EDPS

Generative AI & data Protection

EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

Manuel García Sánchez

HoS Consultation and Audits

30/11/2023

Generative AI

- **Generative AI** is a subset of artificial intelligence aiming at **generating new content**, including text, images, videos and code, among others, through learning patterns obtained from pre-existing data.
- it relies on the use of the so-called **foundation models**, which serve as a base model for other AI systems that will be 'fine-tuned' from it. These are generally models that are trained on diverse and extensive datasets. **Large language models** are a specific type of foundation model trained on massive amounts of text data that can generate natural language responses to a wide range of inputs.
- Generative AI is receiving significant attention and is subject to particular scrutiny due to the **potential harms** involving misinformation, copyright, environmental damage as well as privacy and data protection considerations.

Generative AI – Personal data processing

- Personal data processing in a generative AI system **can occur at various levels and stages of its development process, implementation and use**, without necessarily being obvious at first sight.

- Personal data can be processed:
    - when creating the **training datasets**,
    - at the **training stage**,
    - **by inference** once the model is created and in use, or
    - through the **inputs and outputs** of the system once it is running.

Generative AI – Personal data processing

- If your use of generative AI systems involves the processing of personal data, **the EUDPR applies in full**. The Regulation is technologically neutral, and applies to all personal data processing activities, regardless of the technologies used **and without prejudice to other applicable legal frameworks**, in particular the future AI Act.

- EUIs must consider carefully how to use generative AI responsibly and beneficially. **All stages of a generative AI solution life cycle should operate in accordance with the applicable legal frameworks**, including the Regulation when the system involves the processing of personal data.

Generative AI – Lawfulness

- The processing of any personal data in generative AI systems **is only lawful if at least one of the grounds for lawfulness listed in the Regulation is applicable**. In addition, for the processing of special categories of personal data to be lawful, one of the exceptions listed in the Regulation must apply.

- **The use of consent as a legal basis might apply for some use cases**. For that consent to be valid, it needs to meet all the legal requirements, including the need for a **clear affirmative action, be freely given, specific, informed and unambiguous**. We may need to consider the impact of consent withdrawal.

Generative AI – Data minimisation

- There is a **false believe that the principle of data minimisation has no place in the context of artificial intelligence**. However, data controllers have an obligation to limit the processing of personal data to that which is strictly necessary for the purposes of the processing.

- Controllers must ensure that **the staff involved in the development and / or procurement of models are aware of the different technical procedures available to minimize** the use of personal data.

- **Models must be trained with high quality datasets** including only the personal data necessary to fulfil the purpose of the processing.

- When using systems **designed or operated by third-party service providers**, controllers should include in their assessments of the models **data minimisation considerations**.

Generative AI – Data accuracy

- Data controllers must ensure data accuracy at all stages of the development and use of an artificial intelligence system.

- This implies **having control over the content of the datasets used for training models**, including when these are sourced from third parties. Contractual assurances and certifications must be obtained on the procedures used to ensure the accuracy of the data used.

- It is equally **important to have control over the output data**, including the inferences made by the model, which requires regular monitoring of that information.

- Although generally not data protection oriented, **metrics on statistical accuracy, when available, can offer an indicator for the accuracy of the data the model uses as well as on the expected performance**.

Generative AI – DPIAs

- The Regulation requires that a DPIA has to be carried out when the personal data processing **is likely to result in a high risk to fundamental rights and freedoms of natural persons, and always before the start of the processing**. As a result of the assessment, appropriate technical and organizational measures must be taken to mitigate the identified risks.

- The controller **is obliged to seek the advice of the data protection officer (DPO) when carrying out a DPIA**. It might be advisable to seek the views of those affected by the model, either the data subjects themselves or relevant representatives.

- **Regular monitoring and updates of the DPIAs need to be carried out**, since the functioning of the model may exacerbate identified risks or create new ones.

Generative AI – Individual rights

- Appropriate **information and transparency policies** help mitigate risks to individuals and ensure compliance with the requirements of the Regulation.

- This implies having **comprehensive information about the content of the datasets** used at different stages of development, including the origin of the data, the curation/tagging procedure, as well as any associated processing. In particular, EUIs should ensure to have adequate information on those datasets provided their suppliers and that such information is reliable and regularly updated.

- As the right to information includes the obligation to provide, in cases of **profiling and automated decisions**, meaningful information about the logic of such decisions, their meaning and possible consequences on the individual, it is important to **maintain up-to-date information, not only about the functioning of the algorithms used but also about the processing datasets.**

Generative AI – Individual rights

- The particular characteristics of the generative AI systems mean that the exercise of individual rights **can present particular challenges**, not only in the area of the right of access, but also in relation to the rights of rectification, erasure and objection to data processing.

- One of the most relevant elements **is the difficulty in gaining access to the personal data undergoing processing** because of the arduous of isolating them within the complex structures of the model.

- Keeping a **detailed record of the processing of personal data**, as well as **managing datasets in a way that allows traceability of their use**, can allow for the proper exercise of individual rights. Data minimisation techniques can also help to mitigate the risks related to not being able to ensure the proper exercise of individual rights in accordance with the Regulation.

Generative AI – data security

- The Regulation requires EUIs to implement appropriate technical an organisational measures to ensure a level of security appropriate to the risk. You should, in addition to the traditional security controls for IT systems, integrate **specific controls tailored to the already known vulnerabilities of these systems** - model inversion attacks, prompt injection, jailbreaks - in a way that facilitate **continuous monitoring and assessment of their effectiveness**.

- Use only **datasets provided by trusted sources and carry out regularly verification and validation procedures, including for in-house datasets**. Train your staff on how to identify and deal with security risks linked with the use of generative AI systems.

Generative AI – Bias

- In general, artificial intelligence solutions **tend to magnify existing human biases and to incorporate new ones**, which create new ethical challenges and legal compliance risks.

- It is essential that the datasets you use to create and train your models **ensure an adequate and fair representation of the real world** - without bias - while also implementing **accountability and oversight mechanisms that allow for continuous monitoring to prevent the occurrence of biases**.

- EUIs, as public authorities, should put in place **safeguards to avoid overreliance in the results provided by the systems that can lead to automation and confirmation biases**.

Generative AI – The role of the DPO

- From the organisational perspective, the implementation of generative AI systems compliance with the basic principles of the Regulation **should not be a one-person effort.**

- the DPO may liaise with other relevant functions within the organisation, notably the **Legal Service, the IT Service and the LISO in order to ensure that the EUI works within the parameters of good data governance and compliance with the Regulation**.

- When possible, **the creation of an AI task force** and the preparation of an action plan, including **awareness raising and training actions** at all levels of the organisation and the **preparation of internal guidance** may contribute to the achievement of these objectives.

EDPS

EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

@EU_EDPS

European Data Protection Supervisor

EDPS