Workshop – Summary & Outcomes

Accountability Requirements for AI Applications

March 2022

Preliminaries and Workshop Goals

The Problem

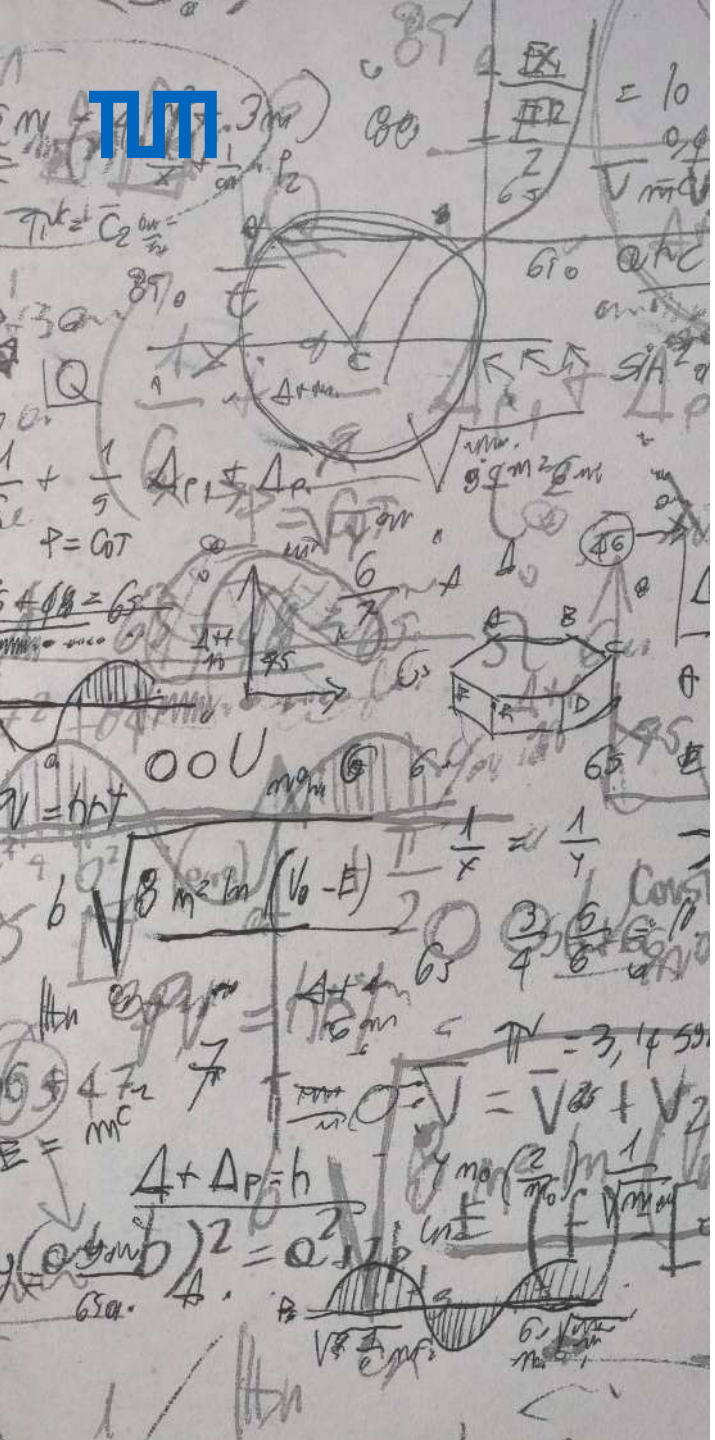
The pressing issue of who to hold accountable for AI decisions and predictions is still a challenge in existing research and even more for practice.

The growth of AI applications into highly **complex systems** intensifies challenges of responsibility clarification.

The **opacity** of complex AI systems hinders providing reasonable explanation to justify or disprove accountabilities.

The transfer and application of **existing legal and ethical frameworks** to the specific context of AI has been challenging.

Icons from flaticon.com (Freepik, Smashicons)



Definitions

Some preliminary definitions to kick-off from the same baseline.

Accountability = “the fact of being responsible for what you do and able to give a satisfactory reason for it”

I

II

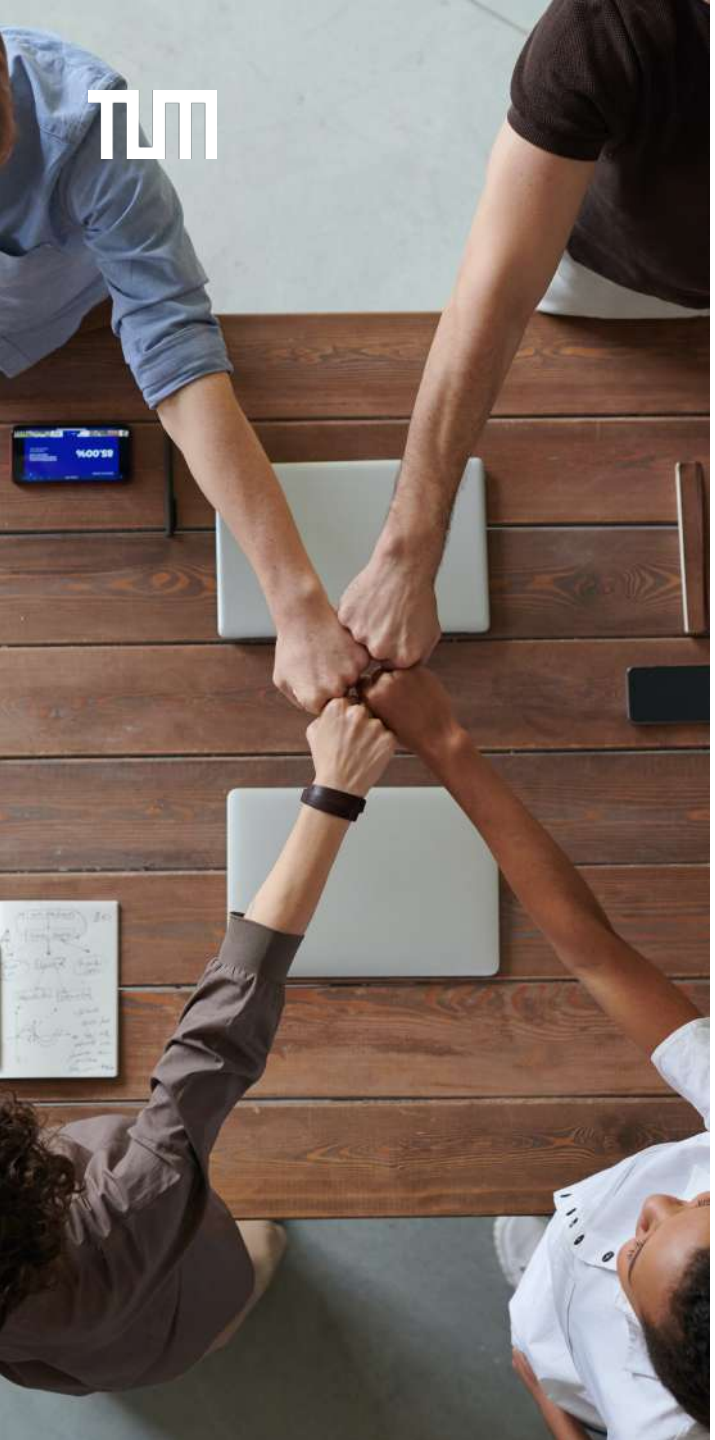
I **Responsibility** = “something that it is your job or duty to deal with”

