TRR 318 CONSTRUCTING EXPLAINABILITY

Erklärungen gemeinsam entwickeln | 02.2024

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 ▼

ChatGPT im Fokus & Einblicke in unsere Public Talks

Der TRR 318 hat zu einer Reihe von Online-Vorträgen zum Thema ChatGPT eingeladen. In diesen Public Talks diskutierten zwei Expert*innen die aktuellen Entwicklungen und Auswirkungen von ChatGPT-Technologien auf die Gesellschaft.

Im ersten Vortrag stellte Professorin **Dr. Isabel Steinhardt** die Faktoren vor, die die Nutzung von ChatGPT-Technologien beeinflussen. Sie beleuchtete, dass soziale Faktoren entscheidend sind und zu unterschiedlichen Nutzungsweisen, Akzeptanzgraden und Kompetenzen führen.

Der zweite Vortrag der Reihe wurde von Professor **Dr.-Ing. Hendrik Buschmeier** gehalten und thematisierte die Dynamik der Interaktion mit KI-Chatbots. Buschmeier untersuchte die komplexe

Beziehung zwischen Mensch und Technologie, insbesondere im Kontext generativer KI wie ChatGPT.
Vorträge ansehen von Steinhardt und Buschmeier

News

Neue Video-Reihe „Co-Constructing Science“

Der TRR 318 hat die neue Video-Reihe „Co-Constructing Science“ gestartet und die erste Episode wurde veröffentlicht. Die Forschung des TRR erfordert die Zusammenarbeit von unterschiedlichen Disziplinen wie Soziologie, Ökonomie und Informatik. Die Video-Reihe gibt daher Einblicke in die interdisziplinäre Forschungsarbeit.
Video ansehen

TRR 318 auf der XAI Conference 2024

Forschende des TRR 318 stellten diesen Sommer ihre neuesten Publikationen auf der World Conference on Explainable Artificial Intelligence (XAI) vor. Die Konferenz bietet eine Plattform für den Austausch von Wissen, Erfahrungen und Innovationen im Bereich der Erklärbaren Künstlichen Intelligenz. Teilprojekte **A03**, **A04** und **C05** nutzen die Gelegenheit, ihre Ergebnisse zu präsentieren und sich international zu vernetzen.
Weiterlesen

Neues Buch von TRR-Professor

Professor Dr. Tobias Matzner (Teilprojekte **B03**, **B06**) hat das Buch „Algorithmen: Technology, Culture, Politics“ veröffentlicht, das dem Verständnis von Algorithmen dient. Das Buch geht auf Themen wie Voreingenommenheit und Verantwortung ein und befasst sich mit selbstfahrenden Autos und GPT.
Weiterlesen

Interview mit TRR Projektleiterin im Magazin „Das kommt aus Bielefeld“

Professorin Dr.-Ing. Britta Wrede (Teilprojekte **A03**, **Ö**) erklärt im Magazin „Das kommt aus Bielefeld“ in der Ausgabe *„Bielefeld INTELLIGENT! Wirtschaft verändert!*“, wie wichtig transparente und erklärbare KI ist. Damit verdeutlicht sie unter anderem die Bedeutung der Forschung des TRR 318.
Weiterlesen

TRR-Sprecherin zu Gast im Podcast

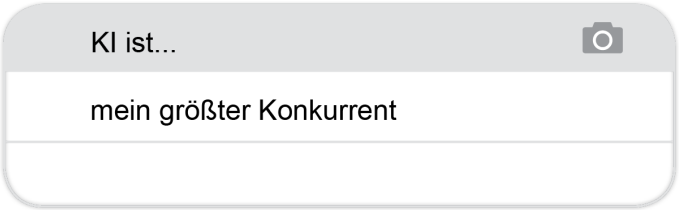
In der aktuellen Folge des Podcasts „Autonomie & Algorithmen“ von Dr. Christiane Attig und TRR-Assoziierten Jun.-Prof. Dr. Benjamin Paaßen gibt Prof. Dr. Katharina Rohlfing, Sprecherin des TRR 318, Einblicke in die Forschung des Transregios. Sie erklärt die interaktive Perspektive auf das Thema „Explaining“, die im Zentrum der Arbeit des TRR 318 steht.

