10 May 2024

GenAI: Compliance and Governance Challenges from a Cybersecurity and Privacy Perspective

Courtney Barton, VP Senior Counsel and Global Privacy and Security Leader
Marriott International

Wim Nauwelaerts, Partner
Alston & Bird LLP

Jari Salomaa, CEO
Valo Security

Privacy+ Security Forum

What is Generative AI (GenAI)?

= **Artificial Intelligence (AI) technology** that:

- Can take a prompt or query from a user (the "**input**") and respond to it with a type of "**output**" that resembles/is equivalent to what a human could create
- Does **not rely on a database of preformulated answers** - the output is based on the general characteristics that the technology has "observed" in the training data, without (necessarily) duplicating the training data
- Can be used to **produce content** such as text, images, or audio

Examples of GenAI Use Cases (1)

Privacy+ Security Forum

- **Coding assistant:** GenAI tools can potentially save hours of coding time per job (valuable for architects, developers)
- **Content creation:** Produce pitches for sales or marketing departments, mock-up versions/features of products or services
- **Document drafting:** Generate policies for HR, legal documents, protocols, instruction manuals, email responses
- **Customer support:** Interact directly with customers in a way (almost) indistinguishable from a human; or (for example) to review and summarize a customer's history

Examples of GenAI Use Cases (2)

Privacy+ Security Forum

- **R&D acceleration:** Summarize, aggregate, reformulate, analyze, and extract insights from (scientific) literature
- **Presenting R&D results:** Elaborate bullet-pointed conclusions into more substantial explanations for publication; make information more understandable
- **Optimizing processes/critical thinking/pattern spotting:** *"Please suggest ways to optimize this process…"* ; *"Please compare…"*
- **Specialized applications**: *e.g.*, drug discovery; medical devices design
- **Idea generation:** *"Generate 10 blog post titles on the following topic…"*

GenAI Governance Challenges

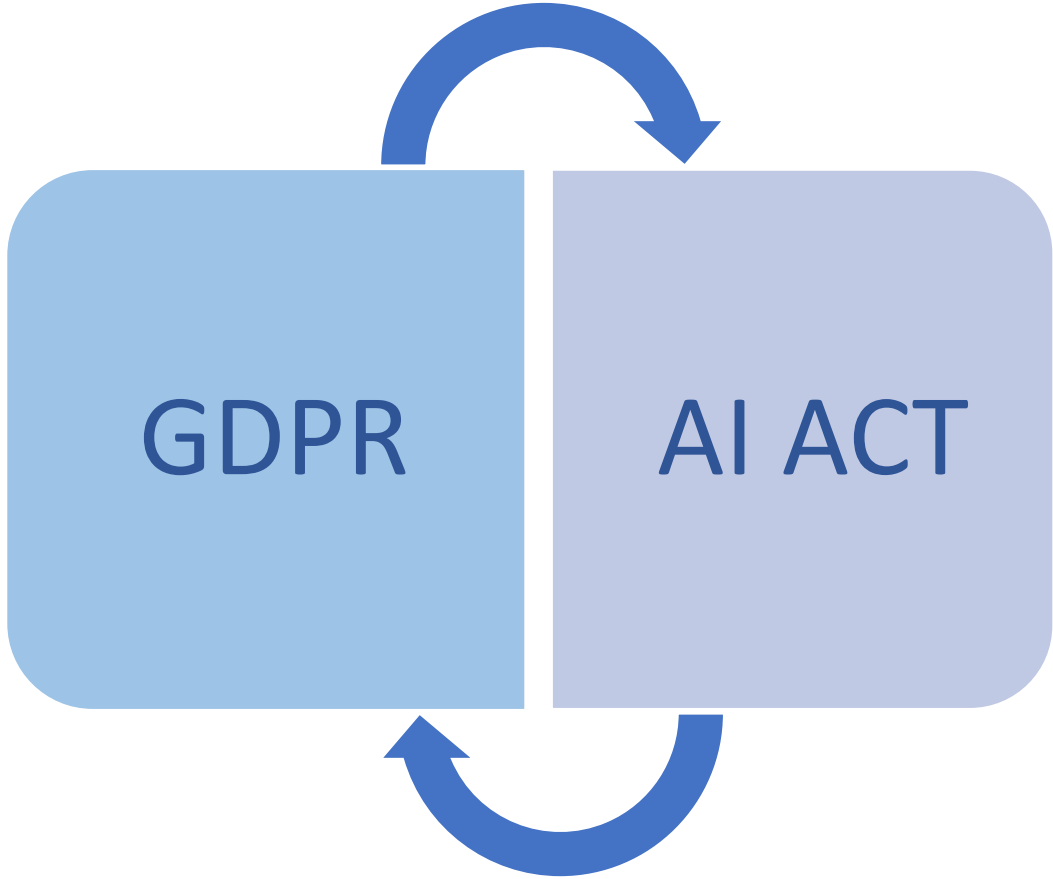
- **NIST AI 600-1** (April 2024 Draft) = resource for GenAI to the AI Risk Management Framework (AI 2 RMF), pursuant to President Biden's Executive Order (EO) 14110 on Safe, Secure, and Trustworthy Artificial Intelligence
- Serves as both a **use-case** and **cross-sectoral profile** of the AI RMF, to assist organizations in deciding how they might best manage AI risk
- NIST AI 600-1 outlines the following **risks associated with GenAI**:

1) Eased access to chemical, biological, radiological, or nuclear (CBRN) weapons
2) Confabulation: Producing inaccurate content
3) Dangerous Recommendations: Facilitating harmful content and actions
4) **Data Privacy**
5) Environmental: Resource-heavy model training impacts.
6) Human-AI Configuration: Problems in human-AI interactions.
7) Information Integrity
8) **Information Security**
9) Intellectual Property
10) Obscene Content
11) Toxicity, Bias, and Homogenization
12) Value Chain and Component Integration

NON-EXHAUSTIVE LIST

GenAI Compliance Challenges in the EU

Privacy+ Security Forum



GenAI & the EU GDPR (1)

Privacy+ Security Forum

- **Compliance** with the EU GDPR may be required in the case of:
  - Providers of GenAI tools that process personal data for training purposes
  - Users of GenAI tools that include personal data in their prompts/input
- **Key data protection principles** under the EU GDPR:
  - ***Lawfulness, fairness and transparency***
    - Legal bases for processing (Consent? Legitimate interest? Additional restrictions on sensitive data processing - how do they apply in the context of GenAI)?
    - Notice and information to data subjects - unique challenge for GenAI models given the volume of data
  - ***Purpose limitation*** - "*To improve the model only*"?
  - ***Data minimization*** - Reduce inputs of personal data via data filtering and use of synthetic training data; ban on inputting personal data?
  - ***Accuracy*** - Outputs of GenAI tools are not always intended as factual information…
  - ***Storage limitation*** - The EU GDPR does not allow personal data to be stored indefinitely…
  - ***Integrity and confidentiality (security)*** - How to deploy GenAI securely?
  - ***Accountability*** - Is the GenAI Provider able to demonstrate its compliance with the EU GDPR?

GenAI & the EU GDPR (2)

Privacy+ Security Forum

- Requirement to carry out a **Data Protection Impact Assessment (DPIA)**?
- **Data subject rights** - Can data subjects effectively exercise their EU GDPR rights *e.g.*, the right to have personal data deleted from the GenAI model?
- Right not to be subject to **a decision based solely on automated processing of personal data** (including profiling) - which produces legal or similarly significant effects - triggers additional considerations (CJEU *SCHUFA* case, C-634/21)
- **Controller/processor** roles of GenAI Providers and Users under the EU GDPR
- **Restricted transfers** of personal data outside of Europe?

GenAI & the AI Act (1)

- Based on European Parliament's text of **16 April 2024** (Corrigendum)
- **Risk-based** rules relating to **AI Systems (AIS)** and **General-Purpose AI Models (GPAIM)**
- New rules apply to **Providers placing GPAIM the EU market**, irrespective of whether those Providers are established or located within the EU or in a third country
- **Recital 99 AI Act**: *"**Large generative AI models** are a typical example for a GPAIM, given that they allow for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks*

GenAI & the AI Act (2)

- **GPAIM** = an **AI model** that:
  - Displays **significant generality**
  - Is capable of competently performing a **wide range of distinct tasks** regardless of the way the model is placed on the market
  - Can be integrated into a **variety of downstream systems or applications**

⇒except AI models that are used for R&D / prototyping activities before they are placed on the market

- GPAIM **do not** constitute AIS on their own

- GPAIM **are typically integrated** into and form part of AIS to make them accessible by individual end-users

⇒Through the addition of further components *e.g.*, a user interface

GenAI & the AI Act (3)

Privacy+ Security Forum

- When a GPAIM is integrated into or forms part of an AI System (**AIS**), this system should be considered a General-Purpose AI System (**GPAIS**), if it has the capability to serve a variety of purposes
- GPAIS can be used directly, or may be integrated into **other AIS**
- GPAIS may be used as High-Risk AI Systems (**HRAIS**) by themselves or be components of other HRAIS
- **AIS** = a machine-based system designed to operate with varying levels of autonomy, and that:
  - May exhibit adaptiveness after deployment
  - Infers, from the input it receives, how to generate outputs (such as predictions, content, recommendations, or decisions),
  - Can influence physical or virtual environments

