UNIVERSITAT DE BARCELONA

MSc in Fundamental Principles of Data Science

1 Ethical Data Science

Foundations

Jordi Vitrià

2020-2021

Why Ethics?

in technology, data science, AI...

Scientific point of view

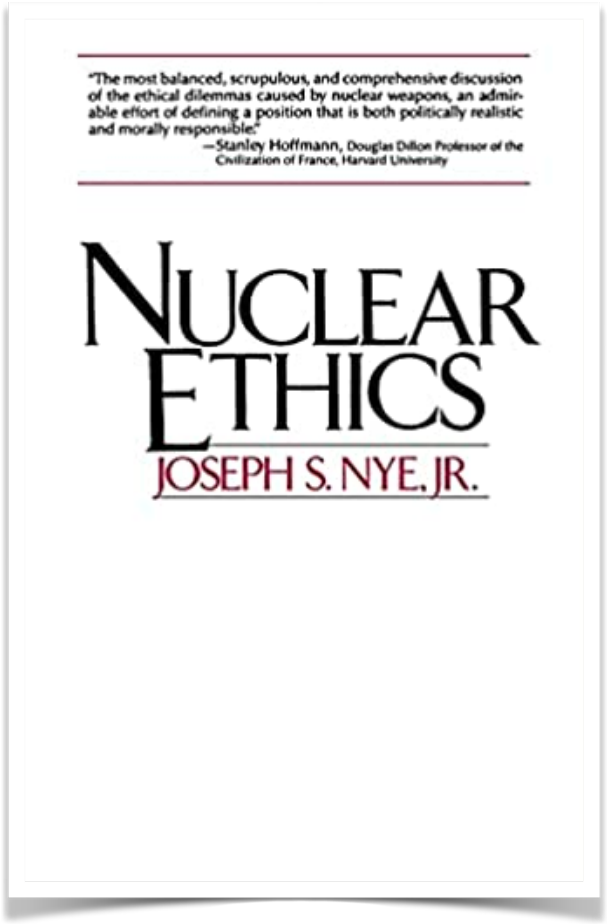
"Everything that is not forbidden by laws of nature is achievable, given the right knowledge"
(Credit: David Deutsch)

But that's the problem.

"Everything" means everything: vaccines and bioweapons, video on demand and Big Brother on the tele-screen. **Something** in addition to science ensured that vaccines were put to use in eradicating diseases while bioweapons were outlawed.

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

Scientific point of view



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.*"

What was the main (unexpected) consequence of the agricultural revolution?
What is the main (unexpected) consequence of the industrial revolution?

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.

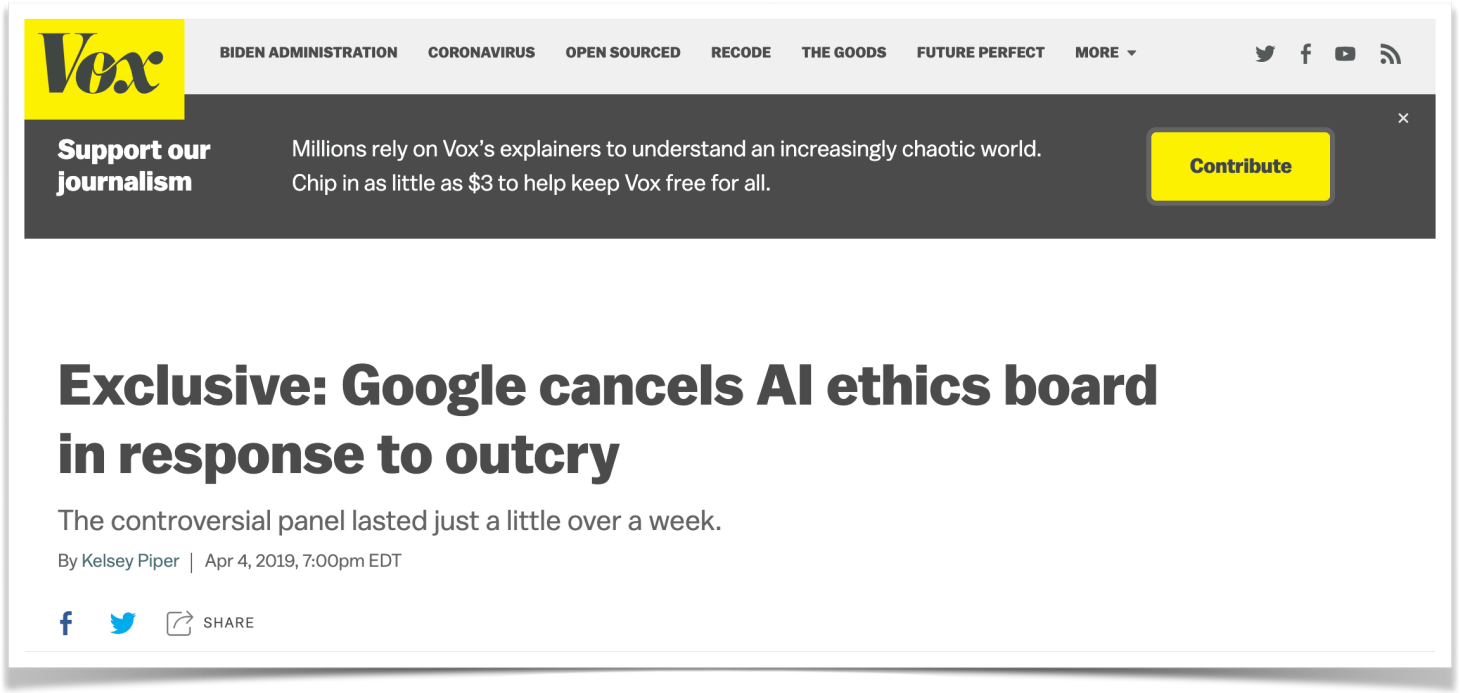
Not only does technology greatly impact our opportunities for living a **good** life, but its **positive and negative impacts are often distributed unevenly** among individuals and groups.

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

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.**

State of the Question (2022)

Industry self-regulation is the process whereby members of an industry, trade or sector of the economy monitor their own adherence to legal, ethical, or safety standards, rather than have an outside, independent agency such as a third party entity or governmental regulator monitor and enforce those standards.



DER TAGESSPIEGEL

LOGIN REGISTRIEREN

POLITIK BERLIN WIRTSCHAFT GESELLSCHAFT KULTUR MEINUNG SPORT WISSEN VERBRAUCHER INTERAKTIV Suchbegriff eingeben

Agenda Brexit Digitalisierung & KI Energie & Klima Gesundheit & E-Health Mobilität & Transport

Politik EU guidelines: Ethics washing made in Europe

Coronavirus in Deutschland – Alle Zahlen im Überblick Hier ansehen

EU guidelines

08.04.2019, 15:48 Uhr

Ethics washing made in Europe

On Tuesday, the EU has published ethics guidelines for artificial intelligence. A member of the expert group that drew up the paper says: This is a case of ethical white-washing. VON THOMAS METZINGER

