EU AiAct - the abridged version

Artificial Intelligence Act
EU Regulation 2024/1689

Paolo Nicoli

paolo-nicoli

pscrapy

⚠️DISCLAIMER I⚠️

I AM NOT A LAWYER

This deck will not save you from needing a qualified lawyer opinion

Don’t @ me

⚠️DISCLAIMER II⚠️

I AM NOT AN ECONOMIST

This deck will not explore any socio-economic effect or impact of the AiAct

If you must @ me let’s do it offline 😛

Outline

A. Context
B. Definitions
    1. AI / GPAI
    2. Risk tiers
    3. Roles
C. Risk Tiers
    1. Prohibited UC
    2. High Risk UC
    3. General-Purpose AI
D. Timeline
E. Appendix

Context

- ➔ First **regulation** for AI, applies automatically to whole EU market
- ➔ Almost anything may be in scope thanks to **very wide** definition of “AI”
- ➔ **Risk-based classification** with stricter requirements for higher risks
- ➔ **Role-based responsibilities** for actors in AI value chain
- ➔ Specific provisions for **General-Purpose AI (Foundation Models)**
- ➔ Some **limitations** and **exceptions** (research, personal use, national security etc)

- ➔ **Already in force**, some provisions are delayed to allow compliance
  On 2025-03-20 the Italian Senate approved Ddl 1146/24, now under review by the Chamber

Definitions

AI System

> *'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that **may** exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*

Almost anything in IT can be in scope:

- Any ML model, any architecture (supervised, unsupervised, reinforcement, …)
- Hardcoded rules, Expert Systems
- Any function or procedure (input, output, machine based)

→ *future-proofing against evolving technology*

***NB:** AI models by themselves are not AI systems: need be part of "system" (input interface, output etc)*

AI Systems are in scope if output is used in the EU, regardless of where they are deployed/served from
→ *GDPR-like extraterritoriality*

AI Systems are subject to risk classification based on the impact of intended use case(*) on EU citizens
→ *focus on protecting citizen rights, health and safety*

(*) and reasonably foreseeable misuse

ref: art3

General-Purpose AI

> *[GPAI is] any AI model [...] that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications*

GPAI are not AI Systems until integrated with other components (interfaces, prompts etc)
→ *GPAI + integration = GPAI System*

GPAI designated as **Systemic Risk** if:

- training compute > 10^{25} FLOPs
- decision by EU Commission based on current SotA, training specs etc

→ *stronger disclosure requirements*

GPAI Models expected to be Systemic Risk:

- Grok-2, Grok-3
- Gemini 1.5 Pro, Veo
- GPT-4, GPT-4 Turbo, GPT-4o, o1, o1-min, o3, o3-mini, Sora
- Llama 3.1-405B, Llama 4-Behemoth
- Claude 3 Opus, Claude 3.5 Sonnet, Claude 3.7 Sonnet
- Mistral Large, Mistral Large 2

ref: art3

Risk Tiers

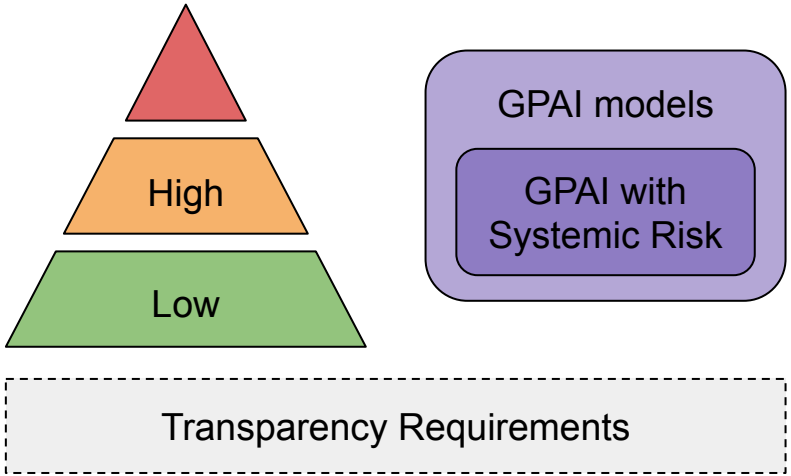
AIS Risk Tiers:

1. **Unacceptable**
   ⇒ 🚫Prohibited🚫
2. **High Risk**
   ⇒ 🚧Heavy regulatory burden🚧
3. **Low/Minimal Risk**
   ⇒ 📜Voluntary Code of Conduct📜

