Explainability as a Non-Functional Requirement

Maximilian A. Köhl, Dimitri Bohlender, Kevin Baum, Markus Langer, Daniel Oster and Timo Speith

SWE 522 FALL 2021

GROUP 6

YAĞIZ ALKAN

MURAT MERT ŞENTÜRK

ÖMER FARUK ÇEVİK

Content

- Definitions
  - What is a non-functional requirement?
  - What is explainability ?
  - Explainable systems
  - Motivation
  - Elicitation
  - Specification
  - Verification
- About the paper
  - Authors
  - Paper statistics

What is a non-functional requirement?

- Recall from our course:

- «Constrain how such services (the functional requirements) should be provided.
  Quality requirements: safety, security, accuracy, time/space performance, usability, …

Others: compliance, architectural, development requirements

Non-functional requirements that do not constrain how the software should satisfy its functional requirements but how it should be developed.
[Bogazici University SWE522 course, SWE522-21F-02-fundamentals.pdf, slides 27-30 »

What is an explanation?

As mentioned by Köhl et al explanation is ‘answer to certain questions, in particular Why queasiton’.

Because it contains the form and attributes defined by technical accounts, a response to a question can constitute an explanation.

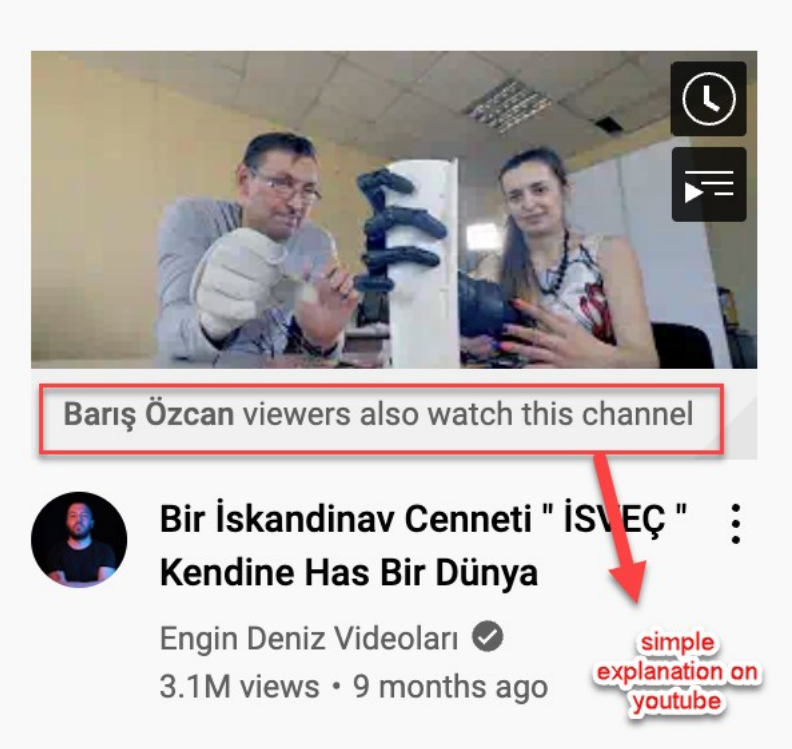
We need an explanation concept that addresses a certain type of stakeholder for our purposes—an explanation for an engineer may not explain anything to a user.

That is, we require a concept that allows for generalization and abstraction from a specific person.

It's also important to consider the context in which an explanation is given. For starters, it has an impact on more than just what has to be conveyed.

What is explainability - 1?

- Let’s start with a basic concrete example



- The image on the right is an example of explanation.
- It is for end users.
- It just explains why a content is suggested to the end user.
- Note that this explanation might be enough just for end users.
- That does not mean that this explanation is suitable for each stakeholder.
- Developers expect detailed code for technical explainability.
- Inspectors expect statistical explanations.

We just started to dig deeper on explainability as a non-functional requirement!

What is explainability - 2?

- Explanation in a software depends on the 5Ws
- What
- When
- Who
- Where
- Why
- If one of the Ws changes, the clarification probably needs to be changed. Because each user type or each stakeholder type needs a distinct description.

Explainable Systems

- People desire sufficient understanding of the systems which they are interacting. For example, when a user mouse over a text box, the tooltip can provide a statement about the text box.
- Descriptions improve usability, help in locating sources of error, and can minimize the chance for human error.
- A lack of explainability, on the other hand, not only gives rise to various moral, social, and legal problems. It further fuels distrust, diminishes user acceptance and satisfaction, and inhibits the adoption of new technologies
- Moreover, European Union debated about a general right to explanation which is partly applied in certain regulations.
- Design decisions that impact the explainability of a system must not be taken by the developer implicitly, but explicitly specified as part of the design process.

