The Moral Compass

A tool for responsible projects

by Gabriella Cinque

The Moral Compass

What for?

A moral compass is a natural feeling that makes people know **what is right and wrong** and how they should behave.

When designing for technology, we may find ourselves lost without knowing anymore if we are taking the right decisions or not.

This tool aims to provide **a moral compass for technologists** by matching ethical theories with possible stakeholders and categories of potential harm, in order to bring up questions when approaching the design of a service or product and therefore anticipate risks.

Turn the wheel and see where it brings you!

Photo by Valentin Antonucci on Pexels.

How to play

1-10 players / 60-90 minutes / Paper + pencils

1. Read out loud a **design ficiton set-up scenario** related to the topic / product / service you want to explore that you had already prepared.

Those who are not comfortable with Ethical theories, you can simplify the exercise by matching only Stakeholders with Harms.

2. Turn the wheel randomly to have an **ethical theory + a type of harm + a stakeholder** in a row. Read out loud the **description** for the selected ethical theory, type of harm and stakeholder.

4. Ask participants to write down alone **scenarios** of what could happen under these circumstances. This step should not take more than 5 minutes.

Encourage participants to think about the experiences and perspective of the stakeholder. Ask them to be specific.
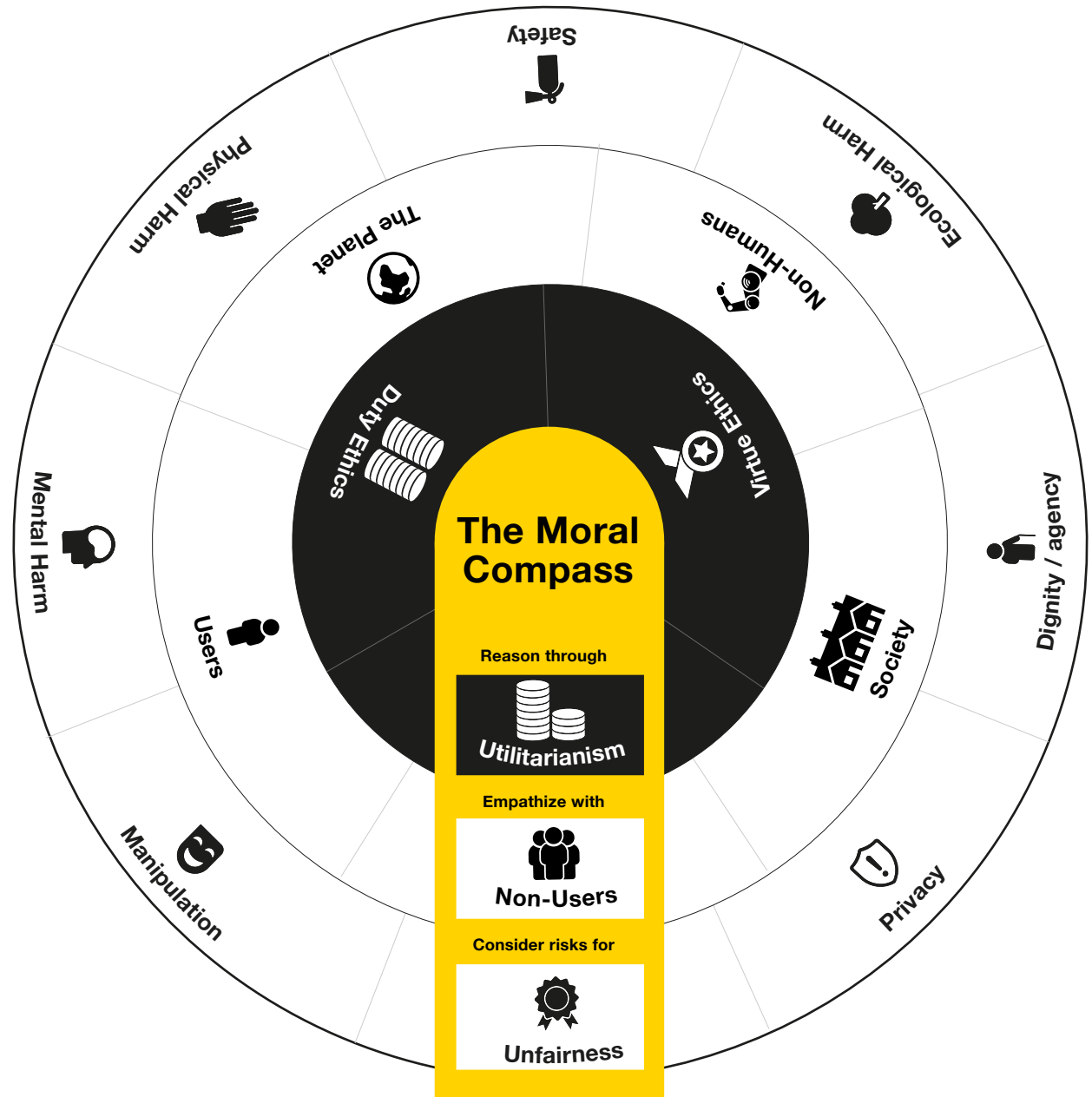
5. Let each participants past their scenarios on a board and ask them to **share** with the rest of the group. Keep a few minutes at the end to let participants **react** on what has been shared.

