




# AI & Fundamental Rights

"AI Beyond the Hype" @ AEA -EAL and SCL 31.5.2019

A close-up profile of a woman with blonde hair, looking towards the left. In the background, a blurred rainbow is visible against a light-colored wall. The woman is wearing a dark green jacket over a blue t-shirt. Her right hand is visible in the lower left, with a ring on her finger.

# FR for AI?

# AI: challenge for FR?





**WHAT ARE WE TALKING ABOUT?**

- UN Universal Declaration of Human Rights Preamble:

"Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world..."

- **Rights as fictions**

See for example Women Rights Movement.





# ANIMAL RIGHTS



## UK:

- 1793: John Cornish was found not guilty of any crime when he pulled out a horse's tongue. The court ruled that Cornish could be prosecuted only if there was evidence of malice towards the horse's owner.
- 1821: Colonel Richard Martin proposed a statute to protect horses – laughter in Parliament
- 1822: Enactment of the Treatment of Horses and Cattle Act
- 1876: Cruelty to Animals Act
- 1911: Protection of Animals Act
- 2004: The Hunting Act (2017, according to a survey, 85% thought fox hunting should remain prohibited)

## India:

- 1960: Prevention of Cruelty to Animals Act 1960



A woman with dark hair, wearing a vibrant red shawl with a colorful floral and geometric pattern, is shown from the chest up. She is looking upwards and to her left with a thoughtful expression. The background is a plain, light-colored wall. A dark grey horizontal band across the middle of the image contains the text 'HUMAN RIGHTS' in large, white, bold, sans-serif capital letters.

# HUMAN RIGHTS

## Example Slavery:

- In the beginning of 1800: Slavery was permitted under international law
- 1807: UK abolishes the slave traffic throughout its colonies
- 1815: Conference of Vienna condemns slavery
- 1857: US Supreme Court ruled that Slavery was legal
- 1948: Universal Declaration of Human Rights: “no one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms”.





# ROBOT RIGHTS?!



## 4 ARGUMENTS:

- Ability to suffer
- Compassion
- Value to humans
- Combination of humans and AI





**ABILITY TO SUFFER**

Consciousness as a prerequisite  
for protection based on the  
ability to suffer.

Definition of consciousness:  
“the way things seem to us”  
= *qualia* (Daniel Dennett)



# Qualia:

1. Sensing stimuli
2. Perceiving sensation
3. Having a sense of self,  
namely a conception of its  
own existence in space and  
time

Article from 2016: “An Artificial Robot Nervous System to Teach Robots How to Feel Pain and Reflexively React to Potentially Damaging Contacts” by Johannes Kühn and Sami Haddadin

A close-up, low-angle shot of the droid R2-D2. The droid is positioned in the center-right of the frame, looking towards the left. Its head is tilted slightly upwards. The background is a workshop or hangar, with various mechanical parts and tools visible. A red tool is prominent in the lower-left foreground. The lighting is dramatic, with strong highlights on the droid's metallic surfaces and deep shadows in the background.

**COMPASSION**



Kant: “(...) duties to inanimate objects  
(...) these allude, indirectly, to our  
duties to (women) and men. The  
human impulse to destroy things that  
can still be used is very immoral... thus  
all duties relating to animals, other  
beings and things have an indirect  
reference to our duties towards  
(woman) mankind.”

Evan Dashevsky (robot ethicist from MIT's Media Lab): “Would you rather live in, say, a Westworld universe filled with humans who feel free to rape and maim the park's mechanical inhabitants, or on the deck of Star Trek: The Next Generation, where advanced robots are treated as equals?”

Role of Physicality: toy mechanical dinosaurs “Pleo”

Uncanny Valley (Masahiro Mori):  
description of the slow rise, sharp drop  
and then relatively fast rise in feelings  
of familiarity , as robots become more  
like humans.





**VALUE TO HUMANITY**



**COMBINATION OF HUMANS AND AI**

Elon Musk:

Humans must merge with AI or  
become irrelevant in the AI age

→ NEURALINK.