Pekka Ala-Pietilä, Chair of the AI HLEG
AI Finland, Huhtamaki, Sanoma
Wilhelm Bauer
Fraunhofer
Urs Bergmann – Co-Rapporteur
Zalando
Mária Bieliková
Slovak University of Technology in Bratislava
Cecilia Bonefeld-Dahl – Co-Rapporteur
DigitalEurope
Yann Bonnet
ANSSI
Loubna Bouarfa
OKRA
Stéphan Brunessaux
Airbus
Raja Chatila
IEEE Initiative Ethics of Intelligent/Autonomous Systems & Sorbonne University
Mark Coeckelbergh
University of Vienna
Virginia Dignum – Co-Rapporteur
Umea University
Luciano Floridi
University of Oxford
Jean-Francois Gagné – Co-Rapporteur
Element AI
Chiara Giovannini
ANEC
Joanna Goodey
Fundamental Rights Agency
Sami Haddadin
Munich School of Robotics and MI
Gry Hasselbalch
The thinkdotank DataEthics & Copenhagen University
Fredrik Heintz
Linköping University
Fanny Hidvegi
Access Now
Eric Hilgendorf
University of Würzburg
Klaus Höckner
Hilfsgemeinschaft der Blinden und Sehschwachen
Mari-Noëlle Jégo-Laveissière
Orange
Leo Kärkkäinen
Nokia Bell Labs
Sabine Theresia Köszegi
TU Wien
Robert Kroplewski
Solicitor & Advisor to Polish Government
Elisabeth Ling
RELX
Pierre Lucas
Orgalim – Europe's technology industries
Ieva Martinkenaite
Telenor
Thomas Metzinger – Co-Rapporteur
JGU Mainz & European University Association
Catelijne Muller
ALLAI Netherlands & EESC
Markus Noga
SAP
Barry O'Sullivan, Vice-Chair of the AI HLEG
University College Cork
Ursula Pachl
BEUC
Nicolas Petit – Co-Rapporteur
University of Liège
Christoph Peylo
Bosch
Iris Plöger
BDI
Stefano Quintarelli
Garden Ventures
Andrea Renda
College of Europe Faculty & CEPS
Francesca Rossi
IBM
Cristina San José
European Banking Federation
George Sharkov
Digital SME Alliance
Philipp Slusallek
German Research Centre for AI (DFKI)
Françoise Soulié Fogelman
AI Consultant
Saskia Steinacker – Co-Rapporteur
Bayer
Jaan Tallinn
Ambient Sound Investment
Thierry Tingaud
STMicroelectronics
Jakob Uszkoreit
Google
Aimee Van Wynsberghe – Co-Rapporteur
TU Delft
Thiébaut Weber
ETUC
Cecile Wendling
AXA
Karen Yeung – Co-Rapporteur
The University of Birmingham

ETHICS AS AN ESCAPE FROM REGULATION?

WASHING ' TO ETHICS SHOPPING?[2]

BEN WAGNER

A strange confusion among technology policy makers can be witnessed at present. While almost all are able to agree on the common chorus of voices chanting 'something must be done,' it is very difficult to identify what exactly must be done and how. In this confused environment it is perhaps unsurprising that the idea of 'ethics' is presented as a concrete policy option. Striving for ethics and ethical decision-making, it is argued, will make technologies better. While this may be true in many cases, much of the debate about ethics seems to provide an easy alternative to government regulation. Unable or unwilling to properly provide regulatory solutions, ethics is seen as the 'easy' or 'soft' option which can help structure and give meaning to existing self-regulatory initiatives. In this world, 'ethics' is the new 'industry self-regulation.'

Rigorous ethical approaches?

This approach does not do justice to many of the proponents of ethical approaches to technology who think long and hard about ethical frameworks for technology development. It is however indicative of the increasingly common role of technology ethics in political debates. For example, as part of a conference panel on ethics, one member of the Google DeepMind ethics team emphasised repeatedly how ethically Google DeepMind was acting, while simultaneously avoiding any responsibility for the data protection scandal at Google DeepMind (Powles and Hodson 2018). In her understanding, Google DeepMind were an ethical company developing ethical products and the fact that the health data of 1.6 Million people was shared without a legal basis was instead the fault of the British government. This suggests a tension between legal and ethical action, in which the appropriate mode of governance is not yet sufficiently defined.

Ethics / rights / regulation

Such narratives are not just uncommon in the corporate but also in technology policy, where ethics, human rights and regulation are frequently played off against each other. In this context, ethical frameworks that provide a way to go beyond existing legal frameworks can also provide an opportunity to ignore them. More broadly the rise of the ethical technology debate runs in parallel to the increasing resistance to any regulation at all. At an international level the Internet Governance Forum (IGF) provides a space for discussions about governance without any mechanism to implement them and successive attempts to change this have failed. It is thus perhaps unsurprising that many of the initiatives proposed on regulating technologies tend to side-line the role of the state and instead emphasize the role of the private sector. Whether through the multi-stakeholder model proposed by Microsoft for an international attribution agency in which states play a comparatively minor role (Charney et al. 2016), or in a proposal by RAND corporation which suggests that states should be completely excluded from such an attribution organisation (Davis II et al. 2017). In fact, states and their regulatory instruments are increasingly portrayed as a problem rather than a solution.

Case in point: Artificial Intelligence

This tension between ethics, regulation and governance is evident in the debate on

There are **hundreds** of documents about ethical guidelines!

nature machine intelligence

PERSPECTIVE

https://doi.org/10.1038/s42256-019-0088-2

The global landscape of AI ethics guidelines

Anna Jobin, Marcello Ienca and Effy Vayena*

In the past five years, private companies, research institutions and public sector organizations have issued principles and guidelines for ethical artificial intelligence (AI). However, despite an apparent agreement that AI should be 'ethical', there is debate about both what constitutes 'ethical AI' and which ethical requirements, technical standards and best practices are needed for its realization. To investigate whether a global agreement on these questions is emerging, we mapped and analysed the current corpus of principles and guidelines on ethical AI. Our results reveal a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy), with substantive divergence in relation to how these principles are interpreted, why they are deemed important, what issue, domain or actors they pertain to, and how they should be implemented. Our findings highlight the importance of integrating guideline-development efforts with substantive ethical analysis and adequate implementation strategies.

Artificial intelligence (AI), or the theory and development of computer systems able to perform tasks normally requiring human intelligence, is widely heralded as an ongoing "revolution" transforming science and society altogether[1,2]. While approaches to AI such as machine learning, deep learning and artificial neural networks are reshaping data processing and analysis[3], autonomous and semi-autonomous systems are being increasingly used in a variety of sectors including healthcare, transportation and the production chain[4]. In light of its powerful transformative force and profound impact across various societal domains, AI has sparked ample debate about the principles and values that should guide its development and use[5,6]. Fears that AI might jeopardize jobs for human workers[7], be misused by malevolent actors[8], elude accountability or inadvertently disseminate bias and thereby undermine fairness[9] have been at the forefront of the recent scientific literature and media coverage. Several studies have discussed the topic of ethical AI[10–13], notably in meta-assessments[14–16] or in relation to systemic risks[17,18] and unintended negative consequences such as algorithmic bias or discrimination[19–21].

National and international organizations have responded to these concerns by developing ad hoc expert committees on AI, often mandated to draft policy documents. These committees include the High-Level Expert Group on Artificial Intelligence appointed by the European Commission, the expert group on AI in Society of the Organisation for Economic Co-operation and Development (OECD), the Advisory Council on the Ethical Use of Artificial Intelligence and Data in Singapore, and the Select Committee on Artificial Intelligence of the UK House of Lords. As part of their institutional appointments, these committees have produced or are reportedly producing reports and guidance documents on AI. Similar efforts are taking place in the private sector, especially among corporations who rely on AI for their business. In 2018 alone, companies such as Google and SAP publicly released AI guidelines and principles. Declarations and recommendations have also been issued by professional associations and non-profit organizations such as the Association of Computing Machinery (ACM), Access Now and Amnesty International. This proliferation of soft-law efforts can be interpreted as a governance response to advanced research into AI, whose research output and market size have drastically increased[22] in recent years.