II **Reasoning** = “the process of thinking about something in order to make a decision“

Source: Cambridge Dictionary

Our Project Approach

In the joint research project between Fujitsu and TUM, we aim at applicable solutions for an AI accountability framework using a risk-based approach.



Workshop Motivation

The workshop's main goal was to gather insights from practice on requirements of an accountability framework for AI systems guiding the project's further activities.

Workshop Methodology

Workshop Structure

The workshop's core idea was to define requirements of an accountability framework for AI systems based on risks and responsibilities.

Welcome

- Introduction from Fujitsu & TUM
- Discussion
- Introduction of case studies

Part I: Risks

- Discussion of case studies in small groups
- Wrap-up in panel

Part II: Management and Responsibilities

- Discussion of case studies in small groups
- Wrap-up in panel

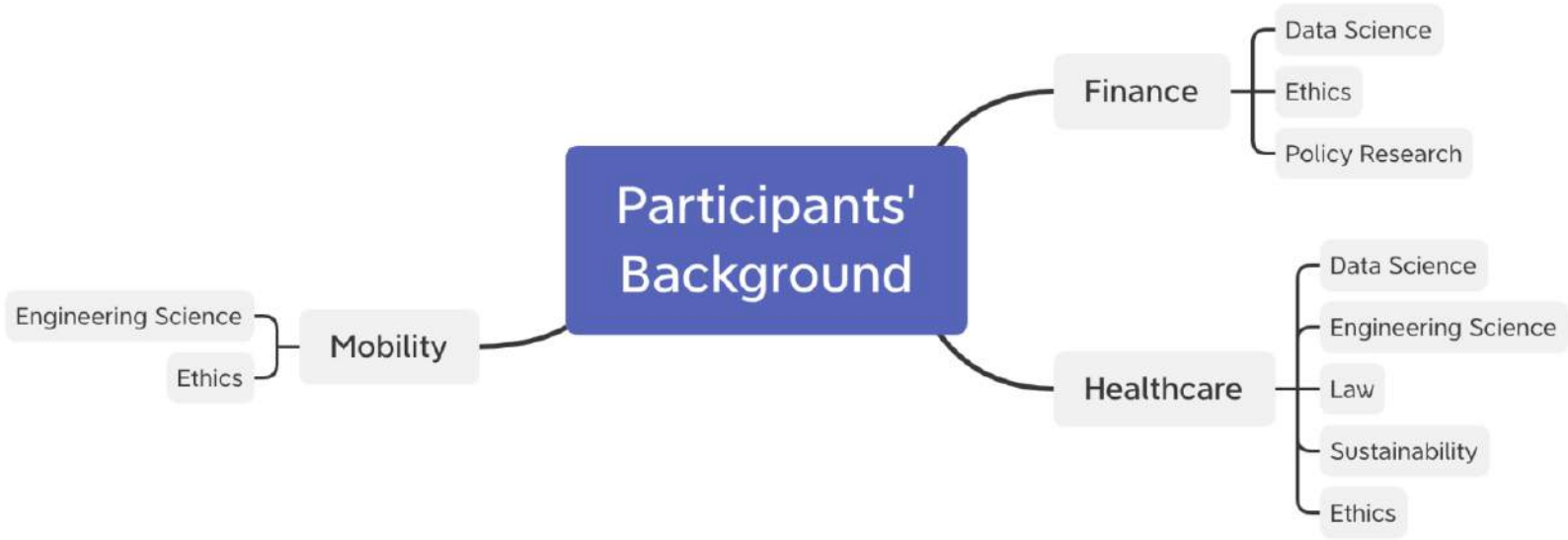
Participants

Participants mainly came from industry and academia brining along a valuable variety in social and professional backgrounds.

3	different consulting firms represented
4	different academic institutions represented
8	academics
10	private sector participants
11	distinct specialties
12	female participants
18	participants in total (incl. 3 organizers)