If you cannot phisically turn the wheel (online workshop), pick 3 items randomly and propose them to the group.

6. Turn the wheel **again**, from 3 to as many times as you wish.

The Moral Compass

Reason through
Utilitarianism

Empathize with
Non-Users

Consider risks for
Unfairness

Harms

Physical Harm

Biggest **death & injury risks** come from accidents or misuse.
Consider also impact on **diet** and **sleep**.
Repetitive actions can also bring to **strain injuries**, as well as overexposure to screens can lead to eye strains.

Mental Harm

Anxiety, **depression**, **addiction**...
Technology can also exacerbate social **exclusion** and cause harm to **relationships** or to **social image**.
Exposure to **objectionable content** (extreme porn, gore, abuse) can cause significant distress.
Mental harm can affect content moderators as well as users.

Manipulation

Huge topic in recent years, focusing on electoral interference, **bots**, **troll farms**, and **fake news**. Risk is increased by **synthetic media** – generated audio and video.
Tech companies also manipulate users through **dark patterns**.
User research and adaptive AI may be unethical if they treats **users as subjects** for behavioural modification.

Safety

Give users **safety controls** (Block, Mute, Report, Ban, etc). But safety tools can also become vectors of abuse: domestic IoT devices offer an emerging vector of spying and harassment. For this reason, safety risks need to be tackled as a product/design + **policy** issue. Frequently overlooked by homogenous teams. Children and other **vulnerable users** have additional safety needs, and often applicable legislation.
Consider also **stress cases**.

Unfairness

Discrimination can be found in pricing, equality of access, denial of service, categorisation, algorithmic bias and more. This is a **social issue** needing social approaches.
Defining **fairness** is important but tough: one measure of fairness might make another worse.
Also examine **power imbalances**: could unfairness exacerbate these?

Privacy

Violations of privacy can come from individuals, companies, authorities...
Compliance with law (GDPR, CCPA, etc) is important, but so are user norms and expectations, which change rapidly. Also, **policies** are not the same everywhere.
The topic is complex for **AI**: we need to collect enough data to train the AI and so improve the service, but at the same time, we can't collect users' data if not essential to provide a service to this user.

Dignity / agency

Studies have shown people **'overtrust'** robots in emergency situations. A right to **explanation** is a key part of democracy: why was I arrested? Why has this candidate won? Automated decisions could erode this principle, while also devaluating expertise and putting people out of jobs, harming their self-worth and dignity.
AI should always be explainable.
People deserve opportunity to **contest & remedy decisions**.

