

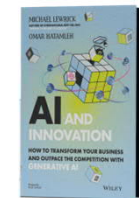
Project:
Team:
Version & Date:



Quick guide: The AI Ethics Canvas offers a structured framework for assessing and mitigating ethical risks. By systematically evaluating fairness, explainability, transparency, empowerment, privacy, and robustness, teams can identify potential issues, develop concrete mitigation strategies, and ensure responsible AI development.

AI Ethics Canvas

Tips and tricks for the
template on book page: 113



Lewrick / Hatamleh
AI and Innovation
978-1394254972

Name of AI Project/Name of Use Case:



Fairness

How is fairness ensured in the development and deployment of the AI system to prevent bias and discrimination?

For Example: Describe how the AI system avoid discrimination against any particular group of people?

Explainability

How is explainability ensured to make the AI system's decision-making process transparent and understandable?

For Example: Explain how your AI system arrives at its decisions in a clear and understandable way?

Transparency

How is transparency ensured in the data collection, processing, and use of AI to build trust with users?

For Example: Describe the level of transparency about the data used to train your AI system and the algorithms it employs?

Empowering

How is empowerment ensured for individuals and society through the development and application of AI?

For Example: Describe how the AI system empowers users and stakeholders to make informed decisions?

Privacy

How is privacy protected in the collection and use of data for AI development and deployment?

For example: Describe how the AI system protect, for example, the privacy of individuals whose data is used?

Robustness

How is robustness ensured in the AI system to make it resilient against adversarial attacks and unexpected inputs?

For example: Describe how is ensured that the AI system is robust against adversarial attacks and unintended consequences?



Potential Risks

What are the potential risks and how is the risk classified?



Outline potential biases that could arise from the data, algorithms, or deployment of the AI system.

Outline potential challenges in understanding and communicating the decision-making process of the AI system to users and stakeholders.

Outline potential issues related to data privacy, algorithm transparency, and accountability in the AI system.

Outline potential negative impacts on individuals or society due to the development or deployment of the AI system.

Outline potential privacy risks associated with data collection, storage, processing, and use in the AI system.

Outline potential vulnerabilities and weaknesses in the AI system that could lead to errors, biases, or malicious attacks.

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks

- ☐ Minimal risks
- ☐ Limited risks
- ☐ High risks
- ☐ Unacceptable risks



Mitigation Strategies

Develop comprehensive mitigation strategies to address each identified risk, including preventive measures, contingency plans, and monitoring mechanisms

What is the mitigation strategy to address this specific risk?

What is the mitigation strategy to address this specific risk?

What is the mitigation strategy to address this specific risk?

What is the mitigation strategy to address this specific risk?

What is the mitigation strategy to address this specific risk?

What is the mitigation strategy to address this specific risk?