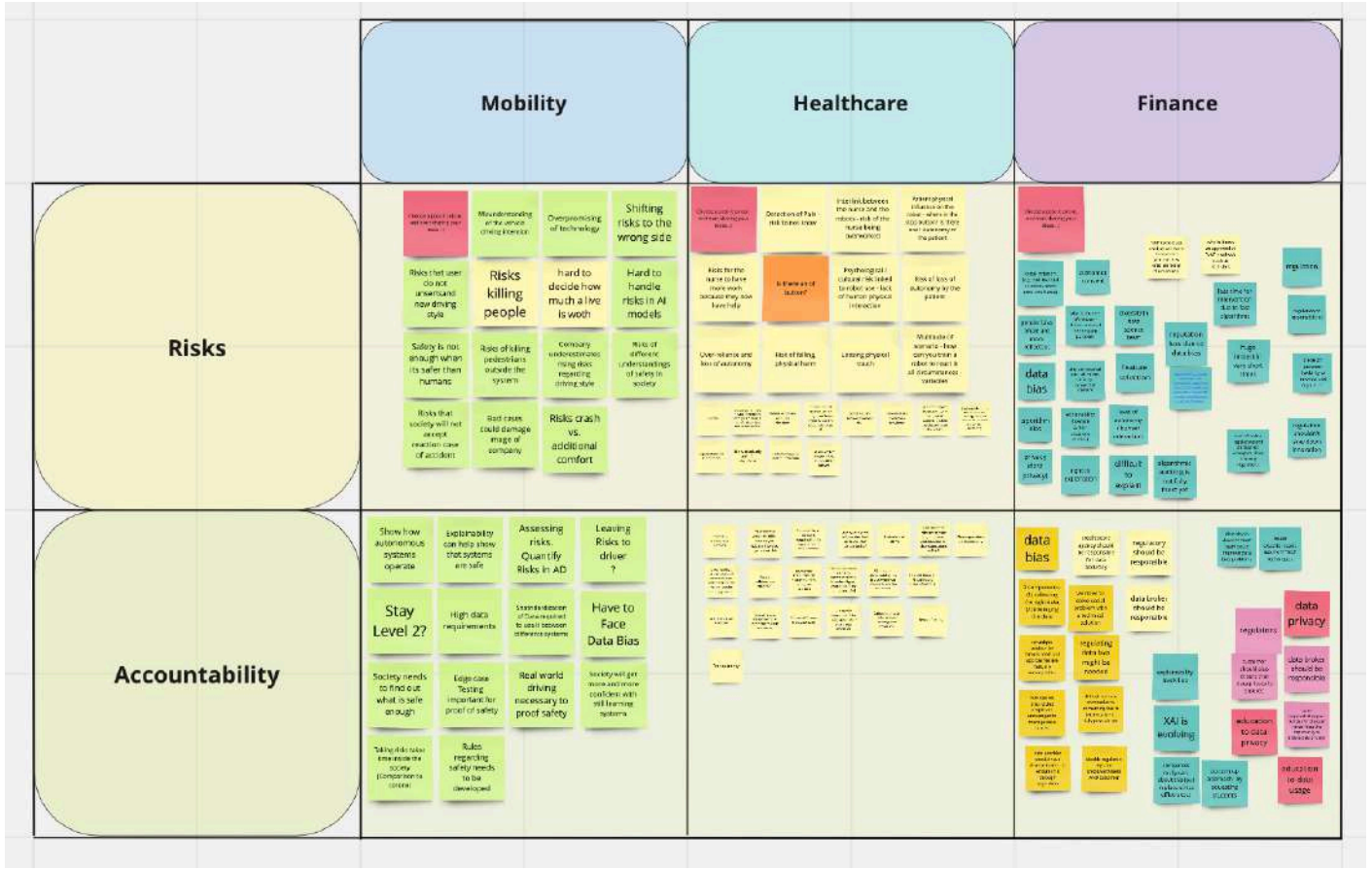
Participant Background

Participants brought great variety and diversity to the panel discussions and analyses during the case study sessions.

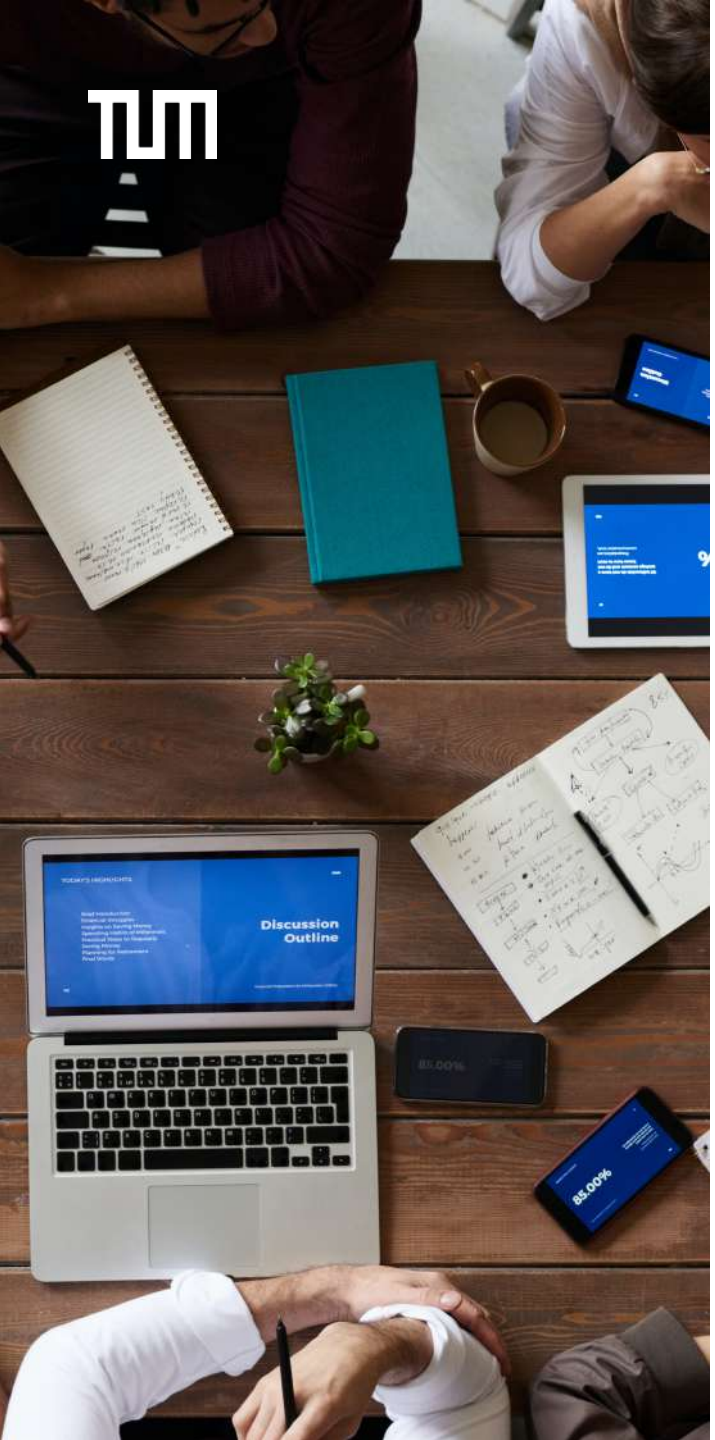


Workshop Tools

We used an online collaborative Miro-board to record and structure ideas from the discussion sessions.



