



# CONSTRUCTING EXPLAINABILITY

## Erklärungen gemeinsam entwickeln | 01.2024

Im Newsletter des Transregios (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 [X](#) 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 ▼](#)

### KI & Bildung

In der heutigen Zeit wachsen Kinder in einer Welt auf, die von Technologie durchdrungen ist – von Sprachassistenten bis hin zu Chatbots. Doch die Themen rund um Künstliche Intelligenz finden selten ihren Weg in die Kinder- und Jugendliteratur. Katharina Rohlfing, Professorin für Psycholinguistik an der Universität Paderborn und Sprecherin des SFB/TRR 318 „Constructing Explainability“, erklärt in einem aufschlussreichen Interview, warum es wichtig ist, KI in Geschichten für junge Leser\*innen zu integrieren. Sie betont, dass Kinder dadurch lernen können, kritisch mit Technologie umzugehen und ihre eigene Rolle in der Gestaltung dieser Entwicklungen zu verstehen. Rohlfing, die



selbst an mehreren bald erscheinenden Kinderbuchprojekten beteiligt ist, hebt hervor, wie diese Literaturform Kindern helfen kann, über die Herausforderungen und Möglichkeiten von KI nachzudenken und zu diskutieren. Erfahrt mehr über die Bedeutung von KI in der Kinderliteratur und die innovativen Projekte, die dieses spannende Thema aufgreifen.

[Weiterlesen](#)

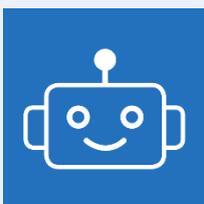
## News



### **Kommentare zum AI Act**

Die EU hat ein KI-Gesetz eingeführt, um den Einsatz von Künstlicher Intelligenz zu regulieren und eine verantwortungsvolle Nutzung sicherzustellen. Wir haben Forschende des TRR 318 nach ihrer Meinung gefragt und ihre Kommentare auf der Webseite gesammelt.

[Weiterlesen](#)



### **Neue unabhängige Forschungsgruppe**

Die neue Forschungsgruppe „Human-Centric Explainable AI“ („Menschenzentrierte Erklärbare KI“) unter der Leitung von Dr. David Johnson ist gestartet und wird die TRR 318 Forschung ergänzen. Die Forschungsgruppe befasst sich mit KI und ihren Auswirkungen auf eine übermäßige Abhängigkeit von KI-gestützter Entscheidungsfindung.

[Weiterlesen](#)



**TRR Workshop** Vom 15. bis 17. Mai fand an der Universität Paderborn der TRR Workshop „Ethics And Normativity of Explainable AI“ statt. Dafür sind führende internationale Forschende zusammengekommen, um über die Ethik künstlicher Intelligenz zu diskutieren.

[Weiterlesen](#)

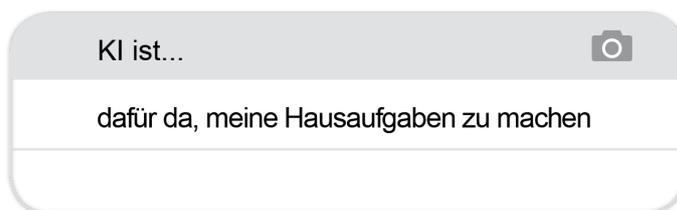


## 2. Folge von „Explaining Explainability“

In der zweiten Folge des Forschungspodcast diskutieren Moderatorin Prof. Britta Wrede und ihre Gäste Prof. Katharina Rohlfing und Prof. Philipp Cimiano über das Thema 'Explaining' und befassen sich mit den Mechanismen des Erklärens sowie dessen Bedeutung für die XAI-Forschung.

[Mehr entdecken](#)

Weitere News



„Obwohl KI zweifellos auch nützlich sein kann, um Hausaufgaben zu erledigen, ist die Bedeutung der Technologie und ihr Potenzial weitaus größer als nur die Erledigung von Hausaufgaben. Mit KI können komplexe Probleme gelöst, medizinische Diagnosen unterstützt und Innovationen vorangetrieben werden. Dennoch beschäftigen sich derzeit viele Lehrende in Schule und Hochschule mit digitalen Technologien und KI. Oft geht es dabei um die Frage, wie KI konkret dazu genutzt werden kann, Lernprozesse zu gestalten oder auch die bestehende Prüfungskultur zu verändern.“

