WASP Software Engineering Course Module 2024
Assignment

Matthias Wagner, Lund University

September 2024

1 Introduction - My Research

My research is towards continuous quality assurance and ML pipelines throughout a system's lifetime, inspired by the AI Act, a new regulation in the European Union with extraterritorial reach.

AI/ML systems are more and more on the rise, also in safety-critical systems, while quality assurance and trustworthiness still pose significant challenges in this field. This is one reason why the European Union aimed to regulate this new technology to achieve trustworthy and human-centric AI while mitigating potentially harmful effects. The result of this effort was the AI Act, which is in force since August 2024. It sets out extensive requirements for high-risk AI systems as well as for general-purpose AI models. Quality assurance research is needed to break down the Act's requirements into practical methods and tools.

My expected contributions are twofold. First, I'm contributing to an empirical foundation for quality assurance in the context of the AI Act. This includes close collaboration with industry partners and thus fosters co-learning between industry and academia. In the second stage, our aim is to develop artifacts for AI Act conformance. This involves methods and tools as part of an automated pipeline with MLOps being a popular term. Data Engineering will also be of high importance. These developed methods and tools should contribute towards trustworthy AI by design.

My research approach, especially for the second stage, will use the concept of automated MLOps pipelines as a foundation and be complemented by requirements engineering and testing. The goal is to achieve automated verification and validation in terms of the requirements set out by the AI Act. These requirements concern several aspects of an AI system and are aimed at the whole system's lifetime: risk and quality management system; record-keeping; data quality and governance; transparency; accuracy, robustness, and cybersecurity; human oversight; and technical documentation.

The first part of my research is focused on existing literature and reviewing it. For the artifact development, we plan to utilize the design-science approach. Industry collaboration is a priority here because they are the ones directly affected and thus actual case studies would be of high value for both industry and academia. This is especially relevant for the artifact validation.

Overall, almost all of the concepts presented in this lecture are very related to my research which is why this essay is a perfect opportunity to reflect on potential future directions.

2 Lecture Concepts

2.1 Robert's Lectures

2.1.1 ML Code as a small fraction of ML systems [3]

ML systems are often still experimental with companies developing ad-hoc solutions that just focus on the ML code without taking into account the whole system context. This can become especially critical for high-risk use cases. The new requirements set out by the AI Act cover the system holistically and can only be fulfilled when taking into account all components of the system. This is why this concept relates very well to my research.

2.1.2 AI/ML Software Testing

The AI Act entails several testing-related requirements. Providers are obliged to have techniques and procedures in place for the examination, testing, and validation of their ML system over the whole system lifetime. Moreover, tested and validated levels of accuracy, robustness, and cybersecurity have to be provided. Despite the unique challenges that ML systems entail in testing, the obligation to consider the whole system's lifetime for continuous compliance, also after deployment, adds another significant challenge. Often times organizations might not even be aware that an adaption of their current testing practices is needed to properly account for ML models.

2.2 Guest Lectures

2.2.1 SE4ML/SE4AI (Volvo)

The idea of utilizing Software Engineering principles for ML systems for a systematic approach of integrating ML components into software is very related to my research. That is because it is very hard to comply with a comprehensive regulation like the AI Act without following a systematic approach including different activities such as risk analysis, requirement engineering, data collection and validation, design, modeling, versioning, testing, integration, deployment, maintenance, etc. – as shown in the lecture. A very interesting concept that I plan to use for my research is MLOps, which focuses on streamlining the process of taking ML models to production and then maintaining and monitoring them.

2.2.2 Behavioral Software Engineering (SAAB)

In my opinion, the concept of Behavioral SE can help organizations drive the change process needed to include ML models in their current software. Cross-communication between different roles is commonly needed, for example, data engineers with software engineers, while also coordinating with dedicated ML engineers. Thus, taking into account the human in the loop when introducing new technologies sounds inevitable to some extent.

3 CAIN Papers

3.1 "A Meta-Summary of Challenges in Building Products with ML Components – Collecting Experiences from 4758+ Practitioners" [2]

This paper is a meta-summary of software engineering challenges that industry practitioners are facing when in integrating ML components into software products. From 341 initial papers, 50 ended up in the final selection with over 4758 industry practitioners involved. The results were grouped as follows: requirements engineering; architecture, design, and implementation; model development; data engineering; quality assurance; process; organization and teams.