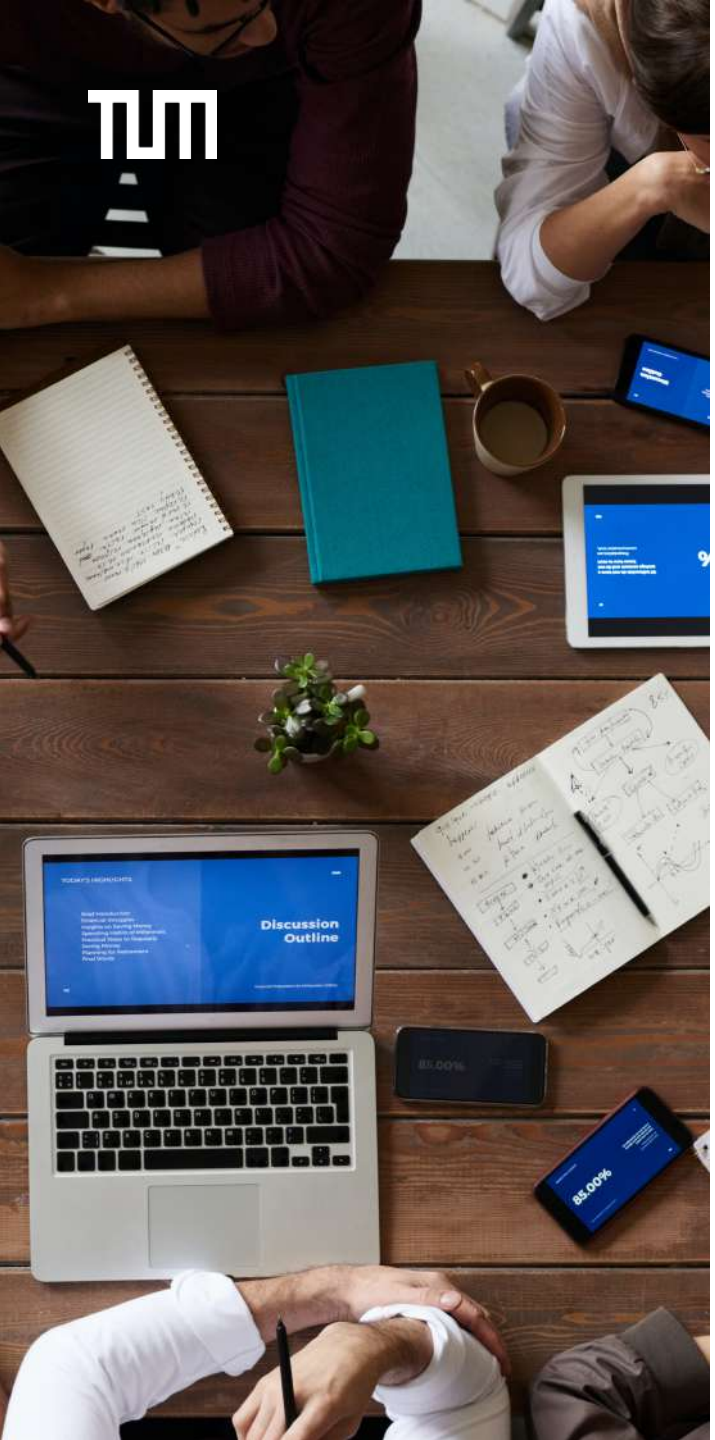
Outcomes



Introductory Question

We discussed the participants' insights on AI and accountability from their daily practice.

In your opinion, what are the **most pressing challenges** in the industry regarding risks of **AI and accountability**?



Most Pressing Challenges

A variety of pressing challenges concerning AI and accountability were identified and examined by the participants during the first joint discussion round.

Education

”Education is key. People need to be educated on the risks and safety of AI, data scientists and developers need to be educated on the ethical challenges of AI and regulators need to be educated on current technological developments.”

Regulation

”Detailed legal acts and legal cases are required.”

Explainability

”There is a gap between what can be explained and what needs to be explained. How can we ensure that people can understand what the system explains?”

Acceptance & Trust

"Deployment and technology acceptance are two different things. Trust is key in acceptance, thus we need to demonstrate trust.”

Data Bias

”AI systems mustn't be biased against certain groups inside society. Non-discrimination and data quality needs to be ensured during development and deployment.”

Privacy

”How can the high data privacy standards be fulfilled in AI systems.”

Accountability

“Accountability needs to be understood in terms of how but also which systems to design. Only because we can do something does not mean, we should do it.”

Safety & Risk

"Technology can never be 100% safe. The question is, how much risk is bearable, what is safe enough and how can we determine suitable thresholds.”

Icons from flaticon.com (Freepik)

Case Studies

We discussed 3 case studies regarding potential risks and responsibilities in break-out sessions.

Mobility

Case: adjusting driving style for self-driving vehicles

Healthcare

Case: robots to support care of elderly people

Finance

Case: algorithmic assessment of people's creditworthiness

Icons from flaticon.com (Good Ware, Kiranshastry)

Use Case: Mobility

This use case on driving styles for autonomous vehicles was discussed regarding identified risks as well as potential responsibilities and strategies to mitigate them.

A vehicle manufacturer has conducted a study on satisfaction with its own self-driving vehicles. The study found that the biggest negative point was the vehicle's slow driving style. Due to their impatience, some customers prefer to drive independently and faster than autonomously. The vehicle manufacturer considers to match the risk through driving characteristics to that of a human driver in situations where there is always some risk.

Risks: Mobility

Various potential risks have been identified by the participants for the use case on driving style adaptation powered by AI-algorithms.

Transparency

Customers and society could have a different understanding of safety and how safe a system has to be. Intransparency of the system could lead to further risks.

Regulations

The company must ensure that the applicable laws continue to be complied with the adaptation of the driving style.

Safety Assessment

Changing the system properties poses a strong risk. Every new feature needs to be tested and validated to ensure that the system works as intended. Changing the driving style could shift the overall risk e.g. to pedestrians.

Society

Changing the driving style and the risk of the overall technology could lead to a debate how much a human life is worth. The debate could slow down the overall technology development.

Trust

The vehicle's new AI enabled driving style could confuse passengers and other traffic participants. If the driving style is not intuitive, new problems regarding trust could arise.

Company Image

The new application could lead to additional deaths in the traffic. Even if the technology is considered safe, society may not accept the damage. Subjective opinions could harm the image of the entire industry.