Will man Aufgaben, gerne auch Hausaufgaben lösen, dann erfordert das zunächst ein Problembewusstsein, dieses Bewusstsein für ein Problem und das Verständnis über eine oder mehrere Lösungsansätze sind dabei eine grundlegend menschliche Herausforderung. Hausaufgaben fördern die Entwicklung von Selbstständigkeit und Eigenverantwortung. Schüler\*innen müssen lernen, ihre Zeit effektiv zu nutzen und



ihre Aufgaben eigenständig zu bewältigen, was wichtige Fähigkeiten für das spätere Leben sind. Hausaufgaben sollten auch ein wichtiges Instrument sein, um den Lernfortschritt der Schüler\*innen zu verfolgen und individuelles Feedback zu geben. Daraus konstituiert sich unter anderem ein Betreuungs- und Vertrauensverhältnis zwischen Lehrkräften und Schüler\*innen. Wird dieser Prozess ausgelagert, dann hat das Auswirkungen auf die persönlichen Beziehungen und das pädagogische Verhältnis. Eine geschickte, sinnvolle wie auch innovative Integration von KI erfordert klare Regeln, das Bewusstsein für ethische Dimensionen im Umgang mit den Technologien und gezielte Integrationen in didaktische Konzepte. Damit Schülerinnen und Schüler einen guten Umgang mit neuen Technologien erlernen können, braucht es auch geeignete Fortbildungsangebote für Lehrkräfte.“

*Prof. Dr. Dan Verständig*

*Professor für Erziehungswissenschaft an der Universität  
Bielefeld, [TRR-Assoziierter](#)*

## Veröffentlicht

### **Forschungsartikel: Humans in XAI: Increased Reliance in Decision-Making Under Uncertainty by Using Explanation Strategies**

In dieser Studie wird diskutiert, dass KI-basierte Entscheidungsunterstützungssysteme zunehmend in der Lage sind, komplexe Entscheidungsprozesse, insbesondere in unsicheren Szenarien, zu erklären. Die Transparenz für den Endnutzer bleibt jedoch begrenzt. Die Forschung zeigt, dass ein geführter Erklärungsansatz das Vertrauen der Nutzer mehr fördert als eine transparente Strategie, und manchmal kann der Verzicht auf Erklärungen paradoxerweise zu einem höheren Vertrauen der Nutzer führen.

[Weiterlesen](#)

### **Forschungsartikel: On Feature Removal for Explainability in Dynamic Environments**

Die Studie untersucht entfernungsbasierte Erklärungen in dynamischen Lernumgebungen und vergleicht zwei Algorithmen auf synthetischen Datenströmen. Es wird gezeigt, dass sie unterschiedliche Erklärungen liefern, insbesondere für korrelierte Merkmale, und es werden Empfehlungen für die Anwendung gegeben.

[Weiterlesen](#)

## Medientipps

### **Podcast „Werkstatt-Gespräch - KI & Bildung“**

Im Podcast „Werkstatt-Gespräch - KI & Bildung“ der Bundeszentrale für politische Bildung unterhalten sich Expert\*innen aus IT, Wissenschaft, Schule und Journalismus, um Klarheit rund um das Thema KI und Bildung zu bringen.

[Podcast hören](#)

### **Podcast „Cornelsen Praxisguide - KI im Unterricht“**

In diesem Podcast „Cornelsen Praxisguide - KI im Unterricht“ berichten Lehrkräfte über ihre Erfahrungen und Praxisbeispiele zum Einsatz von KI im Unterricht.

[Podcast hören](#)

### **Diskussionspapier „Künstliche Intelligenz in der Bildung - Drei Zukunftsszenarien und fünf Handlungsfelder“**

Die fortschreitende Digitalisierung und der Einsatz von Künstlicher Intelligenz in der Bildung ermöglichen neue individualisierte Lernmöglichkeiten. Dieses Diskussionspapier stellt drei Zukunftsszenarien für den Einsatz von KI in Bildungseinrichtungen vor und zeigt fünf Handlungsfelder für eine erfolgreiche Anwendung auf, um den kooperativen Austausch zwischen den relevanten Stakeholdern zu fördern.

[Papier lesen](#)

### **YouTube Video „The EU AI Act Explained“**

Das Video des Kanals „EU Made SIMPLE“ nimmt das neue KI-Gesetz der Europäischen Union unter die Lupe. Außerdem geht es um die Frage, welche Auswirkungen der AI Act auf das beliebte KI-System 'ChatGPT' haben könnte.

