Software Engineering Assignment

1. Introduction

My PhD focuses on Machine Learning for Causal Inference, while this can be seen as removed from software engineering (SE) because we do not produce software directly this is not the whole picture. Indeed our work is intended to be used for decision making and cannot exist in isolation and thus should benefit from software engineering principles.

To explain this and relate SE my work to, I will explain my research area : Machine learning changed a lot of fields including clinical research and decision making in healthcare, two applications that are at the core of my work (I hold a PharmD). If AI and ML are associated with prediction or classification, those tools can be used to improve our understanding of causal relationships in complex contexts. Contrary to traditional statistical methods ML offers more flexibility to uncover causal relations in large and heterogeneous patient cohorts.

In healthcare, Machine learning leveraged Causal inference can overcome limitations of clinical trials (They often lack representativity of real patient) or real world assessment of drugs (Many confounding variables, treatment propensity). Causal inference can be used to evaluate the effect of a drug or treatment on a patient while filtering out confounders, this is crucial when randomized trials may be impossible or unethical.

Ml for Causal inference relies on different approaches such as causal tree or double machine learning. Those methods are used to model complex dependencies between variables and thus achieve a better understanding of treatment effect but also guide clinical decision or healthcare policies

2. Ideas from Roberts's lectures
    a. Behavioral Software Engineering

BSE combines human psychology and software engineering to better understand how developers and users interact with the product. This notion could be applied at different stages of my work:

- Improve collaboration across disciplines: In my space peoples with a lot of different backgrounds have to collaborate to achieve the desired goal and BSE can help identify behaviors that led to a good collaboration or help find pain points that might hinder progression-
- Improvement on the end user side: BSE can help build more intuitive and useful software for healthcare practitioners, ensuring a real world usefulness and not just an AI/ML algorithm with good accuracy that will never be used. There is also a great task to be done where BSE can help us, it's the acceptance of digital medicine by clinicians but more so by patients.

b. QA

QA on the software engineering side is also a major contributor to push my work from “the lab” to a real and safe product. QA ensures that software used for decision making are safe, reliable and performs as intended. By applying testing, verification and validation AQ can show whether our product is compliant with standards and expectations of the medical world. This minimizes error rates, can help improve clinical analysis and more broadly guarantee digital solution safety in health care.

3. Ideas from guest lecture

a. Software Engineering relevance changing with the integration of AI/ML

While not working in SE, we have seen the same questions arise in healthcare so I thought that It would be interesting to draw a parallel : Here also views are conflicted, on one side some see the emergence of AI in health care as a one fit all solution with better diagnosis, task automation all to the benefit of the patient. On the other side there are a lot of new problems that need to be solved, training needs to be implemented, data ownership is more than ever an issue and all of these questions might take some critically useful time away from clinicians.

b. Different area share the same issue when risk or human are involved or may be affected by the end product

Another parallel can be made from what I saw during the SAAB lecture : All applications and software that involves human safety should benefit from rigorous software engineering practices, and most can be transposed to different domains while still using the core principles of software engineering. So what is striking is that SE principles are transversal and that it goes beyond coding but encompasses methods to develop safe and robust solutions wherever they might be used.

4. Publications

a. Data Sovereignty for AI Pipelines: Lessons Learned from an Industrial Project at Mondragon Corporation[1]

i. Core ideas

The article explores the importance of data governance in AI pipelines, showing the difficulties related to data sharing across organizations. The principal hurdles identified are the reluctance to share data due to the fear of losing control and difficulties to guarantee safe access afterwards. The article proposes solutions like the use of data sovereignty technologies like the IDS connector to secure and govern data sharing, thus improving trust between collaborating organizations which is crucial for AI system engineering.

ii. How the paper relates to your own research

My work focuses on clinical research and decision making in healthcare and can directly benefit from data sovereignty principles discussed in the articles. In clinical research data sharing is a major issue. The implementation of data sovereignty practices can not only assure protection of patient data but also improve collaboration and help leverage the vast amount of data now available. The article findings can reinforce security while improving trust.

