Ethical Issues in ML and IoT

Dr. Attlee M. Gamundani (gamundani@unu.edu)
Dr. Serge Stinckwich (stinckwich@unu.edu)

ICTP Workshop on TinyML, October 20th, 2021

UNITED NATIONS UNIVERSITY
Institute in Macau

United Nations University

- **A network for solutions-focused research**: we generate research to encourage the rethinking of existing policies and perspectives
- **A think tank for the UN System:** our research informs policymaking and promotes positive global change
- **An immersive academic institution:** we train the next generation of leaders through various postgraduate degree programmes worldwide

UNITED NATIONS UNIVERSITY
Institute in Macau

UNU System

A global system of research and training institutes, coordinated by UNU Centre in Tokyo

United Nations University Institute in Macau

- The only UN entity in Macau
- UNU-Macau is a research institute at the intersections of Information and Communication Technology (ICT) and sustainable development.
- UNU-Macau conducts policy-relevant research to address keys issues from the UN 2030 Agenda on SDGs
  - What are the (positive/negative) impact of Digital Technologies on the World?
  - How Digital Technologies can help to understand/achieve the Sustainable Development Goals?

UNITED NATIONS UNIVERSITY
Institute in Macau

- Socio-technical systems
- Algorithmic decision-making (ADM)
- Modelling & Simulation
- Participatory Design
- AI Governance and Ethics
- UN Secretary-General's Roadmap for Digital Cooperation

Dr Serge Stinckwich
Head of Research

Dr JeongHyun Lee
ICTD Fellow

Dr Attlee M. Gamundani
Research consultant

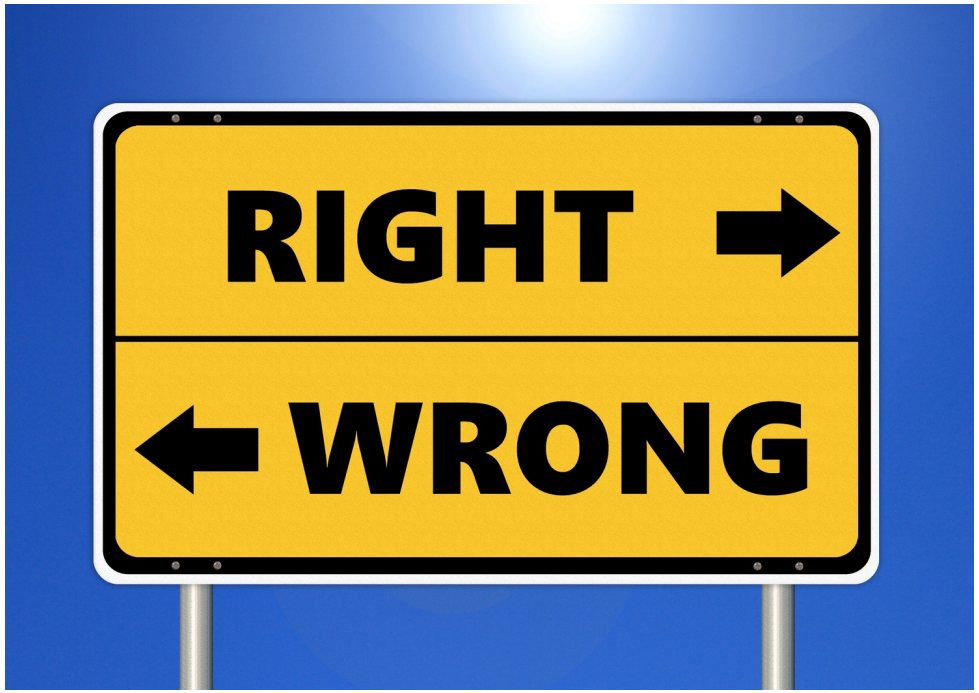
UNITED NATIONS
UNIVERSITY
Institute in Macau

Agenda for this presentation

- Ethics principles overview
- Ethics challenges in Machine Learning
- Fun break
- Ethics challenges in IoT
- AI regulations & role of UN
- What are the best practices for software developer?
- Q&A

UNITED NATIONS
UNIVERSITY
Institute in Macau

Ethics principles overview



What is ethics (or moral philosophy)?

- Branch of philosophy that "involves systematizing, defending and recommending concepts of right and wrong behavior".
- Ethics etymology comes from ancient greek "Ethos" (meaning "relating to one's character")
- Ethics vs morality
  - Are often taken as synonyms
  - Ethics can be seen as a theory of morality

General subjects areas of Ethics

- *Descriptive Ethics*
  - What do people think is right.
  - Studied primarily by sociology, psychology, anthropology, …
- *Meta-ethics*
  - Where our ethical principal come from and what they mean?

General subjects areas of Ethics

- ***Normative or prescriptive ethics***
  - More practical task, that is to arrive at moral standards that regulate right and wrong conduct,
  - Attemps to develop a set of rules that govern human conducts
  - *Example: the action of stealing is morally wrong*
- ***Applied ethics***
  - Analysis of specific controversial moral issues:
    - Animal rights, environmental concerns, capital punishment, euthanasia, …
  - More concrete fields: medical ethics, bioethics or business ethics
  - AI ethics is part of applied ethics

UNITED NATIONS
UNIVERSITY
Institute in Macau

Deontological or duty ethics

- Morality of an action is based on whether that action is right or wrong, rather than based on the consequences of the action
- Action is more important than consequences
- *Example: Since the 2000's large and medium companies have increasingly tried to project a social or environmental image through certain marketing practices (Ethics washing)*

Consequentialism or teleological ethics

- Determine ethical correctness of an action or norm based on their (foreseeable) consequences
- *Example: From this perspective, the motives of the company to invest in CSR (Corporate Social Responsibilities) play no role. The only considerations should be the social or environmental impacts.*

