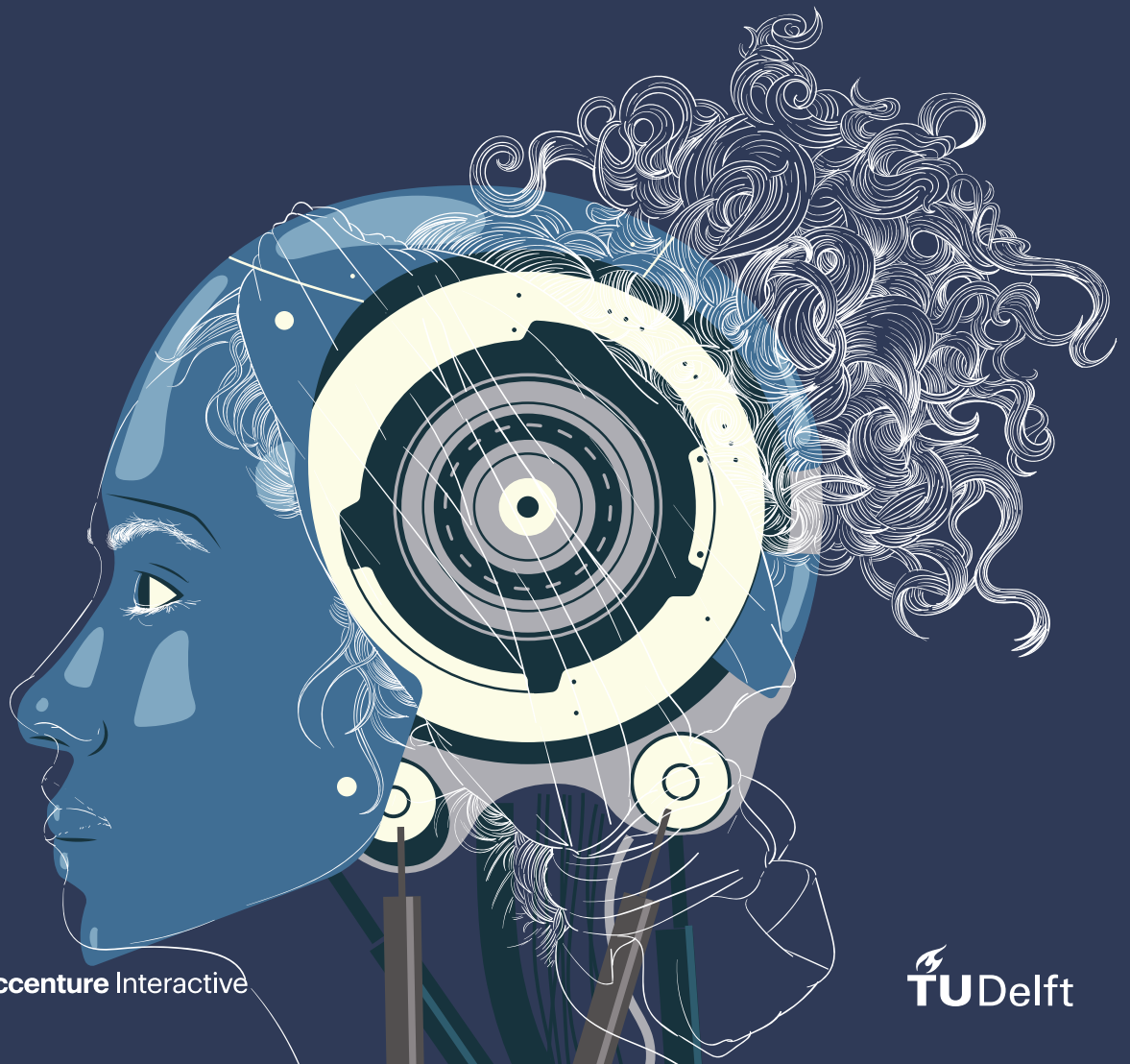


Ethical Toolkit for the Development of AI Applications

Toolkit Material

Mario Alberto Sosa Hidalgo
2019



Author

Mario Alberto Sosa Hidalgo
mariososadi@gmail.com

August, 2019

** The cover was designed using illustrations from Freepik.com

Table of Contents



STRATEGIC BLUEPRINT

Following the right steps for a valuable ethical assessment of an AI application



ETHICS & [A]I MINI-WORKSHOP

Understanding ethics and its relationship with Artificial Intelligence



THE EVIL IN [A]I GAME

Exploring the current ethical situation of AI in an extreme way



AI APP PROJECT OVERVIEW

Defining the vital ethical areas to consider on the project by making concise moral agreement based on the ethical risks



THE RESPONSIBLE ARTIFICIAL INTELLIGENCE DECK

A tool for triggering awareness, discussion, and solutions about the ethical risks of the development of AI



ETHICAL RISK CARDS

Translating the uncovered ethical risks of the AI project into tangible insights



ETHICAL EVALUATION AXES

Qualitative assessment of the ethical risks detected using the Responsible AI Deck



PROJECT'S MORAL CODE

Defining the vital ethical areas to consider on the project by making concise moral agreement based on the ethical risks



Strategic Blueprint of the Ethical Toolkit

*Following the right steps for a valuable
ethical assessment of an AI application*



STRATEGIC BLUEPRINT

Ethical Toolkit for the Development of AI Applications



Educating Phase (Team & Client)

1 Ethical Alignment



Ethics & [A]I Mini-workshop

The ethical alignment workshop is a strategic feature of the toolkit that aims to educate and ethically align both the client and the team involved in the AI application project. The deck explores the topic of ethics and its relevance in an AI context.