Podcast anhören

Save the Date

Die 3. TRR 318 Konferenz “Contextualizing Explanations“ findet am 17. und 18. Juni 2025 in Bielefeld statt. Termin schon jetzt eintragen - weitere Details folgen!

Weitere News



„Die Diskussion, dass Menschen sich regelmäßig von Technologien (Computer, Roboter, KI) bedroht fühlen, ist eigentlich ganz interessant, denn mit jeder neueren Technologie scheinen wir uns nochmal versichern zu müssen, was uns besonders macht. Spannend, dass wir das nicht wissen, sondern immer wieder neu glauben, es erklären und rechtfertigen zu müssen, z.B. indem wir sagen ‚aber diese Technik kann nicht kreativ sein‘, ‚aber diese Technik kann nicht fühlen‘ oder ‚aber diese Technik hat zu wenig Wissen über die Welt‘.
Besonders auffällig ist das an Arbeitsplätzen, an denen Menschen schon zuvor befürchtet haben, durch alle möglichen Technologien ersetzt zu werden, weil Technik einzelne Aufgaben oder manchmal auch ganze Jobs ersetzt.

Faktisch ist es allerdings so, dass sich alle Arbeitsplätze ständig verändern und auch neue Berufsfelder entstehen. Beispielsweise hat bei der Einführung der Computer niemand daran gedacht, dass mal eine große Menge Menschen als Appentwickler arbeiten wird. Es ist sicher, dass KI die Art, wie wir arbeiten verändert. Vermutlich wird sich in vielen Fällen auch ändern, was wir tun, allerdings eher in der Aufteilung zwischen Menschen und KI und nicht so, dass Menschen und KI bei genau gleichen Aufgaben in Konkurrenz zueinander stehen.“

Prof. Dr. Kirsten Thommes
*Projektleiterin der Teilprojekte **A03** und **C02***

Veröffentlicht

Forschungsartikel: Can AI explain AI? Interactive Co-Construction of Explanations among Human and Artificial Agents
Die Studie untersucht, wie fortgeschrittene KI genutzt werden kann, um Menschen das Verständnis komplexer KI-Systeme zu erleichtern. Die Forschenden analysieren, wie Nutzer*innen gemeinsam mit einer KI Erklärungen konstruieren, anstatt diese passiv zu empfangen. Dabei spielen die Interaktionsdynamik, die Benutzeroberfläche und die Erwartungen der Nutzer*innen eine entscheidende Rolle für den Erfolg der Erklärprozesse.
Weiterlesen

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 Endnutzer*innen bleibt jedoch begrenzt. Die Forschung zeigt, dass ein geführter Erklärungsansatz das Vertrauen der Nutzer*innen mehr fördert als eine transparente Strategie, und manchmal kann der Verzicht auf Erklärungen paradoxerweise zu einem höheren Vertrauen der Nutzer*innen führen.
Weiterlesen

Forschungsartikel: Effects of Task Difficulty on Visual Processing Speed
Das Paper untersucht, wie erhöhte Aufgabenschwierigkeit die Geschwindigkeit und Genauigkeit der visuellen Verarbeitung beeinflusst. Mit einem Modell basierend auf der Theorie der visuellen Aufmerksamkeit wird analysiert, wie sich dies auf die Aufmerksamkeitskapazität auswirkt und welche praktischen Implikationen dies für Experimente mit begrenzter Versuchszahl hat.
Weiterlesen

Medientipps

YouTube Video „Was macht fit für KI im Job und wie erleben Datenarbeiter in Indien den KI-Boom?“
Das Video des Kanals "DW Deutsch" zeigt, wie künstliche Intelligenz Berufe verändert, einige überflüssig macht und neue schafft. Es wird erläutert, welche Fähigkeiten notwendig sind, um mit dem Wandel Schritt zu halten, und wie Datenarbeiter in Indien den Aufschwung der KI erleben.
Video ansehen