Reports and guidance documents for ethical AI are instances of what is termed non-legislative policy instruments or soft law[23]. Unlike so-called hard law—that is, legally binding regulations passed by the legislatures to define permitted or prohibited conduct—ethics guidelines are not legally binding but persuasive in nature. Such documents are aimed at assisting with—and have been observed to have significant practical influence on—decision-making in certain fields, comparable to that of legislative norms[24]. Indeed, the intense efforts of such a diverse set of stakeholders in issuing AI principles and policies is noteworthy, because they demonstrate not only the need for ethical guidance, but also the strong interest of these stakeholders to shape the ethics of AI in ways that meet their respective priorities[16,25]. Specifically, the private sector's involvement in the AI ethics arena has been called into question for potentially using such high-level soft policy as a portmanteau to either render a social problem technical[16] or to eschew regulation altogether[26]. Beyond the composition of the groups that have produced ethical guidance on AI, the content of this guidance itself is of interest. Are these various groups converging on what ethical AI should be, and the ethical principles that will determine the development of AI? If they diverge, what are their differences and can these differences be reconciled?

Our Perspective maps the global landscape of existing ethics guidelines for AI and analyses whether a global convergence is emerging regarding both the principles for ethical AI and the suggestions regarding its realization. This analysis will inform scientists, research institutions, funding agencies, governmental and intergovernmental organizations, and other relevant stakeholders involved in the advancement of ethically responsible innovation in AI.

Methods

We conducted a scoping review of the existing corpus of documents containing soft-law or non-legal norms issued by organizations. This included a search for grey literature containing principles and guidelines for ethical AI, with academic and legal sources excluded. A scoping review is a method aimed at synthesizing and mapping the existing literature[27] that is considered particularly suitable for complex or heterogeneous areas of research[27,28]. Given the absence of a unified database for AI-specific ethics guidelines, we developed a protocol for discovery and eligibility, adapted from the Preferred

Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland. *e-mail: effy.vayena@hest.ethz.ch

Data and Ethics

What does ethics have to do with data?

The combination of data analytics, a data-saturated and poorly regulated commercial environment, and the absence of widespread, well-designed standards for data practice in industry, university, non-profit, and government sectors has created a '**perfect storm**' **of ethical risks.**

Thus **no single set of ethical rules or guidelines will fit all data circumstances**; ethical insights in data practice must be adapted to the **needs of many kinds of data practitioners operating in different contexts**.

What does ethics have to do with data?

We can define a **harm** or a **benefit** as 'ethically significant' when it has a substantial possibility of making a difference to certain individuals' chances of having a good life, or the chances of a group to live well: that is, to flourish in society together.

Some harms and benefits are not ethically significant. Say I prefer Coke to Pepsi. If I ask for a Coke and you hand me a Pepsi, even if I am disappointed, you haven't impacted my life in any ethically significant way.

In the context of data practice, the potential harms and benefits are real and ethically significant. But **due to the more complex, abstract, and often widely distributed nature of data practices, as well as the interplay of technical, social, and individual forces in data contexts, the harms and benefits of data can be harder to see and anticipate**.

In this respect, then, data has a broader ethical sweep than engineering of bridges and airplanes. Data practitioners must confront a far more complex ethical landscape than many other kinds of technical professionals...

Ethical Benefits of Data Practices

HUMAN UNDERSTANDING:

Because data and its associated practices can uncover previously unrecognized correlations and patterns in the world, **data can greatly enrich our understanding of ethically significant relationships — in nature, society, and our personal lives**.

Ethical Benefits of Data Practices

SOCIAL, INSTITUTIONAL, AND ECONOMIC EFFICIENCY:

Once we have a more accurate picture of how the world works, **we can design or intervene in its systems to improve their functioning**. This reduces wasted effort and resources and improves the alignment between a social system or institution's policies/processes and our goals.

Ethical Benefits of Data Practices

PREDICTIVE ACCURACY AND PERSONALIZATION:

Not only can good data practices help to make social systems work more efficiently, but they can also used to more precisely **tailor actions to be effective in achieving good outcomes for specific individuals, groups, and circumstances**, and to be more responsive to user input in (approximately) real time.

Ethical Harms of Data Practices

HARMS TO PRIVACY & SECURITY:

Thanks to the ocean of personal data that humans are generating today (or, to use a better metaphor, the many different **lakes, springs, and rivers of personal data** that are pooling and flowing across the digital landscape), most of us do not realize **how exposed our lives are**, or can be, by common data practices.

Ethical Harms of Data Practices

HARMS TO FAIRNESS AND JUSTICE:

We all have a **significant interest in being judged and treated fairly**, whether it involves how we are treated by law enforcement and the criminal and civil court systems, how we are evaluated by our employers and teachers, the quality of health care and other services we receive, or how financial institutions and insurers treat us.

Ethical Harms of Data Practices

HARMS TO TRANSPARENCY AND AUTONOMY:

In this context, *transparency* is the **ability to see how a given social system or institution works**, and to be able to inquire about the basis of life-affecting decisions made within that system or institution.

So, for example, if your bank denies your application for a home loan, transparency will be served by you having access to information about exactly *why* you were denied the loan, and by whom.

Europe's GDPR

Europe's GDPR

The GDPR can be summarised in the following points:

1. It concerns "**Personal Data**": Name, address, localisation, online identifier, health information, income, cultural profile, ...
2. Communication: Who gets the data, why, for how long? (No use for other 'incompatible' purposes. Use as long as necessary.)
3. Consent: Get clear informed consent.
4. **Access: Provide access to my data.**
5. Right to be forgotten (not for research).
6. **Right to explanation for contracts (& right to have a person decide).**
7. Marketing: Right to opt out.
8. Legal: Maintain EU legislation when transferring data out.
9. Need for a "data protection officer" in your organisation.
10. **Impact assessment prior to high-risk processing (new technology, personal information, surveillance, sensitive).**