Ecological Harm

Harming the planet means crossing the lines of the **9 limits of the Earth**:

1. Reduction of biodiversity
2. Climate change exasperation
3. Alteration of the balance of phosphorus and nitrogen
4. Land use
5. Pure water availability
6. Ozone layer preservation
7. Ocean acidification
8. Pollution by new substances
9. Toxic smoke emissions

Stakeholders

Users

Typically **direct** stakeholders, interacting directly with your technology.
Respect and embrace user diversity for best results.

Not all use is **consensual**: think about Facebook shadow profiles, Covid-tracker apps, corporate monitoring…
Roles are **dynamic**: people might switch from user to non-user.
Also try to consider targeted VS not targeted roles, expected VS atypical use.

Non-Users

Usually indirect stakeholders – do not interact directly with the technology but may be affected nevertheless.
A blind spot in user-centred design.

Remember to consider also **excluded** stakeholders, those who cannot or do not use the technology. Reasons for exclusion can include physical, cognitive, social or situational constraints.
For example, a technology that relies heavily on visual elements will exclude stakeholders with low-vision.

Society

It's easier to think of society as groups of people: neighbourhoods, citizens, workers, voters…
May be represented by advocacy groups or activists.

Technology can affect not just individuals and groups of people, but **concepts** like democracy, freedom of expression, public health, literacy, equality…
This is more relevant at large scale.

Non-Humans

What about robots & AIs, do they deserve moral consideration? Could they suffer? Would it be real suffering, or just mimicked suffering?
What about living beings, like animals, or plants?
What new ways of interaction between designers and non-humans can be imagined?

Considering non-humans means switching from an homocentric view to a **biocentric anti-speceist** vision.

The Planet

We can identify the planet with specific **ecosystems**, perhaps starting from those present locally and then building links to related ecosystems (for example the ecosystem of the river near a factory, but also that of the surrounding flora with which it interacts).

Tech industry is a significant contributor to greenhouse emissions. Resource depletion comes mostly from **manufacturing** and **use** phase of the life cycle.

Ethical theories

Each ethical theory has its own caveats. In order to explore risks further, it is important to consider scenarios through multiple lenses.

Duty Ethics

Are you acting accordingly to your duties, no matter what?
You should do certain things, because they are the right thing to do.

Deontology judges human practices based on whether they are consistent with certain **duties** identified. It emphasizes the **principles** behind actions rather than consequences.
Acting rightly requires being motivated by proper universal principles that treat everyone with respect. Principles should be set and agreed at the beginning of each project, and taken as a reference all along.

Some caveats:

There may be disagreement about the principles involved in the decision: what's the right duty?

What if the «right» choice has bad consequences?

What if one's duty is in conflict with that of someone else?

Utilitarianism

What is the greatest good for the greatest number of people?
You should do certain things, because they produce good consequences.

Consequentialism judges humn practices based on their **consequences**, so the amount of happiness and suffering created by a person's actions is what really matters.
Making a decision relates then to weather it will benefit the **majority**.
Sometimes you may even need to break some of the traditional moral rules to achieve such an outcome.

Some caveats:

It is not always possible to predict the outcome of a decision.

Consider that harming a minority and benefitting a majority does not build mutually beneficial relationships.

Virtue Ethics

What makes it good?
You should do certain things, because they are examples of good character.

Instead of focusing on the consequences of actions or duty fulfillment, virtue ethics takes **virtues** (qualities of moral character) as fundamental to the ethical life. For Aristotle, a virtue is an excellence of a person's functioning in a certain area of life, a **stable character trait** that governs a person's actions in some respect.
This requires you to understand what virtues are good for your product, service, project, team, etc. and then decisions are made in light of those particular virtues.

Some caveats:

This does not consider obligations towards stakeholders nor the impacts.

Some virtues may be opposite and conflictual. Ex. Should you be loyal to the company or honest to the customers?

Example

Let's say you are working on a project aiming to build a new website to track goods transportation in France.

