# Governing Autonomous AI Agents in Europe Ethics, Law and Industrial Policy in the Age of Agentic AI

Mihalis Kritikos

European Commission

# Legal/Policy Challenges

AI agents might require **extensive access to data stored on the devices**

**Challenging to determine in advance what personal data is gathered**, how it is used, and for what specific purposes.

Lack of transparency

The continuous adaptation of Agentic AI **can potentially perpetuate and amplify existing biases**

**Potential accountability gap**

ChallengeAgent behavior can evolve dynamically and cross multiple regulatory categories

# Ethical Challenges of Autonomous Agents

| | | | |
|---|---|---|---|
| How much autonomy is acceptable? | Can humans still exercise meaningful oversight? | Can decisions be explained after the fact? | Autonomy may outpace accountability |
| Opacity (“Black Box” Problem) | Bias and fairness challenges | Misuse risks | Expertise Gap |

# Main challenges

**Translate principles into concrete requirements**

Turn values into checklists, documentation and measurable indicators

**Embed ethics into workflows**

Integration into:

- data collection (consent, representativeness)
- model development (bias audits, explainability tools)
- deployment (monitoring)

Not a one-time review → continuous lifecycle process

**Interdisciplinary implementation**

Combine technical, legal, and ethical expertise

Requires new roles (AI ethics officers and auditors)

# Key Challenges

01

**Ambiguity** of ethical concepts (e.g., “fairness” varies by context)

02

**Lack of standardised metrics** across jurisdictions

03

**Trade-offs:** Transparency vs privacy / Accuracy vs fairness

04

**Enforcement gaps** in global frameworks

05

Most frameworks are non-binding or partially enforceable

# No EU Agentic Law (yet)

**Current EU law does not create a separate legal category for "agents“**

Instead, agents are treated as either:

- **AI systems under the AI Act -high-risk systems when deployed in sensitive contexts**
- **General-purpose AI (GPAI) based systems under the AI Act**

Most advanced agents are built on GPAI models.

- **Software products under product safety and liability rules**

# AI Act

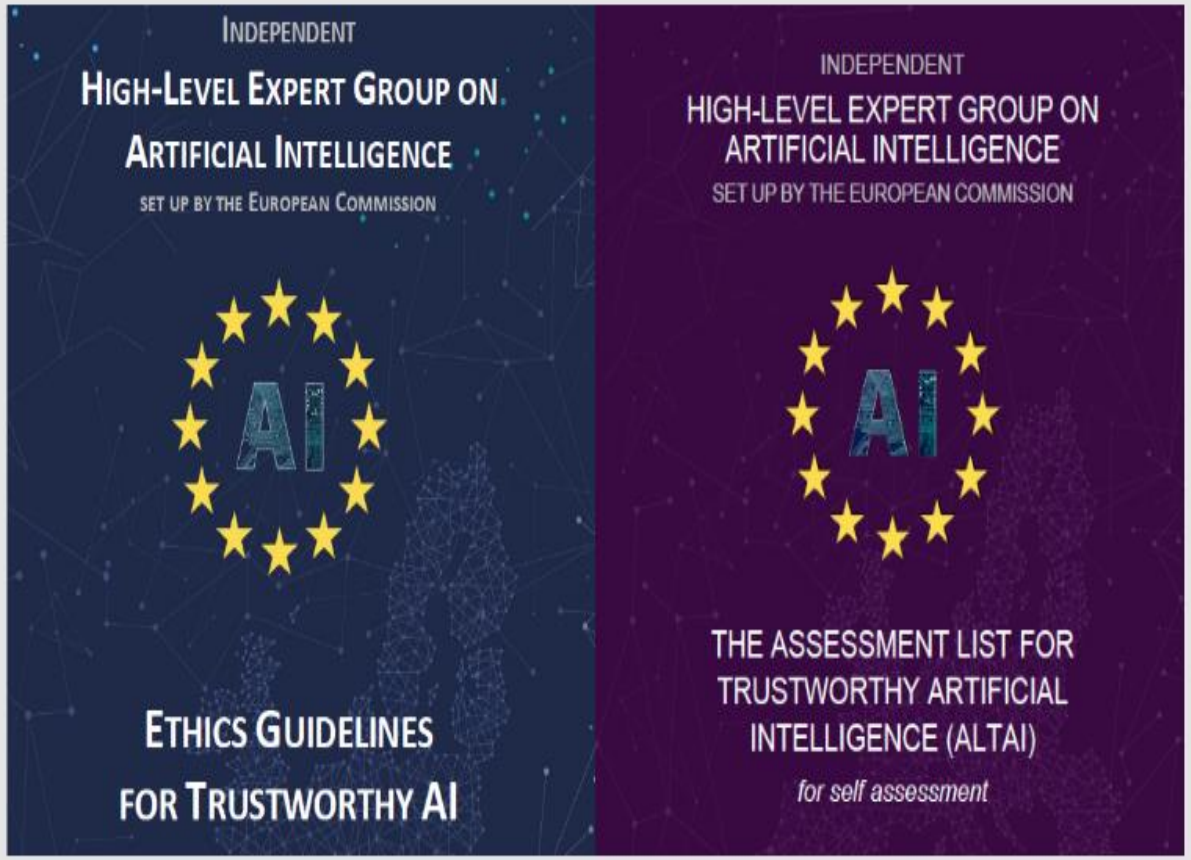
The regulatory framework applies rules across the AI value chain to developers, system providers, and deployers. Key governance components include:

- **Risk-Based Categorization:** Autonomous AI agents are assessed by their domain of application.
- **High-Risk Obligations:** High-risk autonomous agents must feature continuous quality monitoring, robust logging for traceability, and clearly defined human oversight mechanisms with the technical ability to "stop" or override the system.
- **GPAI Foundation Model Rules:** Since most autonomous agents are scaffolding built on top of broad AI models, the model providers are required to assess systemic risks and implement transparency/copyright measures.

**The AI Act governs AI agents through four primary pillars:**

risk assessment, transparency tools, technical deployment controls, and human oversight design.

# EU ethics governance framework




**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

# EU’s ethical approach for AI in research

AI as a **self-standing ethics issue** in Horizon Europe

**‘Ethics by design’ approach** integrated into **Horizon Europe** ethics appraisal process.

**Dedicated evaluation of serious and complex AI-related Horizon Europe** proposals/projects

**Trustworthy and Ethical AI Research** is the general approach for Horizon Europe

# Lifecycle Operationalisation of Ethics

1. **Design Phase:** Ethical risk assessment + Dataset documentation (data sheets)
2. **Development Phase:** Fairness constraints in training / Explainable models
3. **Testing Phase:** Bias & robustness testing / Adversarial evaluation
4. **Deployment Phase:** Human-in-the-loop systems / Transparency reports
5. **Monitoring Phase:** Continuous audits / Incident reporting systems

# 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'**

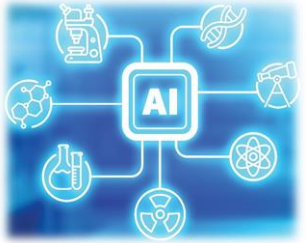
**Addressing ethical issues during research and development**

# EU Operational Guidelines in ethics /June 2026

**GUIDANCE NOTE ON *AI AND INFORMED CONSENT***

**GUIDANCE NOTE ON FUNDAMENTAL *RIGHTS IMPACT ASSESSMENT***

# Framework of AI EU initiatives




## AI Continent

- Apply AI Strategy
- AI in Science Strategy
- Data Union Strategy
- Cloud and AI Development Act
- AI Gigafactories

RAISE
(Resource for AI Science in Europe)

# AI in Science Strategy

- Acceleration and innovation
- Cross-disciplinary application
- Resource coordination
- **Ethical and responsible use**
- Competitiveness
- Infrastructure development
- Talent and collaboration:

Foster collaboration and develop talent to drive AI

# What to watch in the next 12 months

- First major enforcement actions under the AI Act
- Publication of final guidance on high-risk AI classification
- How the EU AI Office supervises frontier model providers
- Enforcement of transparency obligations for generative AI and deepfakes
- Potential simplification measures aimed at improving European AI competitiveness

# EU Omnibus agreement

**AI Act Adjustments:** Clarifies the overlap between the AI Act and product safety laws (like the Machinery Regulation) to avoid duplication. It also extends certain privileges and regulatory sandbox access previously reserved for SMEs to mid-cap companies.

**Delayed Implementation for High-Risk AI**: Sets a staggered timeline for the implementation of high-risk AI rules.

General high-risk systems will apply from **December 2, 2027,** while systems embedded in products like lifts or toys will apply from **August 2, 2028.**

**Citizen Protection & Bans:** Prohibits AI systems that generate child sexual abuse material (CSAM) or non-consensual sexually explicit content, such as AI "nudification" apps

# What Still Needs To Be Done

- Agent-specific accountability frameworks
- Mandatory oversight mechanisms
- Continuous monitoring
- Independent audits
- Governance of multi-agent systems