Ethics for AI Developers

a crash course

4 November 2025

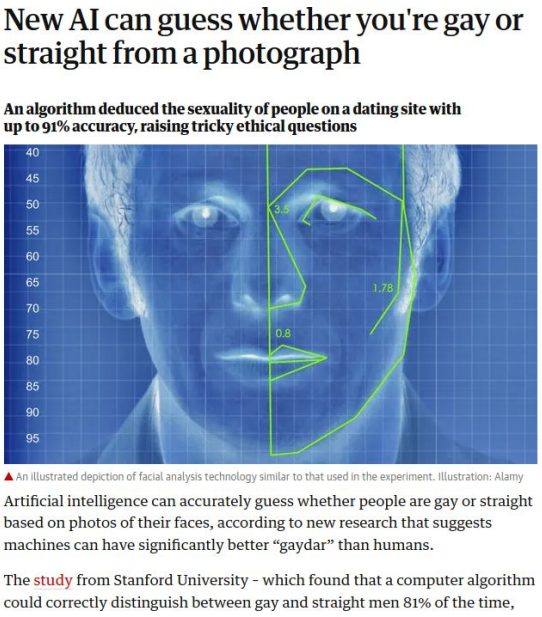
Human-centred Machine Learning course

Giovanni Sileno
g.sileno@uva.nl

University of Amsterdam

Technology in the (bad) news…

Technology in the (bad) news…



New AI can guess whether you're gay or straight from a photograph

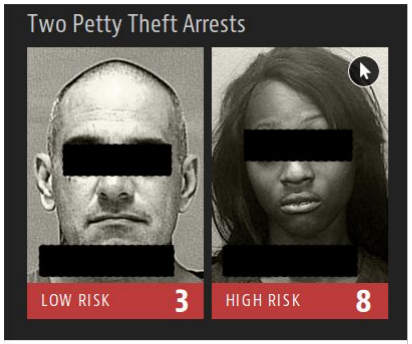
An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions

▲ An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy

Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better "gaydar" than humans.

The study from Stanford University - which found that a computer algorithm could correctly distinguish between gay and straight men 81% of the time,

https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph (2017)



COMPAS: software used in the US predicting future crimes and criminals argued to be biased against African Americans (2016)

Angwin J. et al. ProPublica, May 23 (2016). *Machine Bias: risk assessments in criminal sentencing*

SyRI (System Risk Indication) used in the Netherlands to create risk alerts for welfare frauds by processing and linking personal data of citizens argued to be discriminatory and unlawful (2018)

https://pilpnjcm.nl/en/proceedings-risk-profiling-dutch-citizens-syri

Technology in the (bad) news…

WIRED Technology | Science | Culture | Gear | Business | Politics | More

Privacy

Co-op is using facial recognition tech to scan and track shoppers

Branches of the Southern Co-op are using facial recognition to look for potential shoplifters. The roll-out raises concerns about the creep of surveillance tech in the private sector

https://www.wired.co.uk/article/coop-facial-recognition (2020)

NL#TIMES TOP STORIES HEALTH CRIME POLITICS BUSINESS TECH

Jumbo - Credit: Jumbo / Jumbo

CRIME TECH INNOVATION JUMBO SHOPLIFTING AI » MORE TAGS

SHARE THIS:

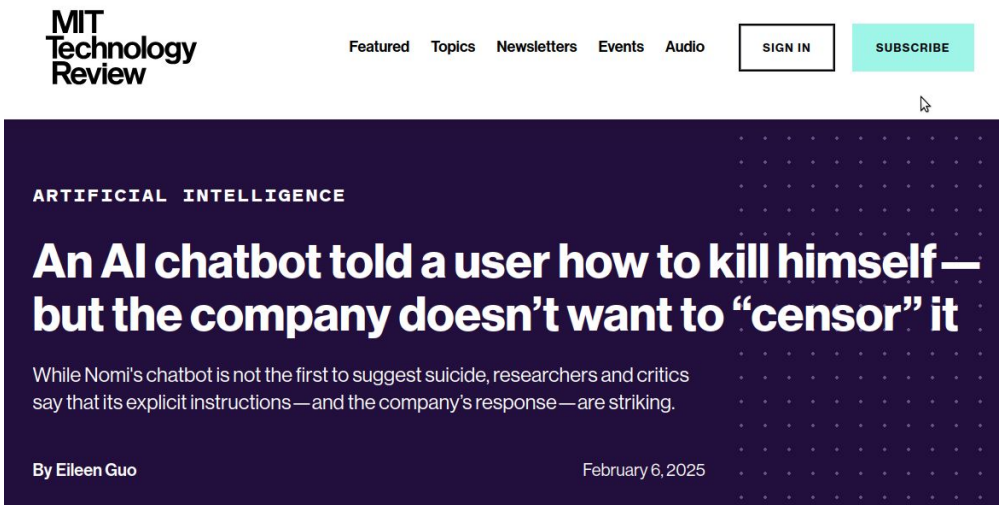
SUNDAY, 11 FEBRUARY 2024 - 08:15

Jumbo takes extra measures against self-scan shoplifting

Supermarket chain Jumbo will take extra measures to stop shoplifting over the next few weeks. For example, there will be more clearly visible camera surveillance and more and smarter random checks at self-checkouts. There will also be extra communication to customers that they must pay for all groceries and that they will be

https://nltimes.nl/2024/02/11/jumbo-takes-extra-measures-self-scan-shoplifting (2024)

Technology in the (bad) news…



Technology in the (bad?) news…

nature computational science

Explore content · About the journal · Publish with us

nature > nature computational science > articles > article

Article | Published: 18 December 2023

Using sequences of life-events to predict human lives

Germans Savcisens, Tina Eliassi-Rad, Lars Kai Hansen, Laust Hvas Mortensen, Lau Lilleholt, Anna Rogers, Ingo Zettler & Sune Lehmann

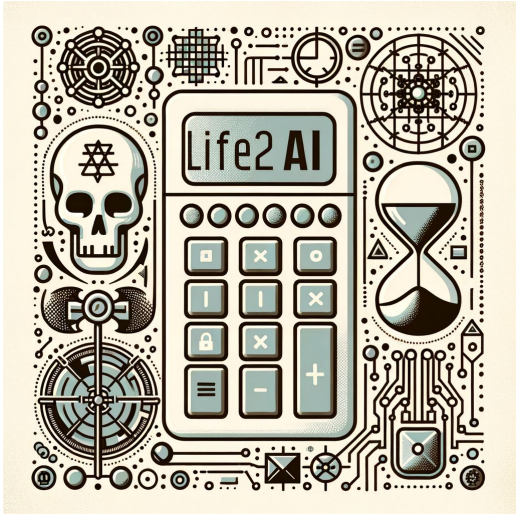
Nature Computational Science **4**, 43–56 (2024) | Cite this article

33k Accesses | **64** Citations | **2289** Altmetric | Metrics

A preprint version of the article is available at arXiv.

Abstract

Here we represent human lives in a way that shares structural similarity to language, and we exploit this similarity to adapt natural language processing techniques to examine the evolution and predictability of human lives based on detailed event sequences. We do this by drawing on a comprehensive registry dataset, which is available for Denmark across several years, and that includes information about life-events related to health, education,



Overview

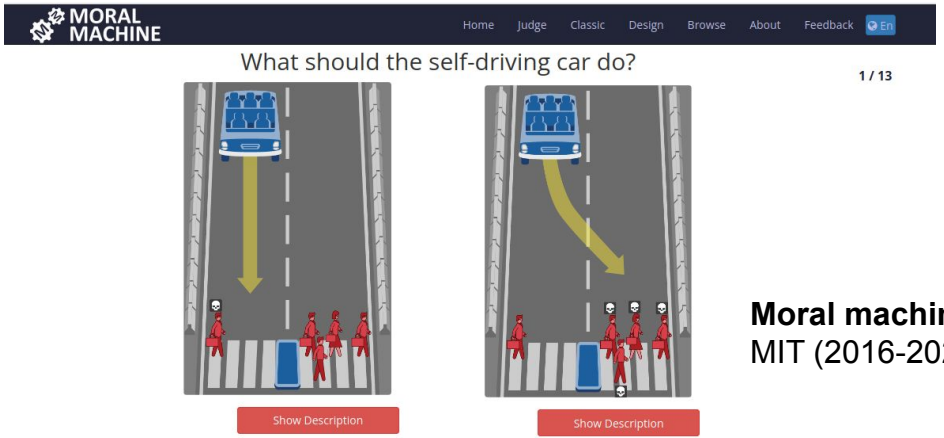
my talk consists of:

- a brief & broad overview on concepts which are relevant to ethics
- a few practical examples of problems that you may encounter as AI designer/developer

Why is this relevant?

- students going into computational tracks (consequently becoming tech researchers/experts) may not have been explicitly exposed to topics related to ethics

➡ risk of naive if not dangerous misunderstandings



Moral machine, MIT (2016-2020)

Why is this relevant?