GPAI Models:

Standard:
⇒ 🛂Disclosure obligations(*)🛂

Systemic Risk
⇒🚨Disclosure & Risk Mitigation🚨



(*) Less rigid if model released as Open-Source (i.e. open-weights)

Risk Tiers vs GPAI

Modular obligations

- GPAI model obligations combine with Risk-tier obligations (and transparency)
- both GPAI-model-only and risk-tier only scenario possible

Prohibited AIS

A. no GPAI
B. GPAI
C. Systemic Risk GPAI

High-Risk AIS

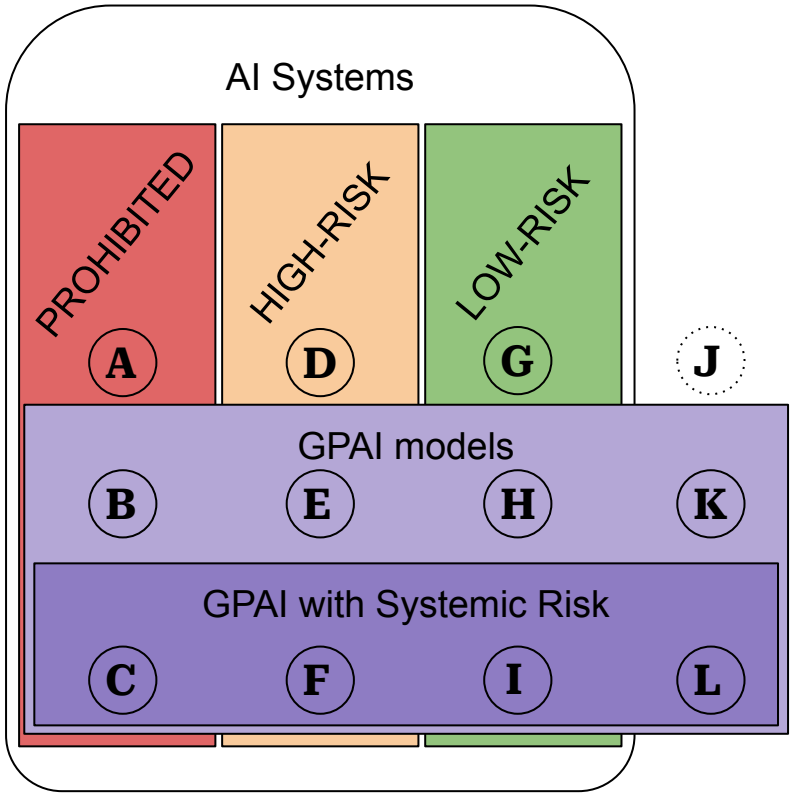
D. no GPAI
E. GPAI
F. Systemic Risk GPAI

Low-Risk AIS

G. no GPAI
H. GPAI
I. Systemic Risk GPAI

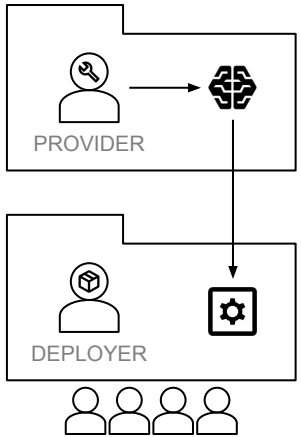
Not an AIS

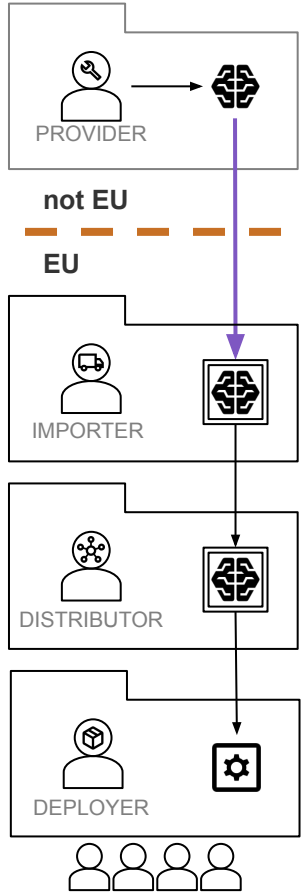
J. no GPAI (out-of-scope)
K. GPAI
L. Systemic Risk GPAI



