The EU AI Act: Setting Global Standards for Artificial Intelligence

Dr Marcin Betkier

AI and Society Seminar Series

Wellington 01/03/2024

Agenda

- Introduction
- Scope and extent of the regulation
- Prohibited systems
- High-risk systems
- General-purpose AI models
- Governance, enforcement & entry into force
- Options for New Zealand

Why a global standard?

- First comprehensive regulation
  - “promoting the uptake of human centric and trustworthy AI”
  - setting the “rules for placing on the market, putting into service and use of AI systems” and
  - general-purpose AI models (GPAI models)
- “Brussels effect” (direct application of the European law and indirect standard setting)
- But, the Council of Europe’s AI Treaty will also be coming soon

Scope (subjects and jurisdictions)

In essence, **a product safety regulation**, but it protects not only health and safety, but also fundamental rights, democracy, rule of law and environment.

It covers:

- **Providers** (those who develop or have an AI system/ GPAI model) placing on the market or putting into service AI systems or placing on the market GPAI models in the Union (located anywhere).
- **Deployers** (those using an AI system under their own authority) of AI systems located in the Union.
- Providers and deployers of AI systems that have their place of establishment or who are located in a third country, where **the output** produced by the system **is used in the Union**.
- **Product manufacturers** placing on the market or putting into service an AI system together with their product under their own name/trademark.
- **Importers and distributors** of AI systems (into the Union).
- **Authorised representatives of providers**, which are not established in the Union.

Exceptions

- Any **research, testing, development** activity prior to placing system/model on the market or put into service.
- AI systems /models / output, developed and put into service **for the sole purpose of scientific research and development**.
- AI systems for **military, defence or national security purposes** (and other areas where EU law does not apply).
- Deployers who are **natural persons** using AI systems in the course of a purely **personal non-professional activity**.
- AI systems released under **free and open source licences,** unless they are
  - placed on the market or put into service as high-risk AI systems, or
  - an AI system that falls under Title II (prohibited practices) and IV (transparency obligations).

AI system definition

- an AI system - “a machine-based system 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”,
- Follows the OECD’s latest definition.
- Broad and functional (technologically neutral).

Regulation categories

- Prohibited systems
- High-risk systems
- Limited risk (transparency obligations only)
- General-purpose AI models
- General-purpose AI models with systemic risk

Prohibited systems

- **Harmful subliminal techniques** or purposefully **manipulative or deceptive techniques** to materially distort behaviour;
- **Exploiting vulnerabilities** of a person or group due to specific characteristics (likely) leading to significant harm;
- **Biometric categorisation** systems that individually categorise a person based on sensitive information (except for labelling/filtering biometric datasets in the area of law enforcement);
- **Social scoring** systems;
- **Real-time remote biometric identification** systems in the public spaces for the purpose of law enforcement (except for searching for victims, prevention of threat to life/terrorist attack, targeting suspects of serious criminal offences – all subject to additional safeguards – assessments, warrants and reporting requirements);
- **Predictive policing** based solely on profiling or personality traits, except when supporting human assessments based on objective, verifiable facts linked to criminality;
- Systems creating **facial recognition databases** based on untargeted scraping (Internet, CCTV); and
- **Inferring emotions in workplaces or educational institutions**, except for medical or safety reasons.

High-risk systems – what are they?

- AI systems being **a product (or its safety component)** covered by certain EU legislation that mandates **a third-party conformity assessment** (Annex II, for example, machinery, toys, lifts, medical devices).
- AI systems used in one of the areas listed in Annex III are presumed as high-risk (e.g. **part of biometrics, critical infrastructure, education, employment, law enforcement, migration, justice/elections**)
- Derogations (do not apply to profiling of natural persons):
  - a narrow procedural task, improving the result of previously completed human activity, detecting decision-making patterns or deviations from those, preparatory tasks
  - Have to be assessed and notified by the provider.

High-risk systems – requirements

- Risk management system (Art 9)
- Obligations on training, validation and testing data sets (e.g. data governance and management practices, representativeness, bias detection and correction) (Art 10)
- Technical documentation showing that the AI system complies with the requirements (Art 11)
- Record-keeping (Art 12)
- Transparency and provision of information to deployers (instructions for use, Art 13)
- Effective human oversight (Art 14)
- Consistent accuracy, robustness and cybersecurity (Art 15)

High-risk systems – obligations on providers, importers and distributors

- **Providers**:
  - Ensure compliance with obligations, conformity assessment, registration,
  - Quality management system that implements and ensures all those system-level obligations starting from design and development phase to “post-market” monitoring
  - Documentation keeping
  - Automatic event logging, traceability, monitoring, corrective actions and accountability.
  - Providers outside the Union appoint authorised representatives to act for them and perform those tasks
- **Importers and distributors** have to verify conformity assessment
- All subjects down the value chain may be treated as providers if they substantially modify the system or re-purpose it to a high-risk area