1. The set-up scenario → 2. Turn the wheel → 3. Write your own specific scenario

France, summer 2024.

Life has returned to normal after years of pandemic, the Olympics bring new enthusiasm and great dynamism.

Goods and people are moving frantically like never before.

Just one year ago, the Logistics Flow Observatory had just been deployed, and now we are already seeing the consequences ...

You get:

- 🗣 **Mental harm**
- 👥 **Non-users**
- 🪙 **Duty Ethics**

Maybe you can ask yourself:

Who are non-users here? What kind of mental harm they could get if they were under a duty ethics perspective?

You get:

- 🧍 **Dignity / Agency**
- 🧍 **Users**
- 🪙 **Duty Ethics**

Maybe you can ask yourself:

How could our users lose dignity or agency while using the service? According to our principles, what duties should we perform to prevent this risk?

You get:

- ✋ **Physical harm**
- 🤖 **Non-Humans**
- 🏅 **Virtue Ethics**

Maybe you can ask yourself:

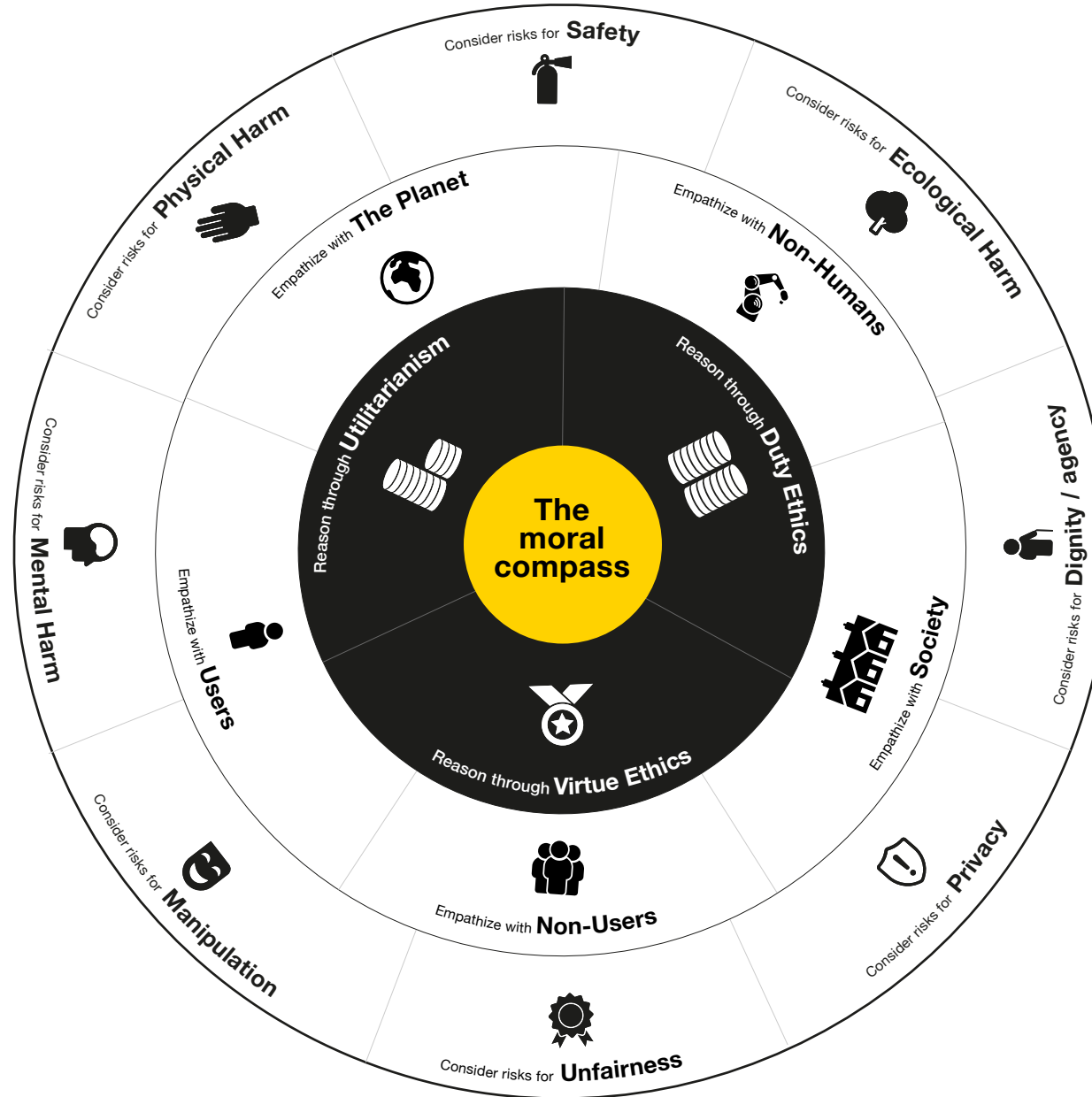What qualities should the web platform have in order to avoid physical harm to humans?

