

The need for an interdisciplinary approach to AI research

Developing research ethics for AI

Workshop @ Comète on Ethical AI

23-24 November 2023

Palaiseau, France

Anaïs Rességuier (PhD), Trilateral Research



EU-FUNDED RESEARCH PROJECTS

TECH Ξ THOS

 irecs


 DARLENE

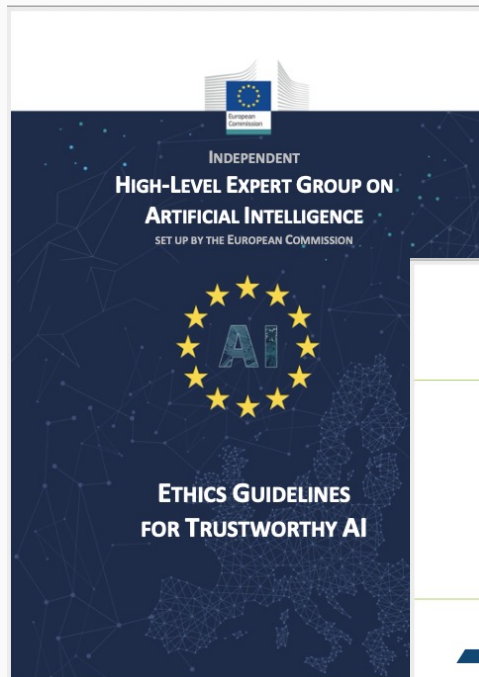
With thanks to Fabienne Ufert

OUTLINE

1. AI ethics and its discontent
2. A distinction within ethics
3. Research ethics for AI

1. AI ETHICS AND ITS DISCONTENT

AI ETHICS



(HLEG-AI, 2019)



(OECD, 2019)

Objectives for AI applications

We will assess AI applications in view of the following objectives. We believe that AI should:

1. Be socially beneficial.
2. Avoid creating or reinforcing unfair bias.
3. Be built and tested for safety.
4. Be accountable to people.
5. Incorporate privacy design principles.
6. Uphold high standards of scientific excellence.
7. Be made available for uses that accord with these principles.

(Google, undated)



(UNESCO, 2022)

