EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

28 October 2025

Generative AI and the EUDPR.

Orientations for ensuring data protection compliance when using Generative AI systems.

(Version 2)

edps.europa.eu

These EDPS Orientations on generative Artificial Intelligence (generative AI) and personal data protection intend to provide practical advice and instructions to EU institutions, bodies, offices and agencies (EUIs) on the processing of personal data when using generative AI systems, to facilitate their compliance with their data protection obligations as set out, in particular, in Regulation (EU) 2018/1725. These orientations have been drafted to cover as many scenarios and applications as possible and do not prescribe specific technical measures. Instead, they put an emphasis on the general principles of data protection that should help EUIs comply with the data protection requirements according to Regulation (EU) 2018/1725.

These revised orientations offer more detailed guidance taking into account the evolution of Generative AI systems and technologies, their use by EUIs, and the results of the EDPS' monitoring and oversight activities.

The EDPS issues these orientations in its role as a data protection supervisory authority and not in its role as market surveillance authority under the AI Act.

These orientations are without prejudice to the Artificial Intelligence Act.

Introduction and scope

1. These orientations are intended to provide practical advice by the European Data Protection Supervisor (EDPS) to Union institutions, bodies, offices and agencies (EUIs) on the processing of personal data in their use of generative AI systems, to ensure that they comply with their data protection obligations in particular as set out in the Regulation (EU) 2018/1725 ('the Regulation', or EUDPR). Even if the Regulation does not explicitly mention the concept of Artificial Intelligence (AI), the right interpretation and application of the data protection principles is essential to achieve a beneficial use of these systems that does not harm individuals' fundamental rights and freedoms.

2. The EDPS issues these orientations in his role as a data protection supervisory authority and not in his new role as market surveillance authority under the AI Act.

3. These orientations do not aim to cover in full detail all the relevant questions related to the processing of personal data in the use of generative AI systems that are subject to analysis by data protection authorities. Some of these questions are still open, and additional ones are likely to arise as the use of these systems increases and the technology evolves in a way that allows a better understanding on how generative AI works.

4. Because artificial intelligence technology evolves quickly, the specific tools and means used to provide these types of services are diverse and they may change very quickly. Therefore, these orientations have been drafted to cover as many scenarios and applications as possible.

5. These orientations are structured as follows: key questions, followed by initial responses along with some preliminary conclusions, and further clarifications or examples.

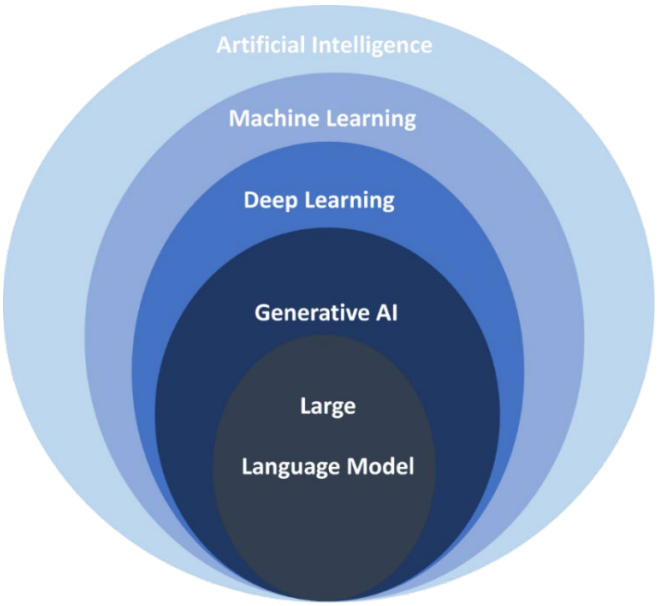
6. The first orientations, issued in 2024, served as a preliminary step towards the development of more comprehensive guidance by the EDPS. In 2025, these orientations have been revisited and expanded, providing further clarification and additional elements to support EUIs in the development and implementation of these systems. The guidance may continue to be updated, refined, and expanded over time to address emerging needs and ensure effective implementation.

1. What is generative AI?

Generative AI is a subset of AI and refers to deep-learning models that can generate high-quality text, images, and other content based on the data they were trained on[1]. To explain the technical background of generative AI, it uses complex machine learning models called deep learning models that mimic the human brain's learning and decision-making processes. These models operate by identifying and encoding patterns and relationships within extensive datasets. They then use this information to comprehend natural language requests from users and create new, relevant content in response[2].

Large language models ("LLM") are a type of machine learning models trained on massive amounts of text data (from billions to trillions of tokens[3]) that can generate natural language responses to a wide range of inputs based on patterns and relationships between words and phrases. This vast amount of text used to train the model may be taken from the Internet, books, and other available sources. Some applications already in use are code generation systems, virtual assistants, content creation tools, language translation engines, automated speech recognition, medical diagnosis systems, scientific research tools, etc.

The relationship between these concepts is hierarchical and can be understood as a series of subsets within a larger framework, where each subsequent term represents a more specialized version of the previous one. Generative AI is a broad category encompassing various forms of content generation, while LLM is a specific application of generative AI.

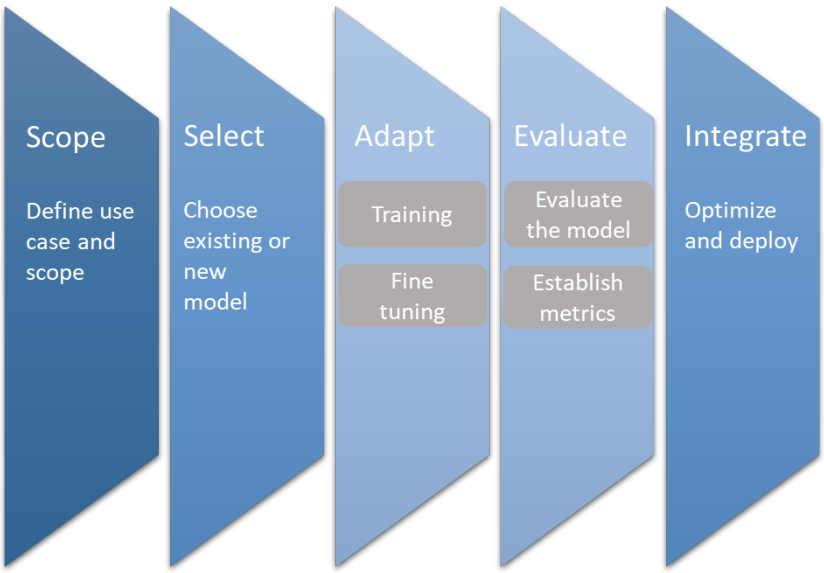


Graphic: *Conceptual hierarchy of Artificial Intelligence*

The generative AI life cycle covers different phases, starting by the definition of the use case and the scope. In some cases, it might be possible to identify a suitable existing pre-trained

[1] IBM Research, What is generative AI, https://research.ibm.com/blog/what-is-generative-AI
[2] IBM Research, https://www.ibm.com/think/topics/generative-ai
[3] Tokens represent commonly occurring sequences of characters. For example, the string "tokenization" is decomposed as "token" and "ization", while a short and common word like "the is represented as a single token.

model to start with, or in other cases a new model may be built from scratch. The following phase involves training the model with relevant datasets for the purpose of the future system, including fine-tuning of the model with specific, custom datasets required to meet the defined use case. To finalise the training, specific techniques requiring human agency are used to ensure more accurate information and controlled behaviour. The following phase aims at evaluating the model and establishing metrics to regularly assess factors, such as accuracy, and the alignment of the model with the use case. Finally, models are deployed and implemented, including continuous monitoring and regular assessment using the metrics established in previous phases.