Virtue ethics

- Focus on the inherent character of a person rather than on nature or consequences of specific actions she or he performs
- Focus on good habits
- *Example: benevolence*

Comparison of main ethical theories

	Consequentialism	Deontology	Virtue Ethics
Description	An action is right if it promotes the best consequences, i.e. maximize happiness	An action is right if is in accordance with a moral rule or principle	An action is right if it is what a virtuous person would do in the circumstances
Central Concern	The results matter, not the actions themselves	Persons must be seen as ends and may never be used as means	Emphasise the character of the agent making the actions
Guiding Value	Good (often seen as maximum happiness)	Right (rationality is doing one's moral duty)	Virtue (leading to the attainment of eudaimonia)
Practical Reasoning	The best for most (means-ends reasoning)	Follow the rule (rational reasoning)	Practice human qualities (social practice)
Deliberation focus	Consequences (what is outcome of action?)	Action (Is action compatible with some imperative?)	Motives (Is action motivated by virtue?)

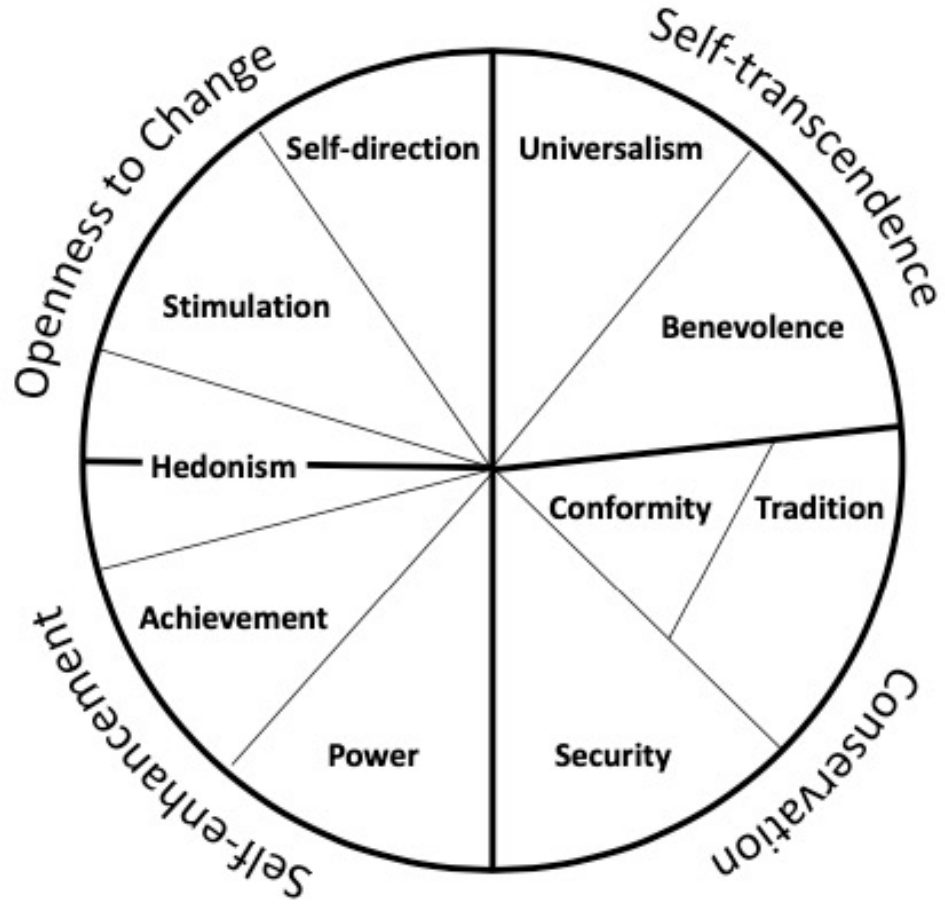
Extract from Virginia Dignum, "Responsible AI", Springer Verlag, 2019

UNITED NATIONS UNIVERSITY
Institute in Macau

Theory of human values

- Main challenges of ethical reasoning: what moral values to consider and how to prioritise them depending on the situation
- Attempt to measure universal human values that are recognized throughout all major cultures: ***Schwartz classification***