GenAI & the AI Act (4)

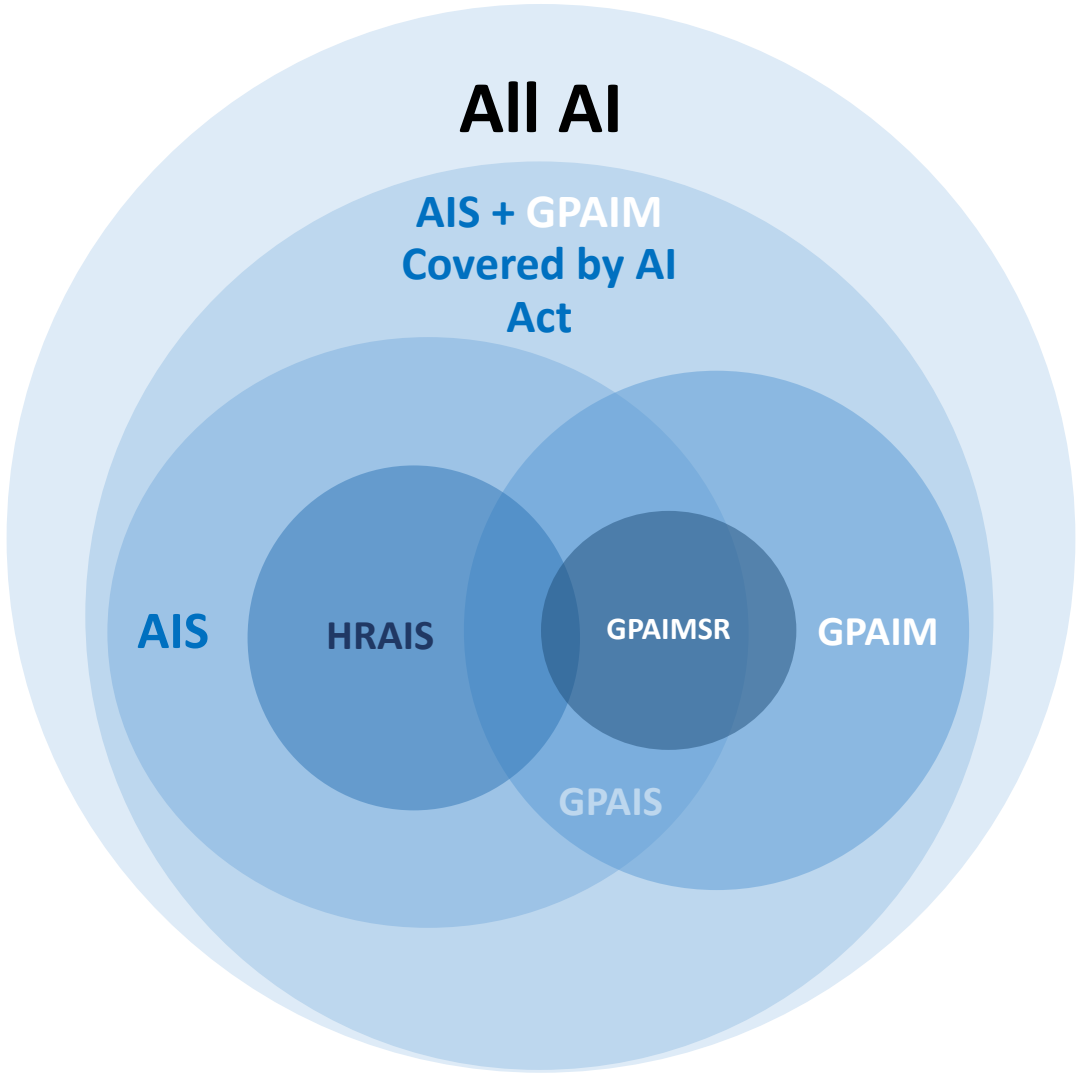
Privacy+ Security Forum

- Additional rules & restrictions apply to **GPAIM 'with Systemic Risk'** (**GPAIMSR**)
- '**Systemic Risk**' =
  - A risk that is specific to the high-impact capabilities of a GPAIM, having a **significant impact on the EU market**
  - Due to a) its reach, or b) actual or reasonably foreseeable **negative effects** on public health, safety, public security, fundamental rights, or the society as a whole
  - That can be propagated at scale **across the value chain**
- The specific rules for GPAIMSR apply also when these models are integrated or form part of an AIS

GenAI & the AI Act (5)

