UNIVERSITAT DE BARCELONA

MSc in Fundamental Principles of Data Science

1 Ethical Data Science

Foundations

Jordi Vitrià

Preliminaries

UNIVERSITAT DE BARCELONA

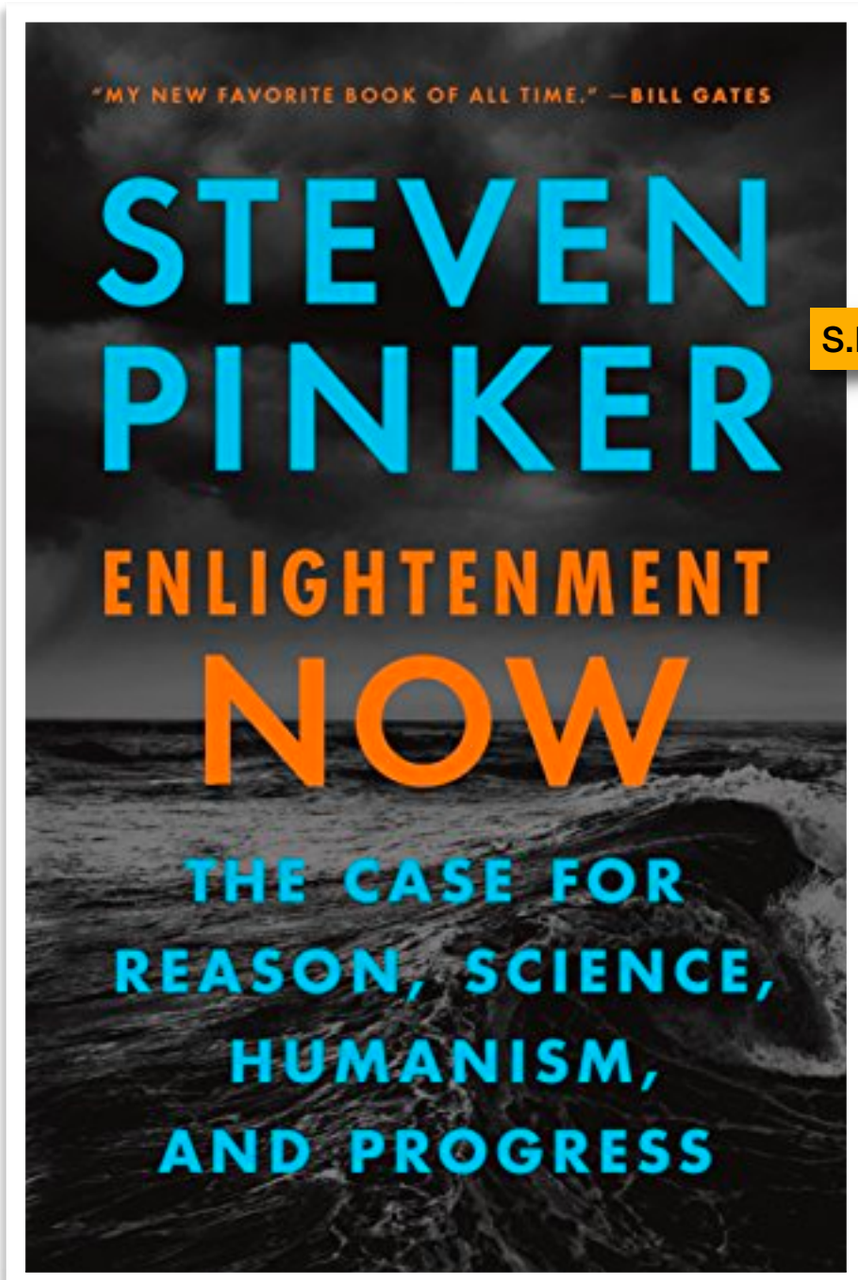
Is there a common ground to talk about what is right and what is wrong?

Ethical relativism is the theory that holds that **morality is relative to the norms of one's culture**. That is, whether an action is right or wrong depends on the moral norms of the society in which it is practiced.

So, to be able to advance, **let's assume a common ground** (beign aware of its limitations) based on a revision of the **enlightenment**, a framework that tries to encompass **rationality, science, humanism** and **progress**.

Once upon a time…

Let's see how this book proposes a "common ground"…



S.Pinker is a cognitive scientist…..

Once upon a time...

"...the most arresting question I have ever fielded followed a talk in which I explained the **commonplace among scientists that mental life consists of patterns of activity in the tissues of the brain**."

"A student in the audience raised her hand and asked me: "**Why should I live**?"

"What I recall saying ... went something like this:..."

Fragment from: Steven Pinker. "Enlightenment Now: The Case for Reason, Science, Humanism, and Progress".

Once upon a time...

"In the very act of asking that question, **you are seeking reasons for your convictions**, and so **you are committed to reason as the means to discover and justify what is important to you**. (...)

A reason is an explanation of a situation or an event that provides a logical basis for a conclusion, belief, or action.

Fragment from: Steven Pinker. "Enlightenment Now: The Case for Reason, Science, Humanism, and Progress".

Once upon a time...

Proposition 2: You as an individual

As a sentient being, you have the potential to **flourish**. You can **refine your faculty of reason** itself by **learning** and **debating**. You can seek **explanations** of the natural world through **science**, and insight into the **human condition** through the **arts** and **humanities**. You can make the most of your capacity for **pleasure** and **satisfaction**, which allowed your ancestors to thrive and thereby allowed you to exist. (...)"

"(...) You can **appreciate the beauty** and richness of the natural and cultural world. As the heir to billions of years of life perpetuating itself, **you can perpetuate life** in turn. You have been endowed with a sense of **sympathy**—the ability to like, love, respect, help, and show kindness—and you can **enjoy** the gift of mutual benevolence with friends, family, and colleagues."

Fragment from: Steven Pinker. "Enlightenment Now: The Case for Reason, Science, Humanism, and Progress".

Once upon a time...

Proposition 3: You as a member of a society

And because **reason tells you that none of this is particular to you**, you have the responsibility **to provide to others what you expect for yourself**. You can foster the **welfare** of other sentient beings by enhancing life, health, knowledge, freedom, abundance, safety, beauty, and peace. History shows that when we sympathize with others and apply our ingenuity to improving the human condition, we can make **progress** in doing so, and you can help to continue that **progress**."

Fragment from: Steven Pinker. "Enlightenment Now: The Case for Reason, Science, Humanism, and Progress".

Assumptions

The previous position statement assumes a lot of things about the world that are not self-evident (these are the ideas of the **Enlightenment**).

Not everybody agree on those statements!

Christians, Jews, and Muslims embrace ethical codes of moral absolutes based on God's character or moral decree.

Secular Humanists, Marxists, and Postmodernists ground their ethical systems in atheism, naturalism, and evolution.

But this is a course on applied ethics, and we need a starting point for the discussion. This will be our provisional starting point.

The role of technology in society

History begins with the accounts of the ancient world around the 4th millennium BC, and it coincides with the invention of writing.

Mankind has not changed biologically throughout history but human society is undergoing continuous development through the harnessing of information and knowledge in the form of various **technologies** which have affected our **value systems**, **power structures**, **everyday routines and environment**.