Graphic: *the generative AI life cycle.*

At this point, it is crucial to distinguish between AI models and AI systems. While AI models are foundational, they do not constitute AI systems on their own[4]. An AI system is a comprehensive framework that can be composed by one or more AI models, alongside with other essential components. The EDPS' Guidance for Risk Management for Artificial Intelligence systems provides extensive information on the complete AI lifecycle. Relevant use cases in generative AI are general consumer-oriented applications (such as ChatGPT and similar systems that can be already found in different versions and sizes[5], including those that can be executed in a mobile phone). There are also business applications in specific areas, pre-trained models, applications based on pre-trained models that are tuned for specific use in an area of activity, and, finally, models in which the entire development, including the training process, is carried out by the responsible entity.

[4] EDPS Guidance for Risk Management for Artificial Intelligence systems (*forthcoming*).
[5] The size of a Large Language Model (LLM) is usually measured by the number of parameters it contains. The size of an LLM is important since some capabilities only appear when the model grows beyond certain limits.

Generative AI, like other new technologies, offers solutions in several fields meant to support and enhance human capabilities. However, it also creates challenges with potential impact on fundamental rights and freedoms that risk being unnoticed, overlooked, not properly considered and assessed. The EDPS' Guidance for Risk Management for Artificial Intelligence systems[6] offers comprehensive direction for EUIs to identify, mitigate and manage risks stemming from data processing activities involving AI systems.

→ The training of a Large Language Model (LLM) (and generally of any machine-learning model) is an iterative, complex and resource intensive process that involves several stages and techniques aiming at creating a model capable of generating human-like text in reaction to commands (or prompts) provided by users. The process starts with the model being trained on massive datasets, most of it normally unlabeled and obtained from public sources using web scraping technologies (data protection authorities already have expressed concerns and outline the key privacy and data protection risks associated with the use of publicly accessible personal data). After that, LLMs are - not in all cases - fine-tuned using supervised learning or through techniques involving human or AI agency (such as Reinforcement Learning with Human Feedback (RLHF), Reinforcement Learning with AI Feedback (RLAIF) or Constitutional AI and Adversarial Testing via Domain experts) to help the system better recognize and process information and context, as well as to determine preferred responses, whether to limit output in reply to sensitive questions and to align it with the values of the developers (e.g. avoid producing harmful or toxic output). Once in production, some systems use the input data obtained through the interaction with users as a new training dataset to refine the model.

---

[6] *(forthcoming)*

2. Can EUIs use generative AI?

As an EUI, there is no obstacle in principle to develop, deploy and use generative AI systems in the provision of public services, providing that the EUI's rules allow it, and that all applicable legal requirements are met, especially considering the special responsibility of the public sector to ensure full respect for fundamental rights and freedoms of individuals when making use of new technologies.

In any case, if the use of generative AI systems involves the processing of personal data, the Regulation 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 legal frameworks, in particular the AI Act. The principle of accountability requires responsibilities to be clearly identified and respected amongst the various actors involved in the generative AI model supply chain.

EUIs can develop and deploy their own generative AI solutions or can alternatively deploy for their own use solutions available on the market. In both cases, EUIs may use providers to obtain all or some of the elements that are part of the generative AI system.

To make sure that EUIs deploy and use a generative AI solution which is in compliance with the Regulation, they can follow the below recommendations:

- **Define Purpose and Legal Basis:** Clearly define the specific purpose for processing of the generative AI and identify the appropriate legal basis for its deployment.

- **Determine and document roles and responsibilities** (see section 3 below): Formally determine and document all roles and responsibilities in relation to the processing operation taking place in the context of the generative AI.

- **Records Registry:** Ensure that the processing activities taking place in the context of the generative AI system are thoroughly documented in your records. Such records should include all necessary information required by Article 31 of the Regulation.

- **Conduct a Generative AI Risk Assessment:** Perform a comprehensive generative AI risk assessment in accordance with the EDPS' Guidance for Risk Management for Artificial Intelligence system.[7]

- **Conduct a Data Protection Impact Assessment ("DPIA")** (see section 6 below)**:** If required, conduct a DPIA and adopt data protection by design and by default measures.[8]

[7] *(forthcoming)*
[8] See the EDPS guidance "Accountability on the ground Part II: Data Protection Impact Assessments & Prior consultation", available at : https://www.edps.europa.eu/sites/default/files/publication/18-02-06_accountability_on_the_ground_part_2_en.pdf

- **Implement core data protection principles:** As required under the Regulation, implement all data protection principles such as transparency, data minimisation, data retention and data security.

- **Uphold Data Subject Rights** (see section 14 below)**:** EUIs must have robust procedures in place to handle data subject rights, such as the right to access, rectification, and erasure of personal data.

- **Perform Third Party Vendor Due Diligence:** In case EUIs use a generative AI model developed by another entity, they should thoroughly assess and ensure that the generative AI complies with all data protection requirements, and request all documentation to perform the verification. If the development of the generative AI model was outsourced by the EUI, the EUI remains a controller[9], and the third party vendor is a processor[10] developing the model on behalf of the EUI. In the latter case, a data processing agreement should be put in place in accordance with Article 29 of the Regulation.

- **Ensure Accountability:** Document all implemented mitigation measures and the final assessment that the generative AI is trustworthy, and compliant with the Regulation, thereby ensuring full accountability.

As AI technologies advance rapidly, EUIs must consider carefully when and how to use generative AI responsibly and beneficially for public good. 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.

→ The terms trustworthy or responsible AI refer to the need to ensure that AI systems are developed in an ethical and legal way. It entails considering the unintended consequences of the use of AI technology and the need to follow a risk-based approach covering all the stages of the life cycle of the system. It also implies transparency regarding the use of training data and its sources, on how algorithms are designed and implemented, what kind of biases might be present in the system and how are tackled possible impacts on individual's fundamental rights and freedoms. In this context, generative AI systems must be transparent, explainable, consistent, auditable and accessible, as a way to ensure fair processing of personal data.

---

[9] Article 3(8) of the Regulation.
[10] Article 3(1) of the Regulation.

3. How to determine roles and responsibilities in Generative AI systems?

Several parties are involved in the development and deployment of generative AI systems, with distinct extents of involvement in the processing of personal data. A critical step is the qualification of the roles and responsibilities of such operators by determining whether they are controllers, processors, or joint controllers within the meaning of the Regulation. Such qualification is necessary to determine the obligations of operators involved in each processing in accordance with the Regulation.

In the context of a generative AI system, personal data processing involves multiple entities and various purposes and operations that depend on the stage of an AI model's life cycle. The different stages can be summarized in two phases, the development phase, which includes all stages before the deployment of the AI model (e.g. pre-training, post-training), and the deployment phase, which covers all stages related to the use of the AI model[11]. Considering the complexity of the supply chain in generative AI, as well as the varying degrees of involvement at each stage of the processing operation, the determination of the roles of controller, processor and joint controller is particularly challenging for generative AI systems. Therefore, EUIs should conduct a thorough assessment on a case-by-case basis and document the results in their records of processing activities.

It is worth clarifying that the terms "provider", "developer" and "deployer" widely-used in the tech industry and in other legal frameworks (such as the AI Act) do not correspond to the data protection concepts of "controller", "processor", "joint controller" as defined in the Regulation[12]. EUIs are advised to follow the existing EDPS Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725[13].

Those data protection roles in the generative AI systems can be defined as follows:

A **controller** is the entity which determines the purposes and means of the processing of personal data[14]. As analysed in the EDPS Guidelines, determining the essential means of the processing, such as what type of personal data is collected, the duration of the processing, etc should be sufficient. In generative AI, an entity that determines "why" and "how" processing takes place includes an organization that decides to develop an AI system, use a service provider for development, and/ or deploy a generative system for a specific purpose.

