

Strategy: the Digital Ethics Canvas

Fundamentals of the strategy

What is the goal of this strategy?

Reduce ethical debt by systematically assessing **ethical risks** at all stages of the design and development process.

When to use it?

- During the **design** and the **development** stages.
- **Repeat** the analysis several times **throughout the engineering process**.

Assessment questions

The strategy includes assessment questions under **five ethical lenses**: Welfare, Fairness, Autonomy, Privacy and Sustainability. See below and/or refer to the Digital Ethics Canvas (see links at the bottom).

Applying the strategy

1. **Evaluate the benefits**: List expected benefits
2. **Evaluate the risks**: Identify risks using the five ethical lenses, evaluate their probability and severity
3. **Reduce the risks**: For all risks, especially the highest, outline mitigation measures
4. **Review**: Assess if benefits outweigh risks and decide whether/how to proceed with the project
5. **Repeat**: Continuously reassess ethical risks

Evaluating & reducing risks using the Digital Ethics Canvas

WELFARE

RISK

- Can it be used in harmful ways? Towards vulnerable populations?
- Kind of impacts from errors?
- Type of protection against attacks/misuse?

MITIGATION

FAIRNESS

RISK

- How accessible?
- Kind of biases may affect results?
- Different outcomes for different users/groups?
- Can contribute to discrimination?

MITIGATION

SUSTAINABILITY

RISK

- What is the carbon footprint?
- Types of resources it consumes and produces?
- Type of human labor involved?

MITIGATION

AUTONOMY

RISK

- Understand how it works + limits?
- Choices in use (consent, settings)?
- Affects user autonomy/agency (manipulation, influence)?

MITIGATION



PRIVACY

RISK

- Data collected? Personal/sensitive data?
- Who has access to the data?
- How is the data protected?
- Does it disclose/can it be used to disclose private information?


MITIGATION

Evaluate risks

1. **Identify the risk**: what could go wrong and for whom?
2. **Evaluate the level of the risk**:
 - a. Probability: low/mid/high 
 - b. Severity: low/mid/high 
 - c. Resulting level:

		Severity		
		low	mid	high
Probability	low	low	low	mid
	mid	low	mid	high
	high	mid	high	high
















Mitigate risks

1. **Identify mitigation possibilities**: can the risk be reduced? If yes how and by whom?
2. **Evaluate the mitigation potential**: low/mid/high 

To remember

- Some risks may relate to **more than one lens**.
- Some risks **may not be covered**.
- The **process** is more important than the result.
- **Research** may be necessary to evaluate risks.
- Predicting risks is always uncertain and it is **safer to overestimate** risks than to miss real ones.

Strategy: the Digital Ethics Canvas

DIGITAL ETHICS CANVAS		
CONTEXT	SOLUTION	BENEFITS
<hr/> <hr/>		
WELFARE		
RISK <ul style="list-style-type: none">Can the solution be used in harmful ways, in particular with regards to vulnerable populations?What kind of impacts can errors from the solution have?What type of protection does the solution have against attacks or misuse?  	MITIGATION	
FAIRNESS		
RISK <ul style="list-style-type: none">How accessible is the solution?What kinds of biases may affect the results?Can the outcomes of the solution be different for different users or groups?Could the solution contribute to discrimination against people or groups?  	MITIGATION	
AUTONOMY		
RISK <ul style="list-style-type: none">Can users understand how the solution works and what its limits are?Are users able to make choices (e.g. consent, settings) in their use of the solution and how?How does the solution affect user autonomy and agency?  	MITIGATION	
PRIVACY		
RISK <ul style="list-style-type: none">What data does the solution collect?Is it collecting personal or sensitive data?Who has access to the data?How is the data protected?Could the solution disclose / be used to disclose private information?  	MITIGATION	
SUSTAINABILITY		
RISK <ul style="list-style-type: none">What is the carbon footprint of the solution?What types of resources does it consume (e.g. water) and produce (e.g. waste)?What type of human labor is involved?  	MITIGATION	

This work is licensed under CC BY-SA 4.0

Digital Ethics Canvas 2024 - C. Hardebolle, V. Macko, V. Ramachandran, Y. Du, N. Bartemiejczyk, A. Haber, P. Jermann

