TrustMLRegulation.
Managing Trust and Distrust in Machine Learning with Meaningful Regulation.

Meaningful regulation for technology is understood as consisting of regulatory frameworks, institutional designs and best practices of governance and oversight that are both politically and technically feasible. To make these regulations effective, sufficient trust in the functional correctness of the regulated technologies is also needed, while at the same time still allowing freedom to facilitate innovation.

To this end, this IEAI related project started out with the goal of exploring the border that divides the domains of trust and distrust in autonomous and intelligence systems. In particular, this collaboration between the Department of Electrical and Computer Engineering and the Munich Center for Technology and Society at TUM looked at the governance challenges for industrial corporations during the development of such systems, addressing the following questions

1. What could be suggestions for effective regulation, certification and oversight of AI-based applications? Are they feasible in terms of how they fit into the existing institutional landscape? What are the limits of their application and are there ways to overcome such limits?

2. Are those suggestions for regulation and governance that are institutionally feasible also manageable on a technical level? How can selected Machine Learning methods be tested and evaluated in a way that fits the institutional and political requirements of such suggestions?

Potential unintended consequences and several technical characteristics, such as unpredictability and lack of transparency, pose considerable challenges to the current governance infrastructure. This concern has been recognized and taken up by different regulatory initiatives ranging from industry-based self-governance guidelines, standards-setting bodies, academic communities and policy actors on national, international and global regulatory levels.

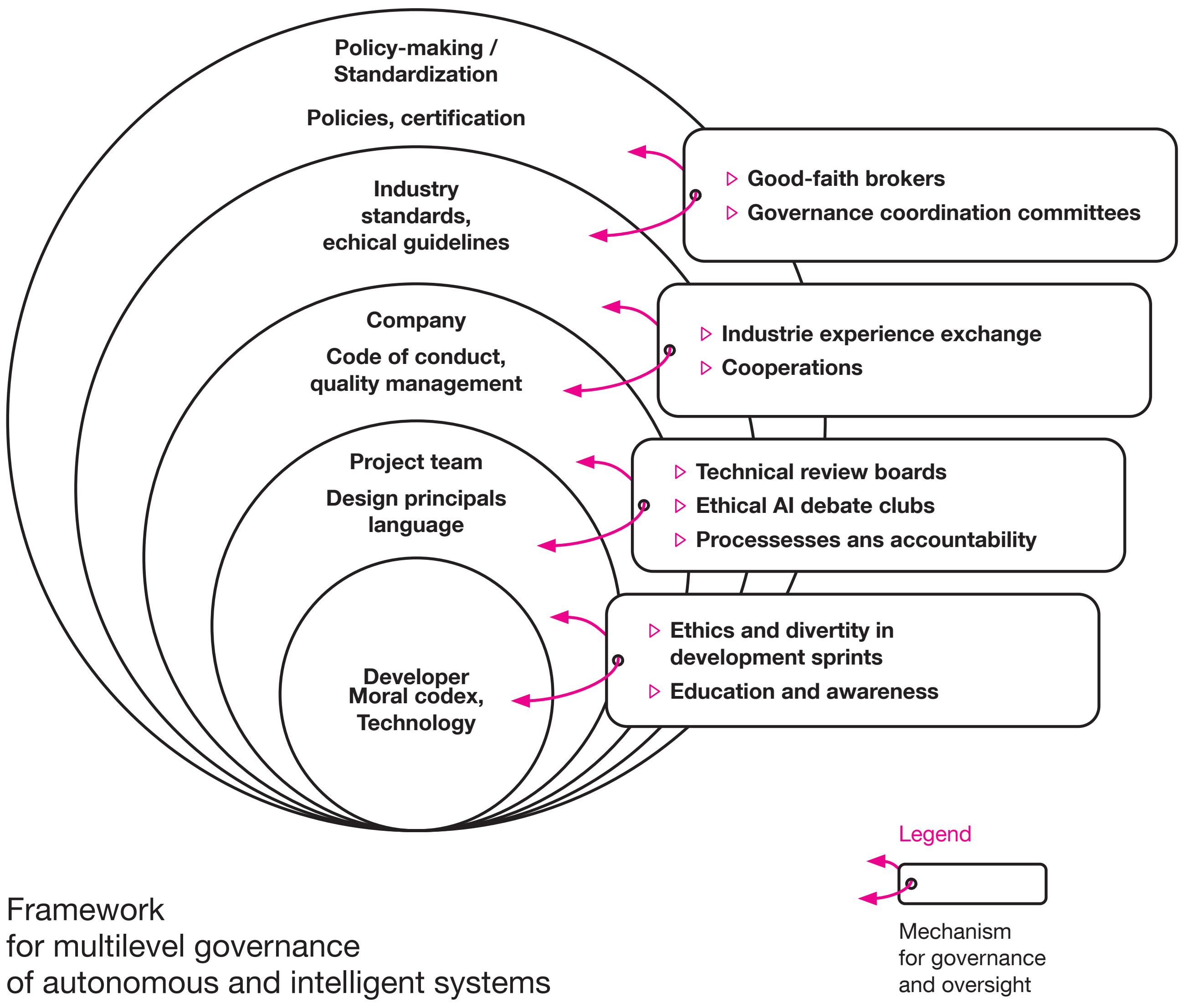
The team started by mapping these initiatives and identified common threads, as well as controversies and differences between them. This included AI standardization initiatives such as the IEEE AIS, ISO/IEC AIS, SAE AIS, TPC AIS, ITU AIS or the Industry for Open AIS, as well as the EU AIS (EC's European Group on Ethics in Science and NewTechnologies (EGE), Declaration of cooperation on AI, Artificial Intelligence for Europe). From this mapping, the team undertook two policy analysis deep dives in a comparative manner. One analysis looked at industry self-regulation efforts in internal AI ethics guidelines in the insurance sector and the other focused on reconstructing the policy process that has led to the EU High-Level Expert Group on Artificial Intelligence recommendations and their implementations in EU Horizon funding and in tools such as the ALTAI framework.