YouTube Video „NEW GPT-4o: My Mind is Blown“
In dem Video "NEW GPT-4o: My Mind is Blown" geht es darum, wie die neueste Version von GPT-4o beeindruckende Fortschritte in der künstlichen Intelligenz demonstriert und den Zuschauer mit ihren Fähigkeiten verblüfft.
Video ansehen

Was habe ich gelernt?

„Im Rahmen meiner Arbeit beschäftige ich mich intensiv mit allen zugänglichen neuen KI-Technologien, von fortgeschrittenen großen Sprachmodellen (Large Language Models) bis hin zu den neuesten Musik- und Bildgeneratoren. Ich habe die Vollversion von ChatGPT sofort gekauft, als sie verfügbar war – und nutze sie fast täglich. Die ersten Interaktionen waren spielerisch und experimentell: Was kann das System gut? Wo sind die Grenzen? Kann es wissenschaftliche Texte auf demselben Niveau schreiben, wie ich? Nein. Schlechte Gedichte, Kochtipps, Textfeedback, rudimentäre Programmierung? Ja. Brainstorming? Auf jeden Fall!

Im Laufe der Zeit wurde mein Nutzungsverhalten von ChatGPT konstanter. Ich behandle es jetzt wie einen außerirdischen Assistenten. Es erzeugt nachweislich kompetente Texte und verrückten Unsinn, wenn auch nicht zu gleichen Teilen. Die Interaktion mit ChatGPT ist ein iterativer, reflexiver Prozess. Es geht nicht darum, nach einer anfänglichen Anfrage ein perfektes Ergebnis zu erzielen, sondern darum, den sich entfaltenden Dialog zu steuern. In einem kürzlich erschienenen Artikel wurde ChatGPT als 'bullshit' bezeichnet. Dem kann ich nicht widersprechen, aber das ist nicht das ganze Bild. Der effektive Einsatz von ChatGPT erfordert die Steuerung einer ungleichen

Zusammenarbeit, die systematische Bewertung und Überprüfung der Antworten, die Bereitstellung klarer Rückmeldungen und die Delegation der Aufgaben, die das System tatsächlich gut ausführen kann.“

Nils Klowait
Doktorand im Teilprojekt Ö

TRR digital

Oder direkt per Mail mit Fragen oder Feedback an uns.

Newsletter abonnieren.

Top: german version ▲
Footer: Impressum ▼

Developing explanations together | 02.2024

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!

ChatGPT in Focus & Insights from our Public Talks

The TRR 318 hosted a series of public talks on the topic of ChatGPT. In these public talks, two experts discussed the current developments and impact of ChatGPT technologies on society.

In the first talk Professor **Dr. Isabel Steinhardt** presented the factors that influence the use of ChatGPT technologies. She highlighted that social factors are decisive and lead to different ways of use, degrees of acceptance and competencies. Using the concept of the digital divide, she showed that unequal access and unequal use of technologies ultimately reinforce social differences. The second lecture in the series was given by Professor **Dr.-Ing. Hendrik Buschmeier** and focused on the dynamics of the interaction with AI chatbots. Buschmeier examined the complex relationship between humans and technology, particularly in the context of generative AI such as ChatGPT.