[Video sehen](#)

## Gesucht

### **Teilnehmende für Studien**

Eine Auswahl:

- Gesellschaftsspiele erklären und erklärt bekommen,
- sich von virtuellen Agenten anleiten lassen,
- Alltagsbegegnungen mit künstlicher Intelligenz festhalten,
- Robotern Bewegungen beibringen.

[Zur Webseite](#) für mehr Informationen zu den aktuellen Studien. Oder: Mailingliste mit Einladungen zu neuen Studien [abonnieren](#).

## Schulklassen für KI-Workshops an der Universität Paderborn

Ein TRR-Forschungsteam aus dem Fachbereich Didaktik der Informatik bietet für Schulklassen Workshops zu KI an. Vorwissen ist nicht erforderlich, die Termine werden individuell vergeben. Interessierte Lehrkräfte sind eingeladen, sich per E-Mail anzumelden: [communication@trr318.uni-paderborn.de](mailto:communication@trr318.uni-paderborn.de)

## Was habe ich gelernt?

„Das übergeordnete Ziel unserer Öffentlichkeitsarbeit ist ein Austausch mit der breiten Bevölkerung über die Visionen, Ziele und Forschungsergebnisse des TRR 318. Um dieses Ziel zu erreichen, nutzen wir eine ganze Reihe unterschiedlicher Formate. Mein Fokus liegt dabei auf der Konzeption und Durchführung von Workshops, in denen die Öffentlichkeitsarbeit mit der Forschung verknüpft wird. Dabei sollen die Workshops eine interaktive Umgebung bieten, um Ideen auszutauschen, Konzepte zu vertiefen und praktische Fähigkeiten im Umgang mit KI zu entwickeln. Gleichzeitig sind die Workshops selbst Mittelpunkt der Forschung in meinem Teilprojekt.

Bei der Durchführung der Workshops habe ich schnell festgestellt, dass das Wissen über Künstliche Intelligenz bei Schüler\*innen sehr unterschiedlich ist. Während die meisten den Begriff mit ChatGPT oder Sprachassistenten wie Alexa oder Siri verbinden, ist nur wenigen bewusst, dass beispielsweise auch TikTok, YouTube oder Netflix für personalisierte Angebote auf künstliche Intelligenz zurückgreifen. Es zeigt sich häufig, dass der Begriff verwendet wird, um eine BlackBox zu beschreiben, deren Funktionsweise sich die Schüler\*innen nicht mehr logisch erklären können. Doch nicht nur bei den Vorkenntnissen, sondern auch bei den Ansichten zur Vertrauenswürdigkeit von KI gibt es häufig unterschiedliche Ansichten. Durch die von den Teilnehmer\*innen geprägten Diskussionen verlaufen die Workshops immer unterschiedlich, sodass man selbst



flexibel bleiben muss, damit der Workshop zum Erfolg für alle Beteiligten wird und die Öffentlichkeitsarbeit erfolgreich mit der Forschung verbunden werden kann.“

*Michael Lenke*

Doktorand im **Projekt Ö**

---

TRR digital



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

Newsletter **abonnieren**.

Top: [german version](#) ▲

Footer: [Impressum](#) ▼

## Developing explanations together | 01.2024

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



---

### AI & Education

Nowadays, children are growing up in a world permeated by technology - from voice assistants to chatbots. However, the

topics surrounding artificial intelligence rarely find their way into children's and young adult literature. Katharina Rohlfing, Professor of Psycholinguistics at Paderborn University and spokesperson for the SFB/TRR 318 "Constructing Explainability", explains in an insightful interview why it is important to integrate AI into stories for young readers. She emphasizes that children can learn to deal critically with technology and understand their own role in shaping these developments. Rohlfing, who herself is involved in several upcoming children's book projects, emphasizes how this form of literature can help children think about and discuss the challenges and opportunities of AI. Find out more about the importance of AI in children's literature and the innovative projects that address this exciting topic.

[Read more](#)



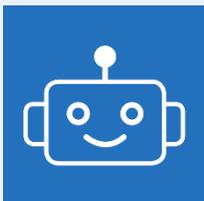
## News



### Comments on the AI Act

The EU has introduced an AI law to regulate the use of artificial intelligence and ensure its responsible use. We asked TRR 318 researchers for their opinions and collected their comments on the website.