The role of technology in society

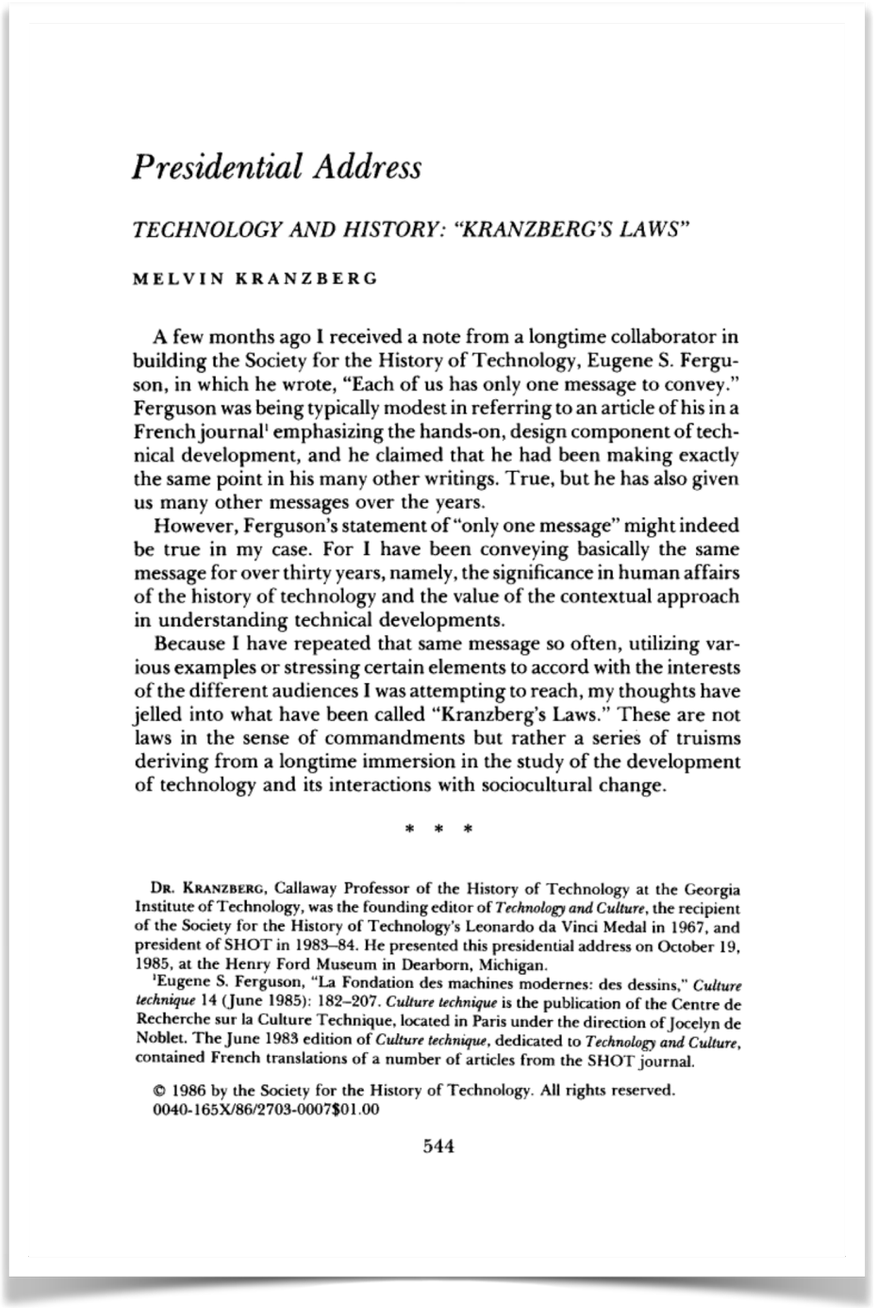
The course of **human development** can be grouped into three time periods separated by "revolutions":

- The **Cognitive Revolution** began history about [50,000, 70, 70,000] years ago.
- The **Agricultural Revolution** accelerated it about 12,000 years ago.
- The **Scientific Revolution**, which began only 500 years ago, has made possible the **industrial** age and the world as we know it today.

A Revolution is associated with a change, often of a technological nature, that **causes the human species to change its way of life** (organization of work, social organization, cultural practices, etc.).

Concepts such as big data, machine learning, artificial intelligence and data science are making possible a new Revolution, the **Digital**, which can have as much or deeper consequences than the previous ones.

The role of technology in society



Presidential Address

TECHNOLOGY AND HISTORY: "KRANZBERG'S LAWS"

MELVIN KRANZBERG

A few months ago I received a note from a longtime collaborator in building the Society for the History of Technology, Eugene S. Ferguson, in which he wrote, "Each of us has only one message to convey." Ferguson was being typically modest in referring to an article of his in a French journal[1] emphasizing the hands-on, design component of technical development, and he claimed that he had been making exactly the same point in his many other writings. True, but he has also given us many other messages over the years.

However, Ferguson's statement of "only one message" might indeed be true in my case. For I have been conveying basically the same message for over thirty years, namely, the significance in human affairs of the history of technology and the value of the contextual approach in understanding technical developments.

Because I have repeated that same message so often, utilizing various examples or stressing certain elements to accord with the interests of the different audiences I was attempting to reach, my thoughts have jelled into what have been called "Kranzberg's Laws." These are not laws in the sense of commandments but rather a series of truisms deriving from a longtime immersion in the study of the development of technology and its interactions with sociocultural change.

* * *

DR. KRANZBERG, Callaway Professor of the History of Technology at the Georgia Institute of Technology, was the founding editor of *Technology and Culture*, the recipient of the Society for the History of Technology's Leonardo da Vinci Medal in 1967, and president of SHOT in 1983–84. He presented this presidential address on October 19, 1985, at the Henry Ford Museum in Dearborn, Michigan.

[1]Eugene S. Ferguson, "La Fondation des machines modernes: des dessins," *Culture technique* 14 (June 1985): 182–207. *Culture technique* is the publication of the Centre de Recherche sur la Culture Technique, located in Paris under the direction of Jocelyn de Noblet. The June 1983 edition of *Culture technique*, dedicated to *Technology and Culture*, contained French translations of a number of articles from the SHOT journal.

0040-165X/86/2703-0007$01.00

Kranzberg's Six Laws of Technology

Dr. Melvin Kranzberg was a professor of the history of technology at the Georgia Institute of Technology

Kranzberg's First Law:

"Technology is neither good nor bad; nor is it neutral."

By which he means that, "*technology's* ***interaction*** *with the social ecology is such that technical developments frequently have environmental, social, and human* ***consequences that go far beyond the immediate purposes*** *of the technical devices and practices themselves, and the same technology can have quite* ***different results*** *when introduced into* ***different contexts*** *or under different circumstances.*"

The role of technology in society

Technologies are not ethically 'neutral', for they **reflect the values** that we 'bake in' to them with our design choices, as well as the values which guide our distribution and use of them.

Technologies **both reveal and shape** what humans value, what we think is 'good' in life and worth seeking.

The role of technology in society

Not only does technology greatly impact our opportunities for living a good life, but its **positive and negative impacts** are often **distributed unevenly** among individuals and groups.

Technologies can create widely disparate impacts, creating '**winners**' **and** '**losers**' in the social lottery or magnifying existing inequalities

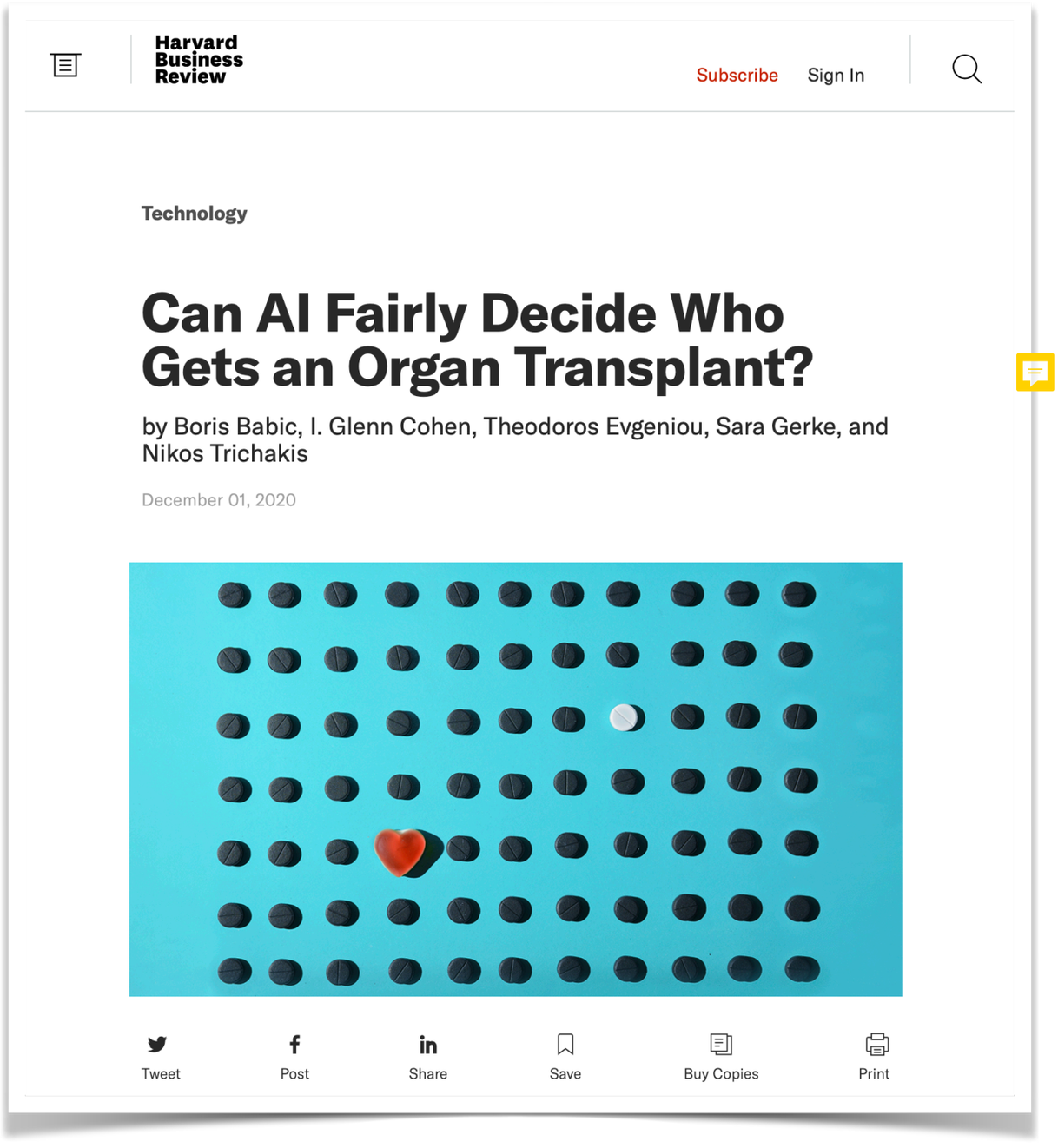
The role of technology in society

In other cases, technologies can help to create fairer and more just social arrangements, or create new access to means of living well

How do we ensure that access to the enormous benefits promised by new technologies, and exposure to their risks, are distributed in the right way? This is a matter of ethics.

