Isabel Feustel

Ulm University
Institute of Communications Engineering
Albert-Einstein-Allee 43
89081 Ulm

isabel.feustel@uni-ulm.de
www.uni-ulm.de/nt/staff/
research-assistants/isabel-feustel/

1 Research interests

My research is focused on the field of **explainable AI (XAI)**, which aims to address the challenge of providing transparency to AI systems. I am particularly focused on the development of dialogue systems that enable **natural interaction with explanations**. By employing **computational argumentation** approaches, my objective is to create methods that facilitate meaningful dialogue between users and AI systems, allowing for a greater understanding of the systems' reasoning processes.

1.1 Enabling XAI explanations through dialogue

In recent years, the need for transparency in AI systems has significantly increased, leading to the growing popularity of the field of explainable AI (XAI) (Das and Rad, 2020). Ensuring that AI systems are understandable to users is crucial for building trust and facilitating effective use (Schmidt et al., 2020). One promising approach to achieving this is through dialogue systems, which can enable more dynamic and interactive explanations (Sokol and Flach, 2020).

Dialogue systems offer several advantages for the provision of explanations. These include the ability to segment information into manageable parts, thereby facilitating the comprehension of complex concepts; the capacity to elicit questions based on the specific needs of the user, which results in a more personalized and relevant interaction; and the capability to adapt the system's responses to align with the user's knowledge level and language proficiency, which enhances comprehension and satisfaction.

However, many existing XAI methods are non-conversational, offering explanations that are challenging for non-expert users to comprehend. Current conversational approaches in XAI like Slack et al. (2023), Shen et al. (2023) or Feldhus et al. (2023) often rely on basic question-answering systems and lack sophisticated dialogue management capabilities. This limitation neglects the importance of context in maintaining coherent and meaningful interactions. In order to address these issues, we proposed a generic dialogue architecture that integrates XAI explanations into a dialogue system (Feustel et al., 2023). Subsequently, we implemented a prototype based on this architecture.

Recognizing that effective explanations often require more than just model-specific details, we incorporated a knowledge module containing domain-specific information. This module is essential for providing comprehensive reasoning about the AI's domain, thereby facilitating a more profound comprehension of the foundation of the underlying process.

1.2 Integrating Domain Knowledge

The incorporation of domain expertise prompted the need to ascertain an effective methodology for integrating this knowledge into a dialogue system and establishing a connection with XAI explanations. The proximity of the areas of argumentation and XAI presents an opportunity for exploration, as arguments and explanations share comparable characteristics (Vassiliades et al., 2021). We determined that computational argumentation offers a suitable framework for representing domain facts, as it allows for structured and logical presentations of information.

Utilizing our expertise in argumentative dialogue, we determined that argumentative tree structures could be readily adapted to effectively address this integration challenge (Feustel et al., 2024). We extended our prototype system to include domain specific arguments and conducted a small study to evaluate the system's effectiveness. The results indicated positive trends, suggesting that integrating domain knowledge into the dialogue system has a positive effect on the dialogue.

1.3 Future Directions

In future research we want to explore several key areas to enhance the capabilities of the explanatory dialogue systems.

Firstly, we aim to improve the Natural Language Understanding (NLU) to achieve a more generic understanding of explanation requests, as we observed a high error rate in the current system that was NLU-related, resulting in users not being understood correctly. This involves developing advanced models capable of accurately interpreting and processing a wide range of user queries, re-

The 20th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems, pages 78–80

gardless of the specific wording or context.

Additionally, we plan to advance Natural Language Generation (NLG) techniques. Currently, our and other XAI systems rely on template-based system responses, which can result in rigid responses. By exploring more sophisticated NLG methods, such as those powered by large language models, we aim to generate more fluid and contextually appropriate responses. This improvement would also include the ability to paraphrase arguments to better fit the dialogue context, thereby enhancing the coherence and relevance of the information provided to users.

Another important area of focus is the annotation of arguments to enable better selection for specific user requests. By refining the way arguments are annotated and categorized, dialogue systems can more effectively retrieve and present the most pertinent information based on the user's needs. This involves developing detailed and nuanced annotation schemas that capture the essential qualities of arguments, ensuring that the system can make informed decisions about which arguments to present in various contexts.

