Workshop 3 - Risks of AI systems: Determining responsibilities within organisations

The TUM Institute for Ethics in Artificial Intelligence's team, as part of their collaborative project with Fujitsu Global "Towards an Accountability Framework for AI systems" presented a workshop on the risks associated with AI systems and determining responsibilities within organisations. The workshop's aim was to **bridge the gap between theoretical requirements for AI systems' accountability on the AI lifecycle vs their practical application in the real world.** Experts from diverse fields came together to discuss this challenge, ensuring a variety of perspectives were considered. The essential elements of our discussions and preliminary findings are presented here. This summary is to outline major outcomes and findings that were identified during the workshop.

AI Development Stakeholders

To break the ice, we first discussed the mapping of stakeholders from theoretical to practice. From the theoretical map showed to them, major points came out of the conversations:

- Stakeholder analysis is a valuable tool for understanding complex stakeholder relationships and can inform effective stakeholder engagement and management strategies.
- Stakeholder relationships can be categorised into three main types: cooperative, competitive, and conflicting.
- Biases, assumptions and values of the (moral) person conducting the stakeholder analysis have to be considered as they can influence their look on the situation.

Measures in the Ideal World vs in Real Use Cases

In the workshop, participants engaged in a discussion on the (ethical) obligations and corresponding measures for organisations involved in the AI lifecycle. The exercise focused on identifying necessary actions at each stage of the AI development process, and participants were asked to envision an "ideal world" where companies acted in full compliance with their obligations, and then to envision practices in the "real world" based on use case led discussions.

Problem Understanding	**Data Handling**	**Model Building**	**Model Monitoring**
Define legal, business, and ethical environment	Assess data quality regarding official requirements, ethicality, and adequacy	Understand the system, data, and third party components	Determine monitoring criteria, KPIs and objectives

Define the use case and its context	Provide documentation	Define targets	Determine resolution strategies & fallback plans
Create an adequate team	Assess and minimise risks	Evaluate the model in regards to cost vs quality, performance, and user needs	Constant monitoring of the model
Plan Processes	Privacy	Monitor risks, KPIs, etc.	Re-evaluate objectives and conflict of interests
Initiate risk assessment	Respect needs of minorities	Solve identified issues	Communicate transparently about risks, reports, etc.
Ensure outside engagement	Ensure transparency on data collection processes	Document processes	Change mindset and guide people
	Define data set limitations	Reevaluate team adequacy	Enable auditing
	Respect data security and regulations	Ensure adaptability	Ensure transparency of system's purpose and operations
		Ensure participation and collaboration	Avoid reputational issues (FDA wall of shame)
		Consider new responsibilities and processes changes that might arise	Run tests to ensure usefulness, usability and accuracy with participatory design
		Clarify technical objectives	
		Understand limitations of the model	

Table 1: Summary of the exercise outcomes. Measures identified for the real world use case are marked in pink, measures identified for both the ideal world and use cases are marked in blue.

Table 1 summarises the measures for operationalizing responsible AI identified during the workshop for both the ideal considerations and the use case based discussions. It shows that a few of the actions considered for the ideal world were also found useful when regarding them with a concrete use case in mind. In the table, those overlapping measures are represented in blue boxes. On the other hand, some steps mentioned for the use cases were new to the discussion, they are represented in pink boxes. We can thus see that while the tasks that are proposed in theory seem to help make AI applications more responsible, continuous improvement of measures and adaptation to requirements of specific contexts is required.

Measures Implemented in Company Processes

The survey conducted during the workshop indicates that companies have implemented over half of the ideal steps discussed, indicating a willingness and effort on their part. Big interest in ensuring transparency was reported for model building, and traceability through record keeping and incident tracking were reported tasks for model monitoring. Major challenges to executing the ideal tasks include lack of training and legal uncertainty. Interestingly, participants did not believe the implementation of the ideal tasks for accountability would lead to the realisation of responsible AI, and continuous improvement was seen as necessary. Despite progress, missing standardisation and best practices are still a significant concern for companies.

Conclusion and Outlook

In summary, measures can be implemented throughout the AI development process, with general alignment to ideal considerations. The AI lifecycle provides guidance, but integrating measures can be challenging due to interconnected stages. Companies are making progress, but standardisation and best practices are lacking. The proposed theoretical tasks improve AI responsibility, but continuous monitoring and improvement are still required.