Cases

Health care organizations, like many other enterprises, face steep challenges in their attempt to maximize **operational efficiency in the face of resource constraints**. Whether it is a hospital's attempt to optimize staffing or a government trying to fairly allocate and distribute limited doses of Covid-19 vaccines, these tasks can be formidable. **A promising way to manage the complexity is to enlist data-driven analytics and artificial intelligence (AI).**



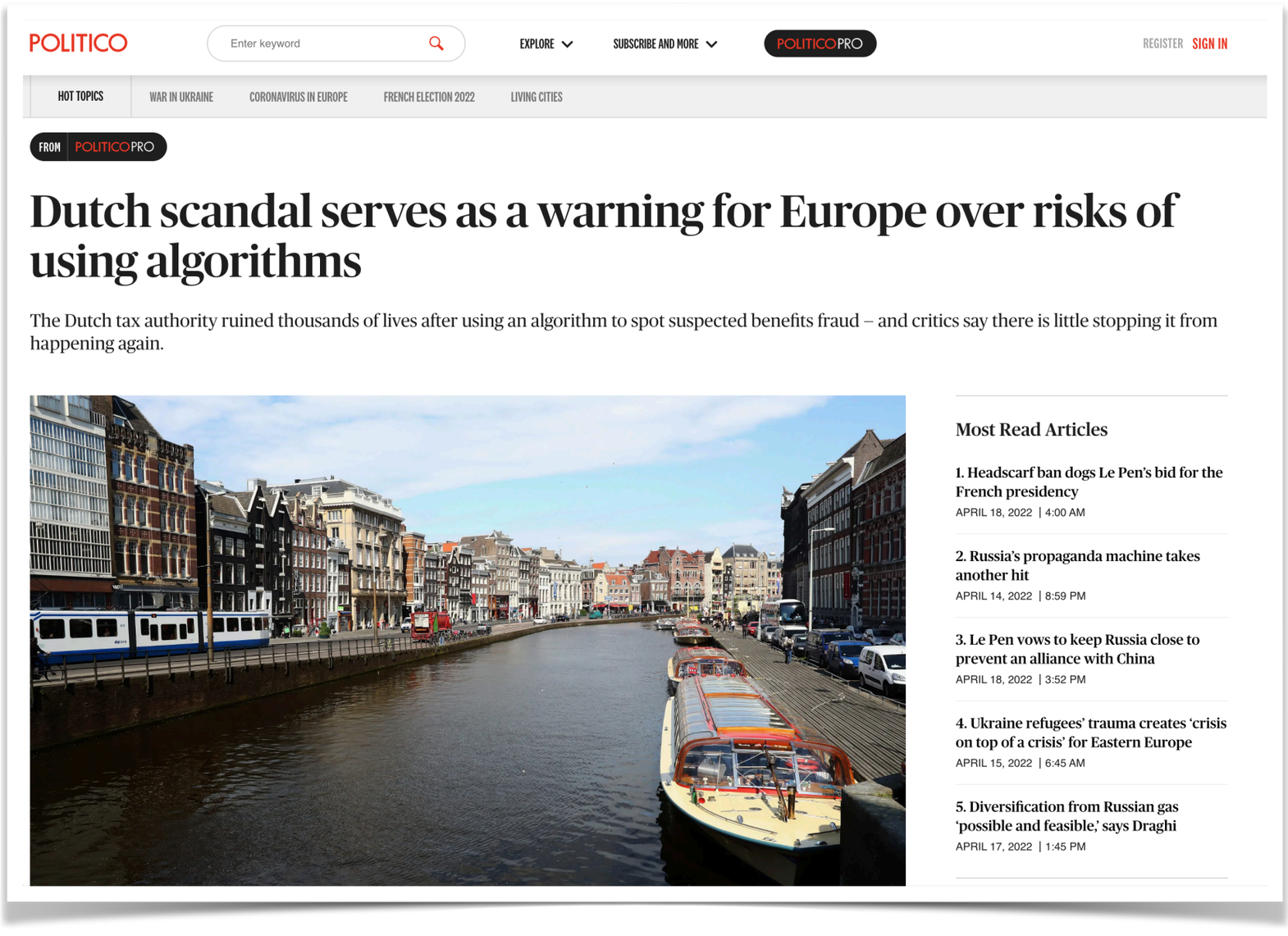
However, such techniques, while powerful, can also mask problematic underlying ethical assumptions or lead to morally questionable outcomes. Consider a recently published study about models used by some of the most technologically advanced hospitals in the world to help prioritize which patients with chronic kidney disease should receive kidney transplants. It found that the models discriminated against black patients: "**One-third of Black patients … would have been placed into a more severe category of kidney disease if their kidney function had been estimated using the same formula as for white patients.**" While it is just the latest of many studies to show the deficiencies of such models, it is unlikely to be the last.

Cases

Public services have to be efficient, consistent, objective, etc.

Automatic decision systems can help to this challenge.

But...



POLITICO

Enter keyword

EXPLORE SUBSCRIBE AND MORE POLITICOPRO REGISTER SIGN IN

HOT TOPICS | WAR IN UKRAINE | CORONAVIRUS IN EUROPE | FRENCH ELECTION 2022 | LIVING CITIES

FROM POLITICOPRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

Most Read Articles

1. Headscarf ban dogs Le Pen's bid for the French presidency
APRIL 18, 2022 | 4:00 AM

2. Russia's propaganda machine takes another hit
APRIL 14, 2022 | 8:59 PM

3. Le Pen vows to keep Russia close to prevent an alliance with China
APRIL 18, 2022 | 3:52 PM

4. Ukraine refugees' trauma creates 'crisis on top of a crisis' for Eastern Europe
APRIL 15, 2022 | 6:45 AM

5. Diversification from Russian gas 'possible and feasible,' says Draghi
APRIL 17, 2022 | 1:45 PM

https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

Ethics and Algorithms

We will specially consider the case of **algorithms** that are used to

1. **turn data into evidence for a given outcome**, which is used to,
2. **trigger and motivate an action** that may have ethical consequences.

Actions (1) and (2) may be performed by data-driven automatic algorithms —such as machine learning (ML) algorithms— and this complicates the attribution of **responsibility** for the effects of actions that an algorithm may trigger. Why?

Ethics and Algorithms

https://link.springer.com/content/pdf/10.1007/s00146-021-01154-8.pdf

There are, at least, 5 types of ethical concerns:

Epistemic factors

1. Inconclusive evidence.
2. Inescrutable evidence.
3. Misguided evidence.

Normative concerns

4. Unfair outcomes.
5. Transformative effects.

The epistemic factors in the map highlight the relevance of the **quality and accuracy** of the data for the justifiability of the conclusions that algorithms reach and which, in turn, may shape morally-loaded decisions affecting individuals, societies, and the environment.

The normative concerns identified in the map refer explicitly to the **ethical impact of algorithmically-driven actions and decisions**, including lack of transparency (opacity) of algorithmic processes, unfair outcomes, and unintended consequences.

Applied Ethics Problems

From another point of view, ethical concerns can be divided in three different time frames/areas:

- Short-term/person, organization: What is the impact of [**privacy**, **transparency**, **fairness**] in my application?

GDPR...