The Evil in [A]I Game

This module, which has the format of a game, explores the current ethical situation of AI in an extreme way. It aims to detonate a change of perspective around AI and to strengthen the knowledge acquired during the "Ethics & AI" workshop.

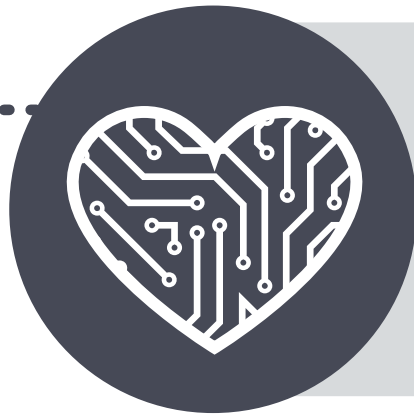
Executing Phase (Project Team)

2 Project Vision & Values



AI Project General Checklist

The AI Project Checklist is a detailed and pictorial way of discussing the new AI application, thinking about the stakeholders involved, the type of AI models used, the ethical principles to be considered for the development team and the client.



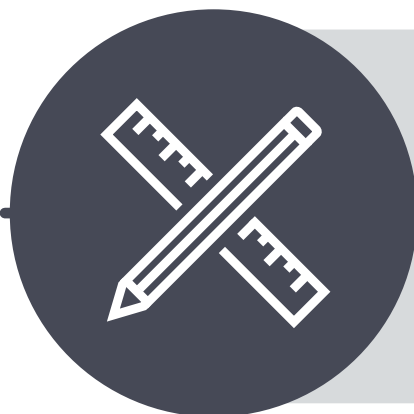
The Responsible Artificial Intelligence Deck

The Responsible Artificial Intelligence Deck, is a deck of cards intended to trigger ideas regarding the ethical risks that could occur with the development of an AI app. Furthermore, this deck is designed to help AI Applications designers and developers to think about the Ethics of AI and the impacted stakeholders in a more visual way.



Ethical Risks Cards

The Ethical Risks Cards objective is to write down the possible ethical risks ideated using the Responsible AI Deck that could be evaluated later on using the Ethical Evaluation Axes canvas.



Ethical Evaluation Axes

Based on the scale of the axes, or on areas of the canvas that are important for the client, it would be decided which specific ethical risks should include in the decision-making processes of the project. This gives a a holistic view of the ethical risks of the implementation and the principal ethical areas to take into consideration.

3 Moral Code or Best Practices



Moral Code of the Project

By defining the ethics of the project, the team will integrate the most relevant risks in a reflection manner for an AI application. This is an structured way of to culminating the ethical strategy by making concise moral agreements and tangible statements.

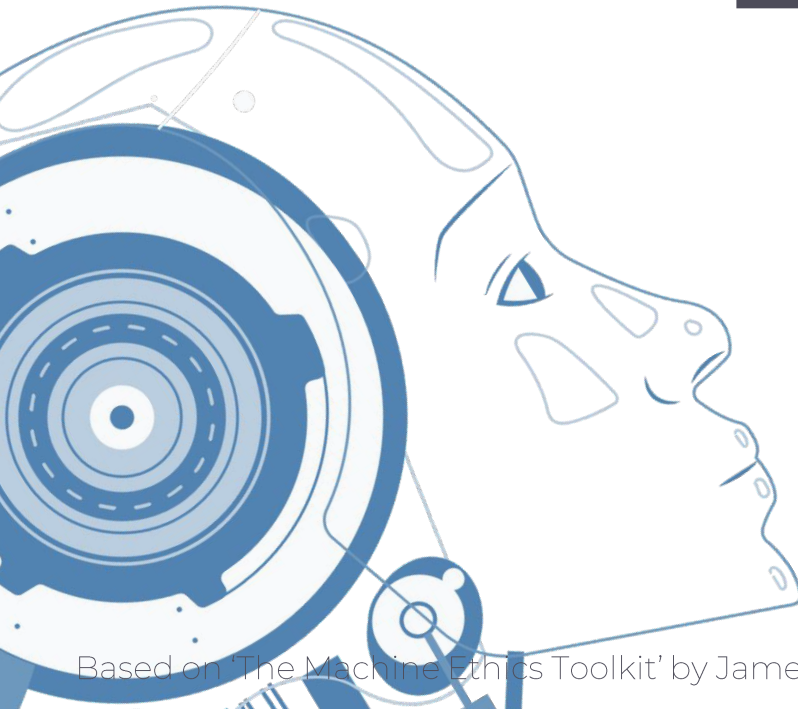


Ethics & [A]I Mini-Workshop

*Understanding ethics and its
relationship with Artificial Intelligence*



Ethics & [A]I Workshop



Let's introduce ourselves.

Activity 1 (*5 min*)
Your worldview

If **YOU** were an autonomous car,
where in the world would you like
to go?

What is **ethics** exactly?

Ethics is mostly focused on the
normative issues that help us **decide**
how to act in a spectrum of situations.

What is the **Ethics of AI**?

Formally speaking, AI ethics is
commonly classified into
roboethics and **machine ethics**.

Roboethics: branch of ethics of technology that concerns the morals of humans as they develop and use AI.

Why the Ethics of AI is **relevant**?

Societal consequences (risks) of
implementing AI in human-use
applications.

For example: Labor automation,
human profiling, discrimination due
to statistical biases, etc.

What **will you learn** in this workshop?

1. **Three classic ethical** approaches.
2. Understanding the **importance** of ethical decision making.

Split in **groups of 3**

Activity 2 (15 min)

Quick Ethical Dilemmas

Best Friend's Big Day Dilemma

Your best friend is getting married in an hours time. You are already all fancy dressed up for the wedding when you discover that yours best friend's partner has been having an affair. You have concrete evidence to prove their guilt.