By focusing on these improvements, we posit that significant advancements can be made towards more sophisticated, transparent, and user-centric dialogue systems.

2 Spoken dialogue system (SDS) research

I believe that in the next 5 to 10 years, the field of dialogue research is expected to see significant advancements in creating more flexible and natural dialogue systems. These systems will be capable of adapting to individual user styles, making interactions more personalized and effective. We will also see the emergence of multilingual and culturally adaptable systems, which can truly focus on users from diverse backgrounds. This will foster global communication and accessibility. Moreover, there will be renewed discussions on human-like systems, exploring the ethical and social implications of developing systems that closely mimics human behavior.

With the integration of large language models (LLMs), there may be a fundamental rethinking of traditional dialogue system frameworks, leading to more intuitive and seamless conversational experiences. We need to think about how LLMs can be integrated into traditional dialogue system architectures to leverage their full potential. However, we also need to be aware of the limitations they bring, such as biases in training data and the potential for generating misleading or inappropriate content.

Additionally, I see a future with more open domain dialogues, allowing users to engage in a wider variety of topics without the constraints of pre-defined domains. I think these open domain applications might function as microservices, where a single speech interface processes the intent and directs the user to the appropriate application to fulfill their request. Virtual agents will increase in prevalence, necessitating a high need for natural speech interaction to ensure user satisfaction and effectiveness across various tasks and applications. Improved assistant systems will further support users in various tasks, from simple queries to complex problem-solving, enhancing productivity and user satisfaction across different applications.

3 Suggested topics for discussion

- **Personalization and User Modelling:** Best practises for tailoring dialogue to individual users. What can be personalized and what should not be personalized? Which aspects of a user can already be modelled and how can we model more complex aspects? E.g. Mental Model
- **Evaluation of Dialogue:** How can we evaluate non-task-oriented dialogues? How can we engage participants to interact with the system without influencing the results?
- **Error-Communication:** There are various aspects where a (dialogue) system can fail (e.g. wrong AI prediction, wrong intent classification, ..). Can we somehow track these failures? How should systems react if the users notices some wrong behavior? Can we implement feedback loops to optimize the dialogue policy?

References

Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* .

Nils Feldhus, Qianli Wang, Tatiana Anikina, Sahil Chopra, Cennet Oguz, and Sebastian Möller. 2023. InterroLang: Exploring NLP models and datasets through dialogue-based explanations. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics, Singapore, pages 5399–5421. https://doi.org/10.18653/v1/2023.findings-emnlp.359.

Isabel Feustel, Niklas Rach, Wolfgang Minker, and Stefan Ultes. 2023. Towards interactive explanations of machine learning methods through dialogue systems. *The 13th International Workshop on Spoken Dialogue Systems Technolog* .

Isabel Feustel, Niklas Rach, Wolfgang Minker, and Stefan Ultes. 2024. Enhancing model transparency: A dialogue system approach to xai with domain knowledge. *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue* .

Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29(4):260–278.

Hua Shen, Chieh-Yang Huang, Tongshuang Wu, and Ting-Hao Kenneth Huang. 2023. ConvXAI: Delivering heterogeneous AI explanations via conversations to support human-AI scientific writing. In *Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, New York, NY, USA, CSCW '23 Companion, page 384–387. https://doi.org/10.1145/3584931.3607492.

Dylan Slack, Satyapriya Krishna, Himabindu Lakkaraju, and Sameer Singh. 2023. Explaining machine learning models with interactive natural language conversations using TalkToModel. *Nature Machine Intelligence* https://doi.org/10.1038/s42256-023-00692-8.

Kacper Sokol and Peter Flach. 2020. One explanation does not fit all: The promise of interactive explanations for machine learning transparency. *KI-Künstliche Intelligenz* 34(2):235–250.

Alexandros Vassiliades, Nick Bassiliades, and Theodore Patkos. 2021. Argumentation and explainable artificial intelligence: a survey. *The Knowledge Engineering Review* 36:e5.

Biographical sketch

Isabel Feustel is a PhD student at Ulm University supervised by Dr. Dr. Wolfgang Minker and Dr. Stefan Ultes. She obtained a Master's Degree in Media Informatics at Ulm University in 2019, where her research focused on communication styles within dialogues. She began her PhD by exploring argumentative dialogues and focused her studies on explanatory dialogues.