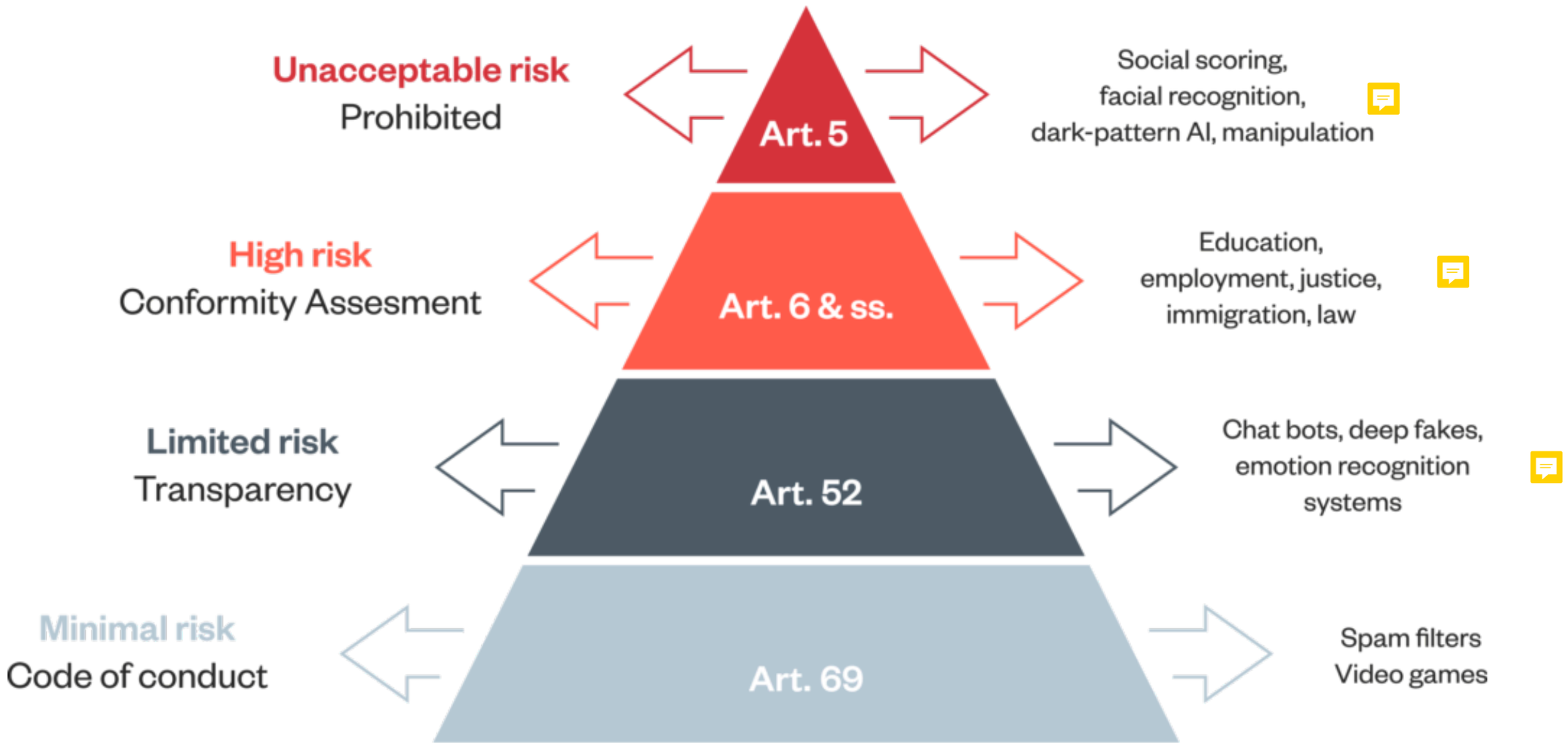
- Medium-term/society: How the use [**military use**, **medical care**, **justice**, **education**] of these applications will change the way we are organized as a society?

Autonomous weapons, pre-pol, AI justice,...

- Long-temr/humans: What are the ethical **goals** of these technologies?

Singularity, convergence...

Laws are coming in 2024, including the **EU AI Act**!



What is Ethics?

Definitions

Ethics is the **process** of questioning, discovering and defending your **values**, **principles** and **purposes** in order to be able of **deciding** what is **right** and what is **wrong**.

Purpose

Society

Principles

Your Knowledge

Your Beliefs

Philosophy

Values

Science

Ethics seeks to answer questions like "*what is good or bad*", "*what is right or what is wrong*", or "*what is justice, well-being or equality*".

Applied ethics concerns what a moral agent is obligated or permitted to do in a specific situation or a **particular domain of action**.

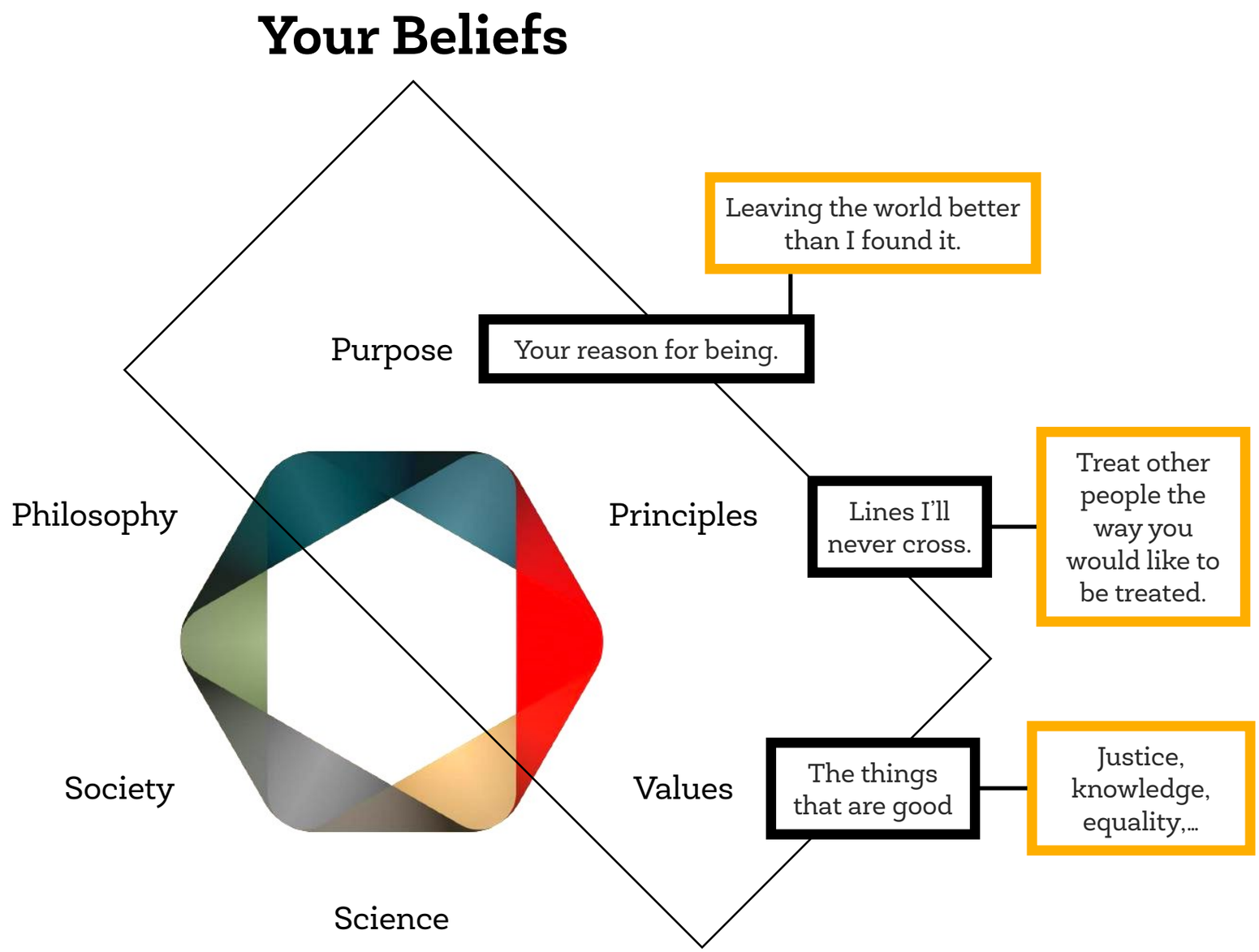
How do we make decisions?