What would you do?

A. Tell your best friend, even if you ruin the wedding, to avoid them marrying to a cheater.

B. Say nothing since your role there is to be supportive with your friend's happiness.

Robin Hood Dilemma

You witness a man rob a bank, but instead of keeping the money for himself, he donates it to a local orphanage. You know this orphanage has been struggling for funding, and this money will allow the children to receive proper food, clothing and medical care. If you report the crime, the money will be taken away from the orphanage and given back to the bank.

What would you do?

A. Turn in the robber in to the authorities and make the right thing according to the law.

B. Say nothing as the money went to a good cause.

Ethical dilemmas **are tough**, man...

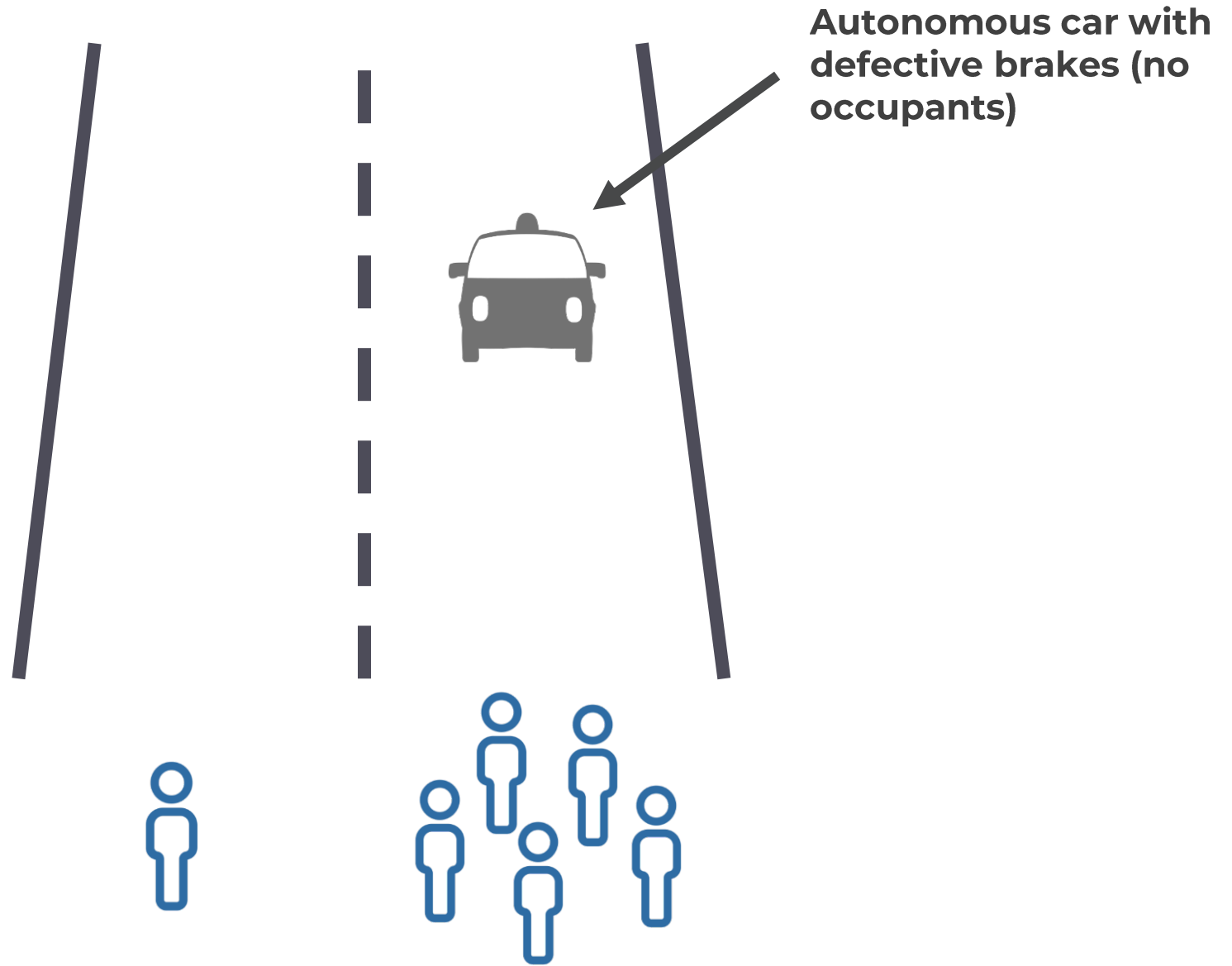
- 1. You have to make a decision**
- 2.** There are several decision choices for you to make
- 3.** There is no perfect solution

