

## Improving the Explainability of Autonomous and Intelligent Systems with Conversational Interfaces

Nico Hochgeschwender<sup>1</sup>

### 1 Abstract

Autonomous and intelligent systems such as self-driving cars, smart grids, advanced satellites or service robots are to a large extent *software-intensive systems* where increasingly advanced functionality is provided by novel learning, reasoning, planning and decision-making abilities. Those abilities are not only exploited by autonomous systems to achieve non-trivial goals, but also to autonomously adapt their course of action to the various and changing requirements induced, for example, by the systems task, its complex, cyber-physical environment or by varying user requests.

Even though autonomous and intelligent systems are successfully deployed in real-world environments they lack the ability to communicate and explain *why* and *how* they have made particular decisions. Therefore, we argue that every autonomous system needs to be *explainable*. To this end, an explainable system should be – at any point in time – capable to express and explain its internal decisions, functional abilities and limitations, and its past, current and potential future behavior to different stakeholders (e.g. developers, users, lawyers or certification agencies). It is important to emphasize that those explanations need to take into account the full software stack of autonomous systems which poses, among others, several representational challenges for platforms, frameworks and infrastructures as one needs to investigate which knowledge is required for the explanations.

Implementing explainable systems is not only crucial for users to develop trust, but also of paramount importance for developers, insurers and lawyers to find out what and why something went wrong in the case of, for example, an accident or failure.

In order to deal with the varying explainability requirements imposed by different stakeholders we propose a user-centered approach to design and develop advanced *user interfaces*. Thereby, our goal is to develop interfaces which are capable to satisfy stakeholder needs by providing configurable means for communication and explanation which make use of

---

<sup>1</sup> German Aerospace Center (DLR), Simulation and Software Technology, Intelligent and Distributed Systems, 51147 Cologne, Germany nico.hochgeschwender@dlr.de

concepts and terms from the stakeholders domain. To this end, we propose to exploit conversational interfaces (also known as chatbots) as a technology to improve the explainability of autonomous systems. Conversational interfaces offer a great opportunity to communicate with autonomous and intelligent systems in a convenient way using human languages.

The idea that chatbots can contribute to improving the transparency of autonomous systems is supported by our recent work described in [1] where we demonstrated how a conversational interface can enhance the task of visualizing and assessing complex software architectures as they occur, for example, in autonomous systems.

However, to fully exploit the advantages of conversational interfaces (e.g. accessibility) for improving the explainability of autonomous systems we need to better understand when, how and why users interact with chatbots for the sake of seeking explanations. We argue that conversational analysis techniques could serve as a tool to investigate and assess those requirements. For example, conversation analysis describes the interaction with a chatbot as sequentially organized actions. The different actions and interactional practices such as questioning, devices (e.g. upper vs. lower case writing) and turn formats (combination of practices and devices) need to be properly investigated from an explanation point of view. Persons seeking explanations usually show their lack of information by asking questions. Therefore, a conversational interface should be capable not only to deal with different question types (e.g., known-answer and unknown-answer questions), but also to deal with varying epistemic stances (expectations of the knowledge of the other speaker). This is in particular important if a chatbot should be used by different stakeholders with different background knowledge and possibly varying concepts and terms.

We argue that the above mentioned considerations are important for developers to design useful conversational interfaces which go beyond the intent and task-based conversational interfaces currently available.

## References

- [1] Stefan Bieliauskas and Andreas Schreiber. *A Conversational User Interface for Software Visualization*. In *Proceedings of the IEEE Working Conference on Software Visualization (VISSOFT)*. 2017.