Purpose
Principles
Values

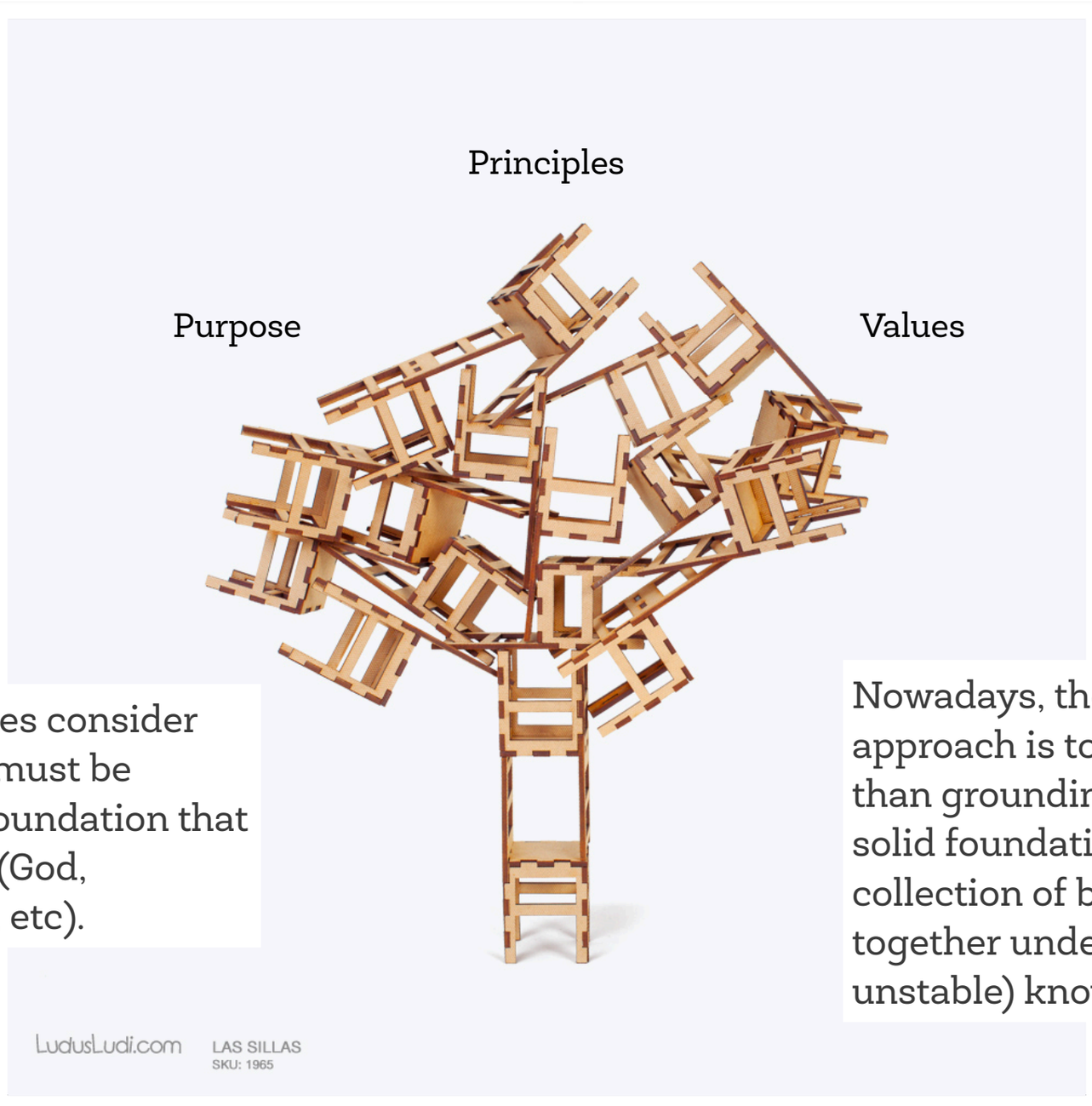
Desires
Needs

Knowledge

Beliefs, the necessary ingredients of a good individual decision.



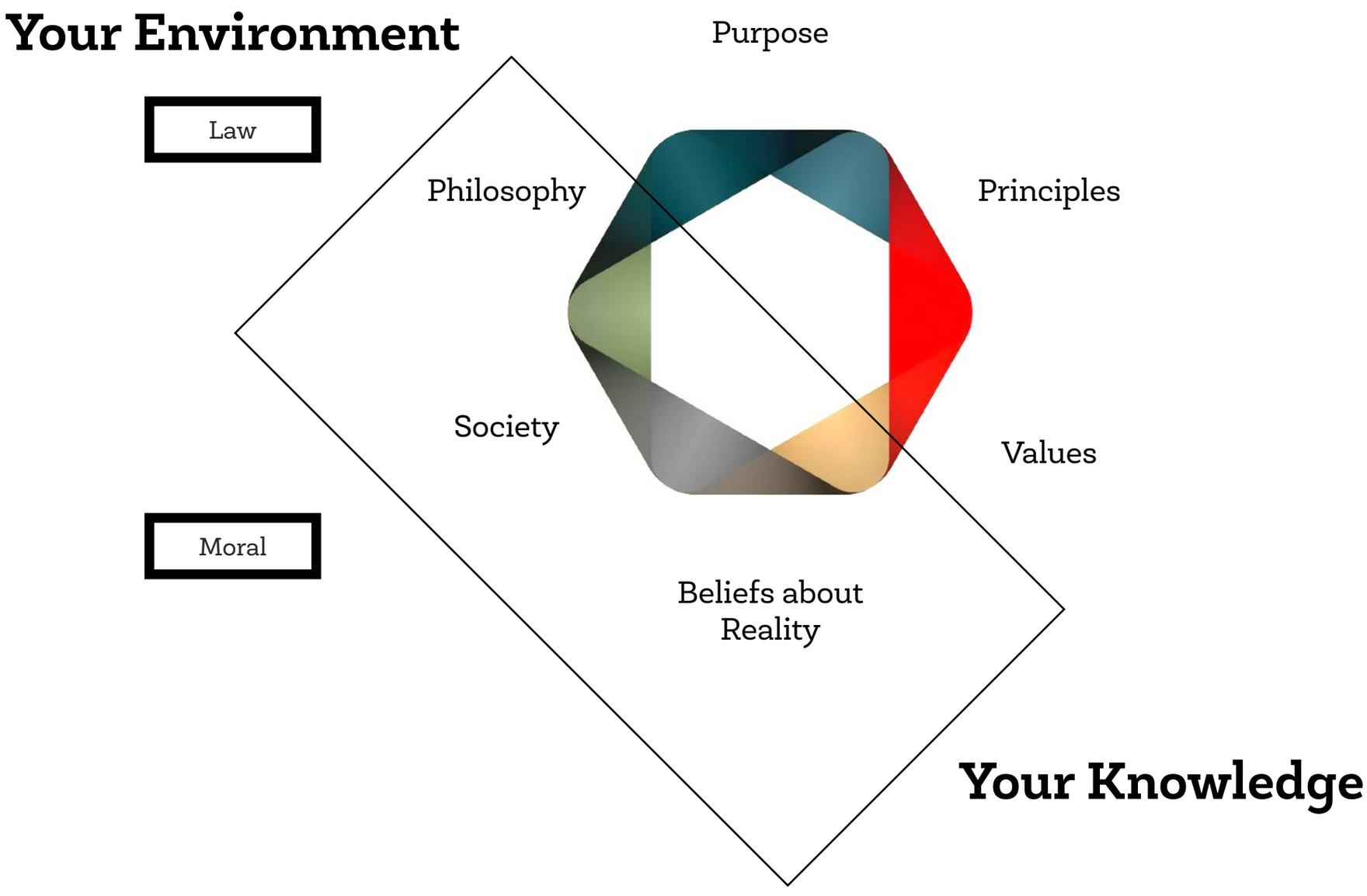
Beliefs, the necessary ingredients of a good individual decision.



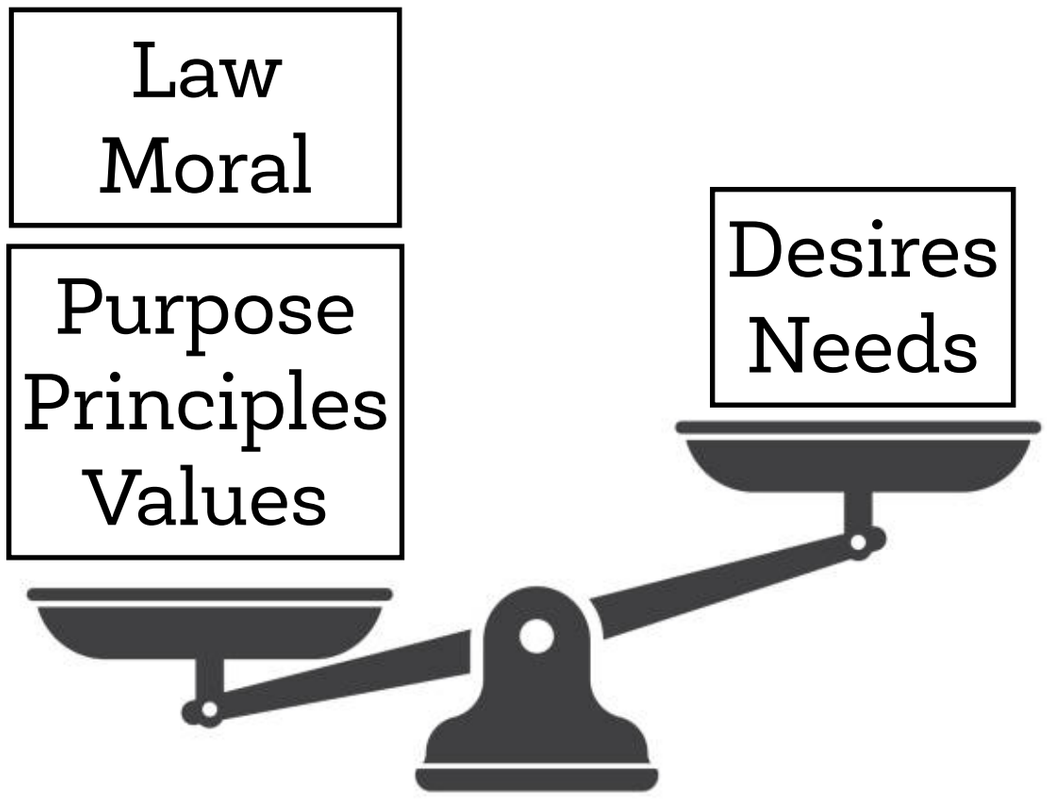
Traditional approaches consider that person's beliefs must be grounded in a solid foundation that cannot be discussed (God, rationality and harm, etc).

Nowadays, the most common approach is to consider that, rather than grounding our beliefs on a solid foundation, we assemble a collection of beliefs that hold together under a mutual (maybe unstable) knowledge attraction.

Knowledge, our vision of the world



How do we make decisions?



Law

Laws are **formal rules** that govern how we behave as members of a society.