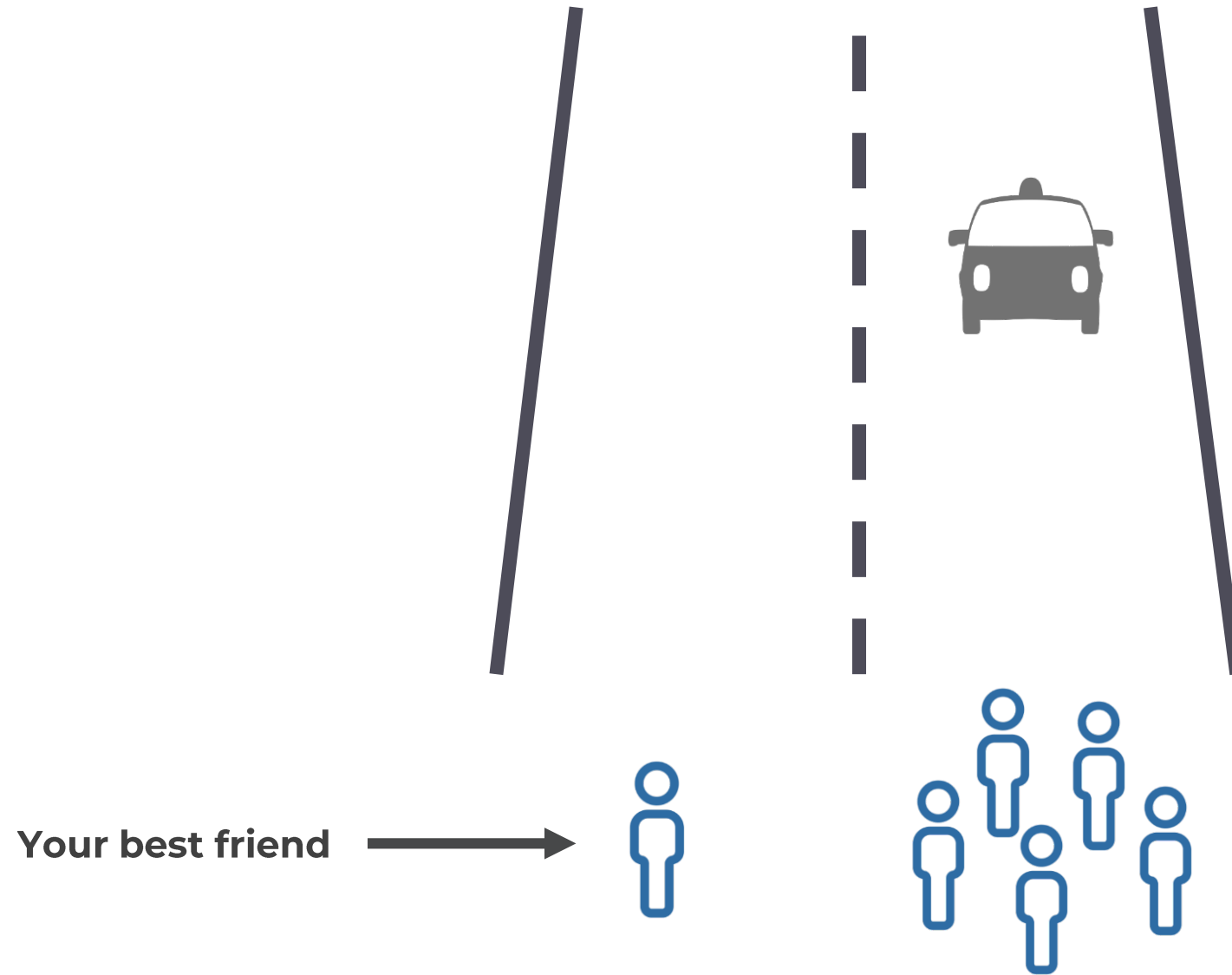
More or less like **ethics**...

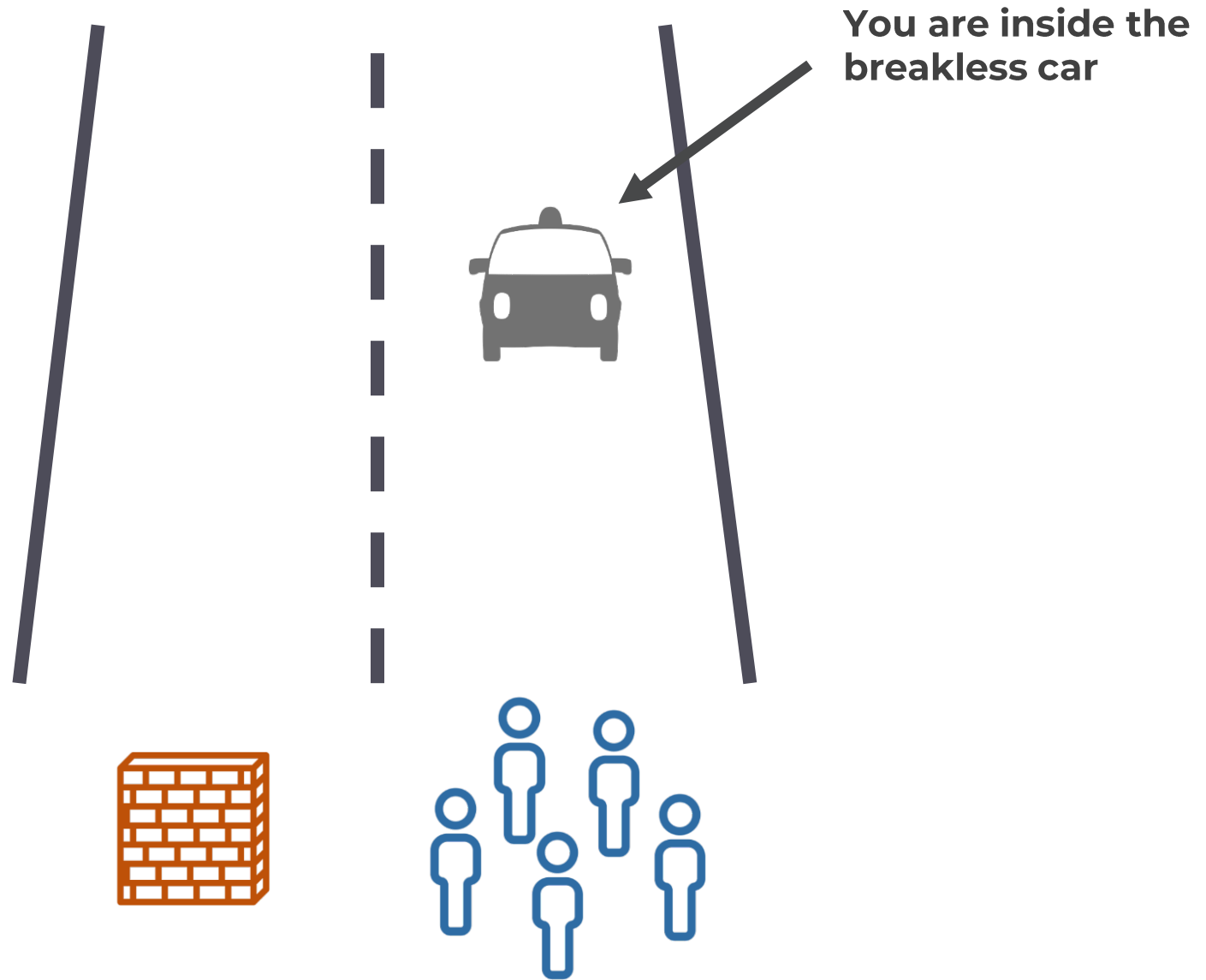
Ethics is **not** about the right or wrong choice, but
about the **more preferred and less preferred
choice in a determined context.**

