Strategy : Harm Modelling

Fundamentals of the strategy

What is the goal of this strategy ?

Anticipate and address potential harms that users and stakeholders of the software may encounter.

When to use it ?

This analysis should be done at every stage of the project *(and if possible for every feature)*, until the harm potentials are satisfactory.

Assessment question

How could my tech **harm** the **users** and **other stakeholders** ?

Applying the strategy

3 elements are needed to be considered to evaluate the potential harms generated by a system:

1. The category of use
2. The types of harms
3. The magnitude of harm

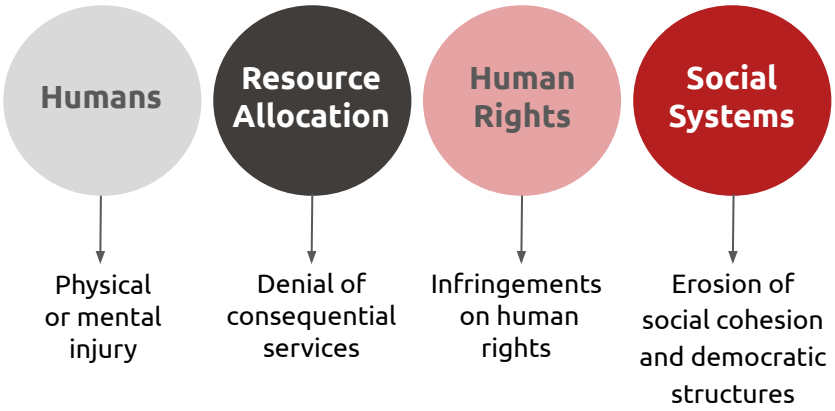
1. Category of use

Harm can arise in 4 categories of use:

- **Malfunction:** How harm could be caused if my system failed to function properly ?
- **Misuse/abuse:** How would someone use features of my software to cause harm?
- **Unintended use:** How using a feature of my software in unexpected ways could cause harm?
- **Intended use:** How would someone use my software correctly and experience harm?

2. Types of Harm

Several taxonomies of harm exist.
We have adapted the **taxonomy of harm** of Microsoft, which focuses on **what is harmed**, as illustrated below.



Not all categories always apply, so you should use the categories that best fit your specific context. This strategy needs to be adapted to the particular project you are working on.

3. Magnitude of Harm

In a second step, it would be helpful to qualify the magnitude of the identified harms. This can be done in terms of :

- **Severity:** how acutely could the technology impact an individual or group's well being ?
- **Scale:** how broadly could the impact on well being be experienced across populations or groups ?
- **Probability:** How likely is the technology to impact individual or group's well being ?
- **Frequency:** How often would an individual or group experience an impact on their well-being from the technology

Strategy : Harm Modelling

Table of harms

This table can be used to identify how the solution you are evaluating could impact stakeholders in harmful ways.
This table has been adapted and simplified from: Types of harm - Azure Application Architecture Guide | Microsoft Learn

Category	Type of harm
Humans	**Physical injury:** technology could hurt people or create dangerous environments *Ideas: errors/failures from software in physical systems, impacts on health, violence incitement…*
	Emotional or psychological injury: technology could lead to severe emotional and psychological distress *Ideas: impacts on identity, reputation, mental health, addiction phenomena…*
Allocation of Resources	**Opportunity loss:** automated decisions could limit access to resources, services, and opportunities essential to well-being *Ideas: discrimination from software in employment, housing, education…*
	Economic loss: automating decisions related to financial instruments, economic opportunity, and resources can amplify existing societal inequities *Ideas: differential pricing, discrimination in access to credit, economic exploitation…*
Human Rights	**Dignity loss:** technology can influence how people perceive, recognize and value one another as humans *Ideas: use of software for public shaming, humiliating outcomes, dehumanization*
	Liberty loss: automating legal, judicial, and social systems can have an impact on people's freedoms and autonomy *Ideas: errors in policing software, social/crowd control, interference with free speech…*
	Privacy loss: the information generated by our use of technology can be used to determine facts or make assumptions about someone without their knowledge *Ideas: private life tracking and interference, forced collection of personal data…*
	Environmental impact: the environment can be impacted by every decision in a system or product life cycle, from resource extraction to waste generation *Ideas: carbon emissions from execution, resource exploitation, electronic waste…*
Social Systems	**Manipulation:** technology's ability to create highly personalized and manipulative experiences can undermine an informed citizenry and trust in societal structures *Ideas: disinformation, behavioral exploitation, deception…*
	Social detriment: technology can affect social structures and have negative effects on people's relationships in communities and in society at large scale *Ideas: deskilling, stereotype reinforcement, political polarization, social conflicts…*