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

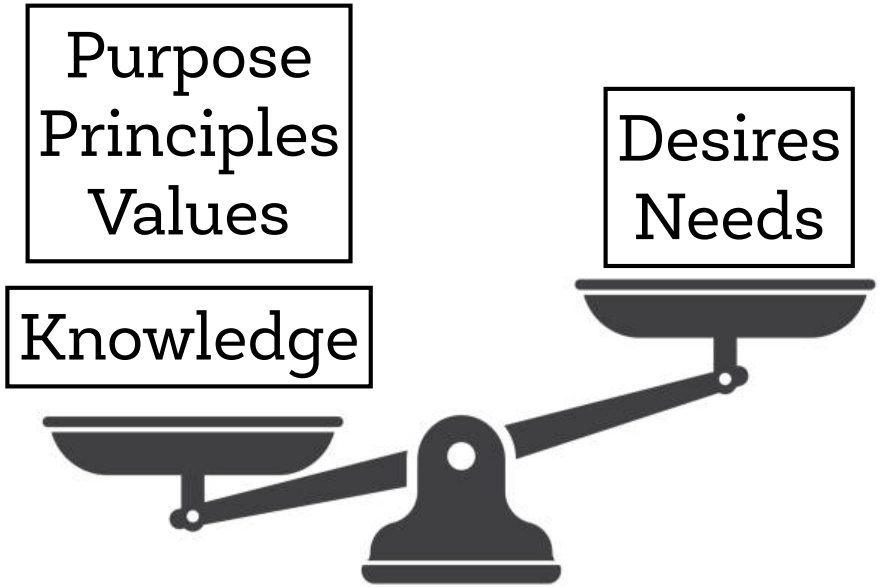
Values

Science

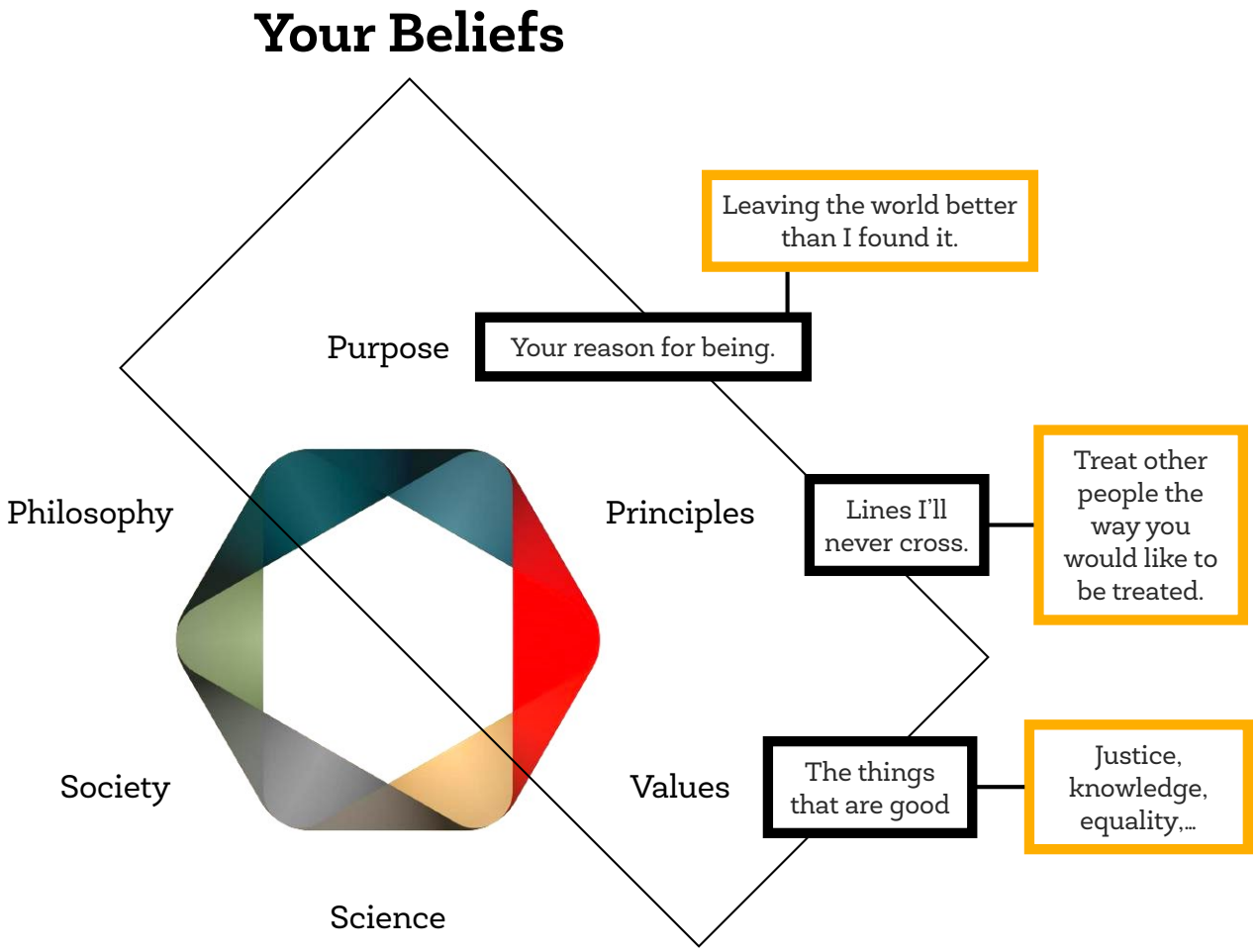
Ethics seeks to answer questions like "*what is good or bad*", "*what is right or what is wrong*", or "*what is justice, well-being or equality*".

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

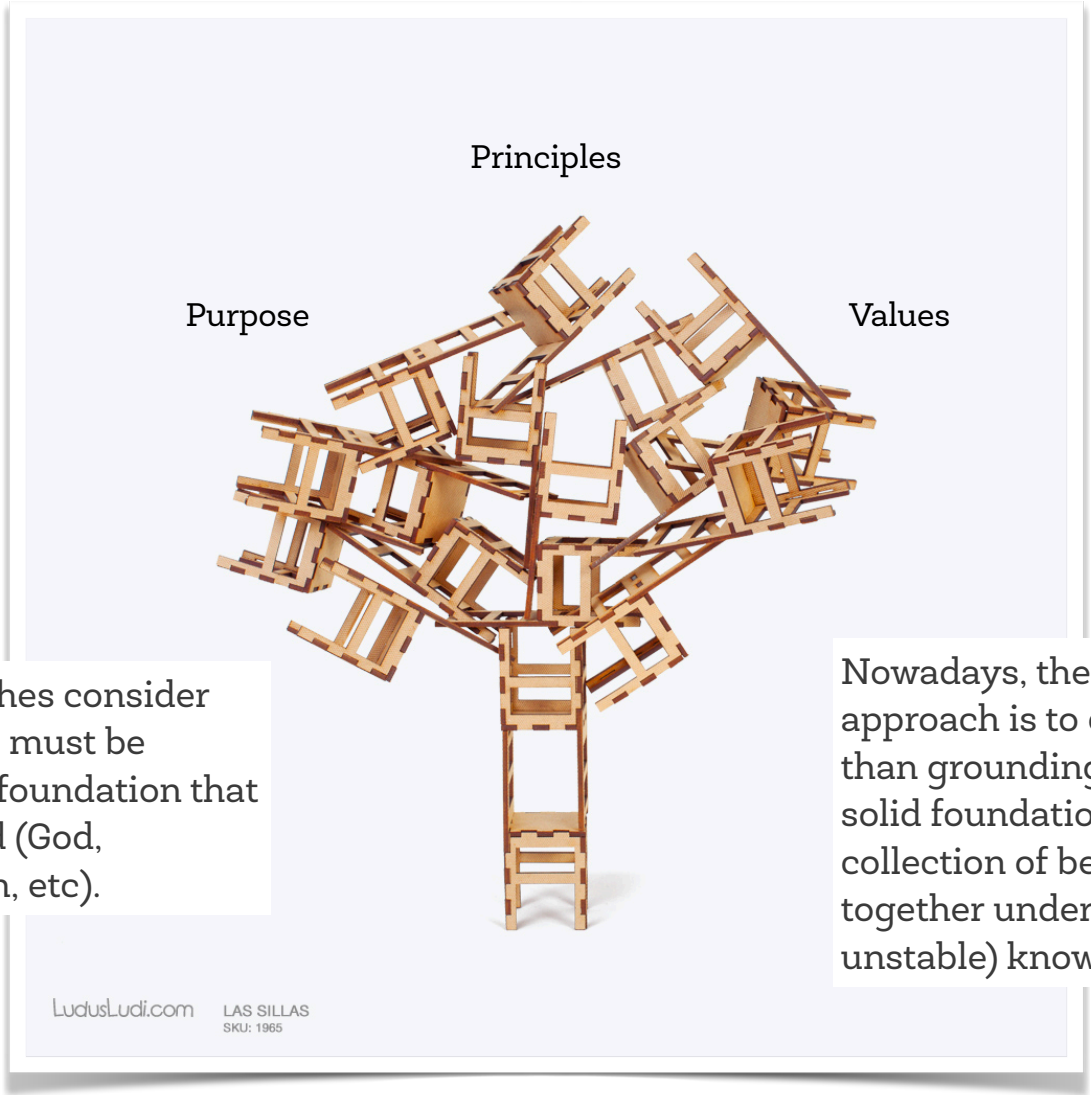
How do we make decisions?