Activity 3 (20 min)

The Trolley 'Autonomous Car' Problem







The most ethical decision according to...



What **did you learn** in this workshop?

1. **Three classic ethical** approaches.
2. Understanding the **importance** of ethical decision making.
3. Ethics is not about the right or wrong choice, but about the **more preferred and less preferred action in a determined context**.



Questions

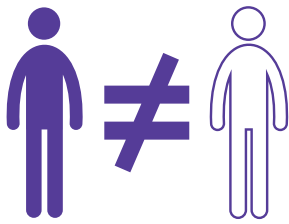


The Evil in [A]I Game

*Exploring the current ethical situation of
AI in an extreme way*



**Discriminates
because of
race**



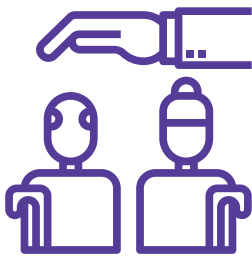
**Unsafe for
Users**



Sexist



**Discriminatory
against Elders**



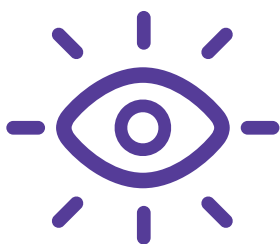
Dishonest



Deceptive



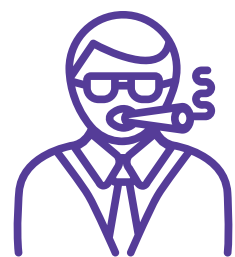
**Exposes Private
Data**



Unhelpful



**Has an
unaccountable
Creator**





THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I

AI Oncologist



AI Lawmaker



AI Judge



AI Doctor



AI Psychologist



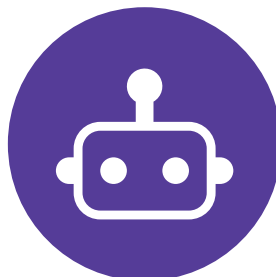
AI Hiring Software



AI Police Software



AI Customer Service Assistant (bot)



Automated Marketing Software





THE
EVIL
IN [A]



THE
EVIL
IN [A]



THE
EVIL
IN [A]



THE
EVIL
IN [A]



THE
EVIL
IN [A]



THE
EVIL
IN [A]



THE
EVIL
IN [A]

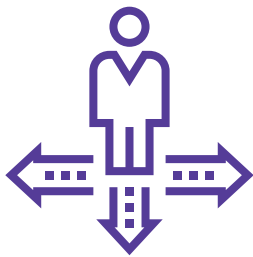


THE
EVIL
IN [A]



THE
EVIL
IN [A]

**Takes own
Decisions**



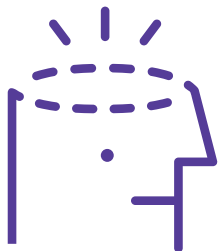
**Unsafe with
Data**



Unexplainable



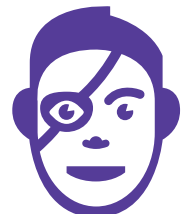
**Nobody knows
how it works**



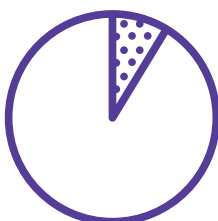
**Threats
people's lives**



**Creator
Exaggerated
Capabilities**



**Discriminates
against
minorities**



Manipulative



Untrustworthy





THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I

**Discriminates
because of
sexual orientation**



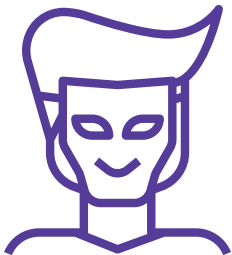
**Discriminates
because of social
background**