Watch Steinhardt and Buschmeier´s talks

News

New video series “Co-Constructing Science“

The TRR 318 has launched the new video series “Co-Constructing Science” and the first episode has been published. The TRR's research requires collaboration between different disciplines such as sociology, economics and computer science. The video series therefore provides insights into interdisciplinary research work.
Watch video

TRR 318 at the XAI Conference 2024

Researchers from TRR 318 presented their latest publications at the World Conference on Explainable Artificial Intelligence (XAI) this summer. The conference provides a platform for the exchange of knowledge, experiences and innovations in the field of Explainable Artificial Intelligence. Subprojects **A03**, **A04** and **C05** take the opportunity to present their results and network internationally.
Read more

New Book by TRR Professor

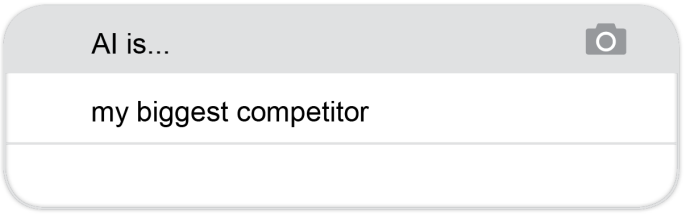
Professor Dr. Tobias Matzner (subprojects **B03**, **B06**) has published the book “Algorithms: Technology, Culture, Politics”, which serves to understand algorithms. The book addresses topics such as bias and responsibility and deals self-driving cars and GPT.

Read more

Save the Date

The 3rd TRR 318 Conference “Contextualizing Explanations” will take place on 17 and 18 June 2025 in Bielefeld. Save the date - more details coming soon!

More News



“The discussion that people regularly feel threatened by technologies (computers, robots, AI) is actually quite interesting, because with every new technology we seem to have to reassure ourselves about what makes us special. It's fascinating that we don't know this, and instead believe we have to explain and justify it again and again, for example by saying ‘but this technology can't be creative‘, ‘but this technology can't feel‘ or ‘but this technology has too little knowledge about the world‘.

This is particularly noticeable in workplaces where people have already feared being replaced by all kinds of technologies, because technology is replacing individual tasks or sometimes even entire jobs. The fact is, however, that all jobs are constantly changing and new professional fields are emerging.

For example, when computers were introduced, no one thought that a large number of people would be working as app developers. It is certain that AI will change the way we work. In many cases, what we do will probably also change, but more in terms of the division between humans and AI and not in such a way that humans and AI are in competition with each other for exactly the same tasks.“

Prof. Dr. Kirsten Thommes
*Project Leader of subproject **A03** and **C02***

Published

Research Paper: Can AI explain AI? Interactive Co-Construction of Explanations among Human and Artificial Agents
The study investigates how advanced AI can be used to facilitate the understanding of complex AI systems. The researchers analyze how users co-construct explanations together with an AI instead of passively receiving them. The interaction dynamics, the user interface and the user's expectations play a decisive role in the success of the explanation processes.
Read more

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: Effects of Task Difficulty on Visual Processing Speed
The paper investigates how increased task difficulty affects the speed and accuracy of visual processing. A model based on the theory of visual attention is used to analyze how this affects attentional capacity and what practical implications this has for experiments with a limited number of trials.
Read more

What have I learned?

”As part of my work, I immerse myself fully in all accessible new AI technologies, from advanced large language models to emerging music and image generators. I purchased the full version of ChatGPT as soon as it was available - and have been engaging with it almost daily. Early interactions were playful and probing: What can it do well? What are its limits? Can it write papers like me? No. Bad poems, cooking advice, text feedback, rudimentary programming? Yes. Brainstorming? Definitely!
Over time, my use of ChatGPT stabilized. I now treat it like an alien assistant. It verifiably generates competent text and mad nonsense, though not in equal proportion. ChatGPT interaction is an iterative, reflexive process. It's not about getting a perfect output after an initial query, but about managing the unfolding dialogue. A recent paper called ChatGPT 'bullshit'. I don't disagree, but that's not the full picture. Effective use of ChatGPT involves overseeing an unequal collaboration, systematically assessing and verifying its responses, providing clear feedback, and delegating tasks it is capable of performing.”

Nils Klowait
PhD student in subproject Ö

TRR digital

Or message us directly for questions or feedback.

Subscribe to the newsletter for free.

UNIVERSITÄT PADERBORN
UNIVERSITÄT BIELEFELD
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