Beliefs, the necessary ingredients of a good individual decision.



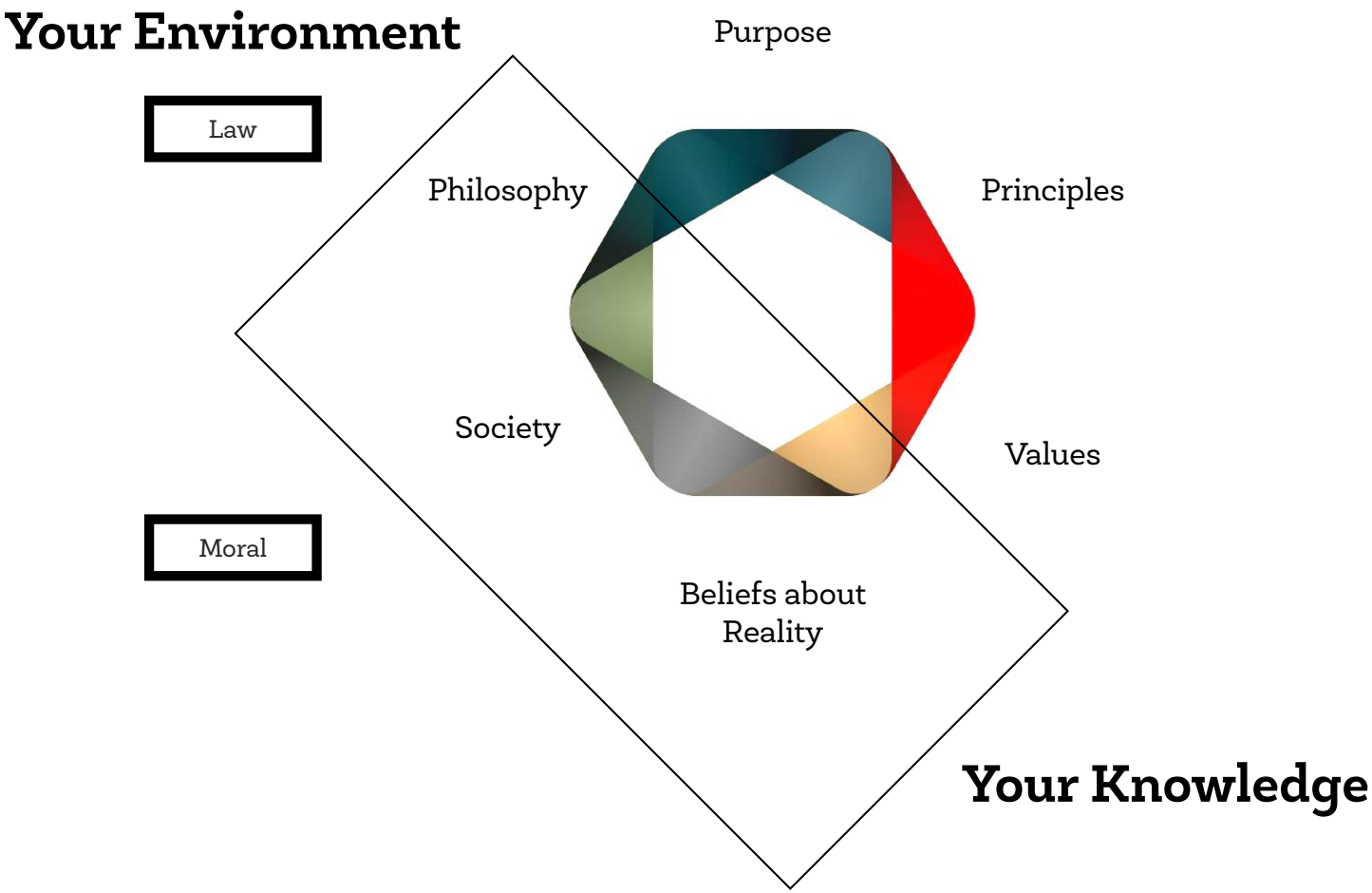
Beliefs, the necessary ingredients of a good individual decision.



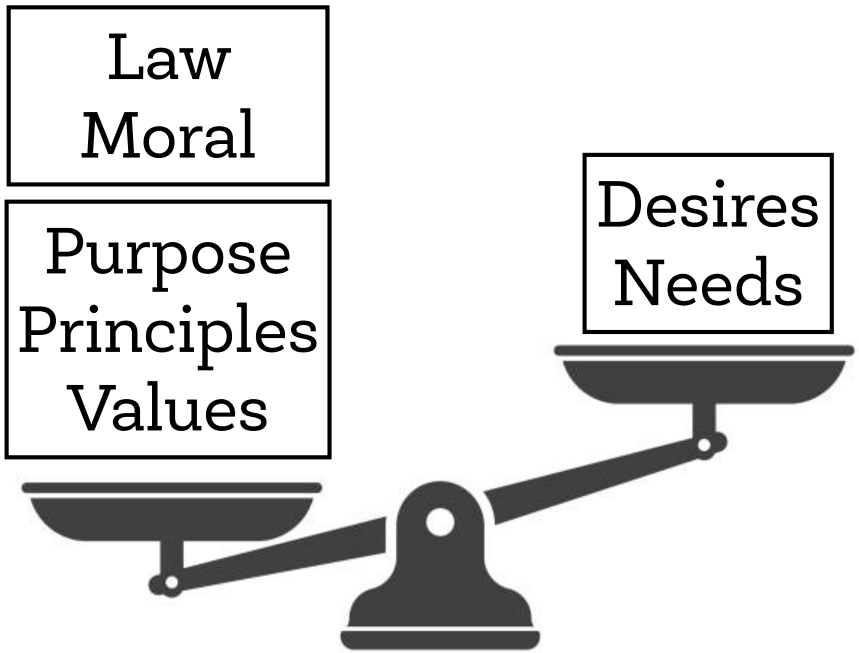
Traditional approaches consider that person's beliefs must be grounded in a solid foundation that cannot be discussed (God, rationality and harm, etc).

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 injust, 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?

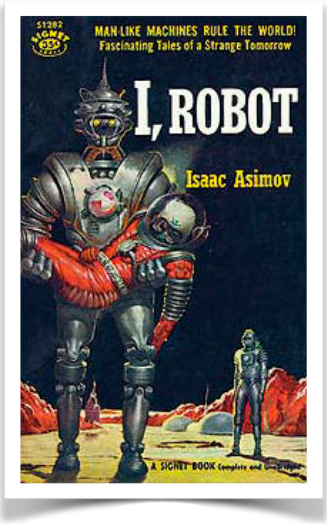
The answer is based on our values, principles and purposes rather than social conventions.

An ethical decision is based on conscious, rational reflection.