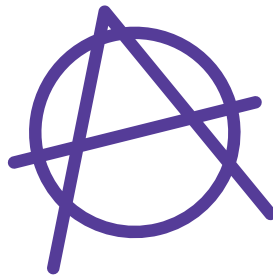
**Doesn't accept
feedback**



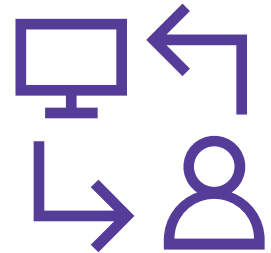
**Fakes a Human
Identity**



Anarchist



Destroys Jobs



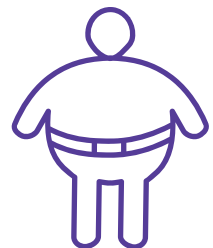
**Makes fun of
Users**



Scams people



**Discriminates
because of
physical
appearance**





THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I

Robot Caregiver



Intelligent Personal Assistant



Facial Recognition Software



AI-powered Education Platform



Smart Traffic Controller



Accounting Smart Algorithm



AI Insurance Policy Software



Smart Sales Assistant



Autonomous Vehicle





THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I



THE
EVIL
IN [A]I

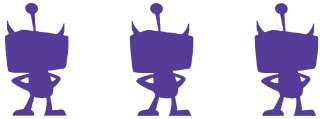
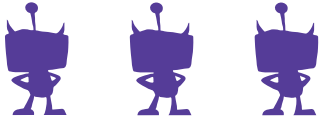
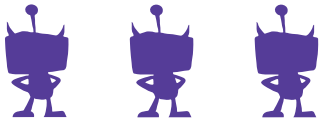


THE
EVIL
IN [A]I

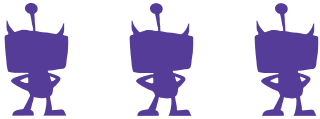
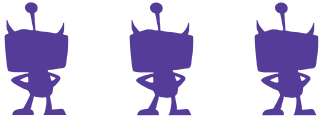
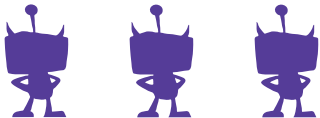
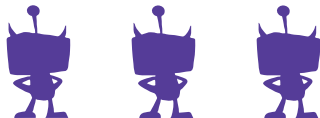


THE
EVIL
IN [A]I

SCORE CARD



SCORE CARD





AI App Project Checklist

Co-defining the ethical impact and ethical specifications of the project



AI APP PROJECT CHECKLIST



Name of the Poject

1 Goal



2 Team Accountable for App



Client

MOBGEN | Accenture Interactive

3 Context of the Project



e.g. AI application for healthcare

4 Stakeholders Impacted



Primary

Unexpected/Unintended

E.g. Business People

Eg. Partners, Customers, other businesses

5 Type of Impact



The AI application will have one or more areas of impact. e.g. AI-poered restaurant reviews impact reputation and financial health.

☐ Financial

☐ Property

☐ Privacy

☐ Emotional

☐ Reputation

☐ Liberty / Freedom

☐ Access to goods

☐ Life / Safety

☐ Rights / IP

6 AI Algorithm & Learning Model used



7 Data Type



Human Data



Non-Human Data

8 Data Source



GDPR compliant

Trustworthy

9 Ethical Principles Checklist (Order by relevance for project **High(+)** / **Med(0)** / **Low(-)**)



Data Privacy



Honest Communication



Human Well-Being



Data Safety



Accountability



Governance



Explainability



Fairness



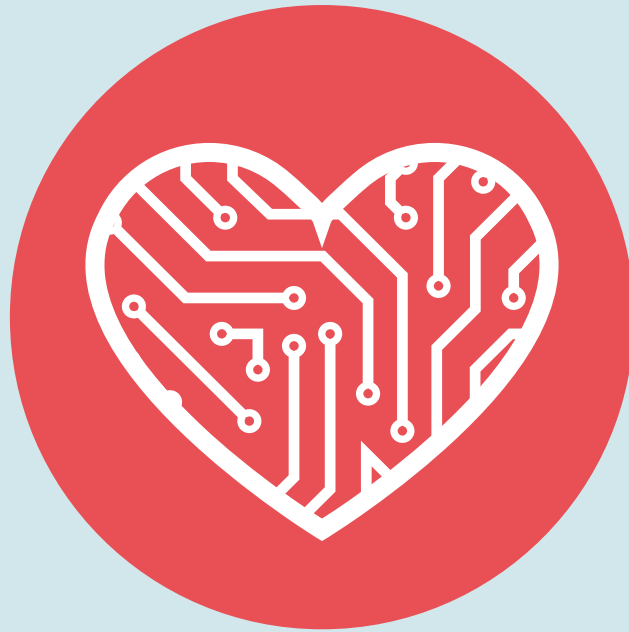
User Safety



Transparency



Value Alignment



The Responsible Artificial Intelligence Deck

*A tool for triggering awareness,
discussion, and solutions about the
ethical risks of the development of AI*



Stakeholders



Children



Stakeholders



Teenagers



Stakeholders



Elders



Stakeholders



Students



Stakeholders



Employees



Stakeholders



The Unemployed



Stakeholders



Business People



Stakeholders



Athletes



Stakeholders

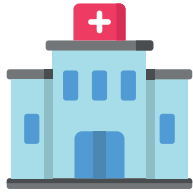


Healthcare Staff





AI Context

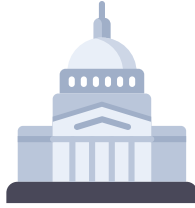


Healthcare

The adoption of AI in healthcare is growing radically. Examples include utilising AI to analyse unstructured data like photos, videos, physician notes to enable clinical decision making; use of intelligence interfaces to enhance patient engagement; and predictive modelling to manage patient flow and hospital capacity/resource allocation.



AI Context



Government

Applications of artificial intelligence to the public sector are broad and growing. It can be used to further public policy objectives (in areas such as emergency services, health and welfare), as well as assist the public to interact with government (through the use of virtual assistants, for example).



AI Context



Law

Current developments in AI in legal practice come in various applications. Some of the applications in this field are: due diligence (litigators perform due diligence with the help of AI tools to uncover background information), litigation prediction, legal analytics, and intellectual property management.