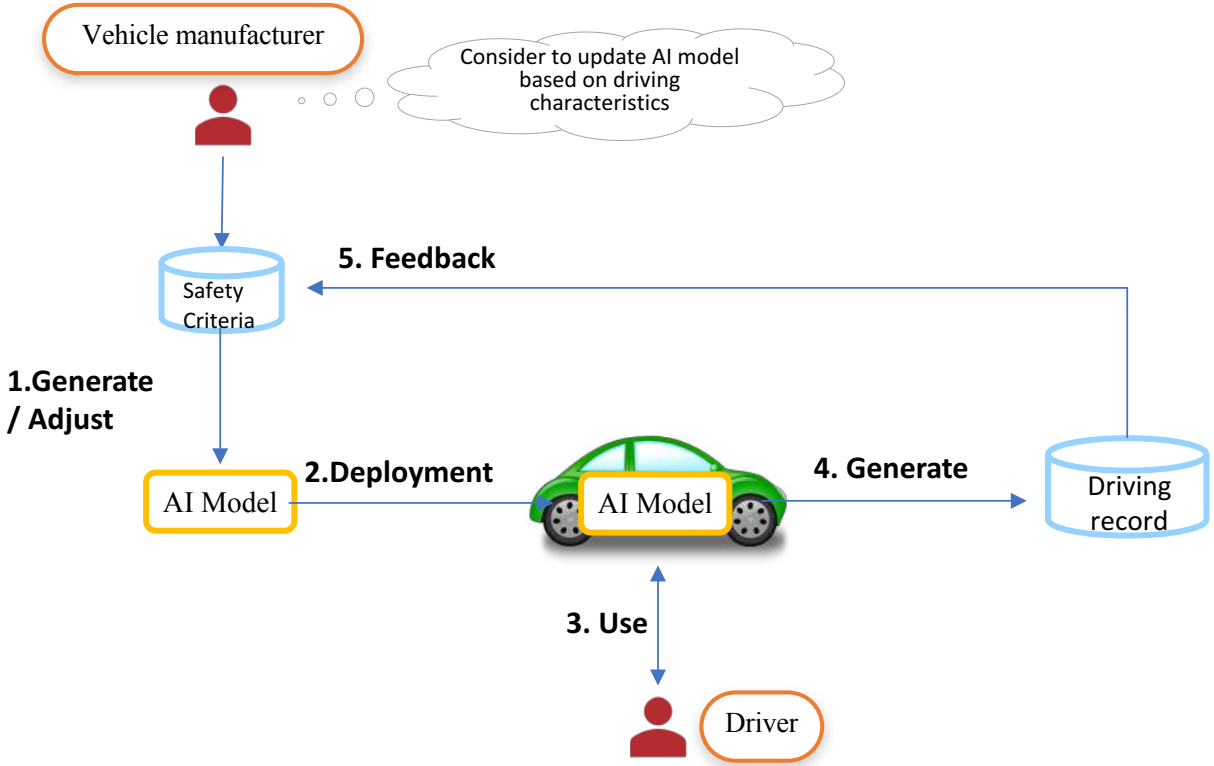
Explainability

Today AI algorithms are opaque and therefore carry risks. The user and the developer need to know what the system is doing in specific situations to debug the system or to react to misbehavior.

Icons from flaticon.com (Freepik)

Use Case: Mobility

A system diagram was shown to the participants in the second discussion round to map the different actors involved in the use case and their activities.



Responsibilities: Mobility

Management strategies and responsibility distribution have been identified for the use case on driving style adaption powered by AI-algorithms.

Data Standardization

A regulation of data standardization might be needed. The quantity and quality of data are critical to the success of AI systems.

In order to share data between different companies and applications, global regulation and standardization is needed.

Responsible actors: companies, agency, regulator, data broker

Transparency & Testing

The stakeholders need to understand how the system works and handles decisions. To increase the trust and acceptance of the customers and regulators the AI application needs to be transparent.

Nevertheless, the system must be tested extensively in the real world to be considered safe and reliable.

Responsible actors: company, regulator

Society

The society needs to discuss edge-case scenarios and how safe an AI system needs to be. A new human driver for example also causes higher risks than an advanced driver. How much risk is acceptable and is the society willing to accept systems that are still in a learning process?

Responsible actors: society, regulator, government, influencer

Icons from flaticon.com (Freepik)

Use Case: Healthcare

This use case on robots used for care purposes was discussed regarding identified risks as well as potential responsibilities and strategies to mitigate them.

An elderly care home decides to use healthcare robots to support the nursing staff in their daily tasks. The robot can help lift and help individuals walk from point A to point B under a nurse's supervision. It adapts its posture to the weight and height of the patient.

Risks: Healthcare

Various potential risks have been identified by the participants for the use case on care robots.

Management

Because the robot would require supervision from the nursing team, risks lie with nurses' undereducation on the robot technology. Additionally, management could believe the nurses can now take more work on as they receive technological help, putting the nurses at risk of overwork.

Psychological

Due to the use of a robot for movements, the reduction of human physical interaction might be a wellness risk for the patients.

Technical

The robot could present a risk to the user's physical wellbeing if it was for example not equipped to recognize signs of physical pain, or trained on data sets showing bias, thus having the robot not able to properly assess variations in weight and height of a person due to individual characteristics.

Education

As mentioned earlier, there is a risk for nurses not to know the tool as they would like, creating situations of distress for them. This also applies to patients who will be using the robot, which are at risk of not fully understanding the extend of its usefulness, its workings, and which medical data are necessary to its functioning.

Privacy

As for any AI-powered technology, patient's data will be required to have the robot run as it should. Due to cybersecurity and simple use risks, data privacy to external entities, and non-medical staff internal entities are to be considered.

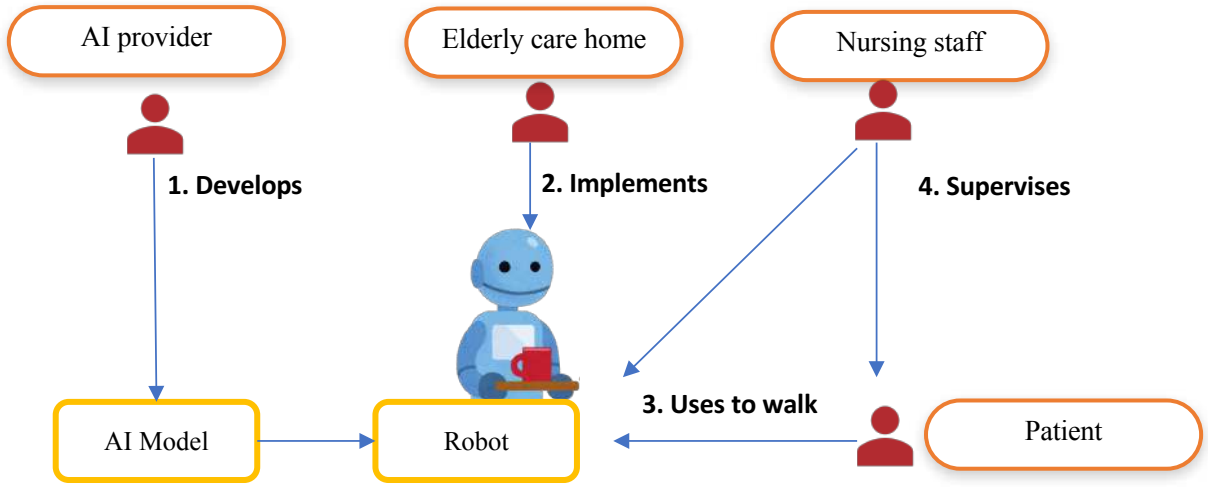
Autonomy

The patient could be at risk to see the robot's use imposed on them without enlightened consent, or in contrast be requiring the robot when unnecessary, and thus not recover or loose movement abilities prematurely. Additionally, the autonomy of the nurse to decide to use this new tool or not has been identified as a risk.