Schwartz classification of basic values

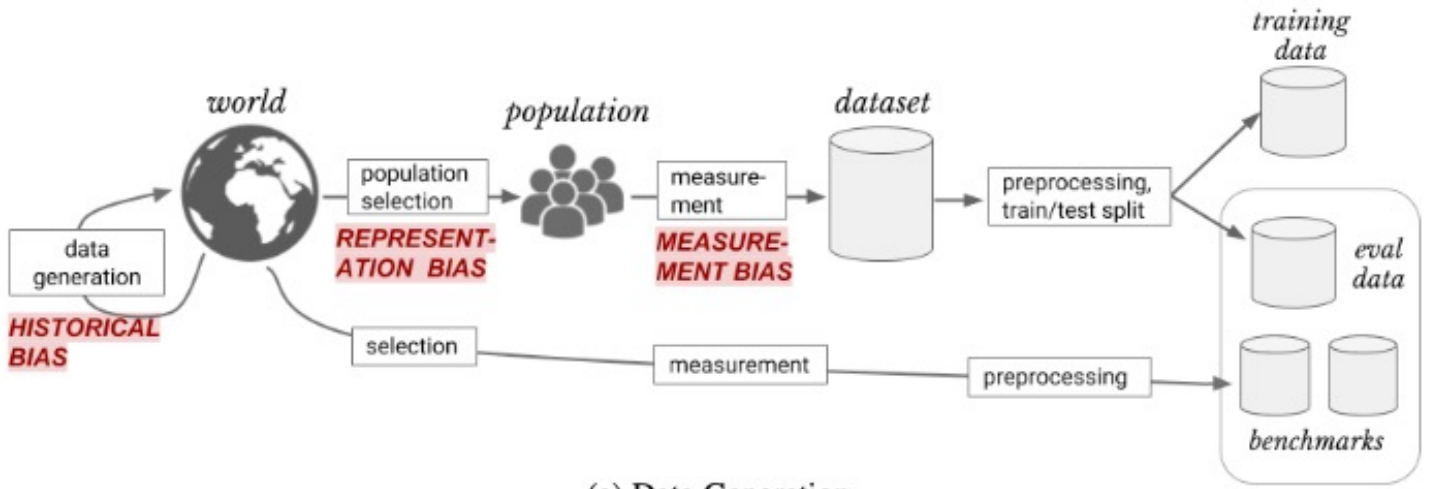


UNITED NATIONS UNIVERSITY
Institute in Macau

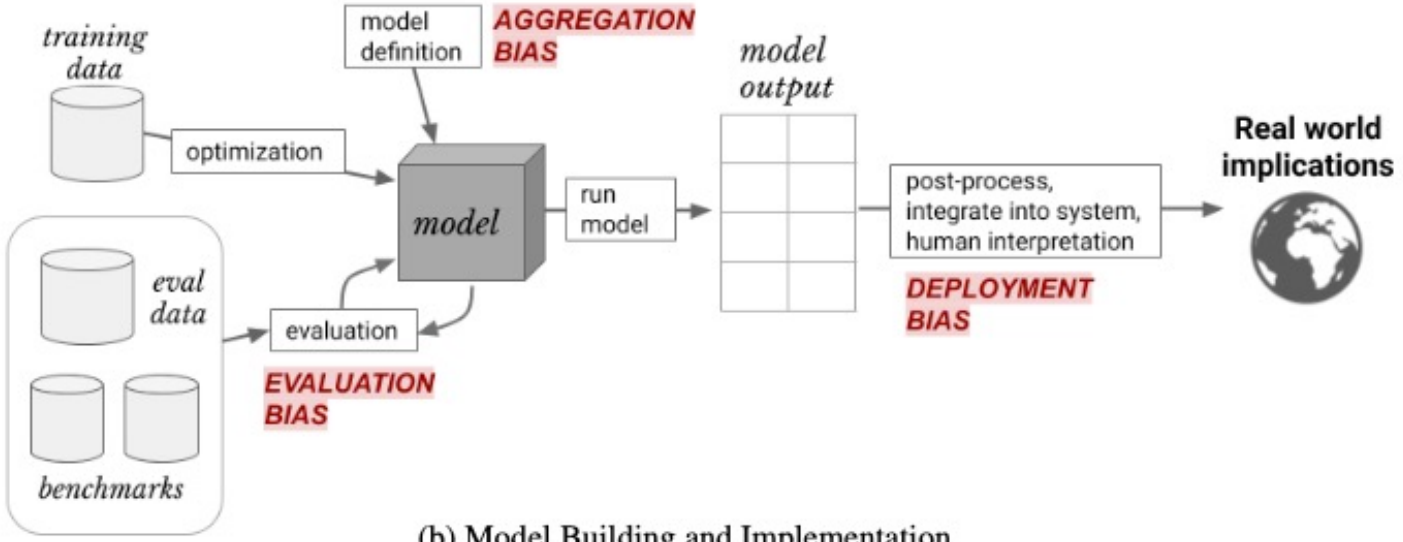
Unintended consequences of ML due to Bias

- There are many ways to understand bias (statistics, social science, …).
- Common rhetoric is that various unwanted consequences of ML algorithms arise in some way from "**biased data**".
- Empirical findings has shown that data-driven methods can unintentionally encode human biases and introduce new ones: **Machine Learning can amplify bias !**
- Biased data is the product of many factors.