AI Context



Education

“Education will be profoundly transformed by AI,” says UNESCO Director-General Audrey Azoulay. Some of the most interesting applications of AI in Education are: teacher-AI collaboration, differentiated and personalized learning, universal access for all students, and tutoring outside the classroom.



AI Context

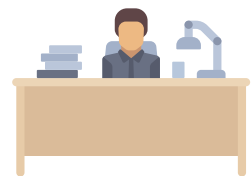


Sustainability

AI capabilities are being used in various ways to further societal goals. After the passage of Hurricane Harvey in 2017, many streets in downtown Houston were flooded and impassable—but others weren't. An AI application that combined satellite imagery with object detection software enabled rescue workers to identify safe escape routes for those trapped by the rising waters.



AI Context



The Workplace

AI is often used in the workplace to assist employers with recruitment through the use of algorithms to make hiring decisions. According to a 2017 survey by the talent software firm CareerBuilder, approximately 55 percent of U.S. human resource managers opined that AI will become a regular part of their work within the next five years.



AI Context



Customer Service

AI assists with this in a number of areas: by providing answers to simple customer questions 24/7, sorting messaging, prioritising queries to a human agent quickly where necessary, and by freeing up human agents to facilitate higher value calls.



AI Context

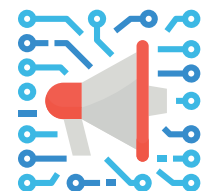


Sales

The biggest vendors from Adobe to Salesforce to Microsoft to Oracle are jockeying for position to bring automation and intelligence to sales. AI allows the salesperson to be able to handle more complex sales and be able to do it in a more expedient manner.



AI Context



Marketing

AI is having a real impact on content marketing – from creating incredibly detailed marketing strategies based on huge amounts of data to generating content automatically, marketers are starting to use AI more and more to help them every step of the way.



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



AI Ethical Principles



Data Privacy

Is the AI dataset compliant with GDPR?

AI companies should address the privacy considerations regarding user data as an input or output of any smart system. Some regulations might apply (i.e. GDPR)



AI Ethical Principles



Data Safety

Is the AI dataset GDPR compliant?

Developers should ensure that the data used to train the AI model won't produce harmful consequences for any of the stakeholders involved in the project. Some regulations might apply (i.e. GDPR)



AI Ethical Principles



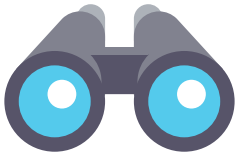
Explainability

Can it be explained why the AI system did certain action?

Every AI system has to be explainable enough for people to understand the reason why it made certain action.



AI Ethical Principles



Transparency

How does the system works?

AI systems should be transparent enough to show users the way they work and how they perform to make certain action.



AI Ethical Principles



User Safety

How is this AI system safe for humans?

Every AI application should be safe enough for all users, stakeholders involved in their context of use, or simply any person interacting with it.



AI Ethical Principles



Honest Communication

Are you clear and honest enough about communicating the real capabilities of the AI system?

AI companies should be honest, straightforward, and clear about the capabilities and evolution of their AI applications.



AI Ethical Principles



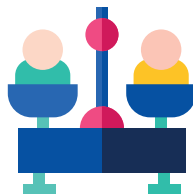
Accountability

Who is responsible of a bad output from the AI system?

Companies should be accountable for the possible mistakes their AI applications make, as well as to provide strategies for such failovers.



AI Ethical Principles



Fairness

Is this system ethically fair? (bias)

Developers must ensure that their AI applications are as unbiased as possible. Strategies for fairness failures should be addressed during the development of the AI applications.



AI Ethical Principles



Value Alignment

How aligned are the developers' ethical views with those of the stakeholders impacted?

Developers should align their ethical views with those of the stakeholders impacted with the AI system.



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



Human Values



Trust

To believe that someone is good and honest and will not harm you, or that something is safe and reliable.



Human Values



Calmness

That someone could be peaceful, quiet, or relaxed; without hurried movement or any type of anxiety.



Human Values



Identity

Who a person is, or the qualities of a person or group that make them different from others.



Human Values



Privacy

The right that someone has to keep their personal life or personal information secret or known only to a small group of people



Human Values

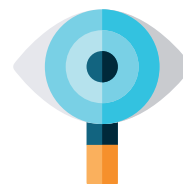


Universal Usability

The fact of something being easy to use for a broad spectrum of people



Human Values



Awareness

Knowledge that something exists, or understanding of a situation or subject at the present time based on information or experience.



Human Values



Support

To help someone emotionally or in a practical way.



Human Values



Freedom

The condition or right of being able or allowed to do, say, think, etc. whatever you want to, without being controlled or limited.



Human Values



Human Connection

The state of being related to someone.



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



**The
Responsible
Artificial
Intelligence
Deck**



Stakeholders



Artists



Risks



AI Ethical Principles



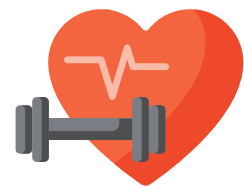
Governance

What is the level of autonomy of the AI system?

Developers must address the level of autonomy that an AI application would have and which bodies would control (or monitor) it when it is delivered.



AI Ethical Principles



Human Well-Being

Is this AI system intended to enhance humans or automate jobs?

To what extent an AI application is made to help humans improve their actions and not only to automate processes.



AI Context



Insurance

Claims management in auto insurance is one of the early use cases of AI application along with the insurance value chain. Major carriers have experimented with deploying AI to track and detect when motorists are engaging in distracted or unsafe driving. The ultimate goal is to predict driver patterns and the likelihood of future accidents, or for rewarding safe driving.