Roles

- **Provider**
  Anyone(*) that develops a AI system or GPAI model (or that has an AI system or GPAI model developed) and places them on the market under its own name or trademark
- **Deployer**
  Anyone(*) using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity
- **Importer**
  Anyone(*) in the EU that places on the market an AI system whose provider is outside the EU
- **Distributor**
  Anyone(*) in supply chain making an AI system available in EU, other than Provider and Importer





(*) A natural or legal person, public authority, agency or other body

ref: art3

Roles - “Promotion rules”

Any other supply-chain role will be promoted to Provider if they

- *put their name or trademark on a high-risk AI system already placed on the market or put into service, regardless of any contractual arrangements stipulating that the obligations are otherwise allocated*
- *make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in such a way that it remains a high-risk AI system*
- *modify the intended purpose of a non-high-risk AI system, including a GPAI system, in such a way that the AI system concerned becomes a high-risk AI system*

ref: art 25

Scope Limitation

- **Personal Use:**
  [AiAct does not apply to] Deployers who are natural persons using AI systems in the course of a purely personal non-professional activity
- **R&D:**
  [AiAct does not apply to] AI systems or AI models developed and put into service for the sole purpose of scientific research and development
- **NatSec:**
  [AiAct does not apply to] AI systems or AI models used for military, defence or national security purposes

Risk Tiers - Prohibited UC

Prohibited UC

Explicit list of use cases where the risk has been deemed unacceptable

- **Subliminal manipulation**
- **Personal vulnerability exploitation**
- **Social Scoring** (e.g. black mirror)
- **Criminal Profiling** (e.g. minority report)
- **Scraped facial recognition DB** (e.g. clearviewAI)
- **Emotion detection** (workplace or education)
- **Biometric categorisation**
  detecting protected attributes (race, religion, etc)
- **Real-time, remote biometric identification in public spaces for law enforcement (*)**

refs art5

(*) Strict list of allowed exception

Prohibited UC - Examples

Subliminal Manipulation
e.g. Hidden emotion detection capability in smartTV to advertise specific products when most vulnerable

Personal vulnerability exploitation
e.g. Detecting if retail chain customer are recovering alcoholics to targets them with discounted drinks

Social Scoring
e.g. Barring individuals from public grants based on social score computed from parking fines

Criminal Profiling
e.g. Terrorist profiling based on individuals data

Scraped Facial Recognition DB
e.g. ClearviewAI [news]

Emotion detection (workplace/education)
e.g. System to detect if call center employees get angry during calls

Biometric categorization
e.g. System to classify sexual orientation of customers using photograph

Real-time Remote Biometric Identification
e.g. street surveillance system to detect littering violations

refs: guidelines

Prohibited UC - Obligations & Sanctions

Obligations:

DON'T.

Just, don't.

Sanctions

≤ €35M

or

7% global turnover

(whichever highest)

refs art99

Risk Tiers - High-Risk UC

High Risk UC

Two conditions classify a UC as High-Risk:

- **Safety component under EU harmonization legislations**
  - Safety component/product
  - Product requiring conformity assessment

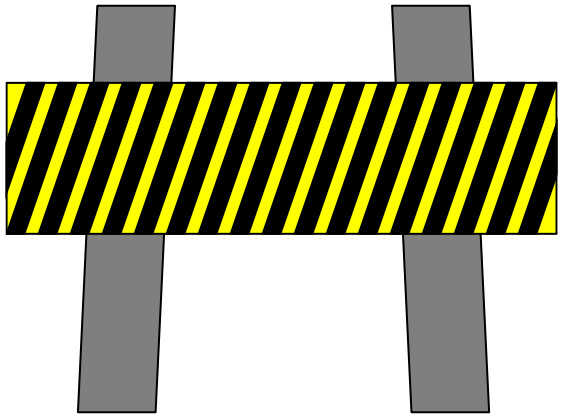
  (e.g. radio/aviation/rail equipment, medical devices, aviation, rails, etc …)
- **Explicit UC listed in AI Act AnnexIII**