Base on this work, the project had the following outcomes

- ▷ Proposals of a practical framework for multilevel governance of autonomous and intelligent systems (see Figure). The framework identifies actors on five different levels of decision-making with a particular focus on industry actors, standardization and regulation and enables the identification and development of potential tools for guidance on each level and mechanisms for governance and oversight of the tools and technologies.

- ▷ Development of analogies to the governance of life sciences and agile approaches with good practices in these fields, exemplified using the case study of the discussion of the United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems between 2017 and 2019. Results show that linguistic analysis allows for the tracking of changes, implying that a change of the topics and thus the focus of a governance process can be achieved by the inclusion of actors from lower levels.

Feedback from four semi-structured interviews was used in the creation process of the governance framework. Interviews highlighted the benefit of the framework to identify potential actors in decision-making processes and to frame governance processes and thus to set up effective governance structures for complex technologies such as autonomous and intelligent systems. Out of these findings, the team has submitted several proposal to fund further aspects related to this work, looking in-depth at effective levels of trustworthy and transparent AI. ●


Policy-making / Standardization
Policies, certification
Industry standards, echical guidelines
Company
Code of conduct, quality management
Project team
Design principals language
Developer
Moral codex, Technology
▷ Good-faith brokers
▷ Governance coordination committees
▷ Industrie experience exchange
▷ Cooperations
▷ Technical review boards
▷ Ethical AI debate clubs
▷ Processesses ans accountability
▷ Ethics and divertity in development sprints
▷ Education and awareness
Legend
Mechanism for governance and oversight

Framework
for multilevel governance
of autonomous and intelligent systems

2020 Papers and Project Highlights

- ▷ A Practical Multilevel Governance Framework for Autonomous and Intelligent Systems (working paper) (Pöhler, L., Diepold, K., Wallach, W.).
- ▷ A sentiment, trust and content analysis of the news coverage on Artificial Intelligence in Germany (Master's Thesis, working paper) (Morandell, J.)
- ▷ Governing artificial intelligence (Master's Thesis) (Pelepets, M.)
- ▷ Governance of Autonomous and Intelligent Systems (AIS) – A Practical Multilevel Governance Framework (Master's Thesis) (Pöhler, L.)
- ▷ AI Ethics in Insurance. Interpretation and Application of Ethical Principles (Master's Thesis) (Pai, C.)
- ▷ Practices of Governing and Making Artificial Intelligence (Master's Thesis) (Samaniego, J.M.)

2020 Conferences

- ▷ Digitization, Automation and Society, TUM-UQ online Workshop, July 2020
- ▷ Digital Superpowers and Geopolitics, TU Vienna Lecture Series, January 2021

Principal Investigators

- ▶ Department of Electrical and Computer Engineering, TUM
- ▶ Munich Center for Technology and Society, TUM