- students going into computational tracks (consequently becoming tech researchers/experts) may not have been explicitly exposed to topics related to ethics

→ risk of naive if not dangerous misunderstandings

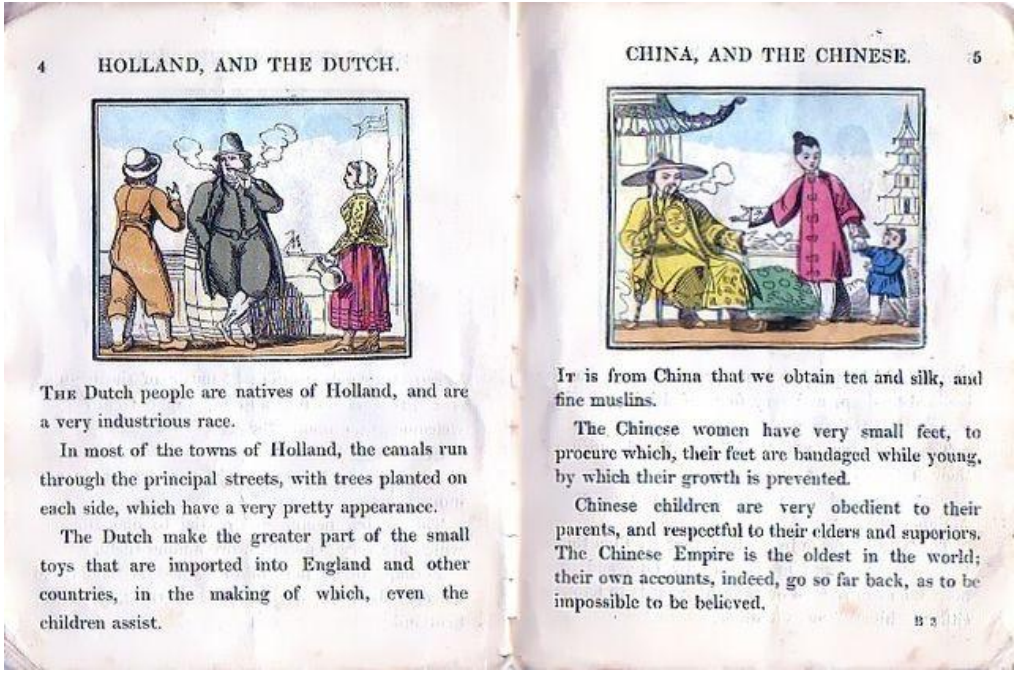
- on the other hand, this is not “rocket science”, these concepts are accessible to all of us **just because we are humans, social beings!**

→ getting acquainted with ethics increases our *“human capital”*

Morals & co.

Morals & co.

- ***“mores”***, latin word for *social norms*, *costumes*, *habits* of a community



4 HOLLAND, AND THE DUTCH.

THE Dutch people are natives of Holland, and are a very industrious race.

In most of the towns of Holland, the canals run through the principal streets, with trees planted on each side, which have a very pretty appearance.

The Dutch make the greater part of the small toys that are imported into England and other countries, in the making of which, even the children assist.

CHINA, AND THE CHINESE. 5

IT is from China that we obtain tea and silk, and fine muslins.

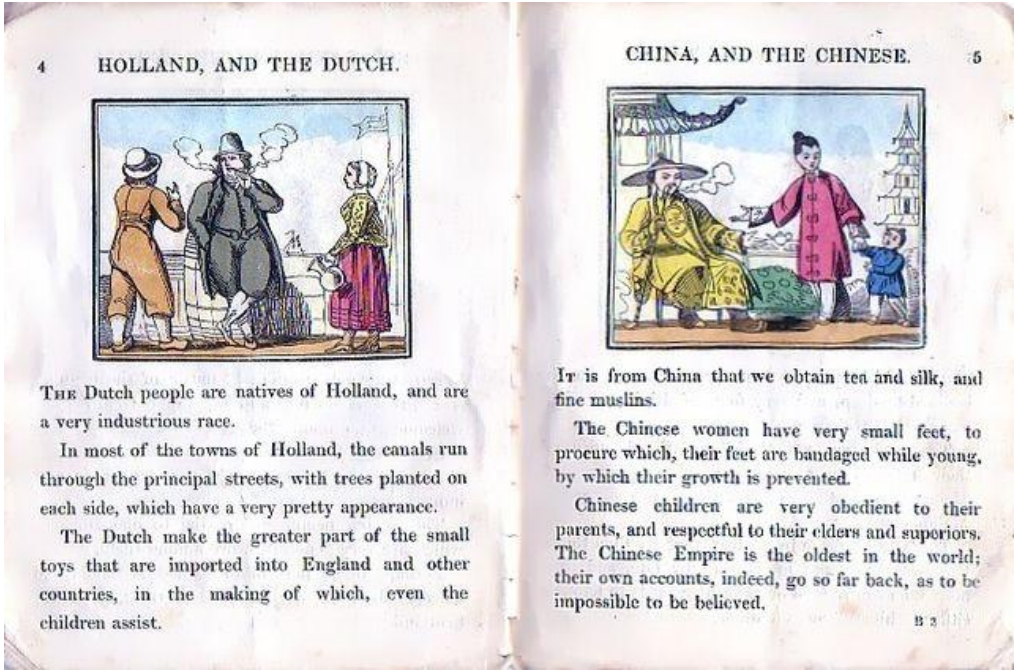
The Chinese women have very small feet, to procure which, their feet are bandaged while young, by which their growth is prevented.

Chinese children are very obedient to their parents, and respectful to their elders and superiors. The Chinese Empire is the oldest in the world; their own accounts, indeed, go so far back, as to be impossible to be believed.

B 3

Morals & co.

- ***"mores"***, latin word for *social norms*, *costumes*, *habits* of a community
- ***"ethos"*** is its greek equivalent (*character)*



4 HOLLAND, AND THE DUTCH.

THE Dutch people are natives of Holland, and are a very industrious race.

In most of the towns of Holland, the canals run through the principal streets, with trees planted on each side, which have a very pretty appearance.

The Dutch make the greater part of the small toys that are imported into England and other countries, in the making of which, even the children assist.

CHINA, AND THE CHINESE. 5

IT is from China that we obtain tea and silk, and fine muslins.

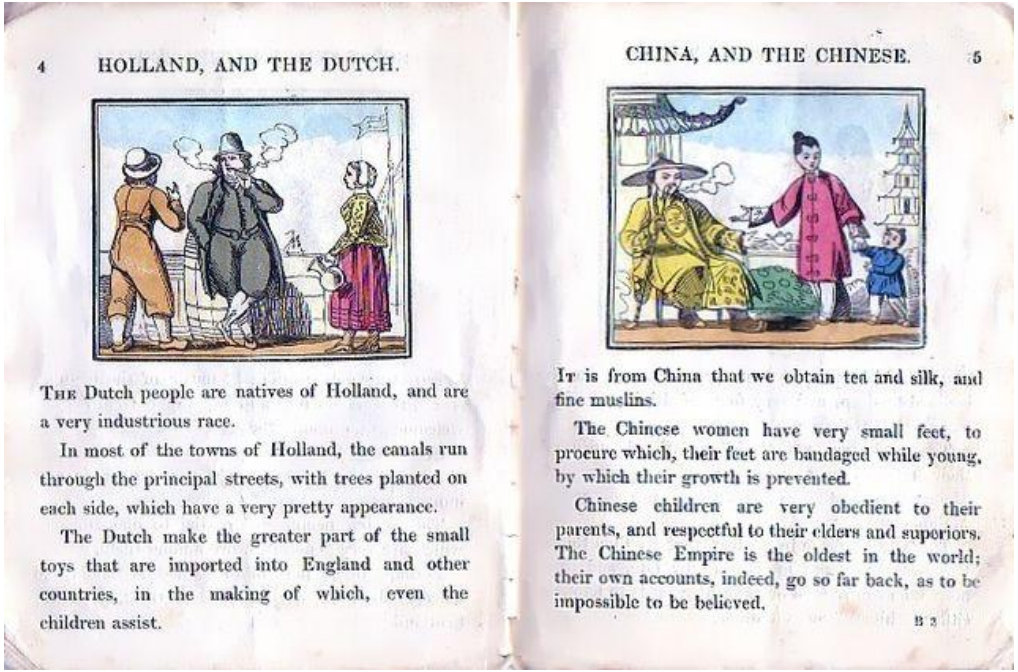
The Chinese women have very small feet, to procure which, their feet are bandaged while young, by which their growth is prevented.

Chinese children are very obedient to their parents, and respectful to their elders and superiors. The Chinese Empire is the oldest in the world; their own accounts, indeed, go so far back, as to be impossible to be believed.

B 3

Morals & co.

- ***"mores"***, latin word for *social norms*, *costumes*, *habits* of a community
- ***"ethos"*** is its greek equivalent (*character)*



4 HOLLAND, AND THE DUTCH.

THE Dutch people are natives of Holland, and are a very industrious race.

In most of the towns of Holland, the canals run through the principal streets, with trees planted on each side, which have a very pretty appearance.

The Dutch make the greater part of the small toys that are imported into England and other countries, in the making of which, even the children assist.

CHINA, AND THE CHINESE. 5

IT is from China that we obtain tea and silk, and fine muslins.

The Chinese women have very small feet, to procure which, their feet are bandaged while young, by which their growth is prevented.

Chinese children are very obedient to their parents, and respectful to their elders and superiors. The Chinese Empire is the oldest in the world; their own accounts, indeed, go so far back, as to be impossible to be believed.

B 3

Note the difference between **descriptive** and **prescriptive** views!

Morals & co.