Motivation

- There is no agreed definition of explainability in software development.
- There is no metrics to evaluate a system's explainability performance.
- No explicit specification of explainability to take them into account during development.
- Especially in artificial intelligence, a black-box model, explainability is a hot topic.
- It is `art`, not `science`. Tailored approach is required.
- Sometimes even domain experts and system engineers struggle to understand certain aspects of a system.
- While explainability is the rage, the concept itself, and in which context which techniques are appropriate, remains under-specified.
- This paper's aim is elicitation, specification, and verification of explainablity as a Non-Functional Requirement

Explanation

- We need specified explanation or interpretation for a targeted stakeholder group. Explanation for an engineer may not explain anything to a user. The author proposes two definition:
- *Definition 1 (Explanation For): E* is an *explanation* of explanandum *X* with respect to aspect *Y* for target group *G*, in context *C*, if and only if the processing of *E* in context *C* by any representative1 *R* of *G* makes *R* understand *X* with respect to *Y*
- *Definition 2 (Explainable System):* A system *S* is *explainable* by means *M* with respect to aspect *Y* of an explanandum2 *X*,for target group *G* in context *C*, if and only if *M* is able to produce an *E* in context *C* such that *E* is an explanation of *X* with respect to *Y* , for *G* in *C*.

Explainability Requirements

TG: What are the relevant target groups

E: What are the explananda, events or decisions?

A: Which aspects of the explananda must be explained to which target group, e.g., why is a decision justified?

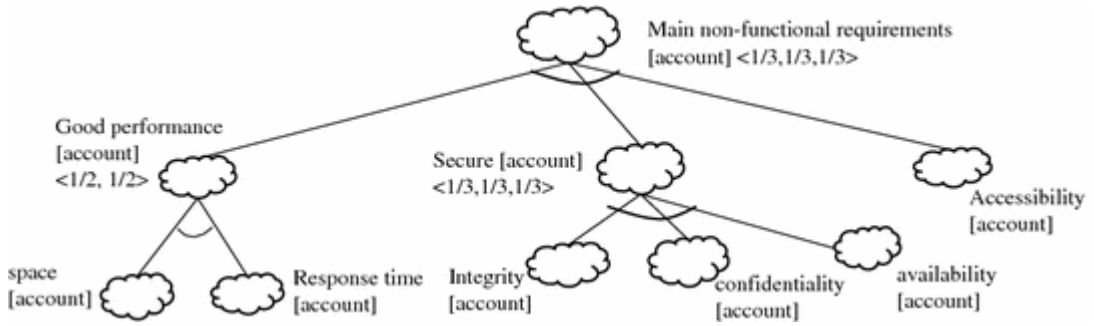
C: In which context may an aspect Y need explanation, and what are the implied constraints? For example, explanations might have to be aural in a driving situation.

Explainability Requirement:

A system must be explainable for target group **TG** in context **C** with respect to aspect **A** of explanandum **E**.

Specification of Explanation

- The authors propose Softgoal Interdependency Graphs



https://link.springer.com/chapter/10.1007/978-3-319-24315-3_20

Decomposition and elicitation of sub-softgoals lies at the heart of building SIGs.

Verification of Explanation

- Explainability of the overall system is then satisficed by satisficing the resulting explainability sub-softgoals.
- The authors of the paper suggest selecting sample of a particular stakeholder with different characteristics such as selecting lawyers different age, gender, experience with a given problem, and provide them with explanations. Then, collecting feedbacks from the participants to check whether or not they understood the explanation in desired way.

  It seems necessary to gain deeper insights into the representatives' processing of explanations. For instance, one could use the think-aloud technique trying to understand how people perceive a given explanation. After processing the explanation, the representatives could try to use self-explanation to answer their own questions based on the explanation.

-

Side effects of explanation

- Authors suggest to understand explainability requirements as *Non-Functional Requirements* (NFRs) of a specific kind that must be *satisficed* rather than satisfied. Because In general, a trade-off between the degree of explainability and other goals must be made.
- Conceptually, requiring a system to be explainable does not entail a specific function that the system must be capable of performing, but rather constrains how it may be implemented.
- The explanation as a NFR may conflict with requirements like privacy, e.g., explanations may leak personal information about the applicants.
- Explainability requirements may conflict with other softgoals such as performance, development cost, precision, or security. A less explainable system may be cheaper to build or could offer a higher performance.

Authors

- Saarland University, Saarbrücken, Germany
  - Maximilian A. Köhl
  - Kevin Baum
  - Markus Langer
  - Daniel Oster
  - Timo Speith
- RWTH Aachen University, Aachen, Germany
  - Dimitri Bohlender

Authors

- Statistics

	Since 2016		
	Citations	h-index	i-10 index
Maximilian A. Köhl	74	6	2
Markus Langer	531	12	13
Kevin Baum	110	6	3
Daniel Oster	71	4	3
Timo Speith	82	4	3

Paper Statistics

- Cited by 26 papers
- Has 46 references
- The publish date of the references ranges from 1977 to 2019

Questions

- We would be glad to answer your questions