Icons from flaticon.com (Freepik)

Use Case: Healthcare

A system diagram was shown to the participants in the second discussion round to map the different actors involved in the use case and their activities.



Use Case: Healthcare

Management strategies and responsibility distribution have been identified for the use case on care robots.

Explanations

The providers should provide clear and detailed explanations of the limitations and risks, hardware, and software workings of the tool to all actors involved with it, whether the patients, the medical staff, or the institution buying it. Additionally, situations in which the end-user or supervisor are to be accountable for the tool need to be defined according to the law and specific regulations by the provider.

Responsible actors: developer, provider, patients, medical staff, medical institution

Data

The provider should ensure that appropriate training dataset are used depending on the target populations for a most adequate performance and fairness demonstration.

Moreover, regulators are to control the compliance to current GDPR regulation and reflect on the possible need to make it evolve to fit best with AI technology evolution.

Responsible actors: regulator, provider

Feedback

Ensuring a proper feedback system between the AI, the provider, the developer, and the users is necessary to fix issues when they arise and meliorate the product to the specific needs of the target population, taking into account multidisciplinary point of views, cultural aspects, and full transparency on all sides.

Responsible actors: developer, provider, users

Icon from flaticon.com (Freepik)

Use Case: Finance

This use case on algorithmic assessment of creditworthiness was discussed regarding risks as well as potential responsibilities and strategies to mitigate them.

A credit score agency developed a new and unconventional AI-enabled algorithm that, based on personal features, assesses the creditworthiness of people and offers this information to other companies, such as from the finance and insurance industry. The algorithmic model is neither disclosed to customers that buy the credit information nor to the public that is subject to creditworthiness investigations.

Risks: Finance

Various potential risks have been identified for the use case on algorithmic assessment of creditworthiness.

Privacy

Data privacy poses a strong risk. Consumer consent can be challenging, in particular, if the algorithm uses historic data.

Efficiency

The algorithms efficiency and speed leads to less time for intervention and therefore large impacts in short times.

Bias & Diversity

Several forms of bias (e.g., gender/socio-economic bias in data/algorithm) can impact the algorithm's fairness and effectiveness and, hence, lead to reputation loss of the engaged actors.

Regulation

Regulations are needed to define responsibilities, however, there is a trade off between flexibility/speed of innovation and innovation governance.

Autonomy

The algorithm's level of decision autonomy and degree of human interaction or intervention can pose risks, in particular given the fast operation times.

Algorithmic auditing

Automated auditing is highly encouraged for such a use case, given its operation in the finance sector. However, methods for algorithmic auditing are still to be advanced.

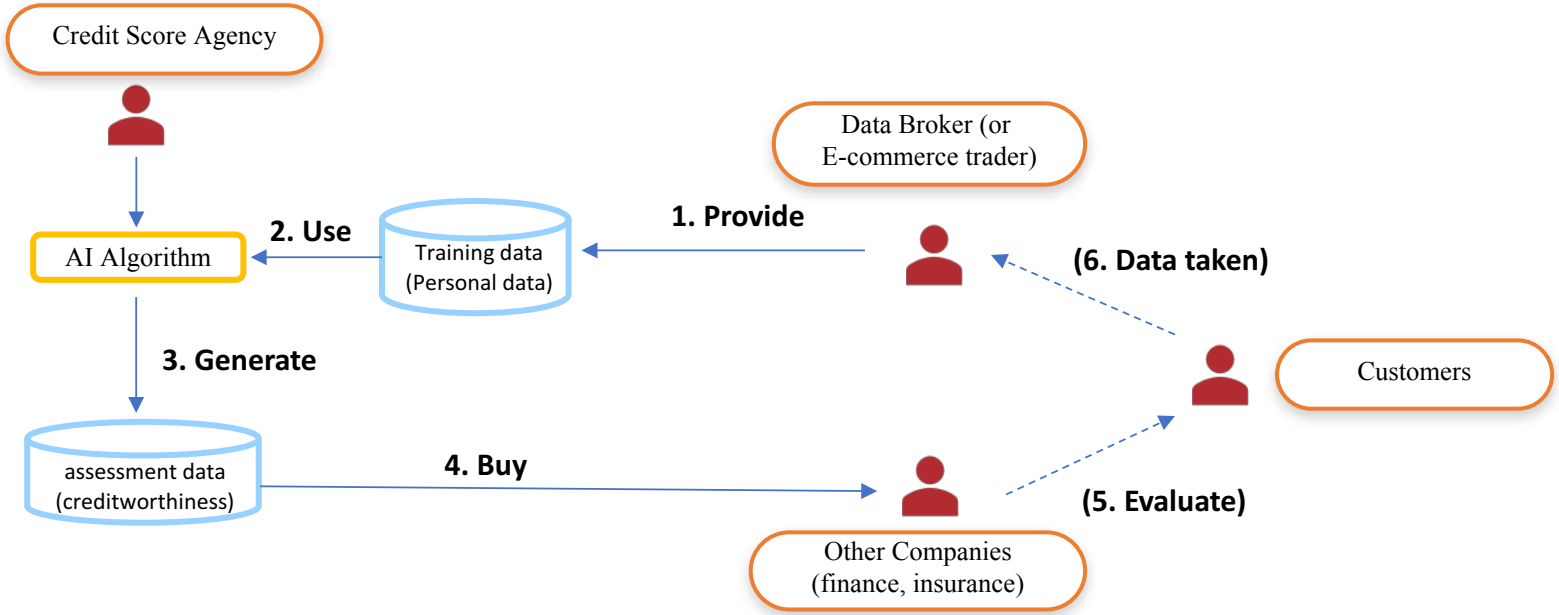
Explainability

Missing transparency can cause problems, in particular, for operation in the finance sector. Transparency and explainability are needed to ensure and demonstrate compliance with regulations

Icons from flaticon.com (Freepik)

Use Case: Finance

A system diagram was shown to the participants in the second discussion round to map the different actors involved in the use case and their activities.