  (e.g. biometric identification, employment services, critical infra, etc…)

refs blog, art6, annexI, annexIII

High Risk UC - AnnexIII

- Biometrics
- Emotion Recognition
- Critical Infrastructure
- Education
- Employment
- Access to essential services
- Law enforcement
- Migration, asylum and border control
- Administration of justice
- Democratic process



refs: art6, annexIII

High Risk UC - AnnexIII Examples

- **Biometrics:**
  e.g. System to identify pedestrian walking in front of store on social media to DM ads
- **Emotion Recognition:**
  e.g. System to detect if customers are interested in shop display arrangements
- **Critical Infrastructure:**
  e.g. System to monitor road and bridges structural integrity
- **Education:**
  e.g. system ranking application to training institutions; system to detect student cheating during exams
- **Employment:**
  e.g. system for job applications screening; system for job performance review
- **Access to essential services:**
  e.g. systems for access to healthcare; systems for creditworthiness prediction; systems for risk assessment and pricing of health/life insurance; emergency calls, patient triage
- **Law enforcement**
- **Migration, asylum and border control**
- **Administration of justice**
- **Voting**

refs: art6, annexIII

High Risk UC - AnnexIII Exceptions & Gotchas

Possible to avoid High-Risk classification thanks to 4 exceptions:

- **Narrow procedural task**
  (e.g. OCR, document classifier)
- **Improve downstream from human task**
  (e.g. marketing copy tone-of-voice correction)
- **Detect decision-making pattern**
  (e.g. anomaly detection in essay grading patterns)
- **Preparatory task upstream from human**
  (e.g. document translation)

⇒ **However AnnexIII AI System is always High-Risk if it performs profiling of natural persons**

refs: recital53

High Risk UC - AnnexIII Profiling Definition

> *Profiling is any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*

Profiling definition is imported from three EU laws concerning data protection, all having same definition:

- Regulation (EU) 2016/679, aka General Data Protection Regulation (GDPR)
- Directive (EU) 2016/680, aka Law Enforcement Directive (LED)
- Regulation (EU) 2018/1725 aka Data Protection Regulation for EU institutions (EUDPR)

refs: link

High Risk UC - Provider Obligations

Art.16 System Requirements: ensure AI system is compliant with AiAct requirement → *see later*

Art.17 Quality Management System: complex of policies, procedures & technical specifications to ensure compliance → *big undertaking*

A.18-19 Document & Logs Keeping: keep all compliance documents available to authorities for 10 years, all logs for at least 6 months

A.20 Corrective Actions and Disclosure: fix issues ASAP and notify relevant parties (authorities, importers, distributors, deployers)

A.21 Cooperation with Competent Authorities: provide documentation, proofs of compliance, access to logs when required

A.22 Authorised Representatives: providers from outside EU that release High-Risk systems in EU must appoint representatives

A.27 Fundamental Rights Impact Assessment: categorize intended audience into groups likely to be affected and assess & mitigate possible risks

A.49 EU Database for High-Risk AI Systems: register AI System into central EU Database

High Risk UC - System Requirements

Art.9 Risk Management System: identify, address and monitor risks to health, safety and rights for users

Art.10 Data and Data Governance: evaluate and correct data for bias, document all processing choices, use representative datasets.

Art.11 Technical Documentation: to be written before release and demonstrating compliance.
→ *use extensive model card template*

Art.12 Record Keeping: allow lifetime logs, risk events tagging, explicit reference for biometric matches

Art.13 Transparency to Deployers: documentation with clear deployment instructions, risks & failure modes, metrics, minimum hardware, etc

Art.14 Human Oversight: allow explainability, clear usage warnings for risk scenarios, possibility of manual overrides

Art.15 Accuracy, Robustness, Cybersecurity
Follow technical guidelines (TBD) to ensure reliability, detect anomalies and monitor performance

High Risk UC - Other Roles Obligations

Deployers:

- Follow Provider instruction
- Human oversight
- Monitoring and incidents report *(providers & authorities)*
- Inform subjects of HighRisk processing
- Fundamental Rights Impact Assessment *(essential service access UC)*

Importers:

- Ensure Provider followed AiAct obligations
- Do not import until AiAct conformity
- Indicate reachable address
- Do not compromise AiAct conformity

Distributors:

- Ensure Provider followed AiAct obligations
- Do not distribute until AiAct conformity
- Do not compromise AiAct conformity
- Recall or ensure corrective actions

High Risk UC - Obligations & Sanctions

Obligations:

- **Providers [art 16]**
- **Deployers [art 26]**
- **Importers [art 23]**
- **Distributors [art 24]**