Additionally, an entity who decides to create the training data set on the basis of data collected on its own account may be qualified as controller as well. For entities that rely on

[11] EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, paragraph 18, available at: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf .
[12] ICO Generative AI fifth call for evidence: allocating controllership across the generative AI supply chain.
[13] EDPS Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU). 2018/1725, available at: https://www.edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf
[14] Article 3(8) of the Regulation.

pre-trained data sets created by third parties on their own account, it is important to identify the processing for which such third parties are controllers.

A **joint controller** is the entity which, jointly with others, determines both the purposes and means of the processing of personal data[15]. In the context of generative AI, the entities involved in the processing should determine whether they process data for their own purposes, or for a common purpose shared with another entity. For instance, when organizations decide on a shared objective for developing, training and fine-tuning an AI system, and they both have a significant influence in "why" and "how" the data is processed to achieve that objective, they are joint controllers. In case of joint controllership, joint controllers are under an obligation to define their respective obligations in a transparent manner by concluding an agreement in accordance with Article 28 of the Regulation.

A **processor** is the entity which processes personal data on behalf of the controller.[16] This can be an AI developer who develops the generative AI system on behalf of an EUI as part of a service, and provides the technical infrastructure for the generative AI system to the EUI without any control over the purpose and essential means of its use.

The terms "provider", "developer" and "deployer", commonly used in the tech sector and in frameworks like the AI Act, do not align with the data protection concepts of "controller", "processor" and "joint controller" under Regulation (EU) 2018/1725; EU institutions should therefore rely on the EDPS Guidelines for the correct interpretation of these roles.

[15] EDPS Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725, section 5.
[16] Article 3(12) of the Regulation.

→ EUI-D develops a generative AI tool that can be deployed to support Human Resources (HR) departments of other EUIs - for example, to assist in drafting job descriptions, screening CVs, or summarising interview notes. The generative AI tool integrates LLMs sourced from a third-party provider (e.g., OpenAI, or another LLM developer). EUI-D collects and ingests datasets for the training of the system. During the development stage, EUI-D qualifies as a controller as it determines the purposes (e.g., creating a tool to streamline recruitment workflows) and means of processing (e.g., selecting the data and defining the prompt structure). The third party LLM provider, while a key part in the tool's development, does not determine neither the specific purposes nor means of this initial training process carried out by EUI-D and is therefore not considered a controller or processor for the development of the generative AI tool by EUI-D. Conversely, the third-party LLM developer would be considered a processor if it had developed the LLM at the request of EUI-D in the context of providing a service to the latter.

After the tool is developed, EUI-X purchases the generative AI tool from EUI-D, and implements it to support its HR department. EUI-X inputs internal HR data into the system, including personal data of its applicants and EU staff (e.g., names, qualifications, job histories, internal assessments). It also defines the prompts and system configurations to guide how the tool operates for this specific purpose. EUI-D does not have access to the system of EUI-X and its data. In this new setup, EUI-X qualifies as a controller, as it determines the purposes (e.g., streamlining its own recruitment workflows) and the essential means (e.g., selecting its own data and prompts) of its own distinct processing and usage of the data. EUI-D has merely provided the product and operates as a separate controller, since it is not involved in the processing operation of EUI-X and is only responsible for its own and distinct processing operation of developing the tool. Conversely, EUI-X and EUI-D would be joint controllers for the processing taking place at the development stage, if they would jointly develop the generative AI tool to optimise processing operations in their respective HR departments, and would jointly feed the AI system to that purpose.

4. How to know if the use of a generative AI system involves personal data processing?

Personal data processing in a generative AI system can occur on various levels and stages of its lifecycle, without necessarily being obvious at first sight. During the development stage[17], personal data could be processed as part of the training, testing and validation datasets. During the deployment phase, personal data could be processed as input (prompts including personal data) and output (inferences including personal data) of the AI model or system, but also due to model memorization of training data (reproduction of personal data).

When a developer or a provider of a generative AI system claims that their system does not process personal data (for reasons such as the alleged use of anonymised datasets or synthetic data during its design, development and testing), it is crucial to ask about the specific controls that have been put in place to guarantee this. Essentially, EUIs may want to know what steps or procedures the provider uses to ensure that personal data is not being processed by the model.

The European Data Protection Board (EDPB), in its Opinion 28/2024 (Section 3.2), has clarified the circumstances under which an AI model trained with personal data may be considered anonymous by a supervisory authority: in sum, the EDPS considers that, for an AI model to be considered anonymous, using reasonable means, both (i) the likelihood of direct (including probabilistic) extraction of personal data regarding individuals whose personal data were used to train the model; as well as (ii) the likelihood of obtaining, intentionally or not, such personal data from queries, should be insignificant[18] for any data subject. By default, the EDPS considers that AI models are likely to require a thorough evaluation of the likelihood of identification to reach a conclusion on their possible anonymous nature. This likelihood should be assessed taking into account 'all the means reasonably likely to be used' by the controller or another person, and should also consider unintended (re)use or disclosure of the model. [19]

The use of web scraping techniques to collect personal data entails significant risks, considering that individuals may lose control of their personal information when this is collected without their knowledge, against their expectations, and for purposes that are different from those of the original collection. The EDPS notes that the processing of personal data that is publicly available remains subject to EU data protection legislation. In that regard, the use of web scraping techniques to collect data from websites and their use for training purposes will have to comply with all relevant data protection principles, such as lawfulness,

[17] "The development of an AI model covers all stages before any deployment of the AI model, and includes, inter alia, code development, collection of training personal data, pre-processing of training personal data, and training. The deployment of an AI model covers all stages relating to the use of an AI model and may include any operations conducted after the development phase." EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, para 18.

[18] CJEU judgment of 19 October 2016, Case C-582/14, Breyer v Bundesrepublik Deutschland (ECLI:EU:C:2016:779), paragraph 46, and CJEU judgment of 7 March 2024, Case C-479/22 P, OC v European Commission (ECLI:EU:C:2024:215), paragraph 51.

[19] For further details, see EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models.

transparency, data minimisation and the principle of accuracy, insofar as there is no assessment on the reliability of the sources. A primary challenge to ensuring the legality of web scraping is establishing a valid lawful ground under Article 5 of the Regulation. While web scraping per se is not prohibited, EUIs may face significant challenges to identify an appropriate lawful ground in the context of this data collection technique. For instance, relying on the lawful ground of public interest (Article 5(1)(a)) requires that the legal basis for the processing is laid down in EU law. This would mean that an EUI has to justify why web scraping techniques are necessary to perform its tasks prescribed by EU law.

If a lawful ground is identified for web scraping, EUIs also have to ensure that they meet their transparency obligations in accordance with Articles 14-16 of the Regulation. It is fundamental that individuals have a thorough understanding of how their personal data is processed through web scraping. Considering the difficulties that may be encountered to ensure that collection of personal data via web scraping is in line with the Regulation, the EDPS recommends EUIs to use different sources of personal data, where possible. If web scraping techniques are used for personal data collection, EUIs should put in place safeguards to minimise the impact such collection has on the rights and freedoms of individuals. For instance, data collection could be limited to freely accessible data that were made manifestly public by the individual.[20]

Regular monitoring and the implementation of controls at all stages can help verify that there is no personal data processing, in cases where the model is not intended for it. Web scraping techniques for data collection should be used with caution, and safeguards should be put in place to limit their impact on the rights and freedoms of individuals.

→ EUI-X, a fictional EU institution, is considering the acquisition of a product for automatic speech recognition and transcription. After studying the available options, it has focused on the possibility of using a generative AI system to facilitate this function. In this particular case, it is a system that offers a pre-trained model for speech recognition and translation. Since this system will be used for the transcription of meetings using recorded voice files, it has been determined that the use of this system requires the processing of personal data and therefore it must ensure compliance with the Regulation.

