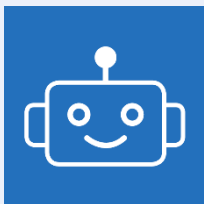
[Watch video](#)



## Questions for Dr. David Johnson

TRR 318 Associate Dr. David Johnson leads the independent research group “Human-Centric Explainable AI” and explores how explanations are co-constructed in human-AI interactions. In an interview, he provides insights into the research and goals of his research group.

[Read more](#)



## New Open Source Software

Researchers from the C03 project have developed *shapiq*, an open source tool for improving explainable AI. The software complements existing methods by not only analyzing individual features, but also taking their interactions into account. In this way, *shapiq* brings more transparency to complex models such as neural networks and decision trees and strengthens trust, fairness and optimization in AI.

[Read more](#)





## 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

AI is...



the beginning of the end of the social structure

“AI is a radical change in society, with profound implications for how sociality is understood and practiced. However, this is not the first time something like this has happened. In the past, our society has processed the profound impacts of other innovations, such as written language, allowing us to communicate without having a human being in front of us. With written language, we communicate with people who are far away, unknown to us, or even anonymous. Like nowadays, in these times people feared that this would lead to the end of sociality. On the contrary, experience has shown the development of more articulate and complex forms of sociality”.

*Professor Elena Esposito*  
*Project Leader **B01***



### Research Paper: Explain with, rather than explain to: How explainees shape their own learning

In this article, researchers from TRR 318, together with colleagues from University College London, investigate the role of people who receive explanations and their influence on their own learning success. For this purpose, adult explanatory dialogues about (present and absent) unknown objects were analysed. When the unknown object was present, more questions were asked about its functions.

[Read more](#)

## What have I learned?

"In our research on explainable AI, we are investigating how people experience AI in everyday life and which explanatory moments are important to them. I find it particularly exciting to see how diverse the perspectives on AI are, whether through personal experiences or everyday encounters. Our interviews and group discussions show that it is essential to actively involve users without a technical background in the development of AI. In addition to the question 'What is technically possible?', it is always necessary to reflect on 'How and for whom is AI designed?'.

During the discussions, we learned that explainability is understood in very different ways in everyday life and can often only be viewed in a context-sensitive way. So far, I have learned that a technical explanation of an AI can be perceived as helpful and even as a participatory moment. Sometimes, however, it is simply a matter of an AI system functioning without justification or back and forth and the interaction between users and AI succeeding smoothly in the context of everyday life. Exploring this variety of impressions further will be an exciting challenge."

*Patrick Henschen*

*PhD student in project **B03***



TRR digital



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

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



UNIVERSITÄT  
PADERBORN



UNIVERSITÄT  
BIELEFELD

Gefördert durch



Deutsche  
Forschungsgemeinschaft

**TRR 318 „Constructing Explainability“**

Teilprojekt Ö „Fragen zu erklärbaren Technologien“

Universität Bielefeld  
Universitätsstraße 25  
33615 Bielefeld

[communication@trr318.uni-paderborn.de](mailto:communication@trr318.uni-paderborn.de)

Wenn Sie diese E-Mail (an: [trr318\\_news@lists.uni-paderborn.de](mailto: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](mailto:trr318_news@lists.uni-paderborn.de)), you can unsubscribe free of charge [here](#).