- **individuals behave according and against *mores***
  - generally no problem for ***descriptive norms*** (if not for identity matters, and so “qualification”/categorization processes)
  - ***prescriptive norms*** address correct/incorrect behaviour
    ⇒ grounds for *judgment*

Morals & co.

- **individuals behave according and against *mores***
  - for identitarian purposes, ***descriptive norms may become prescriptive!***

punks vs traders: spot the (structural) difference

Morals & co.

- **individuals behave according and against *mores***
- **behaviour is moral** if *justifiable according to a **moral standard*** (morality comes always with an evaluative framework)
- moral judgments are given from a collective standpoint (the one associated to the mores)

adultery for puritans (The Scarlet Letter, Hawthorne)

Whose *mores*?

- *mores* were defined above collectively, but one could apply the same concept at individual level:

  "I have my habits, and my ways to evaluate behaviour as good

Whose *mores*?

- *mores* were defined above collectively, but one could apply the same concept at individual level:

  "I have my habits, and my ways to evaluate behaviour as good

- vice-versa, organizations or communities can be seen as “individuals” (collective agencies)! eg. *at UvA we do like that*.

Normality vs normativity

- descriptive/prescriptive evaluative frameworks can be reinterpreted in ***agentive*** terms:

normality	be (not)	believe (not) to be	the case
normativity	ought (not) to be	desire (not) to be	
	“objective”	*subjective*	

Normality vs normativity?

The boundary between descriptive and prescriptive norms is a delicate one.

Suppose you are a shop owner:

- you don't want fraudsters in your shop

Normality vs normativity?

The boundary between descriptive and prescriptive norms is a delicate one.

Suppose you are a shop owner:

- you don't want fraudsters in your shop
- fraudsters typically dress a red tie

Normality vs normativity?

The boundary between descriptive and prescriptive norms is a delicate one.

Suppose you are a shop owner:

- you don't want fraudsters in your shop
- fraudsters typically dress a red tie

➡ you don't want people with red tie in your shop

Normality vs normativity?

The boundary between descriptive and prescriptive norms is a delicate one.

Suppose you are a shop owner:

- you don't want fraudsters in your shop
- fraudsters typically dress a red tie

➡ you don't want people with red tie in your shop

Is this good? is it right? for whom?

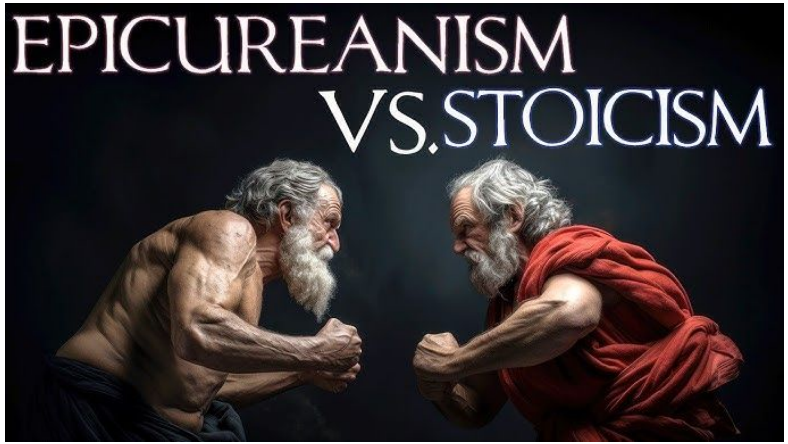
The relation between ethics and morals

Ethics

- In philosophy the word **ethics** was introduced rather to identify a specific branch of philosophical discourse (like *ontology*, *epistemology*, *aestethic*, ...)

Ethics

- In philosophy the word **ethics** was introduced rather to identify a specific branch of philosophical discourse (like *ontology*, *epistemology*, *aestethic*, ...)
- Ethical schools (stoics, epicureans, skeptics, ...) were seen as providing different ***principles*** to define what would be considered to be good or bad



Ethics

- In philosophy the word **ethics** was introduced rather to identify a specific branch of philosophical discourse (like *ontology*, *epistemology*, *aestethic*, ...)
- Ethical schools (stoics, epicureans, skeptics, ...) were seen as providing different ***principles*** to define what would be considered to be good or bad

- There was no “winner”: no ethics school is better than the other

Ethics

- In philosophy the word **ethics** was introduced rather to identify a specific branch of philosophical discourse (like *ontology*, *epistemology*, *aestethic*, ...)
- Ethical schools (stoics, epicureans, skeptics, ...) were seen as providing different ***principles*** to define what would be considered to be good or bad

- There was no “winner”: no ethics school is better than the other
- Rather, individuals are seen as forming their own morality on the basis of mores ***and*** ethics

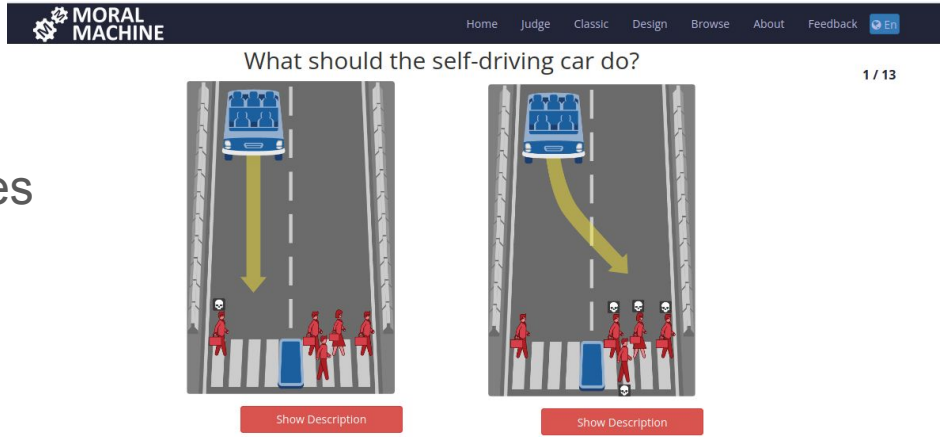
Ethics vs mores

- Yet, a common point of all these schools is that ***one may behave ethically, even going against mores*** (collective morals)

Ethics vs mores

- Yet, a common point of all these schools is that ***one may behave ethically, even going against mores*** (collective morals)

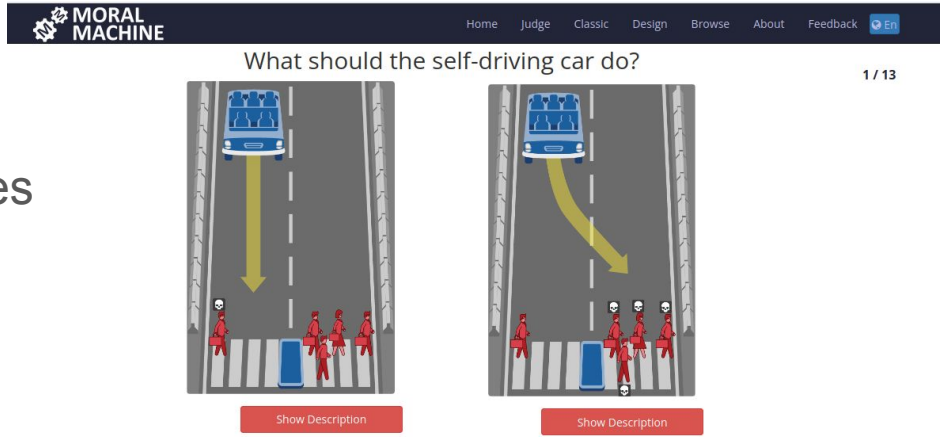
Let's aggregate global preferences over moral dilemmas.
Is the resulting decision-making "moral"? Is it "ethical"?



Ethics vs mores

- Yet, a common point of all these schools is that ***one may behave ethically, even going against mores*** (collective morals)

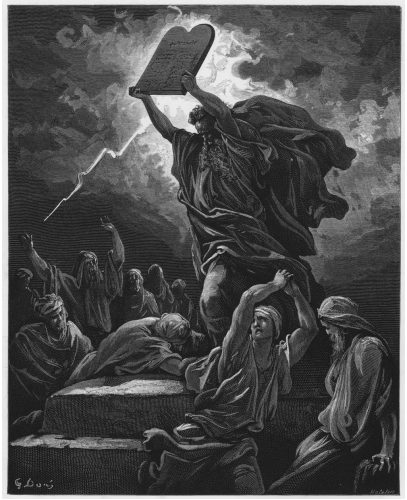
Let's aggregate global preferences over moral dilemmas.
Is the resulting decision-making "moral"? Is it "ethical"?