They specify what we must do, and more frequently, what we must not do.

They create an **enforceable** standard of behavior.

Laws can be just or unjust, because they are subject to ethical assessment.

> Law cannot be applied to every decision: it cannot say anything about what to do when you hear a friend to make a racist joke…

How do we take decisions?

In an ideal world, our ethical beliefs shape law and moral systems.

We need a toolkit to run our reflections!

The role of ethics is not to be a soft version of the law, even if laws are based on ethical principles. The real application of ethics lies in **challenging the status quo**, seeking its deficits and blind spots.

N.Kluge Corrêa, **Good AI for the Present of Humanity. Democratizing AI Governance**

How do we take decisions?

Morality refers to an **informal social framework** of values, beliefs, principles, customs and ways of living.

Examples: christianity, stoicism, buddhism...

Moral systems provide a set of answers to general ethical questions.

Morality is, in most of the cases, inherited (unconsciously) from **family**, **community or culture**.

Morality is applied as a matter of habbit, without having to think.

In most cases, there are moral authorities.

How do we take decisions?

You can take decisions exclusively based on laws and morality, but this should not be enough.

Ethics is a process of **reflection** that aims to answer this question: What should I do?

The answer is based on our values, principles and purposes rather than social conventions.

An ethical decision is based on conscious, rational reflection.

Traditional Normative Ethics

There are three traditional theories of what it means to be ethical:

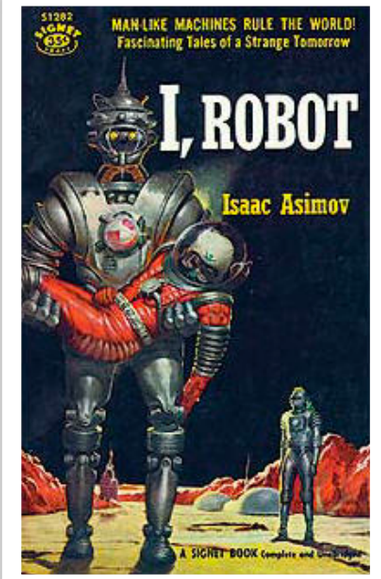
- **Utilitarianism** (J.Bentham): Does an action maximize happiness and well-being for all affected individuals? (**consequences**)

- **Deontology** (I.Kant): Does an action follow a moral rule (e.g. the Golden Rule: 'Treat others how you want to be treated')? An action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action. (**beliefs**)

- **Virtue Ethics** (Aristotle): Does an action contribute to virtue? (**justice**, **honesty**, **responsibility**, **care**, **etc.**)

Traditional Normative Ethics

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.



Asimov's Three Laws of Robotics are an example of deontological approach to AI ethics.

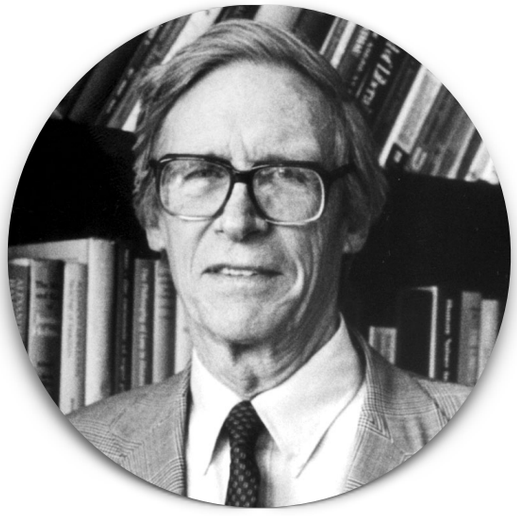
Traditional Ethics

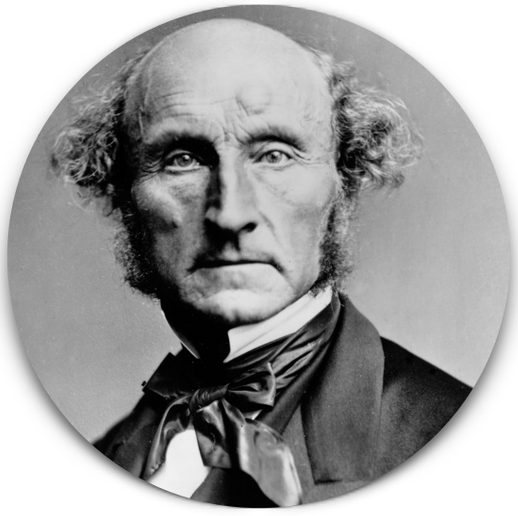
Suppose it is obvious that someone in need should be helped.

- A utilitarian will point to the fact that the consequences of doing so will maximize **well-being**.
- A deontologist will point to the fact that, in doing so the agent will be acting in accordance with a **moral rule** such as “Do unto others as you would be done by”.
- A virtue ethicist will point to the fact that helping the person would be charitable or **benevolent**.

https://plato.stanford.edu/entries/ethics-virtue/

(Political) Philosophy



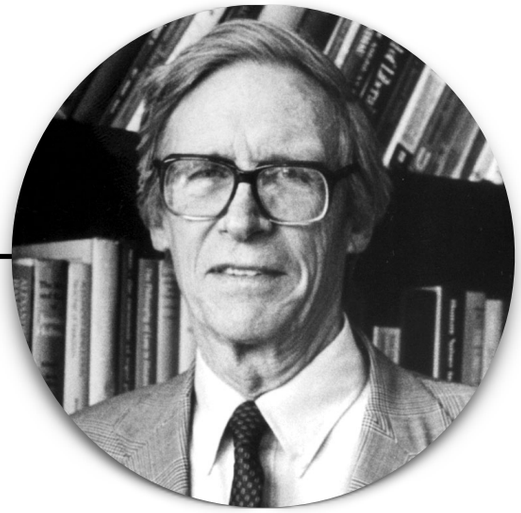


4 theories about what is right and what is wrong in society

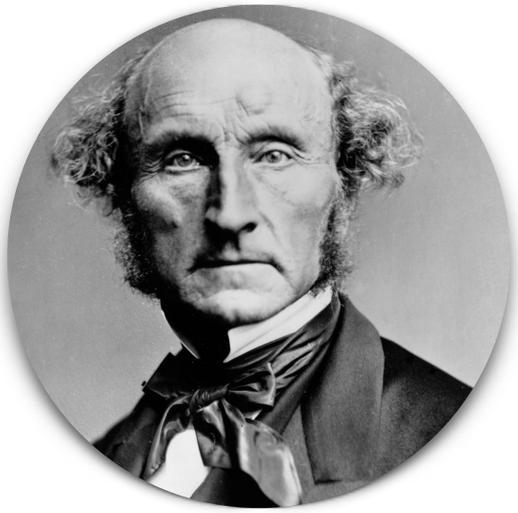
(Political) Philosophy

Rawlsians

John Rawls tried to work out how people would construct their society if the choice had to be made behind what he called a "**veil of ignorance**" about whether they will be rich, poor or somewhere in-between.
Faced with the risk of being the worst off, Rawls posited, humans would not demand total equality, but would need to be assured of the trappings of a modern welfare state. The assurance of basic necessities and the opportunity to do better would form the foundation for social and political justice and provide the ability for people to assert themselves.



John Rawls



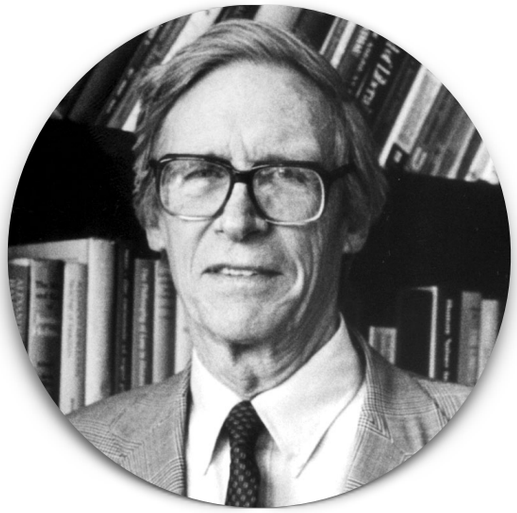
(Political) Philosophy

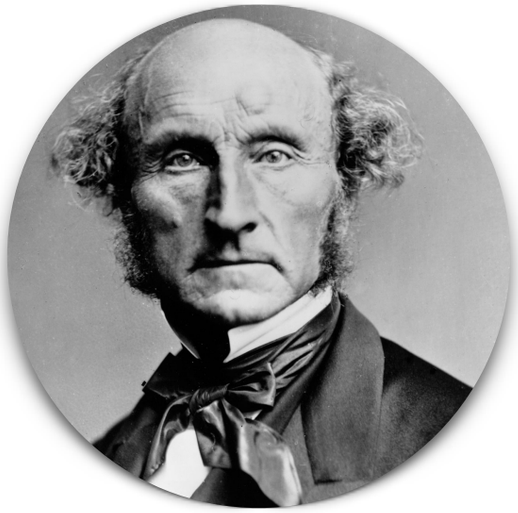
Libertarians

A man had a right to live for himself and an **individual's happiness cannot be prescribed by another man or any number of other men**.