# ETHICS GUIDELINES FOR TRUSTWORTHY AI



Europese Commissie  
Commission européenne

*Glyn Lowe Photoworks*





# 3 COMPONENTS

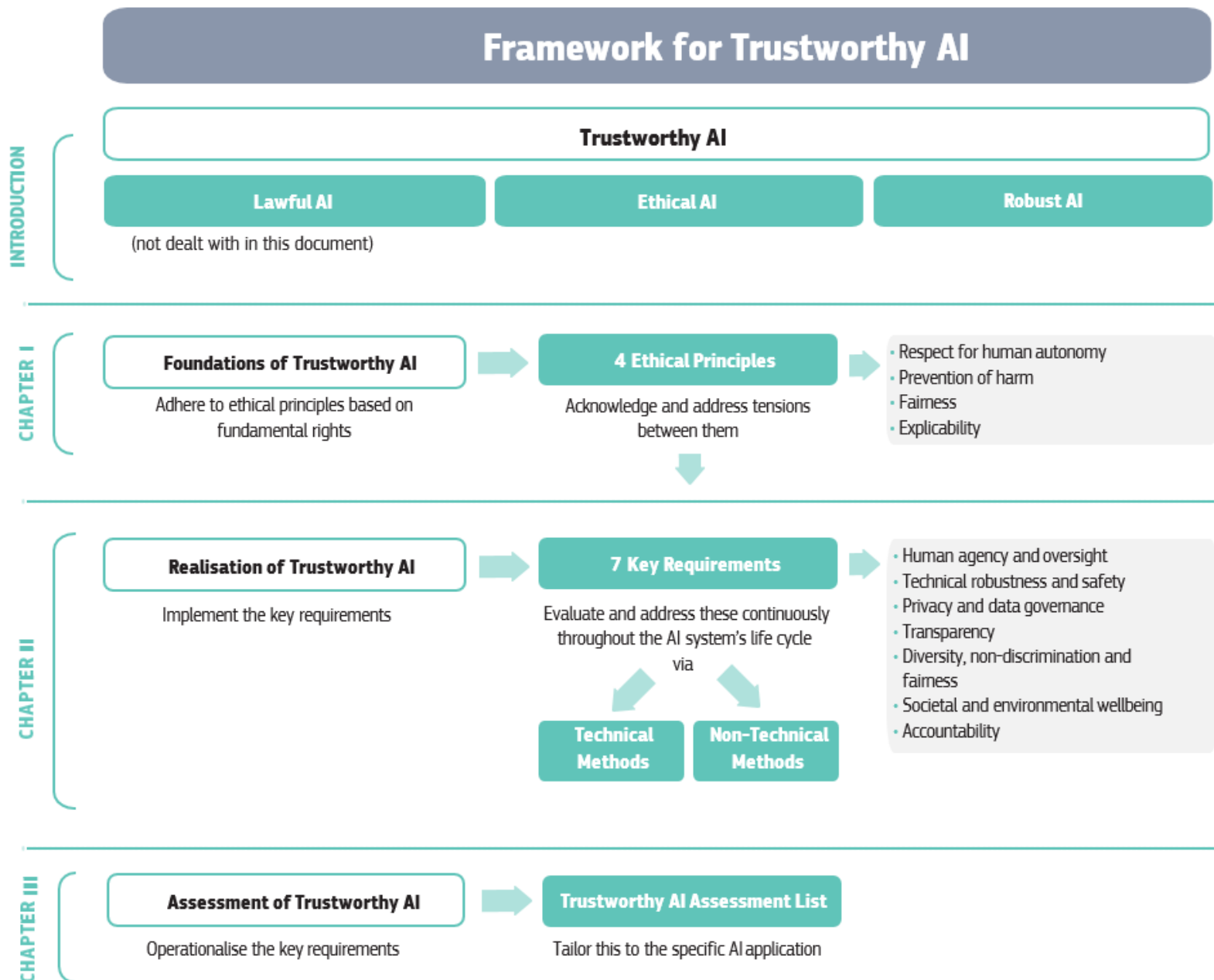


Figure 1: The Guidelines as a framework for Trustworthy AI



A person with long brown hair is shown from the chest up, wearing a white t-shirt and a blue denim jacket. The t-shirt has the text "RIGHT TO BE" and "WITONG" printed on it. The person is holding the edges of the denim jacket. The background shows a scenic view of a lake and hills under a clear blue sky.

# **FROM FUNDAMENTAL RIGHTS TO ETHICAL PRINCIPLES**

be free

go your own way

## **Fundamental Rights as a basis for Trustworthy AI**

- Respect for human dignity
- Freedom of the individual
- Respect for democracy, justice and the rule of law
- Equality, non-discrimination and solidarity
- Citizens' rights

## **Ethical Principles in the context of AI systems**

- The principle of respect for human autonomy
- The principle of prevention of harm
- The principle of fairness
- The principle of explicability

A red flag is waving on a silver pole against a clear blue sky. The flag is positioned diagonally across the frame, with its top right corner near the top right of the image. A dark grey horizontal band is superimposed over the middle of the image, containing the text "RED FLAGS" in white, bold, sans-serif capital letters.

# RED FLAGS



## **Examples of critical concerns raised by AI:**

- Diversity, non-discrimination and fairness
- Identifying and tracking individuals with AI
- Covert AI systems
- AI enabled citizen scoring in violation of fundamental rights
- Lethal autonomous weapon systems (LAWS)
- Potential longer-term concerns (“black swans”)

A woman with dark curly hair, wearing a brown jacket over a yellow shirt, blue jeans, and orange sandals, is captured mid-jump in front of a textured green wall. Her arms are outstretched, and she has a joyful expression. The wall has some peeling paint at the bottom. The text 'THANK YOU!' is overlaid in large white letters, and '@EWLA1 @kathamiller' is below it.

# THANK YOU!

@EWLA1 @kathamiller