

THE ETHICS OF LAW AND THE LAWS OF ETHICS

Mireille Hildebrandt



DigitalSingleMarket

@DSMeu

Following



What are the 5 ethical principles for
#ArtificialIntelligence?

- 1 Do good
- 2 Do no harm
- 3 Keep human agency
- 4 Be fair
- 5 Operate transparently

! Contribute to the Draft Ethics guidelines
for trustworthy **#AI** until 18 January
ec.europa.eu/digital-single...



DRAFT ETHICS GUIDELINES FOR TRUSTWORTHY AI

Working Document for stakeholders' consultation

Brussels, 18 December 2018

Framework for Trustworthy AI

CHAPTER I

Ethical Purpose

Ensure respect of fundamental rights, principles and values when developing, deploying and using AI



CHAPTER II

Realisation of Trustworthy AI

Ensure implementation of ethical purpose as well as technical robustness when developing, deploying and using AI

Requirements for Trustworthy AI

To be continuously evaluated, addressed and assessed in the Design & Use phase of AI through



Technical Methods

Non-Technical Methods

CHAPTER III

Assessment List for Trustworthy AI based on Use Cases

These Guidelines therefore set out a **framework for Trustworthy AI**:

- **Chapter I** deals with **ensuring AI's ethical purpose**, by setting out the fundamental rights, principles and values that it should comply with.
- From those principles, **Chapter II** derives **guidance on the realisation** of Trustworthy AI, tackling both ethical purpose and technical robustness. This is done by listing the requirements for Trustworthy AI and offering an overview of technical and non-technical methods that can be used for its implementation.
- **Chapter III** subsequently **operationalises** the requirements by providing a concrete but non-exhaustive assessment list for Trustworthy AI. This list is then adapted to specific use cases.

These Guidelines therefore set out a **framework for Trustworthy AI**:

- **Chapter I** deals with **ensuring AI's ethical purpose, by setting out the fundamental rights**, principles and values that it should comply with.
- From those principles, **Chapter II** derives **guidance on the realisation** of Trustworthy AI, tackling both ethical purpose and technical robustness. This is done by listing the requirements for Trustworthy AI and offering an overview of technical and non-technical methods that can be used for its implementation.
- **Chapter III** subsequently **operationalises** the requirements by providing a concrete but non-exhaustive assessment list for Trustworthy AI. This list is then adapted to specific use cases.

- “In the context of AI, respect for human dignity entails that all people are treated with respect due to them as individuals, rather than **merely as data subjects**.”

OOPS

- “Given that, on the whole, AI’s benefits outweigh its risks, we must ensure to follow the road that **maximises the benefits of AI while minimising its risks**”
- “In the final version of these Guidelines, a mechanism will be put forward to allow stakeholders to voluntarily endorse them.”

- “(...) compliance with fundamental rights, principles and values entails that these are duly operationalised by implementing them **throughout the AI technology’s design, development, and deployment**. Such implementation can be addressed both by technical and non-technical methods.”

Medium Education

MEMBER FEATURE STORY

Don't Teach Your Kid to Code. Teach Them to Communicate.

The jobs of the future don't exist yet — but we know they'll require some serious social skills

 Greg Satell [Follow](#)
Jan 7 · 5 min read ★



Photo: FlamingoImages/iStock/Getty Images Plus

“The jobs of the future will not depend as much on knowing facts or crunching numbers as **on humans collaborating with other humans to design work for machines.** Collaboration will increasingly become a competitive advantage.”

What's next?

- 1. AI and the new 'choice architecture'**
- 2. The laws of ethics**
- 3. The ethics of law**
- 4. Legal protection by design**

What's next?

- 1. AI and the new 'choice architecture'**
- 2. The laws of ethics**
- 3. The ethics of law**
- 4. Legal protection by design**

AI and the new 'CA'

1. Code-driven (symbol manipulation & various types of logic) IFTTT - deductive
 2. Data-driven (machine learning, deep learning) – inductive
 3. Hybrids are the new black
- Smart contracts and regulation (self-executing code, private enforcement)
 - Predictive targeting (pricing, insurance, policing, legal tech, fintech)

AI and the new 'CA'

- Human agency combines:
 - Deduction (bias, presumptions, prior knowledge, innate, Vorurteil)
 - Induction (experience, feedback, testing, surprise, uncertainty)
 - Abduction (inferring new patterns, bias, assumptions, knowledge)
- Abductive reasoning = speculative reasoning
 - Natality
 - Lateral thinking
 - Play rather than game

AI and the new 'CA'

A combination of deduction and induction is not abduction:

- ML and DL do not move beyond inductive inferences
- This entails a number of 'cognitive biases', such as
 - availability bias (low hanging fruit), clustering bias (spurious correlations), confirmation bias (overfitting), recency bias (low hanging fruit), survival bias (data is not reality), etc.

AI and the new 'CA'

- There is no ability to detect such bias, because systems can only:
 - 'experience' the data it is 'fed' to them and
 - make inferences based on a constructed hypothesis space
 - no own experience, nothing is at stake, only if programmed

AI and the new 'CA'

- Here we have mindless data- and code-driven agency:
 - deterministic or
 - emergent behaviours
- This mindless agency co-determines our **choice architecture (CA)**
 - What choices are enabled or precluded?
 - What **types** of choices are enabled or precluded?
 - For whom?
 - Depending on what?