Bias in ML pipelines



(a) Data Generation



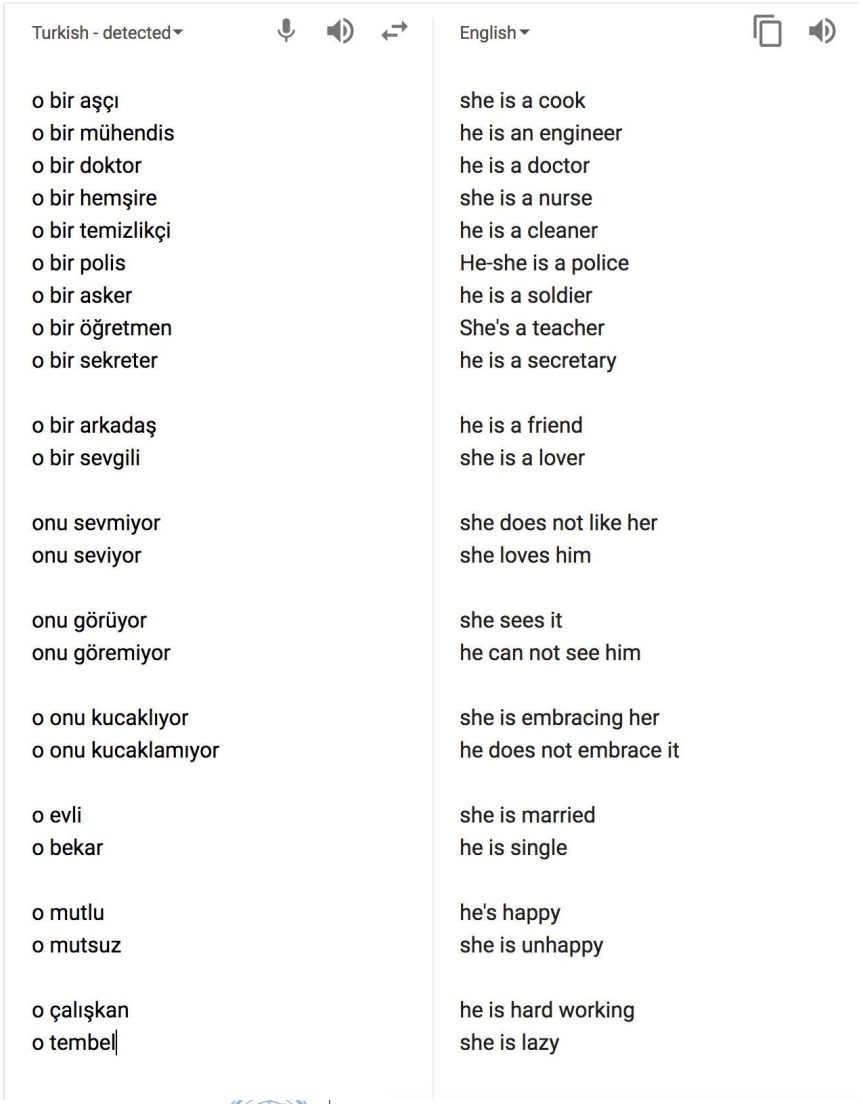
(b) Model Building and Implementation

Extract from Harini Suresh, Jogn V. Guttag, A Framework for Understanding Unintended Consequences of Machine Learning", 2020

UNITED NATIONS UNIVERSITY
Institute in Macau

Historical bias

- Comes from the fact that people are biased, processes are biased, the society is biased.
- It can exist even given perfect sampling and feature selection.
- Any dataset involving humans can have this kind of bias: medical data, sales data, etc …
- https://nypost.com/2017/11/30/google-translates-algorithm-has-a-gender-bias/



UNITED NATIONS UNIVERSITY
Institute in Macau

Representation bias

- Arises from how we sample from a population during data collection process.
- Particulary common problem in datasets
- Example: lack of geographical diversity in datasets like ImageNet results in bias towards Western cultures. The same regarding ethnicity and gender. “groom” category show mostly white people.

Deployment bias

- Arises when there is a mismatch between the problem a model is intended to solve and the way in which it is actually used.
- Often the case, when the model is built in a quite isolated way but it used at the end in a complicated socio-technical environment
- Example: ***Risks assessment tools***. First define to predict a person's likelihood of commit a future crime, then used to determine the length of a sentence (see Collins, **Punishing risk**, 2018)

UNITED NATIONS
UNIVERSITY
Institute in Macau

Fun break

https://www.moralmachine.net/

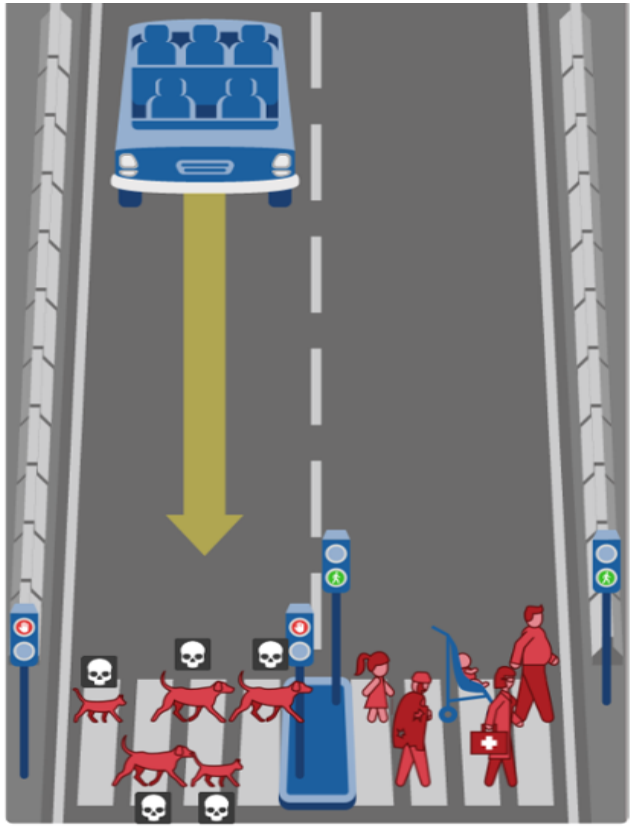
ALERT! : The scenarios presented are hypothetical, they involve making tough decisions on who lives or dies as a result of a decision being made by the Self-driving car in motion. If you are not comfortable to participate in the activities due to the sensitive nature of such typical scenarios, you are excused.

Scenario 1 – What should the self driving car do?

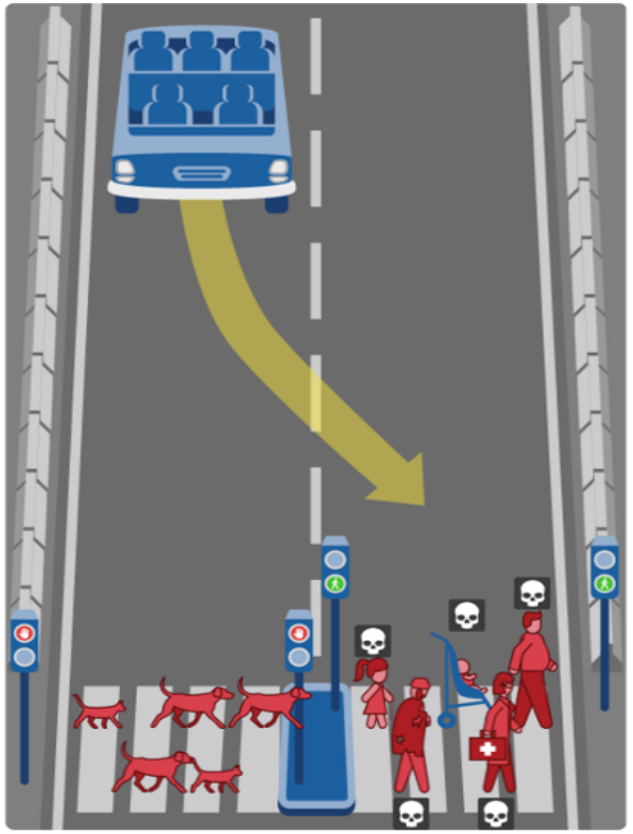
In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...
Dead:

- 2 cats
- 3 dogs

Note that the affected pedestrians are flouting the law by crossing on the red signal.



Left (A)



Right (B)

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...
Dead:

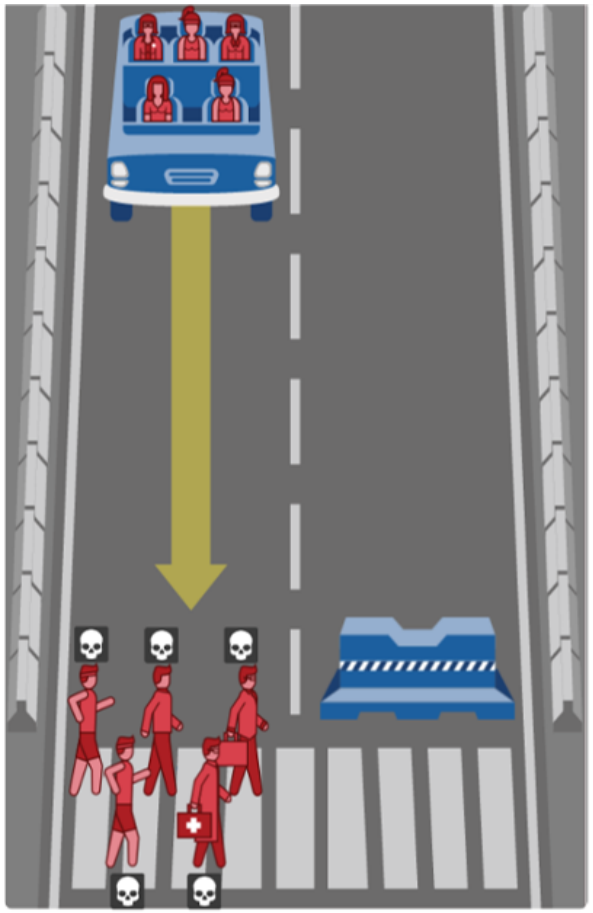
- 1 girl
- 1 baby
- 1 large man
- 1 homeless person
- 1 female doctor

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

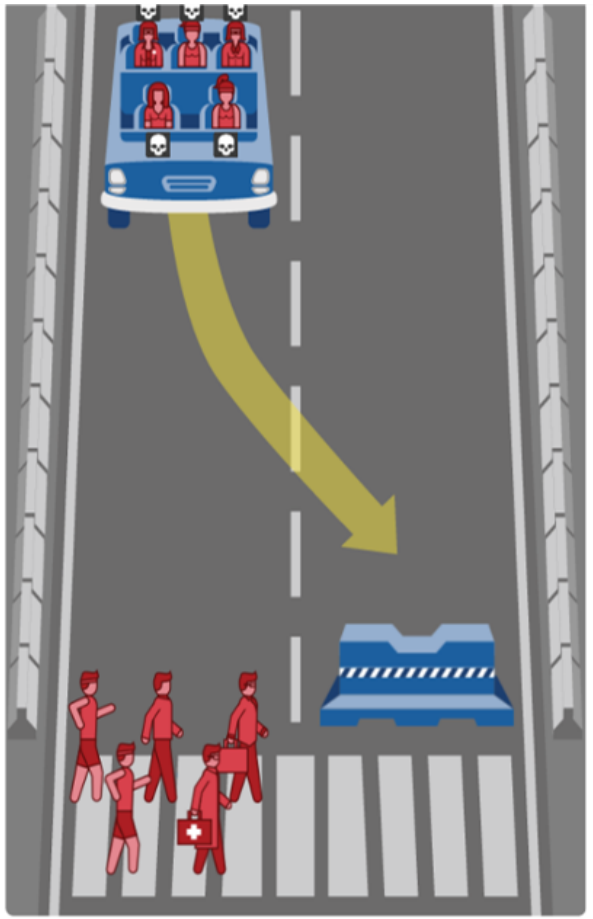
Scenario 2 – What should the self driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...
Dead:

- 2 male athletes
- 1 man
- 1 male executive
- 1 male doctor



Left (A)



Right (B)

In this case, the self-driving car with sudden brake failure will swerve and crash into a concrete barrier. This will result in ...
Dead:

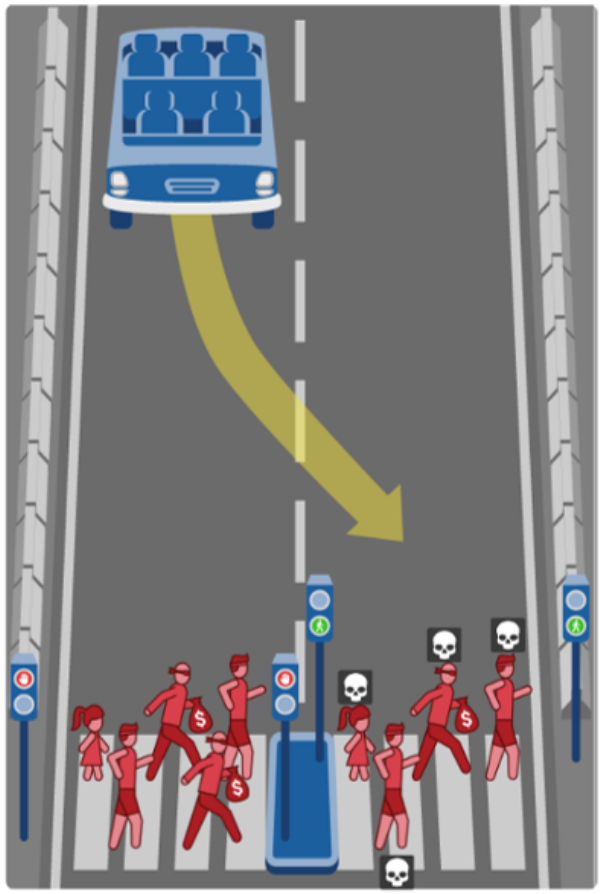
- 2 female athletes
- 1 woman
- 1 female executive
- 1 female doctor

Scenario 3 – What should the self driving car do?

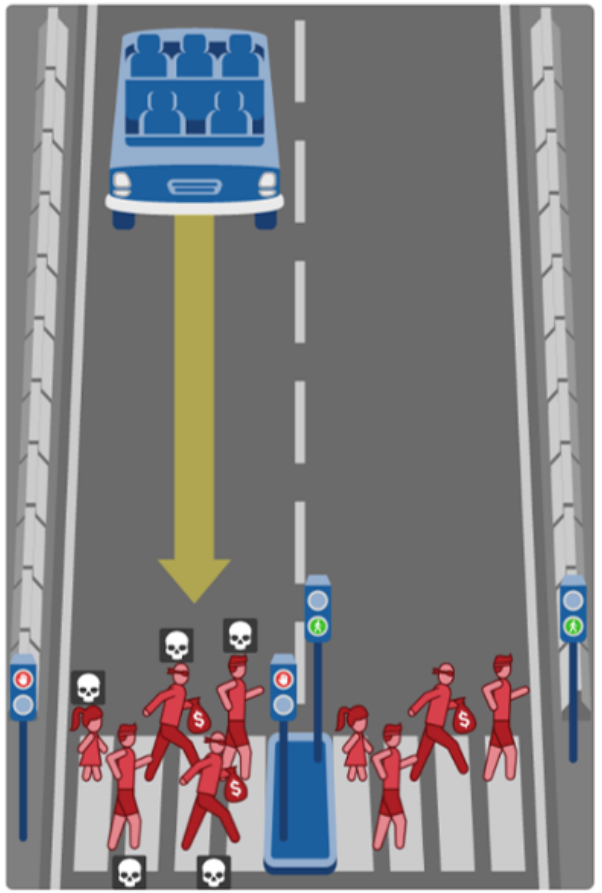
In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...
Dead:

- 1 girl
- 1 criminal
- 2 male athletes

Note that the affected pedestrians are abiding by the law by crossing on the green signal.



Left (A)



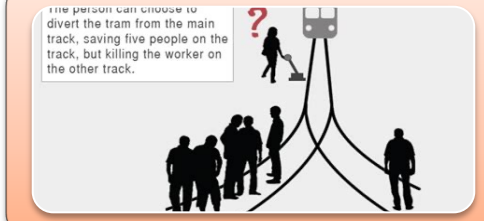
Right (B)

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...
Dead:

- 1 girl
- 2 criminals
- 2 male athletes

Note that the affected pedestrians are flouting the law by crossing on the red signal.

Interactive Activity - conclusion



Original Trolley Dilemma

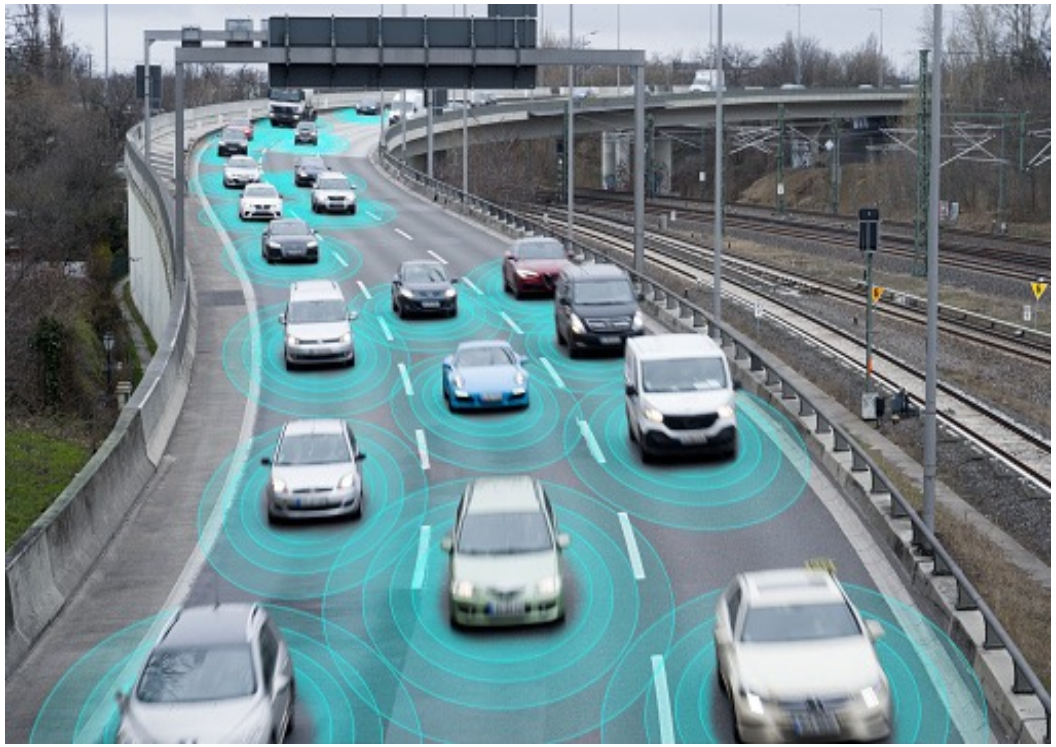
Justification of decisions; Responsibility