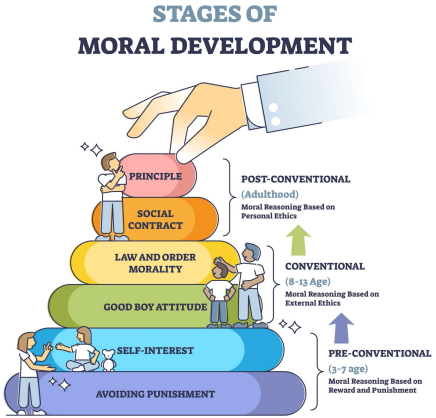


The question of where morality comes from require further investigation…

“Natural” morals…

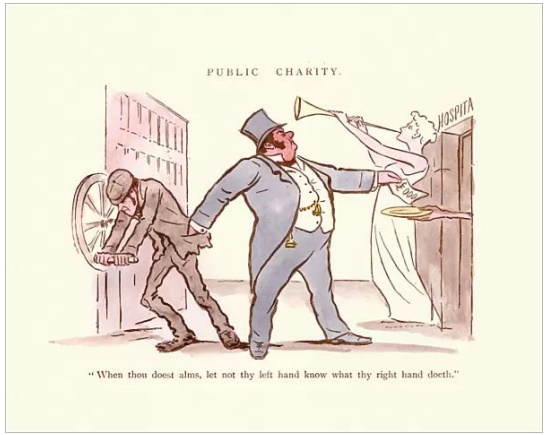
- the simplest way to look at morals is as a **given set** of acceptable and unacceptable behaviours
- examples:
  - eternal truths expressed by religions
  - natural law, as embedded within ourselves
  - innate moral judgments observed in babies





…vs morals as socially constructed…

- morals map to **societal layers** (*niches, strata*, *classes*, *communities, …*)
- **behaving “well” is source/signal of (good) reputation within a class… but not necessarily in others!**
- see e.g. charity events for higher-classes, or mafia/gang-like practices



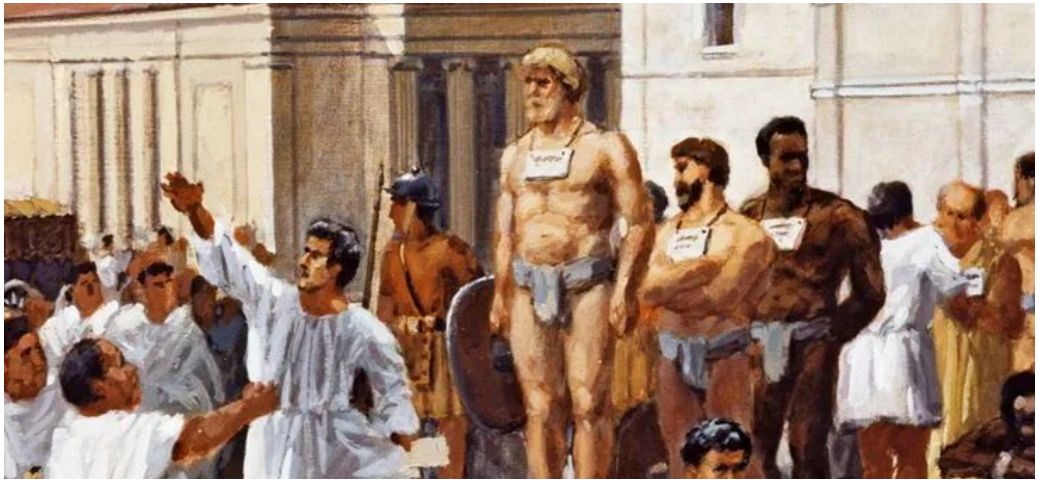
…vs morals as applied ethics…

- another way to look at morals would be as a set of acceptable/unacceptable behaviour ***derived from ethical principles***

(stoics applying emotional detachment)

…vs morals as *materially* constructed!

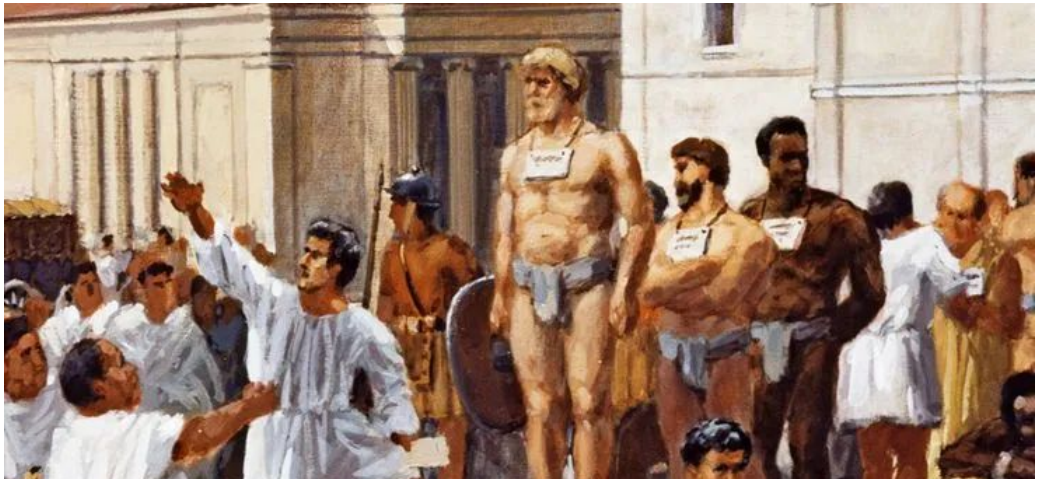
- but then why slavery exists?



- slavery is accepted as long as it sustains (the power of) the people in power.

…vs morals as *materially* constructed!

- but then why slavery exists?



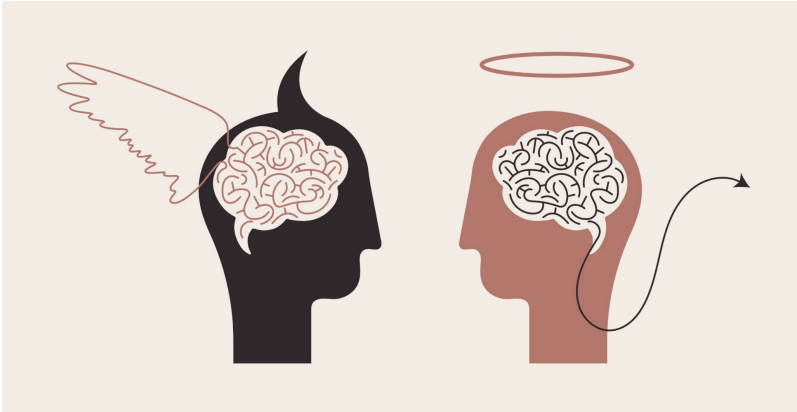
- slavery is accepted as long as it sustains (the power of) the people in power.

→ **morals are socio-economically and historically conditioned**

Norms and values?

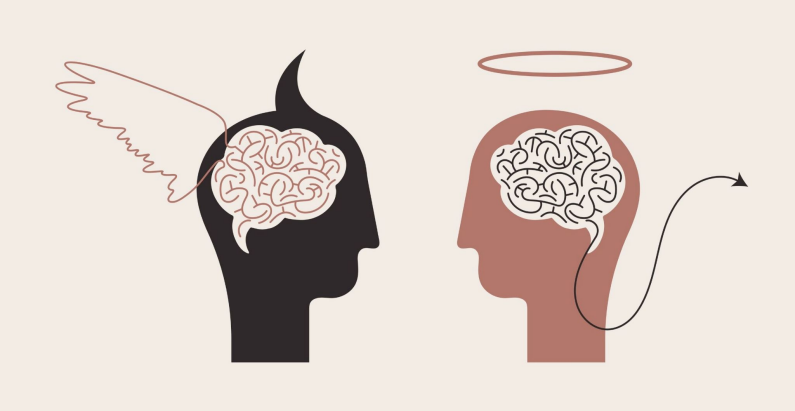
Norms vs values in ethics

- There is a distinction between **right/wrong** (norms) vs **good/bad** (values)



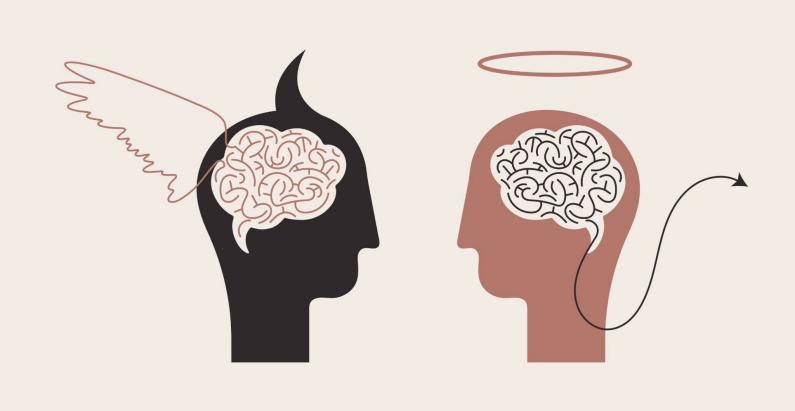
Norms vs values in ethics

- There is a distinction between **right/wrong** (norms) vs **good/bad** (values)
- **Deontology** studies what features make an action right or wrong.



Norms vs values in ethics

- There is a distinction between **right/wrong** (norms) vs **good/bad** (values)
- **Deontology** studies what features make an action right or wrong.
- **Axiology** studies what makes things good (or have value) or bad (or have disvalue or less value) ⇒ **theories of value**



Norms vs values in ethics

- There is a distinction between **right/wrong** (norms) vs **good/bad** (values)
- **Deontology** studies what features make an action right or wrong.
- **Axiology** studies what makes things good (or have value) or bad (or have disvalue or less value) ⇒ **theories of value**

(What comes first?)