AI ETHICS

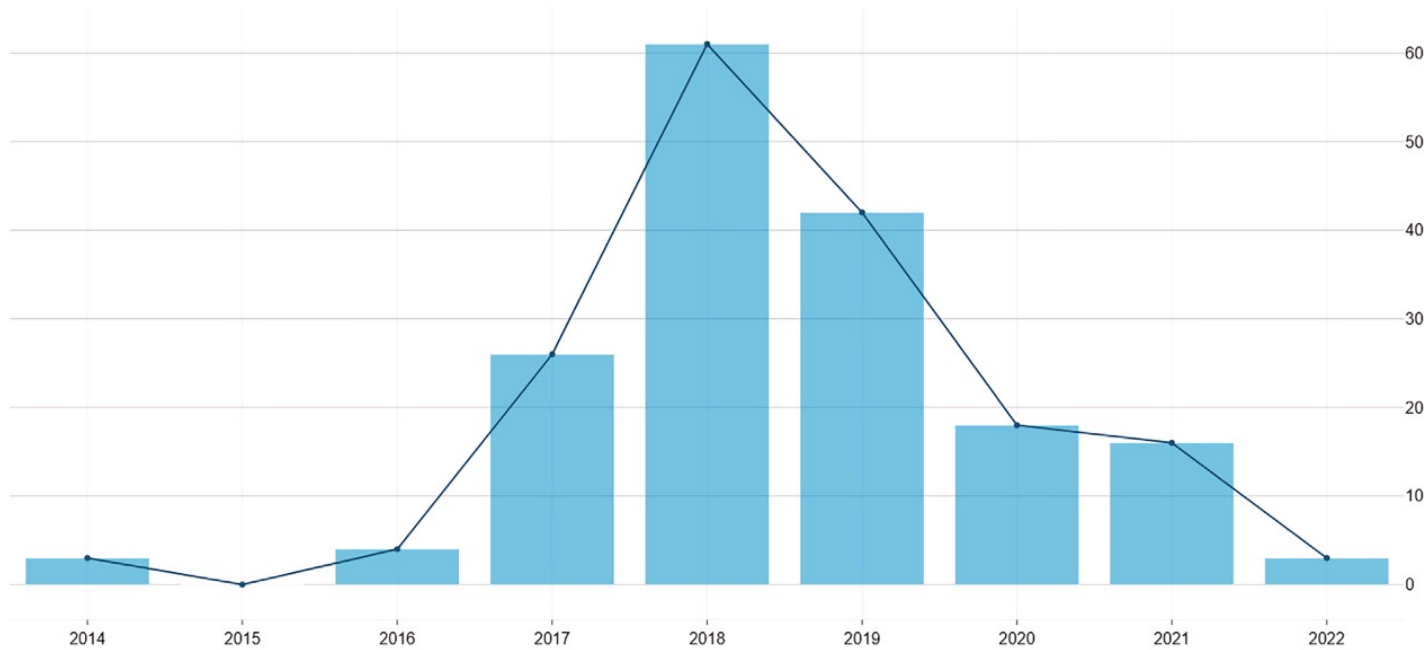


Figure 4. Number of publications per year

Corrêa et al., 2023, "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance"

DISCONTENT WITH AI ETHICS

- ***AI ethics being misused to limit regulation***

(Wagner, 2018; Ochigame, 2019, Resseguier and Rodrigues, 2020, etc.)

- ***AI ethics as securing existing power relations***

(D'Ignazio, Klein, 2020, etc.)

DISCONTENT WITH AI ETHICS

- ***AI ethics being misused to limit regulation***

(Wagner, 2018; Ochigame, 2019, Resseguier and Rodrigues, 2020, etc.)

- ***AI ethics as securing existing power relations***

(D'Ignazio, Klein, 2020, etc.)

AI ethics lack teeth

DISCONTENT WITH AI ETHICS

- ***AI ethics being misused to limit regulation***

(Wagner, 2018; Ochigame, 2019, Resseguier and Rodrigues, 2020, etc.)

- ***AI ethics as securing existing power relations***

(D'Ignazio, Klein, 2020, etc.)

AI ethics lack teeth

ETHICS WASHING

HOW TO ADDRESS THESE CRITIQUES?

- Developing regulatory landscape, especially with the upcoming AI Act
- *What about within ethics?*

