Generative AI and the AI Act

Yordanka Ivanova,
Legal and policy officer
European Commission, CNECT A2
**Generative Artificial Intelligence**

**What is generative AI?**
Generative AI refers to the use of AI to create new content, like text, images, music, audio, and videos.

**How does it work?**
Generative AI models are trained on very large datasets from which they learn the patterns and structure and then generate new synthetic content that has similar characteristics.

**Examples**
- Open AI’s GPT- 4 – ChatGPT consumer interface
- Open AI’s – DALL-E image generator
- META’s Llama
- Google’s PALM-E 2 - Bard chatbot

- DeepMind's Chinchilla
- Anthropic’s Claude -2
- Microsoft’s Copilot
- Stability AI – Stable Diffusion-2 image generator
Opportunities & risks

**Opportunities**
- Scale and speed
- Low cost entry
- Domain adaptation and fine-tuning
- General purpose/multi-tasks
- Innovation and creativity
- Efficiency and economic growth
- Potential to solve global challenges (e.g., health, climate)

**Risks and challenges**
- Black box
- Unreliability (‘hallucinations’)
- Potential for misuse
- New systemic risks (e.g. disinformation, bias, security and safety)
- Unexpected potentially dangerous capabilities
- Impact on labour market
- Negative environmental impact
EU approach to (generative) AI

EXCELLENCE AND TRUST

- 2021 Proposal for a legal framework on AI
- 2021 Coordinated Plan on AI (review)
Coordinated plan on AI

Support excellence and development of AI in EU, including large powerful models

AI testing and experimentation facilities

European Commission

AI on demand platform

Network of Excellence in AI

Data spaces

EuroHPC Joint Undertaking
Proposal for an AI Regulation

Unacceptable risk
- e.g. social scoring, harmful manipulation

High risk
- e.g. recruitment, medical devices, employment

‘Transparency’ risk
- (chatbots, deep fakes)

Minimal or no risk

*Not mutually exclusive

Generative AI prohibited if used for prohibited practice

If generative AI is used for high-risk purpose, the system would be high-risk and permitted subject to compliance with AI requirements and ex-ante conformity assessment

People should be informed they are interacting with AI (chatbots) or deep fakes
State-of-play of the AI Act

1. European Commission
   AI Act Proposal: 21 Apr 2021

2. Council
   General approach: 6 Dec 2022

3. Parliament
   Plenary vote: 14 June 2023

4. Ongoing trilogues
Council position [Dec 2022]

**General Purpose AI system (GPAI)** is defined as an AI system that - irrespective of how it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems;

- Explicit obligation for GPAI providers to **provide information and cooperate with downstream providers** of high-risk AI systems

- In addition, GPAI would be subjected to **adapted requirements and similar procedure as for high-risk AI** (e.g. conformity assessment, quality management, registration, post-market monitoring) **with 2 exceptions:**
  - GPAI providers who have explicitly excluded GPAI from being used for all high-risk use cases
  - GPAI providers that are small and medium enterprises

- **COM to adopt two implementing acts:**
  - to determine the exact information to be given to the downstream providers and
  - to adapt and specify the high-risk requirements (data quality, technical documentation, transparency etc.) for GPAI, taking into account the value chain, technical feasibility and other relevant factors
European Parliament position [June 2023]

**General Purpose AI system (GPAI)** is defined as an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed.

**Foundation model** is defined as AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks;

---

3-pillar approach for GPAI/ foundation models/generative AI including:

i) **Obligation for upstream providers of GPAIs to provide information and cooperate with downstream providers of high-risk applications to enable compliance with the AI Act;**

ii) **For foundation models:** new requirements in the development stage, including appropriate design, testing and analysis, documentation of non-mitigable risks; model evaluation to assess cybersecurity, safety robustness, explainability etc., good data governance and use of environmental standards; technical documentation, registration and quality management

iii) **For generative AI:** three additional obligations to:

- ensure transparency and mark AI-generated content,
- put safeguards against generation of illegal content
- document and provide detailed summary of the training data protected under copyright law.
Key issues and next steps

Key issues for the Trilogue on General purpose AI / foundation models / generative AI

- Where regulate? (model vs. system)
- How to reconcile with risk-based approach? (e.g. tiered approach)
- What requirements and obligations?
- What governance and enforcement?

Next steps

- Aim to **conclude Trilogues** by end-2023 for adoption this legislature
- **Prepare implementation** (AI standards, guidelines, regulatory sandboxes)
- **AI Pact** - voluntary commitments to apply the AI Act ahead of the legal deadline
- **International outreach** – G7 Hiroshima process, US-EU TTC, UN digital compass and follow up to UNESCO recommendation on AI, Council of Europe (CAI) convention on AI, UK AI safety Summit etc.
Thank you