---

[20] For additional information on safeguards that could be put in place, please check: https://www.cnil.fr/en/legal-basis-legitimate-interests-focus-sheet-measures-implement-case-data-collection-web-scraping

5. What is the role of DPOs in the process of development or deployment of generative AI systems?

Article 45 of the Regulation establishes the tasks of the data protection officer (DPO). DPOs inform and advise on the relevant data protection obligations, assist controllers to monitor internal compliance, provide advice where requested regarding DPIAs, and act as the contact point for data subjects and the EDPS.

In the context of the implementation by EUIs of generative AI systems that process personal data it is important to ensure that DPOs, within their role, advise and assist in an independent manner on the application of the Regulation, have a proper understanding of the lifecycle of the generative AI system that the EUI is considering to procure, design or implement and how it works. This means, obtaining information on when and how these systems process personal data, and how the input and output mechanisms work, as well as the decision-making processes implemented through the model. As the Regulation points out[21], the DPO has to provide advice to controllers when conducting data protection impact assessments. Finally, the DPO should be involved in the review of compliance issues in the context of data sharing agreements signed with model providers.

It should be borne in mind that the responsibility to ensure that all processing operations carried out in the context of generative AI are compliant with the Regulation remains with the controller.[22] In that respect, controllers must ensure that all processes are properly documented and that transparency is guaranteed, including updating records of processing and, as a best practice, carrying out a specific inventory on generative AI - driven systems and applications.

From the organisational perspective, the implementation of generative AI systems in compliance with the Regulation should not be a one-person effort. There should be a continuous dialogue among all the stakeholders involved across the lifecycle of the product. Therefore, controllers should liaise with all relevant functions within the organisation, notably the DPO, Legal Service, the IT Service and the Local Informatics Security Officer (LISO) in order to ensure that the EUI works within the parameters of trustworthy generative AI, good data governance and complies with the Regulation. The creation of an AI task force, including the DPO, and the preparation of an action plan, including awareness raising actions at all levels of the organisation and the preparation of internal guidance may contribute to the achievement of these objectives.

[21] Article 39(2) of the Regulation.
[22] EDPS Position paper on the role of Data Protection Officers of the EU institutions and bodies, p.12-13, available at: https://www.edps.europa.eu/sites/default/files/publication/18-09-30_dpo_position_paper_en.pdf

→ As an example of contractual clauses, the European Commission, through the "Procurement of AI Community" initiative, has brought together relevant stakeholders in procuring AI solutions to develop wide model contractual clauses for the procurement of Artificial Intelligence by public organizations. It is also relevant to consider the standard contractual clauses between controllers and processors under the Regulation[1].

6. An EUI wants to develop or implement generative AI systems. When should a DPIA be carried out?

The principles of data protection by design and by default[23] aim to protect personal data throughout the entire life cycle of data processing, starting from the inception stage. By complying with this principle of the Regulation, based on a risk-oriented approach, the threats and risks that generative AI may entail can be considered and be mitigated sufficiently in advance. Developers and deployers may need to carry out their own risk assessments and document any mitigation action taken.

The Regulation requires that a DPIA[24] must be carried out before any processing operation that is likely to result in a high risk[25] to fundamental rights and freedoms of individuals. The Regulation points out the importance of carrying out such assessment, where new technologies are to be used or are of a new kind in relation to which no assessment has been carried out before by the controller, in the case of generative AI systems for example.

The controller is obliged to seek the advice of the DPO when carrying out a DPIA. Because of the assessment, appropriate technical and organisational measures must be taken to mitigate the identified risks given the responsibilities the context and the available state-of-the-art measures.

It may be appropriate, in the context of the use of generative AI to seek the views of those affected by the system, either the data subject themselves or their representatives in the area of intended processing. In addition to the reviews to assess whether the DPIA is rightly implemented, regular monitoring and reviews of the risk assessments need to be carried out, since the functioning of the model may exacerbate identified risks or create new ones. Those risks are related to personal data protection, but are also related to other fundamental rights and freedoms.

All the actors involved in the DPIA must ensure that any decision and action is properly documented, covering the entire generative AI system lifecycle, including, actions taken to manage risks and the subsequent reviews to be carried out.

[23] Article 27 of the Regulation

[24] Articles 39 and 89 of the Regulation.

[25] The classification of an AI system as posing "high-risk" due to its impact on fundamental rights according to the AI Act, does trigger a presumption of "high-risk" under the GDPR, the EUDPR and the LED to the extent that personal data is processed.

It is the EUIs' responsibility to appropriately manage the risks connected to the use of generative AI systems. Data protection risks must be identified and addressed throughout the entire life cycle of the generative AI system. This includes regular and systematic monitoring to determine, as the system evolves, whether risks already identified are worsening or whether new risks are appearing. The understanding of risks linked to the use of generative AI is still ongoing so there is a need to keep a vigilant approach towards non-identified, emerging risks. If risks that cannot be mitigated by reasonable means are identified, it is time to consult the EDPS.

→ The EDPS has established a template allowing controllers to assess whether they have to carry out a DPIA [annex six to Part I of the accountability toolkit]. In addition, the EDPS has established an open list of processing operations subject to the requirement for a DPIA. Where necessary, the controller shall carry out a review to assess if the data processing is being performed in accordance with the data protection impact assessment, at least when there is a change to the risks represented by processing operations. If following the DPIA, controllers are not sure whether risks are appropriately mitigated, they should proceed to a prior consultation with the EDPS.

---

7. When is the processing of personal data during the design, development and validation of generative AI systems lawful?

The processing of personal data in generative AI systems may cover the entire lifecycle of the system, encompassing all processing activities related to the collection of data, training, interaction with the system and systems' content generation. Collection and training-related processing activities include obtaining data from publicly available sources on the Internet, directly, from third parties, or from the EUIs' own files. Personal data can also be obtained by the generative AI model directly from the users, via the inputs to the system or through inference of new information. In the context of generative AI systems, the training and use of the systems relies normally on systematic and large scale processing of personal data, in many cases without the awareness of the individuals whose data are processed.

The processing of any personal data by EUIs is lawful if at least one of the grounds for lawfulness[26] listed in the Regulation is applicable. In addition, for the processing of special categories of personal data to be lawful, one of the exceptions[27] listed in the Regulation must apply. An appropriate lawful ground should be identified for each individual processing operation. In that regard, distinct lawful grounds should be identified for processing carried out during the development phase, and the deployment phase, as the purposes of the processing are distinct in each phase.

Service providers of generative AI models may use legitimate interest under the EU General Data Protection Regulation[28] (GDPR) as a lawful ground for data processing, particularly with regard to the collection of data used to develop the system, including the training and validations processes.[29] Such lawful ground is not applicable in the context of the Regulation. Nonetheless, EUIs have a specific responsibility, as part of their accountability obligations to demonstrate compliance with Article 4(1)(a), and Article 5 of the Regulation, and to verify that the AI model that they are deploying has not been developed by unlawfully processing personal data.[30]

As stated above, EUIs have an obligation to identify a distinct lawful ground for the processing carried out at every stage of the generative AI lifecycle, during both the development and deployment phase. Out of the five lawful grounds of Article 5 of the Regulation, not all would be appropriate to apply to processing taking place in the context of the generative AI. EUIs should make a thorough assessment on a case-by-case basis. The most common lawful ground that EUIs would rely upon would likely be Article 5(1)(a) of the Regulation, where processing is necessary for performance of a task carried out in the public interest or in the exercise of official authority vested in the Union institution or body. The lawful ground of Article 5(1)(b),

[26] Article 5 of the Regulation.
[27] Article 10(2) of the Regulation.
[28] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
[29] For more information on the use of legitimate interest in the context of AI, please check EDPB opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI model.
[30] EDPB opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI model, para 129.

namely compliance with a legal obligation to which the controller is subject to would be challenging to rely on, since it would require a specific law that clearly imposes an obligation to EUIs to process personal data. Such requirement would be difficult to meet especially during the development phase of the generative AI system.