[Read more](#)



### New independent research group

The new research group "Human-Centric Explainable AI", led by Dr David Johnson, has started and will complement the TRR 318 research. The research group focuses on AI and its implications for over-reliance on AI-assisted decision-making.

[Read more](#)



## TRR Workshop

From May 15 to 17, the TRR Workshop "Ethics And Normativity of Explainable AI" took place at Paderborn University. Leading international researchers came together to discuss the ethics of artificial intelligence.

[Read more](#)

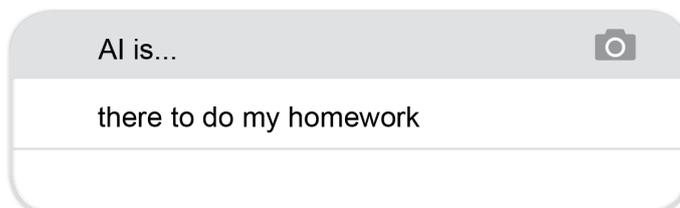


## 2nd episode of "Explaining Explainability"

In the second episode of the research podcast, host Prof. Britta Wrede and her guests Prof. Katharina Rohlfing and Prof. Philipp Cimiano discuss the topic of 'Explaining' and look at the mechanisms of explaining and its significance for XAI research.

[Discover more](#)

[More News](#)



"Although AI can undoubtedly also be useful for doing homework, the importance of the technology and its potential is far greater than just doing homework. AI can be used to solve complex problems, support medical diagnoses and drive innovation. Nevertheless, many teachers in schools and universities are currently focusing on digital technologies and AI. The question is often how AI can be used specifically to shape learning processes or change the existing examination culture.

If you want to solve tasks, including homework, then this first requires an acknowledgement of the problem; this awareness of a problem and the understanding of one or more



approaches to solving it are a fundamentally human challenge. Homework promotes the development of independence and personal responsibility. Students must learn to use their time effectively and manage their tasks independently, which are important skills for later life. Homework should also be an important tool for tracking students' learning progress and providing individualized feedback. Among other things, this results in a relationship of mentoring and trust between teachers and students. If this process is outsourced, it has an impact on personal relationships and the pedagogical relationship. A skillful, meaningful and innovative integration of AI requires clear rules, an awareness of ethical dimensions in dealing with the technologies and targeted integration into didactic concepts. In order for students to learn how to use new technologies well, teachers also need suitable training opportunities."

*Prof. Dr. Dan Verständig*

*Professor of Educational Science at Bielefeld University,  
TRR-Associate*

## Published

### **Research Paper: Humans in XAI: Increased Reliance in Decision-Making Under Uncertainty by Using Explanation Strategies**

This study discusses that AI-powered decision support systems are becoming better at explaining complex decision-making processes, especially in uncertain scenarios. However, end-user transparency remains limited. Research shows that a guided explanation approach increases user reliance more than a transparent strategy, and sometimes, not providing explanations can paradoxically lead to higher user trust.

[Read more](#)

### **Research Paper: On Feature Removal for Explainability in Dynamic Environments**

The study investigates removal-based explanations in dynamic learning environments and compares two algorithms on synthetic data streams. It is shown that they provide different explanations, especially for correlated features, and recommendations for application are given.

[Read more](#)

## What have I learned?

"The overarching goal of our public relations work is to engage with the general public about the visions, goals and research results of TRR 318. We use a whole range of different formats to achieve this goal. My focus is on the conception and implementation of workshops in which public relations work is linked to research. The workshops are intended to provide an interactive environment for exchanging ideas, deepening concepts and developing practical skills in dealing with AI. At the same time, the workshops themselves are the focus of the research in my sub-project.



When conducting the workshops, I quickly realized that students' knowledge of artificial intelligence varies greatly. While most associate the term with ChatGPT or voice assistants such as Alexa or Siri, only a few are aware that TikTok, YouTube or Netflix, for example, also use artificial intelligence for personalized offers. It often turns out that the term is used to describe a black box that students can no longer logically explain how it works. However, there are often differing views on the trustworthiness of AI, not just in terms of prior knowledge. Due to the discussions shaped by the participants, the workshops always run differently, so that you have to remain flexible in order to ensure that the workshop is a success for all participants and that public relations work can be successfully combined with research."

*Michael Lenke*  
*PhD student in [project Ö](#)*

---

**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](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](#).