Dipartimento di Ingegneria Informatica
Automatica e Gestionale Antonio Ruberti
Sapienza
Università di Roma

CPS – As defined by STOA

Cyber-physical systems (CPS) are intelligent robotics systems, linked with the Internet of Things, or technical systems of networked computers, robots and artificial intelligence that interact with the physical world.

CPS - Expectations

By 2050, these systems may interact with us in many domains, driving on our roads, moving alongside us in our daily lives and working within our industries.

Due to the wide range of situations where we will be interacting with CPS, understanding the impacts of these systems is essential.

CPS - Expectations

Due to the wide range of situations where we will be interacting with CPS, understanding the impacts of these systems is essential.

In addition, it is essential to have systems that behave in an ethically acceptable way: the impacts are beneficial !!

We start from Asimov, of course

Isaac **Asimov** in his **1942** short story 'Runaround' (set in 2015), introduced the Three Laws of Robotics — **engineering safeguards and built-in ethical principles**

Asimov's Laws of Robotics

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 the 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 Laws.

Machine Ethics?

We may say that machine ethics pertains to science fiction, or say that ethics is a prerogative of human behavior, and that machines do not/cannot have it.

It is more appropriate to speak about ethical behavior.

Human rights

Sandewall 2018

Universal Declaration of Human Rights ONU 1948

From Art 1. All human beings are **born free** and **equal in dignity and rights**. They are **endowed with reason and conscience** and should act towards one another in a **spirit of brotherhood**.

Human rights

- Freedom
- Dignity
- Reason, being endowed with reason
- Free will
- Morality

Adding an MBS

PAS Planning and Action System based on a World Model

Add a software module: a “Moral Belief System” (MBS)

Autonomous modifications of the MBS

Free will?

John McCarthy 2000 makes a proposal on how to endow a robot with free will

But: Can a robot be conscious?

Pioneers of Machine ethical studies

Anderson & Anderson AI Magazine '07

- Distinction between Computer ethics and Machine ethics
  - Interdisciplinary research area
  - Analysis of ethic behavior and difficult decisions/choices
    "act utilitarianism"
  - Relativism of ethics
  - Investigation on how to reproduce ethic behavior
  - Explicit representation of ethical principles
  - Example of "abstraction" of ethical principles

A prominent work

Abhilash Thekkilakattil Gordana Dodig-Crnkovic 2015

Ethical aspects of CPSs are extra-functional properties, that cover the whole life cycle

- **design**
- **development**
- **deployment/production**
- **use**

One of the ethical challenges involved is the question of identifying the **responsibilities of each stakeholder** associated with the development and use of a CPS.

Various levels of autonomy

- Automatic Systems
- Semi-automatic Systems
- Semi-autonomous systems
- Autonomous systems

Can we predict the behavior of an automatic system?

Yes

Can we predict the behavior of an autonomous system?

No

Embedded EthiCS, Grosz et al, Harvard

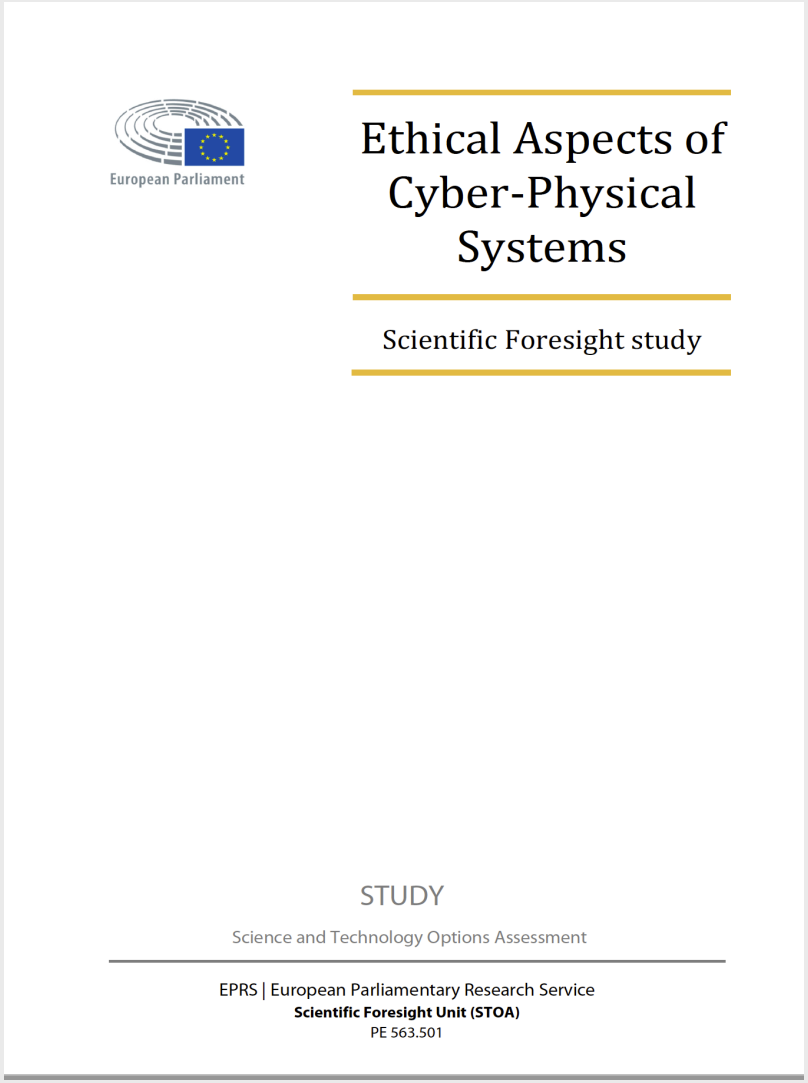
An alternative and more integrative approach to incorporating ethical reasoning into computer science education

In contrast to stand-alone computer ethics or computer-and-society courses, it makes ethical reasoning an integral component of courses throughout the standard computer science curriculum

Students can learn to think not only about
what technology they could create, but also
whether they should create that technology

Interdisciplinary teaching/teachers

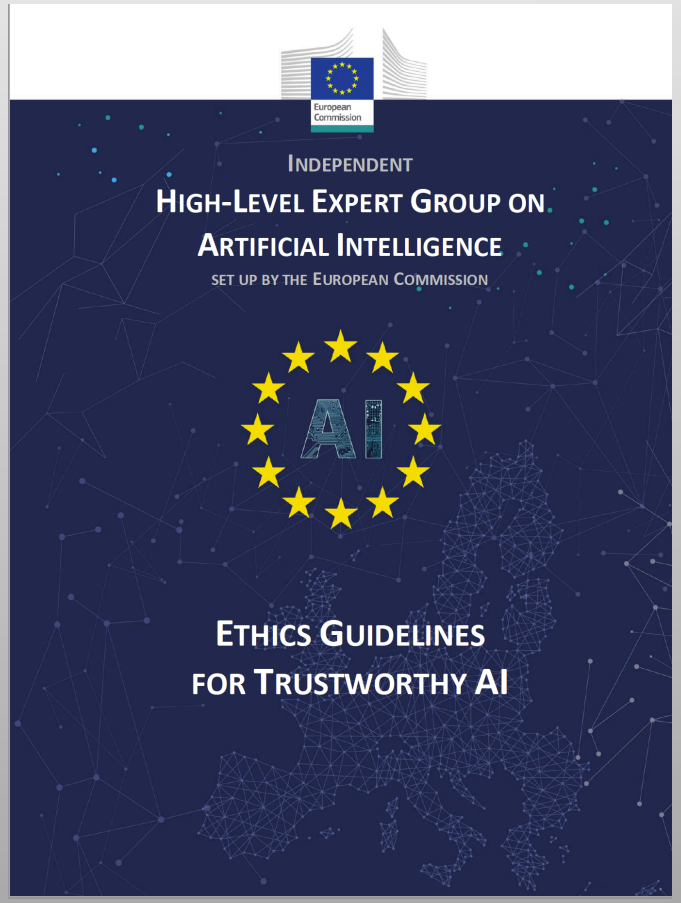
STOA 2016


European Parliament
Ethical Aspects of Cyber-Physical Systems
Scientific Foresight study
STUDY
Science and Technology Options Assessment
EPRS | European Parliamentary Research Service
Scientific Foresight Unit (STOA)
PE 563.501

Seven areas of concern

- Disabled people and daily life of elderly people
- Healthcare
- Agriculture and food supply
- Manufacturing
- Energy and critical infrastructures
- Logistic and transport
- Security and safety

High Level Expert Group on AI



Foundations

- **LAWFUL - ETHICAL - ROBUST**
- Ethical Principles:
  - Respect for human autonomy
  - Prevention of harm
  - Fairness
  - Explicability

Requirements

- Developers
- Deployers
- End users and broader society

Seven equally important and interconnected requirements, to be implemented and evaluated throughout the systems lifecycle.

Seven requirements

Principle of Respect for human autonomy

Support human autonomy and decision making
Enable and not hamper fundamental rights
Human agency
Human oversight

HITL, HOTL, HIC

Principle of Prevention of harm

Physical and mental integrity
of humans must be ensured

Resilience to attacks and security

Fallback plan and general safety

Accuracy

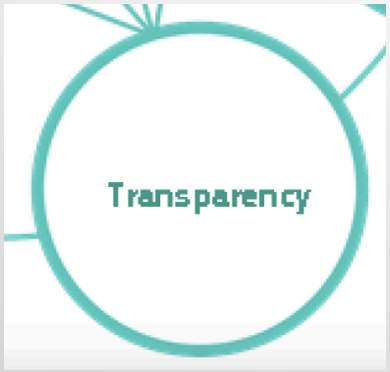
Reliability and Reproducibility

Principle of Prevention of harm

Privacy and data protection guaranteed throughout the entire lifecycle

Quality and integrity of data

Access to data

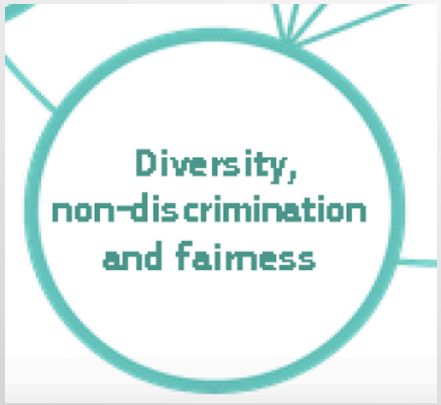


Principle of Explainability

Traceability

Explainability

Communication



Principle of Fairness

Avoidance of unfair bias

Accessibility and universal design

Stakeholder participation

Principle of Fairness and Prevention of harm

Sustainable and environmentally friendly

Social impact

Society and democracy

Principle of Fairness

Auditability

Minimization and reporting of negative impacts

Trade offs

Redress

A Summary by F. Rossi

- Properties of the technology
  - Bias (Careful with the examples)
  - Value alignment (Ethical, moral social and legal constraints)
  - Black box (Must be able to explain decisions)
- Governance and rules
  - Data issues (Privacy, storage, ownership, use)
  - Accountability (Who is responsible if something goes wrong?)
- Impact on jobs
  - How to cope with jobs transformation?
- Impact on society
  - People machine and machine-machine interactions
- Deep fake
  - AI can generate content that looks real but it is not
- Autonomous weapons, surveillance systems
  - Are these acceptable uses?
- Superintelligence
  - Will we loose control?

Readings 1

United Nations, General Assembly resolution 217 A (1948)
Universal Declaration of Human Rights. http://www.un.org/en/universal-declaration-human-rights/index.html

Ethics, Human Rights, the Intelligent Robot, and its Subsystem for Moral Beliefs - E. Sandewall - Int. J. of Social Robotics (2019)

Free will - even for robot - J. MaCarthy – J. Exp. and Theor. Artif. Intell. (2000)

Machine Ethics: Creating an Ethical Intelligent Agent – M. Anderson, S. L. Anderson - AI Magazine (2007)

Ethics Aspects of Embedded and Cyber-Physical Systems - A.Thekkilakattil, G. Dodig-Crnkovic (2015)

Embedded EthiCS: Integrating Ethics Across CS Education - Harvard pilot project – Barbara Grosz et al - CACM August 2019

Readings 2 – Docs from the EU

Ethical Aspects of Cyber-Physical Systems - Scientific Foresight study - EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 563.501 + Annexes (2016)

Ethics Guidelines for Trustworfhy AI – High-Level Expert Group on AI – European Commission (2019)

Thank you for your attention!