Responsibilities: Finance

Management strategies and responsibility distribution have been identified for the use case on algorithmic assessment of creditworthiness.

Data bias

A regulation of data bias might be needed. However, challenges arise with this approach, as technical integration of fairness in data and algorithms is still an ongoing field of research.

Diversity in development teams could raise the awareness for data bias issues and, hence, help managing this risk.

Responsible actors: credit score agency, regulator, data broker

Data privacy

The actors developing, deploying and using the investigated algorithm should ensure that data privacy is provided. Regulators should check compliance to data privacy standards.

The data subjects simultaneously bear some responsibility for enabling all possible measures to protect their own data. Therefore, education to data privacy and data usage is key.

Responsible actors: credit score agency, regulator, data broker, customer, data subject

Explainability

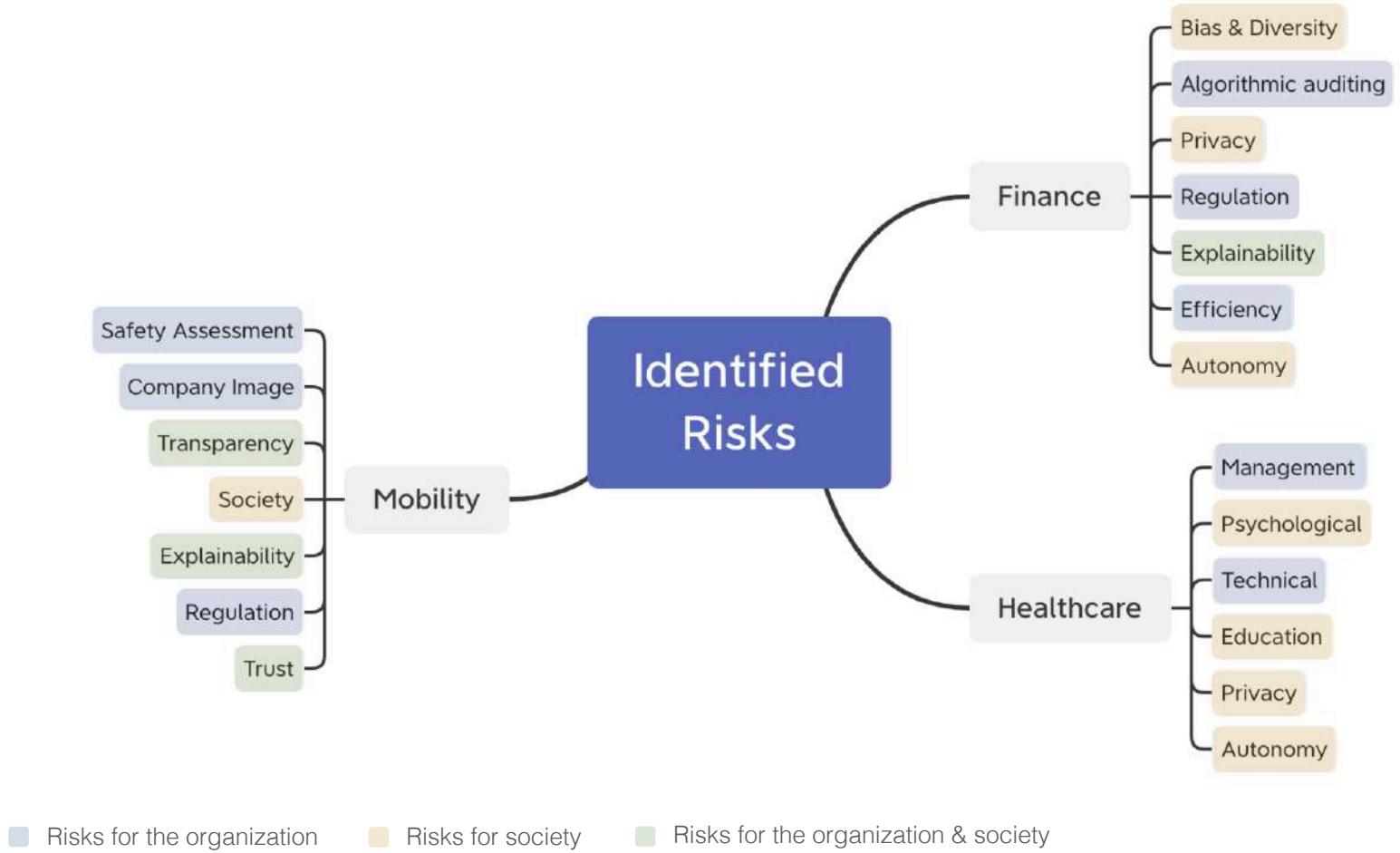
XAI techniques are emerging and companies are increasingly interested in fostering algorithmic transparency.

Education can help strengthen and accelerate awareness and mitigation of explainability issues. A “bottom-up” approach to train students on explainability issues and methods could help solving this issue for the future.

Icons from flaticon.com (Freepik)