Some issues to “decide” actions

- Human vs. Non-human (object, animal…)
- “Categories” of humans (age, gender, size, class…)

UNITED NATIONS UNIVERSITY
Institute in Macau

From trolley problems to autonomous cars: AI as a Socio-Technical Systems



UNITED NATIONS UNIVERSITY
Institute in Macau

And what about IoT?

- End of 2020: more than 50 billions things connected
- Similar situation than AI: wide-known benefits of IoT but not that much studies of IoT vulnerabilities/harms except security/privacy issues
- IoT ethics intersect with security, privacy, legal, etc …
- Combining to AI, IoT can generate a lot of unintented consequences
- Lots of potential issues in LMIC where IoT policies are under-developed (UN role)

UNITED NATIONS UNIVERSITY
Institute in Macau

Privacy threats of IoT

- Identification of person
- Behavioral profiling
- Localization, tracking and location-based personalization
- Combination of information sources
- Lifecycle transitions (are data erased when used by a new user?)

UNITED NATIONS UNIVERSITY
Institute in Macau

Ethical issues in IoT

- **Ambiguity**: people are not able to distinguish natural/artificial artifacts
- **Public vs private border line**: absence of clear-cut boundaries, there is no distinction between private and public data. Collecting data without user's consent
- **Unpredictable behavior**: might interfere with human actions and decisions
- **Loss of control**: governance and human control will be ceased due to huge number of devices
- **Life threats**: a breach on IoT network can harm directly our lives (e.g data breach on an autonomous car)
- **Equality of access, responsibility, transparency, and more**

AI regulations & legislations

- Ethics starts where law ends.
- European Union: data rights and privacy laws, EU General Data Protection Regulation (GDPR)
- Global Partnership for AI (GPAI): international multi-stakeholder forum to debate policy implications of AI globally
- IEEE: Ethics Certification Program for Autonomous & Intelligent Systems, which develops metrics towards the implementation of certification regarding transparency, accountability and algorithmic bias
- Google Responsible Innovation Team

Role of UN towards sustainable AI

- Role of UN as an active convener to facilitate discussions on AI and sustainability among diverse and multiple stakeholders.
- Both positive and negative impacts of AI on SDGs should be analyzed
- Climate change should not be separated from the conversation about the impact of AI on the world
- UN establish an inter-agency working group on AI (IAWG-AI), co-led by UNESCO and ITU (October 2020)

UNITED NATIONS
UNIVERSITY
Institute in Macau

Sustainable Development Goals & and Ethics

The case for comprehensive engineering

...ethical issues themselves are **interconnected and often cannot be dealt with in isolation**.

..**privacy and intellectual property** are concerned with **preventing access** to information, whereas transparency, accountability and equity concern the provision of access to information (sometimes the very same information that is protected on privacy grounds).

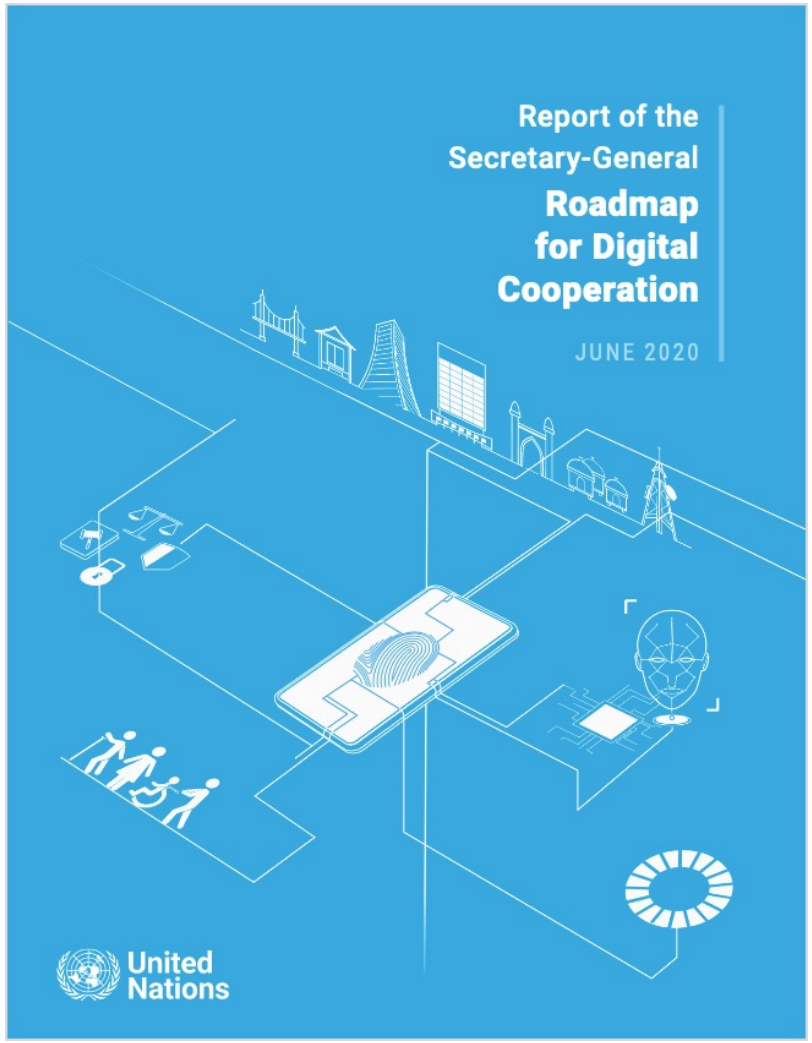
Furthermore, **democratic governance** and **fair regulation** of the internet may pertain to both issues.