*Need for more
practical and effective
implementation of AI
ethics*



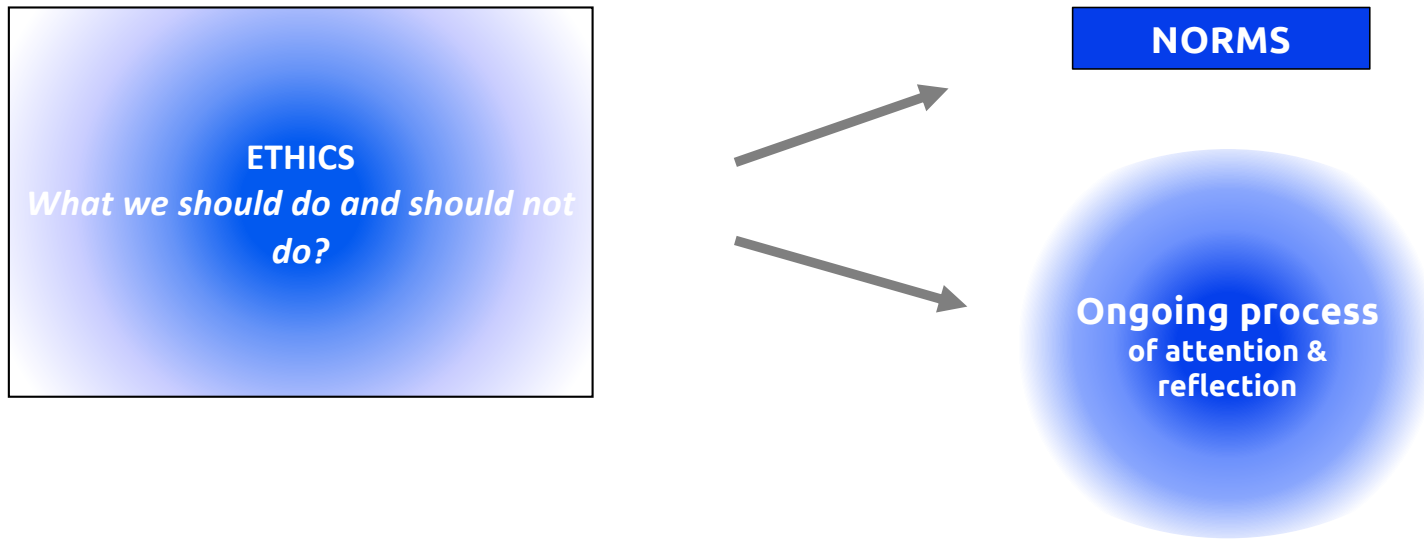
**RESEARCH
ETHICS FOR AI**

2. A DISTINCTION WITHIN ETHICS

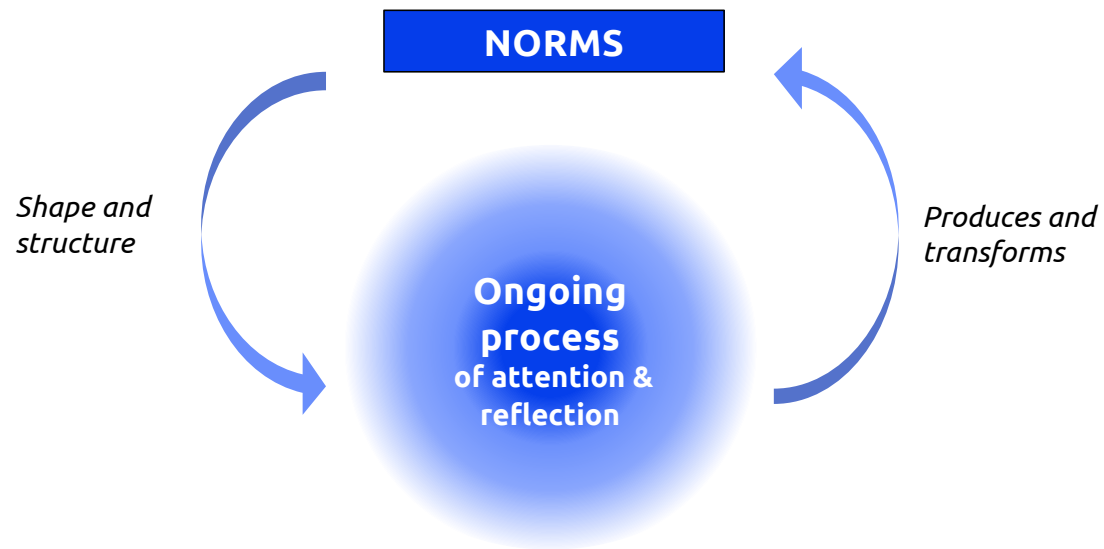
A DISTINCTION WITH ETHICS



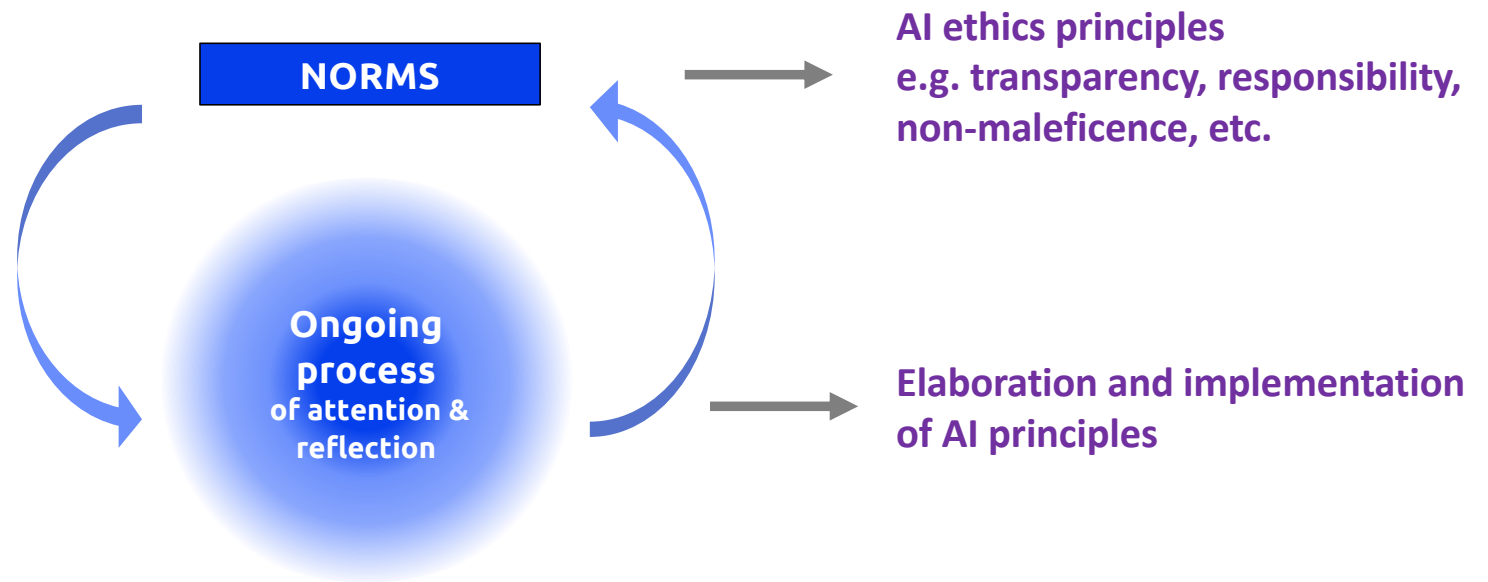
A DISTINCTION WITH ETHICS



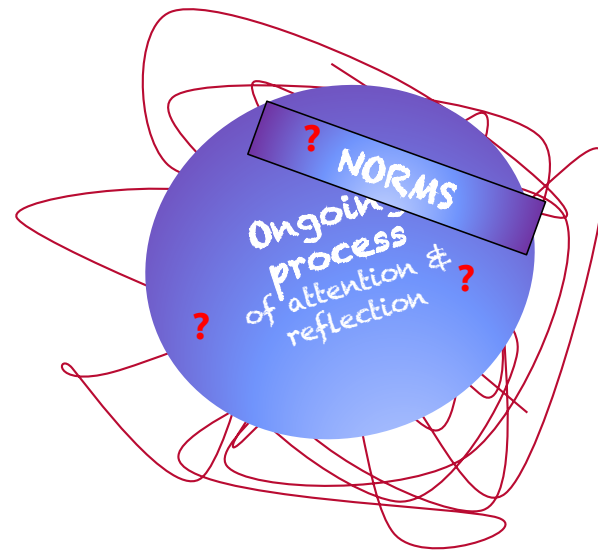
A DYNAMIC RELATIONSHIP