Traditional Normative Ethics

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

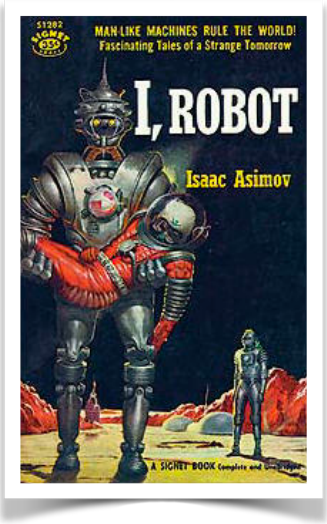
- **Utilitarianism** (J.Bentham): Does an action maximize happiness and well-being for all affected individuals? (**consequences**)
- **Deontology** (I.Kant): Does an action follow a moral rule (e.g. the Golden Rule: 'Treat others how you want to be treated')? An action should be based on whether that 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, responsability, care, etc.**)



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

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.

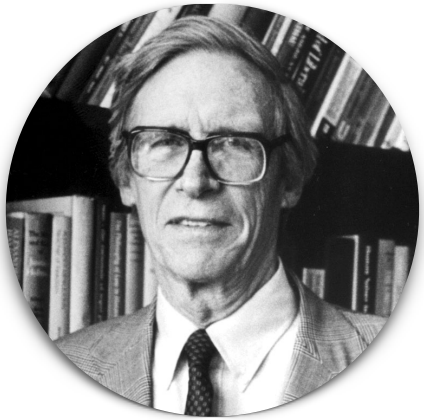
Traditional Ethics

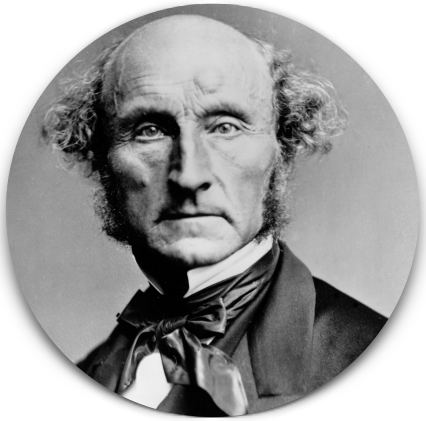
Suppose it is obvious that someone in need should be helped.

- 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



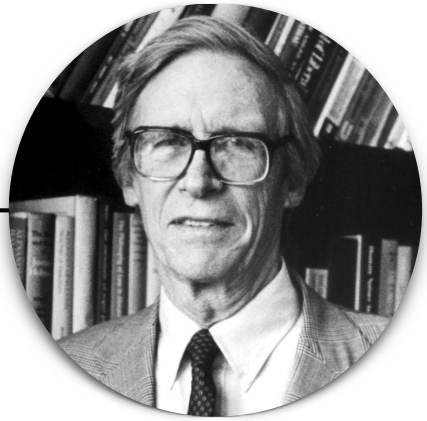


4 theories about what is right and what is wrong in society

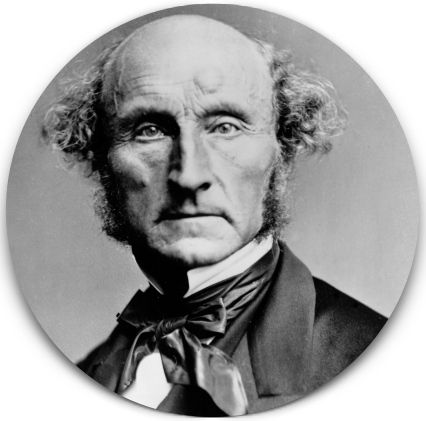
(Political) Philosophy

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



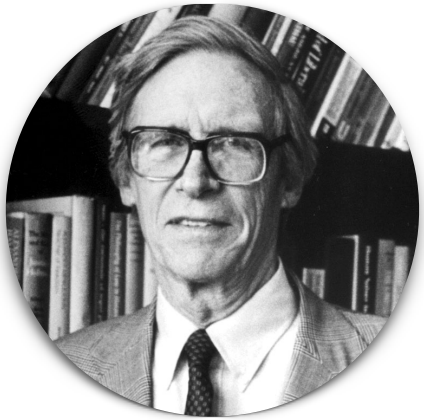
(Political) Philosophy

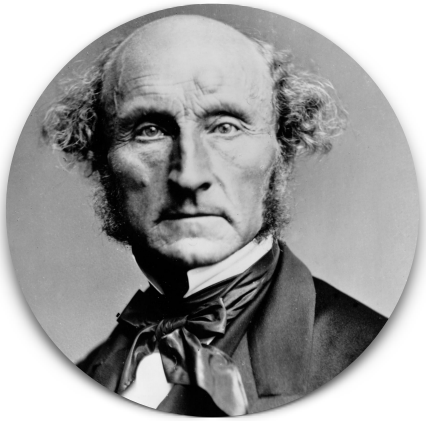
Libertarians

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**.

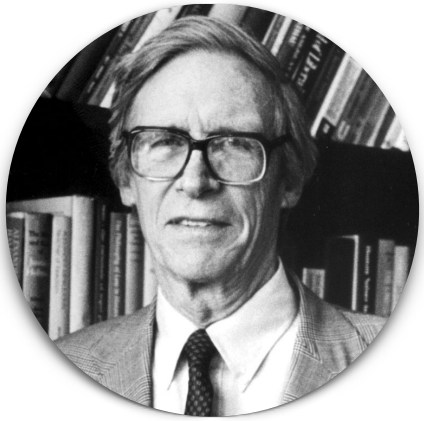
For libertarianism, the only proper limit to one person's enjoyment of these rights is his or her duty to respect the similar rights of others.