What's next?

1. AI and the new 'choice architecture'
2. **The laws of ethics**
3. The ethics of law
4. Legal protection by design

The Laws of Ethics

1. Utilitarian ethics (rule, act; methodological individualism)
 - Moral machines (MIT platform), ethics as a preference
2. Deontological ethics (autonomy; categorical imperative)
 - Respecting the dignity of AGI? Coding the CI?
3. Virtue ethics (perfectionism and the golden mean)
 - Practical wisdom of machines?



An abstract graphic with a solid orange background. It features several large, irregular orange circles of varying sizes. At the bottom, there is a horizontal band with a pattern of small, overlapping circles in shades of orange, light blue, and white. The text "BEING PROFILED:COGITAS ERGO SUM" is printed in a bold, black, sans-serif font across the lower portion of the orange area.

BEING PROFILED:COGITAS ERGO SUM

The Laws of Ethics

Taking the *cogitas* seriously (no *cogito* without a *cogitas*):

- **Relational** ethics based on a relational self
- **Ecological** ethics based on an ecological understanding of agency

The Laws of Ethics

Enhancing and protecting human agency:

- capability approach of Sen [agent oriented, but environment dependent]
 - Interpersonal environment
 - Institutional environment
 - Technological environment
- affordances approach of Gibson [environment oriented, but agent dependent]
 - What human agency does an oral society afford?
 - What human agency does a society of the script or the printing press afford?
 - What human agency does a data- or code-driven environment afford?

The Laws of Ethics

1. Utilitarian ethics:
 - What choice architecture (CA) enhances the ability to **form** preferences?
 - What CA the ability to **exercise** preferences?
2. Deontological ethics:
 - What CA respects human agency **as** autonomous?
 - Even when using others **as** instruments?
3. Virtue ethics:
 - What CA affords institutions and individuals
 - Practical **wisdom**?
 - Virtuous **action**?

The Laws of Ethics

CAs of/for:

- **Users (individual persons)**
- **Users (business or government agencies)**
- **Retailers (selling software and/or hardware, IoT)**
- **Those who put AI applications on the market**
- **Those who develop AI applications**

The Laws of Ethics

- **CAs redistribute risks**
- **CAs redistribute types of choices**
- **CAs redistribute the capability to exercise choices**
- **CAs reduce or enhance the affordances of an environment**

What's next?

1. AI and the new 'choice architecture'
2. The laws of ethics
3. The ethics of law
4. Legal protection by design

The ethics of law

- Governments should treat:
 - each individual person
 - with equal respect and concern.
- This underpins both:
 - **Democracy** ('1 person 1 vote')
 - **The Rule of Law** (checks and balances, human rights)

The ethics of law

- Governments should be transparent by default
- Citizens should be opaque by default
- Governments are bound to act from the perspective of the **general interest**
 - Legality principle (constitutional, administrative, criminal law)
- Citizens are **free to act strategically**, unless they harm others
 - Freedom to conduct a business (freedom to contract, transfer property)

The ethics of law

- This has consequences for the CA that a law should present:
 - To big players
 - To individual citizens
- This has consequences for algorithmic governance:
 - Imposing a CA that makes behaviour 'legal by design', or
 - Developing and requiring a CA that provides 'legal protection by design'



Advertising and academia are controlling our thoughts. Didn't you know? | George Monbiot

[George Monbiot](#) Mon 31 Dec 2018 06.00 GMT

By abetting the ad industry, universities are leading us into temptation, when they should be enlightening us

The ethics of law

- Advertising companies try to develop a choice architecture:
 - to reduce the agency of consumers
- Supported by university departments of psychometrics that try to develop a scientific discipline to overcome e.g.
 - consumer scepticism,
 - resistance against fake news,
 - warning signs of addiction.

The ethics of law

- What CA should law develop to prevent or divert a manipulative CA?
- Let's take the GDPR:
 - Prohibition of forced consent for additional processing
 - Withdrawal of consent as easy as provision
 - Data protection impact assessment
 - Data protection by design
 - Qualified prohibition automated decisions

What's next?

1. AI and the new 'choice architecture'
2. The laws of ethics
3. The ethics of law
4. **Legal protection by design**

Legal protection by design

- LPbD is NOT same as LbD
 - translation of legal norms into technical constraints (techno-regulation)
- LPbD is rather connected with 'values by design' and 'value-sensitive design'
 - but, it is about 'legal'
 - about rights and obligations
 - not values and principles

Legal protection by design

1. The scope of LPbD should be determined by way of
 - *democratic* participation, for instance in the context of
 - participatory technology assessment and
 - involvement of the democratic legislature.
2. Those subject to such LPbD should be able
 - to *contest* its application *in a court of law*.

Legal protection by design

■ Let's take the GDPR:

- Prohibition of forced consent for additional processing
- Withdrawal of consent as easy as provision
- Data protection impact assessment
- Data protection by design
- *Qualified prohibition automated decisions*