Sanctions

- **Compliance issues**

≤ €15M

or

3% global turnover

(whichever highest)

- **Reporting issues**

≤ €7.5M

or

1% global turnover

(whichever highest)

Risk Tiers - GPAI

General-Purpose AI - Obligations & Sanctions

GPAI Provider Obligations:

a. Regulatory documentation (annex XI)
   → *includes training details & energy consumption*
b. Documentation for downstream integrations (annex XII)
c. Ensure Copyright compliance
d. Training data summary (template TBD)

Open Source Exception:

GPAI models not classified as systemic risk and released under open-source licence, i.e. open weights & arch info, are exempt from (a) and (b)

GPAI Systemic Risk Provider Obligations:

1. Model evaluation & adversarial testing
2. Assess & mitigate systemic risk in EU
3. Track, document & report incidents
4. Ensure cybersecurity safeguards

Sanctions

→ same for Systemic Risk and not

≤ €15M

or

3% global turnover

(whichever highest)

Risk Tiers - Low Risk & Transparency

Transparency & Low Risk

”Limited” Risk (*)

AI systems intended to interact directly with natural persons

Obligations:

- Inform end-user that they are interacting with an AI system
- GPAI systems outputs shall be detectable and marked in machine-readable form
- End-users shall be informed of artificially generated/edited contents

Sanctions

≤ €15M

or

3% global turnover

(whichever highest)

Low Risk:

anything not Prohibited, HighRisk or GPAI

→ voluntary code-of-conduct

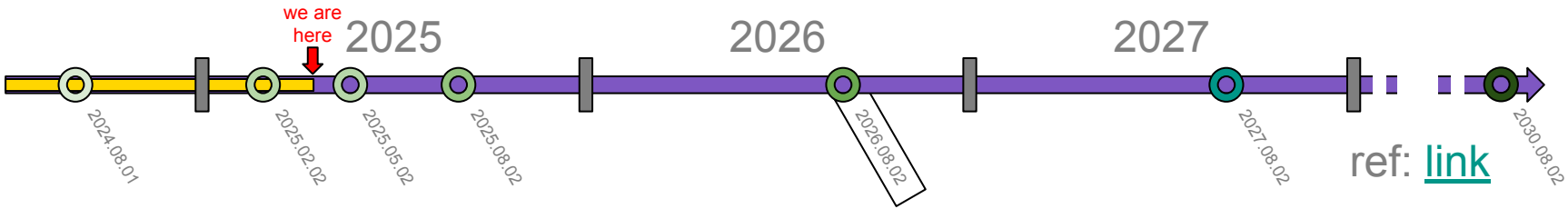
→ no penalties

(*) Not a term defined in AiAct but still widely used in press

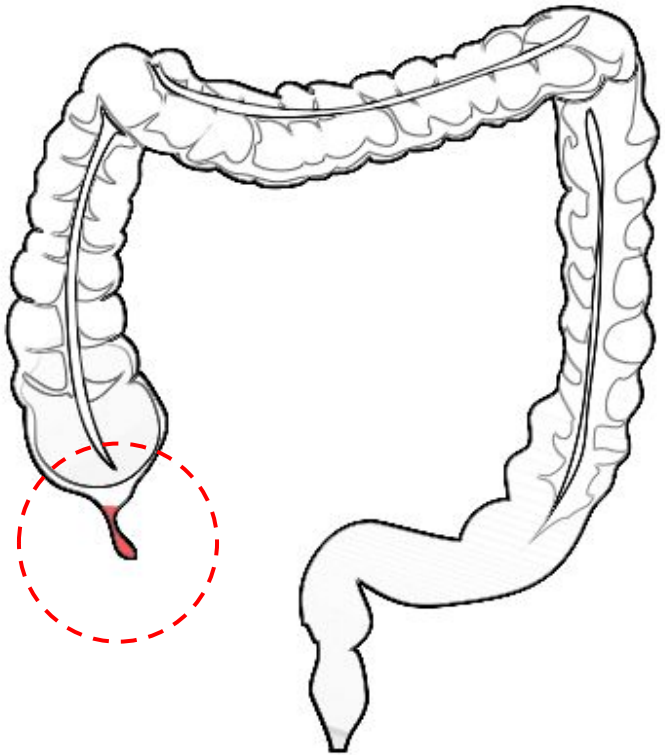
Timeline

Staggered application:

- **1st Aug 2024**: Entry into force
- **2nd Feb 2025**: Prohibitions on certain AI systems start to apply, AI literacy obligations
- **2nd May 2025**: Code of Practices published
- **2nd Aug 2025**: Notified bodies, Governance, Confidentiality, Penalties and GPAI models provisions start to apply
- **2nd Aug 2026**: remainder of the AI Act starts to apply except Article 6(1)
  → HighRisk for AI safety components (e.g. toys, medical devices, etc)
- **2nd Aug 2027**: Article 6(1) starts to apply
- **2nd Aug 2030:** Legacy High-Risk Systems must be compliant



Appendix



Miscellanea

- **AI Literacy:**
  providers & deployers shall ensure tech expertise among staff
- **AI Board, AI Office & scientific panels:**
  multiple organs created within EU institutions to supervise and enforce AI Act
- **High Risk AI Systems Database:**
  central database for tracking High-Risk systems active in EU market, publicly accessible
- **Redress procedures:**
  - *AiAct infringement → Complaint with Market Surveillance Authorities*
  - *High-Risk system decision explanation → affected subject can ask deployer*
- **Regulatory Sandboxes:**
  Controlled environments for testing new products without needing up-front compliance

Biometric Stuff

- **biometric data**: personal data from processing relating to the physical, physiological or behavioural characteristics of a natural person
  e.g. facial images or dactyloscopic data
- **biometric identification:** compare biometric data against DB
  e.g. identify suspect face
- **biometric verification:** 1-to-1 compare of biometric data against previous record
  e.g. authentication or KYC

- **biometric categorisation system:**
  AI system using biometric data as feature for classification
  → *prohibited when target is protected data*
  → *high risk when features are protected data*
- **remote biometric identification:**
  AI systems performing bio-id without subject involvement, typically at distance
  → *prohibited when real-time in public spaces(*)*
  → *high risk otherwise*

ref: art3

(*) Unless law enforcement

Italian Law - Ddl 1146/24

On 20th March, 2025 the Italian Senate approved Ddl 1146/24 to begin the process of ratifying the AiAct

Key provisions

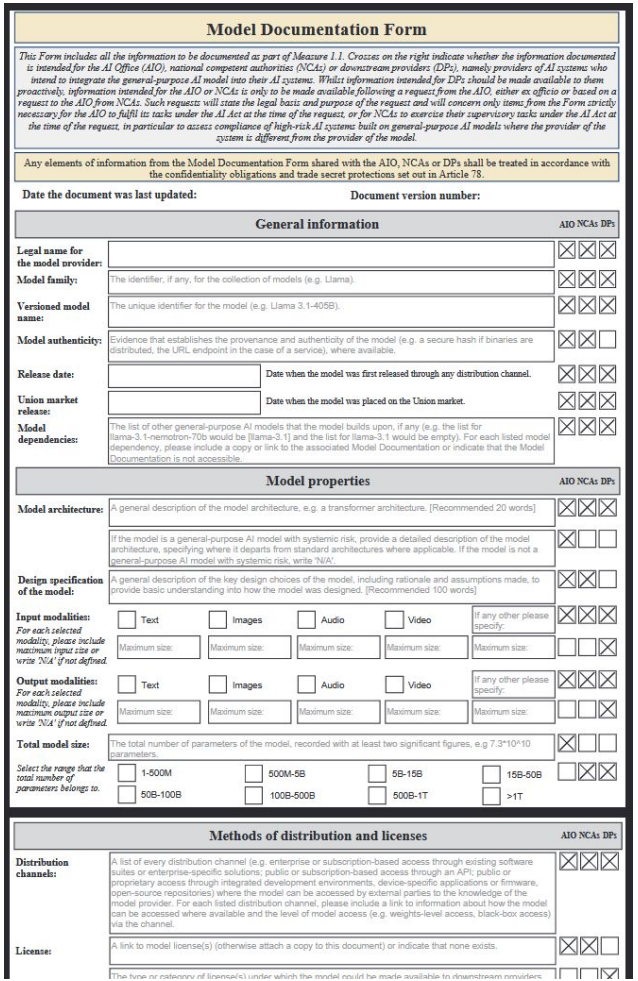
- AI used by PA shall be deployed on servers located in Italy
- AgID is national notification authority
- ACN is national market surveillance authority
- AI in health restricted to “support” role, doctor responsible for decisions
- Use of AI added as aggravating circumstance to penal code for some crimes

ref: link

GPAI - Code of Practice

GPAI Code of Practice is currently being developed as a collaborative effort between AI Office, Member States representatives and business stakeholders, with a tentative release date of **May 2025**