Theory of value (philosophy)

Philosophy has long been debating on the distinction between

- **intrinsic values** (something good in itself)
    - pleasure/absence of pain, satisfaction of appetites, needs, desires?
- **instrumental/final values** (something good for something else)
    - eg. money

Theory of value (economics)

- **Labour**
- Utility (use)
- Exchange
- ...

how much its production costs
cf. **cryptocurrencies** as **blockchain**

Theory of value (economics)

- Labour
- **Utility (use)**
- Exchange
- ...

how much benefit it brings
cf. **software**

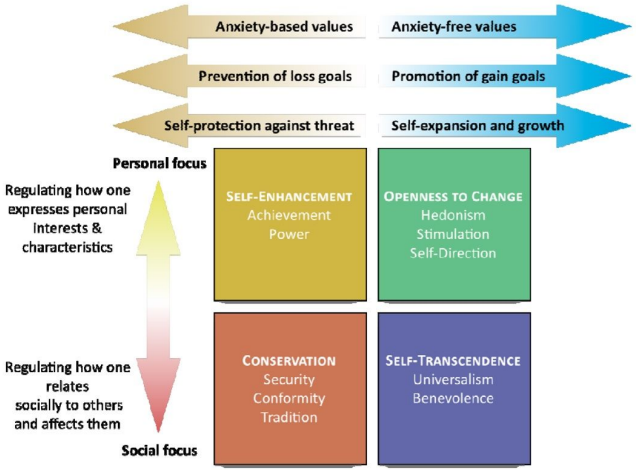
Theory of value (economics)

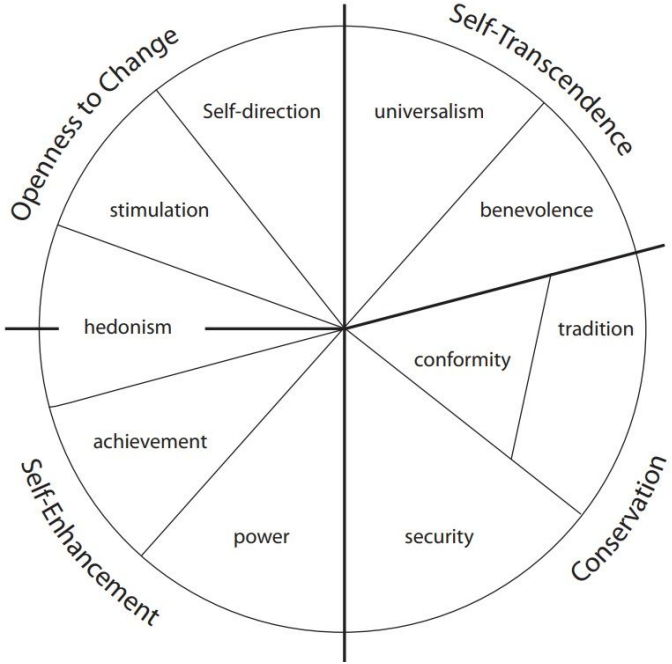
- Labour
- Utility (use)
- **Exchange**
- ...

how much others would pay for it
cf. **NFT (non-fungible tokens)**

Theory of value (psychology)

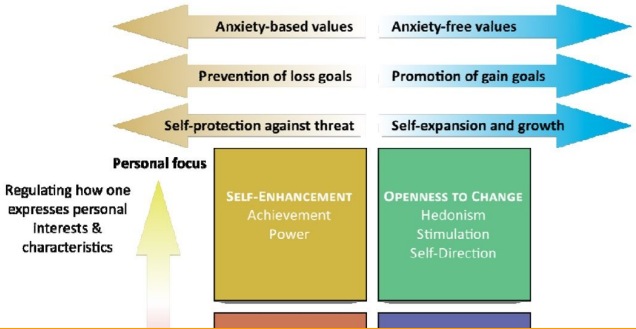
- Values are abstract concepts, but how people express them may be categorized
- eg. Schwartz's value framework

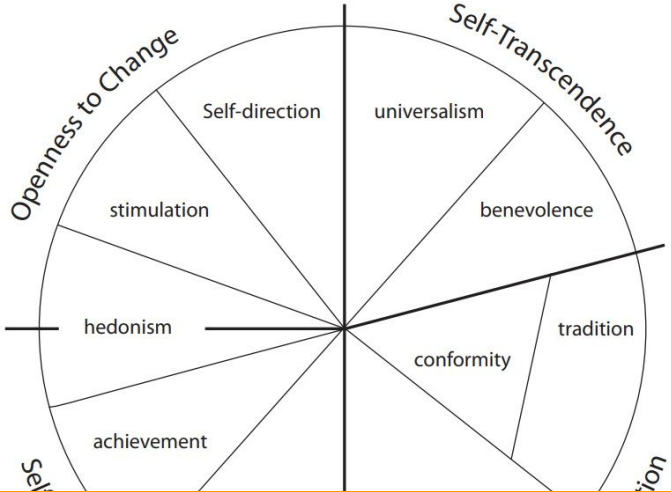




Theory of value (psychology)

- Values are abstract concepts, but how people express them may be categorized
- eg. Schwartz's value framework





several papers are using it to interpret LLMs chatbots.
but what are they capturing?

Theory of value (law)

- Judges embody the law
- For common cases, norms provide sufficient guidance.
- For *hard cases*, values fits into, used for balancing between opposing interests/norms, eg. environmental protection vs economic opportunity

Theory of value (law)

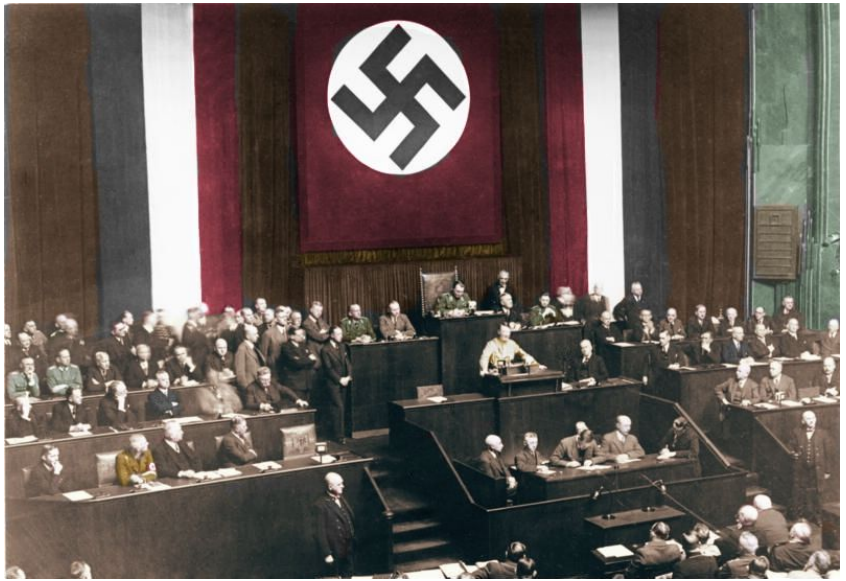
- Increasing reference to human rights and the rule of law in legislation, eg.

  The EU's founding values are 'human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities'; Article 3 Treaty European Union (TEU)

- Talking about values in laws pushes the legislator to take some stance towards them, but cannot guarantee effective application…

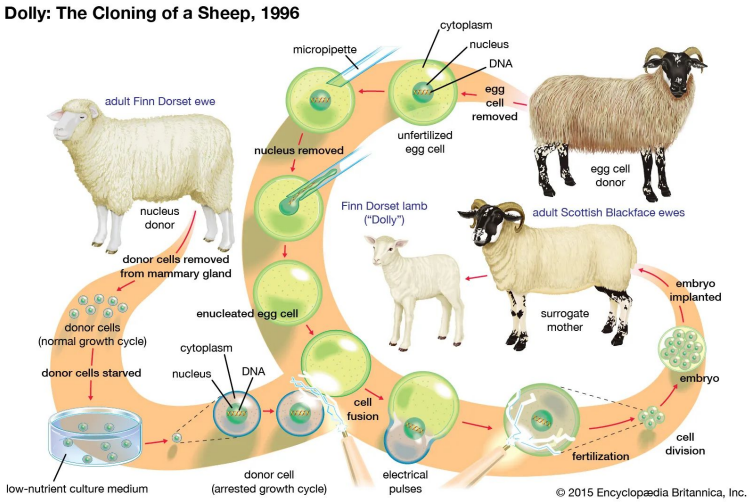
Beyond norms: when laws are written

- There is a distinction between ***state of right*** vs ***state with right/with laws***



Beyond norms: when laws are not there

- When laws are not there (yet), ethics is deemed to play a role (eg. bioethical committees).

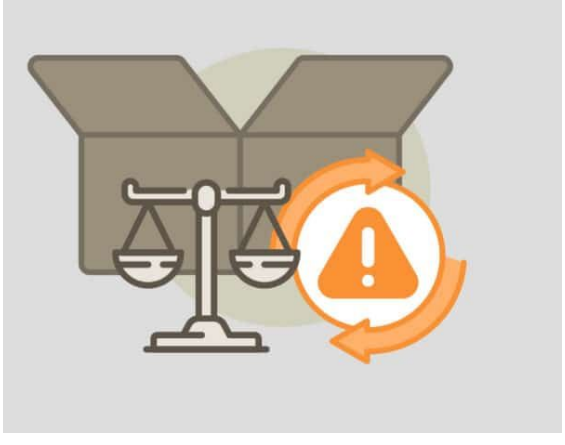


Beyond norms: when laws are not there

- When laws are not there (yet), ethics is deemed to play a role (eg. bioethical committees).
- Part of the AI Ethics track started from a similar perspective.