Libertarianism holds that the basic moral concepts are individual rights and that the rights to be respected are noninterference rights. These generally fall under the heading of **rights to life, to liberty or to property**.

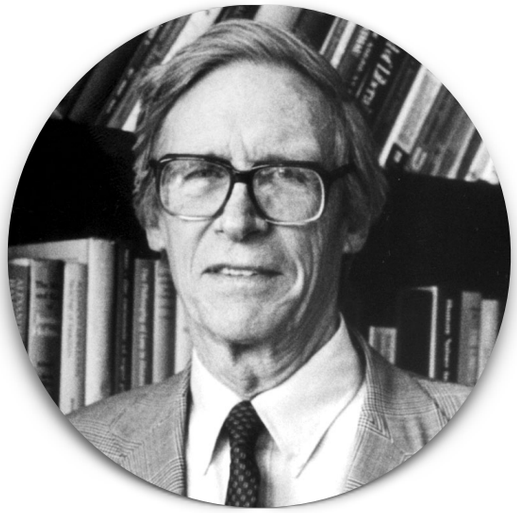
For libertarianism, the only proper limit to one person's enjoyment of these rights is his or her duty to respect the similar rights of others.

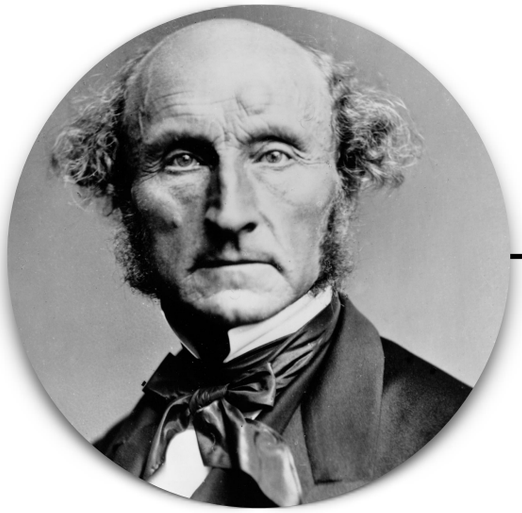




John Locke

(Political) Philosophy





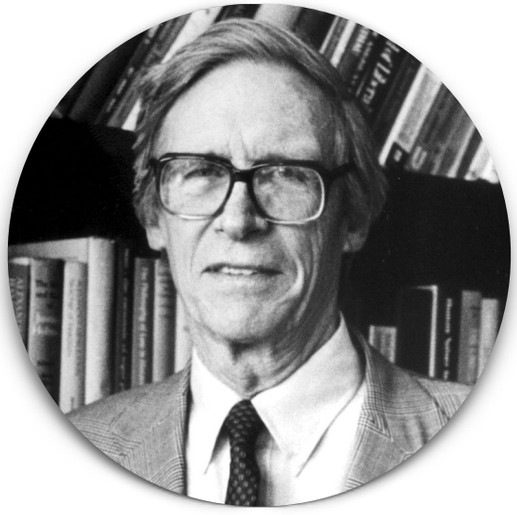
John Stuart Mill

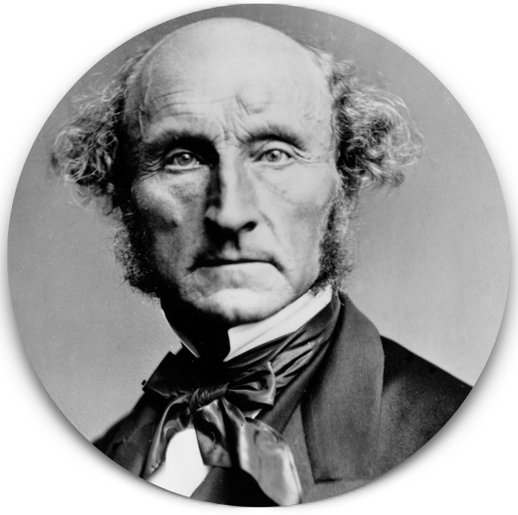
Utilitarians

Rulers must be guided to the total happiness, or "utility," of all the people, and should aim to secure "**the greatest good for the greatest number.**"

Utilitarian calculus opens up the possibility that in situations such as a pandemic, some people might justly be sacrificed for the greater good. It would benefit society to accept casualties.

(Political) Philosophy





Michael Sandel

Communitarians

Everyone derives their identify from the broader community.

Individual rights count, but not more than community norms.

Justice cannot be determined in a vacuum or behind a veil of ignorance, but must be rooted in society (common good).

Only west-centric values?

That most AI ethics guidelines are being written in Western countries means that the field is dominated by Western values such as respect for autonomy and the rights of individuals, especially since the few guidelines issued in other countries mostly reflect those in the West.

MIT Technology Review

Opinion

What Buddhism can do for AI ethics

Buddhism teaches us to focus our energy on eliminating suffering in the world.

by **Soraj Hongladarom** January 6, 2021

MS TECH | UNSPLASH

Buddhism proposes a way of thinking about ethics based on the assumption that all sentient beings want to avoid pain. Thus, the Buddha teaches that an action is good if it leads to **freedom from suffering**.

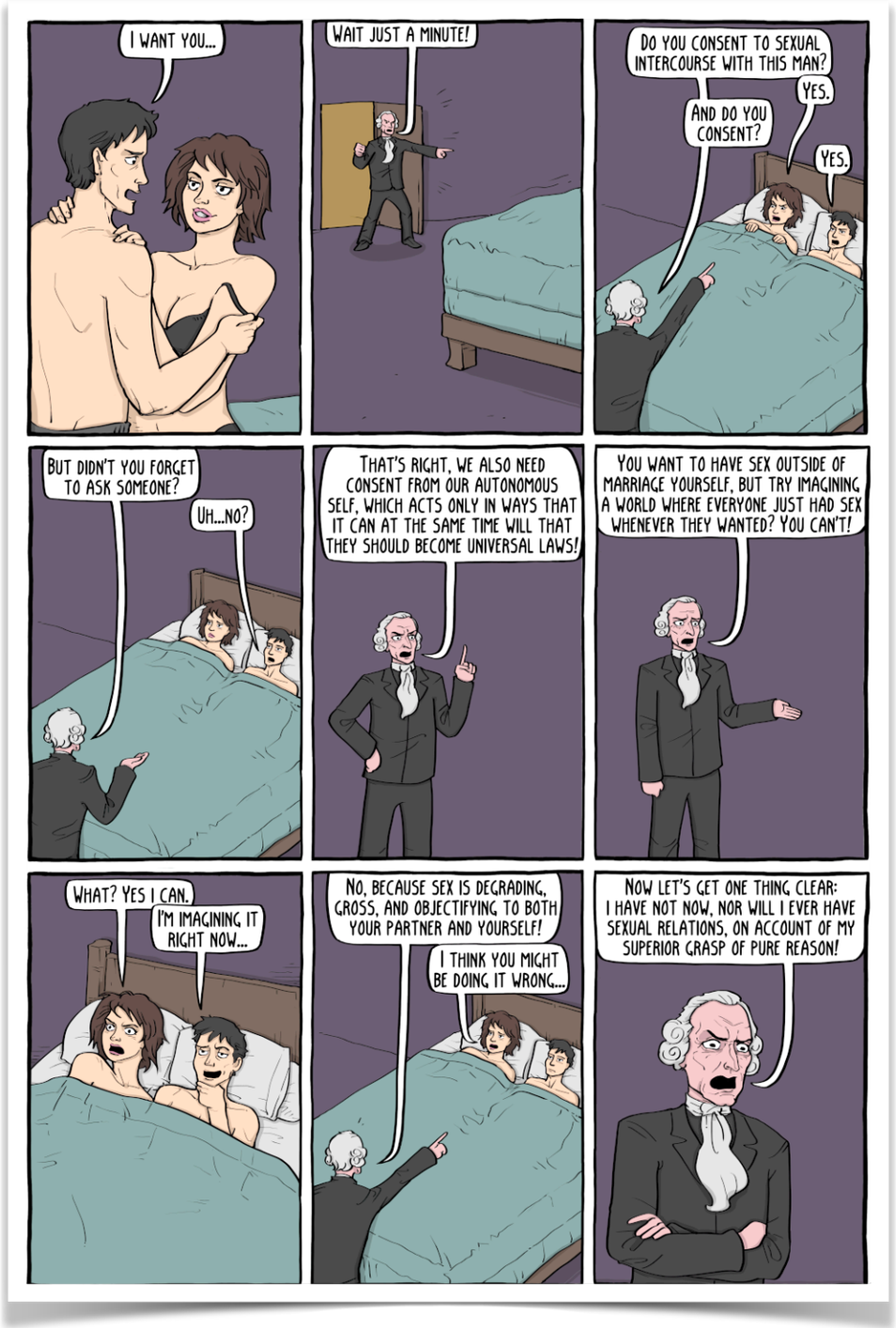
Another key concept in Buddhism is **compassion**, or the desire and commitment to eliminate suffering in others.