When EUIs decide that the processing is carried out for the performance of task carried out in the public interest or the exercise of official authority[31] they should demonstrate that there is either a task in the public interest related to their core functions, or that they are exercising an official authority through the specific powers, tasks and duties vested in them. The legal basis for the processing must be laid down in EU law.[32] As specified in Recital 22 of the Regulation, public interest also includes the processing of personal data necessary for the management and functioning of the EUIs. The specific legal basis in EU law may provide additional instructions concerning aspects of the processing, such as the data categories, or the retention periods for the personal data processed.[33] In addition, the referred EU Law should be clear and precise and its application should be foreseeable to individuals subject to it, in accordance with the requirements set out in the Charter of Fundamental Rights of the European Union and the European Convention for the Protection of Human Rights and Fundamental Freedoms.

Moreover, where a legal basis gives rise to a serious interference with fundamental rights to data protection and privacy, there is a greater need for clear and precise rules governing the scope and the application of the measure as well as the accompanying safeguards. Therefore, the greater the interference, the more robust and detailed the rules and safeguards should be. When relying on internal rules, these internal rules should precisely define the scope of the interference with the right to the protection of personal data, through identification of the purpose of processing, categories of data subjects, categories of personal data that would be processed, controller and processors, and storage periods, together with a description of the concrete minimum safeguards and measures for the protection of the rights of individuals.

The use of consent[34] as a lawful ground may apply in some limited circumstances in the context of the use of generative AI systems. Obtaining consent[35] under the Regulation, and for that consent to be valid, it needs to meet all the legal requirements, including the need for a clear affirmative action by the individual, be freely given, specific, informed and unambiguous. Given the way in which generative AI systems are trained, and the sources of training data, including publicly available information, it would be practically hard to acquire individuals' consent, also in the context of its use by public bodies, such as EUIs. In other words, it is difficult to obtain valid consent in the context of generative AI systems, when personal data are not collected directly from the individual concerned or are collected on a large scale.

---

[31] Article 5(1)(a) of the Regulation.
[32] Article 5(2) of the Regulation.
[33] Accountability on the ground, Part I: Records, Registers and when to do Data Protection Impact Assessments, p. 19, available at: https://www.edps.europa.eu/sites/default/files/publication/19-07-17_accountability_on_the_ground_part_i_en.pdf
[34] Articles 5(1)(d) and 7 of the Regulation.
[35] EDPB Guidelines 05/2020 on consent under Regulation 2016/679, available at https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

In addition, for consent to be valid, individuals must be able to withdraw consent at any time. If consent is withdrawn, all data processing operations that were based on such consent and took place before the withdrawal - and in accordance with the Regulation - remain lawful. However, in this case, the controller must stop the processing operations concerned. If there is no other lawful basis justifying the processing of personal data, the relevant data must be deleted by the controller, posing a significant challenge to the functionality of the generative AI system. This illustrates the inherent difficulties of relying on consent for generative AI systems.

As controllers for the processing of personal data, EUIs are accountable for the transfers of personal data that they initiate and for those that are carried out on their behalf within and outside the European Economic Area. These transfers can only occur if the EUI in question has instructed the entity processing personal data on their behalf or allowed them, or if such transfers are required under EU law or under EU Member States' Law. Transfers can occur at different levels in the context of the development or use of generative AI systems, including when EUIs make use of systems based on cloud services or when they have to provide, in certain cases, personal data to be used to train, test or validate a model. In either case, these data transfers must comply with the provisions laid down in Chapter V[36] of the Regulation, while also subject to the other provisions of the Regulation, and be consistent with the original purpose of the data processing.

Personal data processing in the context of generative AI systems requires a lawful ground in line with the Regulation. If the data processing is based on public interest, or the exercise of official authority- or more rarely - a legal obligation, that legal basis must be clearly and precisely set out in EU law. The use of consent as a lawful ground requires careful consideration to ensure that it meets the requirements of the Regulation, in order to be valid.

→ The GPA Resolution on Generative Artificial Intelligence Systems states that, where required under relevant legislation, developers, providers and deployers of generative AI systems must identify at the outset the legal basis for the processing of personal data related to: a) collection of data used to develop generative AI systems; b) training, validation and testing datasets used to develop or improve generative AI systems; c) individuals' interactions with generative AI systems; d) content generated by generative AI systems.

---

[36] Articles 46 to 51 of the Regulation.

→ EUI-D deploys an HR analytics system to support their recruitment procedures. The system intends to process personal data of applicants, such as CVs and interview scores. The purpose of the processing is to detect patterns in recruitment - such as nationality representation - to optimise the procedures. EUI-D relies on Article 5(1)(a) of the Regulation for such processing. The assessment of EUI-D is that such processing is necessary to carry out its recruitment procedures in a fair and efficient manner, as required by Article 27 of the Staff Regulations.[37] [38]

[37] Regulation No 31 (EEC), 11 (EAEC), laying down the Staff Regulations of Officials and the Conditions of Employment of Other Servants of the European Economic Community and the European Atomic Energy Community.

[38] Article 27 of the Staff Regulations provides that "The principle of the equality of Union's citizens shall allow each institution to adopt appropriate measures following the observation of a significant imbalance between nationalities among officials which is not justified by objective criteria. Those appropriate measures must be justified and shall never result in recruitment criteria other than those based on merit."

8. How to apply purpose limitation in the generative AI lifecycle?

The power of generative AI models lies in their adaptability and versatility across numerous fields. Their broad functionality however should not come at the expense of data protection principles, particularly the principle of purpose limitation[39] In accordance with the Regulation, personal data can only be processed for specified, explicit and legitimate purposes for which it has been collected.

The lifecycle of a generative AI system comprises of distinct stages, including training, testing, fine-tuning, and deployment, each of which may involve potentially the processing of personal data for different purposes. Regardless of the stage, data protection principles should be respected and a purpose should be defined for each processing operation. It is crucial to note that the purpose of data processing during the development phase can be often distinct from its purpose during deployment.

For example, the purpose of collecting data from publicly available sources to train an LLM model with the aim of enabling it to understand and generate human-like text across a wide range of topics, is distinct from subsequent activities. A separate purpose would be to use historical recruitment data to fine-tune that same model to enhance its performance in screening CVs and generating interview questions tailored to job descriptions. This in turn differs from the deployment phase, where the LLM is used within an organization to support HR staff in conducting job screenings and interviews more efficiently. Each of these stages serves a different purpose, involves different categories of personal data, and presents unique risks and data protection compliance obligations. Consequently, it is essential to assess each phase of the generative AI lifecycle separately from a data protection perspective and define a specific purpose of processing.