iii. How your research and its results would fit into a larger AI-intensive software project where one of the core ideas from the paper would benefit the project if applied

We are currently using a data space certified for sensitive data where we can collaborate with people within my organization (Astrazeneca) that falls aligned with what the article proposes but nothing yet has been done for data sharing across organizations. In my previous position at Astrazeneca France I worked on building a data warehouse directly connected to french insurance claims, so data sovereignty is an issue that I tend not to overlook.

iv. Discuss briefly how your research could be potentially adapted/changed to make AI engineering in the project based on the idea of the paper even better/easier

My research could be adapted to improve AI engineering in this project by integrating more control mechanisms into AI pipelines. For example, we could develop AI tools that can not only analyze data, but also continuously monitor the use of that data to ensure that it complies with data sovereignty terms. This would make AI systems more transparent and secure, facilitating collaboration between different entities in the healthcare field while respecting strict data protection requirements.

b. Trustworthy and Robust AI Deployment by Design: A framework to inject best practice support into AI deployment pipelines [2]

i. Core ideas

The article focuses on challenges of deploying AI/ML systems and reviews tools and platforms available to facilitate that process. The main findings is that while there are numerous tools available they are underused or too complex for small businesses to use. The article highlights the importance of adopting good practices for software engineering to ensure high quality implementation and deployment, this is paramount to avoid sub-optimal performances and negative impact and risks for the community.

ii. How the paper relates to your own research

My research can draw from the concept explored in the article, especially on the necessity to deploy reliable AI systems as in health care errors or inefficacy can lead to catastrophic situations. As such the best practices identified in the article can help us build structured AI pipelines so that our solution might not just be efficacious but also reliable.

iii. How your research and its results would fit into a larger AI-intensive software project where one of the core ideas from the paper would benefit the project if applied

As part of a larger software project focused on AI in healthcare, our research could incorporate the deployment best practices identified in the article to ensure the reliability and safety of AI systems. For example, using a framework such as Seldon Core, which offers advanced features such as anomaly detection and online pattern selection, we could develop healthcare decision-making systems capable of adapting in real time to new clinical data. Our WASP research could help deliver AI models specifically tailored to medical use cases, while ensuring that these models meet the high standards of robustness and security required in mission-critical environments.

iv. Discuss briefly how your research could be potentially adapted/changed to make AI engineering in the project based on the idea of the paper even better/easier

Our research could be adapted to make AI engineering in these projects even more efficient by simplifying the integration and adoption of advanced deployment practices. For example, we could develop tools or modules that facilitate the automation of quality testing of AI models in a clinical context, using open frameworks such as TensorFlow Extended (TFX). In addition, by adapting our algorithms to be compatible with cloud-based or on-premise deployment solutions, we could increase their applicability and adoption while reducing technical hurdles for clinical teams that are not familiar with SE.

5. Conclusion

Overall, by using ideas from Behavioral Software Engineering (BSE) and Quality Assurance (QA), my research can improve collaboration between different fields, create easier-to-use tools for healthcare, and ensure that AI systems are safe and reliable. The best practices from software engineering, as mentioned in the articles, can help make AI systems more secure and trustworthy, which is especially important in healthcare. This shows how important it is to combine knowledge from different areas to create effective and dependable solutions in healthcare.

REFS:

[1] *Data Sovereignty for AI Pipelines: Lessons Learned from an Industrial Project at Mondragon Corporation*. CSDL | IEEE Computer Society. (n.d.-a). https://www.computer.org/csdl/proceedings-article/cain/2022/927500a193/1EhshRjQEKc

[2] *Trustworthy and Robust AI Deployment by Design: A framework to inject best practice support into AI deployment pipelines*. CSDL | IEEE Computer Society. (n.d.-b). https://www.computer.org/csdl/proceedings-article/cain/2023/011300a127/1Ovts05lg6A