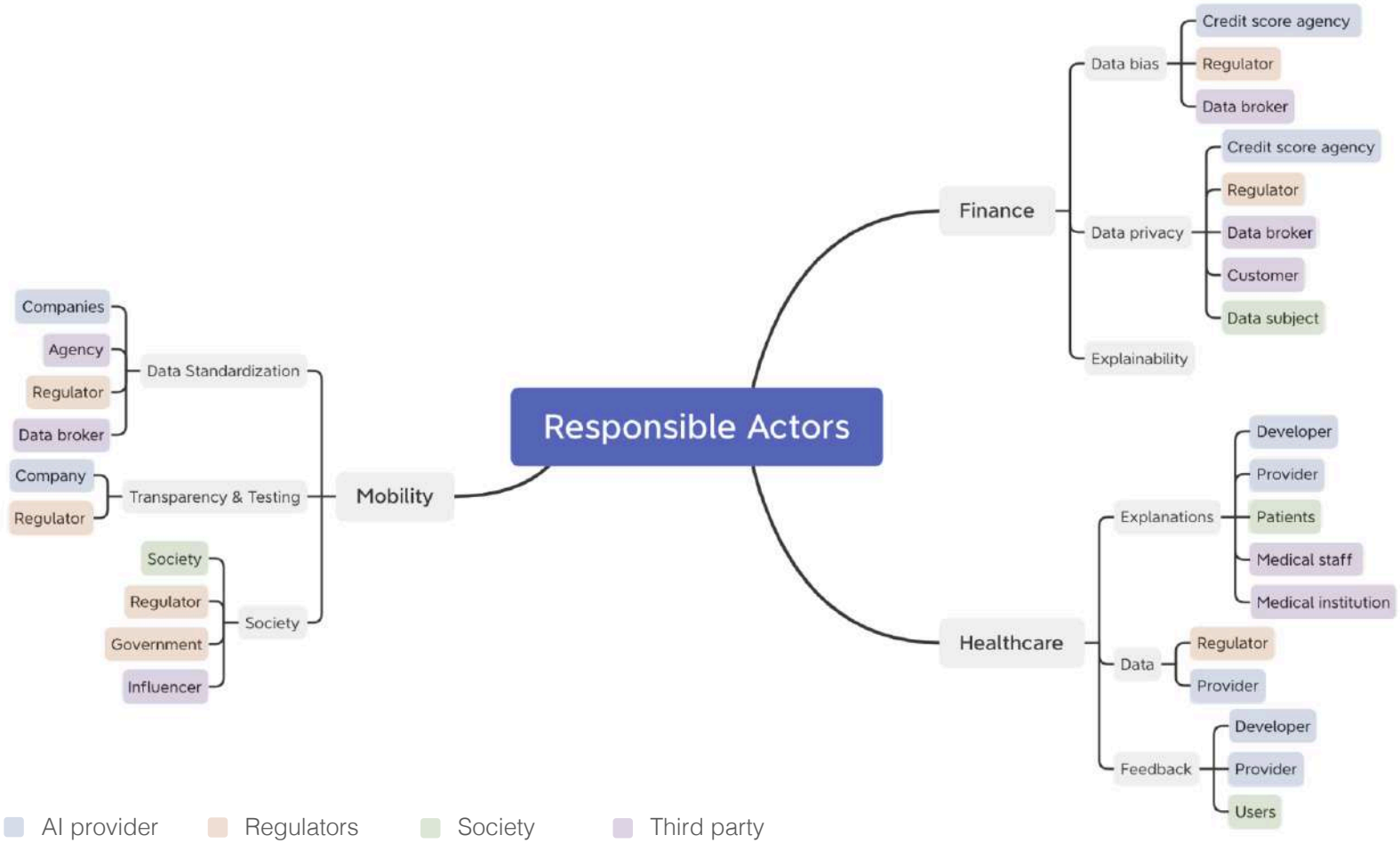
Summary

Various potential risks have been identified during the workshop for the three presented use cases.



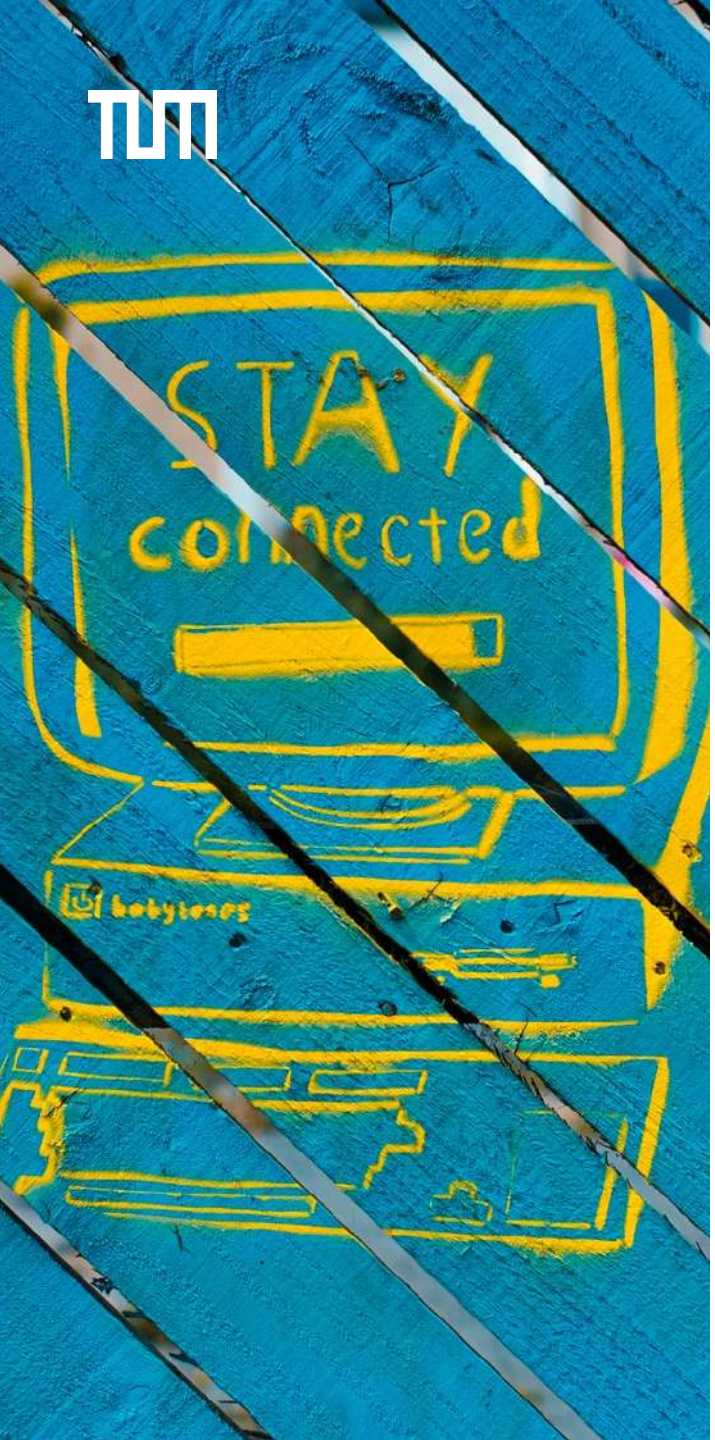
Summary

Management strategies and responsibility distribution have been discussed for the three use cases during the workshop.



Outlook

Our second workshop will take place in June dealing more intensively with risk management and responsibility assessment concepts and methods.



Stay connected!

We are happy to see you again in our next workshop. Until then, stay tuned through our multiple channels.

Stay connected through our websites ieai.mcts.tum.de and mos.ed.tum.de/en/ftm/, subscribe to our newsletter or follow us on twitter, LinkedIn and YouTube.