AI ETHICS



CONFUSION AND WEAKENING OF ETHICS



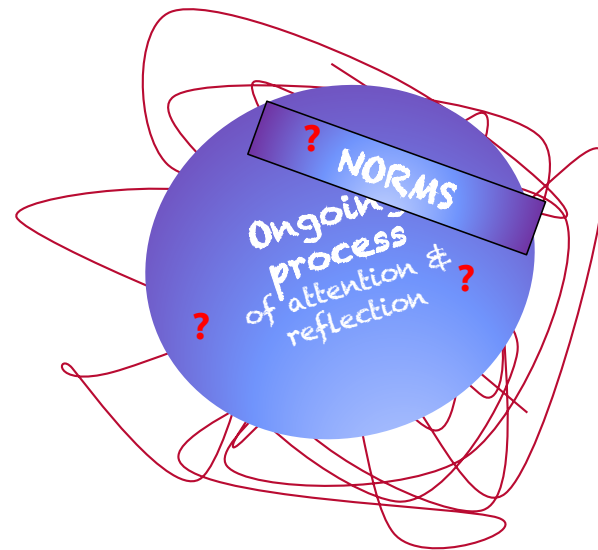
Weak norms:

- *easily manipulated (ethics washing)*

Weak ongoing process:

- *Poor and superficial reflection*
- *Inattention to context*
- *No questioning of power relations*

CONFUSION AND WEAKENING OF ETHICS



Weak norms:

- *easily manipulated (ethics washing)*

Weak ongoing process:

- *Poor and superficial reflection*
- *Inattention to context and its evolution*

AI ethics that lacks teeth

ETHICS WASHING

STRENGTHENING BOTH ASPECTS OF ETHICS FOR AN EFFECTIVE AI ETHICS

NORMS

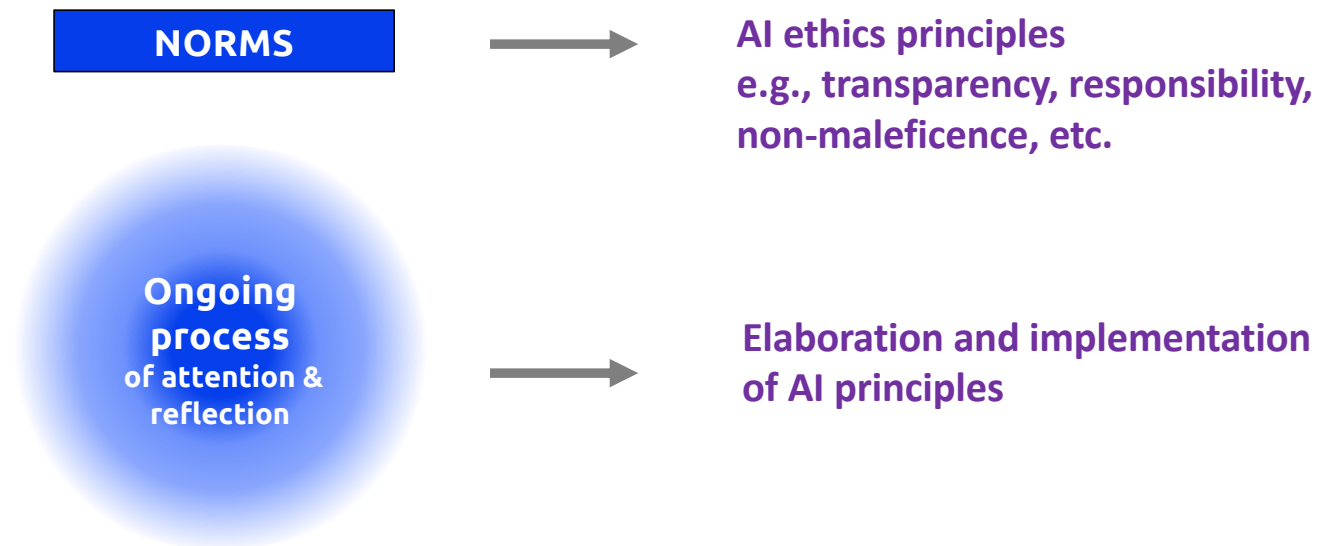
**Ongoing process
of attention &
reflection**