AI Context

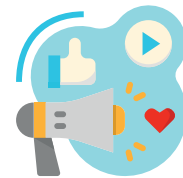


Banking

Large banks have a huge customer success burden, so they naturally look toward automation of customer service with chatbots. Financial institutions like hedge funds are chasing alpha with AI on top of new layers of data sources, and insurance companies are improving risk models with AI.



AI Context

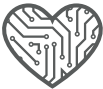


Social Media

With the large volume of communications out on the internet, it's just about impossible to filter out the communications that cause concern without completely slowing down the system or heavily controlling it. AI can provide more rapid understanding of this information and manage a company's social media footprint.



Opportunities



The Responsible Artificial Intelligence Deck



The Responsible Artificial Intelligence Deck



The Responsible Artificial Intelligence Deck



The Responsible Artificial Intelligence Deck



The Responsible Artificial Intelligence Deck



The Responsible Artificial Intelligence Deck





Ethical Risk Cards

*Translating the uncovered ethical risks
of the AI project into tangible insights*



ETHICAL RISK CARD



AI Context

Stakeholder(s) Affected

AI Ethical Principle(s)

Risk Identified

ETHICAL RISK CARD



AI Context

Stakeholder(s) Affected

AI Ethical Principle(s)

Risk Identified

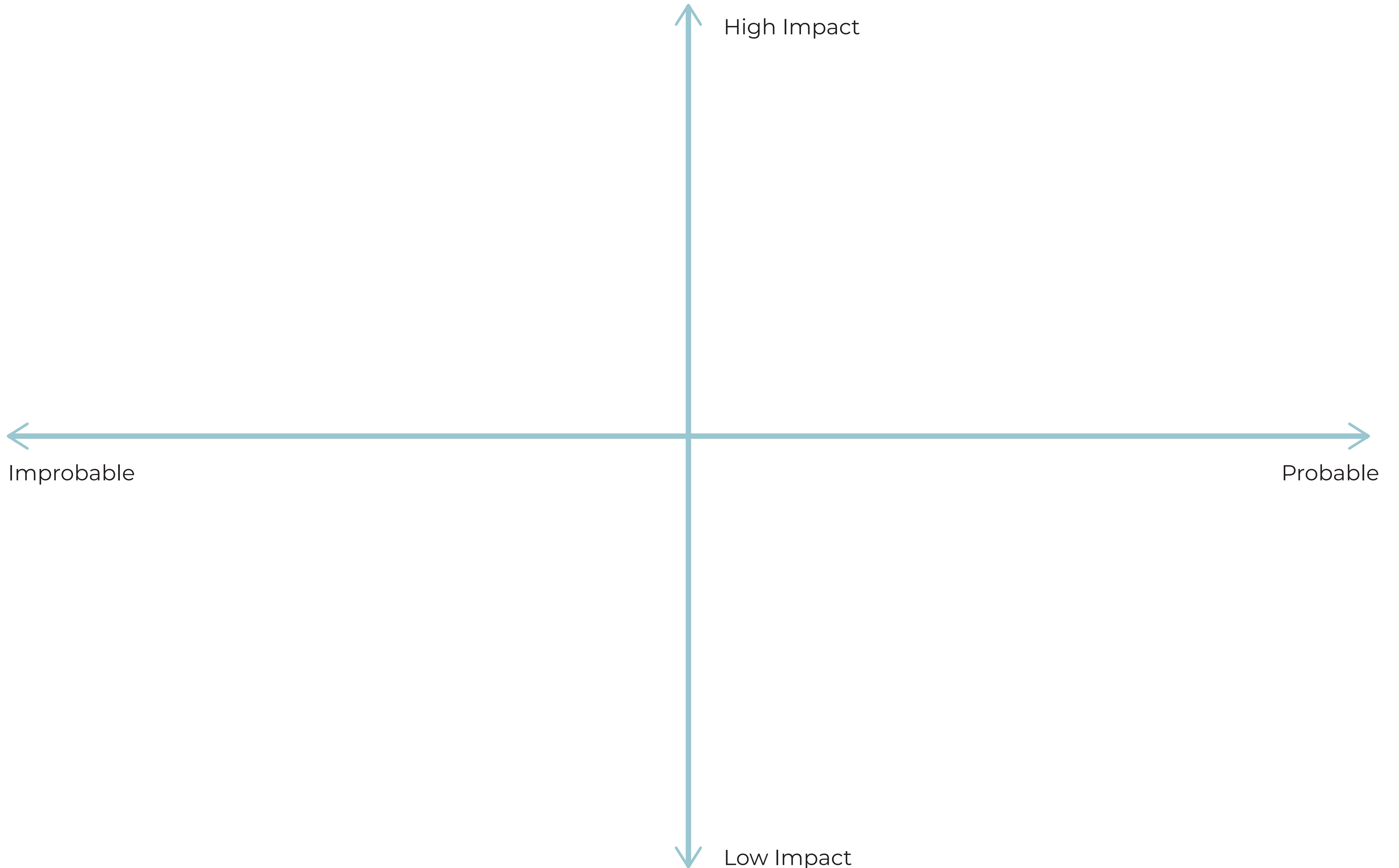


Ethical Evaluation Axes

*Qualitative assessment of the ethical
risks detected using the Responsible AI
Deck*



ETHICAL EVALUATION AXES





Project's Moral Code

*Defining the vital ethical areas to
consider on the project by making
concise moral agreement based on the
ethical risks*



PROJECT'S MORAL CODE



It's always OK to_____.



It's acceptable to_____ if
_____.



It's unacceptable to_____ if
_____.



It's never OK to_____.



Notes:

