A Trustworthy approach to AI: the EU example

Upholding ethical standards in a disruptive technological domain

Dr Mihalis Kritikos, Ethics & Research Integrity Sector

EU initiatives in the field of AI

ETHICS GUIDELINES/REQUIREMENTS

PROPOSAL FOR AN AI LIABILITY DIRECTIVE

COORDINATED PLAN ON AI

PROPOSAL FOR AN AI ACT

A Trustworthy approach to AI

Trustworthy AI

-HLGAI: Ethics Guidelines

-ALTAI List

-EU-funded projects on AI ethics

AI Act

-HRIAs

-focus on the protection of fundamental rights

CEN/CENELEC AI ethics standards

Bias in AI systems and AI aided decision making

Competence Requirements for AI ethicists professionals

Artificial Intelligence trustworthiness characterisation

AI Act as a framework of ethical governance

It aims at **mitigating the harmful use of AI.**

It facilitates the **transparent, ethical use of AI — and keeps machine intelligence under human control.**

Fines for non-compliance are significantly higher than those for the EU's General Data Protection Regulation (GDPR), **ranging up to 30 million Euros, or 6% of annual revenue.** In contrast, the GDPR imposes fines of up to 20 million Euros or 4% of revenue.

Jurisdictional reach: If adopted, the EU AI Act will impose **a set of obligations on both providers and deployers of in-scope AI systems used in or producing effects in the EU, irrespective of their place of establishment.**
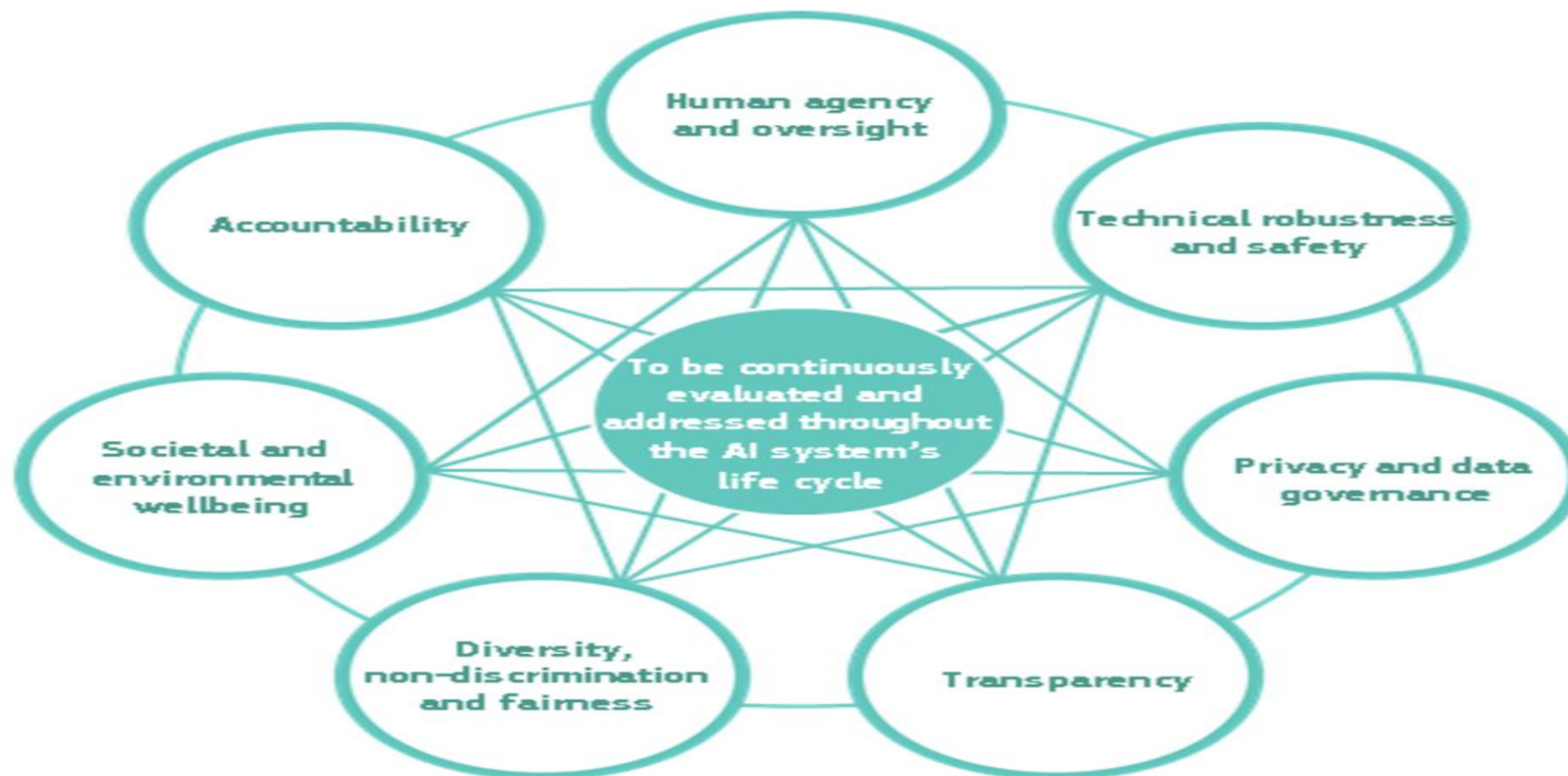
EU (trustworthy) approach to AI

Trustworthy AI should be:

- (1) lawful - respecting all applicable laws and regulations
- (2) ethical - respecting ethical principles and values
- (3) robust - from a technical perspective
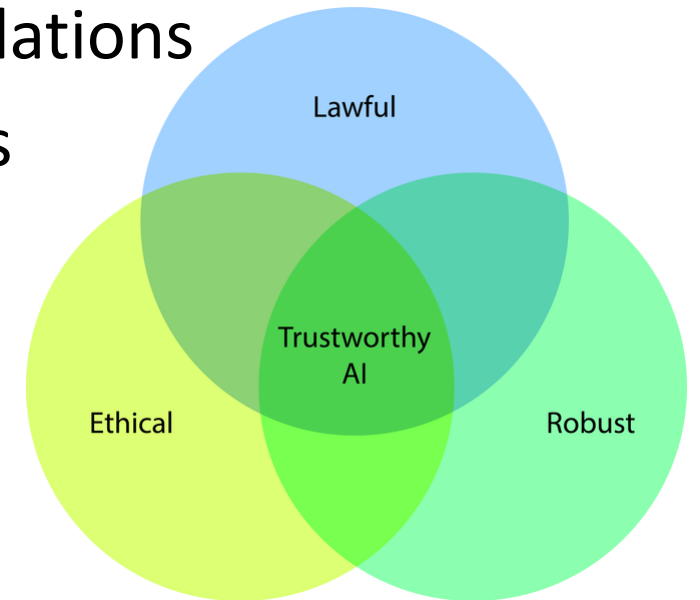
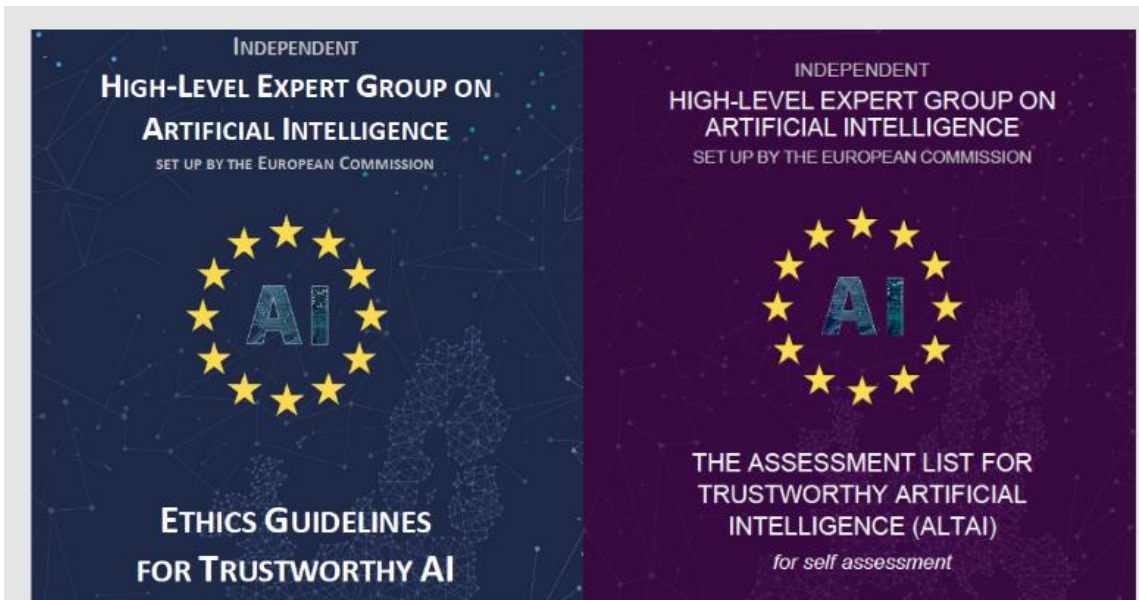# Ethics Guidelines for Trustworthy AI

Ethics Guidelines for Trustworthy AI

4 ethical principles:

- Respect for huma autonomy

- Prevention of harm

- Fairness

7 core requirements:

1. Human agency and oversight

2. Technical robustness and safety

3. Privacy and Data Governance

4. Transparency

5. Diversity, non-discrimination and fairness

6. Societal and environmental wellbeing

7. Accountability

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION

AI

ETHICS GUIDELINES
FOR TRUSTWORTHY AI

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION

AI

THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

The Ethics Issues

1. Human embryonic stem cells & human embryos

2. Humans

3. Human cells/tissues

4. Personal data

5. Animals

6. Non-EU countries

7. Environment, health & safety

8. **Artificial Intelligence**

9. Other ethics issues

Self-Assessment	(Pre-)Screening	Assessment	Check/Review

European Commission

Artificial Intelligence
Ethics appraisal framework

AI is a self-standing ethics issue!

AI is considered as an ETHICS issue for **ALL** activities involving:

- the development, deployment and/or use of AI-based systems and techniques

- AI-based system as components of larger systems.

European Commission

Artificial intelligence
Scientific & ethics evaluation

Trustworthy Artificial Intelligence

- **Scientific experts** to answer a specific question:

  - Do the activities proposed involve the use and/or development of AI-based systems and/or techniques?

  - If so, scientific experts' must assess the **technical robustness*** of the proposed AI-system as part of the excellence criterion.

- **The ethics experts** take into account the assessment on the technical robustness when performing their ethics evaluation.

(*) Technical robustness refers to technical aspects of AI systems and development, including resilience to attack and security, fullback plan and general safety, accuracy, reliability and reproducibility.

European Commission

Ethics Appraisal Process
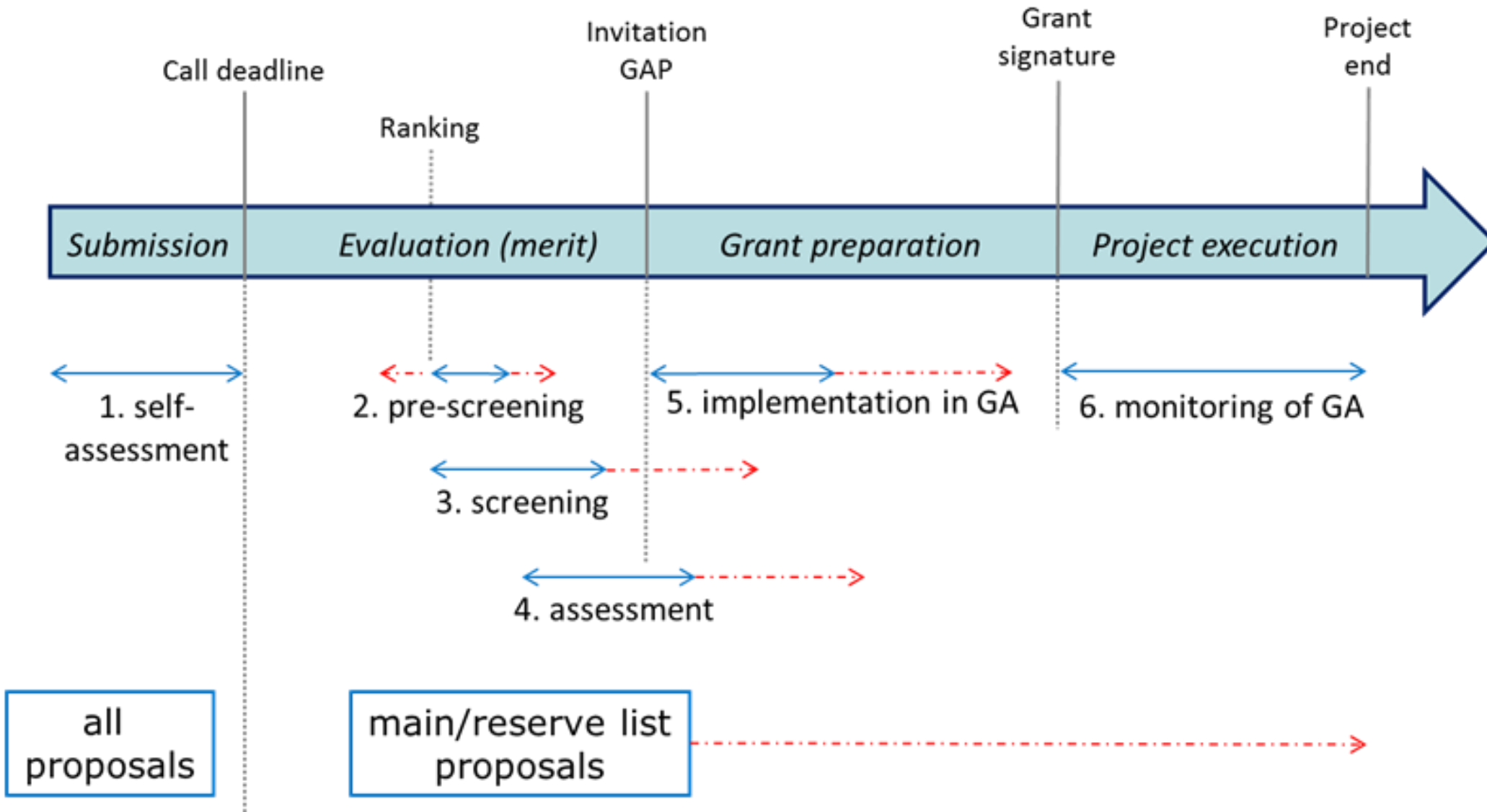
→ Ethics appraisal process for Horizon Europe is a **corporate process**
   → **Adapted** to the need specific to the programme
→ **Implemented** by executive agencies and joint undertakings
! Ongoing serious/complex projects ➜ reviewed by DG RTD Ethics Sector

Executive Agencies

Joint Undertakings

Horizon Europe
+
Other funding programmes:
EDF, DEP, EU4Health, …

Guidance on Ethics By Design for AI

Builds on the work of the **Independent High-Level Expert Group on AI** and their **'Ethics Guidelines for Trustworthy AI'**

Addressing ethical issues during research and development

Ethics By Design and Ethics of Use Approaches
for Artificial Intelligence

Version 1.0
25 November 2021

Structure

Part 1: Principles and requirements: This part defines the ethical principles that AI systems should adhere to and derives requirements for their development;

Part 2: Practical steps for applying Ethics by Design in AI development: This section explains the Ethics by Design concept and relates it to a generic model for the development of AI systems. It defines the actions to be taken at different stages in the AI development in order to adhere to the ethics principles and requirements listed in Part 1;
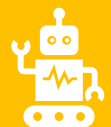
- Part 3: Ethical deployment and use: it presents guidelines for deploying or using AI in an ethically responsible manner.

What is coming

Guidance Note on *human-centered AI: algorithmic bias and fairness*
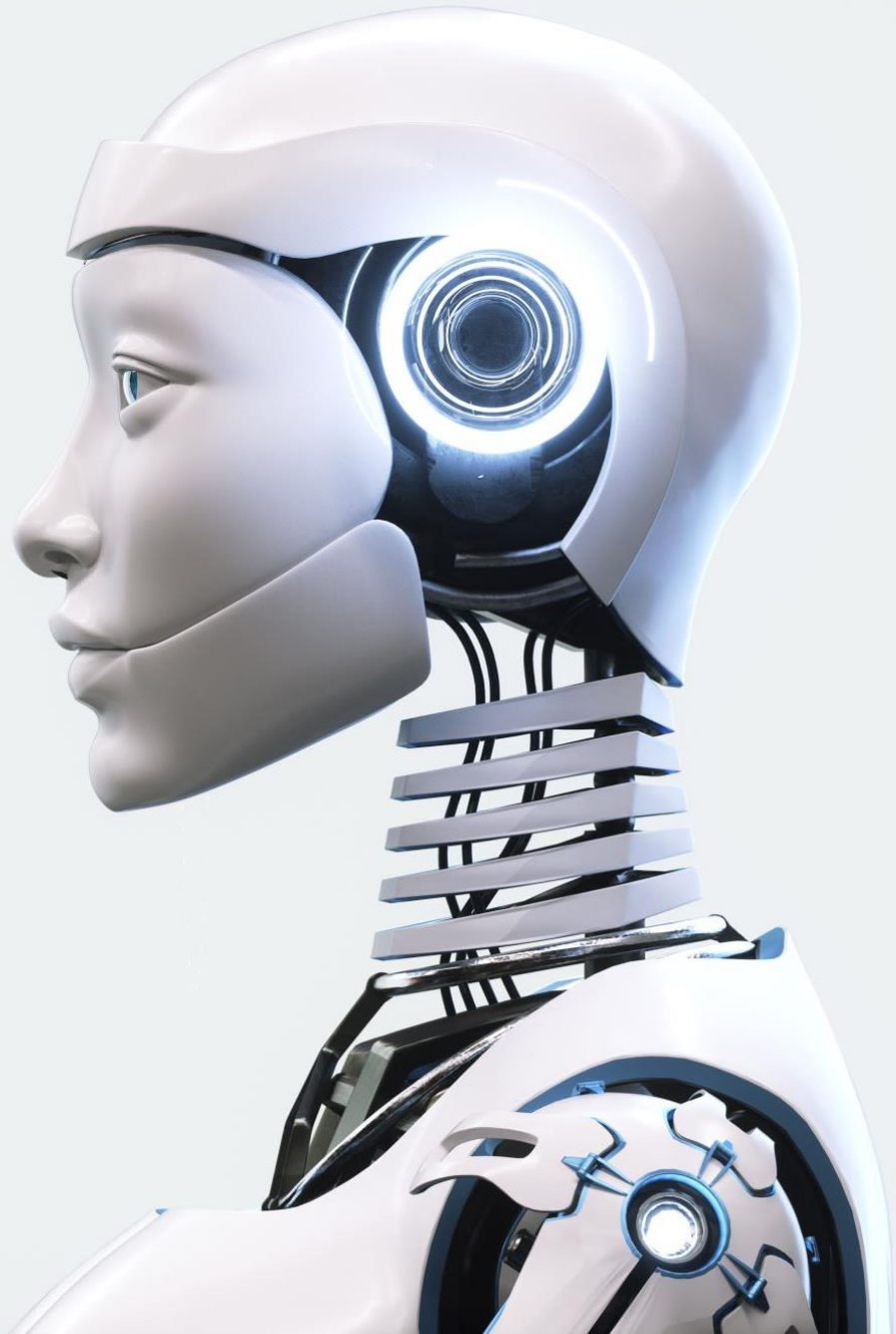
Guidance Note on *AI and informed consent*

Guidance Note on *AI Ethics Audits and Checks*

Guidance Note on *AI Ethics and project lifecycle*

International initiatives on the ethics of AI

UNESCO 'Recommendation on the Ethics of Artificial Intelligence' in November 2021. This framework was adopted by all 193 Member States.

OECD AI principles (first set of intergovernmental policy guidelines on AI).

UNICEF policy guidance on AI for children

Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law

WHO guidance on the ethics and governance of AI for health, when designing, developing, and deploying AI for health.

6 core principles identified by WHO are:

(1) protect autonomy; (2) promote human well-being, human safety, and the public interest; (3) ensure transparency, explainability, and intelligibility; (4) foster responsibility and accountability; (5) ensure inclusiveness and equity; (6) promote AI that is responsive and sustainable.
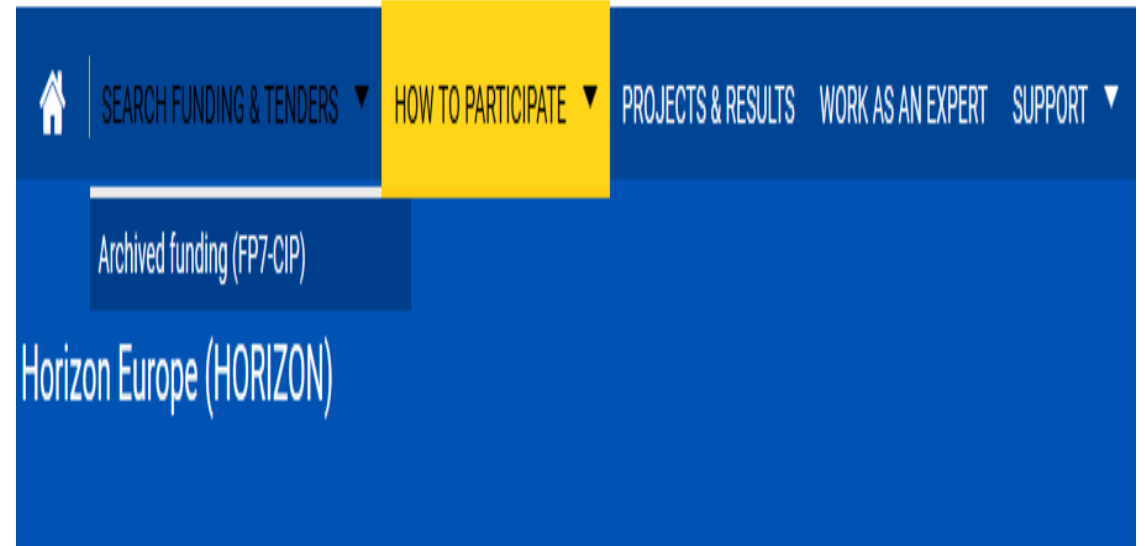
European Parliament

- The 2017 resolution of the European Parliament on the civil law rules on robotics and AI prioritized six main areas for EU legislative efforts: ethics, liability, intellectual property and flow of data, standardization, employment and institutional coordination and oversight.

-Recommendations for a code of conduct for robotics scientists, where the role of ethical design and responsible research was recognized.

Key sources and materials

- **Guidelines on ethics by design for AI**

- **Guidance on Identifying serious and complex ethics issues in EU-funded research (AI section)**

- **Assessment List for Trustworthy Artificial Intelligence** (ALTAI) (Independent High-Level Expert Group on AI)

- **Ethics guidelines for trustworthy AI** (Independent High-Level Expert Group on AI)

- How-to complete your ethics self-assessment ('How-to')

Thank you for your attention!