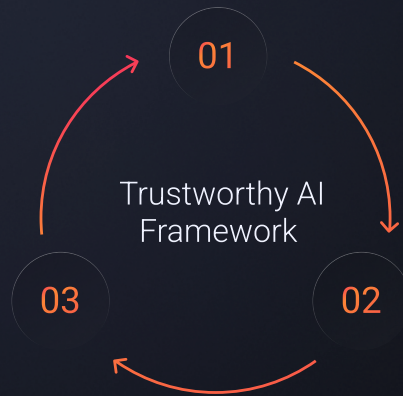


Trustworthy AI Framework

Ensuring Trustworthy AI requires a systematic integration into processes that extends across various roles and teams. Employing the ML6 Trustworthy AI framework involves a structured methodology for managing ethical and legal risks



01

Risk identification & documentation

Identifying ethical & legal risks as integrated part of way of working

- ⊗ Risk assessments
- ⊗ Ethical Sounding board
- ⊗ Employee awareness

02

Risk mitigation

Identifying mitigation measures during design & development

- ⊗ Best practices in development
- ⊗ Solutions fit for purpose
- ⊗ Compliance with applicable regulations

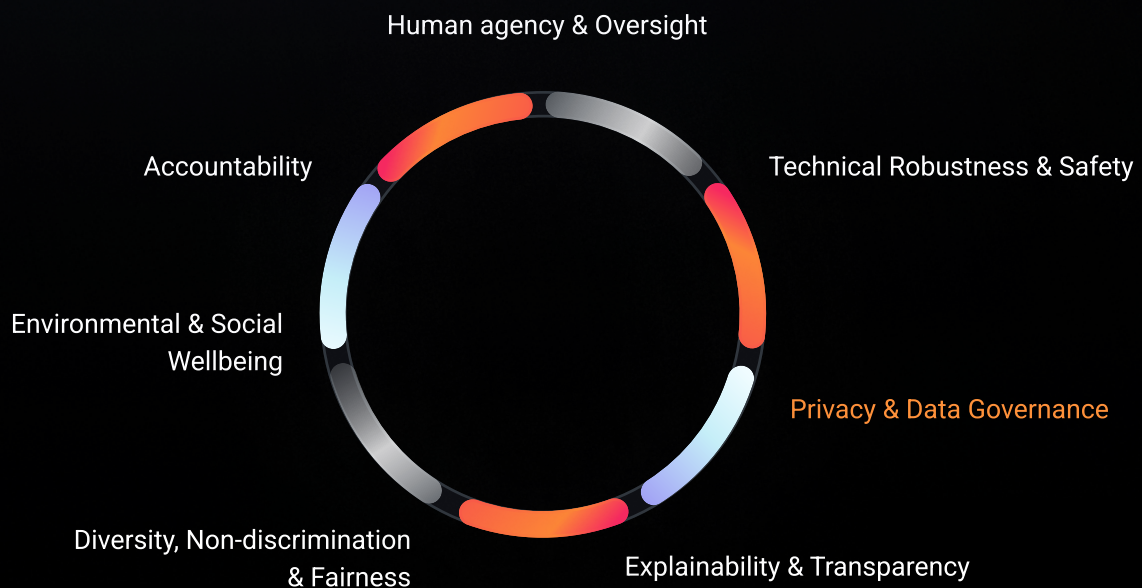
03

Monitoring

Ensuring continuous risk monitoring and control

- ⊗ Monitoring and control
- ⊗ Process integration
- ⊗ Feedback & continuous improvement

In line with the high-level expert group of AI (ALTAI), at ML6 we typically look at 7 dimensions to assess AI systems and potential risks:





Technical Robustness and safety

We strive to mitigate and prevent the risk of unintentional harm by ensuring that AI systems we build are resilient, secure and safe, as well as accurate, reliable & reproducible.



Explainability & Transparency

We see three elements of Transparency: traceability, explainability and open communication about limitations, which are key to creating trust, improving AI systems and gaining new insights



Environmental & social wellbeing

The broader society and the environment should be considered as stakeholders throughout the AI system's life cycle. It is important to raise awareness of the environmental and energy footprint of digital technology, consider societal, economic and political implications and respect democratic processes.



Human Agency & Oversight

AI systems should empower humans, allowing them to make informed decisions and fostering their fundamental rights. At the same time, we need to make sure that proper oversight mechanisms exist where needed.



Privacy & Data Governance

Privacy and the protection of personal information is core to building trustworthy AI systems. Adequate data governance is needed to mitigate risks and ensure the protection, quality and integrity of data.



Diversity, Non-discrimination & Fairness

AI systems may suffer from unfair biases, for example due to historical biases in data they are trained on. Risks of bias need to be identified and mitigated early on and monitored during the development and use of AI systems in order to avoid unfair discrimination.



Accountability

We need to ensure responsibility and accountability for AI systems and their outcomes. At ML6, we hold ourselves to account to work with our clients to put in place the necessary mechanisms to ensure responsibility and accountability of the applications we build.

Depending on specific use cases and sector, some of these dimensions will be more or less relevant for a specific AI application. It is key to identify the dimensions that carry the highest risks for a specific AI system and define risk mitigation strategies early on.