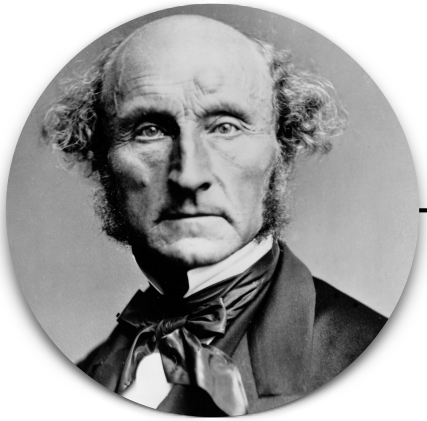




John Locke

(Political) Philosophy





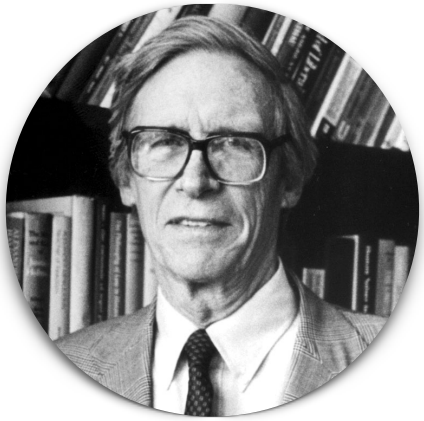
John Stuart Mill

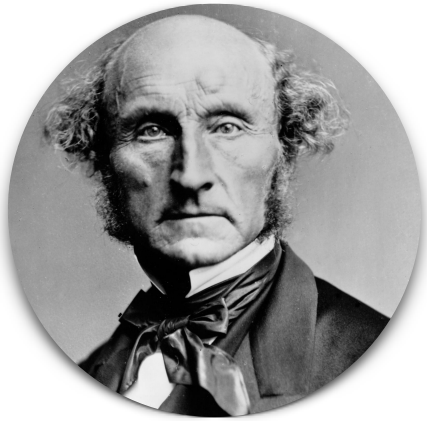
Utilitarians

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

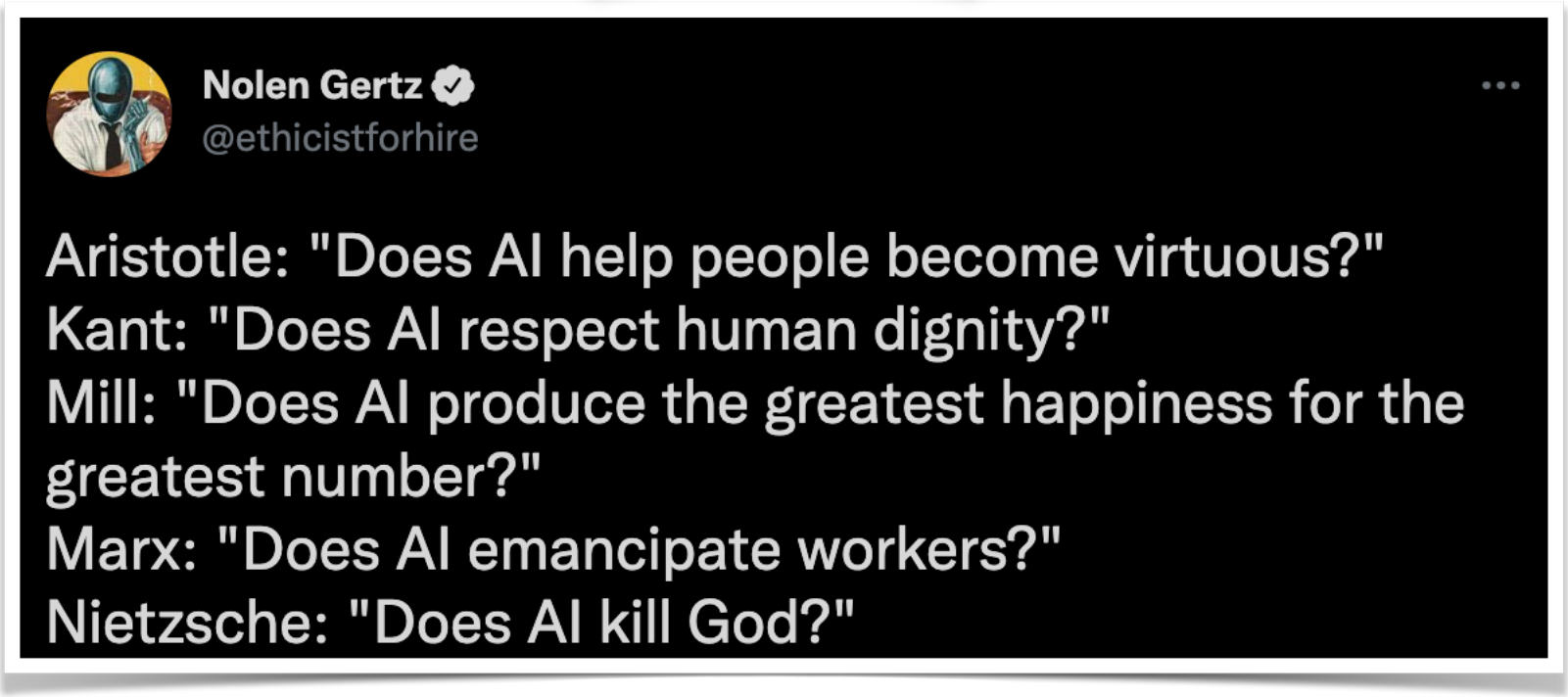
MS TECH | UNSPLASH

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.

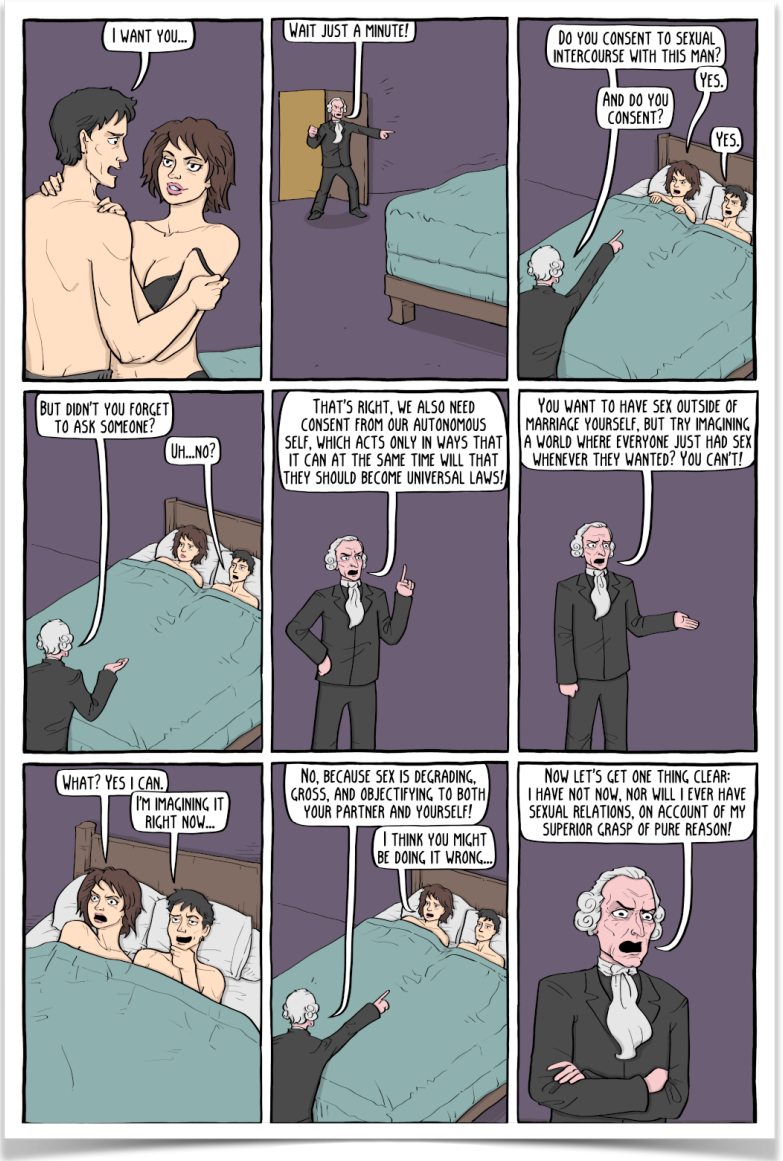
Canonical views of AI ethics?