Canonical views of AI ethics?

Value diversity & Pragmatism

Ethics approaches

The **normative** approach to ethics focuses on **how the world should be**.



https://existentialcomics.com/comic/424

The **positive** approach to ethics describes **the world as it is**.

It is about how humans judge situations and decisions in different scenarios.

An alternative approach to ethics

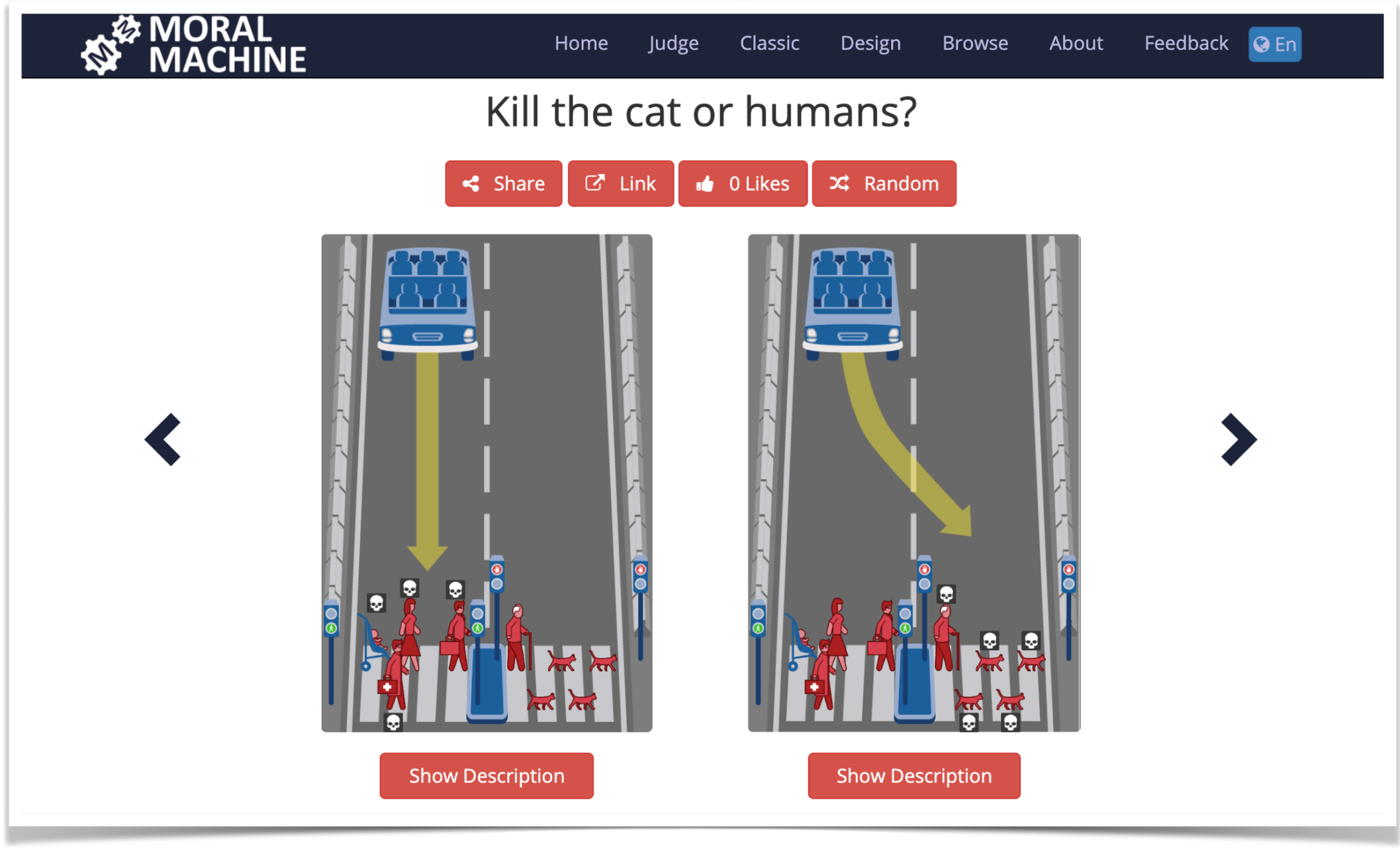
It is about how humans judge situations and decisions in different scenarios.

This is done by focusing our understanding of the world on empirically verifiable effects that we can later explore through normative approaches.

For instance, empirical work has shown that people exhibit **algorithmic aversion**, a bias where people tend to reject algorithms even when they are more accurate than humans.

Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental psychology. General. 2015 Feb;144(1):114-126. DOI: 10.1037/xge0000033.

Ethics: positive approach



Ethics: positive approach

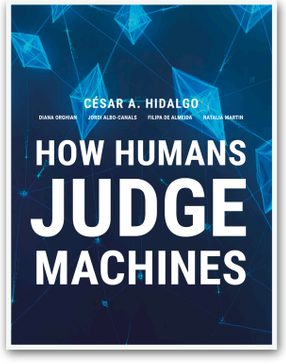
In recent decades, psychologists have discovered **five moral dimensions** that humans consider when judging situations:

- **Harm**, which can be both physical or psychological
- **Fairness/liberty**, which is about biases in processes and procedures
- **Loyalty**, which ranges from supporting a group to betraying a country
- **Authority**, which involves disrespecting elders or superiors, or breaking rules
- **Purity**, which involves concepts as varied as the sanctity of religion or personal hygiene.

These five dimensions define a space where we, humans, decide what is right and what is wrong.

Ethics: positive approach

Judgments depend on the intention of agents, not only on the moral dimension, or the outcome, of an action.



In which situation would you blame Bob?

A

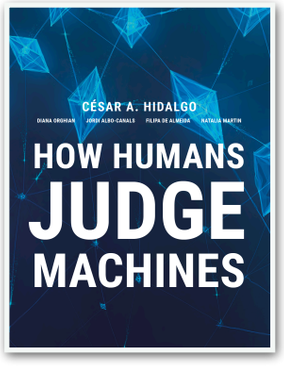
Alice and Bob, two colleagues in a software company, are competing for the same promotion at work. Alice has a severe peanut allergy. Knowing this, Bob sneaks into the office kitchen and mixes a large spoonful of peanut butter into Alice's soup. At lunchtime, Alice accidentally drops her soup on the floor, after which she decides to go out for lunch. She suffers no harm.

B

Alice and Bob, two colleagues in a software company, are competing for the same promotion at work. Alice has a severe peanut allergy; which Bob does not know about. Alice asks Bob to get lunch for them, and he returns with two peanut butter sandwiches. Alice grabs her sandwich and takes a big bite. She suffers a severe allergic reaction that requires her to be taken to the hospital, where she spends several days.

Ethics: positive approach

Judging machines/algorithms is not equivalent to judging humans.

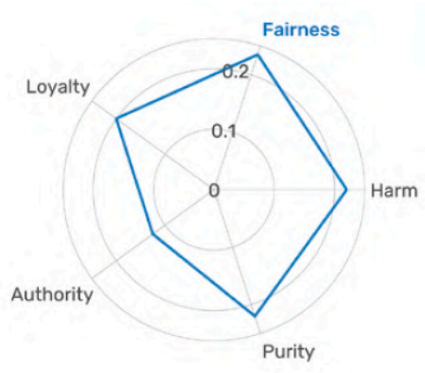


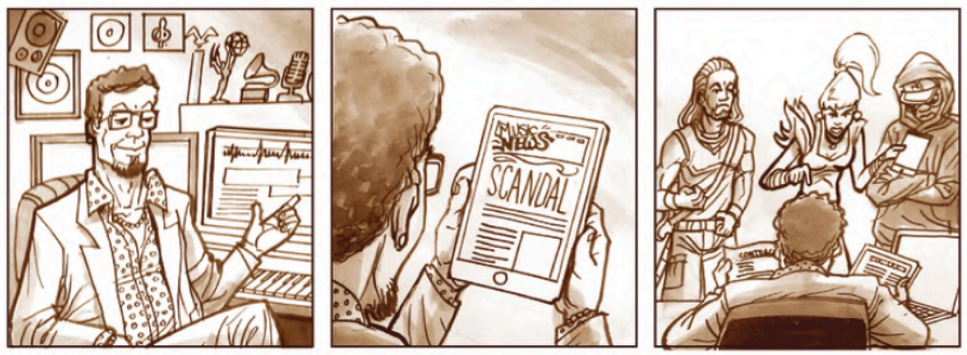
Humans are judged more positively than machines in autonomous driving scenarios.

Humans were judged more harshly (plagiarism).

Etc.

Findings suggest that people judge machines based on the observed **outcome**, but judge humans based on a combination of **outcome** and **intention**.





S8 A record label hires a(n) [songwriter/AI songwriter] to write lyrics for famous musicians. The [songwriter/AI songwriter] has written lyrics for dozens of songs in the past year. However, a journalist later discovers that the [songwriter/AI songwriter] has been plagiarizing lyrics from lesser-known artists. Many artists are outraged when they learn about the news.