Carl has been driving his truck for 15 years. He's seen all sorts of things, but this time it's too much! His company manages to accurately track all movements, including those of Carl, and this stresses him to death. They assured him that the data they collect is anonymous, but nobody can be trusted these days. Carl has become suspicious, paranoid. He feels compelled to be flawless and constantly risks accidents to maximize his travels, fearing repercussions from the company.

And eventually all the companies adopted the Logistics Flow Observatory. It was so convenient to have all the information ready and concise at hand, that over time they stopped asking questions, making decisions only with respect to what the platform told them. The Observatory therefore had an enormous responsibility and duty to provide reliable information, because highly strategic decisions were taken without ever challenging.

I, the Logistics Flow Observatory, am very prud of myself.

Thanks to my ability to deduce habits from journeys, and to my capacity to give good advice, the working conditions of truck drivers have significantly improved, workplace regulations are more respected and there are fewer road accidents. I know that this was not meant to be my job, but what can I do about it: I'm really someone who cares!
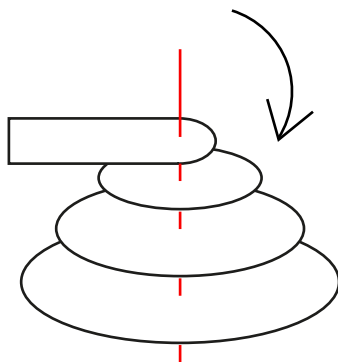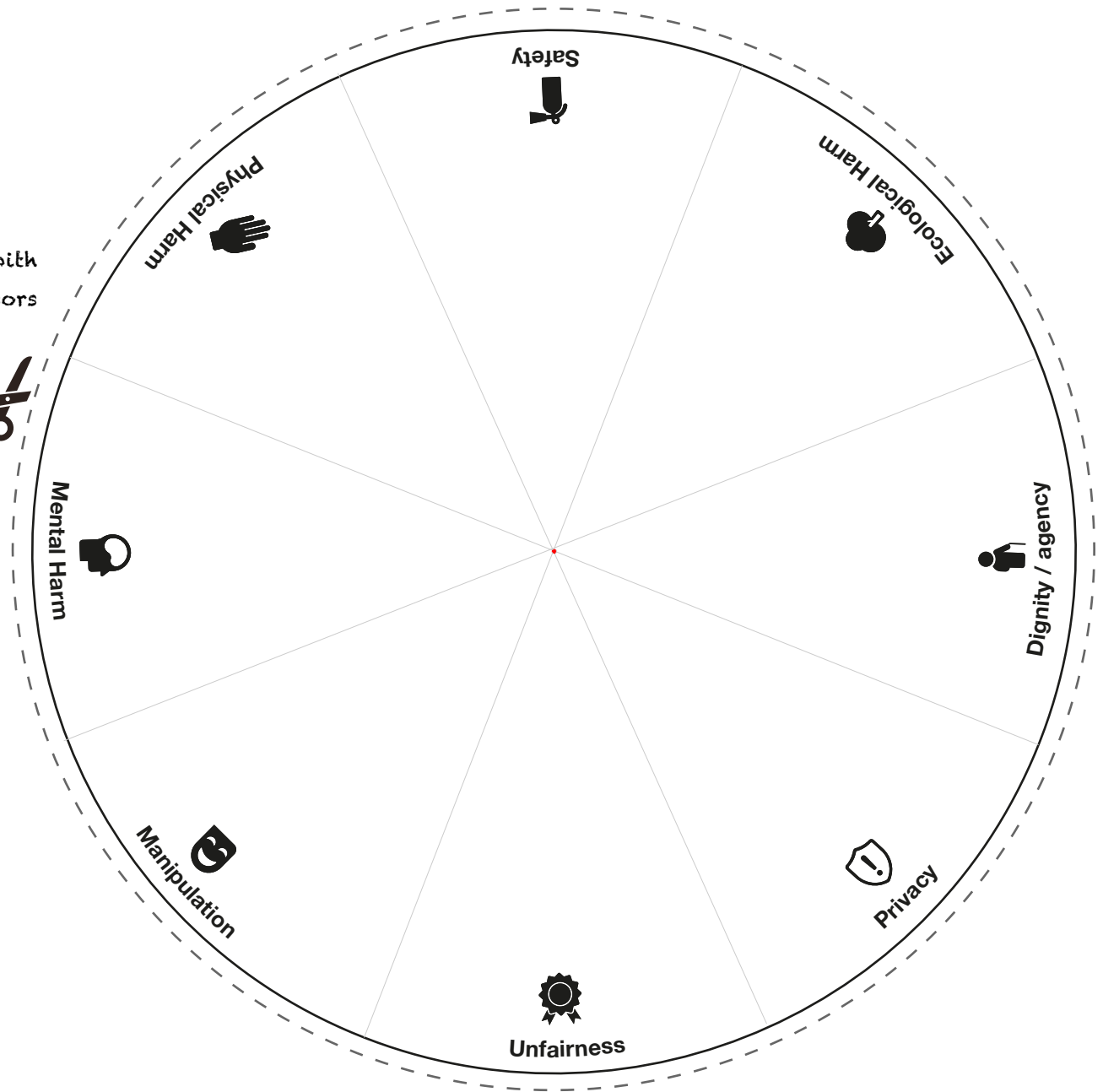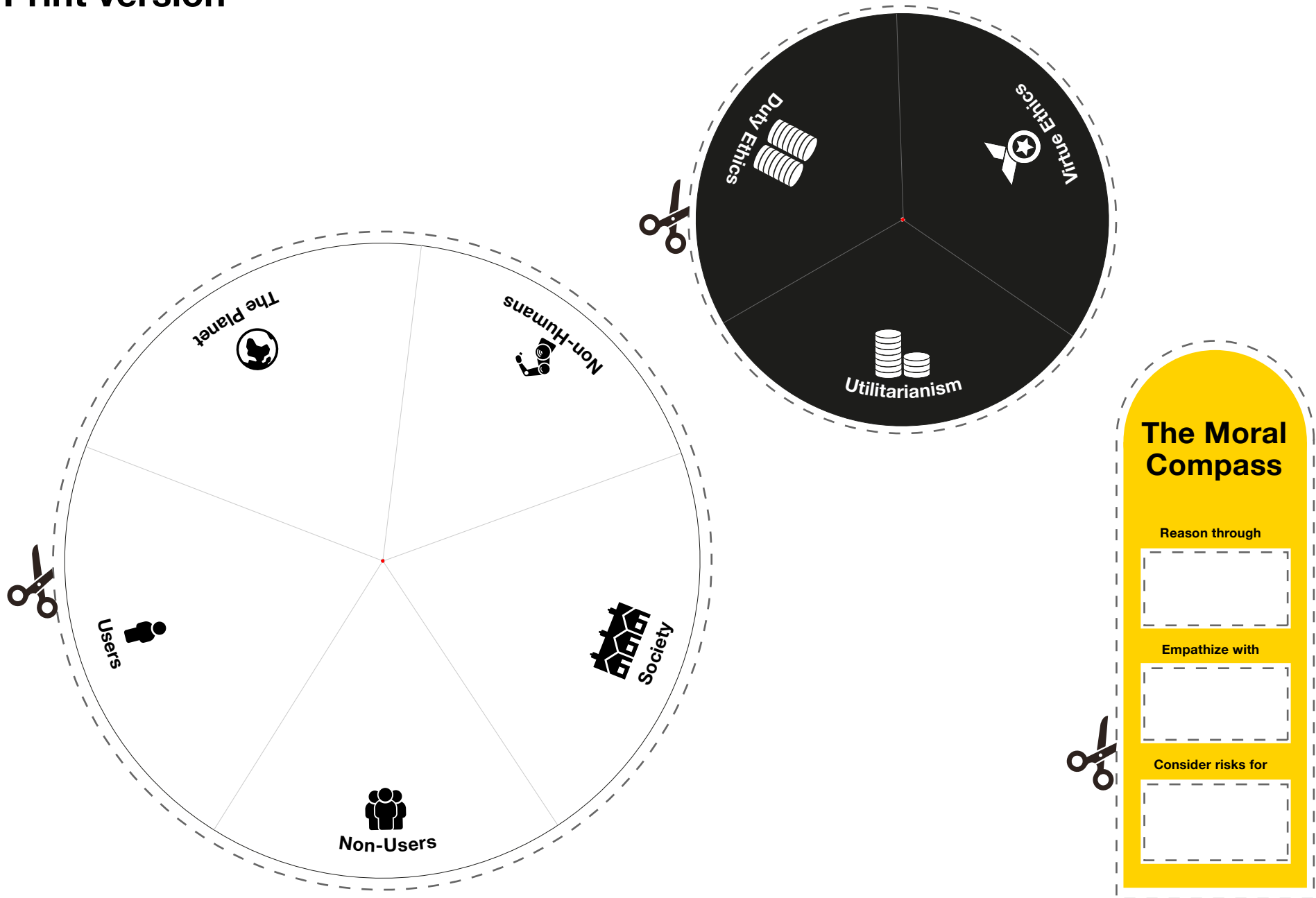
Digital version

