Trustworthy AI
And the European AI act

G.Landais

Trustworthy AI

How can I provide a trustworthy AI ?

→ Am I subject to any rules?
→ Are there ethic ? good practice guideline ?

European legislation

- European AI act
  https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792
  https://artificialintelligenceact.eu/

- Ethics Guidelines for Trustworthy AI
  https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Artificial intelligence act

EU AI Act: first regulation on artificial intelligence
The use of artificial intelligence in the EU is regulated by the AI Act, the world's first comprehensive AI law. Find out how it protects you.

Published: 08-06-2023, Last updated: 19-02-2025

European Parliament

What is it about ?

- Regulation
- Evaluate AI risk
  (**Astronomy and ChatGPT are not considered to be a High Risk**)
- Encourage AI development with sandbox environment
- EU act compliance guidance

Calendar

- 2024/08/01 : begining of regulation.
- 2025/02/02 : Prohibition of AI presenting unacceptable risks.
- 2025/08/02 : establish rules for general purposes

Who is impacted ?

Organisation from Europe or outside Europe that deploy, distribute or provide AI, in particular High-risk AI systems

Consequences

CE

- **High-risk AI** systems **requires CE marking**
- **GPAI** (General Purpose AI) needs transparency subject to a **good practice guideline**

The regulation is not applicable to any activities related to AI research, testing, and development before it is marketed or put into operation.

European Artificial intelligence act

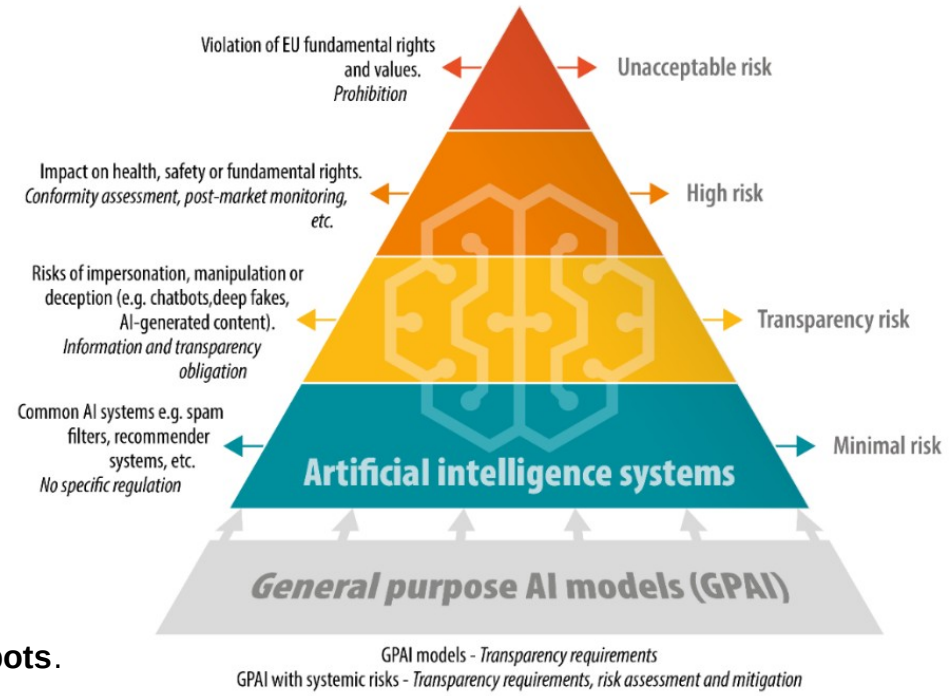
Risks levels

113 articles (+annexes) gathered into 13 main topics
https://artificialintelligenceact.eu/ai-act-explorer/

Prohibited AI practices
- Prohibes subliminal or manipulative techniques
- Protect **Fundamental laws**
- Biometric categorisation or identification in public spaces
- AI evaluating or classifying individuals or groups based on social
- ...

High-risk AI systems
- Critical infrastructure
- **Education** and vocational training
- Access to essential private and public services
- Law enforcement
- Migration and border control management
- Administration of justice and democratic processes
- ...

→ Needs to run a conformity assessment procedure



Transparency risk
- Users must be **made aware that they interact with chatbots**.
- (deployers and providers of) AI **must disclose that the content has been artificially generated**

Minimal risks
(Example SPAM filter) is out of requirements

European Artificial intelligence act

General-purpose AI (GPAI) , what it is ?

https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf

Characteristics:

- Wide range of possible use, both intented and unintented
- Applied to many different tasks without providing fine tuning
- Large scale : memory, power and lot of data
- Often GPAI are building-bloks reuse by other AI systems

Large variety of usage: natural language, used in medicine and healthcare, finance, life sciences, programming ...

GPAI obligations

- GPAI system **transparency** requirements :
  maintain tech. Doc. Including content used for training the AI model

  (Opensource have less obligations)

- Put a policy that **respects the Union copyright law**

- Cooperation with authorities - keeping track of incident (eg: violation of fundamental rights)

- GPAI model providers will be able to **rely on codes of practice**

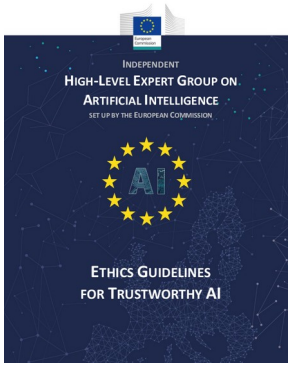
see Article 53: Obligations for Providers of General-Purpose AI Models
https://artificialintelligenceact.eu/article/53/

Ethics Guidelines for Trustworthy AI

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
→ Dedicated for any AI provided in Europe or consume in Europe