Controllers may want to reuse data collected for the initial purpose of training a generative AI model for another processing activity. In such cases, they must determine whether the subsequent processing is compatible with the initial purpose for which the personal data was collected. Article 6 of the Regulation provides criteria for conducting this compatibility assessment, a provision relevant in the context of development and deployment of AI models.[40]

The EDPS recognises that defining a specific and clear purpose for a generative AI model during its development phase might be more challenging than at later stages of deployment. It is inherent in the nature of the generative AI systems to be open-ended and serve for different applications. However, the purpose of the collection must be clearly and specifically identified.[41] Therefore, the purpose should be defined even in the early stages of development of the model, by considering potential use cases and intended functionalities. Controllers should have a clear context for the deployment of the AI model and must include this in the details of the purpose of processing when completing their records. For instance, controllers

[39] Article 4(1)(b) of the Regulation.
[40] EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, adopted on 17 December 2024, paragraph 17.
[41] Article 29 Working Party Opinion 03/2013 on purpose limitation (WP203), p. 15-16, available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

should specify the type of AI model developed, its expected functionalities, and any other relevant context that is already known at that stage.[42]

EUIs should ensure that before processing personal data in generative AI systems, they establish specific, explicit and clear purposes for each different stage of the generative AI lifecycle. The categories of personal data processed in each stage and how the processing can meet the specified purpose should be documented in the records of processing operations, in accordance with Article 31 of the Regulation.

→ EUI-X plans to deploy a generative AI system for translating internal and external documents. Before starting with the processing of any data, EUI-X should define and document the purposes of processing for each phase of the generative AI lifecycle:

- **Training:** Develop a language model using non-personal data.
- **Fine-tuning:** Adapt the model to specific EU terminology using publicly available personal data.
- **Deployment:** Use the live model, which may contain personal data from users and their prompts.

By conducting an analysis, separating the processing activities and determining the distinct purposes, EUI-X can uphold the principle of purpose limitation and ensure compliance with the Regulation.

---

[42] EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, paragraph 64.

9. How can the principle of data minimisation be guaranteed when using generative AI systems?

The principle of data minimisation means that controllers shall ensure that personal data undergoing processing are adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.[43] In the context of artificial intelligence, data controllers have an obligation to limit the collection and otherwise processing of personal data to what is necessary for the purposes of the processing, avoiding indiscriminate processing of personal data. This obligation covers the entire lifecycle of the system, including testing, acceptance and release into production phases. Personal data should not be collected and processed indiscriminately. In that regard, EUIs must also verify that the purpose in question cannot be achieved by processing other data, such as synthetic or anonymised data before deciding that personal data is necessary to process in the first place.[44] If processing of personal data is deemed necessary, EUIs must ensure that staff involved in the development of generative AI models are aware of the different technical procedures available to minimise the use of personal data and that those are duly taken into account in all stages of the development.

EUIs should develop and use models trained with high quality datasets limited to the personal data necessary to fulfil the purpose of the processing. In this way, these datasets should be well labelled and curated, within the framework of appropriate data governance procedures, including periodic and systematic review of the content. Datasets and models must be accompanied by documentation on their structure, maintenance and intended use. When using systems designed or operated by third-party service providers, EUIs should include in their assessments considerations related to the principle of data minimisation.

The use of large amounts of data to train a generative AI system does not necessarily imply greater effectiveness or better results. The careful design of well-structured datasets, to be used in systems that prioritise quality over quantity, following a properly supervised training process, and subject to regular monitoring, is essential to achieve the expected results, not only in terms of data minimisation, but also when it concerns quality of the output and data security.

[43] Article 4(1)(c) of the Regulation.
[44] EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, paragraph 64.

→ EUI-X intends to train an AI system to be able to assist with tasks related to software development and programming. For this, they would like to use a content generation tool that will be available through the individual IT staff members' accounts. EUI-X needs to reflect before training the algorithm to make sure they will not be processing personal data that would not be useful for the intended purpose. For example, they may carry out a statistical analysis to demonstrate that a minimum amount of data is necessary to achieve the result. Furthermore, they will need to check and justify whether they will be processing special categories of personal data. Additionally, they will need to examine the typology of data (i.e. synthesised, anonymised or pseudonymised). Finally, they will need to verify all relevant technical and legal elements of the data sources used, including their lawfulness, transparency and accuracy.

10. Are generative AI systems respectful of the data accuracy principle?

Generative AI systems may use in all stages of their lifecycle, notably during the training phase, huge amounts of information, including personal data.

The principle of data accuracy[45] requires data to be accurate, up to date, while the data controller is required to update or delete data that is inaccurate. Data controllers must ensure data accuracy at all stages of the development and use of a generative AI system. Indeed, they must implement the necessary measures to integrate data protection by design that will help to increase data accuracy in all the stages.

This implies verifying the structure and content of the datasets used for training models, including those sourced or obtained from third parties. 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, including human oversight. Developers should use validation sets[46] during training and separate testing sets for final evaluation to obtain an estimation on how the system will perform. Although generally not data protection oriented, metrics on statistical accuracy (the ability of models to produce correct outputs or predictions based on the data they have been trained on), when available, can offer an indicator for the accuracy of the data the model uses as well as on the expected performance.

When EUIs use a generative AI system or training, testing or validation datasets provided by a third party, contractual assurances and documentation must be obtained on the procedures used to ensure the accuracy of the data used for the development of the system. This includes data collection procedures, preparation procedures, such as annotation, labelling, cleaning, enrichment and aggregation, as well as the identification of possible gaps and issues that can affect accuracy. The technical and user documentation of the system, including model cards, should enable the controller of the system to carry out appropriate checks and actions regularly to ensure the accuracy principle. This is even more important since models, even when trained with representative high quality data, may generate output containing inaccurate or false information, including personal data, the so-called "hallucinations."

Despite the efforts to ensure data accuracy, generative AI systems are still prone to inaccurate results that can have an impact on individuals´ fundamental rights and freedoms. While providers are implementing advanced training systems to ensure that models use and generate accurate data, EUIs should carefully assess data accuracy throughout the whole lifecycle of the generative AI systems and reconsider the use of such systems if the accuracy cannot be maintained.

[45] Article 4(1)(d) of the Regulation.
[46] Validation sets are used to fine-tune the parameters of a model and to assess its performance.

→ EUI-X, following the advice of the DPO, uses a generative AI system to screen job applications for the purpose of summarizing the CVs and documents provided by the candidates in their job application profile within EUI-X talent management system. The goal is to create a concise and standardized summary for every candidate, including all their qualifications, skills and experiences, to be used for the eligibility check by the HR staff. Before inserting information in the generative AI system, EUI-X ensures that all documents provided by the applicants (e.g. CVs, diplomas) are up-to-date. In that respect, if a candidate updated their profile in the talent management system, the data inserted in the AI system are updated as well. To mitigate any hallucinations and inaccuracies in the AI-generated outputs, EUI-X has a manual verification step for each AI generated summary to ensure that they accurately reflect the data provided by the candidates.

11. How to inform individuals about the processing of personal data when EUIs use generative AI systems?

Appropriate information and transparency policies can help mitigate risks to individuals and ensure compliance with the requirements of the Regulation, in particular, by providing detailed information on how, when and why EUIs process personal data in generative AI systems. This implies having comprehensive information - that must be provided by developers or suppliers as the case may be - about the processing activities carried out at different stages of development, including the origin of the datasets, the curation/tagging procedure, as well as any associated processing. In particular, EUIs should ensure that they obtain adequate and relevant information on those datasets used by their providers or suppliers and that such information is reliable and regularly updated. Certain systems (i.e. chatbots) may require specific transparency requirements, including informing individuals that they are interacting with an AI system without human intervention.

As the right to information[47] includes the obligation to provide individuals, in cases of profiling and automated decisions, meaningful information about the logic of such decisions, as well as their meaning and possible consequences on the individuals, it is important for the EUI to maintain updated information, not only about the functioning of the algorithms used, but also about the processing datasets. This obligation should generally be extended to cases where, although the decision procedure is not entirely automated, it includes preparatory acts based on automated processing.

EUIs must provide to individuals all the information required in the Regulation when using generative AI systems that process personal data. The information provided to individuals must be updated when necessary to keep them properly informed and in control of their own data.