Consider risks for **Safety**

Consider risks for **Ecological Harm**

Consider risks for **Physical Harm**

Empathize with **The Planet**

Empathize with **Non-Humans**

Consider risks for **Mental Harm**

Reason through **Utilitarianism**

Reason through **Duty Ethics**

Consider risks for **Dignity / agency**

The moral compass

Empathize with **Users**

Empathize with **Society**

Reason through **Virtue Ethics**

Consider risks for **Manipulation**

Empathize with **Non-Users**

Consider risks for **Privacy**

Consider risks for **Unfairness**

Print version

Cut the dotted line with a cutter or scissors

Use a scrapbooking brad to keep the four layers together and allow them to spin

Safety

Physical Harm

Ecological Harm

Mental Harm

Dignity / agency

Manipulation

Privacy

Unfairness

Print version

Virtue Ethics

Duty Ethics

Utilitarianism

The Planet

Non-Humans

Users

Society

Non-Users

The Moral Compass

Reason through

Empathize with

Consider risks for

This tool has been developed by Gabriella Cinque and takes inspiration mostly from lessons and reading of *Future Ethics* by Cennydd Bowles, but also *The Future Computed: Artificial Intelligence and its role in society* by Brad Smith or *Value Sensitive Design* by Batya Friedman.

It is the result of field work trying to adopt responsible approaches during projects.

The wheel mechanism comes from the wheel of emotions, a tool used especially for autism to help people give voice to their feelings.

Any questions?

gabriella.cinque.design@gmail.com