Value diversity



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

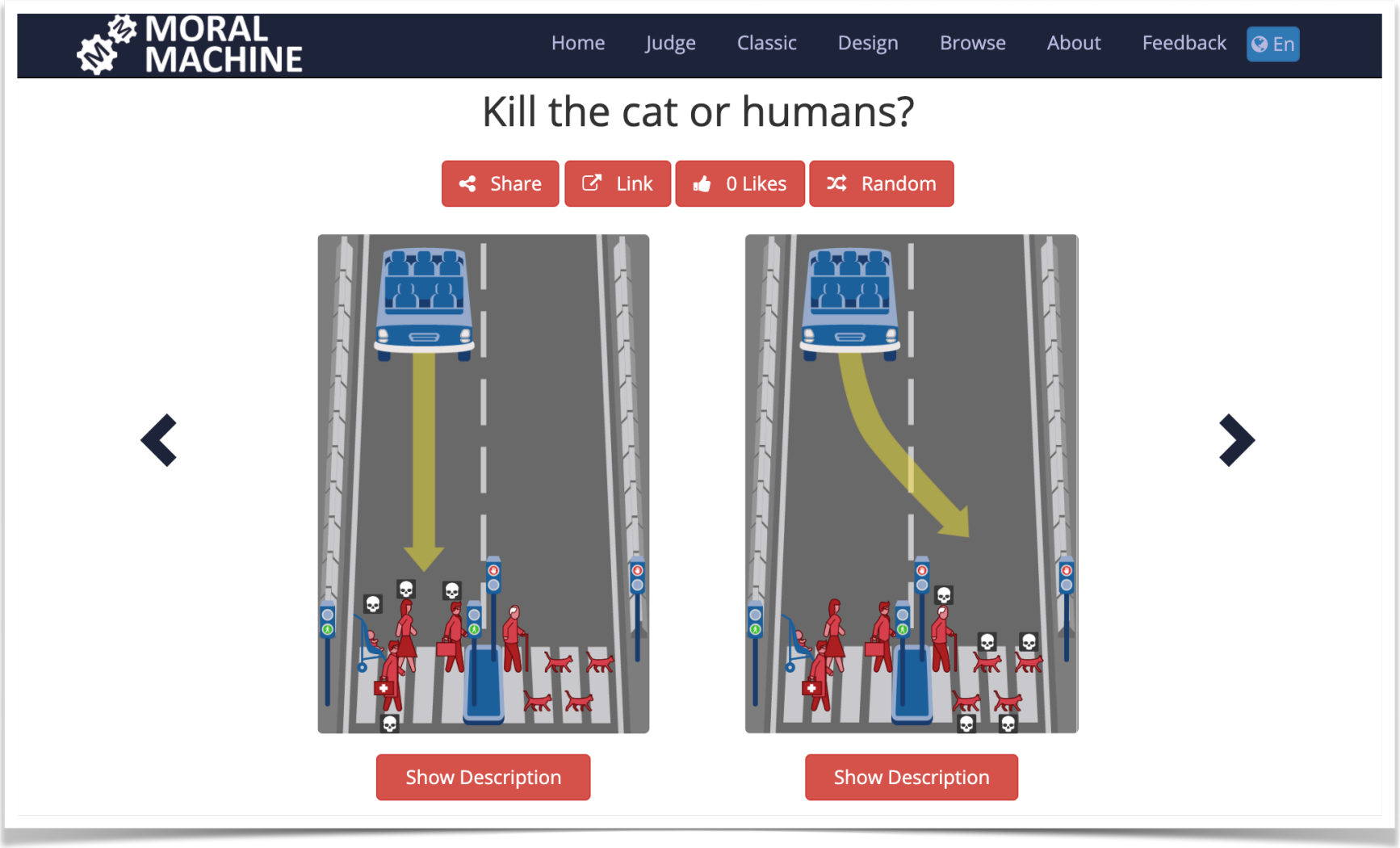
The positive approach to ethics describes the world as it is. It is about how humans judge situations and decisions in different scenarios.

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

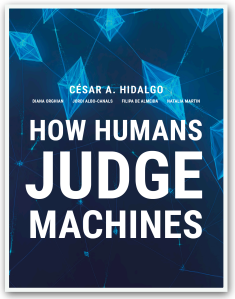
In recent decades, psychologists have discovered **five moral dimensions** that humans consider when judging situations:

- **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

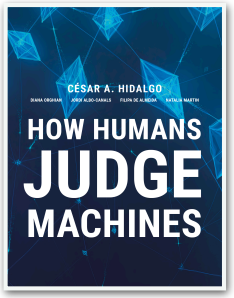
Alice and Bob, two colleagues in a software company, are competing for the same promotion at work. Alice has a severe peanut allergy. Knowing this, Bob sneaks into the office kitchen and mixes a large spoonful of peanut butter into Alice's soup. At lunchtime, Alice accidentally drops her soup on the 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 eqivalent to judging humans.

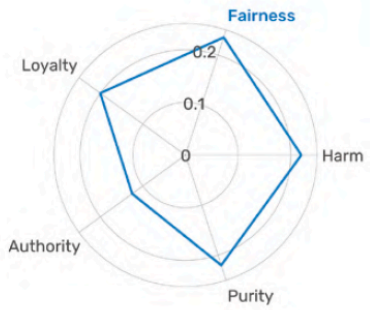


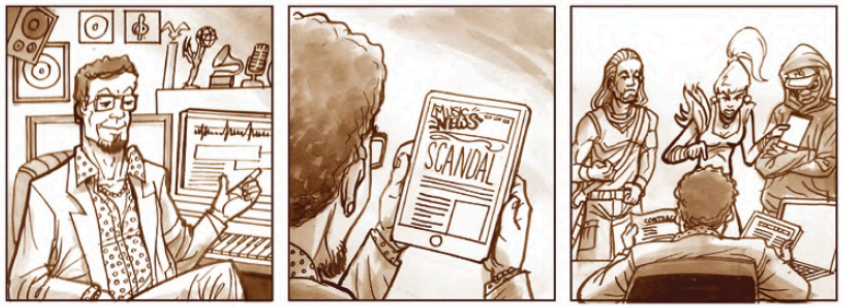
Humans are judged more positively than machines in autonomous driving scenarios.

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**.

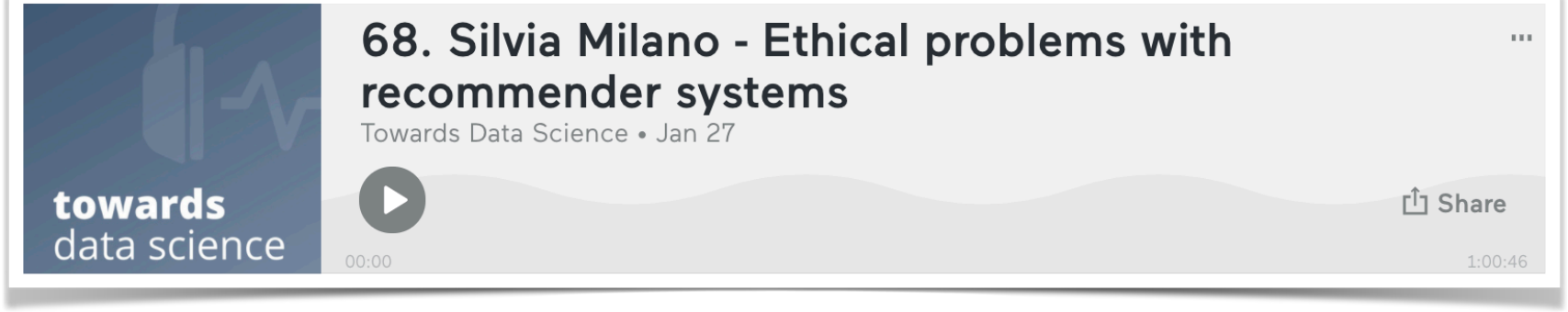




S8 A record label hires a(n) [songwriter/AI songwriter] to write lyrics for famous musicians. The [songwriter/AI songwriter] has written lyrics for dozens of songs in the past year. However, a journalist later discovers that the [songwriter/AI songwriter] has been plagiarizing lyrics from lesser-known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

Additional Resources

Listen to this podcast!



Listen to this podcast!