→ EU-X is preparing a chatbot that will assist individuals when accessing certain areas of its website. The controllers affected, with the advice of the DPO, have prepared a data protection notice, available in the EU-X website. The notice includes information on the purpose of the processing, the legal basis, the identification of the controller and the contact details of the DPO, the recipients of the data, the categories of personal data collected, the retention of the data as well on how to exercise individual rights. The notice also includes information on how the system works and on the possible use of the user´s input to refine the chat function. EU-X uses consent as a legal basis, but users can withdraw their consent at any moment. The notice also clarifies that minors are not permitted to use the chatbot. Before using the EUI's chatbot, individuals can provide consent after reading the data protection notice.

---

[47] Article 14 of the Regulation.

12. What about automated decisions within the meaning of Article 24 of the Regulation?

The use of a generative AI system does not necessarily imply automated decision-making[48] within the meaning of the Regulation. However, there are generative AI systems that provide decision-making information obtained by automated means involving profiling and /or individual assessments. Depending on the use of such information in making the final decision by a public service, EUIs may fall within the scope of application of Article 24 of the Regulation, so they need to ensure that individual safeguards are guaranteed, including at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

In managing AI decision-making tools, EUIs must consider carefully how to ensure that the right to obtain human intervention is properly implemented. This is of paramount importance in case EUIs deploy autonomous AI agents that can perform tasks and make decisions without human intervention or guidance.

EUIs must be very attentive to the weight that the information provided by the system has in the final steps of the decision-making procedure, and whether it has a decisive influence on the final decision taken by the controller. It is important to recognise the unique risks and potential harms of generative AI systems in the context of automated decision-making, particularly on vulnerable populations and children[49].

Where generative AI systems are planned to support decision-making procedures, EUIs must consider carefully whether to put them into operation if their use raises questions about their lawfulness or their potential of being unfair, unethical or discriminatory decisions.

[48] Article 24 of the Regulation.
[49] Global Privacy Assembly (GPA), 20 October 2023, Resolution on Generative Artificial Intelligence Systems, available at https://www.edps.europa.eu/system/files/2023-10/edps-gpa-resolution-on-generative-ai-systems_en.pdf

→ EUI-X is considering using an AI system for the initial screening and filtering of job applications. Service provider C has offered a generative AI system that performs an analysis of the formal requirements and an automated assessment of the applications, providing scores and suggestions on which candidates to interview in the next phase. Having consulted the documentation on the model, including the available measures on statistical accuracy (measures on precision and sensitivity of the model) and in view of the possible presence of bias in the model, EUI-X has decided that it will not use the system at least until there are clear indications that the risk of bias has been eliminated and the measures on precision improve, to the analysis of formal requirements.

In any case, if such system is considered as 'fit for purpose' (i.e. candidates' screening) and compliant with all regulations applicable to the EUI, the EUI should be able to demonstrate that it can validly rely on one of the exceptions under Article 24(2) of the Regulation; that the EUI has implemented suitable measures to safeguard individuals' rights, including the right to obtain human intervention by the EUI, to express her or his point of view and to contest the decision (e.g., non-eligibility).

Information must be provided by the EUI, in accordance with Articles 15(2)(f) and 16(2)(f) of the Regulation, when the data is collected directly or indirectly from the individual respectively, about the logic involved by the AI system, as well as on the envisaged consequences of such processing for the individual. A DPIA must also be carried out prior to the deployment of the AI system by the EUI.

The EUI-X may decide to use, instead of a generative AI system, a 'simpler' online automated tool for the screening of job applications (for instance, an IT tool checking automatically the number of years of professional experience or of education).

13. How can fair processing be ensured and avoid bias when using generative AI systems?

In general, artificial intelligence solutions tend to magnify existing human biases and possibly incorporate new ones, which can create new ethical challenges and legal compliance risks. Biases can arise at any stage of the development of a generative AI system through the training of datasets, the algorithms or through the people who develop or use the system. Biases in generative AI systems can lead to significant adverse consequences for individuals' fundamental rights and freedoms, including unfair processing and discrimination, particularly in areas such as human resource management, public health medical care and provision of social services, scientific and engineering practices, political and cultural processes, the financial sector, environment and ecosystems as well as public administration.

Main sources of bias can come, among others, from existing patterns in the training data, lack of information (total or partial) on the affected population, inclusion or omission of variables and data that should not or should be part of the datasets, methodological errors or even bias that are introduced through monitoring.

It is essential that the datasets used to create and train models ensure an adequate and fair representation of the real world - without bias that can increase the potential harm for individuals or collectives not well represented in the training datasets - while also implementing accountability and oversight mechanisms that allow for continuous monitoring to prevent the occurrence of biases that have an effect on individuals, as well as to correct those behaviours. This includes ensuring that processing activities are traceable and auditable[50] and that EUIs keep supportive documentation. In that regard, it is important that EUIs adopt and implement technical documentation models, which can be of particular importance when the models use several datasets and / or combine different data sources.

Generative AI systems providers try to detect and mitigate bias in their systems. However, EUIs know best their business case and should test and regularly monitor if the system output is biased by using input data tailored to their business needs.

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

The application of procedures and best practices for bias minimisation and mitigation should be a priority in all stages of the lifecycle of generative AI systems, to ensure fair processing and to avoid discriminatory practices. For this, there is a need for oversight and understanding of how the algorithms work and the data used for training the model.

[50] The audit of training data can help to detect bias and other problematic issues by studying how the training data is collected, labelled, curated and annotated. The quality of the audit and its results depends on the access to the relevant information, including the training datasets, documentation and implementation details.

→ EU-X is assessing the existence of sampling bias on the automated speech recognition system. Translation services have reported significantly higher word error rates for some speakers than for others. It seems that the system has difficulties to cope with some English accents. After consulting with the developer, it has concluded that there is a deficit in the training data for certain accents, notably when the speakers are not native. Because itis systematic, EU-X is considering refining the model using its own-generated datasets.

14. What about the exercise of individual rights?

Individuals whose personal data is processed at any stage of a generative AI system's lifecycle, from development to deployment, have rights over their personal data. These rights include the right to be informed, access, erasure, rectification, objection, restriction, data portability, and withdrawal of consent. EUIs developing or deploying the generative AI systems shall implement and maintain effective procedures to enable individuals to exercise these rights whenever personal data is processed. When receiving an individual rights request in the context of generative AI, EUIs should identify whether the request concerns i) training data, ii) post-training data, (including fine-tuning data and data from reinforcement learning from human feedback) iii) user inputs (prompts), and/or) iv) outputs of the generative AI model.

The unique characteristics of the generative AI systems present significant challenges to the exercise of individual rights.[51] Particularly in the context of requests related to training or post-training data, it may be challenging to identify the individual that the training data concerns. First of all, this is because generative AI models, like LLMs, are often trained on diverse and vast datasets from multiple sources. This makes it extremely difficult to determine whether a specific individual's personal data was included in the training data set and subsequently, to trace it. It is also complex to manage personal data generated through inference. In particular, generative AI systems create new inferred information based on learned patterns. Due to their opaque nature, it is difficult to trace the new information back to a specific individual.

Additionally, training data is usually processed through various techniques to make it more suitable for machine learning. For instance, an individual's browsing history may be transformed into a profile that summarises the most visited categories of websites. In such cases, the controller has to take reasonable steps to verify the identity of the individual and respond to their request. If the controller demonstrates that it is unable to identify the individual, Articles 17-22 of the Regulation do not apply. In accordance with Article 12(1) of the Regulation, the controller does not have to store additional information for the sole purpose of handling individual rights requests. However, if the individual provides additional information enabling their identification for the purpose of exercising their rights, in accordance with Article 12(2) of the Regulation, the controller has to handle the request. Finally, as a matter of good practice, the controller should inform the data subject of any additional information that may be provided for their identification.