STRENGTHENING BOTH ASPECTS OF ETHICS FOR AN EFFECTIVE AI ETHICS

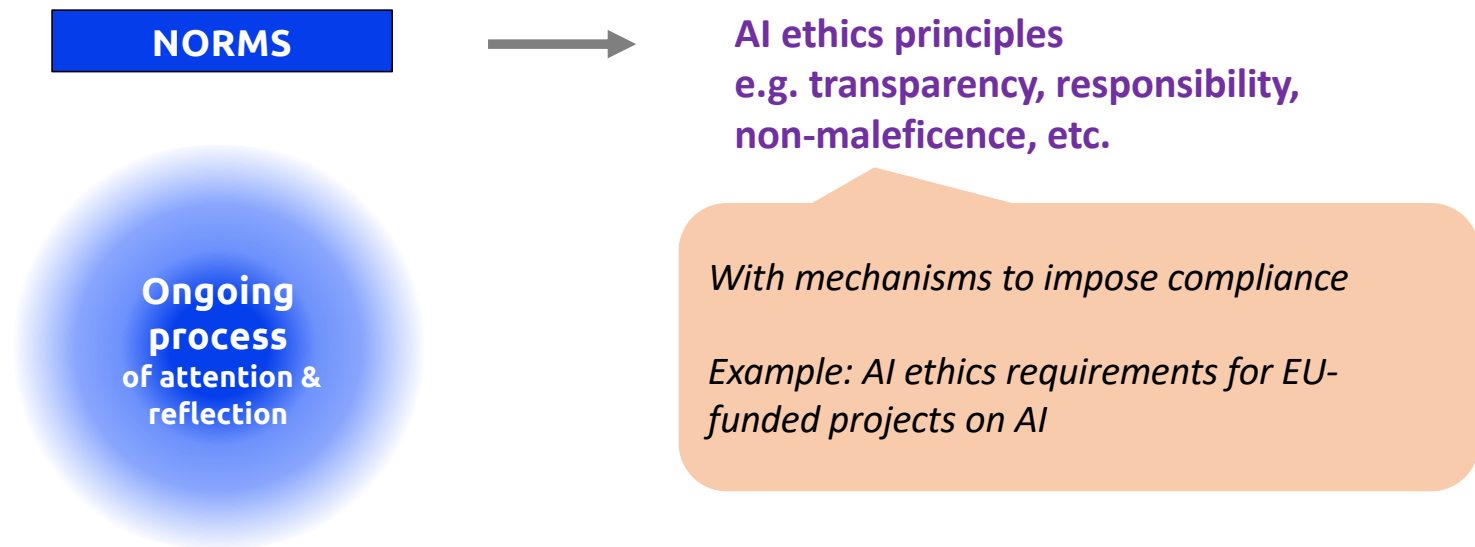


3. RESEARCH ETHICS FOR AI

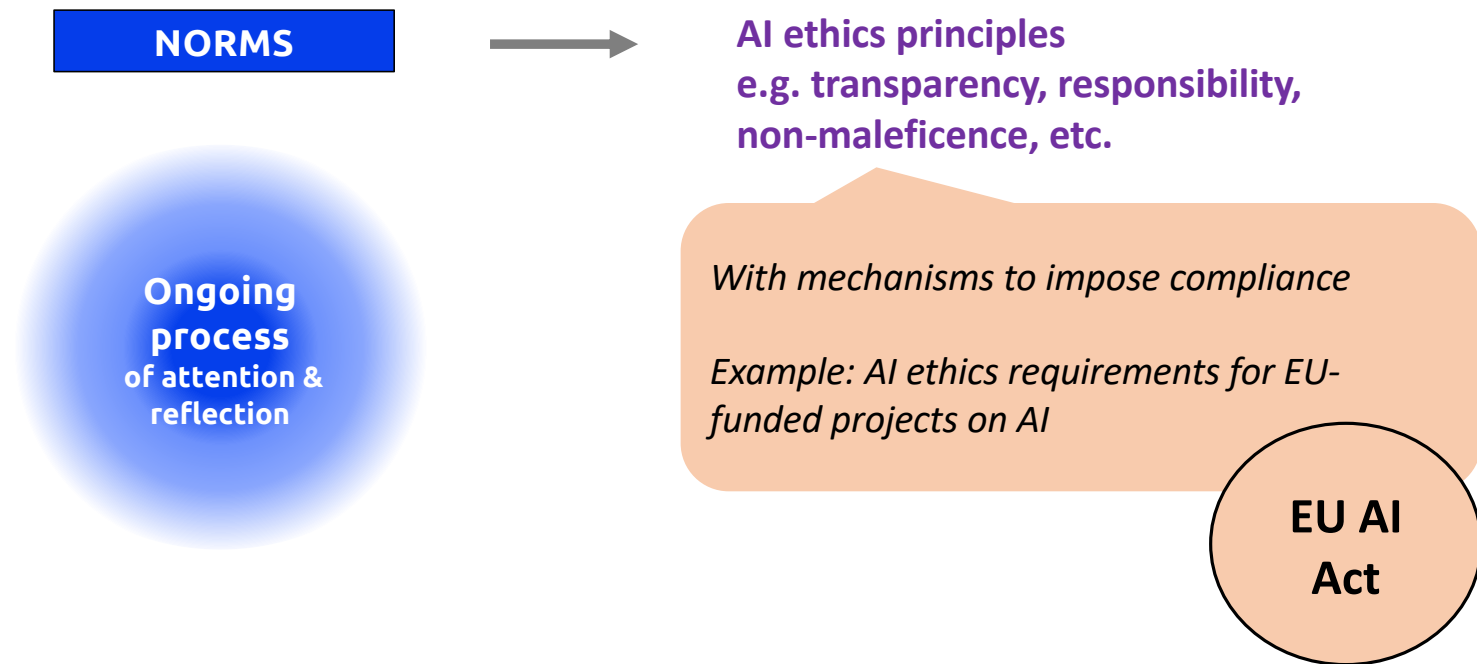
RESEARCH ETHICS FOR AI



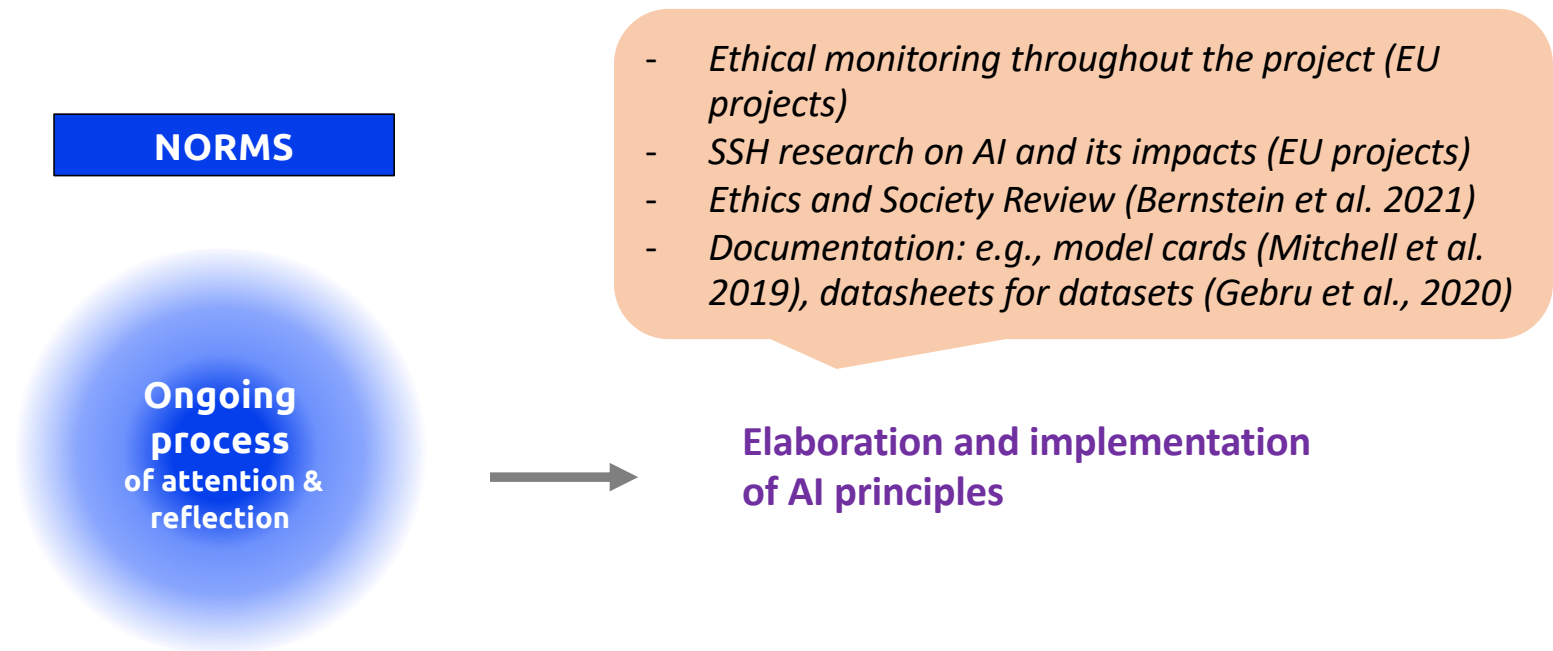
RESEARCH ETHICS FOR AI



RESEARCH ETHICS FOR AI



RESEARCH ETHICS FOR AI



ADAPTATIONS OF RESEARCH ETHICS TO AI

- **Research ethics throughout the project**
 - *as opposed to one-off compliance check at the start of the project*
- **Engage with broad impacts on the society once the system is deployed**
 - *as opposed to impact on human participants only*
- **To be implemented in industry settings**
 - *not only in academic environments*

KEY TAKE AWAYS

TWO ASPECTS OF ETHICS:

- Norms
- Open process of reflection and attention

CONFUSION BETWEEN THESE 2 ASPECTS LEADS TO WEAKENING BOTH:

- Ethics washing
- Weakeness of critical perspective in AI ethics

KEY ROLE OF RESEARCH ETHICS FRAMEWORKS TO STRENGTHEN AND OPERATIONALISE AI ETHICS

RESEARCH ETHICS FOR AI:

- Norms with requirements for compliance
- Elaborate and embed norms throughout research projects