WASP Cloud

Máté Földiák

August 2024

1 Introduction

The emergence and rapid development of distributed autonomous systems in safety-critical domains (such as avionics) necessitate to provide assurance that functional and extra-functional requirements are satisfied. The analysis workflows of extra-functional properties, such as reliability, availability or performance require to synthesize a large variety of feasible design candidates along different priorities or design principles. These design candidates are then subject to thorough multi-disciplinary analysis and optimization in order to find the final design as an appropriate trade-off between potentially contradicting requirements. In this early design phase, systems engineers need to deal with uncertainties of different nature. For example, certain components may still be unknown, or underspecified or extra-functional parameters may be inaccurate (epistemic uncertainty). Moreover, the stochastic behaviour of components may exhibit certain randomness at runtime (aleatoric uncertainty). If not incorporated appropriately, such uncertainties may result in design flaws and suboptimal solutions. The main objective of the PhD project is to provide advanced modelling techniques in the presence of various uncertainties including support for model queries, transformations, or design space exploration. Such techniques would lead to more effective decision-making and design of distributed autonomous cyber-physical systems.

2 Main lectures

Quality Assurance and Testing in SE, QA for ML
This topics fits quite well with my research, as both focuses on the quality of the system, but my topic focuses more on physical systems in early design stage, where there are lots of requirements but very little code.

Here it I kind of missed that verification and validation can be more than validating if a piece of code meets the requirements and needs. It was mentioned that requirements alone can lead to problems, like requirement compatibility, as there is no guarantee that two requirement can be satisfied at the same time, or that the requirements cover the problem to a sufficient extent. But it is also worth mentioning, that there are standards and best practices on how to express requirements (phraseology) and even domain specific requirement modeling languages, mixing formal semantics with natural language.

Also, "in an ideal case the oracle should be generated (from model)" is interesting given that in some extreme cases, the development and testing is isolated as much as possible to avoid issues from using the same artifacts. Like of your model contains an error, and you use it for both development and testing, then the tests will not reveal the error.

(And we have some companies, that have the mentality to claim no responsibility in their contracts, with which the financial effects of failure can be significantly reduced.)

Main lecture 2: Requirements Engineering (for ML)
Here I have one general observation, which is how the goals in engineering and legislation, such as safety and explainability, is treated as very important, but in public communication these are abstracted to a meaningless and generic statement, and ToS contracts all liability is avoided. For me it reads as "rest assured, we don't give a dime neither for your safety nor your rights".

3 Related guest lectures

I was aware that designing around human behaviour and capabilities is important when designing critical systems. (I heard a case when an emergency stop button was designed to activate on a long press, but users panicked and started pushing it rapidly, thus not activating it.) But it was new how human relation dynamics can influence sofrtware development in the context of a critical system.

The second is the mentality towards AI, as in some cases it is indeed capable of solving challenging tasks, and in many places it is quicly extrapolated to every problem, while cases of failures are swept under the rug or downplayed in relevance. It is also interesting, or at least for me, how the cost of advancement is rarely discussed. Many AI project is funded and some of them does yield result, but there are also signs of diminishing return. Especially with LLM. (And there is also a race to introduce minor improvements and tools, such as AI search, leading to Google recommending objectively false and/or unhealthy activities.)

4 Paper 1: Welcome Your New AI Teammate: On Safety Analysis by Leashing Large Language Models

Safety engineering is a core process in designing cyber-physical systems that interacts with humans. As such, analysing safety scenarios is a critical part for both OEMs (Original Equipment Manufacturer) and regulatory bodies. However changes in the safety requirements can lead to excessive work in ensuring that the system complies with the new goals. The paper presents a pipeline that integrates LLMs into a specific approach to safety evaluation, namely Hazard analysis and risk assessment. This work is relevant as LLMs show signs of limitations in many specific and expert-level tasks, and safety analysis is one, where such mistakes are very costly.

I selected this paper as it explores the possibility of augmenting critical design tasks with AI systems in an industrial context. There were two key points I aimed to observe, (i) what is the underlying approach for safety by the associated company, and (ii) how the paper evaluates their findings. This is to guide both how I should present my results (if I agree with how they did it), or what type of change in direction should I endorse.

Here the discussion is how to assist engineers in safety critical context, which is something I already worked on, but in an academic setting. As with any type of domain, safety engineering contains both repetitive or algorithmic, and creative tasks. As such, it is reasonable to investigate if an AI system can augment or replace any of that. The selected paper gives a mixed impression on what the goal actually is. While the high level goal seems to be the aid in creative tasks, the results and actual application is on the baseline level of the safety analysis process. Comparing this to my own research, the targeted the more algorithmic part, where a natural language description is mapped to a formal language. Many, including the authors of the paper, covertly imply that AI will or can replace humans in creative tasks, however both my experience and my take on this paper suggest otherwise. Like here, if AI is good in creative tasks, then why are you only using it to create the *initial* point for an engineers' brainstorming session?

As mentioned before, my side-project included a mapping from natural language to a formal language. There we assumed that the engineers already have the properties they want to evaluate over a system model, and the task is to map it to the modeling language. This task is definitely not on the creative side of thing, but it involves some place for interpreting descriptions, which should not be happening if engineers are doing a proper job. Based on the selected paper, the idea would be to implement a chain of AI calls, that can refine a high level requirement to smaller, simpler ones, that are equal in function to the high-level one. This would be in line with our results, as the mapping showed significant difficulties with complex formal descriptions. But for now, the focus is populating the dataset with more complex formal queries, so that the fine-tuning can be more efficient.

5 Paper 2: Approach for Argumenting Safety on Basis of an Operational Design Domain

This paper is again focuses on the safety aspects of systems with AI components. The general problem presented in the paper is that Autonomous AI systems operating in an open environment can encounter many things, some of which can be novel to the system. However, contrary to human operators, an AI system cannot recognize the situation as unknown, and it can result in faulty operation. The paper addresses this issue by defining a process that ensures the correct and complete development of the Operational Design Domain. The ODD is the domain the AI system is able to recognize and operate in it as specified by requirements. The problem is that this domain may be different from the Operational Domain, which is where the AI will operate. The proposed solution aggregates previous experience, data, standards, expert opinions, and many other things from the target and similar domains.

While this paper is not tightly connected to my research in technical terms, it provides valuable insight into the (ideal) process of designing safe AI systems. A appreciate that it takes a refined stand on how development and testing should be done, and provide a sound approach, instead of the ”tweak it until it works” approach. The paper focuses on how limitations can be addressed (mostly) before release of the product, which should be mandatory. It also has a relevant aspect, that it focuses more on the development process, similarly to my work, which aims to make the development of safety critical systems easier by providing tools to consider safety aspects earlier. Advocating for inputs and collaboration with experts from different domains is something that, in my amateur opinion, is something we tend to move away from, yet it can be very insightful.

An other characteristic that is nice, is they present a proposal, that is approached from many side of the problem. It includes a contextual aspect, as what is the problem from the perspective of the system, how that problem manifests itself in the context of the operational system. The second thing, is the approach, that the process includes most of the system lifecycle, from creation and operation of other similar systems, the design on the current system, and post-release monitoring. For me, this approach is something to keep in mind, as (i) whatever I make, will not be flawless, thus any guidance on how to evaluate it from the most complete perspective is appreciated, and (ii) it encourages inputs from a wide range of sources, some are less relevant or easy to ignore on first glance. This approach can can serve as an example on how to avoid having tunnel vision on how to tackle a problem.