Beyond norms: when laws are not there

- When laws are not there (yet), ethics is deemed to play a role (eg. bioethical committees).
- Part of the AI Ethics track started from a similar perspective.
- It has been argued however that most of the problems are already covered by existing laws (eg. ***product liability***).



Core components of ethical frameworks

Main ethical constructs

- **deontic**: you perform an action because it is due

Main ethical constructs

- **deontic**: you perform an action because it is due
- **consequentialist**: you pursue an end, and for achieving it you perform an action

Main ethical constructs

- **deontic**: you perform an action because it is due
- **consequentialist**: you pursue an end, and for achieving it you perform an action
- **virtue ethics**: you behave as a virtuous person would do (exemplary role-models)

Main ethical constructs

- **deontic**: you perform an action because it is due
- **consequentialist**: you pursue an end, and for achieving it you perform an action
- **virtue ethics**: you behave as a virtuous person would do (exemplary role-models)
- **contractualism**: you behave on the basis a (often unspoken) social contract between all members of society of which you are part.

Main ethical constructs

- **deontic**: you perform an action because it is due
- **consequentialist**: you pursue an end, and for achieving it you perform an action
- **virtue ethics**: you behave as a virtuous person would do (exemplary role-models)
- **contractualism**: you behave on the basis a (often unspoken) social contract between all members of society of which you are part.
- **discourse ethics**: you behave as it is acceptable (in argumentation) to all those affected by the consequences of your action.

see eg. Mingers, John and Walsham, Geoff. Towards Ethical Information Systems: The Contribution of Discourse Ethics. MIS Quarterly, 34 (4) (2010)

Main ethical constructs

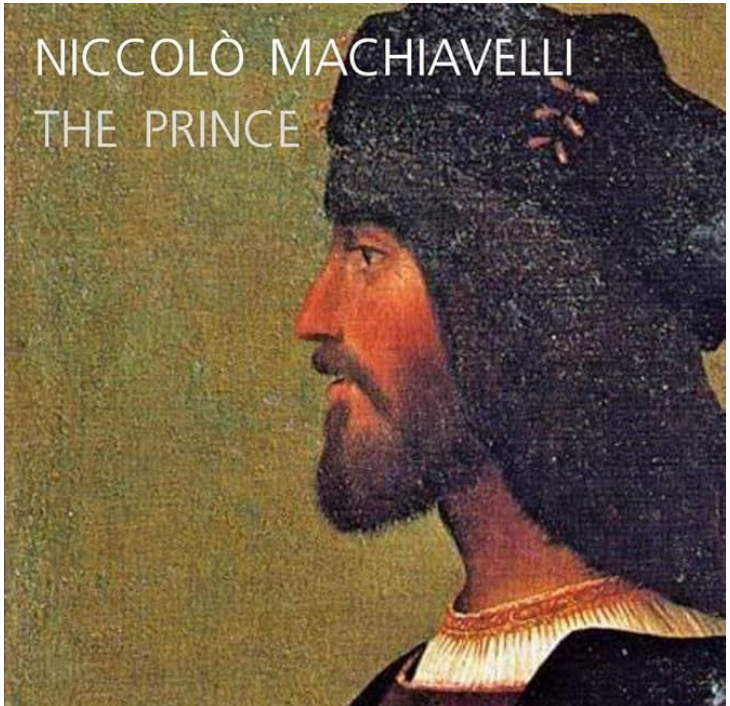
These express different forms of normativity (possibly conflicting)!

- **deontic**: you perform an action because it is due
- **consequentialist**: you pursue an end, and for achieving it you perform an action
- **virtue ethics**: you behave as a virtuous person would do (exemplary role-models)
- **contractualism**: you behave on the basis a (often unspoken) social contract between all members of society of which you are part.
- **discourse ethics**: you behave as it is acceptable (in argumentation) to all those affected by the consequences of your action.

see eg. Mingers, John and Walsham, Geoff. Towards Ethical Information Systems: The Contribution of Discourse Ethics. MIS Quarterly, 34 (4) (2010)

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

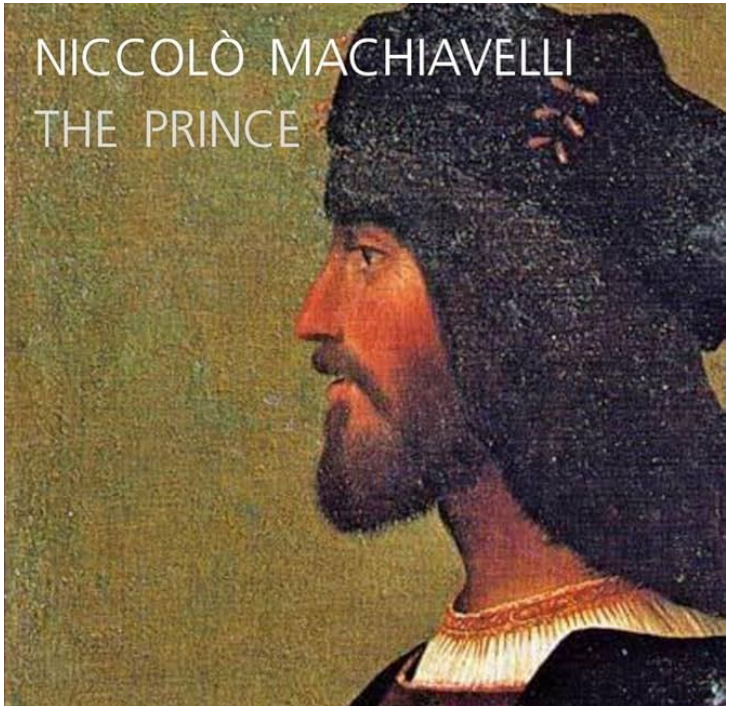


“the ends justify the means”

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

instrumental reasoning patterns…



“the ends justify the means”

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

- consequentialism directly connects with **utilitarianism**…
  - (consequences are evaluated in terms of *utility function*)
- …as well as with **optimization** methods:
  - (you *maximize* utility, expressed eg. as an aggregated reward function)

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

- consequentialism directly connects with **utilitarianism**…
  - (consequences are evaluated in terms of *utility function*)
- …as well as with **optimization** methods:
  - (you *maximize* utility, expressed eg. as an aggregated reward function)

*AI is often **by-design** consequentialist!*

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

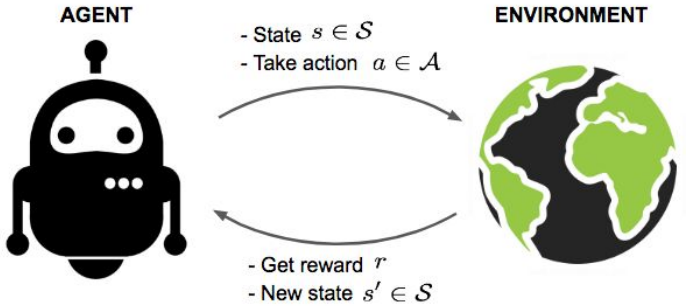
- If the core policy of the agent is seen as its identity, consequentialism can be seen as a *meta-identity*, as it drives a modification of the agent's identity.

Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

- If the core policy of the agent is seen as its identity, consequentialism can be seen as a *meta-identity*, as it drives a modification of the agent's identity.

→ **reinforcement learning** as well as **pure economic rationality** (maximizing profit/minimizing losses) takes a specific normative stance.



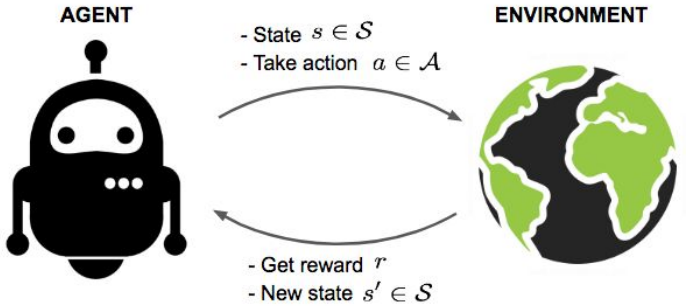
Consequentialism & co.

consequentialist: *you pursue an end, and for achieving it you perform an action*

- If the core policy of the agent is seen as its identity, consequentialism can be seen as a *meta-identity*, as it drives a modification of the agent's identity.

→ **reinforcement learning** as well as **pure economic rationality** (maximizing profit/minimizing losses) takes a specific normative stance.

Is this adequate to behave ethically?



An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

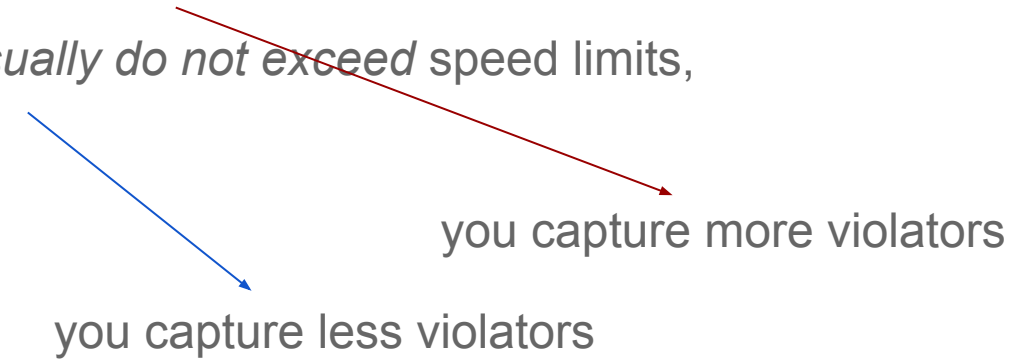
- Where to put the speed cameras?
  - areas in which people *usually exceed* speed limits,
  - areas in which people *usually do not exceed* speed limits,
  - …?