The 3 components of a Trustworthy AI

- Lawfull AI: ensuring compliance with all applicable laws and regulations
- Ethical AI : including fundamental rights, ethical principles
- Robust AI : include tech. + societal perpsective robustness
  technical robustness includes the whole AI life-cycle
  human respect and representativeness



The document is made of 3 parts

- Ennounce ethical principles
  repect of human, acknowledge benefits and AI risks
  ”**even with good intentions, unintentional harm can occurred”**
- List 7 requirements
- Accessment list : Questions to help AI provider and covering the 7 requirements

Ethics Guidelines for Trustworthy AI

7 requirements for Trustworthy AI

- **Human agency and oversight** ← technology must be in favor of human (and supervised by human)

  👇 Can you describe the **level of human control** or involvement?
  👇 Which detection and response **mechanisms** did you establish to assess whether something **could go wrong** ?

- **Robustness and safety** ← safe and reproducibility aspects

  👇 Did you consider different types and natures of vulnerabilities, such as **data pollution**, physical infrastructure, **cyber-attacks**?

- **Privacy and data governance** ← data protection, rights, **copyrights** and data integrity

  👇 Did you assess who can access users' data, and under what circumstances?

- **Transparency** ← **Human must be aware of their interaction with AI**

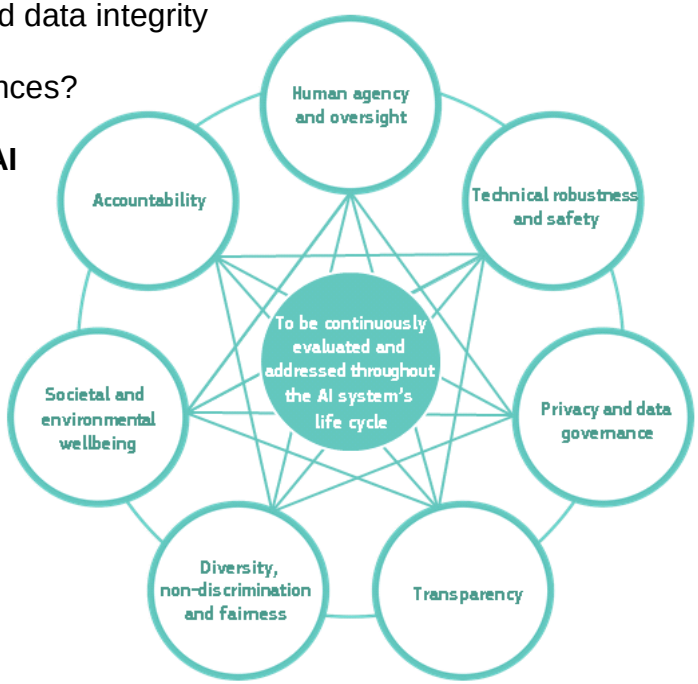
  👇 Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?

- **Diversity, non-discrimination and fairness**
  ← ex: no racial, gender or handicap discrimination

  👇 Did you consider diversity and representativeness of users in the data?

- **Societal and environmental** ← ecology aspect

- **Accountability** ← require responsibility and transparency



Ethics Guidelines for Trustworthy AI

EU AI Act Compliance Checker

https://artificialintelligenceact.eu/assessment/

Example: I'm a deployer of a AI model in Open source that generates images for educational and vocational training.

I'm placing on the market General purpose AI models

Results:

Your results

Providers must submit notification to NCA (National Competent Authority)

If a provider considers their AI system to not pose a significant risk (see Article 6 point 2a) they must register their system in the EU database before that system is placed on the market or put into service (see Article 49 point 2).

They must also document their assessment and provide this documentation to the National Competent Authorities (NCA) upon request (see Article 6 point 4).

If a market surveillance authority finds that the AI system has been misclassified (see Article 80), your system would be subject to the 'high-risk' obligations described in Chapter III Section 2 and you may be subject to fines under Article 99.

My AI system 'output is in used in the EU

AI Literacy obligations

You must take measures to ensure a sufficient level of AI literacy for your staff (and other people dealing with the operation and use of AI systems on their behalf), taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, according to Article 4.

Your results

Providers must submit notification to NCA

If a provider considers their AI system to not pose a significant risk (see Article 6 point 2a) they must register their system in the EU database before that system is placed on the market or put into service (see Article 49 point 2).

They must also document their assessment and provide this documentation to the National Competent Authorities (NCA) upon request (see Article 6 point 4).

If a market surveillance authority finds that the AI system has been misclassified (see Article 80), your system would be subject to the 'high-risk' obligations described in Chapter III Section 2 and you may be subject to fines under Article 99.

General Purpose AI model obligations

You need to follow these obligations for Providers of General Purpose AI (GPAI) models under Article 53. In summary, you must:

- Create and keep technical documentation for the AI model, and make it available to the AI Office upon request.
- Create and keep documentation for providers integrating AI models, balancing transparency and protection of IP.
- Put in place a policy to respect Union copyright law.
- Publish a publicly available summary of AI model training data according to a template provided by the AI Office.

Also, consider whether the GPAI is used as, or a component of, an AI system. If so, obligations on high risk AI systems may apply directly or indirectly under Recital 85.

AI Literacy obligations

You must take measures to ensure a sufficient level of AI literacy for your staff (and other people dealing with the operation and use of AI systems on their behalf), taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, according to Article 4.

Conclusion

What I retain ?

- Respect of fundamental laws
- Recognize the benefit and the risks
  - Risk level, the logo CE
- Different level of risks and obligation
  - Deploying AI may require initiating an endorsement process
- An ethic for trustworthy AI
- Science is not impacted
  - Respects of copyrights (transparency)