Challenges may also be encountered with regard to the exercise of the right to erasure or rectification. EUIs could be concerned that erasing or rectifying an individual's data from the training dataset could affect the model's performance. However, removing or changing a data point from a massive training dataset will unlikely have an impact on the generative AI model's ability to fulfil its training purposes, given that ample data from other individuals are still processed. The primary challenge would be more related with the technical and computational difficulties of removing the concerned data. To respond to the request for erasure or rectification, a controller does not need to erase or alter all machine learning models

[51] Chapter III of the Regulation.

based on the data concerned for erasure or rectification, unless the model itself contains such data or can be used to infer it.

Individual requests may also concern outputs from the generative AI model (new content that the AI model creates based on a user's prompt[52]). Such requests are, in principle, more common than requests related to training data, as AI generated outputs may have a direct impact on the individual's rights. For instance, an individual may request the rectification of inaccurate personal data generated about them. The controller has to take steps to verify whether the personal data processed are inaccurate and if the right to rectification is applicable, to take appropriate action. If deletion or rectification of personal data affects the model itself, the model may have to be re-trained.

Keeping a traceable record of the processing of personal data, as well as managing datasets in a way that allows traceability of their use, may support the 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.

EUIs, as data controllers, are responsible for and accountable for implementing appropriate technical, organisational and procedural measures to ensure the effective exercise of individual rights. Those measures should be designed and implemented from the early stages of the lifecycle of the system, allowing for detailed recording and traceability of processing activities.

→ EU-X has included in the data protection notice for the chatbot a reference to the exercise of individual rights, including access, rectification, erasure, objection and restriction of processing in accordance with the Regulation. The notice includes contact details of the controller and EU-X DPO, as well as a reference to the possibility of lodging a complaint with the EDPS. Following a request of access from an individual concerning the content of his conversations with the chatbot, EU-X replied, after carrying out the relevant checks, that no content is preserved from the said conversations beyond the established retention period, 30 days. The conversations, as indicated to the individual, has not been used to train the chatbot model.

---

[52] Generally speaking, it should be noted that responsibility for the output depends on its origin. The user of the system may be responsible, if the personal data in the output is derived from statistical inferences based on personal data provided by the user in their prompts. On the other hand, the provider is responsible if the output stems from the model's original training datasets.

→ EUI-X deploys a generative AI chatbot to assist employees with internal HR matters. An individual submits to EUI-X a request for rectification of their personal data, claiming that the generative AI system produces outputs that entail false information about their employment history. EUI-X investigates the rectification request, examines the training data, and initially concludes that it is unable to find the internal data source that could lead to the incorrect information linked with the specific individual. The error could have originated from inaccurate data taken from outdated online public sources, or it could be an inference where the model incorrectly connected the individual's employment history with another individual. EUI-X informed the individual that the error does not exist in their internal databases, that they cannot trace the false information in the system, and that they are therefore unable to rectify the requested data point. The individual follows up by sending to EUI-X a full script of the chat with the generative AI, that includes the specific prompts that they used and the chabot's false information about their employment history with their name. EUI-X is able now to investigate the model's behaviour and treats the rectification request as valid. They implement a correction by feeding the model with a new instruction that will prevent the system from repeating the same false output about the specific individual in the future.

15. What about data security?

The use of generative AI systems can amplify existing security risks or create new ones, including bringing about new sources and transmission channels of systemic risks in the case of widely used models. Compared to traditional systems, generative AI specific security risks may derive from unreliable training data, the complexity of the systems, opacity, problems to carry out proper testing, vulnerabilities in the system safeguards etc. The limited offer of models in critical sectors for the provision of public services such as health can amplify the impact of vulnerabilities in these systems. The Regulation requires EUIs to implement appropriate technical and organisational measures to ensure a level of security[53] appropriate to the risk for the rights and freedoms of natural persons.

Controllers 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[54], prompt injection[55], jailbreaks[56] - in a way that facilitates continuous monitoring and assessment of their effectiveness. Controllers are advised to only use datasets provided by trusted sources and carry out regularly verification and validation procedures, including for in-house datasets. EUIs should train their staff on how to identify and deal with security risks linked to the use of generative AI systems. As risks evolve quickly, regular monitoring and updates of the risk assessment are needed. In the same way, as the modalities of attacks can change, proper access to advanced knowledge and expertise must be ensured. A possible way to deal with unknown risks is to use "red teaming[57]" techniques to try to find and expose vulnerabilities.

When using Retrieval Augmented Generation[58] with generative AI systems, it is necessary to test that the generative AI system is not leaking personal data that might be present in the system's knowledge base.

The lack of information on the security risks linked to the use of generative AI systems and how they may evolve requires EUIs to exercise extreme caution and carry out detailed planning of all aspects related to IT security, including continuous monitoring and specialised technical support. EUIs must be aware of the risks derived from attacks by malicious third parties and the available tools to mitigate them.

[53] Article 33 of the Regulation.
[54] A Model inversion attacks takes place when an attacker extracts information from it through reverse-engineering.
[55] Malicious actors use prompt injection attacks to introduce malicious instructions as if they were harmless.
[56] Malicious actors use jailbreaking techniques to disregard the model safeguards.
[57] A red team uses attacking techniques aiming at finding vulnerabilities in the system.
[58] AI systems upon which a Large Language Model bases its answers in a knowledge base prepared by the generative AI system owner (e.g. an EUI) with internal sources and not in the knowledge stored by the LLM itself.

→ EU-X, following a security assessment, has decided to implement the ASR system on premises, instead of using the API services provided for the developer of the model. EU-X will train its IT staff on the use and further development of the system, in close cooperation with the provider. This may include training on how to refine the model. In addition, EU-X will get the services of an external auditor to verify the proper implementation of the system, including on security.

16. Do you want to know more?

- **EDPS and EDPB work on AI**
    - EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, adopted on 17 December 2024
    - 45th Closed Session of the Global Privacy Assembly - Resolution on Generative Artificial Intelligence Systems - 20 October 2023
    - EDPS TechDispatch #2/2023 - Explainable Artificial Intelligence
    - EDPS at work: data protection and AI (includes links to several documents published by the EDPS alone or in cooperation with other authorities)
    - EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)
    - EDPS Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments
    - Large Language Models (EDPS website, part of the EDPS "TechSonar" report 2023-2024)
- **Other relevant documents**
    - Article 29 Data Protection Working Party: Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01), 3 October 2017
    - CNIL: Artificial Intelligence
    - Data Protection Authority of Belgium: Artificial Intelligence Systems and the GDPR, December 2024
    - German Data Protection Conference: Artificial Intelligence and Data Protection, 6 May 2024
    - Autoriteit Persoonsgegevens (Dutch Data Protection Authority): Guide to scraping by individuals and private organisations, 2 April 2025
    - Information Commissioner's Office: ICO consultation series on generative AI and data protection, 2024

- Information Commissioner's Office: Guidance on AI and data protection, March 2023
- Spanish Data Protection Authority: Artificial Intelligence: accuracy principle in the processing activity
- Italian Data Protection Authority: Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale – September 2023 (Italian)
- The Hamburg Commissioner for Data Protection and Freedom of Information: Checklist for the use of LLM-based chatbots - 15/11/2023
- AI Security Concerns in a nutshell (DE Federal Office for Information Security, March 2023)
- Multilayer Framework for Good Cybersecurity Practices for AI (ENISA, June 2023)
- Ethics Guidelines for Trustworthy AI (EC High-Level Expert Group on AI, 2019)
- Living Guidelines on the responsible use of Generative AI in research (ERA Forum Stakeholders' document, March 2024)
- OECD AI Incidents and Hazards Monitor (AIM)
- OECD Catalogue or tools and metrics for trustworthy AI