(You have only a limited amount of speed cameras)

An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

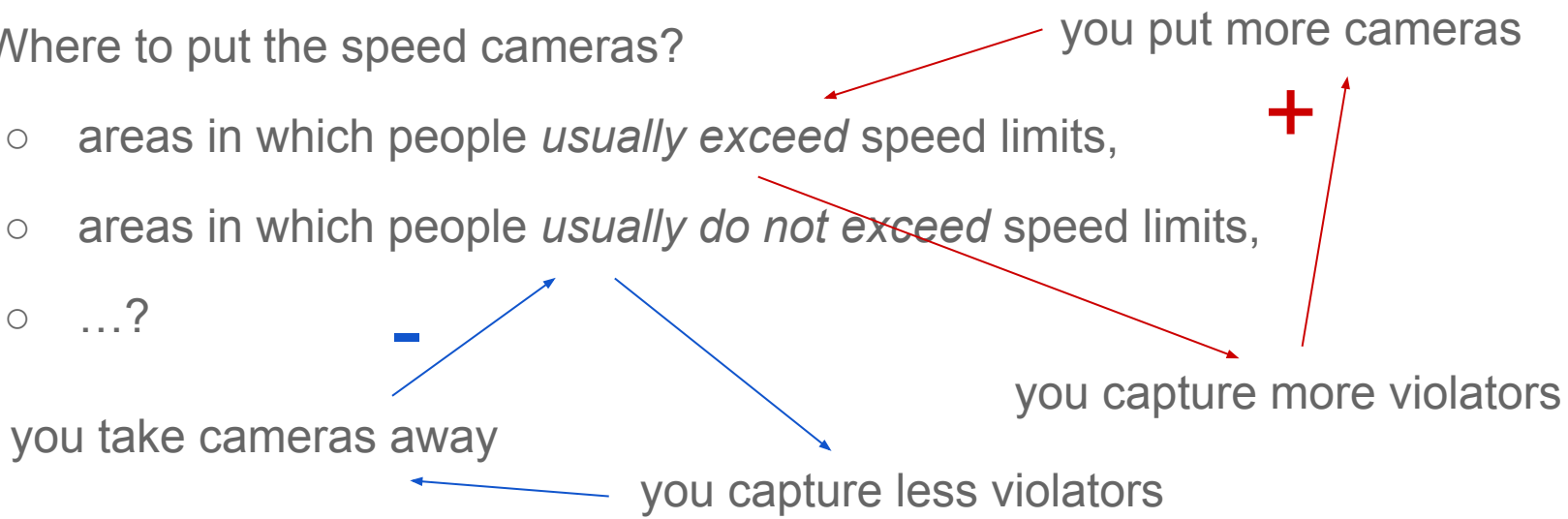
- Where to put the speed cameras?
  - areas in which people *usually exceed* speed limits,
  - areas in which people *usually do not exceed* speed limits,
  - …?



An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

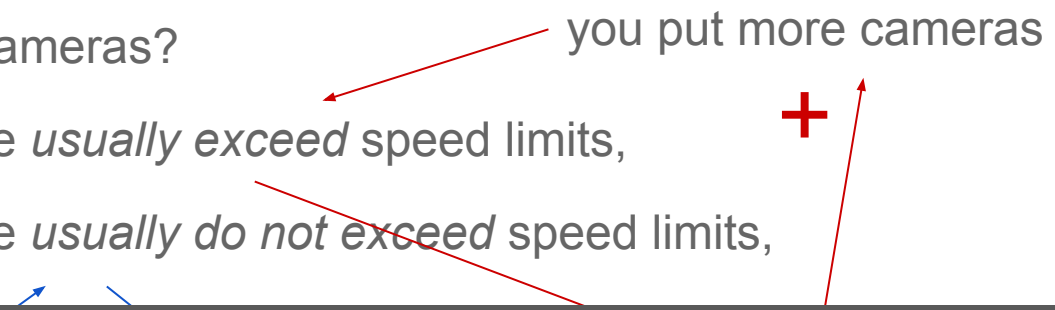
- Where to put the speed cameras?
  - areas in which people *usually exceed* speed limits,
  - areas in which people *usually do not exceed* speed limits,
  - …?



An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

- Where to put the speed cameras?
  - areas in which people *usually exceed* speed limits,
  - areas in which people *usually do not exceed* speed limits,



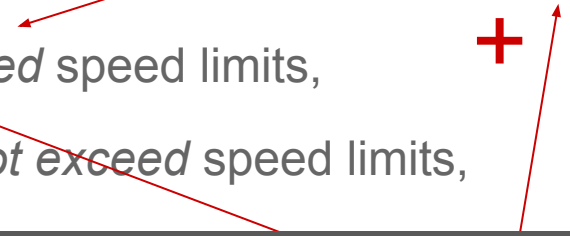
This is an example of *self-fulfilling prophecy*:
if we look for what we expect (only) where we expect,
then we'll (only) see what we expect.

An exercise of monitoring design

In your city the number of car accidents is rapidly increasing. You are asked to find who is exceeding the speed limits (to punish violators).

- Where to put the speed cameras?
  - areas in which people *usually exceed* speed limits,
  - areas in which people *usually do not exceed* speed limits,

you put more cameras



This is an example of *self-fulfilling prophecy*:
if we look for what we expect (only) where we expect,
then we'll (only) see what we expect.

blindness consequent to blind optimization attitude!

An exercise of monitoring design (2)

You are asked to help the police to identify venues of synthetic drug production (an actual Master IS project a few years ago).

An exercise of monitoring design (2)

You are asked to help the police to identify venues of synthetic drug production (an actual Master IS project a few years ago).

- Synthetic drug is usually produced in barns rented for a few months, then abandoned, and chemical residuals thrown in the canals.
- Agriculture is not rentable at the moment, barn owners may be more lenient in checking who is renting their barn

An exercise of monitoring design (2)

You are asked to help the police to identify venues of synthetic drug production (an actual Master IS project a few years ago).

- Synthetic drug is usually produced in barns rented for a few months, then abandoned, and chemical residuals thrown in the canals.
- Agriculture is not rentable at the moment, barn owners may be more lenient in checking who is renting their barn.

Students: **“Let us build a *risk indicator*: if an area is becoming poorer we may expect barns be rented for drug production”**

what is wrong with this?

Circumstantial vs direct evidence

- One way to reduce these issues is to:
  - strongly focus on **direct evidence** related to the ***modus operandi***, the addressed behaviour.
  - avoid as much as possible using **circumstantial evidence**, co-occurring properties like socio-economic features

Circumstantial vs direct evidence

- One way to reduce these issues is to:
  - strongly focus on **direct evidence** related to the ***modus operandi***, the addressed behaviour.
  - avoid as much as possible using **circumstantial evidence**, co-occurring properties like socio-economic features

- Eventually, students looked at data related to the pattern: barns on rent, chemical residuals in canals, assuming that residuals were thrown not nearby, but also not too far. Triangulating the data they found relevant instances, some of those confirmed. (the project took a 9)

what about a “terrorism-risk” indicator in border control based on a picture?

Preventing ethical issues?

You define *behavioural* ***boundaries****,* for instance as deontic directives.

For instance:

- **You are not allowed to clone humans.** No matter what.
- **You have to use direct evidence for judgments.** No matter what.

Preventing ethical issues?

You define *behavioural* ***boundaries***, for instance in terms of virtue ethics (behave as the virtuous person would do).

For instance:

- **Hippocratic Oath** for doctors
  https://en.wikipedia.org/wiki/Hippocratic_Oath
- **Archimedian Oath** for engineers
  https://sefi2024.eu/conference/archimedean-oath/

Preventing ethical issues?

You define *behavioural* ***boundaries***, for instance in terms of virtue ethics (behave as the virtuous person would do).