Privacy+ Security Forum

SPECTRUM OF GPAIM - AIS INTEGRATION UNDER THE AI ACT



GenAI & the AI Act (6)

Privacy+ Security Forum

OBLIGATIONS FOR PROVIDERS OF GPAIM

a) **Keep Technical Documentation** on GPAIM (Including its training and testing process, the results of its evaluation, which shall contain, at a minimum, the information set out in Annex XI to the AI Act)

b) **Make available information and documentation to Providers of AIS** who intend to integrate the GPAIM into their AIS (To enable AIS Providers to have a good understanding of the capabilities and limitations of the GPAIM and to comply with their obligations under the AI Act)

c) **Put in place a policy to comply with EU copyright law** and related rights (In particular to identify and comply with any reservation of rights expressed pursuant to the EU Copyright Directive)

d) **Draw up** and make publicly available a **summary about the content used for training** of the GPAIM

e) **Appoint an Authorized Representative** in the EU, if the GPAIM Provider is established outside of the EU

Obligations a) and b) do not apply to Providers of GPAIM released under a free and open-source license (and that are not GPAIMSR)

GenAI & the AI Act (7)

Privacy+ Security Forum

OBLIGATIONS FOR PROVIDERS OF GPAIMSR - In addition to the obligations for GPAIM Providers:

a) **Evaluate** the GPAIMSR in accordance with standardized protocols and tools reflecting the state of the art, including **conducting** and documenting adversarial **testing** of the model with a view to identifying and mitigating systemic risks

b) **Assess and mitigate possible systemic risks** at EU level that may stem from the development, the placing on the market, or the use of GPAIMSR

c) **Document and report** to the AI Office / national competent authorities, relevant information about **serious incidents** involving the GPAIMSR and possible corrective measures to address them

d) **Ensure** an adequate level of **cybersecurity protection** for the GPAIMSR

e) **Notify the European Commission** following confirmation that a model qualifies as an GPAIMSR because it has high impact capabilities

A GPAIM shall be presumed to have **high impact capabilities** when the cumulative amount of computation used for its training measured in floating point operations (FLOPS) is greater than 10^{25}

GenAI & the AI Act (8)

Privacy+ Security Forum

- GPAIM which are in conformity with **HARMONIZED STANDARDS** (published in the Official Journal of the EU) will be **presumed to be in conformity** with the requirements and obligations of GPAIM Providers in the AI Act

- The European Commission has asked the European Committee for Standardization and the European Committee for Electrotechnical Standardization (**CEN-CENELEC**) to develop harmonized standards in support of the AI Act, with a deadline set for **25 April 2025**

- Standards will have to

  a) **Be clear** and **consistent**, including with the standards developed in accordance with relevant EU harmonization legislation

  b) **Aim to ensure** that GPAIM placed on the EU market meet the AI Act's requirements or obligations

GenAI & the AI Act (9)

- The European Commission may adopt **IMPLEMENTING ACTS ESTABLISHING COMMON SPECIFICATIONS** for the requirements and obligations of GPAIM Providers in the AI Act
- GPAIM which are in conformity with the common specifications will be **presumed to be in conformity** with the requirements of the AI Act
- Providers of GPAIM that **do not comply** with the common specifications will have to justify, on a case-by-case basis, that they have adopted technical solutions that meet the requirements of the AI Act

GenAI & the AI Act (10)

Privacy+ Security Forum

- The AI Office has the task of encouraging and facilitating the drawing up of **CODES OF PRACTICE** at EU level in order to contribute to the proper application of the AI Act
- The **AI Office may invite all Providers of GPAIM**, as well as relevant national competent authorities, to participate in the drawing-up of Codes of Practice. Civil society organizations, industry, academia and other relevant stakeholders, such as downstream providers and independent experts, may support the process
- 9-month **deadline** = tight!
- Providers of GPAIM may rely on Codes of Practice to **demonstrate compliance** with the AI Act

GenAI & the AI Act (11)

Privacy+ Security Forum

SUPERVISION, INVESTIGATION, ENFORCEMENT AND MONITORING

- The **European Commission** will have **exclusive powers** to supervise and enforce rules applicable to **Providers of GPAIM**, and to entrust the implementation of these tasks to the AI Office
    - The AI Office may take the necessary actions to monitor the effective implementation and compliance with the AI Act by Providers of GPAIM, including their adherence to Codes of Practices
    - Has the power to:
        - Request documentation and information
        - Conduct (compliance) evaluations
        - Request mitigation measures
- The European Commission may impose on Providers of GPAIM **fines** not exceeding 3 % of their annual total worldwide revenues or EUR 15,000,000, whichever is higher