Current draft: link
Timeline: link



Model Documentation Form

This Form includes all the information to be documented as part of Measure 1.1. Crosses on the right indicate whether the information documented is intended for the AI Office (AIO), national competent authorities (NCAs) or downstream providers (DPs), namely providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Whilst information intended for DPs should be made available to them proactively, information intended for the AIO or NCAs is only to be made available following a request from the AIO, either ex officio or based on a request to the AIO from NCAs. Such requests will state the legal basis and purpose of the request and will concern only items from the Form strictly necessary for the AIO to fulfil its tasks under the AI Act at the time of the request, or for NCAs to exercise their supervisory tasks under the AI Act at the time of the request, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model.

Any elements of information from the Model Documentation Form shared with the AIO, NCAs or DPs shall be treated in accordance with the confidentiality obligations and trade secret protections set out in Article 78.

Date the document was last updated: **Document version number:**

General information	AIO	NCAs	DPs
Legal name for the model provider:			
Model family: The identifier, if any, for the collection of models (e.g. Llama).			
Versioned model name: The unique identifier for the model (e.g. Llama 3.1-405B).			
Model authenticity: Evidence that establishes the provenance and authenticity of the model (e.g. a secure hash if binaries are distributed, the URL endpoint in the case of a service), where available.			
Release date: Date when the model was first released through any distribution channel.			
Union market release: Date when the model was placed on the Union market.			
Model dependencies: The list of other general-purpose AI models that the model builds upon, if any (e.g. the list for llama-3.1-nemotron-70b would be [llama-3.1] and the list for llama-3.1 would be empty). For each listed model dependency, please include a copy or link to the associated Model Documentation or indicate that the Model Documentation is not accessible.			

Model properties	AIO	NCAs	DPs
Model architecture: A general description of the model architecture, e.g. a transformer architecture. [Recommended 20 words]			
If the model is a general-purpose AI model with systemic risk, provide a detailed description of the model architecture, specifying where it departs from standard architectures where applicable. If the model is not a general-purpose AI model with systemic risk, write 'N/A'.			
Design specification of the model: A general description of the key design choices of the model, including rationale and assumptions made, to provide basic understanding into how the model was designed. [Recommended 100 words]			
Input modalities: ☐ Text ☐ Images ☐ Audio ☐ Video If any other please specify: *For each selected modality, please include maximum input size or write 'NA' if not defined.* Maximum size: Maximum size: Maximum size: Maximum size: Maximum size:			
Output modalities: ☐ Text ☐ Images ☐ Audio ☐ Video If any other please specify: *For each selected modality, please include maximum output size or write 'NA' if not defined.* Maximum size: Maximum size: Maximum size: Maximum size: Maximum size:			
Total model size: The total number of parameters of the model, recorded with at least two significant figures, e.g 7.3*10^10 parameters.			
Select the range that the total number of parameters belongs to. ☐ 1-500M ☐ 500M-5B ☐ 5B-15B ☐ 15B-50B ☐ 50B-100B ☐ 100B-500B ☐ 500B-1T ☐ >1T			

Methods of distribution and licenses	AIO	NCAs	DPs
Distribution channels: A list of every distribution channel (e.g. enterprise or subscription-based access through existing software suites or enterprise-specific solutions; public or subscription-based access through an API; public or proprietary access through integrated development environments, device-specific applications or firmware, open-source repositories) where the model can be accessed by external parties to the knowledge of the model provider. For each listed distribution channel, please include a link to information about how the model can be accessed where available and the level of model access (e.g. weights-level access, black-box access) via the channel.			
License: A link to model license(s) (otherwise attach a copy to this document) or indicate that none exists.			
The type or category of license(s) under which the model could be made available to downstream providers			

Sources & Possible Biases

Most of this deck was first sourced through https://artificialintelligenceact.eu/
The website is maintained by the Future of Life Institute (FLI), a non profit working on international policy.

Its mission statement is "*Steering transformative technology towards benefitting life and away from extreme large-scale risks*" and among its funders is Vitalik Buterin, co-founder of Ethereum.

Thank You!

Paolo Nicoli

paolo-nicoli

pscrapy