For instance:

- **Hippocratic Oath** for doctors
  https://en.wikipedia.org/wiki/Hippocratic_Oath
- **Archimedian Oath** for engineers
  https://sefi2024.eu/conference/archimedean-oath/

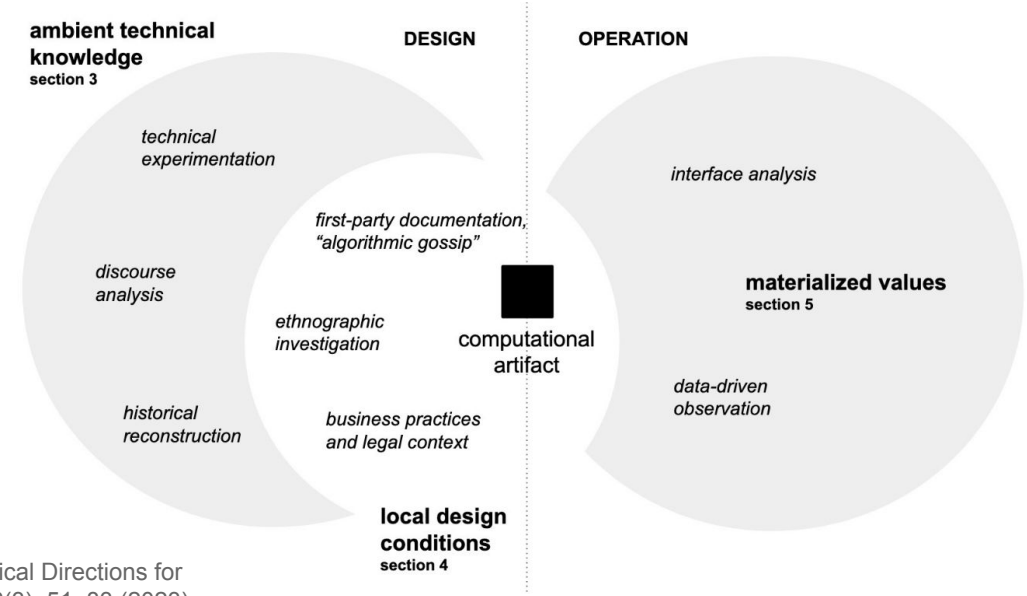
promoting *professional deontology!* (through contractualism)

Identifying ethical issues?

- By taking a designer stance, we cover an internal (prescriptive) view on developed systems, but looking at them **externally** is important too!

Identifying ethical issues?

- By taking a designer stance, we cover an internal (prescriptive) view on developed systems, but looking at them **externally** is important too!
- First, because the reasons why systems behave in certain ways is often inaccessible or not intelligible.
- *Encircling* as a possible option?



Rieder, B., Gordon, G., Sileno, G., Mapping Value(s) in AI: Methodological Directions for Examining Normativity in Complex Technical Systems., Sociologica, 16(3), 51–83 (2023)

Identifying ethical issues?

- By taking a designer stance, we cover an internal (prescriptive) view on developed systems, but looking at them **externally** is important too!
- Second, because even if systems are accessible, **there is a difference between what systems are designed to do, and what they are doing in practice** (a.k.a. “motivating” vs “revealed” aspects) and **how this impact is received by the affected**.

We need instruments to evaluate artefacts and systems independently from the official commitments/requirements (a **descriptive** branch of ethics)

Identifying ethical issues? Several dimensions involved!

- prescriptive vs descriptive ethics
- internal vs external views
- motivating vs revealed aspects
- non-technical vs technical perspectives

Conclusions

In sum

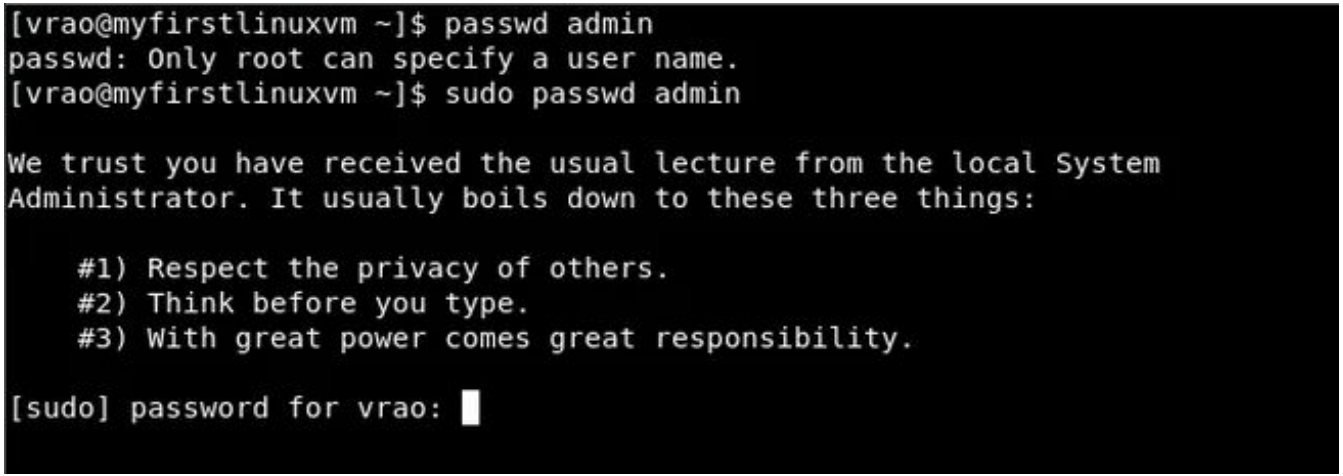
- We looked at the main concepts related to ethics, a few examples of problematic applications, and some insights on algorithmic fairness. Many relevant points were overlooked (eg. privacy, consent, participation).
- These examples demonstrated that ethics does not provide one-fits-all heuristics for ethical concerns. Actually there is no “solution” in **absolute sense**.

In sum

- We looked at the main concepts related to ethics, a few examples of problematic applications, and some insights on algorithmic fairness. Many relevant points were overlooked (eg. privacy, consent, participation).
- These examples demonstrated that ethics does not provide one-fits-all heuristics for ethical concerns. Actually there is no “solution” in **absolute sense**.
- Yet, discussing ethics assists in forming adequate reasons to justify why we accept (or do not accept) an algorithm to work in a certain way.
- This choice is a matter of **individual** and **collective** responsibilities…

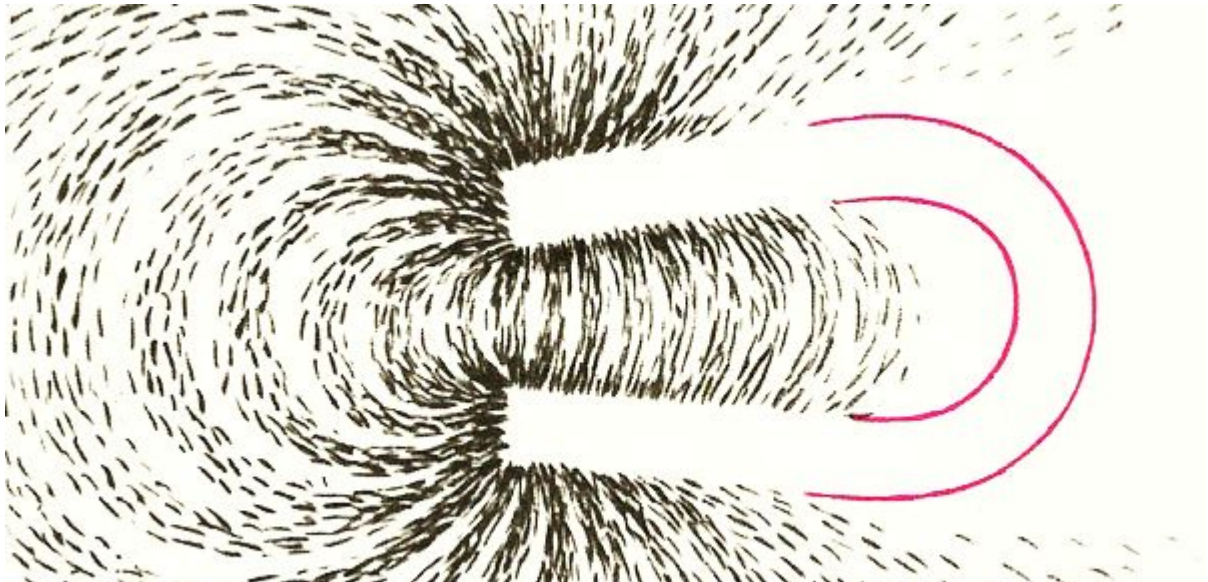
1. With power comes responsibility

- as AI designer/developer, **you have to consider prescriptive ethics** to spell out which norms/values you want your system to realize



2. Your actions have consequences

- as AI designer/developer, and as an social actor, **you have to consider descriptive ethics** to assess which norms/values in practice systems realize in the world



3. Protect your agency: do not be instrumental

- Yet, individual responsibility is only half of the story. Developers are only the last part of the chain.
- Example: *who has responsibility over weapons?*
  - the **designer** who conceived them
  - the **worker** that materially produce them
  - the **company** that sells them
  - the **country** that allows their production and sale.

3. Protect your agency: do not be instrumental

- Yet, individual responsibility is only half of the story. Developers are only the last part of the chain.
- Addressing individual behaviour only puts shame on people who for any reasons are materially obliged to do it.

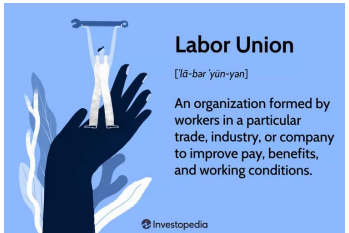
3. Protect your agency: do not be instrumental

- Yet, individual responsibility is only half of the story. Developers are only the last part of the chain.
- Addressing individual behaviour only puts shame on people who for any reasons are materially obliged to do it.
- *What to do then?*

  Participate to steer who has discourse, political and economic power!

ELECTRONIC FRONTIER FOUNDATION

TECH WORKERS COALITION TWC



Ethics for AI Developers

a crash course

4 November 2025

Human-centred Machine Learning course

Giovanni Sileno
g.sileno@uva.nl

University of Amsterdam