Solutions to these problems are **interrelated** and arguably cannot be dealt with in isolation.

(Joreon van den Hoven, 2016)

UNITED NATIONS UNIVERSITY
Institute in Macau

UN Secretary-General's Roadmap for Digital Cooperation



- Define 5 pillars for AI: trustworthiness, human rights, safety, sustainability and promotion of peace
- https://cs.unu.edu/news/news/ai-trustworthiness.html

UNITED NATIONS UNIVERSITY
Institute in Macau

What are the best practices as ML Developer regarding ethical issues?

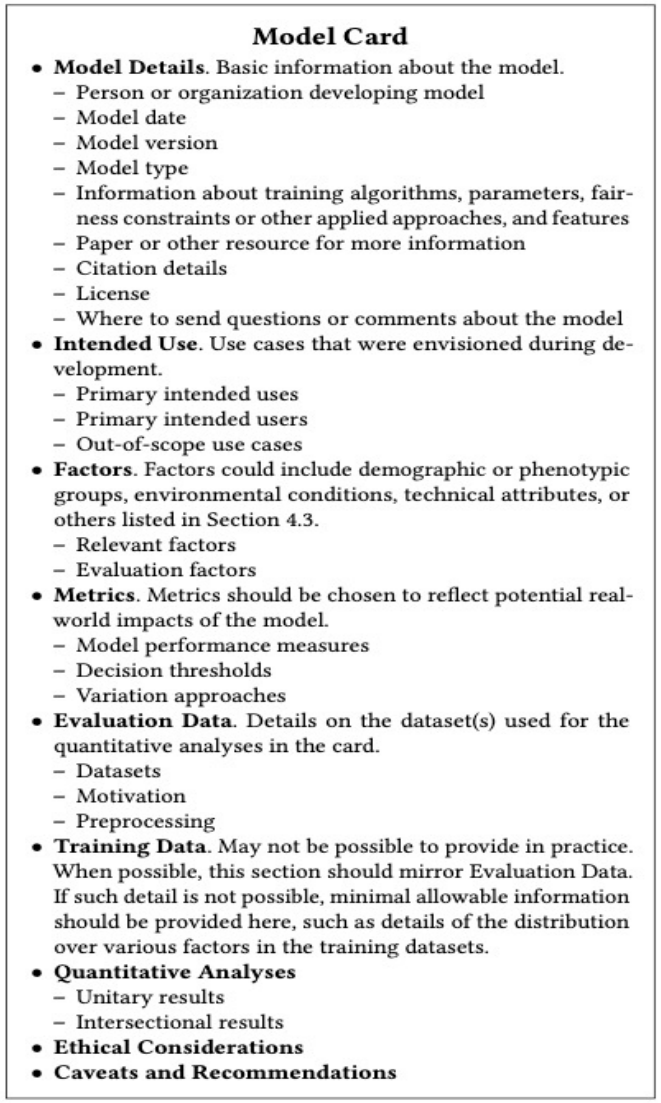
- Hundreds of principles, policies, available: https://oecd.ai/en/

- How to operationalize ethics in Software?
  - Learn more about ethical issues (the role of this introduction)
  - Design an AI-IoT system with these issues in mind
  - Try to engage users from the beginning

UNITED NATIONS UNIVERSITY
Institute in Macau

Models Cards for Model reporting

Short document to describe machine learning models that provide benchmarked evaluation of performance in a variety of relevant conditions.



Model Card

- **Model Details**. Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use**. Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors**. Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics**. Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data**. Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data**. May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

UNITED NATIONS UNIVERSITY
Institute in Macau

Model Card example: toxicity in text

- Developed by jigsaw in 2017, using Convolutional Neural Network - TOXICITY classifier provided perspective API trained to predict the likelihood that a comment will be perceived as toxic.

- **Ethical Considerations**
  - The Perspective API uses a set of values to guide their work. These values are Community, Transparency, Inclusivity, Privacy, and Topic-neutrality.
  - Because of privacy considerations, the model does not take into account user history when making judgments about toxicity.

Conclusion: Still a lot to do …

- In the last 2 years, unprecedented explosion of interest in the academic community on fairness and ML, but still difficult to understand what is fairness in ML
- Lots of empirical work, more theoretical foundations are needed
- How do we operationalize, AI Ethics principles?
- How do we contextualize?
- The power of diversity

UNITED NATIONS
UNIVERSITY
Institute in Macau

Bibliography to know more

- Virginia Dignum, "***Responsible AI – How to Develop and Use AI in a Responsible Way***", Springer Verlag, 2019
- Rachel Thomas, "***Data Ethics***" chapter in Jeremy Howard, Sylvain Gugger, "Deep Learning for Coders with fastai & PyTorch", O'Reilly
- ***Secretary-General's Roadmap for Digital Cooperation*** report, United Nations
- Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Wasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru, "***Model Cards for Model Reporting***", FAccT19

http://macau.unu.edu/

unumacau

UNITED NATIONS
UNIVERSITY
Institute in Macau