The study identified a lack of AI literacy causing unrealistic expectations as well as vagueness in the specification of ML problems that makes the mapping of business goals to performance metrics difficult. Additional restrictions through regulatory requirements were also identified as challenging. Despite being difficult in many cases, transitioning from a model-centric to a pipeline-driven system-wide view was considered important. This stands contrary to the common practice of ad-hoc development of ML components without well-defined processes. Unfortunately, monitoring of deployed models is often an afterthought at best. Often times teams are faced with a limited engineering infrastructure that lacks proper support for the development of ML models. The same is true for proper data management tools that would be needed for validating and improving data quality and, for example, allowing for data versioning and provenance tracking. For quality assurance in general, standard processes and guidelines for aspects like fairness, security, and safety are often lacking. Besides monitoring, the testing of ML systems is also commonly neglected and often also difficult due to a lack of specifications. Acquiring teams with diverse skill sets and managing the resulting interdisciplinary collaboration was also identified as challenging.

This paper relates to my research by shedding light on the most common and pressing challenges faced in the industry related to the integration of ML components. This can be compared with the regulatory requirements set out by the AI Act and thus help to prioritize research towards aspects that are the most challenging/relevant to the industry. It helps me to adapt my research towards the requirements that are likely going to be most challenging to operationalize for companies.

It is easy to imagine how the core findings of this paper would benefit a larger AI-intensive software project when applied. Common pitfalls could be counteracted from the beginning, provided that enough resources are available for its implementation. My research aims to contribute to developing guidelines, methods, and tools that meet some of the requirements of the AI Act. The results of this study and my research aims therefore complement each other very well in the context of a large AI-intensive software project.

3.2 "What About the Data? A Mapping Study on Data Engineering for AI Systems" [1]

This paper is a mapping study on AI data engineering, a sub-field of AI engineering, including 25 out of 259 papers for the final selection. The study identifies which life cycle phases of the AI engineering life cycle are covered by the literature, shows proposed technical solutions, and highlights lessons learned. AI systems need high-quality data that needs to be collected and prepared before the AI model can be trained on it. At the same time, most organizations lack a proper data infrastructure.

The presented technical solutions range from synthetic data generation to data validation tools, and data processing frameworks. For example, data validation should include anomaly detection before being fed into an ML pipeline. There is also a proposal for a pipeline that allows to adapt to changes in the incoming data for the continuous training of deployed ML models. The mapping study also identified several data architectures proposed by the literature, for example, an end-to-end MLOps architecture that includes a separate data engineering zone. Another example is a reference architecture for automated data annotation.

The findings of this study resonate with my AI Act related quality assurance research as this new regulation sets out extensive requirements for the data quality and governance for high-risk AI systems. Requirements concern data management, including data acquisition, collection, analysis, labeling, storage, filtration, mining, aggregation, and retention. Moreover, quality criteria are demanded for the training, validation, and testing data. More covered aspects include bias detection and mitigation, the consideration of the geographic and contextual setting of the data, data gap and shortcomings identification, and data vetting for errors. The relevancy, representativeness, and completeness of data also have to be fostered.

Fulfilling these requirements without an adequate data engineering process in place will be next to impossible. Therefore, studies like this one are very valuable as they show state-of-the-art approaches that are already out there and point out areas that require more research.

When considering larger AI-intensive software projects in this context, we can see that my research, which is related to requirements coming from new regulatory requirements, can profit from the technical solutions, architectures, and best practices provided by this mapping study to contribute towards a legally compliant state-of-the-art AI data pipeline.

References

[1] Heck, P.: What About the Data? A Mapping Study on Data Engineering for AI Systems. In: 2024 IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI (CAIN). pp. 43–52 (Apr 2024)

[2] Nahar, N., Zhang, H., Lewis, G., Zhou, S., Kästner, C.: A Meta-Summary of Challenges in Building Products with ML Components – Collecting Experiences from 4758+ Practitioners. In: 2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN). pp. 171–183 (May 2023)

[3] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.F., Dennison, D.: Hidden Technical Debt in Machine Learning Systems. In: Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)