High-risk systems – obligations on deployers

- Use systems according to the instructions for use (from providers)
- Ensure adequate human oversight and, when appropriate, the quality input data
- Monitor operation of the system, keep the logs
- Additional obligations on particular types of deployers (financial institutions, employers, public authorities, law enforcement authorities)
- Fundamental rights impact assessment for systems used by public authorities, provision of public services and credit scoring (Article 29a)

Conformity assessment

- **Notifying authorities** – organise the assessment, designation, notification and monitoring of conformity assessment bodies;
- **Conformity assessment bodies** - perform third-party conformity assessment activities, including testing, certification and inspection;
- **Notified bodies** – conformity assessment bodies that are “notified” (approved). They have to conform with requirements as to organisation, personnel, procedures, etc.
- **The Commission** provides for standardisation
- AI systems are also registered in a **pan-EU database**

Transparency obligations for providers and deployers of certain AI systems

- Lightweight obligations outside of high-risk area on providers, deployers and users(!) of some AI systems and GPAI models
- Transparency obligations (Art 52) – details at the next lecture
  - AI systems intended to directly interact with natural persons
  - AI systems generating synthetic audio, image, video or text content
  - Emotion recognition systems/ biometric categorisation systems
  - Generating deep fakes

General Purpose AI models

- GPAI model means an AI model, including when trained with a large amount of data using self-supervision at scale, that **displays significant generality** and is **capable to competently perform 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**.
- This does not cover AI models that are used before release on the market for research, development and prototyping activities;

General Purpose AI models - types

- GPAI model
- GPAI model with systemic risk:
  - It has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks; or
  - The EU Commission decides so in particular procedure;
  - Systemic risk is presumed when the cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25.
  - Criteria for the Commission (related to scale) are in Annex IXc
  - Secondary regulation to follow (make it more precise and adapt)

General Purpose AI models - obligations

- GPAI model providers shall:
  - Appoint an authorised representative (if are outside the Union)
  - Put in place policy to comply with the EU copyright law
  - Publicly share summaries of the content used to train their models
  - Write technical documentation for downstream AI systems about the capabilities and limitations of the model and its parameters
  - The last two do not apply to free, open licence and publicly available models
- GPAI model with systemic risk:
  - Perform model evaluation to identify systemic risk,
  - Assess and mitigate possible systemic risks
  - Keep track of, document, and report information about serious incidents and possible corrective measures
  - Ensure an adequate level of cybersecurity protection.
  - No open-source exceptions
  - Compliance with these requirements can be achieved through codes of practice

Governance and enforcement

- AI Office
- European Artificial Intelligence Board
- Advisory bodies:
  - Advisory forum - a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society and academia
  - Scientific panel of independent experts
- National authorities
- Market monitoring and surveillance by surveillance authorities
- The Commission & AI Office have powers to request documentations, provide evaluations, request measures to comply, impose fines
- Fines - tiered system:
  - 35m EUR / 7 % of total worldwide annual turnover (banned uses)
  - 15m EUR / 3% (obligations)
  - 7.5m EUR/ 1.5% (information requirements)

Entry into force

- 24 months with regard to most of the Regulation
- 6 months for prohibitions
- 12 months for provisions concerning notifying authorities and notified bodies, governance, GPAI models, confidentiality and penalties
- 36 months for high-risk AI systems covered by Annex II.
- For the existing systems to be brought to compliance:
  - 4 years for AI systems already placed on the market or put into service by public authorities
  - 2 years (3 years in total) for GPAI models placed on the market before entry into force application of the provisions related to GPAI models

Innovation measures

- At least one regulatory sandbox in each Member State
- controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service
- Such regulatory sandboxes may include testing in real world conditions supervised in the sandbox
- Repurposing personal data for training allowed (!)
- Testing high-risk systems in real world outside sandboxes

New Zealand

- DIA works on government framework related to AI
- How to regulate (and what we can potentially learn/borrow from the AI Act)
  - We know where is the line of political compromise (and arguments used)
  - Risk approach – the results/output of AI systems is prone to errors and that poses a higher risk of peculiar type
  - They require ongoing monitoring tools & human supervision
  - If we want to deploy AI systems in high-risk applications quickly, we need to put assurance systems in place (a lighter version of AI Act)
  - We should look at the no-go/restricted areas
  - We also have copyright and privacy problems

- Regulatory options:
  - Wait and see (how those regulation will land)
  - Influencing global standards (which way?)
  - Regulate for quicker adoption of AI systems (lightweight regulation)

Thank you!

marcin@betkier.com